Pricing, Risk and Solvency Requirements: An Analysis of Investment Guarantees Embedded in Individual Pension Products - A Regime Switching Approach

Inauguraldissertation zur Erlangung des akademischen Grades

eines Doktors der Wirtschaftswissenschaften

der Universität Mannheim

Wojtek Piaskowski

vorgelegt im Herbstsemester 2009

Dekan: Prof. Dr. Hans H. Bauer

Referent: Prof. Dr. Peter Albrecht

Korreferent: Prof. Dr. Hans-Jochen Bartels

Tag der mündlichen Prüfung: 30. September 2009

Contents

List of Figures

List of Tables

Acknowledgments

This thesis has been written during my doctoral study at the Chair of Risk Theory, Portfolio Management and Insurance at the University of Mannheim and my work at the Centre for European Economic Research (ZEW) and has been co-financed with grants from the University of Mannheim and the German Academic Exchange Service (DAAD).

First of all, I owe special acknowledgment to my supervisor, Prof. Dr. Peter Albrecht, who sparked my interest in embedded guarantees. Thank you very much for sharing your knowledge with me and for providing numerous valuable comments. This dissertation would not be what it has become without your support. I am also indebted to Prof. Dr. Hans-Jochen Bartels for answering my questions regarding the Option Pricing Theory. I owe gratitude to Prof. Dr. Raimond Maurer for his support during my application for the DAAD scholarship as well as to Dr. Michael Schröder for granting me academic freedom during my time at the ZEW.

The years of my PhD study would not have been the same without my colleagues Timo Klett, Dr. Jochen Mandl, and Martin Scheffel, and without our discussions, lunch times, coffee breaks, and shared exasperations which accompany any academic work. I dedicate special thanks to Traudel Walther for your smile and cheerfulness, and for being the good spirit of Professor Albrecht's chair. I am greatly indebted to Maja Wujcik, the nicest mathematician I have ever known, for reading the manuscript and for the many useful comments she provided. I would further like to express my gratitudes go to Jakub Kaczkarczyk for his excellent research assistance and to Heidi, as well as Hela Hellerich for careful proof-reading of this thesis.

Furthermore, I would like to use this opportunity to thank Judith Abbé, Katharina Lörch, Ingo Mistele, and Monika Schenk. Thank you very much for the time we spent together not talking about theorems, estimation algorithms, statistical tests, programming bugs, or martingale measures. Your long-lasting friendship means a lot to me. I am indebted to Piotr Nowak for your wise advice and encouragement in all of life's circumstances. A special thanks for your Open Door Theory.

I am, in particular, deeply grateful to my family for your love, care, support, and prayers in all these years. I thank my parents Elżbieta and Marian Piaskowski for starting my education. Every following year of my PhD study, I have realized more and more how important it was for my academic development. I owe you roots to know who I am, wings to unfold, and a warm nest to which it is always a great pleasure to come back. My special gratitude goes to Agatka Rakow for all the hours happily spent together since our childhood. You are the best sister ever!

Most of all, however, I would like to thank to my savior Jesus Christ *inter alia* for giving me the opportunity and the ability to write this thesis.

Chapter 1

Introduction

The demographic aging process has turned into a serious problem for numerous societies over the past decades. As a result, retirement products, such as individual saving plans, in which contributions are invested in the financial markets, are becoming increasingly popular. By purchasing such products, people hope to participate in the return chances of the financial markets. However, the risk that the investment fails to cover the planned life standard in the old age can also be considerable. One of the main goals of saving for retirement is to make up for the loss of income in old age. Achieving this goal can easily be threatened if the stock market crashes shortly before retirement, thereby destroying a great amount of ones savings. In such a scenario, a retiree's future financial well-being is seriously jeopardized. In order to protect the contribution payers against this risk, the modern pension frames include embedded guarantees.

The design of investment guarantees varies according to country. Generally speaking, we can distinguish between deterministic and stochastic guarantees on the one hand, and between maturity guarantees and multi-period guarantees on the other hand. A provider of a deterministic guarantee assures his clients that the investment portfolio will yield at least an *ex-ante* defined rate of return. Deterministic guarantees are embedded, for instance, in the German state-subsidized retirement investment plans established in 2001 by the Retirement Savings Act (Altersvermögensgesetz), called “Riester contracts”. In these contracts, the provider guarantees that, at maturity, the contributor will receive at least the sum of the premiums paid throughout the duration of the contract. This corresponds to a guaranteed return of 0% p.a. The provider of a stochastic guarantee, in contrast, has to achieve a return of an *ex-ante* defined stochastic index. Stochastic guarantees are used, e.g., in Brazil where there exist products with a guaranteed return of 6% above inflation. A special case of a stochastic guarantee is when the index is market based. For instance, in the Polish second pillar pension saving accounts, the index is related to the weighted average return of all funds operating in the second pillar, computed on a three-year basis. At the end of each quarter the savings account has to yield the lower of the two indexes: half of the weighted average return or the weighted average return minus 4%. An overview of current guarantee schemes can be found in Table 1.1.

In a maturity guarantee, the return of the personal account has to be higher or equal to the guaranteed return only at the expiration date of the contract. Otherwise, the guarantee provider is obliged to cover the difference between the guaranteed and the realized return. However, if at any point in time before the contract expires, the cumulated return would be lower than the cumulated guarantee return, but exceeded the guaranteed return at the time of contract expiration, the provider does not have to pay anything. An example for such a guarantee is the aforementioned German Riester contract. In the case of a multi-period guarantee, the guaranteed return has to be realized at the end of each sub-period. One example for this is the Polish

Table 1.1: Overview of guarantee schemes

Country	**Guarantee**
Argentina	$\min[70\% M_{av}; M_{av} - 2\%]$
Belgium	3.25% on employers' contributions
	3.75% on employees' conttribution
Brazil	$I + 6\%$
Chile	$\min[50\% M_{av}; M_{av} - 2\%]$
Colombia	$\min[M_{av}; r_{RP}]$
El Salvador	$\min[50\% M_{av}; M_{av} - 2\%]$
Germany	0%
Italy	DB
Japan	0%
New Zealand	4%
Malaysia	2.5%
Peru	$\min[50\% M_{av}; M_{av} - 2\%]$
Poland	$\min[50\% M_{av}; M_{av} - 4\%]$
United Kingdom	DB
Uruguay	2% (public pension plans)
	$\min[50\% M_{av}; M_{av} - 2\%]$ (private pension plans)

Note:
This table shows investment guarantees embedded in individual pension accounts using the example of certain chosen countries. M_{av} denotes the average rate of return of all pension plans in this market segment, r_{RP} – the rate of return of a reference portfolio, I - the inflation rate, DB – the benefit of a defined benefit plan, respectively.
Sources: Fischer (1998, p. 3-4), Pennacchi (1999, p. 222, 224-225), Sin (2002, p. 13), Turner and Rajnes (2003, p. 255-259), Lachance and Mitchell (2003, p. 160)

second pillar pension funds.

The aim of this thesis is to analyze deterministic maturity guarantees embedded in individual pension products, regardless of the legal definition, i.e., regardless of whether these products are obligatory or voluntary, whether they are provided by the state or by private pension companies. In particular, we focus on four issues: guarantee pricing, shortfall risk analysis, solvency requirements, and expected return. Other risk sources, such as mortality, early contract cancelation, or problems as hedging are left for further research.

Even though there is a comprehensive literature dealing with guarantees embedded in both unit-linked and with-profit life insurance products, beginning with the seminal work of Brennan and Schwartz (1976, 1979a) and Boyle and Schwartz (1977), who applied the option pricing theory of Black and Scholes (1973) and Merton (1973) to price equity-linked life insurance with asset value guarantees. However, literature addressing guarantees embedded in pension plans is still rare. Fischer (1999) proposed a lattice model to price guarantees embedded in Colombian pension plans. In this guarantee the owner of the pension plan has the right to receive the lower of two values: the average return of pension plans on the Colombian market or the return of a benchmark portfolio. Pennacchi (1999) priced investment guarantees in both public and private pension funds available in Uruguay. In the first case, a return of 2% p.a. is guaranteed, in the second case, the guarantee is the lower of the two values: half of the average return of all pensions funds available in this market or the average return of pension plans minus 2%. Bacinello (1997) priced an option to switch from the defined-benefit retirement plan to the defined-contribution retirement plan embedded in the Italian pension plans. Bacinello (2000) extended the model with the option of switching from the defined-contribution to the defined-benefit system. Lachance and Mitchell (2003) studied a similar problem. They price an option to return from the defined-contribution to defined-benefit system as, introduced in Florida, USA.

Gründl, Nietert, and Schmeiser (2004) priced guarantees embedded in German Riester contracts. Kling, Russ, and Schmeiser (2006) extended this model with the possibility of canceling the contract: the purchaser of the Riester product can cancel the contract and retain the guarantee. In an extreme case, he can pay the first contribution and then cancel the contract. This procedure can be repeated in each period, in order to maximize the value of the guarantee.

Maurer and Schlag (2003) and Gründl, Nietert, and Schmeiser (2004) discussed the shortfall risk associated with the Riester contract. They apply lower partial moments and the mean excess loss to quantify this risk. Furthermore, they analyze solvency rules and find them inadequate.

All of above mentioned papers (with the exception of Fischer (1999) who uses a binomial distribution) assume that the log-returns of the risky portfolio, which backs the guarantee, are normally distributed

$$y_t = \left(\mu - \frac{1}{2}\sigma^2\right) + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2). \tag{1.1}$$

However, a simple long-run observation of the time series shows that such an assumption is questionable. One simply has to recall e.g. the (drastic) collapse of the stock prices during the oil crises of the 1970s, the black October of 1987, the displosion of the New Economy bubble in 2000, or the current subprime crisis. A brief look at the development of the German stock market (see Figure 1.1) shows that in the period from 2000 to 2003 the drift was negative, while in the period from 1995 to 1999, the drift was positive.

This shows that the parameters for the bear and the bull market could be estimated separately. The model (1.1) would look then as follows

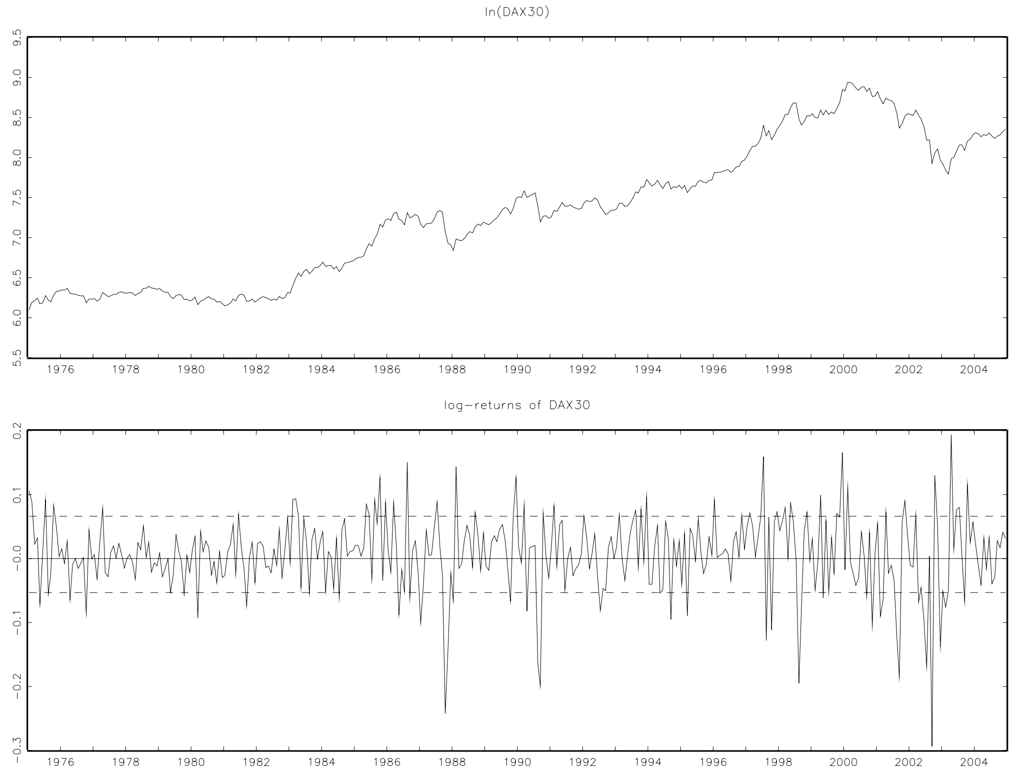
$$y_t = \begin{cases} \left(\mu_{bear} - \frac{1}{2}\sigma^2_{bear}\right) + \varepsilon_{t,bear}, & \varepsilon_{t,bear} \sim N(0, \sigma^2_{bear}), \quad \text{for } t \in BEAR \\ \left(\mu_{bull} - \frac{1}{2}\sigma^2_{bull}\right) + \varepsilon_{t,bull}, & \varepsilon_{t,bull} \sim N(0, \sigma^2_{bull}), \quad \text{for } t \in BULL \end{cases} \tag{1.2}$$

with $BEAR = \{t : \ \mu(t) < 0\}$ and $BULL = \{t : \ \mu(t) > 0\}$.

$$y_t = \left(\mu(Z_t = j) - \frac{1}{2}\sigma^2(Z_t = j)\right) + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2(Z_t = j)), \quad j = 1, \ldots, K \tag{1.3}$$

$$p_{ji} = \Pr[Z_t = j | Z_{t-1} = i], \quad 0 \leq p_{ji} \leq 1, \quad \forall i, j = 1, \ldots, K \tag{1.4}$$

Figure 1.1: Monthly DAX30 and its log-returns: 1975-2004



Note:
The top panel depicts the DAX30 on a logarithmic scale.The bottom panel presents the log-returns of DAX30. The dashed lines show the one standard deviation from the mean bound.

where $\mu(Z_t = j)$ and $\sigma^2(Z_t = j)$ denote the drift and the diffusion parameter, respectively, depending on the j-th regime ($j = 1, \ldots, K$), ε_t denotes the innovation and p_{ji} denotes the transition probability to the j-th regime at time t if the process was in the i-th regime in the previous period.

In this thesis, the Markov switching model will be used to describe the stochastic behavior of returns of risky asset portfolios which back the investment guarantee. We have chosen this model as it is capable to capture the widely observed non-normality (excess kurtosis) of financial return series [1]

[1] See for example Haas et al. (2004).

in contrast to the commonly used arithmetic Brownian motion, and is intuitively easy to interpret. Furthermore, it is a very parsimonious model with respect to the number of parameters used, and, at the same time, allows for many different specifications. To the best of our knowledge, it represents the first attempt at using a regime switching model (which especially implies stochastic volatility model) in the literature dealing with guarantees embedded in pension products.[2]

Using the Markov switching model, we have to address the fact that this model implies the incompleteness of the underlying financial market. This affects option pricing, since, in this situation, in an arbitrage-free market, the equivalent martingale measures is not unique. The general consensus is that the market "chooses" the "right" martingale measure. However, since the guarantees discussed in this thesis are not traded on the market, their prices cannot be observed. Thus, the guarantee provider has to make a suitable choice regarding the equivalent probability measure based, among others, on the grade of his risk aversion. We decided to choose the Esscher measure, which is well known in the actuarial science. Reasons for this choice are fourfold: (1) The process under the Esscher martingale measure $\mathcal{Q}$ remains in the same class of models as the process under the real-word probability measure $\mathcal{P}$. (2) The option price reduces to the well-known Black and Scholes (1973) formula for the case of one switching regime (i.e. $K = 1$). (3) The Esscher transform approach is conform with maximizing the expected utility with the constant risk aversion utility function $u(x) = \frac{x^\gamma}{\gamma}$ $(0 < \gamma < 1)$. (4) The Esscher probability measure allows to price the uncertainty whether the market is in a stable or a turbulent phase.

[2]We used this approach in our previous paper, see Piaskowski (2005). This thesis is an extension of that work.

Apart from pricing, risk management and solvency requirements are very important issues. In this case, we follow the approach proposed by Maurer and Schlag (2003), who use shortfall risk measures to quantify the risk associated with investment guarantees. Additionally, we propose using the mean excess loss to quantify solvency requirements for pension plans embedding investment guarantees.

This thesis is organized as follows: Chapter 2 introduces stochastic models used in finance with particular focus on the geometric Brownian motion with Markov switching, which will be used in this thesis to describe the behavior of the market prices of risky assets. Additionally, this Chapter shows how to estimate the parameters of this model. Chapter 3 analyzes whether German time series can be described with Markov switching models. As Markov switching models violate the assumptions of standard tests for nested models such as the likelihood ratio test or the Wald test, these tests cannot be used. Instead, we use tests developed by Hamilton (1996) and Garcia (1998). Even though these tests have been known for over a decade, most authors have used the standard tests, due to their computational simplicity. The main findings of this Chapter is that German financial time series are better described by Markov switching models than by commonly used geometric Brownian motion and mean-reverting models. Chapter 4 shows how to price an investment guarantee when the portfolio value follows the geometric Brownian motion with Markov switching. The option pricing model is based on the Esscher transform martingale measure developed by Gerber and Shiu (1994b) and Webb (2003). Chapter 5 analyzes the shortfall risk of the guarantee with respect to the shortfall risk measures: the shortfall probability, the shortfall expected value, the shortfall standard deviation, the mean expected loss, and the conditional shortfall standard deviation. Additionally, we propose using the mean excess loss to quantify solvency capital requirements for investment guarantees. Chapter 6 sums up the main results and provides a brief outlook for further research.

Part I

Stochastic Model

Chapter 2

Stochastic models in finance

This Chapter introduces the basic Markov switching model which is fundamental for this dissertation. Before this some stylized facts regarding financial time series will be presented in Section 2.1 and some mathematical notation and definitions used in this thesis will be introduced in Section 2.2. Section 2.3 defines Markov chains and discusses features of Markov chains which will be used throughout this dissertation. Section 2.4 discusses some common stochastic models used in finance. Section 2.6 introduces finite mixtures of normal distributions. Section 2.7 defines the Markov switching model and gives some examples. Section 2.8 shows how to estimate a switching regime model via the EM algorithm. Section 2.9 concludes the Chapter by performing an empirical analysis on the basis of the Markov switching models.

Sections 2.2-2.6 are rather technical such that the reader can skip them and go directly to definition of the Markov switching model if he is not interested in the technical details.

Table 2.1: Most extreme log-returns of DAX30 in 1975-2004

Date	Return	Deviation (as $\frac{r-\mu}{\sigma}$)	Probability ($\times 10^{-6}$)	Frequency (in years)
28/09/2001	-0.1859	-3.2328	612.81	136
31/08/1998	-0.1949	-3.3840	357.16	233
28/09/1990	-0.1994	-3.4606	269.46	309
30/10/1987	-0.2423	-4.1807	14.53	5,734
30/09/2002	-0.2933	-5.0381	0.24	354,503

Note:
The table depicts the five most extreme events. The third column shows by how many standard deviations the returns are departed from the mean. The fourth column shows the probability of such an event, if the normal density held. The fifth column shows how often (in years) such an event would occurred, if the normal distribution held.

2.1 Stylized facts about financial time series

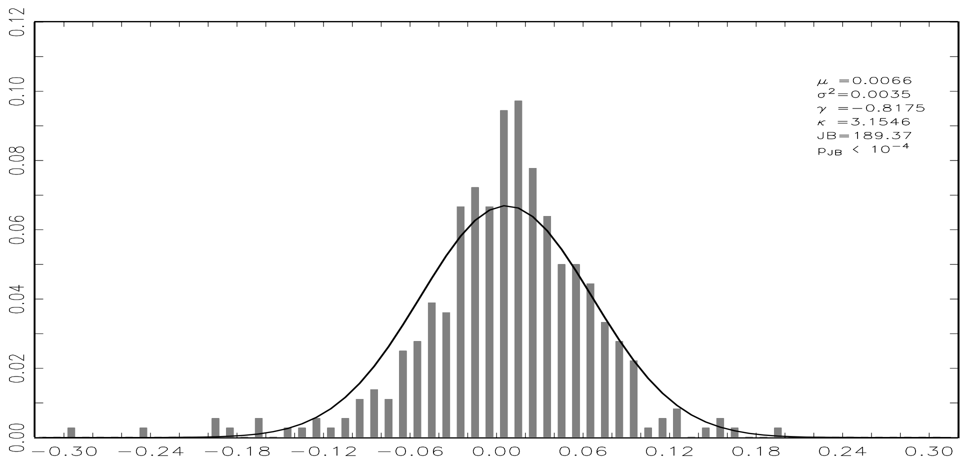
2.1.1 Asymmetry and leptokurtosis

The standard assumption used in finance is that the returns of financial time series are independent, identical normals. However, there is some empirical evidence contradicting this assumption. The first to address this issue was Mandelbrot (1963), who studied daily and monthly prices of cotton traded in New York from 1816 to 1940. He noticed that extreme events occur much more frequently than is allowed by the normal distribution. In the literature, this phenomenon is called *fat tails* or *leptokurtosis*.[1] Another prevalent observed phenomenon is the skewness of the financial time series.

Figure 2.1 presents the histogram of the monthly log-returns of DAX30 - the German blue chip index - from January 1975 to December 2004. In this period, the mean monthly return was equal to 0.0066 and its variance was equal to 0.0035. The solid line represents the density of normal distribution

[1] The distribution of a random variable is called leptocurtic if it has a positive excess kurtosis.

Figure 2.1: Normal distribution vs. monthly log-returns of DAX30 (1975-2004)



Note:
The figure shows the histogram of the monthly log-returns of DAX30. In the top right corner the moments of the time series (μ - mean, σ^2 - variance, γ - skewness, and κ - excess kurtosis), the Jarque-Bera test statistic (JB) and its p value (p_{JB}) are presented. The solid line represents the normal density with parameters estimated form this time series (μ and σ^2). It is straightforward to see that the log-returns of DAX30 are left-skewed and leptokurtic.

with these parameters. Evidently the log-returns of DAX30 are not normally distributed. 52% of the probability mass is located to the right of the mean. Moreover, the left tail is extraordinarily thick. Table 2.1 shows the five most extreme events which occurred in these 30 years. If the log-returns were normally distributed, the return of a figure equal or less than -18.59% in a month would be lower then 0.6‰, which means that it occurred once every 136 years. However, this event has actually occured five times in a period of 30 years. The most negative monthly return occurred in September 2002 and amounted to -29.33%. However, if the normal distribution held, the loss of almost 30% in a month would occur once in 254,503 years (see Table 2.1). The observation that log-returns of DAX30 are not normally distributed can be confirmed by a glance at the skewness and kurtosis. The empirical skewness of DAX30 returns amounts to -0.8175 and their excess kurtosis to 3.1546. According to the normal distribution, both values should equal zero. Thus, the time series is left skewed and leptokurtic. As the Jarque-Bera test statistic equals 189.37 and is significant at all commonly used confidence levels, the assumption of normal distribution can thus be rejected (see Figure 2.1).

2.1.2 Conditional heteroskedasticity

Another observation made by Mandelbrot (1963) concerned the so-called *volatility clusters*. This means that the volatility of time series is persistent or in other words: large changes in the returns are followed by large changes, low changes are followed by low changes. Clusters in the volatility are easily discerned by studying the DAX30 returns plotted on the time line (see bottom panel of Figure 1.1). Between 1976 and 2004, there were periods of small amplitude i.e. from 1975 to 1985, when the German stock market stagnated, from 1988 to 1989, and from 1991 to 1996, during a period a rapid growth phase, in 2000 - at the begin of the New Economy crash, and in 2004, with the market rebounding from the crash. This stands in contrast to the following periods which were characterized by a high amplitude of stock returns: in 1975, due to the increase the stock prices after the OPEC oil crisis; from 1986 to 1987, as a result of the market being in a turbulent phase which ended with the black October of 1987; in 1990, when prices fell rapidly; from 1997 to 1999, when the dot-com bubble rose; from 2001 to 2002, during the New Economy crash; and in 2003 - the first year of the subsequent growth phase. As one can see the current volatility clearly depends on the past volatility. This phenomenon is also called *conditional heteroskedasticity*. A lot of research has addressed this problem, see for instance the literature on GARCH models beveloped by Engle (1982) and Bollerslev (1986).

2.1.3 Leverage effect

Black (1976) found that empirical data feature a negative correlation between the returns and volatility, i.e. that in contrast to the positive returns, the negative returns are followed by an increase in volatility. Black (1976) called this phenomena a leverage effect. This stylized fact was addressed by several econometric models. E.g., Nelson (1991) introduced the E-GARCH and Zakoïan (1994) and Glosten, Jagannathan, and Runkle (1993) the T-GARCH model.

2.1.4 Non-continuous trading

In financial literature, it is very often assumed that the development of pricees is continuous over time. It would imply, for instance, that the volatility on non-trading days should be equal to the volatility on trading days. French and Roll (1986) have found that the hourly volatility of American stocks was 70 times higher during trading time, compared to when the market was closed.

2.1.5 Mean reversion

Mean reversion implies that the drift of the stochastic process is positive when the last realization of the stochastic process is lower than its long-time mean μ and negative, when the last realization of the stochastic process is higher than its long-time mean μ. This implies that as the time horizon goes to infinity ($t \rightarrow \infty$) a mean-reverting stochastic process y_t converges towards its long-time mean μ ($\mathbb{E}[y_t] \to \mu$) (Brigo and Mercurio 2006, p. 59). This phenomenon can be observed e.g. for interest rates and has the following economic explanation. In a time of high interest rates, the demand for credit declines, due to the high price of lending money. As a result, the rates decline and the borrowers are more willing to obtain funds for new investments. Consequently, the interest rates increase causing once again a slow-down of the economy (Hull 2006, p. 651).

2.2 Mathematical preliminaries

Notation 2.1 *Throughout the entire thesis, the following notation will be used:*

1. $\boldsymbol{I}$ – *the identity matrix, i.e. a square matrix with ones on the diagonal and zeros on all other entries,*
2. $\boldsymbol{\iota}_i$ – *the i-th column vector of the identity matrix,*
3. $\boldsymbol{E}$ – *a square matrix of ones,*
4. $\mathbf{1}$ – *a column vector of ones,*
5. $\boldsymbol{O}$ – *a square matrix of zeros,*
6. $\mathbf{0}$ – *a column vector of zeros,*
7. $\mathbb{I}_A$ – *an indicator function defined as*

$$\mathbb{I}_A = \begin{cases} 1 & \textit{if } A \in \Omega \\ 0 & \textit{otherwise,} \end{cases}$$

8. $\mathbb{N} = \{0, 1, \dots\}$ – *the set of normal numbers (including* 0*),*

9. $\mathbb{R}$ *– the set of real numbers,*

10. $\Pr_{\mathcal{P}}[A]$ *– a probability of the event* A *with respect to the probability measure* $\mathcal{P}$*, if the index* $\mathcal{P}$ *is suppressed, then the probability measure* $\mathcal{P}$ *is considered to be the "real world" probability measure,*

11. $\mathbb{E}_{\mathcal{P}}[A]$ *– an expected value of the event* A *with respect to the probability measure* $\mathcal{P}$*, if the index* $\mathcal{P}$ *is suppressed, then the probability measure* $\mathcal{P}$ *is considered to be the "real world" probability measure.*

In consideration of the fact that the term "stochastic process" will frequently be referred to, a clarifying definition of this terminus would be appropriate. Before we do so, we provide definitions of σ-algebra, measurable space, probability measure, and probability space.

Definition 2.2 (σ-algebra) *Let* Ω *be a given set and* $\mathcal{F}$ *be a family of subsets of* Ω *with the following properties*

(i) $\emptyset \in \mathcal{F}$,

(ii) $F \in \mathcal{F} \Rightarrow F^C \in \mathcal{F}$, *where* $F^C = \Omega \backslash \mathcal{F}$ *(i.e.* F^C *is the complement of* F *in* Ω*),*

(iii) $A_1, A_2, \dots \in \mathcal{F} \Rightarrow A := \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Then $\mathcal{F}$ *is called a* σ*-algebra* $\mathcal{F}$ *on* Ω *(Øksendal 2003, p. 7).*

Definition 2.3 (Filtration) *Let* Ω *be a given set and let* T *be a fixed positive number. Assume that for each* $t \in [0, T]$ *there exists a* σ*-algebra* $\mathcal{F}_t$ *and that if* $s \leq t$*, then every set in* $\mathcal{F}_s$ *is a* σ*-algebra* $\mathcal{F}_t$ *as well. Then the collection of* σ*-algebras* $\mathcal{F}$ *is called a filtration (Shreve 2004, p. 51).*

Definition 2.4 (Measurable space) *Let Ω be a set and $\mathcal{F}$ be a σ-algebra then the pair $(\Omega, \mathcal{F})$ is called a measurable space (Øksendal 2003, p. 7).*

Definition 2.5 (Probability measure) *Let $(\Omega, \mathcal{F})$ be a measurable space and let $\mathcal{P}$ be a function $\mathcal{P} : \mathcal{F} \to [0, 1]$ such that properties*

(i) $\Pr[\emptyset] = 0$ *and* $\Pr[\Omega] = 1$,

(ii) *If $A_1, A_2, \dots \in \mathcal{F}$ and $\{A_i\}_{i=1}^{\infty}$ is disjoint (i.e. $A_i \cap A_j = \emptyset$ if $i \neq j$) then* $\Pr\left[\bigcup_{i=1}^{\infty} A_i\right] = \sum_{i=1}^{\infty} \Pr[A_i]$

hold. Then $\mathcal{P}$ is called probability measure on the measurable space $(\Omega, \mathcal{F})$ (Øksendal 2003, p. 8).

Definition 2.6 (Probability space) *Let Ω be a given set, let $\mathcal{F}$ be a σ-algebra and let $\mathcal{P}$ be a probability measure, then the triplet $(\Omega, \mathcal{F}, \mathcal{P})$ is called a probability space (Øksendal 2003, p. 8).*

Now we can define a stochastic process.

Definition 2.7 (Stochastic process) *Let X_t be a (n dimensional) random variable, then a parametrized collection of random variables*

$$(X_t)_{t \in \mathcal{T}}$$

on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ assuming values in $\mathbb{R}^n$ is called a stochastic process (Øksendal 2003, p. 10). $\mathcal{T}$ is a set of time points and will be chosen as $\mathcal{T} = \mathbb{R}_+ \cup \{0\}$ in the following.

Definition 2.8 (Measurable function) *Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and let $f : \Omega \to \mathbb{R}^n$. If for all open sets $\mathcal{U} \in \mathbb{R}^n$ preimage $f^{-1}(\mathcal{U}) \in \mathcal{F}$, then function f is called $\mathcal{F}$-measurable (Øksendal 2003, p. 10).*

Definition 2.9 (Adapted process) *Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration on Ω and let $(X_t)_{t\geq 0}$ be a stochastic process. If for each $t \geq 0$ stochastic variable X_t is $\mathcal{F}_t$-measurable, then the stochastic process $(X_t)_{t\geq 0}$ is called $\mathcal{F}_t$-adapted (Musiela and Rutkowski 2007, p. 35)*

2.3 Markov chains

2.3.1 Definition of the Markov chain

Markov chains are a powerful and very commonly used mathematical tool. They can be applied to many fields of economics, such as insurance (birth-death process), finance (random walk), logistics (queuing problem) or econometrics (Markov chains Monte Carlo). In this Section, some properties of discrete Markov chains will be discussed, which are relevant for the Markov switching models defined in Section 2.7. Further discussion on discrete Markov chains can be found in Cox and Miller (1965, Chapter 3), Kijima (1997, Chapters 2-3), Norris (1997, Chapter 1), and Rolski et al. (1999, Chapter 7). For a discussion of continuous time Markov chains, see Cox and Miller (1965, Chapter 4-5), Kijima (1997, Chapter 4,), Norris (1997, Chapter 2-3), and Rolski et al. (1999, Chapter 8). An applications-oriented discussion is given in Anderson (1991).

Before providing a definition of the Markov chain, we will clarify the definitions of the state space, the stochastic matrix and the stochastic vector.

Definition 2.10 (State space) *Let $\mathcal{K}$ be a countable set, then it is called a state space and all elements $k \in \mathcal{K}$ are called states (Norris 1997, p. 1).*

Definition 2.11 (Stochastic matrix) *Let $\boldsymbol{P}$ be a matrix and let p_{ji} be an element of the matrix $\boldsymbol{P}$. If all elements $p_{ji} \geq 0$ and all rows of $\boldsymbol{P}$ sum to* *unity (i.e. $\boldsymbol{P}\mathbf{1} = \mathbf{1}$), then the matrix $\boldsymbol{P}$ is called a stochastic matrix (Cox and Miller 1965, p. 85).*

Since the row elements of the stochastic matrix $\boldsymbol{P}$ sum to unity, they can be interpreted as (in this case: conditional) probabilities.

Definition 2.12 (Transition probability) *Let $(Z_t)_{t\in\mathbb{N}}$ be a stochastic process with the state space $\mathcal{K}$ and let $i, j \in \mathcal{K}$ be two arbitrary states, then the probability $\Pr[Z_t = j|Z_{t-1} = i] = p_{ji}$ is called (one step) transition probability from state i to state j.*

Definition 2.13 (Stochastic vector) *Let $\boldsymbol{p}$ be a column vector and let p_i be an element of the vector $\boldsymbol{p}$. If all elements $p_i \geq 0$ and sum to unity (i.e. $\boldsymbol{p}'\mathbf{1} = 1$), then the vector $\boldsymbol{p}$ is called a stochastic vector or discrete stochastic measure, equivalently.*

By analogy, the elements of the stochastic vector $\boldsymbol{p}$ can be interpreted as (in this case: unconditional) probabilities.

Definition 2.14 (Probability distribution) *Let $(Z_t)_{t\in\mathbb{N}}$ be a stochastic process with the state space $\mathcal{K}$, let $j \in \mathcal{K}$ be an arbitrary state, and let $\Pr[Z_t = j]$ be the probability of being in state j in time t, then the stochastic vector $\boldsymbol{p}_t = (\Pr[Z_t = j])_{j\in\mathcal{K}}$ is a probability distribution. If $t = 0$, then $\Pr[Z_0 = j]$ is called initial probability and $\boldsymbol{p}_0$ is called initial probability distribution.*

Let us now define the Markov chain.

Definition 2.15 (Markov chain) *Let $\mathcal{K}$ be a state space and $(Z_t)_{t\in\mathbb{N}}$ be a stochastic process with the state space $\mathcal{K}$. If*

$$\Pr[Z_t = z_t|Z_{t-1} = z_{t-1}, \ldots, Z_0 = z_0] = \Pr[Z_t = z_t|Z_{t-1} = z_{t-1}] \tag{2.1}$$

for all $(z_t, z_{t-1}, \ldots, z_0) \in \mathcal{K}^{t+1}$ *and for all* $t \geq 2$*, then the stochastic process* $(Z_t)_{t\in\mathbb{N}}$ *is called a Markov chain (of the first order) and the property (2.1) is called Markov property (of the first order) (Rolski et al. 1999, p. 310).*

To determine a unique Markov chain, the initial probability distribution $\boldsymbol{p}_0 = (\Pr[Z_0 = i])_{i\in\mathcal{K}}$ has to be known or defined. Then the probability distribution of the Markov chain in the first period is given by

$$\boldsymbol{p}_1 = \boldsymbol{P}'\boldsymbol{p}_0 \tag{2.2}$$

and so on. The probability of one particular "path" of the Markov chain equals

$$\Pr[Z_t = z_t | Z_{t-1} = z_{t-1}] \times \cdots \times \Pr[Z_1 = z_1 | Z_0 = z_0] \times \Pr[Z_0 = z_0].$$

Note that in the Markov chain, the probability of the occurrence of the event z_t is only dependent on the value taken in the preceding period (i.e., z_{t-1}). This implies that in order to forecast tomorrow's value (i.e. value in t) of the Markov chain, only today's observation is required (i.e. observation in $t-1$). Therefore, it is often stated that the Markov chain has one period memory. This does not mean that the information from previous periods (i.e. $t-1, \ldots, 0$) is "forgotten", but rather that the addition of this information to the information contained in todays observation does not improve the forecast quality.

Hereafter, only homogeneous Markov chains will be considered.

Definition 2.16 (Homogeneous Markov chain) *Let* $(Z_t)_{t\in\mathbb{N}}$ *be a Markov chain. If the transition probabilities*

$$p_{ji} = \Pr[Z_t = j | Z_{t-1} = i], \quad \forall\ i, j \in \mathcal{K}, \ \sum_{j\in\mathcal{K}} p_{ji} = 1 \tag{2.3}$$

are time invariant, then $(Z_t)_{t\in\mathbb{N}}$ *is a homogeneous Markov chain (Rolski et al. 1999, p. 310).*

Example 2.17 *Consider the price process $(S_t)_{t\in\mathbb{N}}$ of a stock with the corresponding rate of return process $(R_t)_{t\in\mathbb{N}}$. Furthermore the price of the stock can rise by u% (state A) or fall by d% (state B). Additionally, it is known that the probability of the price rising tomorrow if it has increased today is $p_A = \Pr[R_{t+1} = A|R_t = A]$ and the probability of the price falling tomorrow if it has decreased today is $p_B = \Pr[R_{t+1} = B|R_t = B]$. Evidently the stock's rate of return process is a Markov chain with the state space $\mathcal{K} = \{A, B\}$ and the transition matrix*

$$\boldsymbol{P}_{K\times K} = \begin{array}{cc} \begin{pmatrix} p_A & 1-p_B \\ 1-p_A & p_B \end{pmatrix} & \begin{array}{c} {\scriptstyle A} \\ {\scriptstyle B} \end{array} \\ \begin{array}{cc} {\scriptstyle A} \qquad & \qquad {\scriptstyle B} \end{array} & \end{array} . \tag{2.4}$$

In analogy to the Markov chain with a one period memory a Markov chain with a memory of n periods can be defined.

Definition 2.18 (Markov chain of the nth order) *Let $\mathcal{K}$ be a state space and $(Z_t)_{t\in\mathbb{N}}$ be a stochastic process which can only take values from the state space $\mathcal{K}$ and $n \geq 1$. If*

$$\Pr[Z_t = z_t|Z_{t-1} = z_{t-1}, \ldots, Z_0 = z_0] = \Pr[Z_t = z_t|Z_{t-1} = z_{t-1}, \ldots, Z_{t-n} = z_{t-n}] \tag{2.5}$$

then the stochastic process Z_t is called a Markov chain of the n-th order or Markov chain with n period memory.

Please note that each Markov chain of the higher order (i.e. $n > 1$) can be reduced to the Markov chain of the first order.

Theorem 2.19 *If the stochastic variable $(Y_t)_{t\in\mathbb{N}}$ is a Markov chain of the order $n > 1$ and the state space $\mathcal{K}$, then it is possible to define a new stochastic process $(Z_t)_{t\in\{n,n+1,\ldots\}}$ such that*

$$Z_t = (Y_{t+n-1}, Y_{t+n-2}, \ldots, Y_t), \tag{2.6}$$

and the new stochastic process is a Markov chain with the state space $\mathcal{K}^n$.

Proof. As the stochastic process $(Y_t)_{t \in \mathbb{N}}$ is a Markov chain of the n-th order, from the property (2.5) it follows

$$\Pr[Y_{t+1} = y_{t+1} | Y_t = y_t, \ldots, Y_0 = y_0] = \\ \Pr[Y_{t+1} = y_{t+1} | Y_t = y_t, \ldots, Y_{t-n+1} = y_{t-n+1}]. \quad (2.7)$$

Note that the chain rule for conditional probabilities states that

$$\begin{aligned} & \Pr[Y_{t+n} = y_{t+n}, \ldots, Y_{t+1} = y_{t+1} | Y_t = y_t, \ldots, Y_0 = y_0] \\ & = \Pr[Y_{t+1} = y_{t+1} | Y_t = y_t, \ldots, Y_0 = y_0] \\ & \times \Pr[Y_{t+2} = y_{t+2} | Y_{t+1} = y_{t+1}, \ldots, Y_0 = y_0] \\ & \times \ldots \\ & \times \Pr[Y_{t+n} = y_{t+n} | Y_{t+n-1} = y_{t+n-1}, \ldots, Y_0 = y_0]. \end{aligned} \quad (2.8)$$

Applying the Markov property (2.7) on both sides of the equation (2.8) yields

$$\Pr[Y_{t+n} = y_{t+n}, \ldots, Y_{t+1} = y_{t+1} | Y_t = y_t, \ldots, Y_0 = y_0] = \\ \Pr[Y_{t+n} = y_{t+n}, \ldots, Y_{t+1} = y_{t+1} | Y_t = y_t, \ldots, Y_{t-n+1} = y_{t-n+1}].$$

From (2.6) it results

$$\Pr[Z_{t+1} = (y_{t+n}, \ldots, y_{t+1}) | Z_t = (y_{t+n-1}, \ldots, y_t), \ldots, Z_0 = (y_{n-1}, \ldots, y_0)] = \\ \Pr[Z_{t+1} = (y_{t+n}, \ldots, y_{t+1}) | Z_t = (y_{t+n-1}, \ldots, y_t)].$$

which is the Markov property of the first order (the proof based on Kijima 1997, p. 3-4, 11-12). ■

Example 2.20 *Consider the stock price process from Example 2.17. This time the probability of the price increasing or decreasing is conditional on* *the price behavior in two previous periods. It is known that if yesterday and today the stock price increased, then the probability of it rising tomorrow is* $p_{AA} = \Pr[R_{t+1} = A|R_t = A, R_{t-1} = A]$. *If the price has risen yesterday and falls today then the probability of it rising tomorrow is* $p_{AB} = \Pr[R_{t+1} = A|R_t = B, R_{t-1} = A]$. *If its price has fallen yesterday and rises today, then with probability* $p_{BA} = \Pr[R_{t+1} = A|R_t = A, R_{t-1} = B]$ *it will increase tomorrow. Eventually, if it decreased yesterday and today, then the probability is* $p_{BB} = \Pr[R_{t+1} = A|R_t = B, R_{t-1} = B]$ *that it will rise tomorrow.*

It is obvious that the stock's rate of return is a Markov chain of the second order with the state space $\mathcal{K} = \{A, B\}$. *One can easily construct a new rate of return process* $(R^\star_t)_{t\in\mathbb{N}}$ *with four states:* $\alpha = (A, A)$ *- the price has increased yesterday and today by* $u\%$, $\beta = (A, B)$ *- the price has increased yesterday by* $u\%$ *and fallen today by* $d\%$, $\gamma = (B, A)$ *- the price has decreased yesterday by* $d\%$ *and increased today by* $u\%$ *and* $\delta = (B, B)$ *- the price has decreased yesterday and today by* $d\%$. *Then the new return process* $R^\star_t$ *is a Markov chain of the first order with the state space* $\mathcal{K}^\star = \{(A, A), (A, B), (B, A), (B, B)\} = \{\alpha, \beta, \gamma, \delta\}$ *and the following transition matrix*

$$\boldsymbol{P}_{K\times K} = \begin{array}{cc} \begin{pmatrix} p_{AA} & 0 & p_{BA} & 0 \\ 1-p_{AA} & 0 & 1-p_{BA} & 0 \\ 0 & p_{AB} & 0 & p_{BB} \\ 0 & 1-p_{AB} & 0 & 1-p_{BB} \end{pmatrix} & \begin{array}{c} \alpha \\ \beta \\ \gamma \\ \delta \end{array} \\ \begin{array}{cccc} \alpha & \beta & \gamma & \delta \end{array} & \end{array} . \tag{2.9}$$

2.3.2 Transition probabilities

Equation (2.3) defines the one period transition probabilities from the initial state i to the target state j. In order to determine the two period (or two step) transition probability $p^{(2)}_{ji} = \Pr[Z_t = j|Z_{t-2} = i]$ it is sufficient to let the stochastic process in time $(t-1)$ "run" through all possible states $k \in \mathcal{K}$

and add all probabilities together

$$\Pr[Z_t = j | Z_{t-2} = i] = \sum_{k \in \mathcal{K}} \Pr[Z_t = j | Z_{t-1} = k] \times \Pr[Z_{t-1} = k | Z_{t-2} = i], \quad \forall\, i, j \in \mathcal{K}. \tag{2.10}$$

For the general case, the following Theorem holds true.

Theorem 2.21 (Chapman-Kolmogorov equation) *Let* $p_{ji}^{(n)} = \Pr[Z_t = j | Z_{t-n} = i]$ *be the* n *step transition probability of a Markov chain* $(Z_t)_{t \in \mathbb{N}}$*, then for* $m, n \in \mathbb{N} \backslash \{0\}$

$$p_{ji}^{(n+m)} = \sum_{k \in \mathcal{K}} p_{jk}^{(n)} p_{ki}^{(m)}. \tag{2.11}$$

The equation (2.11) is called Chapman-Kolmogorov equation (Kijima 1997, p. 14).

Proof. Theorem 2.21 can be proved through induction with the first step (2.10). ∎

Remark 2.22 *The Theorem 2.21 in the matrix notation yields*

$$\boldsymbol{P}^{m+n} = \boldsymbol{P}^m \boldsymbol{P}^n,$$

where $\boldsymbol{P}^n$ *is an* n *period transition probability matrix and* $\boldsymbol{P}^0 = \boldsymbol{I}$*.*

Corollary 2.23 *From equation (2.2) and Theorem 2.21 it follows that the probability distribution in time* $t = n$ *is given by*

$$\boldsymbol{p}_n = (\boldsymbol{P}^n)' \boldsymbol{p}_0. \tag{2.12}$$

2.3.3 Stopping time

Assume that one is interested in a point in time at which an event will occur. This time is called stopping time.

Definition 2.24 (Stopping time) *Let $(Z_t)_{t\in\mathbb{N}}$ be stochastic process with a state space $\mathcal{K}$. A random variable τ is called stopping time of (Z_t) if, for each n, the occurrence of the event $\{\tau \leq n\}$ is determined by $Z_0, \ldots, Z_n$, i.e. there exists a function $f(\cdot)$ such that*

$$\mathbb{I}_{(\tau \leq n)} = f(Z_0, \ldots, Z_n)$$

(Kijima 1997, p. 19).

Example 2.25 *Recall the stock price process from Example 2.17. Assume that a risk-averse agent would like to invest in that stock. One possible coutios investment strategy would be to buy the stock today and hold it as long as the price increases. On the first day on which the price falls, the individual sells the stock. The day on which the stock is sold is called stopping time.*

Three types of stopping times are particularly interesting: the first passage time, the first return time and the sojourn time.

Definition 2.26 (First passage time) *Let $(Z_t)_{t\in\mathbb{N}}$ be a homogeneous Markov chain on the state space $\mathcal{K}$ and let $i, j \in \mathcal{K}$ be two arbitrary states. If the Markov chain starts from state $i \neq j$ $(Z_0 = i)$, then*

$$\tau_j^{(p)} = \begin{cases} \inf\{n \geq 1 : Z_n = j\} & \exists\ n \geq 1,\ Z_n = j \\ +\infty & \forall\ n \geq 1,\ Z_n \neq j \end{cases}$$

is called a first passage time (Norris 1997, p. 19).

Definition 2.27 (First return time) *Let $(Z_t)_{t\in\mathbb{N}}$ be a homogeneous Markov chain on the state space $\mathcal{K}$ and let $j \in \mathcal{K}$ be an arbitrary state. If the Markov chain starts from state j $(Z_0 = j)$, then*

$$\tau_j^{(r)} = \begin{cases} \inf\{n \geq 1 : Z_n = j\} & \exists\ n \geq 1,\ Z_n = j \\ +\infty & \forall\ n \geq 1,\ Z_n \neq j \end{cases}$$

is called a first return time.

Definition 2.28 (Sojourn time) *Let $(Z_t)_{t\in\mathbb{N}}$ be a homogeneous Markov chain on the state space $\mathcal{K}$ and let $j \in \mathcal{K}$ be an arbitrary state. If the Markov chain starts from state j ($Z_0 = j$), then*

$$\tau_j^{(s)} = \begin{cases} \inf\{n \geq 1 : Z_n \neq j\} & \exists\ n \geq 1,\ Z_n \neq j \\ +\infty & \forall\ n \geq 1,\ Z_n = j \end{cases}$$

is called a sojourn time (Anderson 1991, p. 16).

Both the first passage and the first return time represent the first moment n on which the stochastic process (Z_t) enters the particular state j if in all time points $1, \ldots, n-1$ the stochastic process (Z_t) was in any other state $i \neq j$. The difference being that in the case of the first return time the process (Z_t) starts in the state j (i.e. $Z_0 = j$) and in case of the first passage state the process starts from a particular state $i \neq j$ (i.e. $Z_0 = i$). The sojourn time describes when the process leaves the initial state j (i.e. $Z_0 = j$).

Stopping times can be used for instance to compute a transition probability from the initial state i to a particular state j (equal or unequal i) in n steps.

Definition 2.29 (Transition probability from state i to j in n steps) *Let $(Z_t)_{t\in\mathbb{N}}$ be a homogeneous Markov chain on the state space $\mathcal{K}$ with transition matrix $\boldsymbol{P} = (p_{ji})_{i,j\in\mathcal{K}}$ and let τ_j be the first passage or first return time. Then define*

$$f_{ji}^{(n)} = \Pr[\tau_j = n | Z_0 = i]$$

as the probability that the Markov chain $(Z_t)_{t\in\mathbb{N}}$ goes from the initial state i to the state j in n steps. If $i \neq j$, then $f_{ji}^{(n)}$ is called transition probability of *the first passage time and if $i = j$, then $f_{ji}^{(n)}$ is called the transition probability of the first return time.*

Remark 2.30 *Note that the expected stopping time can be simply computed as*

$$\mathbb{E}[\tau_j] = \sum_{n=1}^{\infty} n \times \Pr[\tau_j = n].$$

In this thesis, the expected sojourn time will be of particular interest, as it is used to price the option on a risky asset when the underlying asset follows a Markov switching geometric Brownian motion (see Section 2.7.2.1).

Theorem 2.31 (Expected sojourn time) *Let $(Z_t)_{t\in\mathbb{N}}$ be a homogeneous Markov chain on the state space $\mathcal{K}$ with transition matrix $\boldsymbol{P} = (p_{ji})_{i,j\in\mathcal{K}}$ with $p_{jj} < 1$ and let $j \in \mathcal{K}$ be an arbitrary state. Then the expected sojourn in state j is given by*

$$D_j = \frac{1}{1 - p_{jj}}, \tag{2.13}$$

see (Kim and Nelson 1999, p. 71-72).

Proof. Note that

$$\begin{aligned}
D_j = \mathbb{E}[\tau_j^{(s)}] &= \sum_{t=1}^{\infty} t \Pr[\tau_j^{(s)} = t] = \sum_{t=1}^{\infty} t\ (1 - p_{jj}) p_{jj}^{t-1} = (1 - p_{jj}) \sum_{t=1}^{\infty} t\ p_{jj}^{t-1} \\
&= (1 - p_{jj})[p_{jj} + 2p_{jj}^2 + 3p_{jj}^3 + 4p_{jj}^4 + \dots] \\
&= (1 - p_{jj})[p_{jj} + \ p_{jj}^2 + \ p_{jj}^3 + \ p_{jj}^4 + \dots \\
&\qquad\qquad\qquad + \ p_{jj}^2 + \ p_{jj}^3 + \ p_{jj}^4 + \dots \\
&\qquad\qquad\qquad\qquad\qquad\qquad + \dots]
\end{aligned}$$

Observe that as $p_{jj} < 1$, we can use the sum of the geometric sequence $\sum_{k=t}^{\infty} p_{jj}^k = \frac{p_{jj}^t}{1-p_{jj}}$. Thus

$$D_j = (1 - p_{jj}) \sum_{t=0}^{\infty} \frac{p_{jj}^t}{1 - p_{jj}} = \sum_{t=0}^{\infty} p_{jj}^t = \frac{1}{1 - p_{jj}},$$

which completes the proof. ■

2.3.4 Stationarity of Markov chains

This Section will show that if a Markov chain has an equilibrium distribution, then it is unique and how to compute it. First, let us define a stationary distribution.

Definition 2.32 (Stationarity) *Let $\boldsymbol{P} = (p_{ji})_{i,j\in\mathcal{K}}$ be a stochastic matrix and let $\boldsymbol{p}$ be a probability distribution. If*

$$\boldsymbol{p} = \boldsymbol{P}'\boldsymbol{p}, \tag{2.14}$$

then $\boldsymbol{p}$ is called stationary, equilibrium or equivalently invariant distribution (Norris 1997, p. 33).

Definition 2.33 (Ergodicity) *Let $(Z_t)_{t\in\mathbb{N}}$ be a homogeneous Markov chain on the state space $\mathcal{K}$ with transition matrix $\boldsymbol{P} = (p_{ji})_{i,j\in\mathcal{K}}$. If it holds*

(a) for all states $j \in \mathcal{K}$ there exists a limit

$$\pi_j = \lim_{n\to\infty} p_{ji}^{(n)}, \tag{2.15}$$

(b) π_j is strictly positive and independent of the state i,

(c) $\boldsymbol{\pi} = (\pi_j)_{j\in\mathcal{K}}$ is a probability distribution, i.e. $\boldsymbol{\pi}'\mathbf{1} = 1$,

then the Markov chain $(Z_t)_{t\in\mathbb{N}}$ is called ergodic and $\boldsymbol{\pi}$ is called stochastic distribution (Rolski et al. 1999, p. 281).

Remark 2.34 *The equation (2.15) in the matrix notation yields*

$$\lim_{n\to\infty} \boldsymbol{P}^n = \begin{pmatrix} \boldsymbol{\pi}' \\ \vdots \\ \boldsymbol{\pi}' \end{pmatrix} = \boldsymbol{\pi}\mathbf{1}', \tag{2.16}$$

where $\boldsymbol{P}$ denotes a transition matrix and $\boldsymbol{\pi}$ denotes an ergodic distribution.

Theorem 2.35 *Let $(Z_t)_{t\in\mathbb{N}}$ be a homogeneous Markov chain on the state space $\mathcal{K}$ with transition matrix $\boldsymbol{P} = (p_{ji})_{i,j\in\mathcal{K}}$. If the Markov chain (Z_t) is ergodic, then the stochastic vector $\boldsymbol{\pi} = (\pi_j)_{j\in\mathcal{K}}$ is the unique solution of the system of linear equations*

$$\boldsymbol{\pi} = \boldsymbol{P}'\boldsymbol{\pi}. \tag{2.17}$$

Equation (2.17) is called balance equation for the matrix $\boldsymbol{P}$ (Rolski et al. 1999, p. 283).

Proof. Note that (2.17) is equivalent to

$$\pi_j = \sum_{k\in\mathcal{K}} \pi_k p_{jk}, \quad \forall\ j \in \mathcal{K}. \tag{2.18}$$

Thus, it is sufficient to prove the equation (2.18). From (2.15) it follows that

$$\pi_j = \lim_{n\to\infty} p_{ji}^{(n)}.$$

Now use the Chapman-Kolmogorov equation (2.11)

$$\lim_{n\to\infty} p_{ji}^{(n)} = \lim_{n\to\infty} \sum_{k\in\mathcal{K}} p_{ki}^{(n-1)} p_{jk}.$$

We interchange the limit and the summation

$$\lim_{n\to\infty} \sum_{k\in\mathcal{K}} p_{ki}^{(n-1)} p_{jk} = \sum_{k\in\mathcal{K}} \lim_{n\to\infty} p_{ki}^{(n-1)} p_{jk}.$$

We re-use the definition of ergodicity (2.15)

$$\sum_{k\in\mathcal{K}} \lim_{n\to\infty} p_{ki}^{(n-1)} p_{jk} = \sum_{k\in\mathcal{K}} \pi_k p_{jk}.$$

As it has be proven that the left-hand-side of equation (2.17) equals its right-hand-side, it remains to be proved that this solution is unique. We now suppose that another probability function $(\boldsymbol{\pi}^*) = (\pi_j^*)_{j\in\mathcal{K}}$ exists, which is unequal to $\boldsymbol{\pi}$. By induction it is straightforward to prove that

$$\pi_j^* = \sum_{k\in\mathcal{K}} \pi_k^* p_{jk}^{(n)}, \quad \forall\ j \in \mathcal{K}, \tag{2.19}$$

with the first step as above. We now take the limit from (2.19)

$$\pi_j^* = \lim_{n\to\infty} \sum_{k\in\mathcal{K}} \pi_k^* p_{jk}^{(n)}.$$

It remains to interchange the summation and limit, and to use the definition of ergodicity (2.15)

$$\pi_j^* = \lim_{n\to\infty} \sum_{k\in\mathcal{K}} \pi_k^* p_{jk}^{(n)} = \sum_{k\in\mathcal{K}} \lim_{n\to\infty} \pi_k^* p_{jk}^{(n)} = \pi_j.$$

Accordingly, it has been proven that $\pi_j^* = \pi_j$ ($\forall\ j \in \mathcal{K}$), thus vector $\boldsymbol{\pi}$ is the unique solution of equation (2.17) (Rolski et al. 1999, p. 283). ■

Theorem 2.36 (Ergodic distribution) *Let $\boldsymbol{P}$ be the transition matrix of the ergodic Markov chain $(Z_t)_{t\in\mathbb{N}}$, then the matrix $(\boldsymbol{I} - \boldsymbol{P}' + \boldsymbol{E})$ is invertible and the stochastic vector from the equation (2.17) is given by*

$$\boldsymbol{\pi} = (\boldsymbol{I} - \boldsymbol{P}' + \boldsymbol{E})^{-1}\mathbf{1} \tag{2.20}$$

(Rolski et al. 1999, p. 288).

Proof. At first it is necessary to prove that the matrix $(\boldsymbol{I} - \boldsymbol{P}' + \boldsymbol{E})$ is invertible. This is equivalent with the following statement: $(\boldsymbol{I} - \boldsymbol{P}' + \boldsymbol{E})\boldsymbol{x} = \mathbf{0}$ implies that $\boldsymbol{x} = \mathbf{0}$. From equation (2.14) it follows $(\boldsymbol{I} - \boldsymbol{P}')\boldsymbol{\pi} = \mathbf{0}$. Therefore from $(\boldsymbol{I} - \boldsymbol{P}' + \boldsymbol{E})\boldsymbol{x} = \mathbf{0}$ it results that $0 = \boldsymbol{\pi}'(\boldsymbol{I} - \boldsymbol{P}' + \boldsymbol{E})\boldsymbol{x} = 0 + \boldsymbol{\pi}'\boldsymbol{E}\boldsymbol{x} = 0$. As $\boldsymbol{\pi}$ is a distribution, it follows $\boldsymbol{\pi}'\boldsymbol{E} = \mathbf{1}'$. Therefore $\mathbf{1}'\boldsymbol{x} = 0$ and consequently $\boldsymbol{E}\boldsymbol{x} = \mathbf{0}$. This implies that $(\boldsymbol{I} - \boldsymbol{P}')\boldsymbol{x} = \mathbf{0}$. This is equivalent to $\boldsymbol{P}'\boldsymbol{x} = \boldsymbol{x}$. From this, it follows that for all $n \geq 1$ it is true that $(\boldsymbol{P}')^n\boldsymbol{x} = \boldsymbol{x}$. As $\boldsymbol{\pi}$ is ergodic, from the equation (2.15) it follows $\lim_{n\to\infty} \boldsymbol{P}^n = \boldsymbol{\pi}\mathbf{1}'$. This means that for $n \to \infty$ it holds $\boldsymbol{x} = (\boldsymbol{P}')^n\boldsymbol{x} \to \mathbf{1}\boldsymbol{\pi}'\boldsymbol{x}$, i.e. $x_i = \sum_{j=1}^m \pi_j x_j$ ($\forall\ i = 1, \ldots, m$). As the right-hand side of this equation is independent from i, it holds true that $\exists\ c \in \mathbb{R},\ \boldsymbol{x} = c\mathbf{1}$. Above it was shown that $0 = \mathbf{1}'\boldsymbol{x}$, therefore $\mathbf{1}'\boldsymbol{x} = \mathbf{1}'(c\mathbf{1}) = cm = 0$. Thus, $c = 0$ and, as a result, $(\boldsymbol{I} - \boldsymbol{P}' + \boldsymbol{E})$ is invertible.

From equation (2.17) it follows that $(\boldsymbol{I}-\boldsymbol{P}')\boldsymbol{\pi} = \mathbf{0}$. Thus $(\boldsymbol{I}-\boldsymbol{P}'+\boldsymbol{E})\boldsymbol{\pi} = \boldsymbol{E}\cdot\boldsymbol{\pi} = \mathbf{1}$. Left multiplying by $(\boldsymbol{I}-\boldsymbol{P}'+\boldsymbol{E})$ yields the equation (2.20) (Rolski et al. 1999, p. 288). ■

Corollary 2.37 *For a two state Markov chain the ergodic distribution has the form*

$$\boldsymbol{\pi} = \begin{pmatrix} \frac{1-p_{22}}{2-p_{11}-p_{22}} \\ \frac{1-p_{11}}{2-p_{11}-p_{22}} \end{pmatrix} \tag{2.21}$$

(Hamilton 1994, p. 683).

2.4 Continuous stochastic models in finance

2.4.1 Diffusion models

The basic stochastic process in the continuous time is called the Wiener process.

Definition 2.38 (Wiener process) *If for a stochastic process $(W_t)_{t\geq 0}$ it holds that*

(i) $W_0 = 0$,

(ii) *The process W has independent increments, i.e. if $r < s \leq t < u$ then $W_u - W_t$ and $W_s - W_r$ are independent stochastic variables,*

(iii) *For $s < t$ the stochastic variable $W_t - W_s$ is normally distributed with mean 0 and variance $(t-s)$,*

(iv) *W has continuous trajectories,*

then W is called a standard Wiener process (or Brownian motion) (Björk 2004, p. 36).

Based on the standard Wiener process, the Itô process (the stochastic integral) can be defined.

Definition 2.39 (Itô process) *Let $(W_t)_{t\geq 0}$ be a standard Wiener process and let $(X_t)_{t\geq s}$ be a stochastic process defined on the probability space $(\Omega, \mathcal{F}, \mathcal{P})$ of the form*

$$X_t = X_s + \int_s^t \mu(u, X_u)du + \int_{t_0}^t \sigma(u, X_u)dW_u \tag{2.22}$$

with diffusion function $\sigma(t, X_t)$ and drift function $\mu(t, X_t)$, which are both $\mathcal{F}_t$-adapted and

$$\Pr\left[\int_s^t \sigma(s, X_u)^2 du < \infty\right] = 1, \qquad \forall t \geq s, \tag{2.23}$$

$$\Pr\left[\int_s^t |\mu(u, X_u)|\, du < \infty\right] = 1, \qquad \forall t \geq s, \tag{2.24}$$

where $s \leq t$. Then the stochastic process X_t is called Itô process (or Itô integral, or stochastic integral) (Shreve 2004, p. 143).

Remark 2.40 *In the financial literature, the Itô integral is commonly written in the differential form as*

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t \text{ with boundary condition } X_s = x_s. \tag{2.25}$$

Theorem 2.41 (Itô's lemma) *Let X_t be an Itô process given by*

$$X_t = X_s + \int_s^t \mu(u, X_u)du + \int_s^t \sigma(u, X_u)dW_u \tag{2.26}$$

with $\mu(t, X)$ and $\sigma(t, X_t)$ $\mathcal{F}_t$-adapted processes, W_t the standard Wiener process and $s \leq t$. Furthermore, let $g(t, x) \in \mathcal{C}^2([0, \infty) \times \mathbb{R})$ (i.e $g(t, x)$ be twice continuously differentiable on $[0, \infty) \times \mathbb{R}$). Then

$$Y_t = g(t, X_t)$$

is again an Itô process with

$$Y_t = Y_s + \int_s^t \left[\frac{\partial g(u, X_u)}{\partial u} + \mu(u, X_u) \frac{\partial g(u, X_u)}{\partial x} + \frac{1}{2}\sigma^2(u, X_u) \frac{\partial^2 g(u, X_u)}{\partial x^2} \right] du + \int_s^t \sigma(u, X_u) \frac{\partial g(u, X_u)}{\partial x} dW_u \tag{2.27}$$

(Øksendal 2003, p. 44).

Proof. For the proof, see Arnold (1973, p. 108-112). ■

Remark 2.42 *Let W_t be a Wiener process and let $s \leq t$. Then*

$$dt \cdot dt = 0, \qquad dt \cdot dW_t = 0, \qquad dW_t \cdot dW_t = dt, \tag{2.28}$$

$$\mathbb{E}\left[\int_s^t dW_u\right] = 0, \qquad \mathbb{E}\left[\left(\int_s^t dW_u\right)^2\right] = \int_s^t du \tag{2.29}$$

(Øksendal 2003, p. 44).

A standard example of the stochastic process used for describing the behavior of the stock prices (or in general the value of the risky portfolio) is the geometric Brownian motion.

Definition 2.43 (Geometric Brownian motion) *Let $(W_t)_{t \geq 0}$ be a standard Wiener process, let $(X_t)_{t \geq s}$ be a stochastic process, and let $\mathcal{F}_t, t \geq 0$ be an associated filtration. Furthermore, let μ_t and σ_t be associated processes. Then the Itô process*

$$X_t = X_s + \int_s^t \mu_u X_u du + \int_s^t \sigma_u X_u dW_u \tag{2.30}$$

is called the geometric Brownian motion (GBM) (Shreve 2004, p. 147-148).

This process was first proposed by Bachelier (1900)[2] in his PhD thesis, which was unfortunately, misunderstood by his contemporaries.[3] Almost six decades later, Osborne (1959) rediscovered the geometric Brownian motion for finance, oblivious to the fact that it has been used for the same purpose by a French mathematician long before him. As a tribute to Bachelier's genius, it is worth mentioning that he used the geometric Brownian motion five years prior to Einstein (1905). Not being familiar with Bachelier's work, Einstein introduced it to the field of physics. Bachelier also priced options decades before Black and Scholes (1973) and Merton (1973) revolutionized the science of finance with their option pricing model (Mandelbrot and Hudson 2004, p. 53-54).

Proposition 2.44 (Solution of the GBM) *Let $(X_t)_{t\geq s}$ be a GBM, $\mu_t = \mu$, $\sigma_t = \sigma > 0$, then equation (2.30) has the solution*

$$X_t = X_s e^{\int_s^t (\mu - \frac{1}{2}\sigma^2)du + \sigma \int_s^t dW_u} = X_s e^{\left(\mu - \frac{1}{2}\sigma^2\right)(t-s) + \sigma(W_t - W_s)}, \tag{2.31}$$

Proof. Set $u(t, X(t)) = \ln X_t$ and use the Itô rule, which results in

$$\ln X_t - \ln X_s = (\mu - \frac{1}{2}\sigma^2)(t - s) + \sigma \int_s^t dW_u. \tag{2.32}$$

Then take $\exp(\cdot)$ of both sides, which completes the proof. For details, see Shreve (2004, p. 191-193). ■

[2]The English translation of Bachelier's article can by found in Cootner (1964, p. 17-78).

[3]The extent to which Bachelier had been misunderstood can be illustrated by the difficulties he encountered to get a permanent professorship. E.g. his application for a vacancy position in Dijon in 1926 was blackballed due to the criticism of the distinguished mathematician Paul Lévy on his work, which allegedly contained profound mistakes. Eventually, Bachelier was able to get a permanent position at Besançon one year later. It was not until many years has passed that Lévy apologized to Bachelier, admitting to have been at fault (Mandelbrot and Hudson 2004, p. 48-49).

Remark 2.45 *Note that process $Y_t = \ln(X_t)$ defined in equation (2.32) is referred to as the arithmetic Brownian motion with mean $(\mu - \frac{1}{2}\sigma^2)(t-s)$ and variance $\sigma^2(t-s)$.*

Proposition 2.46 (Moments of the GBM) *Suppose that $(X_t)_{t\geq s}$ is a GBM, then it has the conditional expected value*

$$\mathbb{E}[X_t|X_s] = X_s e^{\mu(t-s)} \tag{2.33}$$

and the conditional variance

$$\mathbb{V}ar[X_t|X_s] = X_s^2 e^{2\mu(t-s)} \left(e^{\sigma^2(t-s)} - 1\right). \tag{2.34}$$

Proof. To prove equation (2.33) take the conditional expected value of both sides of equation (2.31)

$$\begin{aligned}\mathbb{E}[X_t|X_s] &= \mathbb{E}\left[X_s e^{(\mu-\frac{1}{2}\sigma^2)(t-s)+\sigma(W_t-W_s)}|X_s\right] \\ &= X_s e^{(\mu-\frac{1}{2}\sigma^2)(t-s)}\mathbb{E}\left[e^{\sigma(W_t-W_s)}\right].\end{aligned}$$

As $W_t - W_s$ is normally distributed with mean 0 and variance $(t-s)$ one can use the moment generating function of the normal distribution

$$\mathbb{E}[e^{\sigma(W_t-W_s)}] = \mathbb{M}(\sigma, t-s) = e^{\frac{1}{2}\sigma^2(t-s)},$$

thus

$$\mathbb{E}[X_t|X_s] = X_s e^{(\mu-\frac{1}{2}\sigma^2)(t-s)} e^{\frac{1}{2}\sigma^2(t-s)} = X_s e^{\mu(t-s)}.$$

To prove equation (2.34), first take the conditional expected value of the second power of both sides of equation (2.31)

$$\begin{aligned}\mathbb{E}[X_t^2|X_s] &= \mathbb{E}\left[X_s^2 e^{2(\mu-\frac{1}{2}\sigma^2)(t-s)+2\sigma(W_t-W_s)}|X_s\right] \\ &= X_s^2 e^{2(\mu-\frac{1}{2}\sigma^2)(t-s)}\mathbb{E}\left[e^{2\sigma(W_t-W_s)}\right] \\ &= X_s^2 e^{2(\mu-\frac{1}{2}\sigma^2)(t-s)} e^{2\sigma^2(t-s)} \\ &= X_s^2 e^{2\mu(t-s)+\sigma^2(t-s)},\end{aligned}$$

then the variance equals

$$\begin{aligned}\mathbb{V}\mathrm{ar}[X_t|X_s] &= \mathbb{E}[X_t^2|X_s] - \mathbb{E}^2[X_t|X_s]\\ &= X_s^2 e^{2\mu(t-s)+\sigma^2(t-s)} - X_s^2 e^{2\mu(t-s)}\\ &= X_s^2 e^{2\mu(t-s)}(e^{\sigma^2(t-s)} - 1).\end{aligned}$$

■

Remark 2.47 *Because the increments of the Wiener process are normally distributed (see Definition 2.38), thus equation (2.32) implies that for $X_s = x_s$ process $(\ln(X_t))_{t\geq 0}$ is also normally distributed. Equivalently, process $(X_t)_{t\geq 0}$ is log-normally distributed.*

The GBM is usually used to describe stochastic behaviour of stock prices. However, it is unsuitable for interest rate models, as it does not take the mean reversion into account (see Section 2.1.5), which is commonly observed in the interest rates time series. Therefore, Vasiček (1977) proposed a model that deals with this problem.

Definition 2.48 (Vasiček process) *Suppose that W_t, $t \geq 0$ is a Wiener process, $\mathcal{F}_t, t \geq 0$ is an associated filtration, μ_t and σ_t are associated processes and $\alpha > 0$. Then the Itô process*

$$X_t = X_s + \int_s^t \alpha[\mu_u - X_u]du + \int_s^t \sigma_u dW_u \tag{2.35}$$

is called the Vasiček process.

Proposition 2.49 (Solution of the Vasiček process) *Suppose that $(X_t)_{t\geq s}$ is a Vasiček process with $\mu_t = \mu$ and $\sigma_t = \sigma > 0$, then equation (2.35) has the solution*

$$X_t = X_s e^{-\alpha(t-s)} + \mu(1 - e^{-\alpha(t-s)}) + \sigma \int_s^t e^{-\alpha(t-u)} dW_u. \tag{2.36}$$

Proof. Set $g(t, X_t) = e^{\alpha(t-s)}X_t$ and apply the Itô rule. ■

Proposition 2.50 (Moments of the Vasiček process) *Suppose that* $(X_t)_{t\geq s}$ *is a Vasiček process with* $\mu_t = \mu$ *and* $\sigma_t = \sigma > 0$*, then it has the expected value of*

$$\mathbb{E}[X_t|X_s] = X_s e^{-\alpha(t-s)} + \mu(1 - e^{-\alpha(t-s)}) \tag{2.37}$$

and the variance of

$$\mathbb{V}ar[X_t|X_s] = \frac{\sigma}{2\alpha}\left[1 - e^{-2\alpha(t-s)}\right] \tag{2.38}$$

Proof. Using Shreve (2004, Theorem 4.4.9, p. 149) for the stochastic variable $Z_{s,t} = \int_s^t e^{-\alpha(t-u)}dW_u$ has the expected value of

$$\mathbb{E}[Z_{s,t}] = 0 \tag{2.39}$$

and the variance of

$$\mathbb{V}\text{ar}[Z_{s,t}] = \int_s^t e^{-2\alpha(t-u)}dW_u = \frac{1}{2\alpha}\left(1 - e^{-2\alpha(t-s)}\right). \tag{2.40}$$

To prove the mean equation (2.37) it is sufficient to take the expected value of both sides of the equation (2.36) and apply (2.39). To prove the variance equation (2.38) it is sufficient take variance of both sides of the equation (2.36)

$$\begin{aligned}\mathbb{V}\text{ar}[X_t|X_s] =& \mathbb{V}\text{ar}\left[X_s e^{-\alpha(t-s)} + \mu(1 - e^{-\alpha(t-s)}) + \sigma\int_s^t e^{-\alpha(t-u)}dW_u\middle| X_s\right]\\ =& \mathbb{V}\text{ar}\left[X_s e^{-\alpha(t-s)} + \mu(1 - e^{-\alpha(t-s)})\middle| X_s\right] + \sigma^2\mathbb{V}\text{ar}\left[Z_{s,t}\right]\\ &+ 2\mathbb{C}\text{ov}\left[X_s e^{-\alpha(t-s)} + \mu(1 - e^{-\alpha(t-s)}), \sigma Z_{s,t}\middle| X_s\right].\end{aligned}$$

As the first variance is zero and both processes are independent, the covariance is zero as well. Thus it remains to apply (2.40), which completes the proof (Shreve 2004, p. 151). ■

Remark 2.51 *It is worth noting that the parameter μ is a long term mean of the Vasiček process*[4] *with respect to $t \to \infty$ and that α is the speed at which the process returns to the long term mean μ.*

As the X_t is normally distributed, the model allows negative interest rates with a positive probability. This is a common point of criticism of the Vasiček model. Cox, Ingersoll, and Ross (1985) addressed this issue using the square root process to exclude negative interest rates.

Definition 2.52 (CIR process) *Let W_t, $t \geq 0$ be a Wiener process, let $\mathcal{F}_t, t \geq 0$ be an associated filtration, let $\mu(t)$ and $\sigma(t)$ be associated processes, and let $s \leq t$. Then the Itô process*

$$X_t = X_s + \int_s^t \alpha[\mu_u - X_u]du + \int_s^t \sigma_u \sqrt{X_u} dW_u \tag{2.41}$$

is called a CIR process.

The disadvantage of this model is that it does not have a closed-form solution of the stochastic differential equation. However, it is possible to compute its expected value and variance.

Note that it is a mean-reverting process if $\alpha > 0$.

Proposition 2.53 (Expected value and variance of the CIR process) *Suppose that X_t is a CIR process, then it has a conditional expected value of*

$$\mathbb{E}[X_t|X_s] = X_s e^{-\alpha(t-s)} + \mu(1 - e^{-\alpha(t-s)}) \tag{2.42}$$

and a conditional variance of

$$\mathbb{V}ar[X_t|X_s] = X_s \frac{\sigma^2}{\alpha} \left(e^{-\alpha(t-s)} - e^{-2\alpha(t-s)}\right) + \frac{\mu\sigma^2}{2\alpha} \left(1 - e^{-\alpha(t-s)}\right)^2 \tag{2.43}$$

Proof. For proof, see Gourieroux and Jasiak (2001, p. 252-253). ■

[4] Which is easy to see by taking the limes from the right-hand side of equation (2.37).

Definition 2.54 (Generalized one-factor process) *Suppose that W_t, $t \geq 0$ is a Wiener process, $\mathcal{F}_t, t \geq 0$ is an associated filtration, and μ_t and σ_t are associated processes. Then the Itô process*

$$X_t = X_s + \int_s^t \alpha[\mu_u - X_u]du + \int_s^t \sigma_u X_u^\gamma dW_u \tag{2.44}$$

is called a generalized one-factor interest rate process.

Remark 2.55 *Note that it is a mean-reverting process if $\alpha > 0$ and a heteroskedastic process if $\gamma > 0$.*

This class includes many other processes, such as Vasiček (1977) ($\gamma = 0$), Cox, Ingersoll, and Ross (1985) ($\gamma = \frac{1}{2}$), Brennan and Schwartz (1980) ($\gamma = 1$). For other one-factor models, compare Chan et al. (1992).

It is worth mentioning that the term-structure determined with the above models at time $t = 0$ can deviate significantly from the term-structure observed on the market (Albrecht and Maurer 2008, p. 517). Some authors have attempted to address this problem. See e.g. Hull and White (1990) who solved this problem for the Vasiček and CIR model.

The one-factor models above discussed assume that the behavior of the interest rate can be described by utilizing only one factor. As the one-factor models are too simplistic, several authors developed models with more than one factor. Brennan and Schwartz (1979b, 1982) use both the short interest rate and the yield of a consol bond with continuous coupon payment and infinite maturity as factors. Fong and Vasiček (1991) and Longstaff and Schwartz (1992) use the short interest rate and its volatility as factors. Schaefer and Schwartz (1984) on the other hand, employ the long interest rate and the spread between the short and long interest rate. There are also some three-factor models e.g. by Kraus and Smith (1993) who use the short interest rate, the drift function and the diffusion of the short rate as factors.

2.4.2 Discretization of diffusion models

By using the lemma of Itô (1951) to solve the equation (2.30), the process for the development of the risky asset

$$S_t = S_{t_0} e^{\left(\mu - \frac{1}{2}\sigma^2\right)(t-t_0)+\sigma(W_t - W_{t_0})}, \qquad 0 \leq t_0 < t,\ S_{t_0} > 0 \tag{2.45}$$

can be determined (Albrecht and Maurer 2008, p. 175). Thus, the log rates of return for one period y_t (i.e. $t_0 = t-1$) are equal to

$$y_t = \ln \frac{S_t}{S_{t-1}} = \left(\mu - \frac{1}{2}\sigma^2\right) + \sigma(W_t - W_{t-1}) \tag{2.46}$$

From Definition 2.38 it follows that $\Delta W = W_t - W_{t_0}$ is normally distributed with the mean 0 and the standard deviation $\sqrt{t-t_0}$ (i.e. $\Delta W \sim N(0, t-t_0)$). Thus, the equation (2.46) can be rewritten as

$$y_t = \left(\mu - \frac{1}{2}\sigma^2\right) + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma^2) \tag{2.47}$$

Please note that the mean of the GBM $\left(\mu - \frac{1}{2}\sigma^2\right)$ is a constant as well, thus equation (2.47) can be rewritten as the Gaussian white noise with the mean $m = \left(\mu - \frac{1}{2}\sigma^2\right)$ and the variance σ^2

$$y_t = m + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma^2). \tag{2.48}$$

From the Itô lemma (Theorem 2.41) the Vasiček process (2.35) has the following solution

$$R_t = e^{-\alpha(t-t_0)} R_{t_0} + \mu(1 - e^{-\alpha(t-t_0)}) + \sigma \int_{t_0}^{t} e^{-\alpha(t-u)} dW_u \tag{2.49}$$

(Brigo and Mercurio 2006, p. 58). Accordingly, the one period case (i.e. $t_0 = t-1$) R_t is given by

$$R_t = e^{-\alpha} R_{t-1} + \mu(1 - e^{-\alpha}) + \sigma \int_{t_0}^{t} e^{-\alpha(t-u)} dW_u. \tag{2.50}$$

From Definition 2.38 it follows that $\Delta W = W_t - W_{t_0}$ is normally distributed with the mean 0 and the standard deviation $\sqrt{t-t_0}$ (i.e. $\Delta W \sim N(0, t-t_0)$). Thus, the equation (2.50) can be rewritten as

$$R_t = e^{-\alpha} R_{t-1} + \mu(1-e^{-\alpha}) + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\alpha}(1-e^{-2\alpha})\right), \tag{2.51}$$

with ε_t as the innovation at time t.

The explicit solution of equation (2.51) can be written as a first order autoregressive process (or equivalently AR(1)-process)

$$R_t = c + \phi R_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\alpha}(1-e^{-2\alpha})\right), \tag{2.52}$$

with a constant $c = \mu(1-e^{-\alpha})$ and the autoregressive coefficient $\phi = e^{-\alpha}$. The AR-version of the Vasiček model has the advantage that it enables the GBM model to be written as a special case of the AR with $c = m$ and $\phi = 0$ (see equation (2.48)).

The disadvantage of the CIR model is that it does not have a closed-form solution of the Itô lemma. However, this can be approximated via the Euler method

$$R_t = R_{t-1} + \alpha(\mu - R_{t-1}) + \varepsilon_t = \alpha\mu + (1-\alpha)R_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma^2 R_{t-1}), \tag{2.53}$$

(Albrecht and Maurer 2008, p. 186).

2.5 Literature overview (MS models)

A number of years ago, the general observation was made, that economic variables as GDP, inflation or stock prices can behave differently in several states (or regimes) of the economy. For instance, the economy can be in an expansion or a recession phase. The first attempt to implement this insight for econometric research was undertaken by Quandt (1958). He studied a model in which the consumption equation C varies in dependence of one of two different states

$$C = \alpha_1 Y + \beta_1$$

$$C = \alpha_2 Y + \beta_2.$$

Here Y denotes the income and α_1, β_1, and α_2, β_2 denote the regression parameters in the first or second regime, respectively. The shift from the first state of the economy to the second one occurs only once, and is dependent on an exogenous variable (say Z). If the value of the variable exceeds some critical value (say z^*), the consumption shifts from the first state to the second one. However, Quandt (1958) assumes that the exogenous variable Z cannot be identified, and thus the shift time t^* cannot be identified either. An essential assumption of Quandt's model is the a priori knowledge that the shift occurs once.[5] The only unidentified variable is the switching time t^* which can be estimated via the maximum likelihood method.

An extension of this model was proposed by Goldfeld and Quandt (1972) who admitted an unknown number of switches. In their model, the shift time(s) depends on some observable exogenous variable Z_t. Only the function determining when the shift occurs is unknown. In this approach, the parameters of this function have to be estimated.[6] Goldfeld and Quandt (1972) also studied another case in which the choice between the first and the second regime occurs with some probabilities π and $1 - \pi$ which are unknown. In this model, the probability of a switch is independent of the

[5] Despite that Quandt (1958, p. 873, footnote 1) studied a case in which the shift occurs only once, he points out that the generalization of the model with two, three, etc. switches is possible. This would, however, make the model more cumbersome.

[6] This is the so-called D-method, see Goldfeld and Quandt (1973, p. 4-6).

previous state.[7] In a further paper, Goldfeld and Quandt (1973) assumed that the probability of the shift can depend on the state which was in effect at the previous time. E.g. the probability that the economy will be in the first state at time t if it was in the second one in the previous period, will be $\Pr[Z_t = 1|Z_{t-1} = 2]$. This group of models is referred to as the hidden Markov models, since the unobservable state variable Z is a Markov chain.

The idea of a regression with a hidden Markov state was implemented in the time series analysis in the fundamental paper *A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle* of Hamilton (1989) who studied the US real GNP as the AR(4) model with regime switching. As the observation of the regime variable Z is impossible, he suggested an approach which makes a probabilistic inference on the regime at the time t. Since the publication of his paper, there has been a lot of studies on the Markov switching behavior of macroeconomic values such as GDP (Hamilton (1989)) or inflation (Kim (1993) among others). This method has found a lot of applications for financial time series, as well.

The first model for stock returns within the Markov switching framework was introduced by Turner, Startz, and Nelson (1989). In their model, the mean and the volatility of the stock returns can be dependent on the regime. Hardy (2001) analysed the model of Turner, Startz, and Nelson (1989) for the monthly returns of the US American and Canadian stocks and discovered that it outperforms several autoregressive and ARCH models, among them the GARCH(1,1)-model. Sola and Timmermann (1994) used this model for the UK daily stock returns and ascertained that it outperforms the GARCH and the EGARCH model. Furthermore, they concluded that the Markov switching model can better explain the skewness of the data than the models

[7] This is the so-called λ-method, see Goldfeld and Quandt (1973, p. 6-7).

from the ARCH family. However, the ARCH models were more advantageous in explaining the kurtosis of the daily stock returns than the regime switching model.

Hamilton and Susmel (1994) introduce an extension of Turner, Startz, and Nelson's model to the regime switching ARCH (SWARCH) models.[8] In their study, the US American weekly stock returns are better forecasted with the SWARCH model with Gaussian innovations than with Gaussian GARCH(1,1). The prediction was even more accurate if they used the t-distributed SWARCH models. Moreover, they show how to implement the asymmetry of the returns in the SWARCH model. They model the leverage effects as proposed by Glosten, Jagannathan, and Runkle (1993). Hamilton and Lin (1996) used a bivariate model which combines Hamilton (1989) and Hamilton and Susmel (1994) to study the relation between stock returns and economic growth. They found proof for two regimes in the stock returns and GDP. Furthermore, they found that increased volatility in stock returns is associated with economic recessions.

In addition to stock returns, there is a comprehensive research literature on the interest rates within the Markov regime framework. The first such study was conducted by Hamilton (1988), who modeled the short- and the long-time interest rate. He found that the dynamics of the American three-month interest rate is better explained by the Markov switching approach than by a linear model. The bivariate model of the three-month T-bills and ten-year T-bonds with cross-equation restriction is best represented by the Markov switching model as well. Gray (1996) considered the generalized regime-switching (GRS) model which nests the linear (e.g. Cox, Ingersoll,

[8] Other GARCH models with Markov switching were introduced by Cai (1994), Gray (1996) and Haas (2004).

and Ross 1985), (G)ARCH of Engle (1982) and Bollerslev (1986), Markov switching AR of Hamilton (1989), SWARCH of Hamilton and Susmel (1994) and, additionally, a new model with Markov switching in GARCH. Gray (1996) found that his GRS model outperforms the other models in- and out-of-sample for short term interest rates. The study of Ahrens (1998) confirmed these results for the German market.[9]

Dahlquist and Gray (2000) studied the impact of the European Monetary System (EMS) on the short-term behavior of interest rates. They found that the volatility of the interest rates can be assigned to two regimes: regimes of stable and of high volatility. High volatility occurs when the exchange rate almost exceeded the boundary of the target zone; which was interpreted by the authors as a speculative attack on the currency.

One should note that the research conducted on the Markov switching models on the financial markets focuses mainly on stock and interest rates models, with only a few studies on other topics. Engel and Hamilton (1990) implemented the regime shift approach in the USD/DEM, USD/FFR, and USD/GBP exchange rate. They found that the Turner, Startz, and Nelson (1989)-like model explains the dynamics of the exchange rates better than the random walk, which is a good parameterization of the "peso problem".

Alizadeh and Namikos (2004) studied the use of the Markov switching models in the stock index futures. They found that the time-varying minimum hedging ratio can be better established by using the Markov switching model than by using the other models (including GARCH) in the case of the UK market, both in- and out-of-sample. In the US American case, the Markov model is superior only in-the-sample.

[9]To the best of knowledge Ahrens (1998) and Haas (2004) were the only ones who conducted research on the German financial markets using the hidden Markov approach.

Crawford and Fratantoni (2003) conducted a very interesting study on the real estate prices in several states of North America. After comparing the ARIMA, GARCH and regime switching models, they concluded that the hidden Markov models are particularly suitable for explaining the historical dynamics of real estate prices. However, the out-of-sample forecasting was better for the ARIMA models, linked perhaps to the small sample of data the authors used. Additionally, the regime switching models were reliable concerning the prediction of the turning points of the market, which have practical relevance for real estate fund managers.

Recently, the Markov switching models have been implemented in insurance. For instance Yin, Liu, and Yang (2006) developed a Markov switching model which determines the limit of ultimate survival probabilities and ultimate ruin probabilities. Additionally, Yang and Yin (2004) formulated a model of the insurance surplus process within the Markov switching scheme.

2.6 Finite mixture distributions

2.6.1 Definition of finite mixture distribution

A finite mixture distribution model was first used in astronomy by Pearson (1894), who used a mixture of two univariate normal distributions with unequal mean and variance parameters. For mixtures of other distribution families, compare, for instance, Feller (1943) (for Poisson distribution), Barndorf-Nielsen (1978) and Shaked (1980) (for exponential distribution) among others. Since then, Pearson's approach has found very wide applications in modern science. It is inter alia applied in biology, physics, marketing (Rossi, Allenby, and McCulloch 2005), public health (Spiegelhalter, Abrams, and Myles 2003) and, in particular, in finance. For modern finance theory, mixture distributions offer a way of dealing with some of stylized facts mentioned in Section 2.1. The following discussion will focus on the finite mixture of normal distributions.

It is common to assume that the log-returns of financial time series are drawn from the normal distribution $N(\mu, \sigma^2)$. However, as discussed in Section 2.1, this assumption is not confirmed by observation on the financial markets. Supposing now that the realization y of the random variable Y is drawn not from a single, but from one of a finite number of distributions $f_1(y),\ f_2(y), \ldots, f_K(y)$, where $f_j(y)$ (j=1,...,K) is a density function. Consequently, the Y is drawn from a *mixture of distributions* $f_j(Y)$.

Definition 2.56 (Mixture of distributions) *Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space, let Y be a random variable and let $\mathcal{K}$ be a state space. Additionally, let vector $\boldsymbol{\pi} = (\pi_i)_{i \in \mathcal{K}}$ be a probability distribution and let $f_i(y)$ be a probability density function ($i \in \mathcal{K}$). If*

$$f(y) = \pi_1 f_1(y) + \pi_2 f_2(y) + \ldots, \tag{2.54}$$

then Y follows a mixture distribution. The density functions $f_i(y)$ are called component densities, π_j is called weight and the vector $\boldsymbol{\pi}$ is called a weight distribution. If the state space is a finite set, the mixture distribution is called a finite mixture distribution (Frühwirth-Schnatter 2006, p. 3-4).

Note that the vector $\boldsymbol{\pi}$ is a stochastic vector, thus the weights are all positive and sum up to unity

$$\sum_{j \in \mathcal{K}} \pi_j = 1, \quad 0 \leq \pi_j \leq 1 \ (j = 1, \ldots, K) \tag{2.55}$$

and they could be interpreted as probabilities that the observation y comes from a particular density $p_j(y)$. For instance, if one would like to create a model for returns of a stock, one component density may represent the bull market time and another one the bear market time. In most practically relevant cases it cannot be said which component the particular observation was drawn from. For instance, it cannot be observed which *state* (e.g. the bull or the bear market) the market is in. For this reason, let us assume that the market state is represented by an additional random variable, the so-called *state variable* Z.

Definition 2.57 (State variable) *Let* $(\Omega, \mathcal{F}, \mathcal{P})$ *be a probability space, let* Z *be a random variable, let* $\mathcal{K}$ *be a state space and let vector* $\boldsymbol{\pi} = (\pi_j)_{j\in\mathcal{K}}$ *be a probability distribution. If the random variable* Z *has the following probability distribution*

$$f(z) = \begin{cases} 1 & \textit{with probability } \pi_1 \\ 2 & \textit{with probability } \pi_2 \\ \vdots & \\ K & \textit{with probability } \pi_K \end{cases}, \tag{2.56}$$

then it is called a state variable.

2.6.2 Moments of mixed-normal process

Theorem 2.58 (Moments of the finite mixture of normals) *Assume that the random variable* Y *follows a finite mixture of normal distributions, then its mean is equal to*

$$\mu = \mathbb{E}[Y] = \sum_{j=1}^{K} \pi_j \mu_j, \tag{2.57}$$

its variance is equal to

$$\mathbb{V}ar[Y] = \sum_{j=1}^{K} \pi_j(\sigma_j^2 + \mu_j^2) - \mu^2 = \sum_{j=1}^{K} \pi_j \sigma_j^2 + \sum_{j=1}^{K} \pi_j(\mu_j - \mu)^2, \tag{2.58}$$

its skewness is equal to

$$\gamma[Y] = \frac{\sum_{j=1}^{K} \pi_j[(\mu_j - \mu)^2 + 3\sigma_j^2](\mu_j - \mu)}{\left(\sum_{j=1}^{K} \pi_j \sigma_j^2 + \sum_{j=1}^{K} \pi_j(\mu_j - \mu)^2\right)^{(3/2)}}, \tag{2.59}$$

and its excess kurtosis is equal to

$$\kappa[Y] = \frac{\sum_{j=1}^{K} \pi_j[(\mu_j - \mu)^4 + 6(\mu_j - \mu)^2\sigma_j^2 + 3\sigma_j^4]}{\left(\sum_{j=1}^{K} \pi_j \sigma_j^2 + \sum_{j=1}^{K} \pi_j(\mu_j - \mu)^2\right)^2} - 3, \tag{2.60}$$

where $\mu_j = [Y|Z = j]$ *and* $\sigma_j^2 = \mathbb{V}ar[Y|Z = j]$ *for* $j = 1, \ldots, K$ *(Haas 2004, p. 13,17,19).*

Proof. Let $g(Y)$ be a function of Y with respect to the component density $f(y|Z = j)$ $(i = 1, \ldots, K)$. In the expected value $E[g(Y)|Z = j]$ for all components $f(y|Z = j)$ $(j = 1, \ldots, K)$ the expected value of the mixture distribution has the form

$$\mathbb{E}[g(Y)] = \mathbb{E}[\mathbb{E}[g(Y)|Z]] = \sum_{j=1}^{K} \pi_j \mathbb{E}[g(Y)|Z = j]$$

(Frühwirth-Schnatter 2006, p. 10). Furthermore

$$\mathbb{E}[g(Y)|Z = j] = \int_{\mathbb{R}} g(Y) f(y|Z = j) dy, \qquad (j = 1, \ldots, K).$$

Particularly for $g(Y) = y$ the mean equals

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|Z]] = \sum_{j=1}^{K} \pi_j \mathbb{E}[Y|Z = j] = \sum_{j=1}^{K} \pi_j \mu_j$$

(Haas 2004, p. 17). For moments of the order $k \geq 2$ the function $g(Y)$ takes the form $(Y - \mu)^k$. Now it is useful to use the property

$$m_k = \mathbb{E}[(Y - a)^k] = \int_{-\infty}^{\infty} (y - a)^k \phi(y, \mu, \sigma^2) dy = \int_{-\infty}^{\infty} y^k \phi(y, \mu - a, \sigma^2) dy \tag{2.61}$$

(Frühwirth-Schnatter 2006, p. 11). In the discrete case, the equation (2.61) takes the form

$$\begin{aligned}
\mathbb{E}[(Y-\mu)^k] &= \mathbb{E}\left[\sum_{n=0}^{k}\binom{k}{n}Y^n\mu^{k-n}\right] = \sum_{n=0}^{k}\binom{k}{n}\mu^{k-n}\mathbb{E}[Y^n] \\
&= \sum_{n=0}^{k}\binom{k}{n}\mu^{k-n}\mathbb{E}[Y^n] = \sum_{n=0}^{k}\binom{k}{n}\mu^{k-n}\mathbb{E}[\mathbb{E}[Y^n|Z]] \\
&= \sum_{n=0}^{k}\binom{k}{n}\mu^{k-n}\sum_{j=1}^{K}\pi_j\mathbb{E}[Y^n|Z=j]
\end{aligned}$$

The variance, then, is equal to

$$m_2 = \mathbb{V}\mathrm{ar}[Y] = \sum_{j=1}^{K}\pi_j(\mu_j^2+\sigma_j^2)-\mu^2. \tag{2.62}$$

To compute the skewness $\gamma[Y]$ and the excess kurtosis $\kappa[Y]$, the following two expectations are needed

$$m_3 = \mathbb{E}[(Y-\mu)^3] = \sum_{j=1}^{K}\pi_j[(\mu_j-\mu)^2+3\sigma_j^2](\mu_j-\mu) \tag{2.63}$$

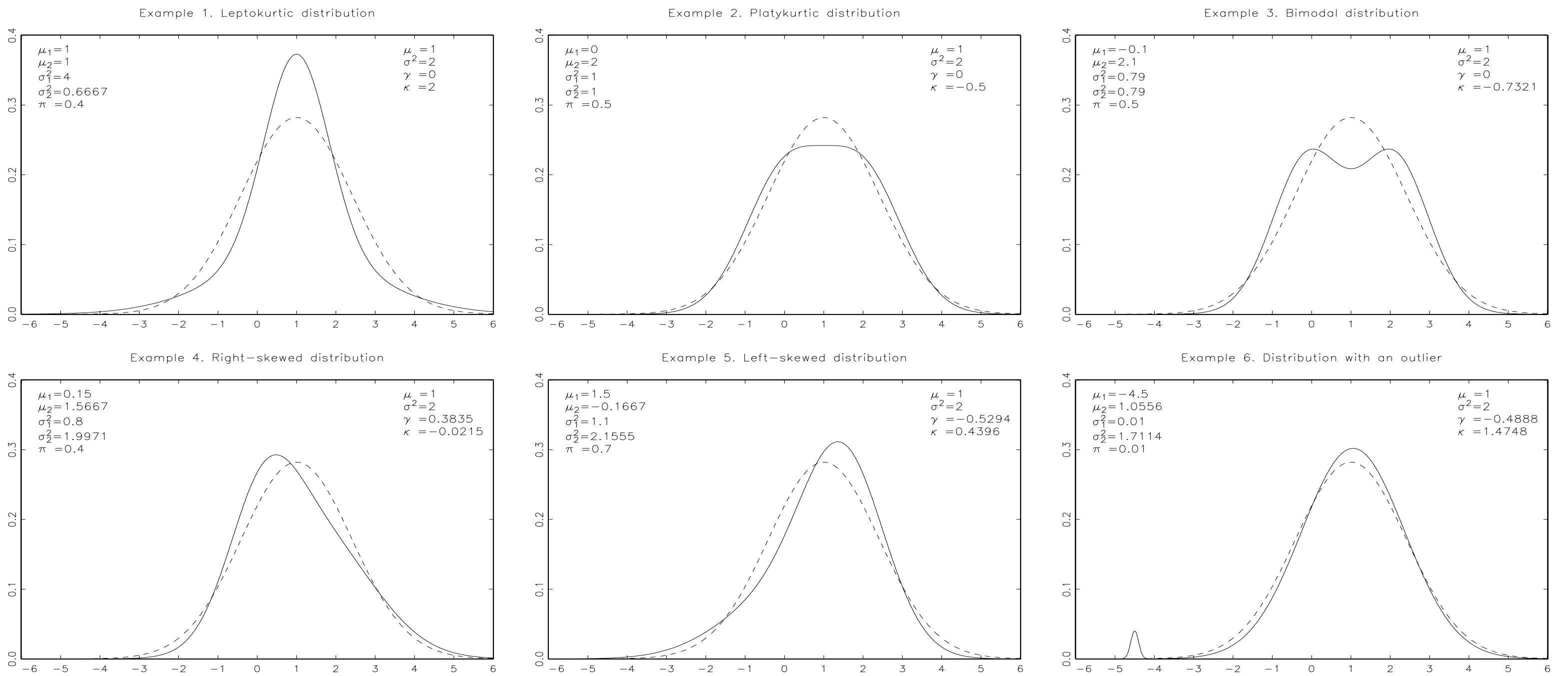
$$m_4 = \mathbb{E}[(Y-\mu)^4] = \sum_{j=1}^{K}\pi_j[(\mu_j-\mu)^4+6(\mu_j-\mu)^2\sigma_j^2+3\sigma_j^4]. \tag{2.64}$$

The skewness is then computed using the expectation (2.63) and the property $\gamma = \frac{m_3}{m_2^{(3/2)}}$. The excess kurtosis can be computed from the expectation (2.64) and the formula for the excess kurtosis $\kappa = \frac{m_4}{m_2^2} - 3$, which completes the proof. ∎

2.6.3 Examples of mixture distributions

This Section aims to demonstrate the potential of mixture distributions in representing a number of different shapes of densities. Figure 2.2 compares the normal density with a mean of 1 and a standard deviation of $\sqrt{2}$ (dashed line) to several mixtures of two normals (solid lines).

Figure 2.2: Examples of mixed normals



Note:

This figure displays six examples of mixed normal distributions which all have the unconditional mean of 1 and the unconditional variance of 2 and are given with the formula $Y \sim \pi \cdot N(\mu_1, \sigma_1^2) + (1-\pi) \cdot N(\mu_2, \sigma_2^2)$. The panels are organized as follows: in the top left corner, of each example, the parameters of the mixed normal distribution are displayed. In the top right corner moments of the mixed normal distribution (μ – mean, σ^2 – variance, γ – skewness, and κ – excess kurtosis) are displayed. The solid line represents a mixture normal distribution with a parameter vector shown in the top left corner. The dashed line represents a normal distribution with a mean of 1 and a variance of 2. Example 1 shows a leptokurtic, Example 2 – a platykurtic, Example 3 – a bimodal, Example 4 – a right-skewed, Example 5 – a left-skewed density, and Example 6 – a density with an outlier, respectively.

Let us fix the means of both composite distributions to be equal. If then their variances are unequal, the mixture will be leptokurtic (see the solid line in Example 1, Figure 2.2), as it will have more probability mass around the mean and in the tails than the normal density with the same unconditional mean and unconditional variance as the mixture (dashed line). The higher the probability weight ascribed to the high volatility distribution (i.e. the higher π) the more probability mass is concentrated in the tails of the mixture. Vice versa, the more weight ascribed to the low volatility density, the more probability mass is concentrated around the mean of the mixture. Keeping in mind that the component means were fixed to be equal, the mixture converges to the normal density if the weight π tends to zero or unity. It can also converge to the normal density independently from the weight parameter if the variances of composite distribution converge to each other. If the weight parameter is fixed (e.g. $\pi = \frac{1}{3}$), the tendency of the mixture being leptokurtic increases with increasing difference between both component variances.

We now consider the following example of a mixture with equal component variances, unequal component means and the parameter $\pi = \frac{1}{2}$. Example 2 in Figure 2.2 showed such a density. As one can see, the mixture is platykurtic because less probability mass is concentrated around the mean and in the tails than in the normal density case. Note that the example is constructed in such a way that the component means are equidistant from the mean of the mixture density (i.e. $\mu_1 = \mu - \delta$ and $\mu_2 = \mu + \delta$), thus the density is symmetric per construction. The lower the distance of the component means to the mixture mean (i.e. if $\delta \to 0$), the more the mixture distribution converges to the normal. If the distance grows, the peak of the mixture becomes flatter and the tails thinner, as is the case of Example 2, in Figure 2.2. If the distance δ exceeds a certain critical point, the mixture becomes bimodal, as shown in Example 3 of Figure 2.2. In this case, the mentioned critical delta lies between 1 and 1.1.

So far three symmetric examples of mixture distributions have been discussed. Let us now focus the attention on some skewed examples. If we relaxed some of above restrictions, e.g. if in Example 2 the component densities were not equally weighted (i.e. $\pi \neq \frac{1}{2}$), or if the component variances were unequal, or if the component means were not equidistant from the mixture mean, then one possible result could be that the mixture density becomes skewed. Note, that the inequality of the component means is *condicio sine qua non* for asymmetry of a mixture of normals. It is not, however, a sufficient condition, as can be seen in Examples 2 or 3. Examples 4 and 5 from Figure 2.2 depict instances for a right-skewed and left-skewed density, respectively. The examples are constructed in such a way that the right-skewed density is simultaneously platykurtic and the left-skewed density is leptokurtic, although reverse cases are, of course, also possible. As there are a number of cases in which the mixture of normals is skewed, a discussion on the behavior of the mixture conditional on its parameter will be omitted. Instead, one more example will be given, since – as was the case with the leptokurtic and left-skewed distributions – this could be very interesting for statistical description of financial time series.

Example 6 in Figure 2.2 presents a mixture density with an outlier. Note that both curves, the normal and the mixture density, are almost equal. The only one difference being that the probability of an extreme event $x \leq -4$ occurring is 0.23‰ for the normal distribution and 1.01% for the mixture of normals (the latter being about 43 times more probable), see Section 2.1.1 and Table 2.1. Thus, if real data were distributed as in Example 6, a "blind" usage of a normal distribution would ignore the possibility of extreme events. Due to the fact that the "worst case" scenario in the financial risk management is of particular interest, the next Section will introduce the Markov switching regime model - the dynamic version of the mixture model.

2.7 Markov switching model

2.7.1 Definition of Markov switching model

Markov switching models were introduced by Hamilton (1989), who modeled quarterly GDP growth rate as

$$y_t = \mu_{z_t} + \phi_1(y_{t-1} - \mu_{z_{t-1}}) + \phi_2(y_{t-2} - \mu_{z_{t-2}}) + \phi_3(y_{t-3} - \mu_{z_{t-3}}) + \phi_4(y_{t-4} - \mu_{z_{t-4}}) + \varepsilon_t \tag{2.65}$$

with $\varepsilon_t \sim N(0, \sigma^2)$. The main idea of the model was that the mean can "shift" between two states, which can be interpreted as the "normal growth phase" with a mean μ_1 and "recession" with mean a μ_2 ($\mu_2 < \mu_1$). As the state of the economy is not observable in the real world, it was modeled with a latent random variable z_t. Hamilton proposed modeling the transition between states as a Markov chain

$$p_{ji} = \Pr[Z_t = j | Z_{t-1} = i], \quad \text{with } \sum_{j=1}^{2} p_{ji} = 1, \text{ and } i \in \{1, 2\}.$$

In this thesis a wider definition will be used.

Definition 2.59 (Markov switching process) *Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space, let $(Z_{t_n})_{t_n \geq 0}$ (where $t_n = n\tau$, $n \in \mathbb{N}$ and τ is a fixed positive number) be a Markov chain with ergodic transition matrix $\boldsymbol{P} = (p_{ji})_{i,j \in \mathcal{K}}$ and state space $\mathcal{K} = \{1, 2, \ldots, K\}$. Furthermore let $(Y_t)_{t \geq 0}$ be a stochastic process* *independent from process* $(Z_{t_n})_{t_n \geq 0}$ *and let*

$$y_{t_n} = \boldsymbol{x}'_{t_n}\boldsymbol{\beta}_j + \varepsilon_{t_n}, \quad \textit{with } \varepsilon_{t_n} \sim N(0, \sigma_j^2), \tag{2.66}$$

where $\boldsymbol{x}'_{t_n} = (1, y_{t_{n-1}}, \ldots, y_{t_{n-r}})$ *denotes the vector of lagged exogenous variables,* $\boldsymbol{\beta}'_j = (\mu_j, \phi_{1(j)}, \ldots, \phi_{r(j)})$ *is the vector of auto-regression coefficients if the state process* Z_{t_n} *is in state* $j \in \mathcal{K}$*,* σ_j^2 *is variance in state* $Z_{t_n} = j$*, and* $r \in \mathbb{N}$ *is the order. The transition probability from past state* i *into current state* j *is given by*

$$p_{ji} = \Pr[Z_{t_n} = j | Z_{t_{n-1}} = i], \quad \textit{with} \quad i, j \in \mathcal{K} \quad \textit{and} \quad \sum_{j \in \mathcal{K}} p_{ji} = 1. \tag{2.67}$$

Consequently the stochastic process (Y_t) *is called Markov switching process, or Hidden Markov process, or regime switching process, equivalently. The matrix* $\boldsymbol{P}$ *is called the transition matrix.*

Notation 2.60 (Markov switching model) *Henceforth, the Markov switching model will be referred to as the MS(m-s)-AR(r) model, where* $m \in \{1, K\}$ *denotes the number of* $\boldsymbol{\beta}$ *vectors,* $s \in \{1, K\}$ *denotes the number of variances and* $r \in \mathbb{N}$ *the order of the auto-regression.*

Example 2.61 *For example MS(1-2)-AR(3) denotes the model*

$$y_{t_n} = \begin{cases} \mu + \phi_1 y_{t_{n-1}} + \phi_2 y_{t_{n-2}} + \phi_3 y_{t_{n-3}} + \varepsilon_{t_n}, & \varepsilon_{t_n} \sim N(0, \sigma_1^2), \qquad \textit{if } z_{t_n} = 1 \\ \mu + \phi_1 y_{t_{n-1}} + \phi_2 y_{t_{n-2}} + \phi_3 y_{t_{n-3}} + \varepsilon_{t_n}, & \varepsilon_{t_n} \sim N(0, \sigma_2^2), \qquad \textit{if } z_{t_n} = 2 \end{cases},$$

MS(3-1)-AR(2) denotes the model

$$y_{t_n} = \begin{cases} \mu_1 + \phi_{1(1)} y_{t_{n-1}} + \phi_{2(1)} y_{t_{n-2}} + \varepsilon_{t_n}, & \varepsilon_{t_n} \sim N(0, \sigma^2), \qquad \textit{if } z_{t_n} = 1 \\ \mu_2 + \phi_{1(2)} y_{t_{n-1}} + \phi_{2(2)} y_{t_{n-2}} + \varepsilon_{t_n}, & \varepsilon_{t_n} \sim N(0, \sigma^2), \qquad \textit{if } z_{t_n} = 2 \\ \mu_3 + \phi_{1(3)} y_{t_{n-1}} + \phi_{2(3)} y_{t_{n-2}} + \varepsilon_{t_n}, & \varepsilon_{t_n} \sim N(0, \sigma^2), \qquad \textit{if } z_{t_n} = 3 \end{cases},$$

MS(3-3)-AR(1) denotes the model

$$y_{t_n} = \begin{cases} \mu_1 + \phi_{1(1)} y_{t_{n-1}} + \varepsilon_{t_n}, & \varepsilon_{t_n} \sim N(0, \sigma_1^2), \qquad \textit{if } z_{t_n} = 1 \\ \mu_2 + \phi_{1(2)} y_{t_{n-1}} + \varepsilon_{t_n}, & \varepsilon_{t_n} \sim N(0, \sigma_2^2), \qquad \textit{if } z_{t_n} = 2 \\ \mu_3 + \phi_{1(3)} y_{t_{n-1}} + \varepsilon_{t_n}, & \varepsilon_{t_n} \sim N(0, \sigma_3^2), \qquad \textit{if } z_{t_n} = 3 \end{cases} \cdot$$

Remark 2.62 *Note that an AR(r) process is a special case of a Markov switching model MS(m-s)-AR(r) with one state (i.e.* $m = s = 1$, $\mathcal{K} = \{1\}$ *and* $\boldsymbol{P} = 1$*).*

Theorem 2.63 *Let* $(Y_t)_{t \geq 0}$ *be a Markov switching process with transition matrix* $\boldsymbol{P}$ *and state variable process* $(Z_{t_n})_{t_n \geq 0}$. *Then the distribution* $\boldsymbol{\pi}$ *of the state variable is unique.*

Proof. From Definition 2.57 it follows that the state variable is distributed by the distribution $\boldsymbol{\pi} = (\Pr[Z_{t_n} = i])_{i \in \mathcal{K}}$. According to Definition 2.59, the transition matrix $\boldsymbol{P} = (\Pr[Z_{t_n} = j | Z_{t_n-1} = i])_{i,j \in \mathcal{K}}$ is ergodic. Thus, as a result from Theorem 2.35, a unique vector $\boldsymbol{\pi} = \boldsymbol{P}'\boldsymbol{\pi}$ exists, which completes the proof. ■

Corollary 2.64 *Note that from Theorem 2.36 and Theorem 2.63 it follows that the stationary distribution of the state variable z is given by*

$$\boldsymbol{\pi} = (\boldsymbol{I} - \boldsymbol{P}' + \boldsymbol{E})^{-1} \mathbf{1}, \tag{2.68}$$

with the special case of

$$\boldsymbol{\pi} = \begin{pmatrix} \frac{1-p_{22}}{2-p_{11}-p_{22}} \\ \frac{1-p_{11}}{2-p_{11}-p_{22}} \end{pmatrix} \tag{2.69}$$

if the number of states equals $K = 2$.

2.7.2 Special cases of Markov switching models

2.7.2.1 Brownian motion with Markov switching

Let us now consider some special cases of the Markov switching model.

Definition 2.65 (GBM with Markov switching) *Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space, let $(Z_{t_n})_{t_n \geq 0}$ (where $t_n = n\tau$, $n \in \mathbb{N}$ and τ is a fixed positive number) be a Markov chain with ergodic transition matrix $\boldsymbol{P}$ and state space $\mathcal{K} = \{1, 2, \ldots, K\}$. Let $(S_t)_{t \geq 0}$ be a stochastic process and $(W_t)_{t \geq 0}$ be a Wiener process, with $\mathcal{F}_t = \sigma\{Z_t, S_t, W_t : \ t \geq 0\}$ being an associated filtration, and $\mu(Z_{t_n})$ and $\sigma(Z_{t_n})$ being associated processes. Then, for each $n \in \mathbb{N}$ the Itô process*

$$S_t = S_{t_0} + \int_{t_0}^{t} \mu(Z_u) S_u du + \int_{t_0}^{t} \sigma(Z_u) S_u dW_u, \qquad \text{for } t \in [t_n, t_{n+1}) \quad (2.70)$$

is called the geometric Brownian motion with Markov switching.

The geometric Brownian motion with Markov switching has the following density function.

Theorem 2.66 (Joint density of GBM with Markov switching) *Let $(S_t)_{t \geq 0}$ be a geometric Brownian motion with Markov switching and let $(Z_{t_n})_{t_n \geq 0}$ (where $t_n = n\tau$, $n \in \mathbb{N}$ and τ is a fixed positive number) be a Markov chain. Then the joint conditional density function $f(S_{t_n}, Z_{t_n} | S_{t_{n-1}} = s, Z_{t_{n-1}} = i)$: $\mathbb{R} \times \mathcal{K} \to \mathbb{R}$ of the pair (S_{t_n}, Z_{t_n}) is given by*

$$\begin{aligned} & f(S_{t_n} = x, Z_{t_n} = j | S_{t_{n-1}} = s, Z_{t_{n-1}} = i) \\ & \qquad = \frac{p_{ji}}{\sqrt{2\pi}\sigma_j \sqrt{\tau} x} \exp\left[-\frac{1}{2} \left(\frac{\ln \frac{x}{s} - (\mu_j - \frac{1}{2}\sigma_j^2)\tau}{\sigma_j \sqrt{\tau}} \right)^2 \right] \end{aligned}$$

for $x > 0$ (Webb 2003, p. 19).

Proof. Let g be a measurable a Borel function. Note that

$$\begin{aligned}
&\mathbb{E}[g(S_{t_n}, Z_{t_n})|S_{t_{n-1}} = s, Z_{t_{n-1}} = i] \\
&= \sum_{j\in\mathcal{K}} \mathbb{E}[g(S_{t_n}, Z_{t_n})|S_{t_{n-1}} = s, Z_{t_{n-1}} = i, Z_{t_n} = j] \times \Pr[Z_{t_n} = j|Z_{t_{n-1}} = i] \\
&\qquad = \sum_{j\in\mathcal{K}} p_{ji}\mathbb{E}\left[g\left(se^{(\mu_j-\frac{1}{2}\sigma_j^2)\tau+\sigma_j(W_{t_n}-W_{t_{n-1}})}, j\right)\right].
\end{aligned}$$

One has to keep in mind that $\frac{(W_{t_n}-W_{t_{n-1}})}{\sqrt{\tau}}$ is standard normally distributed and that the expectation equals

$$\sum_{j\in\mathcal{K}} p_{ji}\int_{-\infty}^{\infty} g\left(se^{(\mu_j-\frac{1}{2}\sigma_j^2)\tau+\sigma_j(W_{t_n}-W_{t_{n-1}})}, j\right)\frac{1}{\sqrt{2\pi}}e^{\frac{-z^2}{2}}dz.$$

Now substitute $x = se^{(\mu_j-\frac{1}{2}\sigma_j^2)\tau+\sigma_j(W_{t_n}-W_{t_{n-1}})}$, which yields

$$\sum_{j\in\mathcal{K}} p_{ji}\int_{0}^{\infty} g(x,j)p_{ji}\frac{1}{\sqrt{2\pi}\sigma_j\sqrt{\tau}x}\exp\left[-\frac{1}{2}\left(\frac{\ln\frac{x}{s}-(\mu_j-\sigma_j^2)\tau}{\sigma_j\sqrt{\tau}}\right)^2\right].$$

By choosing $g(S_{t_n}, Z_{t_n}) = \mathbb{I}_{[S_{t_n}\le s'], Z_{t_n}=j'}$ the proof is completed (Webb 2003, p. 19-20). ■

The geometric Brownian motion with Markov switching has the following density function.

Theorem 2.67 (Density of GBM with Markov switching) *Let $(S_t)_{t\ge0}$ be a geometric Brownian motion with Markov switching and let $(Z_{t_n})_{t_n\ge0}$ (where $t_n = n\tau$, $n \in \mathbb{N}$ and τ is a fixed positive number) be a Markov chain, then the conditional density function $f(S_{t_n}|S_{t_{n-1}} = s, Z_{t_{n-1}} = i) : \mathbb{R}\times\mathcal{K} \to \mathbb{R}$ of the pair (S_{t_n}, Z_{t_n}) is given by*

$$\begin{aligned}
&f(S_{t_n} = x|S_{t_{n-1}} = s, Z_{t_{n-1}} = i) \\
&\qquad = \sum_{j\in\mathcal{K}} \frac{p_{ji}}{\sqrt{2\pi}\sigma_j\sqrt{\tau}x}\exp\left[-\frac{1}{2}\left(\frac{\ln\frac{x}{s}-(\mu_j-\frac{1}{2}\sigma_j^2)\tau}{\sigma_j\sqrt{\tau}}\right)^2\right]
\end{aligned}$$

for $x > 0$ (Webb 2003, p. 20).

Proof. The proof is similar to the proof of Theorem 2.66. ■

2.7.2.2 Vasiček model with Markov switching

The Vasiček process with Markov switching is defined in analogy to the GBM with Markov switching and takes the following form.

Definition 2.68 (Vasiček with Markov switching) *Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space, let $(Z_{t_n})_{t_n \geq 0}$ (where $t_n = n\tau$, $n \in \mathbb{N}$ and τ is a fixed positive number) be a Markov chain with ergodic transition matrix $\boldsymbol{P}$ and state space $\mathcal{K} = \{1, 2, \ldots, K\}$. Let $(R_t)_{t \geq 0}$ be a stochastic process and $(W_t)_{t \geq 0}$ be a Wiener process, with $\mathcal{F}_t = \sigma\{Z_t, R_t, W_t : \ t \geq 0\}$ being an associated filtration, and $\mu(Z_{t_n})$ and $\sigma(Z_{t_n})$ being associated processes. Then for each $n \in \mathbb{N}$ the Itô process*

$$R_t = R_{t_0} + \int_{t_0}^{t} \alpha[\mu(Z_u) - R_u]du + \int_{t_0}^{t} \sigma(Z_u)dW_u, \qquad \textit{for } t \in [t_n, t_{n+1}) \tag{2.71}$$

is called the Vasiček process with Markov switching.

In Section 2.4.2 we have shown how to discretize the Vasiček model to an AR(1) model. Analogously, the Vasiček process with Markov switching can be descretized to a an AR(1) model with Markov switching:

$$\begin{aligned} R_{t_n} &= c(Z_{t_n} = j) + \phi(Z_{t_n} = j)R_{t_{n-1}} + \varepsilon_{t_n}, \\ & \qquad \varepsilon_{t_n} \sim \mathcal{N}\left(0, \frac{\sigma(Z_{t_n} = j)^2}{2\alpha(Z_{t_n} = j)}\left(1 - e^{-2\alpha(Z_{t_n}=j)}\right)\right), \end{aligned} \tag{2.72}$$

$$p_{ji} = \Pr[Z_{t_n} = j | Z_{t_{n-1}} = i], \quad 0 \leq p_{ji} \leq 1, \quad \sum_{j=1}^{K} p_{ji} = 1, \quad \forall i, j = 1, \ldots, K, \tag{2.73}$$

with constants $c(Z_{t_n} = j) = \mu(Z_{t_n} = j)(1 - e^{-\alpha(Z_{t_n}=j)})$ and autoregressive coefficients $\phi(Z_{t_n} = j) = e^{-\alpha(Z_{t_n}=j)}$ (Hamilton 1990, p. 43).

2.8 Estimation of the Markov switching model

2.8.1 Log-likelihood function

The Markov switching model can be estimated with the maximum likelihood estimation (MLE) method.

The estimation of parameters of the model with MLE relies on the maximization of the log-likelihood function. The parameters are estimated if the first derivative (gradient) of the log-likelihood function is equal to a vector of 0s

$$\frac{\partial \mathscr{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}. \tag{2.74}$$

Thus, one has to determine the log-likelihood function:

$$\begin{aligned}
&\mathscr{L}(Y;\boldsymbol{\theta}) = \ln(f;\boldsymbol{\theta})\\
&= \sum_{t_n=1}^{T} \ln\left(\sum_{j=1}^{K}\sum_{i=1}^{K} f(\boldsymbol{y}_{t_n}|Z_{t_n}=j, Z_{t_{n-1}}=i, \mathscr{Y}_{t_{n-1}}) \Pr[Z_{t_n}=j, Z_{t_{n-1}}=i, \mathscr{Y}_{t_{n-1}}]\right)
\end{aligned} \tag{2.75}$$

(Kim and Nelson 1999, p. 65). The first term on the right-hand side is given by

$$\begin{aligned}
&f(\boldsymbol{y}_{t_n}|Z_{t_n}=j, Z_{t_{n-1}}=i, \mathscr{Y}_{t_{n-1}})\\
&\quad= \frac{1}{(2\pi)^{n/2}\det(\boldsymbol{\Sigma}_j)^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{y}_{t_n} - \boldsymbol{x}'_{t_n}\boldsymbol{\beta}_{t_n})'(\boldsymbol{\Sigma}_j)^{-1}(\boldsymbol{y}_{t_n} - \boldsymbol{x}'_{t_n}\boldsymbol{\beta}_{t_n})\right\}
\end{aligned} \tag{2.76}$$

and the second term is the joint distribution of z_{t_n} and $z_{t_{n-1}}$.

Unfortunately, the state variable Z cannot be observed. Thus, the classical version of the MLE algorithm cannot be applied. Instead, Hamilton (1989) proposed using the EM algorithm. In Section 2.8.3 we introduce it, but first we will discuss how to deal with the problem of estimating the latent state variable Z.

2.8.2 Inference about the unobservable state variable

2.8.2.1 Filter of Hamilton (1989)

Hamilton (1989) proposed inferring joint probabilities $\Pr[z_{t_n}, z_{t_{n-1}}]$ from information contained in the observed history $\mathscr{Y}_{t_n} = \{y_{t_1}, y_{t_2} \dots, y_{t_n}\}$.

Algorithm 2.69 (Filter of Hamilton 1989)

F1. At the beginning of period t_n, compute for all $i, j = 1, \dots, K$ the joint probability given for the past information set $\mathscr{Y}_{t_{n-1}}$: $\Pr[Z_{t_n} = j, Z_{t_{n-1}} = i|\mathscr{Y}_{t_{n-1}}]$ from filtered probabilities $\Pr[Z_{t_{n-1}} = i|\mathscr{Y}_{t_{n-1}}]$

$$\Pr[Z_{t_n} = j, Z_{t_{n-1}} = i|\mathscr{Y}_{t_{n-1}}] = \Pr[Z_{t_{n-1}} = i|\mathscr{Y}_{t_{n-1}}] \Pr[Z_{t_n} = j|Z_{t_{n-1}} = i], \tag{2.77}$$

where $\Pr[Z_{t_n} = j|Z_{t_{n-1}} = i]$ denotes the transition probability from equation (2.67) and filtered probabilities $\Pr[Z_{t_{n-1}} = i|\mathscr{Y}_{t_{n-1}}]$ are given from the previous iteration t_{n-1} from step F4.

F2. Compute the marginal density

$$\begin{aligned} f(\boldsymbol{y}_{t_n}|\mathscr{Y}_{t_{n-1}}) &= \sum_{j=1}^{K}\sum_{i=1}^{K} f(\boldsymbol{y}_{t_n}, Z_{t_n} = j, Z_{t_{n-1}} = i|\mathscr{Y}_{t_{n-1}}) \\ &= \sum_{j=1}^{K}\sum_{i=1}^{K} f(\boldsymbol{y}_{t_n}|Z_{t_n} = j, Z_{t_{n-1}} = i, \mathscr{Y}_{t_{n-1}}) \\ &\qquad \times \ \Pr[Z_{t_n} = j, Z_{t_{n-1}} = i|\mathscr{Y}_{t_{n-1}}]. \end{aligned} \tag{2.78}$$

F3. Now the information included in the present observation $\boldsymbol{y}_{t_n}$ can be taken into account. Compute

$$
\begin{aligned}
&\Pr[Z_{t_n}=j, Z_{t_{n-1}}=i|\mathscr{Y}_{t_n}] = \Pr[Z_{t_n}=j, Z_{t_{n-1}}=i|\boldsymbol{y}_{t_n}, \mathscr{Y}_{t_{n-1}}] \\
&= \frac{f(Z_{t_n}=j, Z_{t_{n-1}}=i, \boldsymbol{y}_{t_n}|\mathscr{Y}_{t_{n-1}})}{f(\boldsymbol{y}_{t_n}|\mathscr{Y}_{t_{n-1}})} \\
&= \frac{f(\boldsymbol{y}_{t_n}|Z_{t_n}=j, Z_{t_{n-1}}=i, \mathscr{Y}_{t_{n-1}}) \Pr[Z_{t_n}=j, Z_{t_{n-1}}=i|\mathscr{Y}_{t_{n-1}}]}{\sum_{j=1}^{K}\sum_{i=1}^{K} f(\boldsymbol{y}_{t_n}|Z_{t_n}=j, Z_{t_{n-1}}=i, \mathscr{Y}_{t_{n-1}}) \Pr[Z_{t_n}=j, Z_{t_{n-1}}=i|\mathscr{Y}_{t_{n-1}}]}
\end{aligned}
\tag{2.79}
$$

for all $i, j = 1, \ldots, K$.

F4. Lastly, for all $j = 1, \ldots, K$ *compute the filtered unconditional probability, which will be the input for the next iteration* t_{n+1} *(step F1)*

$$
\Pr[Z_{t_n}=j|\mathscr{Y}_{t_n}] = \sum_{i=1}^{K} \Pr[Z_{t_n}=j, Z_{t_{n-1}}=i|\mathscr{Y}_{t_n}]. \tag{2.80}
$$

F5. Repeat steps F1-F4 for all observations $t_n = t_1, t_2, \ldots, T$.

As input for the first iteration (in t_1*), use the stationary probabilities from equation (2.68) as the filtered probabilities* $\Pr[Z_{t_1}|\mathscr{Y}_{t_1}]$ *(Kim and Nelson 1999, p. 66-68).*

2.8.2.2 Smoother of Kim (1994)

The filtered joint probabilities $\Pr[z_{t_n}, z_{t_{n-1}}|\mathscr{Y}_{t_n}]$ are based on all information available until time t_n (i.e., $y_{t_1}, \ldots, y_{t_n}$), but not on the information of the full sample (i.e., $y_{t_{n+1}}, \ldots, y_T$). The full information set $\mathscr{Y}_T$ can be used by smoothing the probabilities. The smoothing algorithm of Kim (1994)[10] gives the following inferences of the probabilities based on the whole information set $\mathscr{Y}_T$.

[10] Which is an improvement of the smoother proposed by Hamilton (1989) and Lam (1990).

Algorithm 2.70 (Smoother of Kim 1994)

S1. *At time t_n, compute for all $j, k = 1, \ldots, K$ the joint probabilities that the model is in the j-th state in current period ($Z_{t_n} = j$) and in the k-th state in the next period ($Z_{t_{n+1}} = k$) given the whole information set $\mathscr{Y}_T$*

$$
\begin{aligned}
&\Pr[Z_{t_n} = j, Z_{t_{n+1}} = k|\mathscr{Y}_T] \\
&= \Pr[Z_{t_{n+1}} = k|\mathscr{Y}_T] \times \Pr[Z_{t_n} = j|Z_{t_{n+1}} = k, \mathscr{Y}_T] \\
&= \Pr[Z_{t_{n+1}} = k|\mathscr{Y}_T] \times \Pr[Z_{t_n} = j|Z_{t_{n+1}} = k, \mathscr{Y}_{t_n}] \\
&= \frac{\Pr[Z_{t_{n+1}} = k|\mathscr{Y}_T] \times \Pr[Z_{t_n} = j, Z_{t_{n+1}} = k|\mathscr{Y}_{t_n}]}{\Pr[Z_{t_{n+1}} = k|\mathscr{Y}_{t_n}]} \\
&= \frac{\Pr[Z_{t_{n+1}} = k|\mathscr{Y}_T] \times \Pr[Z_{t_n} = j|\mathscr{Y}_{t_n}] \times \Pr[Z_{t_{n+1}} = k|Z_{t_n} = j]}{\Pr[Z_{t_{n+1}} = k|\mathscr{Y}_{t_n}]}
\end{aligned}
\tag{2.81}
$$

whereas $\Pr[Z_{t_{n+1}} = k|\mathscr{Y}_T]$ *was determined in the previous iteration (step S2) at time t_{n+1},* $\Pr[Z_{t_n} = j|\mathscr{Y}_{t_n}]$ *and* $\Pr[Z_{t_{n+1}} = k|\mathscr{Y}_{t_n}]$ *are the filtered probabilities from equation (2.80) (step F4), and* $\Pr[Z_{t_{n+1}} = k|Z_{t_n} = j]$ *is the transition probability from equation (2.67).*

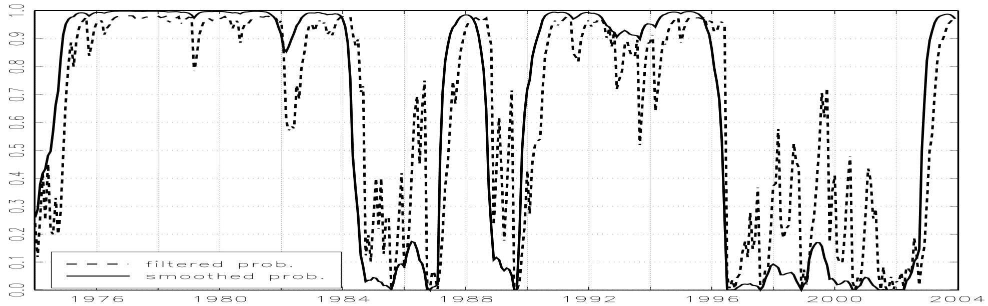
S2. *Now compute the smoothed unconditional probability for all $j = 1, \ldots, K$ which will be the input for the next iteration t_{n+1} (step S1)*

$$\Pr[Z_{t_n} = j|\mathscr{Y}_T] = \sum_{k=1}^{K} \Pr[Z_{t_n} = j, Z_{t_{n+1}} = k|\mathscr{Y}_T] \tag{2.82}$$

S3. *Iterate steps S1-S2 backwards for $t_n = T - 1, \ldots, 1$ (Kim and Nelson 1999, p. 68-70).*

The difference between filtered and smoothed probabilities is depicted in Figure 2.3. It is straightforward to see that the filtered probability curve

Figure 2.3: Filtered and smoothed probabilities: REXP(0%)-DAX30(100%) - MS(1-2)



Note:
The dashed line depicts the filtered probabilities $\Pr[Z_{t_n} = 1|\mathscr{Y}_{t_n}]$ and the solid line the smoothed probabilities $\Pr[Z_{t_n} = 1|\mathscr{Y}_T]$ for the low volatility regime. These probabilities are a product of the EM algorithm; produced by estimating the MS(1-2) model for the DAX30 log-returns.

(dashed line) has a more irregular shape than the smoothed probability curve (solid line). The reason is that to estimate smoothed probabilities one uses more information, i.e. the whole information set $\mathscr{Y}_T$, than to estimate filtered probabilities, i.e., only information $\mathscr{Y}_{t_n}$ available up to the estimation time point t_n.

2.8.3 EM Algorithm

Having shown how to compute probabilities of the unobserved state variable, we can now describe the EM algorithm introduced by Dempster, Laird, and Rubin (1977) to estimate parameters of time series with unobserved variables (or missing observations). The EM algorithm consists of two steps which are repeated until the estimated parameter vector $\hat{\boldsymbol{\theta}}$ converges to the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_{MLE}$

$$\lim_{l \to \infty} \hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_{MLE}. \tag{2.83}$$

Expectation step Establishment of expected realizations of unobserved variable $\widehat{\mathscr{Z}_{T,l}} = \{\widehat{z}_{t_1,l}, \widehat{z}_{t_2,l}, \dots, \widehat{z}_{T,l}\}$ conditional on the estimated parameter vector $\hat{\boldsymbol{\theta}}_{l-1}$ from the last iteration $l-1$.

Maximization step Maximizing the log-likelihood function with respect to parameters of the model ($\hat{\boldsymbol{\theta}}_l$) gives the expected realizations of the unobserved state variable Z ($\widehat{\mathscr{Z}_{T,l}}$) obtained from the expectation step.

Now the particular version of the EM algorithm developed by Hamilton (1990) – for a general case of an r-dimensional MS model of the Kth order

$$\boldsymbol{y}_{t_n} = \begin{pmatrix} \boldsymbol{x}'_{1,t_n}\boldsymbol{\beta}_{1,j} \\ \cdots \\ \boldsymbol{x}'_{r,t_n}\boldsymbol{\beta}_{r,j} \end{pmatrix} + \boldsymbol{\varepsilon}_{t_n}, \quad \text{with } \boldsymbol{\varepsilon}_{t_n} \sim N(0, \boldsymbol{\Omega}_j), \tag{2.84}$$

$$p_{ji} = \Pr[Z_{t_n} = j | Z_{t_{n-1}} = i], \quad \text{with} \quad i, j \in \mathcal{K} \quad \text{and} \quad \sum_{j \in \mathcal{K}} p_{ji} = 1, \tag{2.85}$$

where $\boldsymbol{x}'_{k,t_n} = (1, y_{t_{n-1}}, \dots, y_{t_{n-p_k}})$ denotes the vector of lagged exogenous variables, $\boldsymbol{\beta}'_{k,j} = (\mu_j, \phi_{k,1(j)}, \dots, \phi_{k,p_k(j)})$ is the vector of auto-regression coefficients if the state process Z_{t_n} is in state $j \in \mathcal{K}$, Ω_j is the variance matrix in state $Z_{t_n} = j$, and $p_k \in \mathbb{N}$ is the order of the k-th dimension ($k = 1, \dots, r$) – can be introduced:

Algorithm 2.71 (EM algorithm for the MS models)

EM1. ***Expectation step***

EM1.1. Compute the filtered probabilities as in F1-F5 for $t_n = t_1, t_2, \dots, T$.

EM1.2. Compute the smoothed probabilities as in S1-S3, for $t_n = T - 1, \dots, t_1$.

For the first iteration $l = 1$ take the arbitrary initial guesses of the parameter vector $\hat{\boldsymbol{\theta}}_0$. For all the following iterations $l = 2, \dots$ use the output of the maximization step EM2 as the parameter vector $\hat{\boldsymbol{\theta}}_{l-1}$.

EM2. ***Maximization step***

EM2.1. Compute the transition probabilities[11]

$$p^l_{ji} = \begin{cases} \frac{\sum_{t_n=2+p_{max}}^{T} \Pr\left[Z_{t_n}=j, Z_{t_{n-1}}=i | \mathscr{Y}_T; \hat{\boldsymbol{\theta}}_{l-1}\right]}{\sum_{t_n=2+p_{max}}^{T} \Pr\left[Z_{t_{n-1}}=i | \mathscr{Y}_T; \hat{\boldsymbol{\theta}}_{l-1}\right]} & \text{if } j = 1, \ldots, K-1, \\ & \quad i = 1, \ldots, K \\ 1 - \sum_{j=1}^{K-1} p^l_{ji} & \text{if } j = K, \; i = 1, \ldots, K. \end{cases} \tag{2.86}$$

EM2.2. for each dimension k ($k = 1, \ldots, r$) compute the betas

$$\begin{aligned} \boldsymbol{\beta}^l_{k,j} = & \left(\sum_{t_n=1+p_{max}}^{T} \boldsymbol{x}_{k,t_n} \boldsymbol{x}'_{k,t_n} \Pr[Z_{t_n} = j | \mathscr{Y}_T; \hat{\boldsymbol{\theta}}_{l-1}] \right)^{-1} \\ & \times \left(\sum_{t_n=1+p_{max}}^{T} \boldsymbol{x}_{k,t_n} \boldsymbol{y}'_{k,t_n} \Pr[Z_{t_n} = j | \mathscr{Y}_T; \hat{\boldsymbol{\theta}}_{l-1}] \right) \end{aligned} \tag{2.87}$$

if $m_k = K$ and $p_k > 1$ or

$$\boldsymbol{\beta}^l_{k,j} = \boldsymbol{\mu}^l_{k,j} = \frac{\sum_{t_n=1+p_{max}}^{T} \boldsymbol{y}_{k,t_n} \boldsymbol{y}'_{k,t_n} \Pr\left[Z_{t_n} = j | \mathscr{Y}_T; \hat{\boldsymbol{\theta}}_{l-1}\right]}{\sum_{t_n=1+p_{max}}^{T} \Pr\left[Z_{t_n} = j | \mathscr{Y}_T; \hat{\boldsymbol{\theta}}_{l-1}\right]} \tag{2.88}$$

if $m_k = K$ and $p_k = 1$ or

$$\begin{aligned} \boldsymbol{\beta}^l_{j} = \sum_{j=1}^{K} & \left(\sum_{t_n=1+p_{max}}^{T} \boldsymbol{x}_{k,t_n} \boldsymbol{x}'_{k,t_n} \Pr[Z_{t_n} = j | \mathscr{Y}_T; \hat{\boldsymbol{\theta}}_{l-1}] \right)^{-1} \\ & \times \left(\sum_{t_n=1+p_{max}}^{T} \boldsymbol{x}_{k,t_n} \boldsymbol{y}'_{k,t_n} \Pr[Z_{t_n} = j | \mathscr{Y}_T; \hat{\boldsymbol{\theta}}_{l-1}] \right) \end{aligned} \tag{2.89}$$

if $m_k = 1$ and $p_k > 1$ or

$$\boldsymbol{\beta}^l_{j} = \boldsymbol{\mu}^l_{j} = \sum_{j=1}^{K} \frac{\sum_{t_n=1+p_{max}}^{T} \boldsymbol{y}_{k,t_n} \boldsymbol{y}'_{k,t_n} \Pr[Z_{t_n} = j | \mathscr{Y}_T; \hat{\boldsymbol{\theta}}_{l-1}]}{T - p_{max}} \tag{2.90}$$

[11] Usually, for K=2 probabilities p^l_{11} and p^l_{22} are computed instead of p^l_{11} and p^l_{12}. The computation of $p^l_{21} = 1 - p^l_{11}$ and $p^l_{12} = 1 - p^l_{22}$ is thus enabled.

if $m_k = 1$ and $p_k = 1$. Where m_k indicates whether the beta vector is dependent on the regime ($m_k = K$) or not ($m_k = 1$), p_k denotes the order of the AR(p_k)-process in the k-th dimension, and p_{max} the highest order among all dimensions.

EM2.3. Compute the variances

$$\boldsymbol{\Omega}_j^l = \frac{\sum_{t_n=1+p_{max}}^{T} (\boldsymbol{y}_{t_n} - \boldsymbol{\varepsilon}_{t_n}^{l-1})(\boldsymbol{y}_{t_n} - \boldsymbol{\varepsilon}_{t_n}^{l-1})' \Pr\left[Z_{t_n} = j|\mathscr{Y}_T; \hat{\boldsymbol{\theta}}_{l-1}\right]}{\sum_{t_n=1+p_{max}}^{T} \Pr\left[Z_{t_n} = j|\mathscr{Y}_T; \hat{\boldsymbol{\theta}}_{l-1}\right]} \tag{2.91}$$

for $s = K$ or

$$\boldsymbol{\Omega}^l = \frac{\sum_{t_n=1+p_{max}}^{T} \sum_{j=1}^{K} (\boldsymbol{y}_{t_n} - \boldsymbol{\varepsilon}_{t_n}^{l-1})(\boldsymbol{y}_{t_n} - \boldsymbol{\varepsilon}_{t_n}^{l-1})' \Pr\left[Z_{t_n} = j|\mathscr{Y}_T; \hat{\boldsymbol{\theta}}_{l-1}\right]}{T - p_{max}} \tag{2.92}$$

for $s = 1$. Where s indicates whether the variance matrix is dependent on the regime ($s = K$) or not ($s = 1$) (Hamilton 1990, section 4).

2.9 Empirical results of the MS

2.9.1 DAX-REXP portfolios

At the end of this Chapter, several Markov switching models will be estimated. For this purpose, 13 portfolios were constructed. All portfolios consist of German bonds and stocks. As the proxy for stocks the German Stock Index (DAX30) – a German blue chip index – was used. The bond portfolio was proxied with the German Bond Performance Index (REXP) – a synthetic index for German state bonds. The portfolios were constructed as follows. It was assumed that on 31/12/1974 DEM 195.58 (= €100) was invested in such a way that $x\%$ of the sum was invested in the REXP index and (100-x)% in the DAX30 index (with x=0,10,20,25,30,40,50,60,70,75,80,90,100). The portfolios were held until 31/12/2004 and then sold. From the value of the portfolios, monthly log-returns were computed and the MS models were estimated.

2.9.2 Description of the estimation

For all thirteen portfolios six Markov switching models of the second order were estimated: the heteroskedastic models with regime independent mean equation (i.e. MS(1-2)-type models), homoskedastic models with regime dependent mean equation (i.e. MS(2-1)-type models), and heteroskedastic models with regime dependent mean equation (i.e. MS(2-2)-type models). For all cases, a variant with and without the auto-regression term was estimated. Additionally, the geometric Brownian motion (GBM) and autoregressive model of the first order (AR(1)) were estimated for comparison. The linear models were estimated with the maximum likelihood estimation method. The Markov switching models were estimated using the EM algorithm described in Section 2.8.3. As the EM algorithm does not ensure that the estimated parameter vector lies on the global maximum, the estimation was repeated 200 times for each estimated model and time series. This approach was used inter alia by Rydén, Teräsvirta, and Åsbrink (1998). The initial guess was randomly drawn from the distribution $U(a, b)$, where $U(\cdot, \cdot)$ denotes a uniform distribution over the (a, b) interval. The mean parameters μ_j were drawn from $U(-3\widehat{\mu}, 3\widehat{\mu})$, the variance parameters σ_j^2 were drawn from $U(0, 9\widehat{\sigma}^2)$, the auto-regression parameters $\phi_{1(j)}$ were drawn from $U(-1, 1)$ and the transition probabilities p_{ii} were drawn from $U(0.5, 1)$, where $\widehat{\mu}$ and $\widehat{\sigma}^2$ denote the empirical mean and variance of the estimated sample, respectively. In each estimation iteration, it was tested whether the probability parameters $p_{ji} \in [0.0001, 0.9999]$ and whether the variance parameters $\sigma_j^2 \geq 10^{-10}$ to prevent a collapse of the algorithm. If one of these two boundary restrictions was violated, the value of the parameter was set to be equal to the boundary condition. Each EM algorithm was iterated until the increase of the log-likelihood function fell short of 10^{-8}. After running the EM algorithm 200 times the results were controlled for anomalies. The behavior of smoothed probabilities was tested in particular. If, for the entire sample, the estimated smoothed probabilities were equal to 1 for a particular state (i.e. if $\Pr[Z_{t_n} = j|\mathscr{Y}_T] = 1, \ \forall\ t = 1, \ldots, T, \ \exists\ j \in \mathcal{K}$) the estimation parameter was rejected, due to the over-parametrization.[12] Then, from the remaining estimation runs, the one with the highest log-likelihood function was chosen as being the closest to the true parameter vector.

The procedure was repeated for the models with the restriction $p_{11} = 1 - p_{22}$, as this was required for the tests, see Section 3.4.2.

2.9.3 Estimation results

Tables B.1-B.13 from Appendix B show results of the estimation for all thirteen portfolios.

The parameters were ordered in two states. In the case, of heteroskedastic models, the first state was defined as the "low volatility" state and the second state was defined as the "high volatility" state (i.e. $\sigma_1^2 < \sigma_2^2$). In the homoskedastic case the first state was defined as the "high mean" state and

[12] This means that the state variable was with certainty in one state through the whale sample. Thus, the algorithm yielded an arithmetic Brownian motion (or Vasiček model) without Markov switching and with redundant parameters from the second regime and a redundant transition matrix.

Table 2.2: Significance of the MS parameter (1.1975-12.2004)

		MS(1-1)		**MS(1-2)**		**MS(2-1)**		**MS(2-2)**	
		GBM	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
0%	100%	++		++		++		+	
10%	90%	++		++		++		+	
20%	80%	++		++		++		+	
25%	75%	++		++		++		+	
30%	70%	++		++		++		+	
40%	60%	++		++				+	
50%	50%	++		++				+	
60%	40%	++		++				+	
70%	30%	++		++				+	
75%	25%	++		++				++	
80%	20%	++		++				++	
90%	10%	++	++	++	++			+	
100%	0%	++	++	++	++			+	

Note:
MS(m-s) stands for a Markov switching model with m mean equations and s regimes for the variance. GBM denotes no auto-regression in the mean equation(s) and AR(1) an auto-regression of first order in each mean equation. ++ signifies that all parameters are significantly different from zero at minimum 5% significance level, + denotes that all parameters are significantly different from zero at minimum 5% significance level, excluding the μ_2. These models are treated as fully significant models with parameter $\mu_2 = 0$.

the second state as the "low mean" state (i.e. $\mu_1 > \mu_2$). In the case of the MS(2-2)-type models, the second state is almost always not only "high variance" state but also "low mean". The only exceptions were portfolios with a 75% and 80% bond proportion, where the second state is a "high mean", "high variance" state.

Table 2.2 shows in which models and in which portfolios all parameters are different from zero at the 5% significance level (the two-side t-test was used). This holds true for the GBM model and the MS(1-2) model for all portfolios. For the AR(1) and MS(1-2)-AR(1), all parameters are significantely different from zero in two cases: for the portfolio with a 90% bond proportion and for the pure bond portfolio. For the MS(2-1) model, parameters are non zero for all portfolios with a maximum of 30% stock engagement. For the MS(2-2) model, the parameters are significantly different from zero for a 75% and an

Table 2.3: Time intervals with a high volatility state for the MS(1-2) model

Portfolio composition REXP	DAX30	Periods with a high volatility regime
0%	100%	01/1975-07/1975, 05/1985-03/1988, 10/1989-10/1990, 06/1997-11/2003
10%	90%	01/1975-07/1975, 05/1985-03/1988, 10/1989-11/1990, 06/1997-10/2003
20%	80%	01/1975-07/1975, 05/1985-03/1988, 10/1989-11/1990, 06/1997-10/2003
25%	75%	01/1975-06/1975, 05/1985-03/1988, 10/1989-11/1990, 06/1997-10/2003
30%	70%	01/1975-06/1975, 05/1985-03/1988, 10/1989-11/1990, 06/1997-10/2003
40%	60%	01/1975-05/1975, 05/1985-03/1988, 10/1989-11/1990, 06/1997-10/2003
50%	50%	01/1975-05/1975, 05/1985-03/1988, 10/1989-11/1990, 06/1997-09/2003
60%	40%	01/1975-05/1975, 03/1980-04/1980, 02/1983-05/1983, 05/1985-03/1988, 10/1989-12/1990, 12/1993-01/1994, 05/1997-04/2000, 11/2000-07/2003
70%	30%	01/1975-07/1975, 09/1979-07/1981, 11/1982-08/1983, 05/1985-04/1988, 09/1989-03/1991, 06/1993-09/1994, 05/1997-04/2000, 01/2001-07/2003
75%	25%	01/1975-05/1977, 01/1979-12/2003
80%	20%	01/1975-05/1977, 01/1979-08/2003
90%	10%	01/1975-02/1975, 08/1979-03/1983, 11/1989-10/1990
100%	0%	10/1979-11/1982

Note:
The table shows periods in which the MS(1-2) model features a high volatility state (i.e. periods with the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

80% bond portfolio. For the MS(2-1)-AR(1) and the MS(2-2)-AR(1) models, there are no portfolios with all parameters significantly different from zero.

If one relaxes the assumption that all parameters have to be different from zero and allows one of the states to have a zero intercept, then by all portfolios pass this test for the MS(2-2) model and, obviously, for all the other cases mentioned above. This is a plausible assumption. Merely testing if one of the other parameters equals zero does not make sense. If the auto-regression parameter is zero, than the auto-regression model should be rejected. Likewise, if the variance or one of the transition probabilities equals zero, then the estimations constraints (see Section 2.9.2) have been violated and this model should also be rejected. Quite contrary to the case with one intercept that equals zero, which makes economic sense: It can be interpreted, that in one state the value of the portfolio grows (or falls) and in the second state, it is expected to stay unchanged.

Figures C.1-C.13 from Appendix C show in which state the price process was in the period from 1975-2004. A first glance at the figures shows, that

Table 2.4: Time intervals with a high volatility state for the MS(1-2)-AR(1) model

Portfolio composition		Periods with a high volatility regime
REXP	**DAX30**	
0%	100%	05/1985-03/1988, 10/1989-11/1990, 06/1997-10/2003
10%	90%	05/1985-03/1988, 10/1989-11/1990, 06/1997-10/2003
20%	80%	05/1985-03/1988, 10/1989-11/1990, 06/1997-10/2003
25%	75%	05/1985-03/1988, 10/1989-11/1990, 06/1997-10/2003
30%	70%	05/1985-03/1988, 10/1989-11/1990, 06/1997-10/2003
40%	60%	05/1985-03/1988, 10/1989-11/1990, 06/1997-10/2003
50%	50%	05/1985-03/1988, 10/1989-12/1990, 06/1997-09/2003
60%	40%	03/1980-04/1980, 06/1985-03/1988, 10/1989-12/1990, 05/1997-04/2000, 11/2000-07/2003
70%	30%	02/1975, 10/1979-07/1981, 12/1982-07/1983, 05/1985-04/1988, 09/1989-03/1991, 07/1993-09/1994, 05/1997-04/2000, 01/2001-07/2003
75%	25%	02/1975-05/1977, 01/1979-12/2003
80%	20%	02/1975-05/1977, 01/1979-08/2003
90%	10%	09/1979-05/1983, 11/1989-11/1990
100%	0%	10/1979-10/1982

Note:
The table shows periods in which the MS(1-2)-AR(1) model features a high volatility state (i.e. periods with the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

the MS(1-2), MS(1-2)-AR(1), MS(2-2), and MS(2-2)-AR(1) models behave similarly for all portfolios. For portfolios with a bond proportion between 0% and 50%, the high volatility regime (i.e. periods with the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$) occurs almost in the same periods. The MS(1-2) and the MS(2-2) model feature the second state at the begining of 1975. The high volatility period lasts from the begining of the year until February or March, in the case of the MS(2-2) model, or until May, June or July in the case of the MS(1-2) model (see Tables 2.3 and 2.5 for details). All heteroscedastic models feature the high voaltility state from mid-1985 to March 1988, from October 1989 to the end of 1990, and from June 1997 to mid/end 2003 (see Tables 2.3-2.6 for details). It is straightforward to see that these periods cover the naked-eye-observation of the DAX30 time series made in Section 2.1.2. This shows that portfolios mentioned are dominated by the stock price effects.

For portfolios with a 60% and a 70% bond proportion, there are more (mostly short) periods with a high volatility state. These are periods that

Table 2.5: Time intervals with a high volatility state for the MS(2-2) model

Portfolio composition REXP	DAX30	Periods with a high volatility regime
0%	100%	01/1975-02/1975, 06/1985-03/1988, 10/1989-10/1990, 06/1997-10/2003
10%	90%	01/1975-02/1975, 06/1985-3/1988, 10/1989-10/1990, 06/1997-10/2003
20%	80%	01/1975-03/1975, 06/1985-03/1988, 10/1989-10/1990, 06/1997-10/2003
25%	75%	01/1975-03/1975, 06/1985-03/1988, 10/1989-11/1990, 06/1997-10/2003
30%	70%	01/1975-03/1975, 06/1985-03/1988, 10/1989-11/1990, 06/1997-10/2003
40%	60%	01/1975-03/1975, 07/1985-03/1988, 10/1989-11/1990, 06/1997-10/2003
50%	50%	01/1975-03/1975, 07/1985-03/1988, 10/1989-11/1990, 06/1997-07/2003
60%	40%	01/1975-03/1975, 03/1980-04/1980, 08/1985-03/1987, 05/1987-03/1988, 10/1989-11/1990, 06/1997-04/2000, 11/2000-06/2003
70%	30%	01/1975-05/1975, 04/1979-07/1981, 02/1983-06/1983, 08/1985-04/1988, 09/1989-02/1991, 10/1993-09/1994, 06/1997-06/2003
75%	25%	01/1975-05/1977, 02/1979-01/2004
80%	20%	01/1975-05/1977, 02/1979-10/2003
90%	10%	01/1975, 05/1979-12/1981, 10/1989-10/1990
100%	0%	10/1979-10/1982

Note:
The table shows periods in which the MS(2-2) model features a high volatility state (i.e. periods with the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

occurred from January to June 1975, from May 1979 to June 1981, from November 1982 to June 1983, from May 1985 to April 1988, from September 1989 to February 1991, from June 1993 to September 1994, and from May 1997 to September 2003[13] (see Tables 2.3-2.6 for details). This would mean that the bond effect has begun to influence the behavior of the mixed portfolios.

For the portfolio with a 75% and an 80% bond proportion, the picture changes entirely. There are two very long intervals with a high volatility. The first period lasts from the beginning of 1975 to May or June 1977. Afterwards it occurs about one and a half year with low volatility and a very long period with high volatility from the beginning of 1979 to the end of 2003 or the beginning of 2004 (see Tables 2.3-2.6 for details). For the MS(2-2) and the MS(2-2)-AR(1) models, the second state features not only high volatility but also a high mean, which differs from all other portfolios (see Tables B.10 and

[13]The periods gives the minimum and the maximum of the intervals, for particular portfolios and models they can be shorter, or even disappear.

Table 2.6: Time intervals with a high volatility state for the MS(2-2)-AR(1) model

Portfolio composition REXP	DAX30	Periods with a high volatility regime
0%	100%	06/1985-03/1988, 10/1989-10/1990, 06/1997-10/2003
10%	90%	06/1985-03/1988, 10/1989-10/1990, 06/1997-10/2003
20%	80%	06/1985-03/1988, 10/1989-10/1990, 06/1997-10/2003
25%	75%	06/1985-03/1988, 10/1989-10/1990, 06/1997-10/2003
30%	70%	06/1985-03/1988, 10/1989-10/1990, 06/1997-10/2003
40%	60%	07/1985-03/1988, 10/1989-11/1990, 06/1997-10/2003
50%	50%	07/1985-03/1988, 10/1989-11/1990, 06/1997-08/2003
60%	40%	03/1980-04/1980, 10/1985-03/1988, 10/1989-11/1990, 06/1997-06/2003
70%	30%	05/1979-07/1981, 05/1983, 02/1986-04/1988, 09/1989-02/1991, 11/1993-09/1994, 06/1997-05/2003
75%	25%	02/1975-06/1977, 04/1979-12/2003
80%	20%	02/1975-06/1977, 04/1979-09/2003
90%	10%	10/1979-04/1980, 12/1980-07/1981, 05/1983-06/1983, 05/1986-09/1986, 07/1987-09/1987, 12/1989-10/1990, 02/1994, 08/1997-10/1997, 02/1999-06/1999
100%	0%	01/1980-04/1980, 01/1981-07/1981, 08/1987, 01/1990-04/1990

Note:
The table shows periods in which the MS(2-2)-AR(1) model features a high volatility state (i.e. periods with the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

B.11).

The portfolio with a 90% bond proportion, the MS(1-2), the MS(1-2)-AR(1), and the MS(2-2) model feature similar effects. The high volatility periods occur in early 1975 (this does not apply to the MS(2-1)-AR(1) model), from May 1979 to June 1983, and from November 1989 to November 1990. The MS(2-2)-AR(1) model yields slightly different results, as it features a larger number of high volatility periods than the other heteroskedastic models. These periods are: from October 1979 to April 1980, from December 1980 to July 1981, from May to June 1983, from May to September 1986, from July to September 1987, from December 1989 to October 1990, in February 1994, from August to October 1997 and from February to June 1999 (see Tables 2.3-2.6 for details).

The pure bond portfolio features a similar effect. However, the number of high volatility periods is in this case smaller than in the 90% bond portfolio case. For the MS(1-2), the MS(1-2)-AR(1), and the MS(2-2) model, there is only one period with a high volatility: from October 1979 to November 1982. For the MS(2-2)-AR(1) model, there are four short high volatility periods: from January to April 1980, from January to July 1981, in August 1987, and from January to April 1994.

The MS(2-1) model features only several months with the second regime which is defined as the low mean regime. These are months with extremely negative log-returns. The MS(2-1)-AR(1) model behaves similarly for the portfolios with a bond proportion ranging from 0% to 70%. For portfolios with a bond proportion from 75% to 100% there are a number of months with a low mean regime, which are separated by short- and mid-long periods of the high mean periods.

Figures D.1-D.39 from Appendix D show four moments conditional on the smoothed probability $\Pr[Z_{t_n}|\mathscr{Y}_T]$ for all 13 estimated portfolios and for the MS(1-2), the MS(2-1), and the MS(2-2) model.[14] The conditional mean is equal to the unconditional mean in the case of the MS(1-2) model, as the mean parameter is independent from the regime. For the MS(2-2) model, the conditional mean varies from $\mu(Z_{t_n} = 1)$ to $\mu(Z_{t_n} = 2)$ in the way that it is close to the first state mean in the first state, and close to the second state mean in the second state. There are some periods in which the conditional mean differs significantly from its bounds (i.e. $\mu(Z_{t_n} = 1)$ and $\mu(Z_{t_n} = 2)$) which are mostly (but not necessarily) associated with the points of the regime change. In the case of the MS(2-1) model, the conditional mean is mostly positive and close to its upper bound equal to the $\mu(Z_{t_n} = 1) > 0$. There are a few observations of the conditional mean that are negatively laid out, which are associated with the extremely negative log-returns.

[14] The conditional moments were computed according to formulas (2.57)-(2.60) with $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$ replacing π_j ($j \in \mathcal{K}$).

For the MS(1-2) and the MS(2-2) model, the conditional variance varies between the $\sigma^2(Z_{t_n} = 1)$ (as the lower bound) and $\sigma^2(Z_{t_n} = 2)$ (as the upper bound). The conditional variance usually stays near its bound during the sojourn in the particular regime. The deviation from the state variance is usually (but not necessarily) associated with the regime change. For the MS(2-1) model, the conditional variance usually stays near its lower bound $\sigma^2(Z_{t_n} = 1)$. One should also note that the conditional variance jumps to a significantly higher level for the short periods of one or two months which is always associated to the high deviation of the log-return from its mean. This effect is stronger for negative outliers.

The conditional skewness is constant and positive for the MS(1-2) model. In the case of the MS(2-2) model, the conditional skewness is always non-negative and varies between two bounds. Concerning the MS(2-1) model, the conditional skewness is generally near zero. For short periods, it jumps to high (negative or positive) values but reverts back to zero quickly. However, the positive jumps are rather uncommon and occur almost only if the log-return process is in the low mean regime. The frequency and sojourn of negative jumps increase proportionally with the increase of the bond proportion in the portfolio.

The conditional excess kurtosis behaves similarly for the MS(1-2) and MS(2-2) model. For both models it is strongly negatively correlated with the conditional skewness of the MS(2-2) model (the exception being a 75% and an 80% bond portfolio, where they are strongly positively correlated). This means that if the conditional skewness for the the MS(2-2) model falls, the conditional excess kurtosis for the MS(1-2) and the MS(2-2) model rises and vice versa (for a 75% and an 80% bond portfolio the conditional skewness of the MS(2-2) model and the conditional excess kurtosis for the MS(1-2) and

MS(2-2) model change in the same direction). The excess kurtosis is non-negative and has an lower bound by zero and some upper bound (different for each portfolio). The conditional excess kurtosis for the MS(2-1) model is often near zero. For short periods it jumps to high (negative or positive) values but reverts quickly to zero. However, the negative jumps are rather uncommon and not necesserily associated with the occurrence of the low mean regime. The frequency and sojourn of positive jumps increase with the increase of the bond proportion in the portfolio. The effect of the strong negative correlation between the conditional skewness and conditional excess kurtosis is valid for the MS(2-1) model and for all portfolios, as well.

Chapter 3

Testing Markov switching models

3.1 Introduction

In Section 2.1 we discussed some stylized facts about the financial time series. We mentioned there that they are often asymmetric, leptokurtotic, and heteroscedastic; which implies that they do not follow the normal distribution. Jarque and Bera (1980) constructed a test with the null hypothesis that a time series is normally distributed. Table 3.1 shows that for all portfolios defined in Section 2.9.1 with the exception of the 20%-80% stock-bond portfolio, the hypothesis of a normal distribution has to be rejected on the 5% confidential level. Therefore, we should use other models to describe the stochasticity of these time series.

The aim of this Chapter is to test whether the Markov switching model better describes the rate of returns of German time series than the commonly used normal distribution. We start with the Akaike and Schwatz Information tests (Section 3.2) and come to the conclusion that MS models are almost

Table 3.1: Jarque-Bera test

Stock prop.	**100%**	**90%**	**80%**	**75%**	**70%**	**60%**	**50%**
Bond prop.	**0%**	**10%**	**20%**	**25%**	**30%**	**40%**	**50%**
JB test	189.3726	162.6361	136.8107	124.0592	111.3213	85.6547	59.6324
p value	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$
Stock prop.	**40%**	**30%**	**25%**	**20%**	**10%**	**0%**	
Bond prop.	**60%**	**70%**	**75%**	**80%**	**90%**	**100%**	
JB test	34.2563	13.5356	7.1833	4.3205	13.7683	49.3480	
p value	$< 10^{-4}$	0.0012	0.0276	0.1153	0.0010	$< 10^{-4}$	

Note:
The table depicts the results of the Jarque-Bera test $JB = \frac{T}{6}\left(\gamma^2 + \frac{(\kappa-3)^2}{4}\right)$ with null hypothesis of normal distribution. T denotes the size, γ the skewness, and κ the kurtosis of the sample.

always better than models from the GARCH family and the normal distribution. Thus, in the remainder of the Chapter we omit GARCH models and test only for the MS model. In Section 3.3 we show that Markov switching models cannot be tested using standard statistical tests, such as Wald or a likelihood ratio test. Therefore, we use several special tests for MS models developed by Hamilton (1996) (Sections 3.4 and 3.5) and Garcia (1998) (Section 3.6). In Section 3.7 we conclude that the MS model is superior to all the other models studied in this Chapter.

3.2 Information criterion tests

As mentioned in the introduction, this work aims to use models with stochastic volatility to price the long-term embedded guarantees. It is therefore desirable to find a model which best fits the heteroscedasticity in the mixed portfolios of REXP and DAX30.[1] For this purpose, we estimated several

[1]For the construction of the portfolios see Section 2.9.1.

MS,[2] ARCH, GARCH, E-ARCH, E-GARCH, T-ARCH, T-GARCH models with and without auto-regression term. Furthermore, the linear basis models (i.e. GBM and AR(1), respectively) were estimated, in order to compare these with more sophisticated models. The MS models were estimated using the EM algorithm described in Section 2.8.3, the models from the GARCH family were estimated using the BHHH method (see Bollerslev (1986) or Berndt et al. (1974)), and the linear models with the MLE approach.

A very common method for comparison of non-nested models is the Akaike Information Criterion (AIC) introduced by Akaike (1973)

$$\text{AIC}{=}\mathscr{L}(\boldsymbol{\theta}) - k, \tag{3.1}$$

where $\mathscr{L}(\boldsymbol{\theta})$ denotes the log-likelihood function for the parameter vector $\boldsymbol{\theta}$ and k the number of estimated parameters. Another useful statistic is the Schwarz Bayesian Criterion (SBC) introduced by Schwarz (1978)

$$\text{SBC}{=}\mathscr{L}(\boldsymbol{\theta}) - 0.5k\ln(T), \tag{3.2}$$

where T denotes the number of observations used to estimate the model.[3] The AIC and SBC statistic for estimated models are listed in Tables F.1-F.13 in the Appendix F.

Table 3.2 shows the ranking of models on the basis of the AIC statistic. For all portfolios with a minimum stock proportion of 40% the MS(1-2) model

[2] In the MS family the following models will be tested. The heteroscedastic model with a regime independent mean equation, i.e. MS(1-2) and MS(1-2)-AR(1); the heteroscedastic model with regime switching in the mean equation, i.e. MS(2-2) and MS(2-2)-AR(1); and the homoscedastic model with a regime dependent mean equation, i.e. MS(2-1) and MS(2-1)-AR(1).

[3] In the literature there are several specifications of the AIC and SBC test statistics, which all are equivalent. In this work the definition used by Hardy (2003, p. 62) will be used.

Table 3.2: Ranking of estimated models according to AIC (1.1975-12.2004)

Portfolio composition REXP	DAX30	Models listed in F.1-F.13
0%	100%	**MS(1-2)** $\gg$ E-GARCH(1,1) > GARCH(1,1) > T-GARCH(1,1) $\gg$ GBM
10%	90%	**MS(1-2)** $\gg$ GARCH(1,1) > E-GARCH(1,1) > T-GARCH(1,1) $\gg$ GBM
20%	80%	**MS(1-2)** $\gg$ GARCH(1,1) $\gg$ E-GARCH(1,1) > T-GARCH(1,1) $\gg$ GBM
25%	75%	**MS(1-2)** $\gg$ GARCH(1,1) $\gg$ T-GARCH(1,1) > E-GARCH(1,1) $\gg$ GBM
30%	70%	**MS(1-2)** $\gg$ GARCH(1,1) $\gg$ T-GARCH(1,1) > E-GARCH(1,1) $\gg$ GBM
40%	60%	**MS(1-2)** $\gg$ GARCH(1,1) $\gg$ E-GARCH(1,1) $\gg$ T-GARCH(1,1) $\gg$ GBM
50%	50%	**MS(1-2)** $\gg$ GARCH(1,1) $\gg$ T-GARCH(1,1) > E-GARCH(1,1) $\gg$ GBM
60%	40%	**MS(1-2)** $\gg$ GARCH(1,1) > T-GARCH(1,1) > E-GARCH(1,1) $\gg$ GBM
70%	30%	T-GARCH(1,1) > GARCH(1,1) > **MS(1-2)** > E-GARCH(1,1) *gg* GBM
75%	25%	T-GARCH(1,1) > GARCH(1,1) > E-GARCH(1,1) *gg* **MS(1-2)** $\gg$ GBM
80%	20%	E-GARCH(1,1) > GARCH(1,1) > T-GARCH(1,1) > **MS(2-2)-AR(1)** $\gg$ AR(1)
90%	10%	GARCH(1,1)-AR(1) $\gg$ T-GARCH(1,1)-AR(1) > E-GARCH(1,1)-AR(1) $\gg$ **MS(1-2)-AR(1)** $\gg$ AR(1)
100%	0%	**MS(1-2)-AR(1)** $\gg$ GARCH(1,1)-AR(1) > T-GARCH(1,1)-AR(1) > E-GARCH(1,1)-AR(1) $\gg$ AR(1)

Portfolio composition REXP	DAX30	Models listed in F.1-F.13 which have no auto-regression term
80%	20%	E-GARCH(1,1) > GARCH(1,1) > T-GARCH(1,1) $\gg$ **MS(1-2)** $\gg$ GBM
90%	10%	E-GARCH(1,1) > GARCH(1,1) $\gg$ T-GARCH(1,1) $\gg$ **MS(1-2)** $\gg$ GBM
100%	0%	**MS(1-2)** $\gg$ GARCH(1,1) > E-GARCH(1,1) > T-GARCH(1,1) $\gg$ GBM

Note:
The tables compare the AIC statistics for all models listed in Tables F.1-F.13. The upper table compares all estimated models. The lower table compares only models listed in F.1-F.13 which have no auto-regression term. This table disregards results for portfolios with the bond proportion between 0% and 75% as they are the same as in the upper table.
AIC denotes Akaike Information Criterion $\left[\text{AIC}=\mathscr{L}(\boldsymbol{\theta})-k\right]$, $\mathscr{L}(\boldsymbol{\theta})$ - log-likelihood, k - number of parameters, GBM - Geometric Brownian motion, AR(p) - auto-regressive model of the p-th order, MS(m-s) - Markov switching model with m mean equations and s regimes for the variance, (G)ARCH(p,q) - (Generalised) Auto-regression Conditional Heteroscedasticity with q GARCH processes and p ARCH processes, E-(G)ARCH - exponential (G)ARCH, T-(G)ARCH - threshold (G)ARCH. $x > y$ - model x has the AIC statistic higher by a maximum of 0.5 points than model y, $x \gg y$ - model x has the AIC statistic higher by more than 0.5 points than model y, bold - the best Markov switching model.

is clearly favored over all models from the GARCH family[4] and models without switching characteristics (e.i. GBM and AR(1)). The portfolio with 70% bonds and 30% shares is best represented with the T-GARCH(1,1) and the GARCH(1,1) model. First then the Markov switching models come, with the MS(1-2) model, as the third best one. It should, however, be pointed out that the AIC statistic of the MS(1-2) model is only 0.1503 lower than the AIC statistic of the T-GARCH(1,1) model. This AIC statistic difference is very low, amounting to less than 0.02‰. After the MS(1-2) model come the E-GARCH(1,1) and the GBM model. The situation is similar for the portfolio with the 75%-25% bond-stock proportion, the only difference being that the E-GARCH(1,1) performs better than the MS(1-2) model.

From the portfolio with an 80% and a larger bond proportion, the mean-reverting effects are observable. In the 80% bond portfolio, this effect is not as univocal. In this case, the addition of the auto-regression term to the model may increase the AIC statistic, as is the case in the MS model family. However, there are models, such as the GARCH-typed model, in which the auto-regression term decreases the magnitude of the statistic. Both the E-GARCH and the T-GARCH model types do not have a clear trends. For instance, the E-GARCH(1,1) model performs better than the E-GARCH(1,1)-AR(1) but the E-ARCH(1) performs worse than the E-ARCH(1)-AR(1). For the time series with the 80% bonds and 20% shares, the E-GARCH(1,1) is the best model. This is followed by the GARCH(1,1), T-GARCH(1,1), and MS(2-2)AR(1). The worst performer is the AR(1) process, when used as the reference model. It should be stressed that the difference in the AIC statis-

[4]This section differentiates between the type of models (i.e. GARCH-typed models are all ARCH(p) and GARCH(q,p) models) and the family of models (i.e. GARCH family of models consists of all GARCH-, E-GARCH- and T-GARCH-typed models). Therefore the family of models is a wider concept than the type of models.

tic between the E-GARCH(1,1) and the MS(2-2)-AR(1) equals only 0.4125 which is 0.04‰ of the AIC statistic.

The portfolio with a 10% stock proportion is best fitted by the GARCH(1,1)-AR(1) process. This is followed by both assymetric models – T-GARCH(1,1)-AR(1) and E-GARCH(1,1)-AR(1) – then comes the MS(1-2)-AR(1) and the reference model AR(1) in last place. The situation for the pure bond portolio changes. The Markov switching models with the MS(1-2)-AR(1) model as representative are, once again, the best performers. Then follow the GARCH model family (with the auto-regressive term) and, lastly, the AR(1) model.

The SBC statistic gives similar results to those of the AIC statistic (see Table 3.3). For portfolios with the stock proportion between 50% and 100% the MS(1-2) model is the best choice. The portfolio with 40% shares is best fitted with the GARCH(1,1) model. However, the second best model is the MS(1-2) model, the SBC statistic being only 0.1393 lower (amounting to less than 0.02‰ of the SBC statistic). This is followed by the T-GARCH(1,1), E-GARCH(1,1), and the reference GBM model. The portfolio with a 30% share proportion is also best fitted with the GARCH(1,1) model. However, it is followed by the GBM. Then, listed according to suitability, come the T-GARCH(1,1), MS(1-2), and E-GARCH(1,1) model, with differences in the SBC statistic lower than 0.5 points. The situation of the 75%-25% REXP-DAX30 portfolio is similar. The Markov switching model family, with its representative MS(1-2) model is the worst performer. In the case of the portfolio with 20% stock investment, the ranking of model types is identical to that of the previous portfolio. The difference lies in the fact that the ARCH(1) is the best of the GARCH-typed models and the T-ARCH(1) process the best of the T-GARCH-typed models. Only the best model of the E-GARCH-typed models features the GARCH coefficient. It should be poin-

Table 3.3: Ranking of estimated models according to SBC (1.1975-12.2004)

Portfolio composition REXP	DAX30	Models listed in F.1-F.13
0%	100%	**MS(1-2)** ≫ GARCH(1,1) ≫ E-GARCH(1,1) ≫ T-GARCH(1,1) ≫ GBM
10%	90%	**MS(1-2)** ≫ GARCH(1,1) ≫ E-GARCH(1,1) > T-GARCH(1,1) ≫ GBM
20%	80%	**MS(1-2)** ≫ GARCH(1,1) ≫ E-GARCH(1,1) > T-GARCH(1,1) ≫ GBM
25%	75%	**MS(1-2)** ≫ GARCH(1,1) ≫ T-GARCH(1,1) > E-GARCH(1,1) ≫ GBM
30%	70%	**MS(1-2)** ≫ GARCH(1,1) ≫ T-GARCH(1,1) > E-GARCH(1,1) ≫ GBM
40%	60%	**MS(1-2)** ≫ GARCH(1,1) ≫ T-GARCH(1,1) ≫ E-GARCH(1,1) ≫ GBM
50%	50%	**MS(1-2)** ≫ GARCH(1,1) ≫ T-GARCH(1,1) > E-GARCH(1,1) ≫ GBM
60%	40%	GARCH(1,1) > **MS(1-2)** ≫ T-GARCH(1,1) > E-GARCH(1,1) ≫ GBM
70%	30%	GARCH(1,1) ≫ GBM ≫ T-GARCH(1,1) > **MS(1-2)** > E-GARCH(1,1)
75%	25%	GARCH(1,1) > GBM ≫ T-GARCH(1,1) > E-GARCH(1,1) > **MS(1-2)**
80%	20%	ARCH(1) ≫ GBM ≫ T-ARCH(1) > E-GARCH(1,1) ≫ **MS(1-2)**
90%	10%	GARCH(1,1)-AR(1) ≫ T-GARCH(1,1)-AR(1) > E-GARCH(1,1)-AR(1) ≫ **MS(1-2)-AR(1)** ≫ AR(1)
100%	0%	GARCH(1,1)-AR(1) ≫ **MS(1-2)-AR(1)** ≫ T-GARCH(1,1)-AR(1) > E-GARCH(1,1)-AR(1) ≫ AR(1)

Portfolio composition REXP	DAX30	Models listed in F.1-F.13 which have no auto-regression term
90%	10%	GARCH(1,1) ≫ E-GARCH(1,1) ≫ T-GARCH(1,1) ≫ **MS(1-2)** ≫ GBM
100%	0%	**MS(1-2)** ≫ GARCH(1,1) ≫ E-GARCH(1,1) > T-GARCH(1,1) ≫ GBM

Note:
The tables compare the SBC statistics for all models listed in Tables F.1-F.13. The upper table compares all estimated models. The lower table compares only models listed in F.1-F.13 which have no auto-regression term. This table disregards results for portfolios with the bond proportion between 0% and 80% as they are the same as in the upper table.
SBC - Schwarz Bayesian Information Criterion $\left[\text{SBC}=\mathscr{L}(\boldsymbol{\theta}) - 0.5k\ln(T)\right]$, $\mathscr{L}(\boldsymbol{\theta})$ - log-likelihood, k - number of parameters, T - number of observations, GBM - Geometric Brownian motion, AR(p) - auto-regressive model of the p-th order, MS(m-s) - Markov switching model with m mean equations and s regimes for the variance, (G)ARCH(p,q) - (Generalised) Auto-regression Conditional Heteroscedasticity with q GARCH processes and p ARCH processes, E-(G)ARCH - exponential (G)ARCH, T-(G)ARCH - threshold (G)ARCH. $x > y$ - model x has the SBC statistic higher by a maximum of 0.5 points than model y, $x \gg y$ - model x has the SBC statistic higher by more than 0.5 points than model y, bold - the best Markov switching model.

ted out that the SBC statistic decreases with addition of the auto-regression term for all tested models. There are two exceptions. According to the SBC test, the mean reversion can be found only in portfolios with a maximum of 10% of the stock engagement. The 90%-10% REXP-DAX30 portfolio is best fitted with the GARCH(1,1)-AR(1) model. This is followed by the T-GARCH(1,1)-AR(1), E-GARCH(1,1)-AR(1), the MS(1-2)-AR(1), and the AR(1) process. The Markov switching model family performs better for the pure bond investment portfolio, where its representative MS(1-2)-AR(1) is the second best after the GARCH(1,1)-AR(1) process. They are followed by the T-GARCH(1,1)-AR(1) and the E-GARCH(1,1)-AR(1) model. The worst performer is the reference AR(1) model.

According to Tables 3.2 and 3.3, it can be seen that the portfolios with very little stock proportion exhibit mean reversion effects. However, it would be interesting to see what happens if the mean-reverting effects would be disregarded. The AIC statistic for portfolios with a 20% and 10% stock engagement prefers the E-GARCH(1,1) model. It is followed by the GARCH(1,1), the T-GARCH(1,1), MS(1-2), and the GBM. The SBC statistic for the portfolio with 10% REXP investment shows similar outcomes. The difference is that the GARCH(1,1) model is the best one and the E-GARCH(1,1) follows in second place. Finally, the pure bond strategy is best fitted with the MS(1-2) model, according to both the AIC and the SBC statistics.

In conclusion, according to the information statistics of Akaike and Schwarz, the MS(1-2) model describes the stochastic of the mixed REXP-DAX30 portfolios very well. For the portfolios whose majority is invested in stocks, it outperforms all estimated models from the GARCH family and the reference model GBM as well. If the proportion of bonds lies between 60% and 90% the GARCH model family is a better performer. For the pure bond investment, the Markov switching model with its mean reverting variant MS(1-2)-AR(1) is, once again, the best choice. However, for the cases in which the information statistic was higher for the GARCH model family, the difference to the best MS model was fairly small. As the AIC and the SBC criterion merely indicate the rank of the model but do not state if the difference is significant, it will be assumed that MS and GARCH models fit the portfolios with high bond participation equally well. Therefore the MS models will be preferred in order to render the pricing consistent.

It is worth mentioning that according to the SBC, the linear GBM model is better than the MS family but worse than the best model from the GARCH family for portfolios with a bond ratio between 70% and 80%. This effect did not appear if the model choice was based on the AIC statistic. As mentioned above, the SBC statistic gives the rank of the models; it would therefore be interesting to make a direct comparison between the MS model family and the linear models. Such tests are discussed in the following sections of this Chapter.

3.3 Problems with testing of Markov switching models

As was shown in Section 3.2, the MS models perform better or at least as well as the GARCH models when fitting the German mixed bond-stock portfolios. The remaining part of the Chapter will focus explicitly on testing the regime switching effects in the data. Please note that linear models, (i.e. GBM and AR(1) are nested in the MS models, as they can be regarded as MS models with only one regime, i.e. $K = 1$). When testing the MS model against the linear model, without loss of generality, one should test one of the following null hypotheses

$\mathrm{H}_0^{(1a)}$: $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, $\sigma_1^2 = \sigma_2^2$ with undefined p_{11} and p_{22},

or

$\mathrm{H}_0^{(1b)}$: $p_{11} = 0$ with undefined $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, σ_1^2, σ_2^2

or

$\mathrm{H}_0^{(1c)}$: $p_{11} = 1$ with undefined $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, σ_1^2, σ_2^2

which are equivalent. The $\mathrm{H}_0^{(1a)}$ has the following interpretation. The parameters are equal in both states, hence regardless of the value of Z_t, the result is the same. Hypothesis $\mathrm{H}_0^{(1b)}$ says that it is irrelevant how high parameters $\boldsymbol{\beta}_1$ and σ_1^2 are, thus the time series will never reaches the first state. Hypothesis $\mathrm{H}_0^{(1c)}$ claims the same for parameters from the second state, as the models always remain in the first state.

In literature, a common procedure for testing nested models involves using one of the large sample asymptotic tests, such as the likelihood ratio (LR), the Wald or the Lagrange multiplier (LM) test. These tests are based on the asymptotic distribution theory which says that, under regularity conditions, in a sufficiently large sample, the estimated parameter vector $\widehat{\boldsymbol{\theta}}$ converges to the true parameter vector $\boldsymbol{\theta}_0$. Through the application of the Taylor's expansion one finds that the parameter estimator $\widehat{\boldsymbol{\theta}}$ is equal to the sum of the true parameter $\boldsymbol{\theta}_0$ and the score evaluated at the true value $\left(\frac{\partial \mathscr{L}_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_0}\right)$ divided by the second derivative of the log-likelihood function evaluated at the median $\left(\frac{\partial^2 \mathscr{L}_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_m \boldsymbol{\theta}_m'}\right)$. This approach is based on the assumption that the likelihood function is locally approximately quadratic. This means that the second derivative is approximately constant. In the next step, the application of the central limit theorem is allowed, as the scores have a zero mean (for scores with positive variance). The central limit theorem allows to conclude that the estimator is asymptotically multivariate normal (Hansen 1992, p. S61).

Unfortunately, the MS models violate two crucial assumptions of the asymptotic distribution theory: the local quadrativeness of the likelihood function and the assumption of the positive variance of the scores. The likelihood function is locally quadratic if it is highly probable that the likelihood surface is asymptotically quadratic over the region in which both the null hypothesis and the global optimum lie. In the MS framework, however, (at least some) transition probabilities are not specified under the null hypothesis. This means that in the optimum, the value of the likelihood function is equal for all non-specified parameters, if the null hypothesis is true. Thus, the flatness of the likelihood function is contradictory to the assumption of its local quadrativness. In other cases, the likelihood function has more than one maximum, thus the null hypothesis does not necessarily lie on the same "hill" as the global maximum, which is another violation of the local quadrativness assumption (Hansen 1992, p. S61-S62).

The assumption of the positive variance of the scores is also violated. As mentioned above, the likelihood function is flat under the null. This means that if one intended to test the $\mathrm{H}_0^{(1a)}$, the scores with respect to $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, σ_1^2 and σ_2^2 would all be equal to zero for all possible values of $p_{ji} \in [0, 1]$. Additionally, the likelihood function of the MS models has several local minima, maxima and flection points. Therefore, its scores equal to zero on these points per definition (Hansen 1992, p. S62). These zero scores imply that the information matrix is singular under the null hypothesis (Watson and Engle 1985, p. 341-342).

Given that two regularity conditions of the asymptotic distribution theory are violated, the theory cannot be used. As a consequence, the test statistics are not χ^2 distributed, which causes several theoretical problems with regard to the test statistic (Lee and Chesher 1986, p. 122). Despite the issues discussed above, some authors, such as Hardy (2003, p. 60-62), use the likelihood ratio test, which is very questionable.

3.4 Simple tests

3.4.1 Wald test

As mentioned in Section 3.3 the standard asymptotic tests are not χ^2 distributed. Engel and Hamilton (1990) propose a method to deal with this problem. According to their approach, one assumes that under a null hypothesis, transition probabilities are defined in such a way that $p_{11} = 1 - p_{22}$ or equivalently $\Pr[Z_t = 1|Z_{t-1} = 1] = \Pr[Z_t = 1|Z_{t-1} = 2]$. This means that the probability that the observation y_t comes from the first state ($Z_t = 1$) is independent from the previous realization of the state variable (i.e. independent if $Z_{t-1} = 1$ or $Z_{t-1} = 2$). Thus the model under the null is reduced to a mixed model

$$\mathrm{H}_0^{(2a)}: \quad y_t = \frac{p_{11}}{\sqrt{2\pi\sigma_1^2}} \exp\left[-\frac{(y_t - \boldsymbol{\beta}_1'\boldsymbol{x_t})^2}{2\sigma_1^2}\right] + \frac{1-p_{11}}{\sqrt{2\pi\sigma_2^2}} \exp\left[-\frac{(y_t - \boldsymbol{\beta}_2'\boldsymbol{x_t})^2}{2\sigma_2^2}\right] \tag{3.3}$$

or equivalently

$\mathrm{H}_0^{(2b)}$: $p_{11} = 1 - p_{22}$, $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$ and/or $\sigma_1 \neq \sigma_2$

against the alternative

$\mathrm{H}_1^{(2b)}$: $p_{11} \neq 1 - p_{22}$

instead of $\mathrm{H}_0^{(1a,b,c)}$. The Wald test then has the form

$$WT = \frac{[\widehat{p}_{11} - (1 - \widehat{p}_{22})]^2}{\widehat{var}(\widehat{p}_{11}) + \widehat{var}(\widehat{p}_{22}) + 2\widehat{cov}(\widehat{p}_{11}, \widehat{p}_{22})} \overset{ass}{\approx} \chi^2(1), \tag{3.4}$$

with $\widehat{var}(\widehat{p}_{11})$ and $\widehat{var}(\widehat{p}_{22})$ – variance of the parameter estimate p_{11} and p_{11}, respectively, and $\widehat{cov}(\widehat{p}_{11}, \widehat{p}_{22})$ – covariance of the parameter estimates p_{11} and p_{22}.

Another approach suggested by Engel and Hamilton (1990) is to test if the intercept parameter (μ) is independent from the state variable. The null hypothesis in this case is defined as

$\mathrm{H}_0^{(3)}$: $\mu_1 = \mu_2$, $\sigma_1 \neq \sigma_2$, p_{11}, p_{22} – not specified

with its associated alternative

$\mathrm{H}_1^{(3)}$: $\mu_1 \neq \mu_2$.

The Wald test subsequently takes on the following form

$$WT = \frac{[\widehat{\mu}_1 - \widehat{\mu}_2]^2}{\widehat{var}(\widehat{\mu}_1) + \widehat{var}(\widehat{\mu}_2) - 2\widehat{cov}(\widehat{\mu}_1, \widehat{\mu}_2)} \overset{ass}{\approx} \chi^2(1), \tag{3.5}$$

with $\widehat{var}(\widehat{\mu}_1)$ and $\widehat{var}(\widehat{\mu}_2)$ – variance of the parameter estimate μ_1 and μ_2, respectively, and $\widehat{cov}(\widehat{\mu}_1, \widehat{\mu}_2)$ – covariance of the parameter estimates μ_1 and μ_2.

Engel and Hamilton (1990) only test the intercept parameter but it is also straightforward to test for heteroskedasticity and auto-regression. In this case, the homoskedastic null hypothesis would be

$\mathrm{H}_0^{(4a)}$: $\mu_1 \neq \mu_2$, $\sigma_1^2 = \sigma_2^2$, p_{11}, p_{22} – not specified

against the heteroskedastic alternative

$\mathrm{H}_1^{(4a)}$: $\sigma_1^2 \neq \sigma_2^2$

Subsequently, the Wald test will look as follows

$$WT = \frac{[\widehat{\sigma}_1^2 - \widehat{\sigma}_2^2]^2}{\widehat{var}(\widehat{\sigma}_1^2) + \widehat{var}(\widehat{\sigma}_2^2) - 2\widehat{cov}(\widehat{\sigma}_1^2, \widehat{\sigma}_2^2)} \overset{ass}{\approx} \chi^2(1), \tag{3.6}$$

with $\widehat{var}(\widehat{\sigma}_1^2)$ and $\widehat{var}(\widehat{\sigma}_2^2)$ – variance of the parameter estimate σ_1^2 and σ_2^2, respectively, and $\widehat{cov}(\widehat{\sigma}_1^2, \widehat{\sigma}_2^2)$ – covariance of the parameter estimates σ_1^2 and σ_2^2.

The test for auto-regression will have the following hypothesis

$\mathrm{H}_0^{(5)}$ $\mu_1 \neq \mu_2$, $\sigma_1^2 \neq \sigma_2^2$, $\phi_{i(1)} = \phi_{i(2)}$, $\phi_{j(1)} \neq \phi_{j(2)}$, $(j \neq i,\ i, j = 1, \ldots, r)$,
p_{11}, p_{22} – not specified

and the alternative

$H_1^{(5)}$: $\phi_{i(1)} \neq \phi_{i(2)}$,

with $\phi_{i(k)}$ the i-lagged auto-regression term in the k-th state. The Wald test statistic for the auto-regressive term will look as follows

$$WT = \frac{[\widehat{\phi}_{i(1)} - \widehat{\phi}_{i(2)}]^2}{\widehat{var}(\widehat{\phi}_{i(1)}) + \widehat{var}(\widehat{\phi}_{i(2)}) - 2\widehat{cov}(\widehat{\phi}_{i(1)}, \widehat{\phi}_{i(2)})} \overset{ass}{\approx} \chi^2(1), \tag{3.7}$$

with $\widehat{var}(\widehat{\phi}_{i(1)})$ and $\widehat{var}(\widehat{\phi}_{i(2)})$ – variance of the parameter estimate $\phi_{i(1)}$ and $\phi_{i(2)}$, respectively, and $\widehat{cov}(\widehat{\phi}_{i(1)}, \widehat{\phi}_{i(2)})$ – covariance of the parameter estimates $\phi_{i(1)}$ and $\phi_{i(2)}$.

The results of the Wald test are listed in Tables G.1-G.13 in Appendix G. Table 3.4 gives an overview of the results. It shows which tests can be rejected at the 5% confidence level.

Models MS(1-2) and MS(1-2)-AR(1) tested if the transition probability ($H_0^{(2b)}$) and the variance ($H_0^{(4a)}$) are not dependent on the state. For all samples, both null hypotheses can be rejected. Thus, both models can be used to describe mixed portfolios of REXP and DAX30.

The homoskedastic models MS(2-1) and MS(2-1)-AR(1) were tested for independence of the transitions probabilities ($H_0^{(2b)}$) and the intercept ($H_0^{(3)}$) from the state variable. For the auto-regressive model MS(2-1)-AR(1) the auto-regression parameter ($H_0^{(5)}$) was additionally tested. The null hypothesis that the intercept does not depend on the state variable was rejected for both models and all portfolios. The hypothesis of the regime-independent auto-regression coefficient could be rejected for portfolios with a high bond proportion of 75% and more. The null hypothesis for the transition probabilities cannot be rejected for any sample with the exception of the pure DAX30 portfolio (MS(2-1)-AR(1) model). This implies that the homoskedastic model is not adequate for German mixed bond-stock portfolios, as the transition probabilities are not dependent on the lagged state variable Z_{t-1}.

Table 3.4: Overview of the results of the Wald test (1.1975-12.2004)

Portfolio composition		**MS(1-2)**				**MS(2-1)**					**MS(2-2)**						
		AR(0)		**AR(1)**		**AR(0)**		**AR(1)**			**AR(0)**			**AR(1)**			
REXP	**DAX30**	$H_0^{(2b)}$	$H_0^{(4a)}$	$H_0^{(2b)}$	$H_0^{(4a)}$	$H_0^{(2b)}$	$H_0^{(3)}$	$H_0^{(2b)}$	$H_0^{(3)}$	$H_0^{(5)}$	$H_0^{(2b)}$	$H_0^{(3)}$	$H_0^{(4a)}$	$H_0^{(2b)}$	$H_0^{(3)}$	$H_0^{(4a)}$	$H_0^{(5)}$
0%	100%	+	+	+	+		+	+	+		+		+	+		+	
10%	90%	+	+	+	+		+		+		+		+	+		+	
20%	80%	+	+	+	+		+		+		+		+	+		+	
25%	75%	+	+	+	+		+		+		+		+	+		+	
30%	70%	+	+	+	+		+		+		+		+	+		+	
40%	60%	+	+	+	+		+		+		+		+	+		+	
50%	50%	+	+	+	+		+		+		+		+	+		+	
60%	40%	+	+	+	+		+		+		+		+	+		+	
70%	30%	+	+	+	+		+		+		+		+	+		+	
75%	25%	+	+	+	+		+		+	+	+		+	+		+	
80%	20%	+	+	+	+		+		+	+	+		+	+		+	
90%	10%	+	+	+	+		+		+	+	+		+	+		+	+
100%	0%	+	+	+	+		+		+	+	+		+	+			+

Note:
MS(m-s)-AR(p) denotes a Markov switching model with m mean equations, s regimes for the variance and p auto-regression lags in each mean equation. + - the null hypothesis can be rejected at the 5% confidence level.

$H_0^{(2b)}$: $p_{11} = 1 - p_{22}$ — $H_1^{(2b)}$: $p_{11} \neq 1 - p_{22}$

$H_0^{(3)}$: $\mu_1 = \mu_2$, $\sigma_1^2 \neq \sigma_2^2$, p_{11}, p_{22} - not specified, — $H_1^{(3)}$: $\mu_1 \neq \mu_2$

$H_0^{(4a)}$: $\mu_1 \neq \mu_2$, $\sigma_1^2 = \sigma_2^2$, p_{11}, p_{22} - not specified, — $H_1^{(4a)}$: $\sigma_1^2 \neq \sigma_2^2$

$H_0^{(5)}$: $\phi_{1(1)} = \phi_{1(2)}$, p_{11}, p_{22} - not specified, — $H_1^{(5)}$: $\phi_{1(1)} \neq \phi_{1(2)}$

For the heteroskedastic model with regime dependent mean equation (i.e. MS(2-2) and MS(2-2)-AR(1)) three tests for regime independent transition probabilities ($H_0^{(2b)}$), intercept ($H_0^{(3)}$), and variance ($H_0^{(4a)}$) were conducted. For the model MS(2-2)-AR(1) the independence of the auto-regression coefficient ($H_0^{(5b)}$) was tested additionally. The null hypothesis of independent transition probabilities can be rejected for both models for all samples. The test for homoscedasticity was rejected for both models and all samples, with one exception. For the pure bond portfolio and the model MS(2-2)-AR(1) the homoscedasticity hypothesis cannot be rejected at any common confidence level. The null hypothesis of the regime-independent intercept could not be rejected for either of the models, or the portfolios. The null hypothesis of regime-independent auto-regression coefficient could be rejected only for portfolios with a very low stock proportion of a maximum of 10%. These results show that the MS(2-2)-typed models have regime-dependent transition probabilities, variance, and for models with a high bond proportion, the auto-regression coefficient. On the other hand, the intercept is regime independent, which suggests that the MS(2-2)-typed models are overparametrised.

In conclusion, the results of the Wald test reveal that one should reject the MS(2-1)-typed models and chose the MS(2-2)-typed or MS(1-2)-typed models. Note that the MS(1-2)-typed and MS(2-2)-typed models have a regime-independent intercept. As the MS(1-2)-typed models by the model construction have a regime-independent mean equation and the MS(2-2)-typed models do not reject the $H_0^{(3)}$ null hypothesis. Since both model types are equivalent, the MS(1-2)-type is the better choice, because it avoids overparametrization.

3.4.2 Likelihood ratio test

Besides the Wald test, Engel and Hamilton (1990) proposed a modified form of the likelihood ratio test. They suggest estimating two models with and without restriction on parameters. Thus, it is possible to construct the likelihood ratio statistics

$$LR = 2[\mathscr{L}(\mathrm{H}_1) - \mathscr{L}(\mathrm{H}_0)] \overset{ass}{\approx} \chi^2(k(\mathrm{H}_1) - k(\mathrm{H}_0)). \tag{3.8}$$

where $\mathscr{L}(\cdot)$ denotes the log-likelihood function under the null and under the alternative hypothesis and k the number of parameters of the model under H_0 and H_1, respectively.

Without estimating additional models, testing the MS(2-1)-typed model is fairly straightforward

$\mathrm{H}_0^{(4b)}$: $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$, $\sigma_1^2 = \sigma_2^2$, and $p_{11} \neq 1 - p_{21}$

against the alternative of the MS(2-2)-typed model

$\mathrm{H}_1^{(4b)}$: $\sigma_1^2 \neq \sigma_2^2$

and the MS(1-2)-typed model

$\mathrm{H}_0^{(6)}$: $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, $\sigma_1^2 \neq \sigma_2^2$, and $p_{11} \neq 1 - p_{21}$

against the alternative of the MS(2-2)-typed model

$\mathrm{H}_1^{(6)}$: $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$.

To test the null hypothesis of the regime-independence of the transition probabilities ($\mathrm{H}_0^{(2b)}$) the model with the constraint $p_{11} = 1 - p_{21}$ has to be estimated and compared with the unconstrained model ($\mathrm{H}_1^{(2b)}$).

Tables G.1-G.13 in Appendix G show the output of the likelihood ratio test. A summary of the test results is given in Table 3.5 that shows which tests can be rejected at the 5% confidence level. All six models were tested on whether or not the transition probability is independent from the lagged regime variable Z_{t-1} ($\mathrm{H}_0^{(2b)}$). The test could be rejected for the MS(1-2) and

Table 3.5: Overview of the results of the likelihood ratio test (1.1975-12.2004)

Portfolio composition		**MS(1-2)**		**MS(2-1)**		**MS(2-2)**					
		AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**			**AR(1)**		
REXP	**DAX30**	$H_0^{(2b)}$	$H_0^{(2b)}$	$H_0^{(2b)}$	$H_0^{(2b)}$	$H_0^{(2b)}$	$H_0^{(4b)}$	$H_0^{(6)}$	$H_0^{(2b)}$	$H_0^{(4b)}$	$H_0^{(6)}$
0%	100%	+	+	+	+	+	+		+	+	
10%	90%	+	+	+	+	+	+		+	+	
20%	80%	+	+	+	+	+	+		+	+	
25%	75%	+	+	+	+	+	+		+	+	
30%	70%	+	+	+	+	+	+		+	+	
40%	60%	+	+	+	+	+	+		+	+	
50%	50%	+	+	+	+	+	+		+	+	
60%	40%	+	+	+	+	+	+		+	+	
70%	30%	+				+	+		+	+	
75%	25%	+					+		+	+	
80%	20%	+					+		+	+	
90%	10%	+				+	+		+		
100%	0%	+				+	+		+	+	

Note:
MS(m-s)-AR(p) denotes a Markov switching model with m mean equations, s regimes for the variance and p auto-regression lags in each mean equation. + - the null hypothesis can be rejected at the 5% confidence level.

$H_0^{(2b)}$: $p_{11} = 1 - p_{22}$ $\quad$ $H_1^{(2b)}$: $p_{11} \neq 1 - p_{22}$

$H_0^{(4b)}$: $\mu_1 \neq \mu_2$, $\phi_{1(1)} \neq \phi_{1(2)}$, $\sigma_1^2 = \sigma_2^2$, $p_{11} \neq 1 - p_{22}$, $\quad$ $H_1^{(4b)}$: $\sigma_1^2 \neq \sigma_2^2$

$H_0^{(6)}$: $\mu_1 = \mu_2$, $\phi_{1(1)} = \phi_{1(2)}$, $\sigma_1^2 \neq \sigma_2^2$, $p_{11} \neq 1 - p_{22}$, $\quad$ $H_1^{(6)}$: $\mu_1 \neq \mu_2$, $\phi_{1(1)} \neq \phi_{1(2)}$

MS(2-2) models for all samples. For models MS(1-2)-AR(1), MS(2-1), and MS(2-1)-AR(1), the null hypothesis can be rejected for all samples with a stock proportion ranging from 0% to 30%. For the MS(2-2)-AR(1) model, the null hypothesis of no regime in the transition probabilities can be rejected for all samples excluding the portfolio with with a 90% bonds and a 10% stock engagement.

In addition to the above test, the MS(2-2)-typed models were tested for regime switching in the mean equation (i.e. if $\boldsymbol{\beta}$-vector was independent from the regime, see $\mathrm{H}_0^{(6)}$) and variance ($\mathrm{H}_0^{(4b)}$). The null hypothesis of no regime dependence in the mean equation cannot be rejected for any sample in the case of both models (i.e. with and without auto-regression). This means that the MS(1-2)-typed models fit the tested portfolios better than the MS(2-2)-typed models. The null hypothesis of homoscedasticity was rejected for all samples and both MS(2-2)-typed models, with the exception of the MS(2-2)-AR(1) model and the portfolio with a 90% bond and a 10% stock proportion. In other words, the likelihood ratio test rejects the MS(2-1)-typed models in favor of the MS(2-2)-typed models (with one exception). On this basis we can conclude that the results of the likeklihood ratio test are consistent with the results of the Wald test discussed in Section 3.4.

In conclusion, it can be said that both the Wald and the likelihood ratio test favors the MS(1-2) model for all portfolios, except those with a very high bond engagement. For portfolios with a very high bond proportion, the MS(1-2)-AR(1) model would be more appropriate.

3.5 Tests based on scores

3.5.1 Scores

This section introduces two tests for Markov switching specification which are based on scores. These are the Newey-Tauchen-White score test and the Lagrange multiplier test proposed by Hamilton (1996).

Before introducing the score based tests, it is necessary to define the score. The score vector $\boldsymbol{h}_t$ is the vector of the derivatives of the conditional log-likelihood function $l(y_t|\mathscr{Y}_{t-1};\boldsymbol{\theta})$ with respect to the parameter vector $\boldsymbol{\theta}$.

$$\boldsymbol{h}_t(\boldsymbol{\theta}) = \frac{\partial l(y_t|\mathscr{Y}_{t-1};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \tag{3.9}$$

The score vector can be evaluated at the true parameter vector $\boldsymbol{\theta}_0$

$$\boldsymbol{h}_t(\boldsymbol{\theta}_0) = \left.\frac{\partial l(y_t|\mathscr{Y}_{t-1};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \tag{3.10}$$

or at the maximum likelihood estimate $\widehat{\boldsymbol{\theta}}$

$$\boldsymbol{h}_t(\widehat{\boldsymbol{\theta}}) = \left.\frac{\partial l(y_t|\mathscr{Y}_{t-1};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \tag{3.11}$$

(Hamilton 1996, p. 131). In order to compute the scores, the parameter vector is subdivided into $\boldsymbol{\theta}^{*\prime} = (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_K', \sigma_1^2, \ldots, \sigma_K^2)'$ and $\boldsymbol{\delta}' = (p_{11} \ldots, p_{1K}, \ldots, p_{K1}, \ldots, p_{KK})'$ so that $\boldsymbol{\theta}' = (\boldsymbol{\theta}^{*\prime}, \boldsymbol{\delta}')'$. Hamilton (1996, p. 135) showed that the score vector for the Markov switching models can be computed as

$$\begin{aligned}\boldsymbol{h}_t(\boldsymbol{\theta}^*) =& \frac{\partial l(y_t|\mathscr{Y}_{t-1};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}\\ =& \sum_{j=1}^{K} \boldsymbol{\psi}_{t,j} \Pr[Z_t = j|\mathscr{Y}_t] + \sum_{\tau=1}^{t-1}\sum_{j=1}^{K} \boldsymbol{\psi}_{t,j} \left(\Pr[Z_t = j|\mathscr{Y}_t] - \Pr[Z_t = j|\mathscr{Y}_{t-1}]\right)\\ & (t = 1, \ldots, T),\end{aligned} \tag{3.12}$$

with

$$\boldsymbol{\psi}_{t,j} = \frac{\partial l(y_t|\boldsymbol{x}_t, Z_t = j; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*} \tag{3.13}$$

and $(\Pr[Z_t = j|\mathscr{Y}_t] - \Pr[Z_t = j|\mathscr{Y}_{t-1}])$ the byproduct of the maximum likelihood estimation procedure. The derivatives from equation (2.76) are given by

$$\boldsymbol{\psi}_{t,j} = \frac{\partial l(y_t|\boldsymbol{x}_t, Z_t = j; \boldsymbol{\theta})}{\partial \boldsymbol{\beta}_j} = \frac{(y_t - \boldsymbol{x}_t'\boldsymbol{\beta}_j)\boldsymbol{x}_t}{\sigma_j^2} \quad (j = 1, \ldots, K), \tag{3.14}$$

and

$$\boldsymbol{\psi}_{t,j} = \frac{\partial l(y_t|\boldsymbol{x}_t, Z_t = j; \boldsymbol{\theta})}{\partial \sigma_j} = -\frac{1}{2\sigma_j^2} + \frac{(y_t - \boldsymbol{x}_t'\boldsymbol{\beta}_j)}{2\sigma_j^4} \quad (j = 1, \ldots, K). \tag{3.15}$$

The score with respect to the transition probabilities is given by

$$\begin{aligned}
\boldsymbol{h}_t(p_{ji}) &= \frac{\partial l(y_t|\mathscr{Y}_{t-1}; \boldsymbol{\theta})}{\partial p_{ji}} \\
&= \frac{1}{p_{ji}} \Pr[Z_t = j, Z_{t-1} = i|\mathscr{Y}_t] - \frac{1}{p_{Ki}} \Pr[Z_t = K, Z_{t-1} = i|\mathscr{Y}_t] \\
&+ \frac{1}{p_{ji}} \left\{ \sum_{\tau=2}^{t-1} \Pr[Z_\tau = j, Z_{t-1} = i|\mathscr{Y}_t] - \Pr[Z_\tau = j, Z_{\tau-1} = i|\mathscr{Y}_{t-1}] \right\} \\
&- \frac{1}{p_{Ki}} \left\{ \sum_{\tau=2}^{t-1} \Pr[Z_\tau = K, Z_{t-1} = i|\mathscr{Y}_t] - \Pr[Z_\tau = K, Z_{\tau-1} = i|\mathscr{Y}_{t-1}] \right\} \\
&+ \sum_{Z_1=1}^{K} \frac{\partial \log \Pr[Z_1; \mathscr{Y}_1]}{\partial p_{ji}} \left\{ \Pr[Z_1|\mathscr{Y}_t] - \Pr[Z_1|\mathscr{Y}_{t_1}] \right\} \\
&\quad (i = 1, \ldots, K,\ j = 1, \ldots, K-1,\ t = 2, \ldots, T)
\end{aligned} \tag{3.16}$$

and

$$\frac{\partial l(y_1|\mathscr{Y}_1; \boldsymbol{\theta})}{\partial p_{ji}} = \sum_{Z_1=1}^{K} \frac{\partial \log \Pr[Z_1; \mathscr{Y}_1]}{\partial p_{ji}} \Pr[Z_1|\mathscr{Y}_1] \tag{3.17}$$

where for the probability of initial state Z_1 we use the ergodic probabilities from equation (2.68) (Hamilton 1996, p. 135-137). Please note that in the Markov model of the K-th order, the redundant parameters p_{Ki} ($i = 1, \ldots, K$) have been omitted. For the model with $K = 2$ it is common that probabilities of staying in the regime are estimated (p_{11} and p_{22}) and the probabilities of changing the regime (p_{21} and p_{12}) are treated as the redundant parameters which are omitted. This approach was employed in this work.

3.5.2 Score test of Newey-Tauchen-White

The first order condition for maximum likelihood estimation states that the score has to be equal to zero. If one assumes that for instance, the score $\boldsymbol{h}_t(p_{11})$ were positive, it would imply that $\Pr[Z_t = 1|Z_{t-1} = 1, \mathscr{Y}_{t-1}] > p_{11}$. This in turn, means that according to the information set $\mathscr{Y}_t$ the probability that observations y_t and y_{t-1} came from the first regime were higher that the estimated probability $\widehat{p}_{11}$. It would suggest that the data contain some additional information not captured in the estimated model. Analogically, for a properly specified model, the expected score should be equal to zero.

If one assume that the scores $\boldsymbol{h}_t(p_{11})$ and $\boldsymbol{h}_{t-1}(p_{11})$ were positive serially correlated, it would mean that based on the knowledge that $Z_{t-1} = 1$ and $Z_{t-2} = 1$ the probability $\Pr[Z_t = 1|Z_{t-1} = 1, Z_{t-2} = 1] > p_{11}$. This would be the equivalent to claiming that the probability $\Pr[Z_t = 1|Z_{t-1} = 1, Z_{t-2} = 1]$ was greater $\Pr[Z_t = 1|Z_{t-1} = 1]$ which is a violation of the Markov chain assumption (Hamilton 1996, p. 140).

In other words, if the model is properly estimated it is impossible to forecast the score $\boldsymbol{h}_t(\boldsymbol{\theta}_0)$ knowing the score $\boldsymbol{h}_{t-1}(\boldsymbol{\theta}_0)$. Basing on this idea White (1987) constructed his test for the lack of serial correlation in scores. It is based on the conditional moment test of Newey (1985) and Tauchen (1985). Hamilton (1996) adapted this approach for the Markov switching models and called it Newey-Tauchen-White (NTW) score test.

To compute the NTW test statistic the vectors $\boldsymbol{h}_t(\widehat{\boldsymbol{\theta}})$ have to be constructed (for $t = 1, \ldots, T$) first. This can be done by vertically stacking scores from equations (3.12) and (3.14) for $\boldsymbol{\beta_j}$ $(j = 1, \ldots, K)$, from equations (3.12) and (3.15) for σ_j^2 $(j = 1, \ldots, K)$ and (3.16)-(3.17) for p_{ji} $(j = 1, \ldots, K,$ $i = 1, \ldots, K-1)$.[5]

In the second step, the $(m \times m)$-matrix $\boldsymbol{H}_t(\widehat{\boldsymbol{\theta}}) = [\boldsymbol{h}_t(\widehat{\boldsymbol{\theta}})][\boldsymbol{h}_{t-1}(\widehat{\boldsymbol{\theta}})]'$ has to be constructed. In the next step the elements of the $\boldsymbol{H}_t(\widehat{\boldsymbol{\theta}})$ matrix which are to be tested with the NTW test are chosen. These elements should be collected in the $(l \times 1)$-vector $\boldsymbol{c}_t(\widehat{\boldsymbol{\theta}})$. If one intended to test whether the score $\boldsymbol{h}_t(\widehat{\mu_1})$ can be forecasted on the basis of the previous score $\boldsymbol{h}_{t-1}(\widehat{\mu_1})$, one should choose the (1,1)-element of $\boldsymbol{H}_t(\widehat{\boldsymbol{\theta}})$ which represents $\boldsymbol{h}_t(\widehat{\mu_1}) \cdot \boldsymbol{h}_{t-1}(\widehat{\mu_1})$.

On the basis of the vector $\boldsymbol{c}_t(\widehat{\boldsymbol{\theta}})$ the NTW test statistic can be computed.

$$NTW = \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{c}_t(\widehat{\boldsymbol{\theta}})\right]' \widehat{\boldsymbol{A}}_{(2,2)} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{c}_t(\widehat{\boldsymbol{\theta}})\right] \tag{3.18}$$

where matrix $\widehat{\boldsymbol{A}}_{(2,2)}$ is a (2,2)-sub-matrix of the matrix

$$\widehat{\boldsymbol{A}} = \left(\frac{1}{T}\begin{bmatrix} \sum_{t=1}^{T}[\boldsymbol{h}_t(\widehat{\boldsymbol{\theta}})][\boldsymbol{h}_t(\widehat{\boldsymbol{\theta}})]' & \sum_{t=1}^{T}[\boldsymbol{h}_t(\widehat{\boldsymbol{\theta}})][\boldsymbol{c}_t(\widehat{\boldsymbol{\theta}})]' \\ \sum_{t=1}^{T}[\boldsymbol{c}_t(\widehat{\boldsymbol{\theta}})][\boldsymbol{h}_t(\widehat{\boldsymbol{\theta}})]' & \sum_{t=1}^{T}[\boldsymbol{c}_t(\widehat{\boldsymbol{\theta}})][\boldsymbol{c}_t(\widehat{\boldsymbol{\theta}})]' \end{bmatrix}\right)^{-1}. \tag{3.19}$$

If the model is correctly specified, the test (3.18) statistic is asymptotically $\chi^2(l)$ distributed (Hamilton 1996, p. 131).

Hamilton (1996, p. 139-140) proposed constructing the following Newey-Tauchen-White tests:

1. *The NTW dynamic specification test for autocorrelation across regimes.*

   In this case, the null hypothesis

[5] For $K = 2$ the scores with respect to p_{11} and p_{22} are commonly used instead of scores with respect to p_{11} and p_{12}, respectively.

$\mathrm{H}_0^{(7)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{\mu}_j)\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{\mu}_i)\right] = 0$

is tested against the alternative

$\mathrm{H}_1^{(7)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{\mu}_j)\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{\mu}_i)\right] \neq 0.$

The vector $\boldsymbol{c}_t(\widehat{\boldsymbol{\theta}})$ consists of

$$\frac{\partial l(y_t|\mathscr{Y}_{t-1};\widehat{\boldsymbol{\theta}})}{\partial \mu_j} \cdot \frac{\partial l(y_{t-1}|\mathscr{Y}_{t-1};\widehat{\boldsymbol{\theta}})}{\partial \mu_i} \quad (i,j=1,\ldots,K) \tag{3.20}$$

and the NTW-statistic (3.18) is asymptotically $\chi^2(K^2)$ distributed.

2. *The NTW dynamic specification test for ARCH effects across regimes.*

   In this case, the null hypothesis

   $\mathrm{H}_0^{(8)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{\sigma}_j^2)\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{\sigma}_i^2)\right] = 0$

   is tested against the alternative

   $\mathrm{H}_1^{(8)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{\sigma}_j^2)\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{\sigma}_i^2)\right] \neq 0.$

   The vector $\boldsymbol{c}_t(\widehat{\boldsymbol{\theta}})$ consists of

$$\frac{\partial l(y_t|\mathscr{Y}_{t-1};\widehat{\boldsymbol{\theta}})}{\partial \sigma_j^2} \cdot \frac{\partial l(y_{t-1}|\mathscr{Y}_{t-1};\widehat{\boldsymbol{\theta}})}{\partial \sigma_i^2} \quad (i,j=1,\ldots,K) \tag{3.21}$$

   and the NTW-statistic (3.18) is asymptotically $\chi^2(K^2)$ distributed.

3. *The NTW dynamic specification test for validity of Markov assumptions.*

   In this case, the null hypothesis

   $\mathrm{H}_0^{(9)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{p}_{jj})\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{p}_{ii})\right] = 0$

   is tested against the alternative

   $\mathrm{H}_1^{(9)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{p}_{jj})\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{p}_{ii})\right] \neq 0.$

   The vector $\boldsymbol{c}_t(\widehat{\boldsymbol{\theta}})$ consists of[6]

$$\frac{\partial l(y_t|\mathscr{Y}_{t-1};\widehat{\boldsymbol{\theta}})}{\partial p_{jj}} \cdot \frac{\partial l(y_{t-1}|\mathscr{Y}_{t-1};\widehat{\boldsymbol{\theta}})}{\partial p_{jj}} \quad (j=1,\ldots,K) \tag{3.22}$$

[6] In this case, the scores $\frac{\partial l\left(y_t\middle|\mathscr{Y}_{t-1};\widehat{\boldsymbol{\theta}}\right)}{\partial p_{jj}} \cdot \frac{\partial l\left(y_{t-1}\middle|\mathscr{Y}_{t-1};\widehat{\boldsymbol{\theta}}\right)}{\partial p_{ii}}$ $(i,j=1,\ldots,K)$ could be considered additionally. These are, however, relevant only in very large samples, they will therefore be omitted here (Hamilton 1996, p. 140).

and the NTW-statistic (3.18) is asymptotically $\chi^2(K)$ distributed.

The results of the Newey-Tauchen-White test are listed in Tables H.1-H.13 in Appendix H. Table 3.6 summarizes these results, showing which tests cannot be rejected at the 5% confidence level. Generally speaking, the majority of the tests could not be rejected. This means that the Markov switching specification correctly models the tested portfolios. For the heteroscedastic models without regime switching in the mean equation, only portfolios with a 70% or an 80% bond proportion have additional serial correlation in the intercept (this applies to the model with and without an auto-regression term). No additional ARCH effects were detected for any of the tested portfolios (with and without an auto-regression term in the mean equation). The null hypothesis of the Markov chain assumption could be rejected only for a 90% bond portfolio for both tested models and, additionally, for a 40% bond portfolio for the model with the auto-regression term.

The test for the homoskedastic models with a regime dependent mean equation shows similar results. Hardly any test could be rejected at the 5% confidence level. The assumption of serial correlation in intercept scores could be rejected only in two cases: for portfolios with a 75% and an 80% bond proportion for the model without an auto-regression term. For the MS(2-1)-AR(1) model, the null hypothesis of the lack of serial correlation in the intercept could not be rejected for any tested portfolio. The hypothesis of no additional ARCH effects was rejected only for the model with the auto-regression term for portfolios with an 80% and a 90% bond proportion. The MS(2-1) model did not show any additional ARCH effects. The hypothesis of the Markov chain could not be rejected for any portfolio of either of the MS(2-1)-typed models.

Table 3.6: Overview of the results of the Newey-Tauchen-White test (1.1975-12.2004)

Portfolio composition		**MS(1-2)**						**MS(2-1)**						**MS(2-2)**					
		AR(0)			**AR(1)**			**AR(0)**			**AR(1)**			**AR(0)**			**AR(1)**		
REXP	**DAX30**	$H_0^{(7)}$	$H_0^{(8)}$	$H_0^{(9)}$	$H_0^{(7)}$	$H_0^{(8)}$	$H_0^{(9)}$	$H_0^{(7)}$	$H_0^{(8)}$	$H_0^{(9)}$	$H_0^{(7)}$	$H_0^{(8)}$	$H_0^{(9)}$	$H_0^{(7)}$	$H_0^{(8)}$	$H_0^{(9)}$	$H_0^{(7)}$	$H_0^{(8)}$	$H_0^{(9)}$
0%	100%	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
10%	90%	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
20%	80%	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
25%	75%	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
30%	70%	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
40%	60%	+	+	+	+	+		+	+	+	+	+	+	+	+	+	+	+	+
50%	50%	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
60%	40%	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
70%	30%		+	+		+	+	+	+	+	+	+	+	+	+	+	+	+	+
75%	25%	+	+	+	+	+	+		+	+	+	+	+	+	+	+	+	+	+
80%	20%		+	+		+	+		+	+	+		+	+	+	+	+	+	+
90%	10%	+	+		+	+		+	+	+	+		+			+			+
100%	0%	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Note:

MS(m-s)-AR(p) denotes a Markov switching model with m mean equations, s regimes for the variance and p auto-regression lags in each mean equation. + - the null hypothesis cannot be rejected at the 5% confidence level.

$H_0^{(7)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{\mu}_j)\,\middle|\,\boldsymbol{h}_{t-1}^c(\widehat{\mu}_i)\right] = 0$ $\quad H_1^{(7)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{\mu}_j)\,\middle|\,\boldsymbol{h}_{t-1}^c(\widehat{\mu}_i)\right] \neq 0$ $\quad (i,j = 1,\dots,K)$

$H_0^{(8)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{\sigma}_j^2)\,\middle|\,\boldsymbol{h}_{t-1}^c(\widehat{\sigma}_i^2)\right] = 0$ $\quad H_1^{(8)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{\sigma}_j^2)\,\middle|\,\boldsymbol{h}_{t-1}^c(\widehat{\sigma}_i^2)\right] \neq 0$ $\quad (i,j = 1,\dots,K)$

$H_0^{(9)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{p}_{jj})\,\middle|\,\boldsymbol{h}_{t-1}^c(\widehat{p}_{ii})\right] = 0$ $\quad H_1^{(9)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{p}_{jj})\,\middle|\,\boldsymbol{h}_{t-1}^c(\widehat{p}_{ii})\right] \neq 0$ $\quad (i,j = 1,\dots,K)$

The heteroskedastic models with regime dependent mean equation also fit the German bond-stock portfolios very well. The assumption that there are no additional serial correlations in the intercept could be rejected only for a 90%-10% bond-stock portfolio for both models: with and without an auto-regression in the mean equation. Exactly the same result holds true for the assumption of no additional ARCH effects. It could not be rejected for the portfolio with a 90% bond proportion for both the MS(2-2) and the MS(2-2)-AR(1) models. The null hypothesis of the Markov chain could not be rejected for any of the heteroskedastic models with a regime-dependent mean and no auto-regression. For the model with an auto-regression only for portfolios with a 25% and a 30% bond proportion, this null hypothesis could be rejected.

It is noteworthy that the NWT test at the 5% significance level gives condradictory results. The homoscedastic MS(2-1)-typed models do not show any additional ARCH effects. However, if one added stochastic volatility (i.e. if one wanted to test MS(1-2)-typed and MS(2-2)-typed models) the result of the test would be the same. The same result applies to the dependence of the intercept on the regime. The MS(1-2)-typed models do not show any additional serial correlation in the intercept. However, models with a regime-dependent mean equation (i.e. MS(2-1)-typed and MS(2-2)-typed models) show the same test result. Thus, the test does not provide any unequivocal answer for the model choice problem. It merely shows that there is evidence for the regime effects in German mixed bond-stock portfolios. It could be argued, however, that one should choose a parsimonious model. Due to this decision criterion, one should refuse MS(2-2)-typed models in favor of the MS(1-2)-typed or MS(2-1)-typed models. However, the question as to which of the two is more suitable to describe tested portfolios remains.

Another way to solve this issue would be to look at the 1% confidential level. According to Hamilton (1996), this would be the right method for small sample tests. However, one should be aware that samples used in this work comprise 360 observations, which cannot be considered as a small sample. Nevertheless, the NTW test at the 1% confidence level gives similar results.

3.5.3 Lagrange multiplier test of Hamilton (1996)

Another application of scores is the Lagrange multiplier test (see Hamilton (1996)). Suppose that the $(m \times 1)$-parameter vector $\boldsymbol{\theta}$ were estimated with the constraint that the last m_0 parameter is equal to zero. According to the first order condition of the MLE, the first $(m - m_0)$ elements of the average score $\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{h}_t(\widehat{\boldsymbol{\theta}})$ of the restricted maximum likelihood estimate $\widehat{\boldsymbol{\theta}}$ are zero but the reminding m_0 elements are unequal to zero. The size of the nonzero elements indicates how far the likelihood function might rise if the constraints were relaxed. At the same time, the magnitude of the last m_0 elements allows us to verify the validity of the constraints.

According to Hamilton (1996, p. 132) the LM statistic equals

$$LM = \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \boldsymbol{h}_t(\widehat{\boldsymbol{\theta}})\right]' \left[\frac{1}{T}\sum_{t=1}^{T}[\boldsymbol{h}_t(\widehat{\boldsymbol{\theta}})][\boldsymbol{h}_t(\widehat{\boldsymbol{\theta}})]'\right]^{-1} \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \boldsymbol{h}_t(\widehat{\boldsymbol{\theta}})\right] \qquad (3.23)$$

and is asymptotically $\chi^2(m_0)$ distributed. Hamilton (1996, p. 142) proposed testing Markov switching model against the alternative that allows for autocorrelation in regression residuals

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_{Z_t} + \rho(y_{t-1} - \boldsymbol{x}_{t-1}'\boldsymbol{\beta}_{Z_{t-1}}) + \sigma_{Z_t}\varepsilon_t, \quad \text{with } \varepsilon_t \sim N(0,1), \qquad (3.24)$$

where ρ denotes the autocorrelation coefficient and transition probabilities are given as (2.67). It is straightforward to see that for $\rho = 0$ the model (3.24) is reduced to the Markov switching model.

The null hypothesis of the autocorrelation test would thus be

$\mathrm{H}_0^{(10)}$: $\rho = 0$ (no autocorrelation)

against the alternative that

$\mathrm{H}_1^{(10)}$: $(y_t|\mathscr{X}_t, Z_t, Z_{t-1}; \boldsymbol{\theta}, \rho) \sim N(\{\boldsymbol{x}_t'\boldsymbol{\beta}_{Z_t} + \rho(y_{t-1} - \boldsymbol{x}_{t-1}'\boldsymbol{\beta}_{Z_{t-1}})\}, \sigma^2_{Z_t})$ (Hamilton 1996, p. 143).

To implement the ML test, the score with respect to $\rho = 0$ has to be calculated as

$$\left.\frac{\partial \log p(y_t|\mathscr{X}_t; \theta, \rho)}{\partial \rho}\right|_{\rho=0} = \sum_{i=1}^{K}\sum_{j=1}^{K} \psi_{t,j,i} p(Z_t = j, Z_{t-1} = i|\mathscr{Y}_t; \boldsymbol{\theta})$$

$$+\sum_{\tau=1}^{t-1}\sum_{i=1}^{K}\sum_{j=1}^{K} \psi_{t,j,i}[p(Z_\tau = j, Z_{\tau-1} = i|\mathscr{Y}_t; \boldsymbol{\theta}) - \psi_{t,j,i} p(Z_\tau = j, Z_{\tau-1} = i|\mathscr{Y}_{t-1}; \boldsymbol{\theta})] \tag{3.25}$$

where

$$\begin{aligned}\psi_{t,j,i} &= \left.\frac{\partial \log p(y_t|\boldsymbol{x}_t, \boldsymbol{x}_{t-1}, y_{t-1}, Z_t = j, Z_{t-1} = i; \boldsymbol{\theta}, \rho)}{\partial \rho}\right|_{\rho=0} \\ &= \frac{(y_t - \boldsymbol{x}_t'\boldsymbol{\beta}_j)(y_{t-1} - \boldsymbol{x}_{t-1}'\boldsymbol{\beta}_i)}{\sigma_j^2}\end{aligned} \tag{3.26}$$

Then one constructs the $\boldsymbol{h}_t(\widehat{\boldsymbol{\theta}}, \widehat{\rho})$ as defined in equations (3.12), (3.14), (3.15), (3.16)-(3.17) and stuck it on the score defined in (3.25) and (3.26). Based on the score $\boldsymbol{h}_t(\widehat{\boldsymbol{\theta}}, \widehat{\rho})' = (\boldsymbol{h}_t(\widehat{\boldsymbol{\theta}})', \boldsymbol{h}_t(\widehat{\rho}))'$ one can compute the LM statistically (3.23) which is asymptotically $\chi^2(1)$ distributed.

In addition to the LM autocorrelation test, Hamilton (1996) proposed an LM test for the ARCH effects. The non-restricted model would then be

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_{Z_t} + \sqrt{h_t}\varepsilon_t, \quad \text{where } h_t = \sigma^2_{Z_t}\left[1 + \frac{\xi(y_{t-1} - \boldsymbol{x}_{t-1}'\boldsymbol{\beta}_{Z_{t-1}})^2}{\sigma^2_{Z_{t-1}}}\right],\ \varepsilon_t \sim N(0,1), \tag{3.27}$$

with transition probabilities given as (2.67). It is fairly easy to see that for $\xi = 0$ the volatility of the model equals $\sigma^2_{Z_t}$ and the model (3.27) is reduced to the Markov switching model.

The null hypothesis of the test for the ARCH effects is

$\mathrm{H}_0^{(11)}$: $\xi = 0$ (no autocorrelation)

and the corresponding alternative is

$\mathrm{H}_1^{(11)}$: $(y_t|\mathscr{X}_t, Z_t, Z_{t-1}; \boldsymbol{\theta}, \xi) \sim N(\boldsymbol{x}_t'\boldsymbol{\beta}_{Z_t}, h_t) h_t = \sigma_{Z_t}^2 \left[1 + \frac{\xi(y_{t-1} - \boldsymbol{x}_{t-1}'\boldsymbol{\beta}_{Z_{t-1}})^2}{\sigma_{Z_{t-1}}^2}\right]$

To compute the LM statistic (3.23) one has to construct vector $\boldsymbol{h}_t(\widehat{\boldsymbol{\theta}}, \widehat{\xi})$ as defined above and vector $\boldsymbol{h}_t(\widehat{\xi})$ using equation (3.25) with

$$\begin{aligned} \psi_{t,j,i} &= \left.\frac{\partial \log p(y_t|\boldsymbol{x}_t, y_{t-1}, Z_t = j, Z_{t-1} = i; \boldsymbol{\theta}, \xi)}{\partial \xi}\right|_{\xi=0} \\ &= \left[-1 + \frac{(y_t - \boldsymbol{x}_t'\boldsymbol{\beta}_{Z_t})^2}{\sigma_{Z_t}^2}\right]\left[\frac{(y_{t-1} - \boldsymbol{x}_{t-1}'\boldsymbol{\beta}_{Z_{t-1}})^2}{2\sigma_{Z_{t-1}}^2}\right]. \end{aligned} \tag{3.28}$$

Then the vector $\boldsymbol{h}_t(\widehat{\boldsymbol{\theta}}, \widehat{\xi})' = (\boldsymbol{h}_t(\widehat{\boldsymbol{\theta}})', \boldsymbol{h}_t(\widetilde{\xi}))'$ can be used to compute the LM statistic (3.23) which is asymptotically $\chi^2(1)$ distributed.

In Appendix I, Tables I.1-I.13 the results of the Lagrange multiplier test are listed. Table 3.7 summarizes the results at the 5% confidence level. For the model MS(1-2) the null hypothesis of no additional correlation in the mean equation can be rejected for most portfolios. The exceptions are portfolios with a 70%, a 75%, and an 80% bond proportion. The opposite holds true for the null hypothesis of no additional ARCH effects, which can be rejected for most of the portfolios. Only for portfolios with low bond proportion (from 0% to 25%) and the pure bond portfolio the null hypothesis cannot be rejected. For the MS(1-2)-AR(1) model, both null hypotheses, of no additional auto-correlation in the mean equation and no additional ARCH effects could be rejected for most of the portfolios. The null hypothesis of no autocorrelation could not be rejected for a 40%, a 90%, and a 100% of bond proportion only. The null hypothesis of no additional ARCH effects could not be rejected for portfolios with a 70%, a 90% and a 100% of bond engagement only.

Table 3.7: Overview of the results of the Lagrange multiplier test (1.1975-12.2004)

Portfolio composition		MS(1-2)				MS(2-1)				MS(2-2)			
		AR(0)		AR(1)		AR(0)		AR(1)		AR(0)		AR(1)	
REXP	DAX30	$H_0^{(10)}$	$H_0^{(11)}$	$H_0^{(10)}$	$H_0^{(11)}$	$H_0^{(10)}$	$H_0^{(11)}$	$H_0^{(10)}$	$H_0^{(11)}$	$H_0^{(10)}$	$H_0^{(11)}$	$H_0^{(10)}$	$H_0^{(11)}$
0%	100%	+	+			+	+	+	+	+	+	+	+
10%	90%	+	+			+	+	+	+	+	+	+	+
20%	80%	+	+			+		+	+	+	+	+	+
25%	75%	+	+	+				+	+	+		+	+
30%	70%	+						+	+	+		+	+
40%	60%	+						+	+	+			+
50%	50%	+						+	+	+	+		
60%	40%	+				+		+	+	+	+		
70%	30%					+	+	+	+	+		+	+
75%	25%				+	+	+	+	+	+			
80%	20%					+	+	+	+			+	
90%	10%	+		+	+	+		+	+			+	+
100%	0%	+	+	+	+	+	+	+	+	+	+	+	+

Note:
MS(m-s)-AR(p) denotes a Markov switching model with m mean equations, s regimes for the variance and p auto-regression lags in each mean equation. + - the null hypothesis cannot be rejected at the minimum 5% confidence level.

$H_0^{(10)}$: $\rho = 0$ $H_1^{(10)}$: $(y_t|\mathscr{X}_t, z_t, z_{t-1}; \boldsymbol{\theta}, \rho) \sim N(\{\boldsymbol{x}_t'\boldsymbol{\beta}_{z_t} + \rho(y_{t-1} - \boldsymbol{x}_{t-1}'\boldsymbol{\beta}_{z_{t-1}})\}, \sigma_{z_t}^2)$

$H_0^{(11)}$: $\xi = 0$ $H_1^{(11)}$: $(y_t|\mathscr{X}_t, z_t, z_{t-1}; \boldsymbol{\theta}, \xi) \sim N\left(\boldsymbol{x}_t'\boldsymbol{\beta}_{z_t}, \sigma_{z_t}^2\left[1 + \frac{\xi(y_{t-1} - \boldsymbol{x}_{t-1}'\boldsymbol{\beta}_{z_{t-1}})^2}{\sigma_{z_{t-1}}^2}\right]\right)$

For the MS(2-1) model, the null hypothesis of no autocorrelation cannot be rejected for most of the portfolios. Only for portfolios with a 25%, a 30%, a 40%, and a 50% bond investment, the model is not correctly specified with respect to autocorrelation in the mean equation. The null hypothesis of no additional ARCH effects could not be rejected for 6 out of 13 portfolios. The hypothesis of no additional ARCH effects could be rejected for portfolios with a 0%, a 10%, a 70%, a 75%, an 80%, and a 100% bond engagement. The situation is better for the MS(2-1)-AR(1) model. For this model, both null hypotheses, of no additional autocorrelation and no additional ARCH effects could not be rejected. Thus, it is the most suitable model according to the LM test.

For model MS(2-2), the null hypothesis of no additional autocorrelation could not be rejected for most portfolios. The rejection occurred only for portfolios with an 80% and a 90% bond investment. The null hypothesis of no additional ARCH effects could be rejected for 7 out of 13 models. These are portfolios with a 25%, a 30%, a 40%, a 70%, a 75%, an 80%, and a 90% bond engagement. In the case of the MS(2-2)-AR(1) model, both null hypotheses could be rejected for 4 out of 13 models. The null hypothesis of no additional autocorrelation could be rejected for portfolios with a 40%, a 50%, a 60%, and a 75% bond proportion. The hypothesis of no additional heteroscedastic effects could be rejected for models with a 50%, a 60%, a 75%, and an 80% bond investment.

In conclusion, the ML test shows that the MS(2-1)-AR(1) model shows no additional autocorrelation in the regression residuals for any of the tested portfolios. The remaining models fit the tested samples quite well with respect to the ML autocorrelation test. The exception is the MS(1-2)-AR(1) model which shows additional autocorrelation for all but three samples. According to the additional ARCH effects, the MS(2-1)-AR(1) model is clearly the best, as it passes the LM-ARCH test for all tested portfolios. In contrast, the MS(1-2), MS(1-2)-AR(1), MS(2-1), and MS(2-2) models perform rather poorly. This suggests that for portfolios which fail the ARCH effect test, an additional regime in volatility or model with switching in GARCH could be estimated.[7] The MS(2-2)-AR(1) model performs a little better, as it fails to model the stochastic volatility for only 4 out of 13 tested portfolios.

[7]For Markov switching models with GARCH effects, see Hamilton and Susmel (1994), Cai (1994), Gray (1996) or Haas (2004).

3.6 Tests based on simulation of the test statistic

Several authors used the Monte Carlo simulation for testing Markov switching models. Lam (1990, p. 427-428), Cai (1994, p. 313), and Rydén, Teräsvirta, and Åsbrink (1998, p. 224-225) assume that the true model is a restricted linear model (e.g. AR(p)) with known parameter vector $\boldsymbol{\theta}$. They use $\boldsymbol{\theta}$ to draw a random sample from a restricted model and estimate the non-restricted Markov regime model from the simulated sample. In this manner, they were able to compute the likelihood function for both models, and, eventually, the LR statistic. The repetition of the experiment allowed them to approximate the distribution of the test statistic and its critical values.

The author of this work is skeptical to the bootstrap method used in the context of testing regime switching models. In Section 3.3 it was already mentioned that the likelihood function of the regime models features several local optima. Hence, for each simulated sample, the numerous iteration of the EM algorithm would be needed. Suppose that one wanted to run the bootstrap algorithm 1000 times, and the estimation algorithm 50 times for each output. Then, for all 13 samples and 6 MS models which were estimated in Section 2.9.2 one would have to repeat the EM algorithm 3,900,000 times; which is 250 times more than the number of EM algorithm runs required for estimation of these 6 models for all 13 portfolios, and would last several months. Another possibility would be to run the estimation algorithm only once for each bootstrapped sample as Lam (1990) and Cai (1994) did. This simplification, however, bears the risk that one would not find the global optimum, and that the distribution of the test statistic would be biased. For this reason, the author prefers to use the asymptotic distribution theory to simulate the distribution of the test statistic.

As already mentioned in Section 3.3 the standard asymptotic theory does not hold, as the transition probabilities are not defined under the null hypothesis. Davies (1977,1987) was the first to study the problem of undefined nuisance parameters. He proposed defining the test statistic as a function of the undefined parameter. He then implemented the empirical process theory for supremum likelihood ratio and Lagrange multiplier tests. Andrews and Ploberger (1994) extended the theory with an average exponential likelihood ratio, Wald, and Lagrange multiplier tests. Hansen (1996b) proposed a method for directly computing the critical values via a Monte Carlo simulation. For this simulation, he applied the covariance function of the test statistic. However, these models do not address the problem of identical zero scores. If the first derivative of the likelihood function is zero, Lee and Chesher (1986) proposed looking at derivatives of a higher order. Using this approach, they found that the Lagrange multiplier statistic is χ^2 distributed if higher order derivatives are unequal to zero. Similar results were found for the likelihood ratio and the modified Wald test.

The above mentioned works cannot be applied to tests for Markov switching models, as Hamilton's model includes both the nuisance parameters and the zero scores problem. The following section will introduce the Hansen (1992, 1996a) test constructed for Markov switching models and its extension proposed by Garcia (1998).

3.6.1 Likelihood ratio test of Hansen (1992, 1996)

Before Hansen's (1992, 1996a) test is introduced, a small re-parametrization of the model (2.66) is needed. Let us order the regimes in such a way that $\sigma_1 \leq \sigma_2 \leq \ldots \sigma_K$ then, without a loss of generality, the Markov switching model can be rewritten as

$$\begin{aligned}
y_t &= \mu_1 + \mu_2^* \mathbb{I}_{[Z_t=2]} + \cdots + \mu_K^* \mathbb{I}_{[Z_t=K]} \\
&\quad + (\phi_{1(1)} + \phi_{1(2)}^* \mathbb{I}_{[Z_t=2]} + \cdots + \phi_{1(K)}^* \mathbb{I}_{[Z_t=K]}) y_{t-1} \\
&\quad + \ldots \\
&\quad + (\phi_{p(1)} + \phi_{p(2)}^* \mathbb{I}_{[Z_t=2]} + \cdots + \phi_{p(K)}^* \mathbb{I}_{[Z_t=K]}) y_{t-r} \\
&\quad + (\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_t=2]} + \cdots + \sigma_K^* \mathbb{I}_{[Z_t=K]}) \varepsilon_t \\
&= \mu_1 + \sum_{j=2}^{K} \mu_j^* \mathbb{I}_{[Z_t=j]} + \sum_{i=1}^{r} \left(\phi_{i(1)} + \sum_{j=2}^{K} \phi_{i(j)}^* \mathbb{I}_{[Z_t=i]} \right) y_{t-i} \\
&\quad + \left(\sigma_1 + \sum_{j=2}^{K} \sigma_j^* \mathbb{I}_{[Z_t=j]} \right) \varepsilon_t, \qquad \text{and } \varepsilon_t \sim \mathbb{N}(0,1),
\end{aligned} \tag{3.29}$$

with $\mu_j^* = \mu_j - \mu_1$, $\phi_{i(j)}^* = \phi_{i(j)} - \phi_{i(1)}$ and $\sigma_j^* = \sigma_j - \sigma_1$ for $j = 2, \ldots, K,\ i = 1, \ldots, r$ and with transition probabilities given in (2.67). Then the parameter vector $\boldsymbol{\theta}$ becomes $\boldsymbol{\theta}^{*\prime} = (\mu_1, \mu_2^*, \ldots \mu_K^*, \phi_{1(1)}, \phi_{1(2)}^*, \ldots, \phi_{1(K)}^*, \ldots,$ $\phi_{p(1)}, \phi_{p(2)}^*, \ldots, \phi_{p(K)}^*, \sigma_1^2, \sigma_2^{*2}, \ldots, \sigma_K^{*2}, p_{11}, \ldots, p_{1K}, \ldots, p_{K-1,K}, \ldots, p_{K-1,K})'$.

Hansen (1992, p. S63) divided the parameter vector $\boldsymbol{\theta}^{*\prime} = (\boldsymbol{\gamma}_1', \boldsymbol{\gamma}_2', \boldsymbol{\delta})'$ into three categories, so that vector $\boldsymbol{\gamma}_1$ contains all parameters which are needed to specify the restricted model, vectors $\boldsymbol{\gamma}_2$ and $\boldsymbol{\delta}$ contain nuisance parameters, where $\boldsymbol{\gamma}_2$ is fully identified under the null hypothesis and $\boldsymbol{\delta}$ is not identified. For instance, if one wanted to test the AR(1) model against the MS(2-1)-AR(1) then $\boldsymbol{\gamma}_1' = (\mu, \phi_{1(1)}, \sigma_1)'$, $\boldsymbol{\gamma}_2' = (\mu_2^*, \phi_{1(2)}^*)'$, and $\boldsymbol{\delta}' = (p_{11}, p_{22})'$. Or if one wanted to test MS(2-1) against MS(3-3) then $\boldsymbol{\gamma}_1' = (\mu_1, \mu_2^*, \sigma_1, p_{11}, p_{12}, p_{21}, p_{22})'$, $\boldsymbol{\gamma}_2' = (\mu_3^*, \sigma_2^*, \sigma_3^*)'$, and $\boldsymbol{\delta}' = (p_{13}, p_{23})'$.

The re-parametrization introduced in equation (3.29) allows to write the log-likelihood function as

$$\mathscr{L}_T(\boldsymbol{\theta}) = \mathscr{L}_T(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\delta}) = \sum_{t=1}^{T} l_t(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\delta})$$

with the according null hypothesis

$\mathrm{H}_0^{(12)}$: $\boldsymbol{\gamma}_2 = \mathbf{0}$, with undefined parameter vector $\boldsymbol{\delta}$,

and its alternative hypothesis

$\mathrm{H}_1^{(12)}$: $\boldsymbol{\gamma}_2 \neq \mathbf{0}$ (Hansen 1992, p. S66).

Hansen (1992, p. S66-S67) proposed "eliminating" the problem of the nuisance parameters through the concentration of vectors $\boldsymbol{\gamma}_2$ and $\boldsymbol{\delta}$ into one vector, i.e., he sets $\boldsymbol{\alpha}^H = (\boldsymbol{\gamma}_2', \boldsymbol{\delta}')'$. This results with $\mathscr{L}_T(\boldsymbol{\alpha}^H, \boldsymbol{\gamma}_1)$ and $l_t(\boldsymbol{\alpha}^H, \boldsymbol{\gamma}_1)$, respectively. If one fixed the values of $\boldsymbol{\alpha}^H$ one could find the (pseudo true) values for the vector $\boldsymbol{\gamma}_1$

$$\boldsymbol{\gamma}_1(\boldsymbol{\alpha}^H) = \underset{\boldsymbol{\gamma}_1 \in \Gamma_1}{\arg\max} \lim_{T \to \infty} \frac{1}{T} \mathbb{E} L_T(\boldsymbol{\gamma}_1, \boldsymbol{\alpha}^H),$$

where Γ_1 denotes the compact parameter space for $\boldsymbol{\gamma}_1$. Then, for a sufficiently large sample size T the centered likelihood function is given by

$$\mathscr{L}_T(\boldsymbol{\alpha}^H) = \mathscr{L}_T(\boldsymbol{\gamma}_1(\boldsymbol{\alpha}^H), \boldsymbol{\alpha}^H).$$

Hansen (1992, p. S63-S64, S67-S68) defines the LR test statistic (3.8) as a function

$$LR_T(\boldsymbol{\alpha}^H) = \mathscr{L}_T(\boldsymbol{\gamma}_1, \boldsymbol{\alpha}^H) - \mathscr{L}_T(\boldsymbol{\gamma}_1, \mathbf{0}, \boldsymbol{\delta}) = \sum_{t=1}^{T} [l_t(\boldsymbol{\gamma}_1, \boldsymbol{\alpha}^H) - l_t(\boldsymbol{\gamma}_1, \mathbf{0}, \boldsymbol{\delta})], \quad (3.30)$$

which yields a sequence of Neyman-Pearson likelihood ratio test statistics for the null hypothesis against each simple alternative hypothesis. This is a rarely used definition, but it has the advantage that the LR test statistic for the null hypothesis against the alternative is the lowest upper bound of the likelihood ratio surface

$$LR_T = \sup_{\boldsymbol{\alpha}^H \in A^H} LR_T(\boldsymbol{\alpha}^H),$$

where A^H denotes a compact parameter space over the vector $\boldsymbol{\alpha}^H$. Please note that the LR statistic can be observed, but its mean

$$R_T(\boldsymbol{\alpha}^H) = \mathbb{E}[LR_T(\boldsymbol{\alpha}^H)] \quad (3.31)$$

cannot. Thus, let us define the deviation from the mean process as

$$\widehat{Q}_T(\boldsymbol{\alpha}^H) = LR_T(\boldsymbol{\alpha}^H) - R_T(\boldsymbol{\alpha}^H) = \sum_{t=1}^{T} \widehat{q}_t(\boldsymbol{\alpha}^H), \tag{3.32}$$

where

$$q_t(\boldsymbol{\alpha}^H) = [l_t(\boldsymbol{\gamma}_1, \boldsymbol{\alpha}^H) - l_t(\boldsymbol{\gamma}_1, \mathbf{0}, \boldsymbol{\delta})] - \mathbb{E}[l_t(\boldsymbol{\gamma}_1, \boldsymbol{\alpha}^H) - l_t(\boldsymbol{\gamma}_1, \mathbf{0}, \boldsymbol{\delta})].$$

Then the LR statistic from equation (3.30) can be decomposed as

$$LR_T(\boldsymbol{\alpha}^H) = R_T(\boldsymbol{\alpha}^H) + Q_T(\boldsymbol{\alpha}^H). \tag{3.33}$$

Since the deviation process is stochastic, some errors caused by determining the likelihood ratio can occur. The stochasticity of $Q_T(\boldsymbol{\alpha}^H)$ causes that the likelihood function can be maximized at some value $(\boldsymbol{\gamma}_1', \boldsymbol{\alpha}^{H\prime})'$ other than the null hypothesis $(\boldsymbol{\gamma}_1', \mathbf{0}', \boldsymbol{\alpha}^{H\prime})'$.

Let us standardize the LR process in such a way that it has the same variance for all $\boldsymbol{\alpha}^H \in A^H$

$$\widehat{LR}_T^*(\boldsymbol{\alpha}^H) = \frac{\widehat{LR}_T(\boldsymbol{\alpha}^H)}{\sqrt{V_T(\boldsymbol{\alpha}^H)}} \tag{3.34}$$

and

$$\widehat{LR}_T^* = \sup_{\boldsymbol{\alpha}^H \in A^H} \widehat{LR}_T(\boldsymbol{\alpha}^H), \tag{3.35}$$

where

$$V_T(\widehat{\boldsymbol{\gamma}}_1(\boldsymbol{\alpha}^H), \boldsymbol{\alpha}^H) = \sum_{t=1}^{T} q_t(\widehat{\boldsymbol{\gamma}}_1(\boldsymbol{\alpha}^H), \boldsymbol{\alpha}^H)^2$$

denotes the sample variance with

$$q_t(\widehat{\boldsymbol{\gamma}}_1(\boldsymbol{\alpha}^H), \boldsymbol{\alpha}^H,) = l_t(\widehat{\boldsymbol{\gamma}}_1(\boldsymbol{\alpha}^H), \boldsymbol{\alpha}^H) - l_t(\widehat{\boldsymbol{\gamma}}_1(\boldsymbol{\alpha}^H), \mathbf{0}, \boldsymbol{\delta}) - \frac{1}{T}\widehat{LR}_T(\boldsymbol{\alpha}^H).$$

The standardized stochastic deviation process will subsequently be equal to

$$Q_T^*(\boldsymbol{\alpha}^H) = \frac{Q_T(\boldsymbol{\alpha}^H)}{\sqrt{V_T(\boldsymbol{\alpha}^H)}} \ .$$

Hansen (1992) proved that the standardized likelihood ratio statistic (3.35) is bounded by the standardized deviation process and that this standardized deviation process has a limit:

Theorem 3.1 *Under Assumptions A.1, A.2, and A.3 from Appendix A.1,*

$$\Pr[\widehat{LR}_T^* \geq x] \leq \Pr[\sup_{\boldsymbol{\alpha}^H \in A^H} \widehat{Q}_T^*(\boldsymbol{\alpha}^H) \geq x] \to \Pr[Sup\ Q^* \geq x]$$

(Hansen 1992, p. S69).

Proof. The proof is given in Hansen (1992, p. S66-S69). ■

This theorem gives the bound of the standardized LR statistic as the distribution of the random variable SupQ^*.

To test the $\mathrm{H}_0^{(12)}$ against the $\mathrm{H}_1^{(12)}$ one has to determine the distribution of the random variable SupQ^*. Hansen (1992) proposed using the following theorem

Theorem 3.2 (Theorem 1 from Hansen (1996b)[8]) *Under Assumptions A.4, A.5, and A.6 from the Appendix A.1 and the absence of the serial correlation and heteroscedasticity in the noise function*

$$LR_T \xrightarrow{p} SupC \equiv \sup_{\boldsymbol{\delta} \in \Delta} C(\boldsymbol{\delta}) \tag{3.36}$$

where $\xrightarrow{p}$ *denotes weak convergence with respect to the uniform metric and* $Q(\boldsymbol{\delta})$ *is a chi-square process with a covariance matrix* $\overline{K}(\cdot,\cdot)$*, defined as follows:*

$$\overline{K}(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2) = \iota_k V(\boldsymbol{\delta}_1)^{-1} K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2) V(\boldsymbol{\delta}_2)^{-1} \iota_k' \tag{3.37}$$

where ι_k *is a vector of dimension* k *(the dimension of the parameter vector under an alternative hypothesis) with ones in the positions of the parameters constrained to be zero under the null hypothesis, and zeros on the remaining positions*

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2) = \lim_{T \to \infty} T\ \mathbb{E}\left[\boldsymbol{h}_T^c(\boldsymbol{\alpha}_0^G, \boldsymbol{\delta}_1)\boldsymbol{h}_T^c(\boldsymbol{\alpha}_0^G, \boldsymbol{\delta}_2)'\right], \tag{3.38}$$

where

$$\boldsymbol{h}_T^c(\boldsymbol{\theta}) = \frac{\partial}{\partial \theta_i} \overline{\mathscr{L}}_T(\boldsymbol{\theta}) \tag{3.39}$$

and

$$\overline{\mathscr{L}}_T(\boldsymbol{\theta}) = \frac{1}{T} \sum_t^T \ln l_t(\boldsymbol{\theta})$$

and the variance matrix

$$V(\boldsymbol{\theta}) = \lim_{T\to\infty} T \; \mathbb{E}\left[\boldsymbol{h}_T^c(\boldsymbol{\alpha}_0^G, \boldsymbol{\delta})\boldsymbol{h}_T^c(\boldsymbol{\alpha}_0^G, \boldsymbol{\delta})'\right]$$

$$V(\boldsymbol{\delta}) = V(\boldsymbol{\alpha}_0^G, \boldsymbol{\delta})$$

(Garcia 1998, p. 768).

Proof. For proof of the general version of Theorem 3.2 see Hansen (1996b, p. 425-426). ■

The covariance function (3.37) for the Hansen test is given by

$$\widehat{K}_T^*(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2) = \frac{\widehat{K}_T(\boldsymbol{\alpha}_1^H, \boldsymbol{\alpha}_2^H)}{\sqrt{V_T(\boldsymbol{\alpha}_1^H) V_T(\boldsymbol{\alpha}_2^H)}}, \tag{3.40}$$

with

$$\begin{aligned}\widehat{K}_T(\boldsymbol{\alpha}_1^H, \boldsymbol{\alpha}_2^H) &= \sum_{i=1}^T \widehat{q}_t(\boldsymbol{\alpha}_1^H)\widehat{q}_t(\boldsymbol{\alpha}_2^H) \\ &+ \sum_{k=1}^M w_{kM} \left[\sum_{t=1}^{T-k} \widehat{q}_t(\boldsymbol{\alpha}_1^H)\widehat{q}_{t+k}(\boldsymbol{\alpha}_2^H) \; \sum_{t=1+k}^T \widehat{q}_t(\boldsymbol{\alpha}_1^H)\widehat{q}_{t-k}(\boldsymbol{\alpha}_2^H)\right]\end{aligned}$$

and $\widehat{q}_t(\boldsymbol{\alpha}^H) = q_t(\boldsymbol{\alpha}^H, \widehat{\boldsymbol{\gamma}}_1(\boldsymbol{\alpha}^H))$ and $w_{kM} = 1 - \frac{|k|}{M+1}$ being the Bartlett kernel and M a bandwidth number, which should slowly be increased as the sample size grows (Hansen 1996a, p. 195-196).

Now, it is possible to accomplish the test. Suppose that one can draw Gaussian processes with the covariance function $\widehat{K}_T^*(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)$ given in equation (3.40). According to Theorem 3.2, the supremum of each of these processes has (approximately) the distribution SupQ^*.

Hansen (1996a, p. 196) suggests the following algorithm for obtaining the draws of SupQ^*. One should draw $T + M$ iid $N(0, 1)$ variables $\{u_i\}_1^{T+M}$. Using the sequence of u-s the simulated empirical distribution of

$$\widehat{LR}^*(\boldsymbol{\alpha}^H) = \frac{\sum_{k=0}^{M} \sum_{t=1}^{T} q_t(\boldsymbol{\alpha}^H, \widehat{\boldsymbol{\gamma}}_1(\boldsymbol{\alpha}^H))u_{t+k}}{\sqrt{(M+1)V_n(\boldsymbol{\alpha}^H)}}$$

can be determined, which is Gaussian with a zero mean and a covariance function $\widehat{K}_T(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)$. Simulating a lot of $\widehat{LR}^*(\boldsymbol{\alpha}^H)$ will yield the distribution of the test statistic. Note that the theory gives no information on how to choose M, therefore the simulation should be accomplished with several values of M in order to assess the sensitivity of the test with respect to M. The last unsolved problem is to find $\widehat{\gamma}_1(\boldsymbol{\alpha}^H)$ which can be addressed by conducting the grid search. However, this method is very time-consuming. Thus, there is a trade-off between the precision of the simulation (i.e. the size of the grid-step) and the length of the computation time (Hansen 1992, p. S70).

3.6.2 Likelihood ratio test of Garcia (1998)

The Hansen test has two significant drawbacks. It applies a grid technique, which is very time consuming and is therefore only applicable to several simple cases. Hansen (1992) studied some simple Markov switching models and found that they are insensitive to the choice of the grid. However, this does not necessarily hold true for more complicated models. The second drawback is that the outcome of the Hansen test is the upper bound for the likelihood ratio statistic and is not a critical value. This may imply that the test is conservative (Garcia 1998, p. 766).

For this reason, Garcia (1998) proposed a modification of Hansen's (1992, 1996a) test, which enables us to derive the covariance matrix K analytically. He suggested ordering the vectors in a different manner as Hansen did. Hansen splits the parameter vector $\boldsymbol{\theta}$ into the parameters of interests $\boldsymbol{\alpha}^H = (\boldsymbol{\gamma}_2', \boldsymbol{\delta}')'$ and the nuisance parameter vector $\boldsymbol{\gamma}_1$.

Garcia redefines the parameter of interest as $\boldsymbol{\alpha}^G = (\boldsymbol{\gamma}_1', \boldsymbol{\gamma}_2')'$ and the remaining transition probabilities $\boldsymbol{\delta}$ as the nuisance parameter vector. In Garcia's (1998) approach the null hypothesis

$\mathrm{H}_0^{(12)} \qquad \boldsymbol{\gamma}_2 = \mathbf{0}$, with the undefined transition parameter $\boldsymbol{\delta}$,

and the alternative

$\mathrm{H}_1^{(12)} \qquad \boldsymbol{\gamma}_2 \neq \mathbf{0}$.

remain the same as proposed by Hansen (1992).

Garcia (1998) proceeds as Hansen (1992) does and uses the Theorem 3.2 from Section 3.6.1 to compute the covariance matrix (3.38). In the first step, the scores defined as (3.39) should be computed, since in our thesis, the Markov model (2.66)-(2.67) differs from Garcia's (1998) definition and cannot be taken over. The scores used in this model are given in the following lemma.

Lemma 3.3 *The elements of the score vector $\boldsymbol{h}_T^c(\boldsymbol{\alpha}^G, \boldsymbol{\delta})$, evaluated at the true value of the parameter of interest $\boldsymbol{\alpha}_0^G$ and at the particular given value of the nuisance parameter $\boldsymbol{\delta}$ are given by*

$$\boldsymbol{h}_T^c(\boldsymbol{\alpha}_0^G, \boldsymbol{\delta})_{\mu_1} = \frac{1}{T} \sum_{t=1}^{T} \sum_{Z_t(\boldsymbol{\delta})=1}^{2} \frac{\varepsilon_t}{(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_t(\boldsymbol{\delta})=2]})^2} p_t \tag{3.41}$$

$$\boldsymbol{h}_T^c(\boldsymbol{\alpha}_0^G, \boldsymbol{\delta})_{\mu_2^*} = \frac{1}{T} \sum_{t=1}^{T} \sum_{Z_t(\boldsymbol{\delta})=1}^{2} \frac{\varepsilon_t \mathbb{I}_{[Z_t(\boldsymbol{\delta})=2]}}{(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_t(\boldsymbol{\delta})=2]})^2} p_t \tag{3.42}$$

$$\boldsymbol{h}_T^c(\boldsymbol{\alpha}_0^G, \boldsymbol{\delta})_{\phi_{i(1)}} = \frac{1}{T} \sum_{t=1}^{T} \sum_{Z_t(\boldsymbol{\delta})=1}^{2} \frac{y_{t-i}\varepsilon_t}{(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_t(\boldsymbol{\delta})=2]})^2} p_t, \qquad i = 1, \ldots, r \tag{3.43}$$

$$\boldsymbol{h}_T^c(\boldsymbol{\alpha}_0^G, \boldsymbol{\delta})_{\phi_{i(2)}^*} = \frac{1}{T} \sum_{t=1}^{T} \sum_{Z_t(\boldsymbol{\delta})=1}^{2} \frac{y_{t-i}\varepsilon_t \mathbb{I}_{[Z_t(\boldsymbol{\delta})]}}{(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_t(\boldsymbol{\delta})=2]})^2} p_t, \qquad i = 1, \ldots, r \tag{3.44}$$

$$\boldsymbol{h}_T^c(\boldsymbol{\alpha}_0^G,\boldsymbol{\delta})_{\sigma_1^2} = \frac{1}{T}\sum_{t=1}^{T}\sum_{Z_t(\boldsymbol{\delta})=1}^{2}\frac{1}{2\sigma_1(\sigma_1+\sigma_2^*\mathbb{I}_{[Z_t(\boldsymbol{\delta})=2]})}\left(\frac{\varepsilon_t^2}{(\sigma_1+\sigma_2^*\mathbb{I}_{[Z_t(\boldsymbol{\delta})=2]})^2}-1\right)p_t \tag{3.45}$$

$$\boldsymbol{h}_T^c(\boldsymbol{\alpha}_0^G,\boldsymbol{\delta})_{\sigma_2^{*2}} = \frac{1}{T}\sum_{t=1}^{T}\sum_{Z_t(\boldsymbol{\delta})=1}^{2}\frac{\mathbb{I}_{[Z_t(\boldsymbol{\delta})=2]}}{2\sigma_2^*(\sigma_1+\sigma_2^*\mathbb{I}_{[Z_t(\boldsymbol{\delta})=2]})}\left(\frac{\varepsilon_t^2}{(\sigma_1+\sigma_2^*\mathbb{I}_{([Z_t(\boldsymbol{\delta})=2]})^2}-1\right)p_t \tag{3.46}$$

with $p_t = \Pr[Z_t(\boldsymbol{\delta}) = z_t(\boldsymbol{\delta})|\mathscr{Y}_T;\boldsymbol{\alpha}^G,\boldsymbol{\delta}]$.

Proof. For the proof see Appendix A.2. ■

By using Lemma 3.3 one can compute the covariance of the stochastic variable SupQ^* as given in the following lemma.

Lemma 3.4 *The covariance matrix* $K(\boldsymbol{\delta}_1,\boldsymbol{\delta}_2)$ *of the score vectors, as defined in section 2, is equal to*

$$K(\boldsymbol{\delta}_1,\boldsymbol{\delta}_2) = \frac{1}{\sigma_1^2}\begin{array}{cc} \begin{pmatrix} \boldsymbol{M} & \boldsymbol{X}^{\boldsymbol{\mu},\boldsymbol{\phi}}_{2\times 2r} & \boldsymbol{O}_{2\times 2} \\ \boldsymbol{X}^{\boldsymbol{\phi},\boldsymbol{\mu}}_{2r\times 2} & \boldsymbol{X}^{\boldsymbol{\phi},\boldsymbol{\phi}}_{2r\times 2r} & \boldsymbol{O}_{2r\times 2} \\ \boldsymbol{O}_{2\times 2} & \boldsymbol{O}_{2\times 2r} & \boldsymbol{\Sigma} \end{pmatrix} & \begin{matrix} \boldsymbol{\mu} \\ \boldsymbol{\phi} \\ \boldsymbol{\sigma} \end{matrix} \\ \begin{matrix} \quad\boldsymbol{\mu}' & \qquad\quad \boldsymbol{\phi}' & \qquad \boldsymbol{\sigma}' \end{matrix} & \end{array} \tag{3.47}$$

where

$$\boldsymbol{M} = \begin{pmatrix} 1 & \pi_2(\boldsymbol{\delta}_2) \\ \pi_2(\boldsymbol{\delta}_1) & \min[\pi_2(\boldsymbol{\delta}_1),\pi_2(\boldsymbol{\delta}_2)] \end{pmatrix}$$

$$\boldsymbol{X}^{\boldsymbol{\mu},\boldsymbol{\phi}}_{2\times 2r} = \mu_1 \times (\boldsymbol{M}\otimes \mathbf{1}_{1\times r})$$

$$\boldsymbol{X}^{\boldsymbol{\phi},\boldsymbol{\mu}}_{2r\times 2} = \mu_1 \times (\boldsymbol{M}\otimes \mathbf{1}_{r\times 1})$$

$$\boldsymbol{X}^{\boldsymbol{\phi},\boldsymbol{\phi}}_{2r\times 2r} = (\boldsymbol{M}\otimes(\boldsymbol{R}_{r\times r}+2\mu_1^2))$$

$$\boldsymbol{\Sigma} = \boldsymbol{M}\odot\begin{pmatrix} \frac{1}{2\sigma_1^2} & \frac{1}{2\sigma_1\sigma_2^*} \\ \frac{1}{2\sigma_1\sigma_2^*} & \frac{1}{2\sigma_2^{*2}} \end{pmatrix}$$

and $\boldsymbol{O}$ *and* $\mathbf{1}$ *denote a matrix of zeros and ones respectively,* r *- order of the auto-regression,* $\boldsymbol{R}$ *- an auto-covariance matrix of* $\{y\}$ *of order* r*,* $\otimes$ *- the Kronecker product and* $\odot$ *- the element-by-element matrix-product.*

Proof. For the proof see Appendix A.2. ■

It is noteworthy that the covariance matrix for the Markov switching models (2.66) which are used in this work depend on the ergodic probabilities $\boldsymbol{\pi}$ but not on the transition probabilities (2.67) or the auto-regression coefficients $\phi_{i(1)}$, $\phi^*_{i(j)}$ $(i = 1, \ldots, r,\ j = 1, \ldots, K)$. This is an advantage in comparison to the model used by Garcia (1998). The explicit formula for models used in this work are given in the following lemma.

Lemma 3.5 *Let introduce the following notation.* $\pi_1 = \pi_2(\boldsymbol{\delta}_1)$, $\pi_2 = \pi_2(\boldsymbol{\delta}_2)$ *and* c_{11} *- the auto-covariance of the vector* $\mathscr{Y}_T$*. Thus, the covariance matrix for models MS(1-2) and MS(1-2)-AR(1) is given by*

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\sigma_2^{*2}} = \frac{2\left(\min[\pi_1, \pi_2] - \pi_1\pi_2\right)\sigma_1^2\sigma_2^{*2}}{\left(\pi_1 - 1\right)\pi_1\left(\pi_2 - 1\right)\pi_2}; \tag{3.48}$$

the covariance matrix for model MS(2-1) is given by

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_2^*} = \frac{\left(\min[\pi_1, \pi_2] - \pi_1\pi_2\right)\sigma_1^2}{\left(\pi_1 - 1\right)\pi_1\left(\pi_2 - 1\right)\pi_2} \tag{3.49}$$

the covariance matrix for model MS(2-1)-AR(1) is given by

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2) = \begin{pmatrix} K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_2^*} & 0 \\ 0 & K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\phi^*_{1(2)}} \end{pmatrix}, \tag{3.50}$$

with

$$\begin{aligned} K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_2^*} = {} & \frac{\left(2\mu_1^2 + c_{11}\right)\sigma_1^2}{\left(\pi_1 - 1\right)\pi_1\left(\pi_2 - 1\right)\pi_2\left(\mu_1^2 + c_{11}\right)^2} \\ & \times \left\{\left(\pi_1^2 - 3\pi_2\pi_1 + \pi_2^2\right)\mu_1^2 - c_{11}\pi_1\pi_2 + \min[\pi_1, \pi_2]\left(\mu_1^2 + c_{11}\right)\right\} \end{aligned} \tag{3.51}$$

and

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\phi^*_{1(2)}} = \frac{\left(\left(\pi_1^2 - 3\pi_2\pi_1 + \pi_2^2\right)\mu_1^2 - c_{11}\pi_1\pi_2 + \min[\pi_1, \pi_2]\left(\mu_1^2 + c_{11}\right)\right)\sigma_1^2}{\left(\pi_1 - 1\right)\pi_1\left(\pi_2 - 1\right)\pi_2\left(\mu_1^2 + c_{11}\right)^2}; \tag{3.52}$$

the covariance matrix for model MS(2-2) is given by

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2) = \begin{pmatrix} K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_2^*} & 0 \\ 0 & K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\sigma_2^{*2}} \end{pmatrix}, \tag{3.53}$$

with

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_2^*} = \frac{(\min[\pi_1, \pi_2] - \pi_1\pi_2)\,\sigma_1^2}{(\pi_1 - 1)\,\pi_1\,(\pi_2 - 1)\,\pi_2} \tag{3.54}$$

and

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\sigma_2^{*2}} = \frac{2\,(\min[\pi_1, \pi_2] - \pi_1\pi_2)\,\sigma_1^2\sigma_2^{*2}}{(\pi_1 - 1)\,\pi_1\,(\pi_2 - 1)\,\pi_2}; \tag{3.55}$$

and the covariance matrix for model MS(2-2)-AR(1) is given by

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2) = \begin{pmatrix} K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_2^*} & 0 & 0 \\ 0 & K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\phi_{1(2)}^*} & 0 \\ 0 & 0 & K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\sigma_2^{*2}} \end{pmatrix}, \tag{3.56}$$

with

$$\begin{aligned} K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_2^*} = {} & \frac{(2\mu_1^2 + c_{11})\,\sigma_1^2}{(\pi_1 - 1)\,\pi_1\,(\pi_2 - 1)\,\pi_2\,(\mu_1^2 + c_{11})^2} \\ & \times \left\{ \left(\pi_1^2 - 3\pi_2\pi_1 + \pi_2^2\right) \mu_1^2 - c_{11}\pi_1\pi_2 + \min[\pi_1, \pi_2] \left(\mu_1^2 + c_{11}\right) \right\} \end{aligned} \tag{3.57}$$

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\phi_{1(2)}^*} = \frac{((\pi_1^2 - 3\pi_2\pi_1 + \pi_2^2)\,\mu_1^2 - c_{11}\pi_1\pi_2 + \min[\pi_1, \pi_2]\,(\mu_1^2 + c_{11}))\,\sigma_1^2}{(\pi_1 - 1)\,\pi_1\,(\pi_2 - 1)\,\pi_2\,(\mu_1^2 + c_{11})^2}, \tag{3.58}$$

and

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\sigma_2^{*2}} = \frac{2\,(\min[\pi_1, \pi_2] - \pi_1\pi_2)\,\sigma_1^2\sigma_2^{*2}}{(\pi_1 - 1)\,\pi_1\,(\pi_2 - 1)\,\pi_2}. \tag{3.59}$$

Proof. The proof results straightforward from the application of Theorem 3.2 and Lemma 3.4. ■

As one knows the analytic solution for the covariance matrix $K(\boldsymbol{\delta}_i, \boldsymbol{\delta}_j)$, it is possible to simulate the distribution of the Garcia test statistic. The input is vector $\boldsymbol{\alpha}^H$ and, additionally, for models with an auto-regression, the sample

$\mathscr{Y}_T = (y_1, \ldots, y_T)'$. One should bear in mind that the nuisance vector $\boldsymbol{\delta}$ is undefined under the null hypothesis and, thus, some grid search is needed. Lemma 3.4 shows that if one wants to test the null of no regimes against the alternative of two states, the test statistic does not depend directly on the transition probabilities p_{11} and p_{22}. Without a loss of generality, one can assume that $\pi_1 \leq \pi_2$. Thus, according to Lemma 3.4, the covariance of the test statistic depends only on the ergodic probability π_2 which is given in the equation (2.21) as $\frac{1-p_{11}}{2-p_{11}-p_{22}}$. Therefore, the nuisance parameter vector $\boldsymbol{\delta}$ can be reduced to π_2. From this, it follows that the grid search can only be conducted for the parameter π_2. Thus, we construct a sequence of vectors $\{\boldsymbol{\delta}_i\}_{i=1}^{k}$ so that all $\pi_2(\boldsymbol{\delta}_i) \in \Delta = (0, 1)$ and are equidistant. E.g. let $\pi_2(\boldsymbol{\delta}_i) = 0.001, 0.002, \ldots, 0.999$. The border cases 0 and 1 have to be excluded from the parameter space Δ because at these points, the condition of a non-singular information matrix will be violated. These cases, however, are not of interest for the test, as $\pi_2 = 0$ or $\pi_2 = 1$ can occur only if the Markov chain is reducible and one of the regimes will vanish (Garcia 1998, p. 772-773).

In the second step, one computes the covariance matrices $K(\boldsymbol{\delta}_i, \boldsymbol{\delta}_j)$ for all i, j $(i, j = 1, \ldots, K)$ given by equation (3.47) and collects them to a matrix

$$\boldsymbol{\Omega}_{(k\cdot m)\times(k\cdot m)} = \begin{bmatrix} K_{m\times m}(\boldsymbol{\delta}_1, \boldsymbol{\delta}_1) & \cdots & K_{m\times m}(\boldsymbol{\delta}_1, \boldsymbol{\delta}_k) \\ \vdots & \ddots & \vdots \\ K_{m\times m}(\boldsymbol{\delta}_k, \boldsymbol{\delta}_1) & \cdots & K_{m\times m}(\boldsymbol{\delta}_k, \boldsymbol{\delta}_k) \end{bmatrix}, \tag{3.60}$$

where m denotes the dimension of the vector $\boldsymbol{\gamma}_2$ which determines the dimension of the matrix $K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_1)$.

Then, one computes the matrix $\boldsymbol{B}$ which is the lower triangular matrix from the Cholesky decomposition $\boldsymbol{\Omega} = \boldsymbol{B}\boldsymbol{B}'$. Garcia (1998, p. 786-787, Appendix 4) proposes an algorithm to compute the matrix $\boldsymbol{B}$ analytically, however, this work favors the numeric solution to make the programming code more efficient.

The next step is to draw a $(k \cdot m \times 1)$-vector $\boldsymbol{u}$ of i.i.d. standard normal variables and to construct a $(k \cdot m \times 1)$-vector $G(\boldsymbol{\delta}) = \boldsymbol{Bu}$ of mean zero normal variables with a covariance matrix $\boldsymbol{\Omega}$. Now, the computation of the χ^2 distributed random variable $Q(\boldsymbol{\alpha}^H)$ is fairly straightforward.

Remark 3.6 *Following the remark of Garcia (1998, p. 768, footnote 6) that a $\chi^2(n)$ process $Z(\boldsymbol{\theta})$ can be represented as $Z(\boldsymbol{\theta}) = G(\boldsymbol{\theta}_i)'K(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j)^{-1}G(\boldsymbol{\theta}_j)$ where $G(\boldsymbol{\theta})$ is a Gaussian $(n \times 1)$-vector with a zero mean and a covariance function given by $K(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j) = \mathbb{E}[G(\boldsymbol{\theta}_i) \cdot G(\boldsymbol{\theta}_j')]$.*

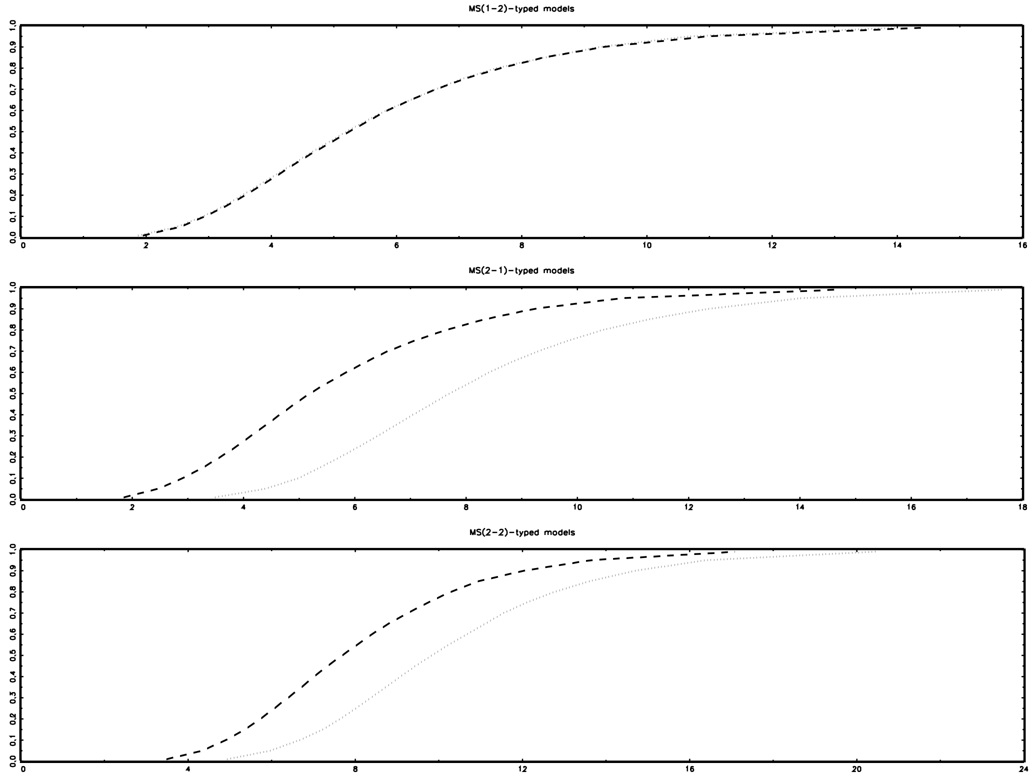
Thus, let us divide vector $G(\boldsymbol{\delta})$ into k vertically stacked $(m \times 1)$-vectors $g(\boldsymbol{\delta}_i)$ and compute the $Q(\boldsymbol{\delta}) = g(\boldsymbol{\delta}_i)'V(\boldsymbol{\delta}_i, \boldsymbol{\delta}_i)g(\boldsymbol{\delta}_i)$ for all $i = 1, \ldots, K$. Eventually, it should be possible to compute the supremum of Q: $\mathrm{Sup}Q = \max_{i=1,\ldots,K} Q(\boldsymbol{\delta}_i)$.

The replication of the algorithm N times allows us to determine the distribution of the $\mathrm{Sup}C$ statistic and its critical values.

Tables J.1-J.13 from Appendix J show the distribution of Garcia's SupQ test statistic. This distribution was simulated within a Monte Carlo approach with 10,000 iterations. The μ_1, μ_2, σ_1, and σ_2 parameters were chosen according to the maximum likelihood estimates listed in Tables B.1-B.13. For a π_2 parameter, a grid search was accomplished. The lower bound of the grid denoted 0.001, the upper bound 0.999 and increment 0.001, respectively. The distribution of the test statistic is similar for all estimated portfolios. The distribution of Garcia's SupC statistic for MS(1-2), MS(1-2)-AR(1), and MS(2-1)[9] models is almost the same (see Figure 3.1 for the example of the

[9] The distribution of the MS(1-2) and the MS(1-2)-AR(1) is not "similar" but actually analytically identical, as they have the same covariance, see equation (3.48).

Figure 3.1: Distribution of the Garcia's SupC statistic for DAX30



Note:
The top panel shows the distribution of the Garcia's SupC statistic for MS(1-2)-typed models, the middle panel for the MS(2-1)-typed models, and, the bottom panel for the MS(2-2)-typed models. The dashed line represents the model without the auto-regression term and the dotted line with the auto-regression term in the mean equation, respectively.

pure DAX30 portfolio). The test statistic distribution for models MS(2-1)-AR(1) and MS(2-2) is similar and stochastically dominated by the distribution of the three above mentioned models. The statistic distribution for the MS(2-2)-AR(1) model is different from all other models discussed here, but is stochastically dominated by them. A closer look at these results shows that the distribution is similar for models with the same number of parameters contained in the $\boldsymbol{\gamma}_2$ vector. Moreover, the distribution of the test statistic seems to be independent from the magnitude of the parameter as it varies in different models and seems to be independent from the type of parameter

Table 3.8: Garcia test (1.1975-12.2004)

Portfolio composition		**MS(1-2)**		**MS(2-1)**		**MS(2-2)**	
REXP	**DAX30**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
0%	100%	67.62***	67.76***	50.99***	54.03***	68.96***	70.37***
10%	90%	63.96***	64.18***	47.52***	50.59***	65.18***	66.59***
20%	80%	59.60***	59.89***	43.74***	46.81***	60.68***	62.10***
25%	75%	57.04***	57.38***	41.70***	44.72***	58.09***	59.47***
30%	70%	54.18***	54.53***	39.52***	42.47***	55.18***	56.53***
40%	60%	47.27***	47.64***	34.57***	37.31***	48.20***	49.45***
50%	50%	38.22***	38.53***	28.58***	30.92***	39.13***	40.17***
60%	40%	27.07***	26.87***	21.17***	22.85***	27.91***	28.51***
70%	30%	15.86***	15.17***	12.42**	13.25*	17.24***	18.05**
75%	25%	12.14**	12.10**	8.24	10.26	13.81**	15.03*
80%	20%	11.23**	11.36**	5.55	11.65	12.51*	15.59*
90%	10%	18.94***	20.30***	9.67*	18.92***	19.87***	22.60***
100%	0%	32.24***	35.71***	17.87***	22.98***	32.32***	36.17***

Note:
MS(m-s)-AR(p) denotes a Markov switching model with m mean equations, s regimes for the variance and p auto-regression lags in each mean equation. *, **, *** - the null hypothesis can be rejected at a 10%, a 5% and a 1% confidence level, respectively.
H_0: the time series follows an auto-regressive model of the p-th order, with no Markov switching, H_1: the time series follows a MS(m-s)-AR(p) model.

contained in the $\boldsymbol{\gamma}_2$ vector (e.g. for the MS(1-2) model $\boldsymbol{\gamma}_2 = \sigma_2^{*2}$ and the MS(2-1) model $\boldsymbol{\gamma}_2 = \mu_2^*$ and the distribution is similar). Furthermore, it seems that the results do not depend on sample $\mathscr{Y}_T$, as the distribution of the SupQ statistic is similar for all tested samples. If it were true that the distribution of the test statistic is dependent on the number of elements in the $\boldsymbol{\gamma}_2$ vector, the impact of other determinants as a magnitude of the parameter could be disregarded. It would imply that the Garcia test statistic could be tabulated, which would considerably simplify the testing of Markov switching models. It would be a very interesting field of study to test this supposition but it is out of the scope of this work.

Table 3.8 shows the results of the Garcia LR test. These results are based on the distribution of the SupC statistic listed in Tables J.1-J.13 in Appendix J. The Garcia test shows that the null hypothesis of no regime switching can be rejected for most of the tested cases at the 5% confidence level. For a heteroscedastic Markov model with no regime in the mean equation, the null hypothesis can be rejected for all portfolios, irrespective of whether the model includes an auto-regression term or not. For a homoskedastic Markov model and a regime-dependent mean equation, the null hypothesis of no regime specification can be clearly rejected for all but three portfolios: with a 75%, an 80%, and a 90% bond engagmet in the case of the MS(2-1) and with 70%, 75%, and 80% bond investment in the MS(2-1)-AR(1) case, respectively. For the heteroscedastic model with a regime-dependent mean equation, most of the portfolios are also better fitted with the regime model than with the linear models. For the MS(2-2) model, the null hypothesis can only be rejected for a portfolio with an 80% bond investment and for the MS(2-2)-AR(1) model, for portfolios with a 75% and an 80% bond engagement.

The Garcia test reports Markov switching in almost all models and samples. However, it gives no answer to the question of which of the models should be used (see discussion in Section 3.5.2).

3.7 Conclusion

This Chapter has presented several tests for Markov switching models. The majority prefers the MS(1-2)-typed models or at least indicates that they are as good as other Markov switching models. Additionally, the tests report that auto-regression effects are only present in portfolios with a very high bond exposition. This result is independent of the type of the model.

The Wald test rejects the MS(2-1)-typed models in favor of the MS(1-

2)-typed and the MS(2-2)-typed models. Given that the Wald test rejects the hypothesis of no regime switching in the intercept, the MS(2-2)-typed models seem to be overparametrized.

The LR test yields that the transition probabilities can be modeled with a Markov chain of the first order for most of the models. This assumption is rejected only for MS(1-2)-AR(1) and MS(2-1)-typed models with a minimum bond engagement of 70% and, in 2 of 13 cases, for the MS(2-2) model. Furthermore, the LR test rejected the MS(2-1)-typed models in favor of MS(2-2)-typed models, but the MS(2-2)-models are rejected in favor of the MS(1-2)-models. This leads to the conclusion that the MS(1-2) model is the best choice (as portfolios with a high bond engagement failed the Markov chain test).

The NTW test favors the MS specification for almost all models and all samples. It states that the MS approach models well-mixed portfolios of German stock and bonds. Unfortunately, it does not provide any hint as to which of these is the best.

The LM test is passed positively only by the MS(2-1)-AR(1) model. The MS(1-2)-AR(1) model fails all LM tests with the exception of portfolios with a 90% and a 100% bond exposition. For the remaining models, only approximately half of the portfolios pass the test. This test is the only one which rejects the null hypothesis of no additional ARCH effects for so many models. This suggests that for these models an extra regime or a Markov switching model with a (G)ARCH term should be tested additionally. As there is no option pricing model for Markov switching with (G)ARCH effects available, the second alternative will be neglected. The estimation of the MS model with three regimes will be discarded, as the testing of the null hypothesis of two regimes against the alternative of three regimes becomes very complicated.

The Garcia test is another test which shows that German stock-bond mixed portfolios exhibit regime switching characteristics. The MS(1-2)-typed models passed the test for all studied portfolios. For the remaining MS models, there were few portfolios with a middle high bond engagement which failed the test.

Both information criterion tests showed that the Markov switching models better fit the studied portfolios than the models from the GARCH family and the linear models. For samples with a middle high bond exposition, certain models from the GARCH family had a slightly higher test statistic. However, the difference is small. Moreover, the SBC statistic shows that for portfolios with a bond exposition between 70% and 80% the linear GBM also ranked better than MS models. This phenomenon is not observable for the AIC statistic. As the tests with the linear null hypothesis reject linearity, this one outcome will be neglected. Moreover, the AIC and SBC tests show that models with a very high bond engagement show an additional auto-regression term. This is true for all tested models, regardless of whether it was an MS, a GARCH or a linear model family.

In conclusion, the majority of the tests used here show that the Markov switching model is very useful in explaining the stochasticity of the tested portfolios. The MS(1-2)-typed models are either the best or at least as good as other MS models. MS(2-2)-typed models also fit the German portfolios. However, they seem to be a little overparametrized. Therefore, in the next part of this dissertation the MS(1-2) model will be used for pricing the guarantees embedded in personal pension products. Admittedly, the samples with 90% or more bond exposition should be modelled with the MS(1-2)-AR(1) model. Unfortunately, there is no option pricing theory in which the underlying instrument follows a Markov switching model with an auto-regression term. Therefore, the auto-regression term will be omitted. As this assumption applies for two samples only, this seems acceptable.

Part II

Investment Guarantees Embedded in Individual Pension Products

Chapter 4

Pricing of investment guarantees

This Chapter shows how to price guarantees embedded in personal pension products. In Section 4.1 we define the model of the financial market considered. Section 4.2 discusses the Contingent Claim Pricing Theorem introduced by Harrison and Pliska (1981). Section 4.3 addresses conditions for the existence of the option price and Section 4.4 its uniqueness. Section 4.5 defines the Esscher risk-neutral probability measure and shows how it can be used to price a European put option. In Section 4.6 we show how to price put options when the price of the underlying follows the geometric Brownian motion or the geometric Brownian motion with Markov switching. In the first case we use the Black and Scholes (1973) price and in the second case, the Bollen (1998)-Hardy (2001) and the Webb (2003) price. In Section 4.7 we price the cost of investment guarantees embedded in personal pension plans and discuss its sensitivity to several factors. Section 4.8 concludes the results.

4.1 Financial market

4.1.1 Money market account and risky stock

First, let us define the financial market. To do this, one has to define assets that can be traded on this market: the risk-free bond and the risky asset.

Definition 4.1 (Frictionless market) *A frictionless market is a market where there are no taxes, no transaction costs, a perfect divisibility of financial instruments, a perfect liquidity, no short-sales constraints, and no borrowing constraints.*

Definition 4.2 (Money market account) *Let $(B_t)_{t\geq 0}$ be a deterministic process defined as follows*

$$dB_t = rB_t dt,$$

where the constant r denotes the risk free rate. Furthermore, let the initial value of B equal to unity ($B_{t_0} = 1$) and let B_t be arbitrarily divisible, then it is called money market account (or risk-free bond).

Definition 4.3 (Risky asset) *Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space, let $(S_t)_{t\geq 0}$ be a positive stochastic process*

$$\frac{dS_t}{S_t} = dX_t,$$

where X_t is a stochastic variable representing the return rate of S_t. Furthermore, let S_t be arbitrarily divisible, then it is called a (non dividend paying) risky asset (e.g., stock or portfolio of stocks).

Now one can define the financial market.

Definition 4.4 (Financial market) *Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space, let $(B_t)_{t\geq 0}$ and $(S_t)_{t\geq 0}$ be a money market account and risky stock, respectively. Then the tuple $\mathcal{M} = (B_t, S_t)_{t\geq 0}$ is called a financial market. Furthermore, it will be assumed that the the financial market is frictionless.*

4.1.2 Contingent claim

Now we can define a contingent claim.

Definition 4.5 (Contingent claim) *Let $\mathcal{M}$ be a financial market defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{P})$, and for given t let H_t be a non-negative random variable measurable with respect to the filtration $\mathcal{F}_t$, and let $f : \mathbb{R} \to \mathbb{R}$ be a function. Then*

$$H_t = f(S_t)$$

is called a contingent claim.

An example of the contingent claim which we are interested in is a European option.

Example 4.6 (European call (put) option) *The European call H^C (put H^P) option is a right, but not an obligation, to buy (sell) the risky stock S at a defined price K, called exercise price, at expiration time (or maturity) T. The payoff of the call is given by the function*

$$H_T^C = (S_T - \mathsf{K})^+ \tag{4.1}$$

and the payoff of the put by

$$H_T^P = (\mathsf{K} - S_T)^+ \tag{4.2}$$

(Elliott and Kopp 2005, p. 6).

Theorem 4.7 (Put-call parity) *Let $\mathcal{M}$ be the financial market and let H_t^C and H_t^P be the call and put option with the exercise price K and maturity time T, respectively. Then the equation*

$$H_t^C - H_t^P = S_t - \mathsf{K}e^{-r(T-t)}$$

holds (Elliott and Kopp 2005, p. 9).

Proof. To prove the put-call parity, we first define two portfolios: (1) a long call and a short put position, both with the same strike price K and expiry date T, and (2) a long position in a stock S_t, and a short position in a discounted zero-bond with the value of $\mathsf{K}e^{-r(T-t)}$, the face value K, and expiry date T. From the definition of the call (see equation (4.1)) and the put option (see equation (4.2)) we know that at time T their values have to be equal

$$H_T^C - H_T^P = (S_T - \mathsf{K})^+ - (\mathsf{K} - S_T)^+ = S_T - \mathsf{K}. \tag{4.3}$$

Thus, the following has to hold true

$$H_t^C - H_t^P = S_t - \mathsf{K}e^{-r(T-t)}. \tag{4.4}$$

Otherwise, arbitrage would be possible, i.e., everyone who buys the underpriced portfolio and sells the overpriced portfolio could make a riskless profit (Elliott and Kopp 2005, p. 9). ■

4.1.3 Self-financing trading strategy

Now let us address the task of defining the trading strategy.

Definition 4.8 (Trading strategy) *Let $\mathcal{M}$ be the financial market defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{P})$, and let $\phi = (\phi_t)_{t_0 \leq t \leq T} = (\varphi_t^B, \varphi_t^S)_{t_0 \leq t \leq T}$ $\in \mathbb{R}^2$ be a measurable, stochastic vector process adapted to the filtration $\mathcal{F}_t$.*

Then ϕ is called a trading strategy. Stochastic variables φ_t^B and φ_t^S can be interpreted as the amount of riskless bond B_t and the amount of the risky asset S_t in the investor's portfolio, respectively (Elliott and Kopp 2005, p. 29).

Furthermore, we can define the portfolio wealth and gain.

Definition 4.9 (Wealth (value) and gains process) *Let $\mathcal{M}$ be a financial market defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{P})$, and let ϕ be a trading strategy. Then, for all $t \in [t_0, T]$ the process,*

$$V_t(\phi) = \varphi_t^B B_t + \varphi_t^S S_t$$

is called a portfolio wealth (portfolio value) and the process

$$G_t(\phi) = \int_{t_0}^{t} \varphi_u^B dB_u + \int_{t_0}^{t} \varphi_u^S dS_u$$

is called a gains process, respectively (Bingham and Kiesel 2004, p. 230).

Remark 4.10 *It is clear that the change in the portfolio value is dependent on the change in the value of the money market account and the change in the stock price*

$$dV_t(\phi) = \varphi_t^B dB_t + \varphi_t^S dS_t, \qquad \forall\ t \in [t_0, T]. \tag{4.5}$$

Definition 4.11 (Self-financing strategy) *Let $\mathcal{M}$ be a financial market defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{P})$, let ϕ be a trading strategy, and let $V_t(\phi)$ be a value process, satisfying the condition*

$$V_t(\phi) = V_{t_0}(\phi) + \int_{t_0}^{t} \varphi_u^B dB_u + \int_{t_0}^{t} \varphi_u^S dS_u = V_{t_0}(\phi) + G_t(\phi) \qquad \forall\ t \in [t_0, T]. \tag{4.6}$$

Then ϕ is called a self-financing strategy (Musiela and Rutkowski 2007, p. 89).

Notation 4.12 *Let $\mathbf{\Phi}$ denote the class of all self-financing trading strategies.*

Remark 4.13 *The intuition behind the self-financing strategy is the following: The agent invests an initial capital $V_{t_0}(\phi)$ in the portfolio of money market account and risky stock. Then he rebalances his investment by (continuously) trading the risk-free bond and the risky stock in such a way that he neither adds additional capital to the portfolio nor withdraws it, i.e.*

$$B_t d\varphi_t^B + S_t d\varphi_t^S = 0 \tag{4.7}$$

(Elliott and Kopp 2005, p. 183-184).

As will be seen later, it is more convenient to work with discounted values than with "real" values.

Notation 4.14 *If we introduce an intrinsic discount process*

$$\beta_t = B_{t_0} e^{-r(t-t_0)},$$

then the process

$$\widetilde{S}_t = \beta_t S_t$$

is called a discounted risky asset. By Analogy, the process

$$\widetilde{V}_t(\phi) = \beta_t V_t(\phi) = \varphi_t^B + \varphi_t^S \widetilde{S}_t$$

is called a discounted wealth, and the process

$$\widetilde{G}_t(\phi) = \int_{t_0}^{t} \varphi_u^S d\widetilde{S}_u$$

is called a discounted gains process. Obviously, the discounted money market account is equal to unity for all $t_0 \leq t \leq T$ (i.e. $\widetilde{B}_t = 1$) (Harrison and Pliska 1981, p. 236).

Proposition 4.15 *Let $\mathcal{M}$ be a financial market defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{P})$, and let ϕ be any trading strategy. Then ϕ is self-financing if and only if*

$$\widetilde{V}_t(\phi) = \widetilde{V}_{t_0}(\phi) + \widetilde{G}_t(\phi), \qquad \forall\ t \in [t_0, T] \tag{4.8}$$

(Harrison and Pliska 1981, Theorem 3.24, p. 238).

Proof. Let $t \in [t_o, T]$. If ϕ is a self-financing strategy, then

$$dV_t(\phi) = \varphi_t^B dB_t + \varphi_t^S dS_t.$$

From this it follows that

$$\begin{aligned} d\widetilde{V}_t(\phi) &= d(\beta_t V_t(\phi)) = -r\widetilde{V}_t(\phi)dt + \beta_t dV_t(\phi) \\ &= -r\beta_t(\varphi_t^B B_t + \varphi_t^S S_t)dt + \beta_t(\varphi_t^B dB_t + \varphi_t^S dS_t) \\ &= \varphi_t^S(-r\beta_t S_t dt + \beta_t dS_t) = \varphi_t^S d\widetilde{S}_t, \end{aligned}$$

which is equivalent to (4.8). The converse direction can be proven by using the definition of the discounted portfolio value $V_t(\phi) = \widetilde{V}_t(\phi)$, reversing the steps above and using (4.8) (Elliott and Kopp 2005, p. 184). ■

4.2 Option pricing

4.2.1 Equivalent martingale measure

To price contingent claims, we have to define the martingale and the martingale probability measure.

Definition 4.16 (Martingale) *Let $(X_t)_{t_0 \le t \le T}$ be a stochastic process on the probability space $(\Omega, \mathcal{F}, \mathcal{P})$. Let*

$$\mathbb{E}_{\mathcal{P}}[X_t|\mathcal{F}_s] = X_s \ (t_0 \le s \le t \le T).$$

Then X is called a $\mathcal{P}$-martingale (martingale under measure $\mathcal{P}$) (Rolski et al. 1999, p. 379).

Definition 4.17 (Equivalent (risk-neutral) martingale measure) *Let $\mathcal{P}$ and $\mathcal{Q}$ be two probability measures, and let $(S_t)_{t \geq t_0}$ be a stochastic process with associated filtration $\mathcal{F}_t$. Then*

- *$\mathcal{Q}$ is an equivalent martingale measure (or risk-neutral martingale measure) with respect to the given probability measure $\mathcal{P}$ if $\mathcal{Q}$ is equivalent to $\mathcal{P}$ ($\mathcal{Q} \sim \mathcal{P}$). This means that both probability measures have the same null set ($\forall\ A \in \Omega\ \mathrm{Pr}_{\mathcal{P}}[A] = 0 \ \Leftrightarrow\ \mathrm{Pr}_{\mathcal{Q}}[A] = 0$)*
- *and the discounted risky stock is a $\mathcal{Q}$-martingale, i.e. $\mathbb{E}_{\mathcal{Q}}[\widetilde{S}_t | \mathcal{F}_s] = \widetilde{S}_s$ ($t_0 \leq s \leq t \leq T$) (Harrison and Pliska 1981, p. 236).*

Notation 4.18 *Henceforth we will denote the set of all equivalent martingale measures of the probability measure $\mathcal{P}$ as $\mathbb{P}$.*

To change one (not necessarily martingale) probability measure to another, we have to use the Radon-Nikodým density.

Definition 4.19 (Radon-Nikodým density) *The Radon-Nikodým density of $\mathcal{Q}$ with respect to $\mathcal{P}$ is defined as the unique $\mathcal{F}_T$-measurable random variable Λ_T, such that for any event $A \in \mathcal{F}_T$ we have*

$$\Pr_{\mathcal{Q}}[A] = \int_A \Lambda_T \, d\mathcal{P}$$

(Musiela and Rutkowski 2007, p. 606).

Remark 4.20 *Definition 4.19 implies that for any $\mathcal{Q}$-integrable random variable X, we have $\mathbb{E}_{\mathcal{Q}}[X] = \mathbb{E}_{\mathcal{P}}[X\Lambda_T]$. Note also that X is $\mathcal{Q}$-integrable if and only if $X\Lambda_T$ is $\mathcal{P}$-integrable. Finally, it is easy to check that $\mathrm{Pr}_{\mathcal{P}}[\Lambda_T > 0] = 1$ and $\mathbb{E}_{\mathcal{P}}[\Lambda_T] = \mathrm{Pr}_{\mathcal{Q}}[\Omega] = 1$ (Musiela and Rutkowski 2007, p. 606).*

Notation 4.21 (Radon-Nikodým derivative) *To emphasize the role of* Λ_T *as the link between the expectations with respect to* $\mathcal{Q}$ *and* $\mathcal{P}$*, it is customary to use the short-hand notation*

$$\Lambda_T = \frac{d\mathcal{Q}}{d\mathcal{P}},$$

which is called the Radon-Nikodým derivative (Musiela and Rutkowski 2007, p. 607).

Definition 4.22 (Radon-Nikodým density process) *Let* $\mathcal{P}$ *and* $\mathcal{Q}$ *be equivalent probability measures, and let* $(\mathcal{F}_t)_{t\in[t_0,T]}$ *be a filtration. Then, for all* $t \in [t_0, T]$,

$$\Lambda_t = \mathbb{E}_{\mathcal{P}}[\Lambda_T|\mathcal{F}_t] = \mathbb{E}_{\mathcal{P}}\left[\frac{d\mathcal{Q}}{d\mathcal{P}}|\mathcal{F}_t\right]$$

is called a Radon-Nikodým density process (Musiela and Rutkowski 2007, p. 607).

Remark 4.23 *It is obvious that the process* $(\Lambda_t)_{t_0\leq t\leq T}$ *is a* $\mathcal{P}$*-martingale.*

Proposition 4.24 *A stochastic process* $(X_t)_{t_0\leq t\leq T}$ *is an* $\mathcal{F}$*-martingale under* $\mathcal{Q}$ *if and only if the process* $(X_t\Lambda_t)_{t_0\leq t\leq T}$ *is an* $\mathcal{F}$*-martingale under* $\mathcal{P}$ *(Musiela and Rutkowski 2007, p. 607).*

Proof. Assume that $(X_t\Lambda_t)_{t_0\leq t\leq T}$ is an $\mathcal{F}$-martingale under $\mathcal{P}$, so that equality $\mathbb{E}_{\mathcal{P}}[X_t\Lambda_t|\mathcal{F}_s] = X_s\Lambda_s$ holds for $t_0 \leq s \leq t \leq T$. Using the Bayes formula

$$\begin{aligned}\mathbb{E}_{\mathcal{Q}}[X_t|\mathcal{F}_s] &= \frac{\mathbb{E}_{\mathcal{P}}[X_t\Lambda_T|\mathcal{F}_s]}{\mathbb{E}_{\mathcal{P}}[\Lambda_T|\mathcal{F}_s]} = \frac{\mathbb{E}_{\mathcal{P}}[X_t\mathbb{E}_{\mathcal{P}}[\Lambda_T|\mathcal{F}_t]|\mathcal{F}_s]}{\mathbb{E}_{\mathcal{P}}[\Lambda_T|\mathcal{F}_s]} \\ &= \frac{\mathbb{E}_{\mathcal{P}}[X_t\Lambda_t|\mathcal{F}_s]}{\Lambda_s} = \frac{X_s\Lambda_s}{\Lambda_s} = X_s,\end{aligned}$$

we conclude that the stochastic process $(X_t)_{t_0\leq t\leq T}$ is an $\mathcal{F}$-martingale under $\mathcal{P}$. The proof of the converse implication goes along the same lines (Musiela and Rutkowski 2007, p. 607). ■

4.2.2 Option pricing formula

So far we considered all self-financing strategies. However, from an economic point of view, only some of these are of importance.

Definition 4.25 (Admissible strategy) *Let $\mathcal{M}$ be a financial market defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{P})$, and let ϕ be a trading strategy. If the discounted portfolio process is non-negative for all $t_0 \leq t \leq T$*

$$\widetilde{V}_t(\phi) \geq 0,$$

self-financing

$$\widetilde{V}_t(\phi) = \widetilde{V}_{t_0}(\phi) + \widetilde{G}_t(\phi),$$

and a $\mathcal{Q}$-martingale

$$\mathbb{E}_{\mathcal{Q}}[\widetilde{V}_t(\phi)|\mathcal{F}_s] = \widetilde{V}_s(\phi),$$

then ϕ is called admissible (Harrison and Pliska 1981, p. 240-241).

Notation 4.26 *Let $\Phi^\star$ denote the class of all admissible trading strategies.*

The non-negativity condition rules out some short-selling strategies. The short selling of the risky asset is allowed in general, but only if the value of the whole portfolio is non-negative. Let us now concentrate on a special class of the admissible strategies: the hedging strategies.

Definition 4.27 (Attainable contingent claim) *Let $\mathcal{M}$ be a financial market defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{P})$, and let H_T be a contingent claim. If there exists an admissible trading strategy $\phi \in \Phi^\star$ such that*

$$\widetilde{V}_T(\phi) = \beta_T H_T$$

then claim H_T is called attainable (replicable, or hedgeable). We say that ϕ generates the contingent claim H_T, and the initial capital $\mathsf{P}_{t_0} = \widetilde{V}_{t_0}(\phi)$ is *called the price of this claim (Harrison and Pliska 1981, p. 240). Then we say that the price* $\mathsf{P}_{t_0} = \widetilde{V}_{t_0}(\phi)$ *is price associated with the contingent claim* H_T.

The idea of the hedging portfolio is that, with a positive start capital V_{t_0} and continuous buying and selling bonds and stocks, the agent can track the value of the contingent claim without investing additional capital after t_0. In this manner the seller of the put can protect himself from potential loss. If he sells a put for the price of V_{t_0} and invests this amount in the self-financing trading strategy, he will avoid additional costs at the expiration time in the case that the buyer would want to deliver the contract to the seller.

Theorem 4.28 (Contingent Claim Pricing Theorem) *Let* $\mathcal{M}$ *be a financial market defined on a filtered probability space* $(\Omega, \mathcal{F}, \mathcal{P})$, *let* H_T *be an attainable contingent claim, and let* $\mathcal{Q} \in \mathbb{P} \neq \emptyset$ *be an equivalent martingale measure. Then, a unique price* P_{t_0} *associated with an attainable claim* H_T *is*

$$\mathsf{P}_{t_0} = \mathbb{E}_{\mathcal{Q}}[\beta_T H_T]$$

(Harrison and Pliska 1981, p. 240).

Proof. Since the contingent claim is attainable, it is true that

$$\widetilde{V}_T(\phi) = \beta_T H_T. \tag{4.9}$$

As each attainable claim is also admissible, the discounted value process is a $\mathcal{Q}$-martingale

$$\mathbb{E}_{\mathcal{Q}}[\widetilde{V}_T(\phi)|\mathcal{F}_s] = \widetilde{V}_s. \tag{4.10}$$

Thus, from equations (4.9) and (4.10) the Theorem 4.28 follows. ∎

4.3 Existence of the solution (Arbitrage)

Until now, we have assumed that the set of all equivalent martingale measures $\mathbb{P}$ is not empty. Let us now study the conditions that have to be fulfilled for the existence of the option price.

The following will define the arbitrage opportunity.

Definition 4.29 (Arbitrage opportunity) *Let $\mathcal{M}$ be a financial market defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{P})$, and let $(\phi_t)_{t_0 \leq t \leq T}$ be a self-financing strategy with zero initial investment*

$$V_{t_0}(\phi) = 0.$$

If the value process $V_t(\phi)$ determined by this trading strategy has a certain non-negative value at the maturity T

$$\mathcal{P}(V_T(\phi) \geq 0) = 1,$$

and there is some positive probability that the value of this portfolio will be positive at the maturity T

$$\mathcal{P}(V_T(\phi) > 0) > 0,$$

then the self financing strategy ϕ is called an arbitrage opportunity (Bingham and Kiesel 2004, p. 232).

From an economic point of view the arbitrage opportunity is practically a money making machine, as it enables the investor to make a profit without investing any start capital. Thus, an arbitrage-free market can be defined as follows.

Definition 4.30 (Arbitrage-free market) *The financial market $\mathcal{M}$ is arbitrage free if there are no arbitrage opportunities in the class of self-financing strategies (Bingham and Kiesel 2004, p. 106).*

Now we can define the condition of existence of the price for the contingent claim.

Theorem 4.31 (First Fundamental Theorem of Asset Pricing) *Assume that the set of equivalent martingale measures is non-empty (i.e. $\mathbb{P} \neq \emptyset$), then the market model $\mathcal{M}$ contains no arbitrage opportunities in the set of trading strategies (Bingham and Kiesel 2004, p. 234).*

Proof. Let the set of equivalent martingale measures $\mathbb{P}$ be not empty, and let $\mathcal{Q}$ be a martingale measure from this set, and let value process $V_t(\phi)$ be an arbitrage opportunity for the self-financing strategy. According to Definition 4.29 of the arbitrage $V_{t_0}(\phi) = 0$, $\Pr_{\mathcal{P}}[V_T(\phi) \geq 0] = 1$ and $\Pr_{\mathcal{P}}[V_T(\phi) > 0] > 0$. As $\mathcal{Q}$ is equivalent to $\mathcal{P}$, then $\Pr_{\mathcal{Q}}[V_T(\phi) \geq 0] = 1$, which is equivalent to

$$\Pr_{\mathcal{Q}}[V_T(\phi) < 0] = 0. \tag{4.11}$$

From the Definition 4.17

$$\mathbb{E}_{\mathcal{Q}}[\widetilde{V}_T(\phi)] = \widetilde{V}_{t_0}(\phi) = 0. \tag{4.12}$$

Equations (4.11) and (4.12) imply that $\Pr_{\mathcal{Q}}[V_T(\phi) > 0] = 0$, which from the equivalence of $\mathcal{Q}$ and $\mathcal{P}$, gives $\Pr_{\mathcal{P}}[V_T(\phi) > 0] = 0$. This contradicts the definition of arbitrage. Therefore, the arbitrage opportunity does not exist (Shreve 2004, p. 231). ■

Remark 4.32 *Thus, if we prove that no arbitrage opportunity does exist, we can state that the price of the contingent claim H_{t_0} with the payout function $f(S_T)$ does exist, such that*

$$\inf_{\mathcal{Q}\in\mathbb{P}} \mathbb{E}_{\mathcal{Q}}[\beta_{T-t_0} f(S_T)] \leq H_{t_0} \leq \sup_{\mathcal{Q}\in\mathbb{P}} \mathbb{E}_{\mathcal{Q}}[\beta_{T-t_0} f(S_T)].$$

We know that bounds exist, because $\mathbb{P}$ is finite.

4.4 Uniqueness of the solution (Completeness of the market)

In the previous section we have shown that the price of the contingent claim exists if the market is arbitrage-free. Now let us study the conditions when the price is unique. First, we introduce the definition of a complete market.

Definition 4.33 (Complete market) *A market $\mathcal{M}$ is complete if every contingent claim is attainable, i.e., for every $\mathcal{F}_T$-measurable random variable H_T there exists a replicating self-financing strategy $\phi \in \Phi$, such that $V_T(\phi) = H_T$ (Bingham and Kiesel 2004, p. 116). If the market is not complete then it is called incomplete.*

Now we can formulate the Second Fundamental Theorem of Asset Pricing.

Theorem 4.34 (Second Fundamental Theorem of Asset Pricing) *Assuming the absence of the arbitrage, the market model is complete if and only if the set of equivalent martingale measures $\mathbb{P}$ is a singleton (i.e., the equivalent martingale measure $\mathcal{Q}$ is unique) (Björk 2004, p. 151, 198).*

Proof. Let the model be complete in order to prove that a unique equivalent martingale measure exists. Furthermore, we assume that there exist two martingale measures: $\mathcal{Q}_1$ and $\mathcal{Q}_2$, which are equivalent to $\mathcal{P}$. Let event A be an element of the filtration $\mathcal{F}_T$. Now consider a contingent claim with the payoff function $H_T = \beta_T^{-1}\mathbb{I}_{[A]}$. As the market is complete, there exists a replicating self-financing strategy ϕ, such that $V_T(\phi) = H_T$. The discounted portfolio value $\widetilde{V}_T(\phi)$ is a martingale with respect to $\mathcal{Q}_1$ and $\mathcal{Q}_2$, because both these measures are risk-free. Thus,

$$\Pr_{\mathcal{Q}_1}[A] = \mathbb{E}_{\mathcal{Q}_1}[\widetilde{V}_T(\phi)] = \mathbb{E}_{\mathcal{Q}_1}[\widetilde{H}_T] = H_{t_0} = \mathbb{E}_{\mathcal{Q}_2}[\widetilde{H}_T] = \mathbb{E}_{\mathcal{Q}_2}[\widetilde{V}_T(\phi)] = \Pr_{\mathcal{Q}_2}[A].$$

Thus, the risk-free measure is unique ($\mathcal{Q}_1 = \mathcal{Q}_2$). The proof in the opposite direction takes much longer and can be found in Shreve (2004, p. 232-234).

■

4.5 Esscher risk-neutral probability measure

In incomplete markets there exist several risk-neutral probability measures. In this dissertation the Esscher martingale measure will be of particular interest.

The Esscher transformation is a well-approved tool among actuaries, and was originally developed by Esscher (1932) to transform a random variable; to give it a new distribution captured at a point of interest. The purpose of this is to enable more accurate approximations to be made at this point. Gerber and Shiu were the first to use the Esscher transform to price European (Gerber and Shiu 1994b) and American options (Gerber and Shiu 1994a). This Section introduces how to use the Esscher martingale measure to price options.

First, we make the following assumption.

Assumption 4.35 *Let $\mathcal{M}$ be a financial market defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{P})$, where the risky stock is a continuously compounded return rate process $(X_t)_{t \geq t_0}$ with stationary independent increments and the initial value $X_{t_0} = 0$, such that*

$$S_t = S_{t_0} e^{X_t}, \qquad \text{for all } t \geq t_0 \tag{4.13}$$

(Gerber and Shiu 1994b, p. 102).

Notation 4.36 *Let*

$$F(x,t) = \Pr_{\mathcal{P}}(X_t \leq x) \qquad \text{for all } t \geq t_0 \text{ and } x \in \mathbb{R}$$

denote the cumulative distribution function of the process X_t, with the associated density function

$$f(x,t) = \frac{d}{dx}F(x,t), \qquad \textit{for all } t \geq t_0 \textit{ and } x \in \mathbb{R}$$

and the moment generating function

$$\mathbb{M}[z,t] = \mathbb{E}\left[e^{zX_t}\right] = \int_{-\infty}^{\infty} e^{zx} f(x,t)dx, \qquad \textit{for all } t \geq t_0 \textit{ and } z \geq 0,$$

respectively (Gerber and Shiu 1994b, p. 102).

Proposition 4.37 *Assume that $\mathbb{M}[z,t]$ is continuous, then*

$$\mathbb{M}[z,t] = \mathbb{M}^t[z,1]$$

(Gerber and Shiu 1994b, p. 102).

Proof. For proof, see Breiman (1968, Section 14.4) or Feller (1971, Section IX.5). ■

In the following, we will introduce the Esscher density function.

Definition 4.38 (Esscher equivalent martingale measure) *Let $\mathcal{P}$ be a probability measure, let $\mathcal{F}_t$ be a filtration, and let $h \in \mathbb{R}$ for which the moment generating function $\mathbb{M}[h,t]$ exists, then the Radon-Nikodým derivative*

$$\left.\frac{d\mathcal{Q}}{d\mathcal{P}}\right|_{\mathcal{F}_t} = \frac{e^{hx} f(x,t)}{\int_{-\infty}^{\infty} e^{hy} f(y,t)dy} = \frac{e^{hx} f(x,t)}{\mathbb{M}[h,t]}$$

defines the Esscher equivalent martingale measure $\mathcal{Q}$ with respect to parameter h (Gerber and Shiu 1994b, p. 102-103).

Notation 4.39 *Hereafter, the following notation will be used*

$$\left.\frac{d\mathcal{Q}}{d\mathcal{P}}\right|_{\mathcal{F}_t} = f(x,t;h)$$

Remark 4.40 *The Esscher transformed moment generating function is*

$$\mathbb{M}[z,t;h] = \int_{-\infty}^{\infty} e^{zx} f(x,t;h)dx = \frac{\mathbb{M}[z+h,t]}{\mathbb{M}[h,t]} \tag{4.14}$$

(Gerber and Shiu 1994b, p. 103).

Proposition 4.41 *Assume that* $\mathbb{M}[z,t;h]$ *is continuous, then*

$$\mathbb{M}[z,t;h] = \mathbb{M}^t[z,1;h]$$

(Gerber and Shiu 1994b, p. 103).

Proof. The proof results from Proposition 4.37. ■

Proposition 4.42 (Existence and uniqueness of the Esscher parameter) *Let r be a constant risk-free rate, then it holds that*

$$r = \ln \mathbb{M}[1,1;h_{\mathcal{Q}}] \tag{4.15}$$

and has the unique solution $h = h_{\mathcal{Q}}$ (Gerber and Shiu 1994a, p. 664 and Gerber and Shiu 1994b, p. 104).

Proof. First of all, we prove the existence of the solution. As the Esscher probability measure is risk-neutral, the discounted stock has to be a $\mathcal{Q}$-martingale

$$S_{t_0} = \mathbb{E}_{\mathcal{Q}}[e^{-r(t-t_0)}S_t].$$

From equation (4.13) it follows that

$$S_{t_0} = e^{-r(t-t_0)} S_{t_0} \mathbb{E}_{\mathcal{Q}}\left[e^{X_{(t-t_0)}}\right].$$

If we omit S_{t_0} and use Remark 4.40 and Proposition 4.41, we get

$$e^{r(t-t_0)} = \mathbb{M}[1,(t-t_0);h_{\mathcal{Q}}] = \mathbb{M}[1,1;h_{\mathcal{Q}}]^{(t-t_0)},$$

which is equivalent to

$$r = \ln \mathbb{M}[1, 1; h_{\mathcal{Q}}],$$

which, in turn, proves the existence of the solution (Gerber and Shiu 1994b, p. 103-104).

Now let us prove the uniqueness of the solution. Consider a function $g(h)$

$$g(h) = \ln \mathbb{M}[1, 1; h]$$

for all h, such that $\mathbb{M}[1, 1; h]$ exists. From Remark 4.40 we have

$$g(h) = \ln \mathbb{M}[h + 1, 1] - \ln \mathbb{M}[h, 1] = \ln \mathbb{E}\left[e^{(h+1)X_1}\right] - \ln \mathbb{E}\left[e^{hX_1}\right].$$

Note that

$$g'(h) = \frac{\mathbb{E}\left[X_1 e^{(h+1)X_1}\right]}{\mathbb{E}\left[e^{(h+1)X_1}\right]} - \frac{\mathbb{E}\left[X_1 e^{hX_1}\right]}{\mathbb{E}\left[e^{hX_1}\right]} = \mathbb{E}[X_1; h + 1] - \mathbb{E}[X_1; h]$$

where $\mathbb{E}[g(X_t); h] = \frac{\mathbb{E}\left[g(X_t)e^{hX_t}\right]}{\mathbb{E}\left[e^{hX_t}\right]}$. Furthermore,

$$\frac{d\mathbb{E}[X_1; h]}{dh} = \frac{\mathbb{E}\left[X_1^2 e^{hX_1}\right]}{\mathbb{E}\left[e^{hX_1}\right]} - \left(\frac{\mathbb{E}\left[X_1 e^{hX_1}\right]}{\mathbb{E}\left[e^{hX_1}\right]}\right)^2 = \mathbb{V}\mathrm{ar}(X_1; h) > 0,$$

as X_t is a non-degenerate random variable. Thus, the first derivative of $\mathbb{E}[X_1; h]$ is a strictly positive and the expected value $\mathbb{E}[X_1; h]$ is a strictly increasing function. From this it results that the function $g(h)$ is strictly increasing, and thus the equation $g(h) = r$ has a unique solution: $h = h_{\mathcal{Q}}$ (Gerber and Shiu 1994a, p. 664). ■

Remark 4.43 *The Esscher equivalent measure is unique. However, this does not mean that other risk-neutral measures do not exist.*

Remark 4.44 *Note that for $t \geq t_0$*

$$\frac{e^{hX_t}}{(\mathbb{M}[h, 1])^t} = \frac{e^{hX_t}}{\mathbb{E}[e^{hX_t}]} = \frac{(S_t)^h}{\mathbb{E}[(S_t)^h]}.$$

.

Thus, we can construct the following factorization rule, which is very convenient, as it saves us some complicated calculations.

Proposition 4.45 (Factorization formula) *Let g be a measurable function and let h, k, t be real numbers, with $t \geq 0$, then it holds that*

$$\mathbb{E}[S_t^k g(S_t); h] = \mathbb{E}[S_t^k; h]\mathbb{E}[g(S_t); k+h]$$

(Gerber and Shiu 1996, p. 188).

Proof.

$$\begin{aligned}\mathbb{E}[S_t^k g(S_t); h] &= \mathbb{E}\left[S_t^k g(S_t)\frac{e^{hX_t}}{(\mathbb{M}[h,1])^t}\right] = \frac{\mathbb{E}[S_t^{k+h}g(S_t)]}{\mathbb{E}[S_t^h]} \\ &= \frac{\mathbb{E}[S_t^{k+h}]}{\mathbb{E}[S_t^h]}\frac{\mathbb{E}[S_t^{k+h}g(S_t)]}{\mathbb{E}[S_t^{k+h}]} = \mathbb{E}[S_t^k; h]\mathbb{E}[g(S_t); k+h]\end{aligned}$$

(Gerber and Shiu 1996, p. 188). ■

Now we can use this factorization rule to find the put price via the Esscher risk-neutral measure.

Theorem 4.46 (Put price via Esscher risk-neutral measure) *Let $\mathcal{M}$ be a financial market defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{P})$ and let $\mathcal{Q}$ be an Esscher equivalent martingale measure, let $F(x, t; h)$ be an Esscher transformed cumulative distribution function with respect to parameter h, and let P be a European put option with the expiration date T and strike price K. Then, the price of this contingent claim is given by*

$$\mathsf{P}_{t_0} = e^{-r(T-t_0)}\mathsf{K}F\left(-\ln\frac{S_{t_0}}{\mathsf{K}}, T-t_0; h_{\mathcal{Q}}\right) - S_{t_0}F\left(-\ln\frac{S_{t_0}}{\mathsf{K}}, T-t_0; h_{\mathcal{Q}}+1\right)$$

Proof. According to the definition of the put option (4.2) and the Contingent Claim Pricing Theorem 4.28, the price is equal to

$$\mathsf{P}_{t_0} = \mathbb{E}_{\mathcal{Q}}[e^{-r(T-t_0)}f(S_T)|\mathcal{F}_{t_0}]$$

with $f(S_T) = (\mathsf{K} - S_T)^+$. If we write,

$$\begin{aligned}
\mathsf{P}_{t_0} &= \mathbb{E}_{\mathcal{Q}}[e^{-r(T-t_0)}(\mathsf{K} - S_T)^+ | \mathcal{F}_{t_0}] \\
&= e^{-r(T-t_0)}\mathsf{K}\mathbb{E}_{\mathcal{Q}}[\mathbb{I}_{[S_T<\mathsf{K}]}] - e^{-r(T-t_0)}\mathbb{E}_{\mathcal{Q}}[S_T\mathbb{I}_{[S_T<\mathsf{K}]}]
\end{aligned}$$

we can use the factorization formula from the Proposition 4.45 with

$$g(S_T) = \mathbb{I}_{[S_T<\mathsf{K}]} = \mathbb{I}_{[S_{t_0}e^x<\mathsf{K}]} = \mathbb{I}_{[x<\ln\mathsf{K}-\ln S_{t_0}]},$$

where $x = X_{T-t_0}$. Then,

$$\begin{aligned}
\mathsf{P}_{t_0} &= e^{-r(T-t_0)}\mathsf{K}\mathbb{E}_{\mathcal{Q}}[\mathbb{I}_{[x<\ln\mathsf{K}-\ln S_{t_0}]}; h_{\mathcal{Q}}] \\
&\quad - e^{-r(T-t_0)}\mathbb{E}_{\mathcal{Q}}[S_T; h_{\mathcal{Q}}]\mathbb{E}_{\mathcal{Q}}[\mathbb{I}_{[x<\ln\mathsf{K}-\ln S_{t_0}]}; h_{\mathcal{Q}}+1].
\end{aligned}$$

As we have $\mathbb{E}_{\mathcal{Q}}[\mathbb{I}_A] = \Pr_{\mathcal{Q}}(A)$ and $\mathbb{E}_{\mathcal{Q}}[S_T; h_{\mathcal{Q}}] = S_{t_0}^{X_{(T-t_0)}}$ we get

$$\begin{aligned}
\mathsf{P}_{t_0} &= e^{-r(T-t_0)}\mathsf{K}\Pr_{\mathcal{Q}}\left(x < -\ln\frac{S_{t_0}}{\mathsf{K}}; h_{\mathcal{Q}}\right) - S_{t_0}\Pr_{\mathcal{Q}}\left(x < -\ln\frac{S_{t_0}}{\mathsf{K}}; h_{\mathcal{Q}}+1\right) \\
&= e^{-r(T-t_0)}\mathsf{K}F\left(-\ln\frac{S_{t_0}}{\mathsf{K}}, T-t_0; h_{\mathcal{Q}}\right) - S_{t_0}F\left(-\ln\frac{S_{t_0}}{\mathsf{K}}, T-t_0; h_{\mathcal{Q}}+1\right),
\end{aligned}$$

which completes the proof. ■

4.6 Option pricing

4.6.1 Black-Scholes market

Now we can use the Esscher martingale measure to prove the well-known Black and Scholes (1973) option pricing formula. This formula can be used to price options in a complete market.

Definition 4.47 (Black-Scholes financial market) *Let $\mathcal{M}_{BS} = (B_t, S_t)$ be a financial market defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{P})$ where the risky stock $(S_t)_{t\geq t_0}$ follows a geometric Brownian motion*

$$S_t = S_{t_0} + \int_{t_0}^{t}\mu S_u du + \int_{t_0}^{t}\sigma S_u dW_u.$$

where $\mu \in \mathbb{R}$ denotes the drift, $\sigma > 0$ the diffusion parameter, and W_t standard Wiener process, respectively. Then the tuple $\mathcal{M}_{BS}$ is called Black-Scholes financial market.

Theorem 4.48 (Black-Scholes option pricing formula) *Let $\mathcal{M}_{BS}$ be a Black-Scholes financial market defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{P})$, then the price of a European put P with the expiration time T and the strike price K is given by*

$$\mathsf{P}_{t_0} = \mathsf{K}e^{-r(T-t_0)}\Phi(-d_2) - S_{t_0}\Phi(-d_1)$$

with

$$d_1 = \frac{\ln \frac{S_{t_0}}{\mathsf{K}} + \left(r + \frac{1}{2}\sigma^2\right)(T - t_0)}{\sigma\sqrt{T - t_0}}$$

and

$$d_2 = d_1 - \sigma\sqrt{T - t_0},$$

where $\Phi(\cdot)$ denotes the standard normal distribution function.

Proof. As the stock process is a geometric Brownian motion, the stock price process follows a log-normal distribution with the moment-generating function

$$\mathbb{M}[z, t] = e^{(\mu z + \frac{1}{2}\sigma^2 z^2)t}.$$

From equation (4.14) it follows that

$$\ln \mathbb{M}[z, t; h] = \left((\mu + h\sigma^2)z + \frac{1}{2}\sigma^2 z^2\right)t. \tag{4.16}$$

Thus the return of the stock has a mean μt under the real probability measure $\mathcal{P}$ and $(\mu + h\sigma^2)t$ under the risk-neutral Esscher measure $\mathcal{Q}$, respectively. Note that the variance of the return is unchanged under both measures: $\sigma^2 t$. Therefore, returns X_t are normally distributed

$$X_t \sim \mathcal{N}((\mu + h\sigma^2)t, \sigma^2 t). \tag{4.17}$$

Equations (4.15) and (4.16) yield the risk-free rate of return

$$r = (\mu + h_{\mathcal{Q}}\sigma^2) + \frac{1}{2}\sigma^2$$

and equivalently the unique Esscher parameter is

$$h_{\mathcal{Q}} = \frac{r-\mu}{\sigma^2} - \frac{1}{2}.$$

Thus, the mean rate of return under the Esscher measure $\mathcal{Q}$ is

$$\mu_{\mathcal{Q}} = r - \frac{1}{2}\sigma^2. \tag{4.18}$$

Equations (4.17) and (4.18) imply that the cumulated distribution function is given by

$$F(x,t;h) = \Phi\left(\frac{x - \left(r - \frac{1}{2}\sigma^2\right)(t-t_0)}{\sigma\sqrt{t-t_0}}\right)$$

From Theorem 4.46 the put price equals

$$\begin{aligned}
\mathsf{P}_{t_0} =& e^{-r(T-t_0)}\mathsf{K}F\left(-\ln\frac{S_{t_0}}{\mathsf{K}}, T-t_0; h\right) - S_{t_0}F\left(-\ln\frac{S_{t_0}}{\mathsf{K}}, T-t_0; h+1\right)\\
=& e^{-r(T-t_0)}\mathsf{K}\Phi\left(\frac{-\ln\frac{S_{t_0}}{\mathsf{K}} - (r-\frac{1}{2}\sigma^2)(T-t_0)}{\sigma\sqrt{T-t_0}}\right)\\
&- S_{t_0}\Phi\left(\frac{-\ln\frac{S_{t_0}}{\mathsf{K}} - (r+\frac{1}{2}\sigma^2)(T-t_0)}{\sigma\sqrt{T-t_0}}\right)\\
=& e^{-r(T-t_0)}\mathsf{K}\Phi\left(-\frac{\ln\frac{S_{t_0}}{\mathsf{K}} + (r-\frac{1}{2}\sigma^2)(T-t_0)}{\sigma\sqrt{T-t_0}}\right)\\
&- S_{t_0}\Phi\left(-\frac{\ln\frac{S_{t_0}}{\mathsf{K}} + (r+\frac{1}{2}\sigma^2)(T-t_0)}{\sigma\sqrt{T-t_0}}\right),
\end{aligned}$$

which completes the proof[1]. ■

[1] This proof is analogous to the proof for the price of the call option given by Gerber and Shiu (1994b, p. 107-108)

4.6.2 Markov Switching market

This Section will discuss option pricing in a special incomplete market: the Markov switching market.

Definition 4.49 (Markov switching financial market) *Let $\mathcal{M}_{MS} = (B_t, S_t)$ be a financial market defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{P})$, let $(Z_{t_n})_{t_n \geq 0}$ (where $t_n = n\tau$, $n \in \mathbb{N}$ and τ is a fixed positive number) be a Markov chain with the transition probabilities*

$$p_{ji} = \Pr[Z_{t_n} = j | Z_{t_{n-1}} = i], \quad \textit{with} \sum_{j \in \mathcal{K}} p_{ji} = 1, \ \textit{and } i \in \mathcal{K},$$

and the state space $\mathcal{K} = \{1, 2, \dots, K\}$. Let $(W_t)_{t \geq 0}$ be a Wiener process, with $\mathcal{F}_t = \sigma\{Z_{t_n}, S_t, W_t : \ t \geq 0\}$ being an associated filtration, and $\mu(Z_{t_n})$ and $\sigma(Z_{t_n})$ being associated processes. Furthermore, let the risky stock S_t follow a geometric Brownian motion with Markov switching

$$S_t = S_{t_0} + \int_{t_0}^{t} \mu(Z_u) S_u du + \int_{t_0}^{t} \sigma(Z_u) S_u dW_u, \qquad \textit{for } t \in [t_n, t_{n+1}),$$

then, the tuple $\mathcal{M}_{MS}$ is called a Markov switching financial market.

Under the classic geometric Brownian motion, the financial market is complete, so that there exists a unique martingale measure $\mathcal{Q}$. However, Chapter 3 shows that the Markov switching model is a better model to describe the stochasticity of the risky portfolio underlying the guarantee. In this model, the variance is stochastic and, thus, the market is not complete anymore. The option pricing in such an economy is not straightforward because there exists no unique equivalent probability measure and, therefore, no unique price of the option. Thus, a reliable choice of the martingale measure has to be made. This section discusses two possible choices: the Bollen-Hardy and the Esscher measure.

4.6.2.1 Bollen-Hardy option pricing formula

Bollen (1998) was the first who priced options in the Markov switching model. His approach bases on the discrete time model. In this thesis, we discuss the extension to the continuous time made by Hardy (2001). This method is based on the model with stochastic volatility. However, the main drawback to this model is that it does not price the switching risk.

Theorem 4.50 (Bollen-Hardy option pricing formula) *Let $\mathcal{M}_{MS}$ be a Markov switching financial market defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{P})$ and the state space $\mathcal{K} = \{1, 2\}$, and let the transition probability be unchanged under the equivalent martingale measure $\mathcal{Q}$, i.e.*

$$\Pr_{\mathcal{P}}[Z_{t_n} = j | Z_{t_{n-1}} = i] = \Pr_{\mathcal{Q}}[Z_{t_n} = j | Z_{t_{n-1}} = i] \qquad \text{for all } i, j \in \mathcal{K}, \tag{4.19}$$

then the Bollen-Hardy price of a European put P with the expiration time T and the strike price K is given by

$$\mathsf{P}_{t_0} = \sum_{i=0}^{N} \Pr(R = i)[\mathsf{K}e^{-r(T-t_0)}\Phi(-d_2(R = i)) - S_0\Phi(-d_1(R = i))]$$

with

$$d_1(R) = \frac{\ln \frac{S_{t_0}}{\mathsf{K}} - \left(r + R\frac{1}{2}\sigma_1^2 + (N - R)\frac{1}{2}\sigma_2^2\right)(T - t_0)}{\sqrt{R\sigma_1^2 + (N - R)\sigma_2^2}\sqrt{T - t_0}}$$

and

$$d_2(R) = d_1(R) - \sqrt{R\sigma_1^2 + (N - R)\sigma_2^2}\sqrt{T - t_0},$$

where $N \in \mathbb{N}$ denotes the number of "switching" periods (i.e. $T - t_0 = N(t_n - t_{n-1})$), R the number of periods when the stock price process is in regime 1 (i.e. $R = \sum_{t_n}^{N} \mathbb{I}_{[Z_{t_n}=1]}$), and $\Phi(\cdot)$ the standard normal distribution function (Hardy 2001, p. 49).

Proof. According to the definition of the put option (4.2) and the Contingent Claim Pricing Theorem 4.28, the price is equal:

$$\begin{aligned}\mathsf{P}_{t_0} &= e^{-(T-t_0)r}\mathbb{E}_{\mathcal{Q}}[(\mathsf{K}-S_T)^+] = e^{-(T-t_0)r}\mathbb{E}_{\mathcal{Q}}[\mathbb{E}_{\mathcal{Q}}[(\mathsf{K}-S_T(R))^+]|R] \\ &= \sum_{i=0}^{N}\Pr_{\mathcal{Q}}(R=i)e^{-(T-t_0)r}\mathbb{E}_{\mathcal{Q}}[(\mathsf{K}-S_T)^+|R=i] = \sum_{i=0}^{N}\Pr_{\mathcal{Q}}(R=i)\mathsf{P}_{t_0}(R=i).\end{aligned}$$

Using the assumption (4.19),

$$\mathsf{P}_{t_0} = \sum_{i=0}^{N}\Pr_{\mathcal{Q}}(R=i)\mathsf{P}_{t_0}(R=i) = \sum_{i=0}^{N}\Pr_{\mathcal{P}}(R=i)\mathsf{P}_{t_0}(R=i).$$

Under the condition that R, the number of periods the state variable has been in the first regime is known, the risky stock follows the geometric Brownian motion with drift $R\mu_1 + (N-R)\mu_2$ and diffusion $\sqrt{R\sigma_1^2 + (N-R)\sigma_2^2}$. Thus, we can use the Black-Scholes formula for the price of the put option.

$$\mathsf{P}_{t_0}(R=i) = \mathsf{K}e^{-r(T-t_0)}\Phi(-d_2(R=i)) - S_0\Phi(-d_1(R=i))$$

with

$$d_1(R=i) = \frac{\ln\frac{S_{t_0}}{\mathsf{K}} - \left(r + R\frac{1}{2}\sigma_1^2 + (N-R)\frac{1}{2}\sigma_2^2\right)(T-t_0)}{\sqrt{R\sigma_1^2 + (N-R)\sigma_2^2}\sqrt{T-t_0}}$$

and

$$d_2(R=i) = d_1(R) - \sqrt{R\sigma_1^2 + (N-R)\sigma_2^2}\sqrt{T-t_0},$$

which completes the proof. ■

4.6.2.2 Webb option pricing formula

The Bollen-Hardy option pricing model assumes that the risk-neutral measure does not change (see assumption (4.19)). Neither Bollen nor Hardy proved that this holds true, which is a drawback to this approach. Webb (2003) proposed several models which allow to price the switching risk. In her thesis, she proposed three martingale measures. They are based on the mean-variance hedging, the Esscher transform, and the minimum entropy. They all have closed analytical solutions and give similar numerical results. However, only the Esscher transform solution can be simulated via the Monte Carlo method. We decided, though, to use this method, as the guarantees we price are path-dependent and can only be solved numerically. Furthermore, the Esscher martingale measure $\mathcal{Q}$ has several additional advantages:

- The process under the new martingale measure remains in the same class of models as the process under the real-word probability measure $\mathcal{P}$. In the considered case, it means that the prices under the Esscher transform follow the geometric Brownian motion with Markov switching, see Gerber and Shiu (1994b, p. 163-165, comment of Michaud).
- The solution reduces to the well-known Black and Scholes (1973) formula for the case of one switching regime ($K = 1$), see Corollary 5.4.4 in Webb (2003).
- The Esscher measure allows the pricing of the switching risk.
- Finally, the Esscher transform approach is conform with maximizing the expected utility with the constant risk aversion utility function $u(x) = \frac{x^\gamma}{\gamma}$ $(0 < \gamma < 1)$, which is commonly used in financial models, see Webb (2003, p. 88). This means, intuitively, that the agent prefers to have more money than less. However, the wealth increase of €1, has a smaller additional utility, the more the agent possesses. The agent with this utility function is risk averse.

Theorem 4.51 (Webb option pricing formula) *Let $\mathcal{M}_{MS}$ be a Markov switching financial market defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{P})$ and the state space $\mathcal{K} = \{1, 2, \ldots, K\}$, and let the transition probability be* *changed under the equivalent martingale measure $\mathcal{Q}$, then the Webb price of a European put* P *with the expiration time* T *and the strike price* K *is given by*

$$\mathsf{P}_{t_0,j_1} = \mathsf{P}_{t_0}(Z_{t_0} = j_1) = \sum_{j_2=1}^{K} \cdots \sum_{j_{N+1}=1}^{K} \underbrace{\prod_{i=1}^{N} p^{(h)}_{j_{i+1}j_i}}_{(\star)}$$
$$\times \underbrace{\left[\mathsf{K}e^{-r(T-t_0)}\Phi(-d_j^-) - S_{t_0} \exp\left\{\sum_{k=1}^{N}\left[(\mu_{j_{k+1}} - r)\tau + h_{j_k}\sigma^2_{j_{k+1}}\tau\right]\right\}\Phi(-d_j^+)\right]}_{(\star\star)} \tag{4.20}$$

with the Esscher transition probabilities

$$p^{(h)}_{ji} = \Pr[Z_t = j | Z_{t-1} = i; h] = \frac{p_{ji}\exp\left\{h_i[\mu_j - \frac{1}{2}\sigma_j^2]\tau + \frac{1}{2}\sigma_j^2\tau h_i^2\right\}}{\sum_{j=1}^{M} p_{ji}\exp\left\{h_i[\mu_j - \frac{1}{2}\sigma_j^2]\tau + \frac{1}{2}\sigma_j^2\tau h_i^2\right\}} \tag{4.21}$$

and parameters

$$d_j^{\pm} = \frac{\ln\left(\frac{S_{t_0}}{\mathsf{K}}\right) + \sum_{i=2}^{N+1}\left(\mu_{j_i} \pm \frac{1}{2}\sigma_{ji}^2\right)\tau + \sum_{i=1}^{N-1}\left(h_{j_i} - h_{j_N}\right)\sigma^2_{j_{i+1}}\tau}{\sqrt{\left(\sum_{i=2}^{N+1}\sigma^2_{j_i}\right)\tau}}$$
$$+ h_{j_N}\sqrt{\left(\sum_{i=2}^{N+1}\sigma^2_{j_i}\right)\tau}, \tag{4.22}$$

where N *denotes the amount of switches in the pricing horizon,* $\tau = (T - t_0)/N$ *the time period between two switches, and* $\Phi(\cdot)$ *the standard normal distribution function (Webb 2003, p. 102-103).*

Proof. For the proof for the call price, see Webb (2003, Chapter 5). Then use the put-call parity, which completes the proof. ■

To assure that the option price under the Esscher measure (4.20)-(4.22) is unique we have to prove that the Esscher parameter vector is unique.

Proposition 4.52 (Uniqueness of the Esscher parameter) *The Esscher parameter vector* $\boldsymbol{h}$ *is unique and can be computed numerically from the equations*

$$\sum_{j=1}^{K} p_{ji} \exp\left\{h_i(\mu_j - \frac{1}{2}\sigma_j^2)\tau + \frac{1}{2}\sigma_j^2\tau h_i^2\right\}\left(\exp\left\{\mu_j\tau + \sigma_j^2\tau h_i\right\} - e^{r\tau}\right) = 0, \tag{4.23}$$

knowing that h_i *is a unique point from the interval*

$$\left(\min_j\left(\frac{r-\mu_j}{\sigma_j^2}\right), \max_j\left(\frac{r-\mu_j}{\sigma_j^2}\right)\right) \tag{4.24}$$

(Webb 2003, p. 89-91).

Proof. The proof is analogous to the proof for Proposition 4.42 and can be found in Webb (2003, p. 89-91) ■

The term $(\star\star)$ in the equation (4.20) determines the price of the put for the given switching path (combination of regimes) $j_1, j_2, \ldots, j_{N+1}$, which is weighted with the probability $(\star)$ that the process will follow that path. Finally, the price is computed for each possible switching path and added together. Note that the term $(\star\star)$ discloses some parallels to the well-known Black and Scholes (1973) formula for the European put option: (a) The terms d_j^+ and d_j^- correspond to d_1 and d_2, respectively. (b) The exponential term after S_0 is consistent with the Black and Scholes formula as well because the stock return by Black and Scholes is equal to the risk-free rate r. For the discounted asset price S_T this yields

$$e^{-rT}S_T = e^{-rT}S_{t_0}e^{rT} = S_{t_0} \tag{4.25}$$

Under the Esscher risk-neutral probability measure, the equity return is equal to $(\mu_{j_n} + h_{j_{n-1}}\sigma_{j_n}^2)\tau$, so this term does not reduce to S_{t_0}.

Moreover, please note that the put price P_{t_0,j_1} depends on the initial state j_1. To determine the price which is independent from the initial regime, the interval τ has to be decreased, due to the fact that the influence of the initial state on the put price decreases as the length of the path grows (see Tables 7.1-7.3 in Webb (2003)). However, this is problematic, since the number of combinations grows exponentially as the length of the path increases. This is particularly problematic with regard to the long-term options, which are studied in this dissertation. For instance, for two states ($K = 2$), 30 years maturity, and one switch per year ($N = 30$), the number of combinations (K^N) equals 1,073,741,824. If the switch occurs every month ($N = 360$), the number of combinations rises to $2.3 \cdot 10^{108}$. Instead of increasing N, the approximate price $\mathsf{P}_{t_0,app}$ could be determined through weighting the initial state-dependent prices in equation (4.20) by the unconditional Esscher probabilities, so that the process stays in the i-th regime

$$\mathsf{P}_{t_0,app} = \sum_{i=1}^{K} \mathsf{P}_{t_0,i} \pi_i^{(h)} \tag{4.26}$$

where the unconditional probabilities $\pi_i^{(h)}$ could be computed with the equation (2.68) using the Esscher conditional probabilities $p_{ji}^{(h)}$ from equation (4.21) instead of the real-world probabilities p_{ji}. For other alternatives, see the discussion in Section 4.7.2.1.

Due to the fact that contributions to retirement saving plans are paid periodically, there exists no closed-form solution of the option pricing, thus the formula (4.20) cannot be applied directly. However, it is possible to simulate the option price with the Monte Carlo simulation with the return mean

$$\left(\mu_j - \frac{1}{2}\sigma_j^2\right)\tau + \frac{1}{2}h_i\sigma_j^2\tau \tag{4.27}$$

and the return standard deviation

$$\sigma_j \tau \tag{4.28}$$

and the Esscher transition probabilities given in the equation (4.21). For the simulation algorithm under the Markov switching regime, see Hardy (2003, p. 98).

4.7 Quantitative results

4.7.1 Outline of the study

4.7.1.1 Design of the guarantee

The German finance industry offers retirement saving plans, in which the contributions are invested in risky portfolios (see Maurer and Schlag 2003, and Gründl et al. 2004). These saving plans are co-financed by the state if the provider includes a guarantee that at least the sum of the charged premiums is paid out to the investor when the contract expires. This is equivalent to a deterministic guarantee rate of 0% on the paid contributions. If the provider's investment strategy fails to generate this minimum return, he is required to finance the difference between the market value of the portfolio and the guaranteed amount. From an economic point of view, this guarantee is a European put option with the following payoff:

$$\mathsf{P}_T = \max\{\mathsf{G}_T - S_T; 0\}, \tag{4.29}$$

with the guarantee value

$$\mathsf{G}_T = \sum_{t_n=0}^{T-1} C_{t_n} e^{g(T-t_n)}, \tag{4.30}$$

where P_T is the value of the option, and S_T is the market value of the risky portfolio at the maturity T, g is the guaranteed rate of return, and C_{t_n} is the contribution paid in at the time t_n. The price of the put at the time $t_n = 0$

equals the expected value of the payoff at the maturity under the risk-free probability measure $\mathcal{Q}$, discounted with the risk-free rate of return r

$$\mathsf{P}_0 = e^{-rT}\mathbb{E}_{\mathcal{Q}}[\max\{\mathsf{G}_T - S_T; 0\}], \tag{4.31}$$

where $\mathbb{E}_{\mathcal{Q}}[\cdot]$ denotes the expected value under the probability measure $\mathcal{Q}$.

4.7.1.2 Investment strategies

In this study we examine several investment strategies. They can be assigned to three categories: buy-and-hold strategies, life-cycle strategies, and the zero-bond strategy. In *buy-and-hold investment strategies* the client invests $x\%$ (with $0\% < x < 100\%$) of his contributions in a well-diversified bond fund and the remaining $(100 - x)\%$ in a well diversified equity fund. While this proportion remains constant throughout the contract duration, the asset allocation in both funds, however, can change over time. In this dissertation, we analyze the following portfolio choices: $x = 0\%, 10\%, \ldots, 100\%$, and, additionally, $x = 25\%$ and $x = 75\%$.

One of the main properties of the buy-and-hold strategies is the constant stock proportion during the whole investment period. This investment decision can, however, be suboptimal. A high proportion of stock implies a higher risk level. This leads: on the one hand to a higher expected profit and, on the other hand, to a higher guarantee cost, as the option price increases along with the increasing risk. On the contrary, the choice of a low stock proportion implies a lower risk level and causes a decrease of the guarantee cost. Simultaneously, it leads to a decrease of the expected return from the portfolio. Given that the goal of retirement savings is to provide a relatively high income during retirement, this product would not be what the client aims to buy. Instead, the investor could choose an investment strategy with

Table 4.1: Life-cycle investment strategies

Years to maturity	**Client's age**	**Aggressive**	**Moderate**	**Conserv.**	**Naive**	**100-x***
		Portfolio proportion invested in bonds				
30 – 26	35 – 39	0%	0%	40%	25%	40%
25 – 21	40 – 44	0%	0%	50%	25%	40%
20 – 16	45 – 49	0%	30%	60%	50%	50%
15 – 11	50 – 54	0%	60%	70%	50%	50%
10 – 6	55 – 59	0%	90%	80%	75%	60%
5	60	50%	90%	90%	75%	60%
4	61	60%	90%	90%	75%	60%
3	62	70%	90%	90%	75%	60%
2	63	80%	90%	90%	75%	60%
1	64	90%	90%	90%	75%	60%

* Example of a 35-year-old investor who buys a 30-year contract.

a decreasing stock proportion as the contract draws to its expiration. Such strategies are called life-cycle strategies.

In this thesis, we will analyze five *life-cycle investment strategies*: the aggressive, the moderate, the conservative, the naive, and the so-called 100-x investment rule (see Table 4.1). In the *aggressive investment strategy*, the client invests the whole portfolio in stocks. In the fifth year before the end of the contract, half of the portfolio's assets will be shifted to bonds. Each following year, the proportion of the portfolio invested in bonds will rise by 10 per cent points, so that in the last year before maturity, 90% of the portfolio will be invested in bonds and the remaining 10% in stocks.

The *moderate investment strategy* is the one proposed by Maurer and Schlag (2003). If the time to maturity is greater than 20, all assets are invested in stocks. In the 20th year before maturity, 30% of the portfolio is shifted to bonds. The last shift occurs 10 years before maturity, when a further 30% of the assets is shifted to bonds. Thus, the bond proportion finally equals 90%.

In the *conservative investment strategy*, the bond proportion is much higher. Every five years the bond proportion rises by 10 per cent points, so that in the last five years of the contract, the stock engagement will have shrunk to 10%. For instance, for a 30 year contract, one would start a portfolio with 60% stocks and 40% bonds. After the first five years, the assets are shifted to a fifty-fifty stock-bond proportion, and so on, so that in the last five years, the proportion of bonds will have reached 90%.

In the *naive investment strategy*, the bond proportion in the portfolio rises every 10 years by 25 per cent points, with the final goal being 75%. For instance, for a 30 year contract one would start with a 25% bond proportion. After 10 years, the stock-bond proportion will be fifty-fifty, and in the last 10 years, the bond proportion will account for 75% of the portfolio assets.

Financial advisers often recommend the *100-x investment rule* to their clients. This means that an x-year old individual should invest $(100 - x)\%$ of his savings in low-risk assets (e.g. bonds) and the remaining $x\%$ in high-risk assets (e.g. stocks). In this study, the rule will be simplified to reduce the time needed for calibration, statistical tests and computation. The bond-stock ratio will be fixed for the year of the investor's round birthday (e.g. 40) and kept constant for five years before and five years after this age. Hereafter, we consider an example of a client who buys a product which expires on his 65th birthday (see Table 4.1).

Last, we describe the *zero-bond investment strategy.* The main idea is to provide a costless guarantee. If the guarantee provider wants to guarantee that the contribution C_{t_n} paid at time t_n $(t_n < T)$ will grow with the guaranteed rate of return g at the maturity T, he should invest $C_{t_n} e^{(g-r_{t_n,T})(T-t_n)}$ in a zero-bond with the risk-free rate of return $r_{t_n,T} > 0$ and time to maturity $T - t_n$. The remaining $C_{t_n}\left[1 - e^{(g-r_{t_n,T})(T-t_n)}\right]$ will be then invested in stocks, in order to participate in the growth chance of the stock market. At maturity, the value of the investment will equal the sum of the amount the provider has guaranteed $C_{t_n}e^{g(T-t_n)}$ and some non-negative amount

$$C_{t_n}\left[1-e^{(g-r_{t_n,T})(T-t_n)}\right]e^{\widetilde{r}_{t_n,T}},$$

where $\widetilde{r}_{t_n,T}\in\mathbb{R}$ is the realization of the stochastic stock return in the period $T-t_n$. For the contract with periodic contributions, the value of the portfolio at the end of the contract will be equal to

$$\begin{aligned}&\sum_{t_n=0}^{T-1}C_{t_n}e^{g(T-t_n)}+\sum_{t_n=0}^{T-1}C_{t_n}\left[1-e^{(g-r_{t_n,T})(T-t_n)}\right]e^{\widetilde{r}_{t_n,T}}\\&=\mathsf{G}_T+\sum_{t_n=0}^{T-1}C_{t_n}\left[1-e^{(g-r_{t_n,T})(T-t_n)}\right]e^{\widetilde{r}_{t_n,T}}\geq\mathsf{G}_T,\end{aligned}$$

where G_T is the guaranteed value given by equation (4.30). Please note that this strategy is only risk-free if the guarantee level g does not exceed the market risk-free rate $r_{t_n,T}$ ($r_{t_n,T}\geq g,\ \forall\ 0\leq t_n<T$) and if the zero-bond is default-free, as well. Both assumptions will be applied in this thesis.

4.7.1.3 Design of the study

Before we present the results of the study, we will give the explanation of the simulation design. An individual retirement account with a single or a periodic contribution payment is assumed. The periodic contribution C_{t_n} of €1200 is paid annually in advance. The single contribution is equal to the net present value of the yearly contributions (i.e. $\sum_{t_n=0}^{T-1}C_{t_n}e^{-r(T-t_n)}$). At the inception of the contract, the client has to fix the contract duration between one and 30 years[2] (T=1,2,...,30). At the contract inception, the client can chose between the guarantee of $g=-2\%$, 0%, 2% or 4% p.a. and one of the investment portfolios defined in Section 4.7.1.2. The investment

[2]For the 100-x investment rule, we assume that the client ends the contract at his 65th birthday.

strategy cannot be changed afterwards. As the contract expires, the contract provider guarantees to pay out the value of the investment portfolio S_T or the guaranteed amount G_T defined in equation (4.30), whichever is higher. The study excludes the surrender and biometrical risk, i.e. if the client cancels the contract (or dies) before the maturity, he (or his inheritors) receives the value of the portfolio. In a such case, no guarantee is given. The credit risk is excluded as well, i.e. we assume that the guarantee provider cannot default.

The provider collects two fees in order to cover the costs and to make a profit: the front-end-sales-charge and the administration charge. The front-end-sales-charge equals 3% of the bond fund units and 5% of the equity fund units. The administration charge is approximated by subtracting 0.5% p.a. from the average return of the investment. These are assumptions made by Maurer and Schlag (2003). In the case of the zero-bond strategy, both fees are charged, however, only from that of the part of the contribution invested in the stock fund. The investment in the zero-bond is charge-free as this part of the investment does not have to be actively managed. Please note that the guarantee is given on the gross contribution, so, e.g., in the one-year maturity case, the guarantee of 2% on the pure stock portfolio is, in fact, a guarantee of 7.13%[3] from the guarantor's point of view, because return from the investment has to cover both the front-end-sales-fee of 5% and the guarantee rate of 2%.

To estimate the distribution parameters of the returns, it is assumed that the bond fund returns have the same distribution as the returns of the German Bond Performance Index (REXP), and the equity fund returns have the same distribution as the returns of the German Stock Index (DAX30). In both cases there are performance indices involved, which means that the

[3]This is calculated from equation $0.95C_{t_n}e^{x} = C_{t_n}e^{0.02}$.

whole income from the investment (dividends, coupon-payment etc.) is reinvested in the portfolio underlying the index. The study compares results for the geometric Brownian motion and the geometric Brownian motion with Markov switching. The parameters for the GBM were estimated with the MLE method, and these of the GBM with regime switching with the EM algorithm (see Algorithm 2.71). Statistical tests have shown that the best representative of the GBM with Markov switching is the MS(1-2) model, i.e. the model with regime-independent mean and regime-dependent variance (see the discussion in Chapter 3).

To calibrate both models, 13 synthetic portfolios were built. It was assumed that on 31/12/1974 the amount of DM 195.58 (equivalent of €100) was invested in each buy-and-hold portfolio defined in Section 4.7.1.2 (i.e. with a 100%-0%, 90%-10%, ..., 0%-100%, and 75%-25% and 25%-75% REXP to DAX proportion, respectively) and that the portfolio was held until 31/12 /2004. Then the monthly log-returns were inferred from the development of the value of these portfolios. The parameters for both models are listed in Tables B.1-B.13 in Appendix B. As a discount rate, the risk-free rate of 0.44% per month (or equivalently 5.42% p.a.) was chosen, which is the average monthly money market rate (Monatsgeld) published by the Federal Bank of Germany (Deutsche Bundesbank) for the period from January 1975 to December 2004. This interest rate is also assumed to be the return of the zero-bond in the zero-bond investment strategy.

The option price for the GBM model was computed under the Black-Scholes martingale measure (see Theorem 4.48) and for the Markov switching model under the Bollen-Hardy (see Theorem 4.50) and Esscher (see Theorem 4.51) risk-neutral measure.

Please note that for the contract with the periodic payment scheme, there exists no analytic solution. For the contract with the single contribution, an analytical solution does exist. However, in the case of the Markov switching model, the use of the pricing formula is very time-consuming. For instance, a 30-year contract with an annual regime switching would require 2^{30} operations which amounts to more than one billion operations (see equation (4.20)). Thus, the option price was simulated via the Monte Carlo method with 1,000,000 iterations. The returns of the portfolio backing the contract were simulated on a monthly basis. Also, the regime change in the Markov switching model can occur at the end of each month (i.e the parameter τ in equation (4.20) is set to be one month).

Please note that different contracts have different cash-flows and different guarantee values. Therefore, the option prices were divided by the net present value of the contributions paid during the contract, i.e.

$$\widetilde{\mathsf{P}}_{t_0} = \frac{e^{-r(T-t_0)}\mathbb{E}_{\mathcal{Q}}\left[\max\{\mathsf{G}_T - S_T; 0\}\right]}{\sum_{t_n=0}^{T-1} C_{t_n} e^{-r(t_n-t_0)}} \tag{4.32}$$

in order to enable a comparison between contracts with different durations and different guarantee rates. The value from equation (4.32) will henceforth be referred to as the *normalized guarantee cost*.

4.7.2 Guarantee cost

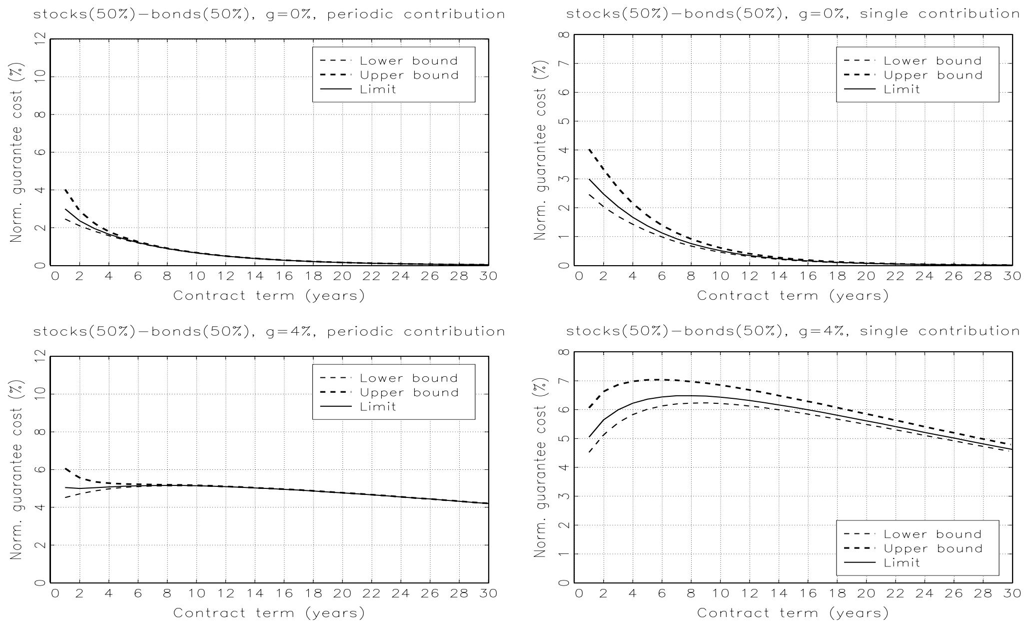
The following sections discuss the impact of different factors on the guarantee cost. The first two are rather technical and contribute to a better understanding of the Webb pricing model. Section 4.7.2.1 discusses the impact of the initial state on the guarantee cost. Section 4.7.2.2 compares the guarantee cost under three valuations models: the Black-Scholes, the Bollen-Hardy, and the Webb approach. Five further sections discuss the sensitivity of the guarantee cost to the change of guarantee level (see Section 4.7.2.3), investment strategy (see Section 4.7.2.4), time (see Section 4.7.2.5), contract term (see Section 4.7.2.6), and contribution payment scheme (see Section 4.7.2.7). They are interesting from the point of view of the guarantee seller, who wants understand the guarantee cost and manage it properly. Lastly, Section 4.7.2.8 discusses the interrelation of the guarantee cost and the expected profitability of the investment portfolio backing the guarantee. This section is particularly interesting from the viewpoint of the potential investor.

4.7.2.1 Impact of the initial state on the guarantee cost

The option price under the Esscher measure depends on the probability of the initial regime (see equation (4.20)). As the regime cannot be observed, the probability of the initial regime is unknown. Figure 4.1 depicts the normalized guarantee cost dependent on the initial state. The thin dashed line represents the normalized cost for the agent who knows with certainty that the market is in the low volatility state at the contract inception ($\Pr[Z_{t_0} = 1] = 1$). The thick dashed line depicts the price for the agent who knows for sure that the market is in the high volatility regime at the start of the contract ($\Pr[Z_{t_0} = 2] = 1$). The solid line represents the normalized cost for the agent who does not know which state the process was in at the beginning. Thus, he assumes that the process was in the low volatility regime with its ergodic probability, i.e., he assumes: $\Pr[Z_{t_0} = 1] = \pi_1$ and $\Pr[Z_{t_0} = 2] = \pi_2$ (see equation (2.69)).

The first two cases represent the lower and the upper bound of the Esscher option price. From Remark 2.34 we know that the limit transition probabilities of the homogeneous Markov chain (here the state variable Z_{t_n}) are the ergodic probabilities, as time goes to infinity. Thus, the case of the uninformed agent is, in fact, the limit of the guarantee cost.

Figure 4.1: Impact of the initial state on the normalized guarantee cost



Note:
This figure illustrates the impact of the initial state on the normalized guarantee cost by means of the fifty-fifty stock-bond portfolio and the MS(1-2) model under the Esscher probability measure. The thin dashed line represents the lower bound ($\Pr[Z_{t_0} = 1] = 1$) and the bold dashed line the upper bound ($\Pr[Z_{t_0} = 2] = 1$) of the normalized cost, respectively. The solid line represents its limit ($\Pr[Z_{t_0} = 1] = \pi_1$). The left column displays the periodic (€1200 up-front annually) and the right column the single contribution (equal to the net present value of periodic contributions) case. The top row shows the low level guarantee ($g = 0\%$ p.a.) and the bottom row the high level guarantee ($g = 4\%$ p.a.), respectively.

Figure 4.1 shows that, in most cases, the shapes of the lower and upper bound and the limit of the price are the same. However, in a few cases, they can differ. For instance, in the case of the fifty-fifty stock-bond portfolio and high level guarantee ($g = 4\%$), the upper bound decreases (first convex and then concave) while the lower bound first increases to a maximum and then decreases (the function is convex) (see the bottom left panel of Figure 4.1).

For low term contracts, the bounds are wide apart from each other. For instance, the cost of a 0% guarantee in the case of a one year maturity contract and the fifty-fifty stock-bond investment has a spread of 1.57 per cent points, according to the initial state. However, as the contract term increases, both bounds converge to the limit and the price spread vanishes. In the case of the periodical contributions, both buonds converge to each other after about 5-10 years (see the left column of Figure 4.1). In the case of the single contribution the convergence is much slower. For instance, for the fifty-fifty stock-bond portfolio with the guarantee level $g = 4\%$, the spread is still 0.24 per cent points after 30 years. This is consistent with the economic intuition: an agent who invests his whole capital at once (single premium) takes more risk than an agent who spreads the capital over time (periodic contribution). Furthermore, the convergence of the price to its limit occurs faster, *ceteris paribus*, (1) the lower the guarantee level g and/or (2) the higher the bond proportion x in the investment strategy.

We can conclude that the choice of the initial state is crucial for the normalized guarantee cost, especially for short time contracts. For contracts with longer maturities, the impact of the initial state is weaker. This should intuitively be expected. Thus, for agents buying a contract in a turbulent market phase, the longer the investment horizon they choose, the more chances they have to compensate initial losses (in comparison to buying the contract in the stable market phase).

The fact that the probability of the initial state cannot be observed constitutes a drawback to this method. There are, however, five ways to deal with this disadvantage. First, the risk averse actuary can choose the upper bound price. Second, the guarantee provider can use the ergodic probabilities to approximate the initial probabilities. If he has a portfolio of different cohorts, this average price would lead to a stable financial situation for the guarantor. Third, the guarantor can only sell contracts with periodic contri-

Table 4.2: Smoothed probabilities in December 2004 (under real-world and Bollen-Hardy measure)

Stock prop.	**100%**	**90%**	**80%**	**75%**	**70%**	**60%**	**50%**
Bond prop.	**0%**	**10%**	**20%**	**25%**	**30%**	**40%**	**50%**
$\Pr[Z_T = 1\vert\mathscr{Y}_T]$	0.9712	0.9703	0.9689	0.9684	0.9676	0.9647	0.9587
$\Pr[Z_T = 2\vert\mathscr{Y}_T]$	0.0288	0.0297	0.0311	0.0316	0.0324	0.0353	0.0413
Stock prop.	**40%**	**30%**	**25%**	**20%**	**10%**	**0%**	
Bond prop.	**60%**	**70%**	**75%**	**80%**	**90%**	**100%**	
$\Pr[Z_T = 1\vert\mathscr{Y}_T]$	0.9313	0.8792	0.8217	0.8873	0.9833	0.9943	
$\Pr[Z_T = 2\vert\mathscr{Y}_T]$	0.0687	0.1208	0.1783	0.1127	0.0167	0.0057	

Table 4.3: Smoothed probabilities in December 2004 (under Esscher measure)

Stock prop.	**100%**	**90%**	**80%**	**75%**	**70%**	**60%**	**50%**
Bond prop.	**0%**	**10%**	**20%**	**25%**	**30%**	**40%**	**50%**
$\Pr[Z_T = 1\vert\mathscr{Y}_T]$	0.9710	0.9701	0.9686	0.9681	0.9672	0.9642	0.9580
$\Pr[Z_T = 2\vert\mathscr{Y}_T]$	0.0290	0.0299	0.0314	0.0319	0.0328	0.0358	0.0420
Stock prop.	**40%**	**30%**	**25%**	**20%**	**10%**	**0%**	
Bond prop.	**60%**	**70%**	**75%**	**80%**	**90%**	**100%**	
$\Pr[Z_T = 1\vert\mathscr{Y}_T]$	0.9297	0.8755	0.8020	0.8715	0.9823	0.9939	
$\Pr[Z_T = 2\vert\mathscr{Y}_T]$	0.0703	0.1245	0.1980	0.1285	0.0177	0.0061	

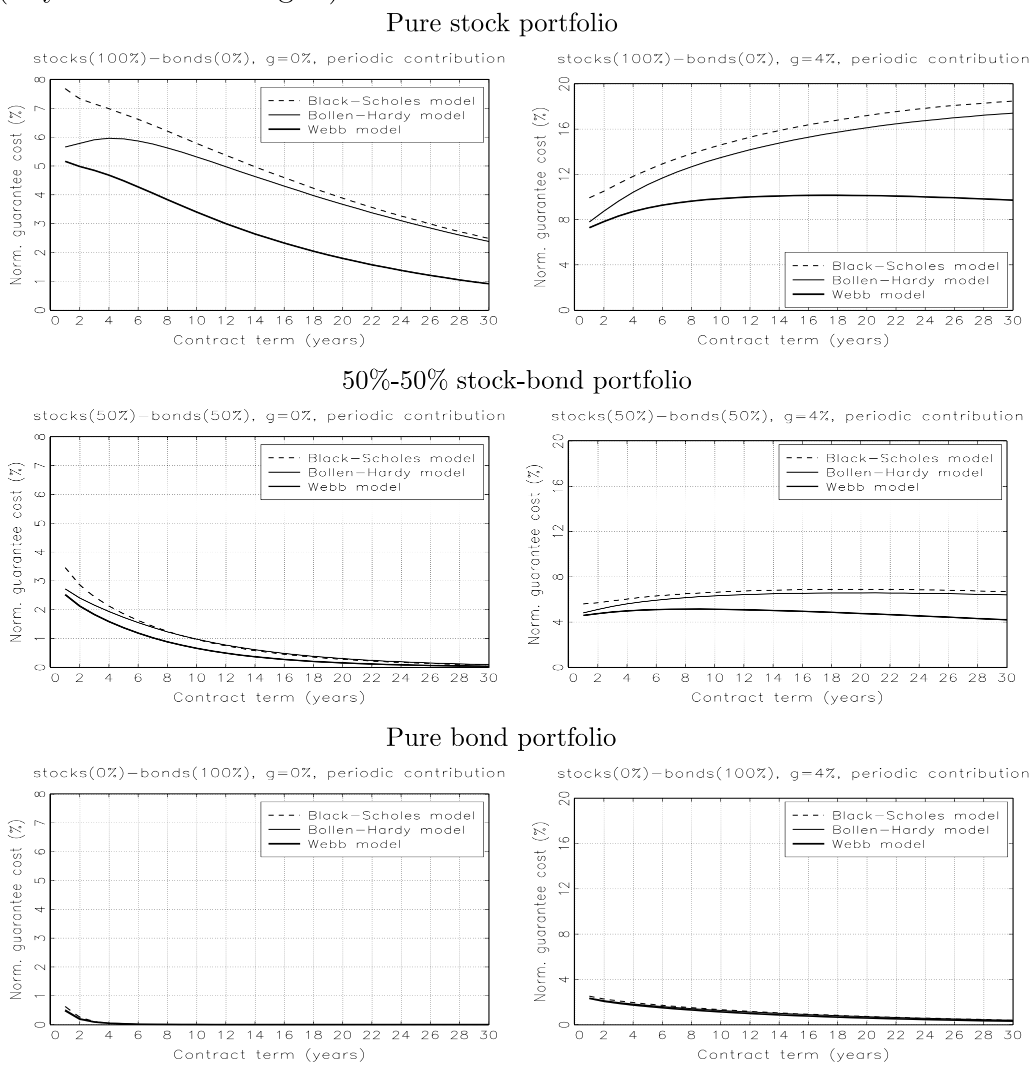
butions and maturities of not less than 10 years. As a result, the prices would converge to the limit and the uncertainty about the initial state would be smaller. Fourth, he can use the additional information from the estimation: the EM algorithm delivers the smoothed probabilities as the by-product of the estimation (see Section 2.8.2.2). Thus, the actuary can interpret them as initial state probabilities. Using this approach, the actuary has to control if the regime does not change. This means that he has to estimate the Markov switching model in constant time periods to receive the "actual" market smoothed probabilities. Last, the actuary can decrease the interval τ (see Theorem 4.51) on which the value of the investment portfolio is to be simulated. In this contribution we use $\tau = 1$ month. The decrease of τ would, however, increase the computation time. For this reason we reject this solution. Instead, we decide to use the additional information from smoothed probabilities. Table 4.2 shows smoothed probabilities under the Bollen-Hardy (and real-world) probability measure and Table 4.3 shows smoothed probabilities under the Esscher probability measure.

4.7.2.2 Impact of the stochastic model on the guarantee cost

This section discusses the impact of the option pricing model on the guarantee cost. Accordingly, three option price models are compared: the Black-Scholes (see Theorem 4.48), the Bollen-Hardy (see Theorem 4.50), and the Webb model (see Theorem 4.51).

Figure 4.2 shows the results. For all possible stock-bond proportions and for all studied guarantee levels, the same is true: the guarantee cost is the highest in the Black-Scholes approach (thin dashed line) and the lowest in the Webb approach (thick solid line). The cost yield according to the Bollen-Hardy model (thin solid line) lies between the other two. Additionally, it can be seen that the cost yield from the Bollen-Hardy model converges to the Black-Scholes cost from below. In all cases the following statements hold true: taking heteroskedasticity into consideration leads to a price decrease. For the low level guarantee of 0%, the price of the Bollen-Hardy model is lower then the price of the Black-Scholes model. The difference decreases as the contract maturity grows. This shows that the Black-Scholes price is the most conservative one. The additional consideration of stochastic volatility lowers the guarantee cost. As shown in Chapter 3, the Markov switching model is more suitable for describing the stochasticity of financial assets than the GBM. This shows that the choice of the less suitable stochastic

Figure 4.2: Impact of the stochastic model on the normalized guarantee cost (buy-and-hold strategies)



Note:

This figure depicts the impact of the stochastic model on the normalized guarantee cost in the example of buy-and-hold strategies with periodic contributions (€1200 up-front annually) and the MS(1-2) model. The dashed line represents the Black-Scholes, the thin solid line – the Bollen-Hardy, and the bold solid line – the Webb model. The top row depicts the pure stock portfolio, the middle row – the fifty-fifty stock-bond portfolio, and the bottom row – the pure bond portfolio with periodic contributions, respectively. The left column depicts contracts with a low level guarantee ($g = 0\%$ p.a.), and the right column contracts with a high level guarantee ($g = 4\%$ p.a.).

model (i.e. GBM) leads to an overpricing of the guarantee.

However, the Markov switching market is incomplete. Thus, the option price is not unique. Two models discussed here yield different prices, therefore, the choice of the suitable martingale measure is crucial for the results. The Esscher measure is a better choice, as it prices the stochastic risk (i.e. it takes into account the stochasticity of the state variable Z_{t_n}) which is neglected by the Bollen-Hardy measure. (For other advantages of the Esscher measure, see Section 4.6.2.2).

Neverless, the Bollen-Hardy model can contribute to an understanding of risk premia. Each of these three models introduces an additional pricing risk. The Black-Scholes option pricing is an approach with the stochastic stock return (arithmetic Brownian motion). This approach assumes that the volatility process of the stock return has a constant parameter σ. The Bollen-Hardy approach introduces a stochastic volatility process (arithmetic Brownian motion with Markov switching) for the stock return. This approach assumes that the volatility process of the stock return has a random parameter σ_1 or σ_2, depending on the latent random variable Z_t (state of the market). This approach, however, does not price the uncertainty linked to the switching parameter Z_t. Thus, the spread between the Black-Scholes and the Bollen-Hardy price quantifies the stochastic volatility risk. The spread between the Bollen-Hardy and the Webb price quantifies the switching risk (i.e., uncertainty if the market is in the stable or in the turbulent phase).

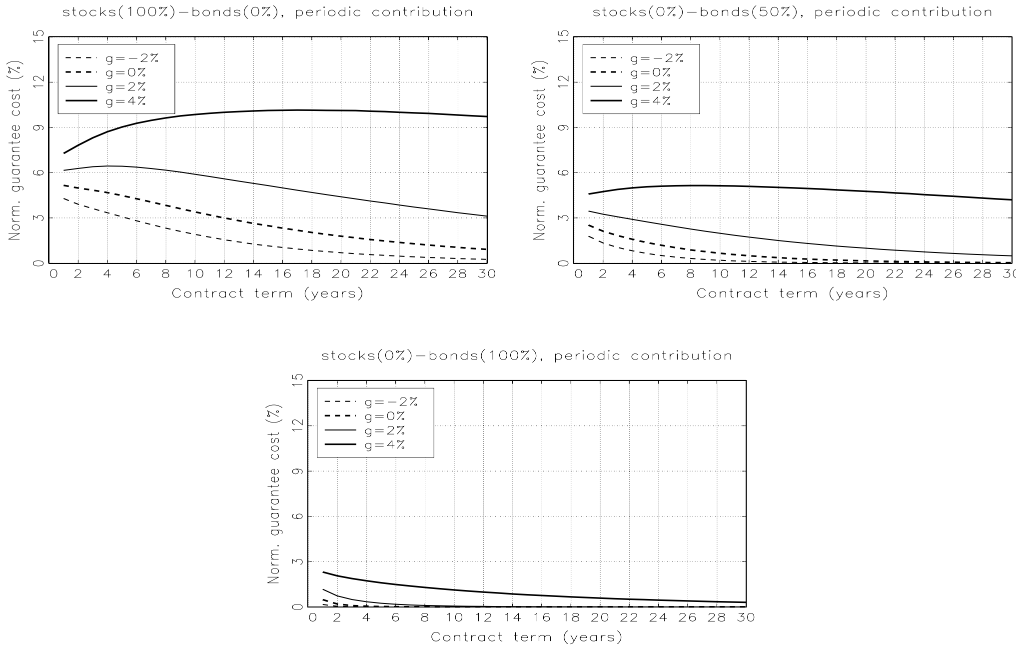
Figure 4.2 clearly shows that neglecting to take into account the stochastic volatility risk leads to an overpricing of the guarantee. This effect is lower, *ceteris paribus*, (1) the lower the contract term and/or (2) the lower the stock proportion in the investment strategy. The effect of neglecting the switching risk leads to more crucial overpricing of the guarantee. Furthermore, this effect grows as (1) the contract term and/or (2) the stock proportion in the strategy increase. This is important as the retirement products we discuss have a long time nature.

4.7.2.3 Impact of the guarantee level on the guarantee cost

This Section analyzes the impact of the guarantee level on the guarantee cost. First, we can state that the cost of guarantees with $g = -2\%$, 0% and 2% behaves similarly. On the contrary, the cost of the 4% guarantee displays a different behavior (see Figure 4.3). The normalized cost of a -2%, 0%, and 2% guarantee is a decreasing function for all stock-bond portfolios. The only exception is the normalized cost of 2% backed by portfolios with a stock proportion of 80% and more. In those cases the normalized cost increases slightly in order to reach a maximum (for a contract with 3 or 4 years duration), and subsequently decreases. In contrast to the other three guarantee levels, the normalized cost of the 4% guarantee has a different shape. It initially increases towards a maximum and decreases afterwards. In only two cases, that of a pure bond and that a the 10%-90% stock-bond portfolio, the curve of the normalized cost of a 4% guarantee decrease for all contract terms.

The guarantee cost increases along with an increase of the guarantee level, which is self-evident. Figure 4.4 shows how the normalized cost reacts if the guarantee level increases by 2 per cent points: from -2% to 0% p.a. (thin dashed line), from 0% to 2% p.a. (thick dashed line), and from 2% to 4% p.a. (solid line). The figure shows that the cost reacts overproportinally to the change in the guarantee level. Particularly, the increase is the highest, when the guarantee level is risen from 2% to 4% p.a. The sensitivity of the cost to the guarantee level is much smaller in the two other cases.

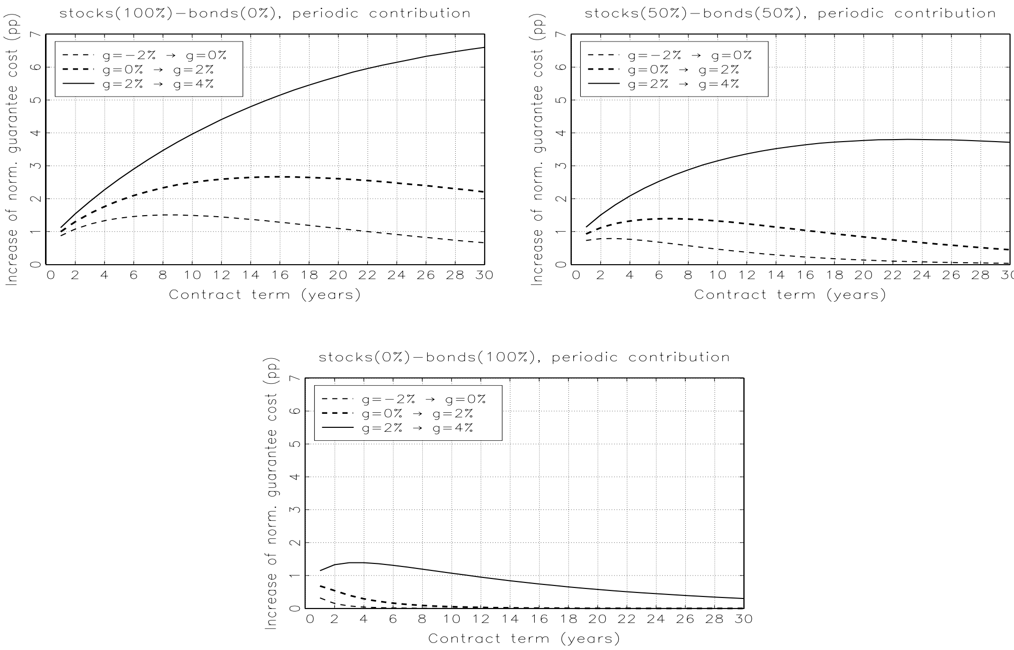
Figure 4.3: Impact of the guarantee level on the normalized guarantee cost



Note:
This figure depicts the impact of the guarantee level on the normalized guarantee cost using the example of the pure stock (top left panel), fifty-fifty stock-bond (top right panel), and the pure bond portfolio (bottom panel) with periodic contributions (€1200 up-front annually). The cost is computed for the MS(1-2) model under the Esscher probability measure. The thin dashed line represents the guarantee of -2% p.a., the bold dashed line – 0% p.a., the thin solid line – 2% p.a., and the bold solid line – 4% p.a., respectively.

In conclusion, the guarantee level of 4% behaves differently from the other three guarantee levels discussed in this contribution, and has a far higher cost as well. Thus, we henceforth refer to the 4% guarantee as the *high level guarantee* and to remaining three guarantee levels as the *low level guarantee*. The low level guarantee will hencefrom be discussed using the example of the 0% guarantee.

Figure 4.4: Sensitivity of the normalized guarantee cost to changes in the guarantee level



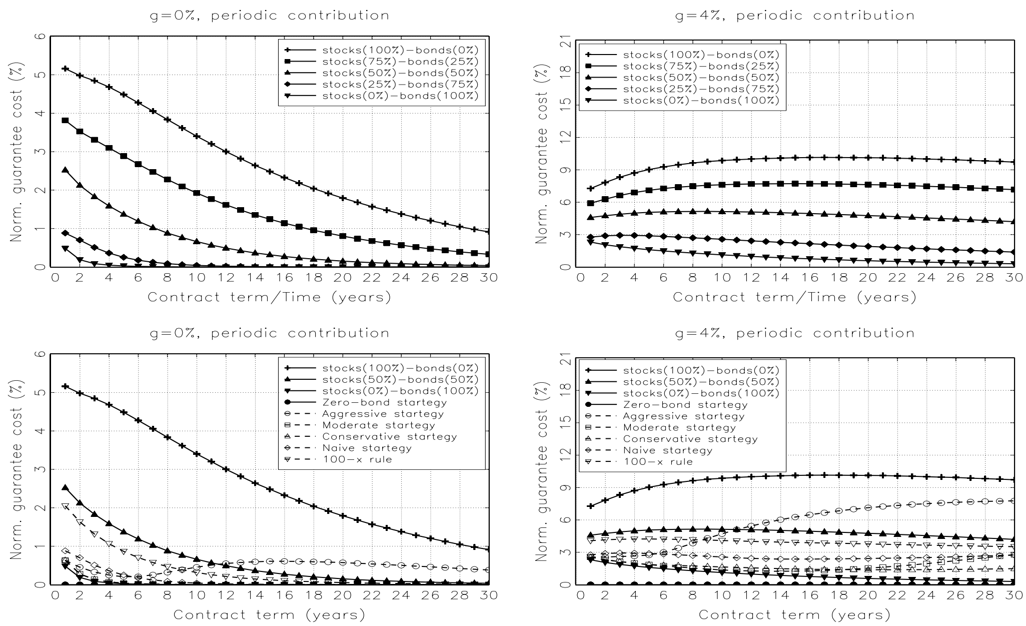
Note:
This figure depicts the sensitivity of the normalized cost to the change of the guarantee level in the example of the pure stock (top left panel), fifty-fifty stock-bond (top right panel), and the pure bond portfolio (bottom panel) with periodic contributions (€1200 up-front annually). The sensitivity is measured in per cent points (pp). The cost is computed for the MS(1-2) model under the Esscher probability measure. The thin dashed line represents the guarantee increase from -2% to 0% p.a., the bold dashed line – from 0% to 2% p.a., and the solid line – from 2% to 4% p.a., respectively.

4.7.2.4 Impact of the investment strategy on the guarantee cost

This Section addresses the impact of the investment strategy on the guarantee price. First, buy-and-hold strategies will be discussed, followed by life-cycle strategies, and lastly, the zero-bond strategy.

From the option price theory it is known that the option price increases with the increase of volatility. Figure 4.5 shows that the normalized cost of the guarantee increases as the proportion of stocks in the investment portfolio

Figure 4.5: Impact of the investment strategy on the normalized guarantee cost



Note:

This figure depicts the impact of the investment strategy on the normalized guarantee cost. The cost is computed for the MS(1-2) model under the Esscher probability measure. The solid line with pluses represents the pure stock investment strategy, the solid line with squares – the 75%-25% stock-bond, the solid line with triangles – the 50%-50% stock-bond, the solid line with diamonds – the 25%-75% stock-bond, the solid line with inverse triangles – the pure bond, the solid line with circles – the zero-bond, the dashed line with circles – the aggressive, the dashed line with squares – the moderate, the dashed line with triangles – the conservative, the dashed line with diamonds – the naive, and the dashed line with inverse triangles – the 100-x investment strategy, respectively. In all cases contributions of €1200 are paid up-front annually. The left column represents the low level guarantee ($g = 0\%$ p.a.) and the right row the high level guarantee ($g = 4\%$ p.a.), respectively.

grows. This is true for all guarantee levels and all contract terms. Only in some cases the increase of the stock proportion does not imply the increase of the cost. This is true for portfolios with a long duration, a low guarantee level, and a low stock proportion. For instance, all contracts with a guarantee level of 0%, a term above 15 years, and a stock proportion of up to 25% have a zero cost. In these cases, the rise of the stock proportion does not affect the price, but increases the profit chance.

The impact of the stock proportion is greater, the higher the guarantee level is. Assume that the guarantee provider believes that his potential clients are willing to pay a maximum of 5% of discounted contributions for the guarantee. Then, he could sell contracts backed with any buy-and-hold strategy and the normalized price of the 0% guarantee would still not exceed 5% (excluding a one-year contract backed by the pure stock strategy). However, in the case of the 4% guarantee, only portfolios with maximum of 50% stocks have a normalized cost that lies below 5%. The guarantor would thus only offer his clients these particular investment portfolios.

Thus, to keep the price below 5%, the guarantor has to reduce the guarantee level (i.e. he offers less protection) or the equity proportion in the portfolio (i.e. he lowers the expected portfolio return). Obviously, clients are interested in both. One way to satisfy both needs and to keep the guarantee price at a moderate level could be to apply one of the life-cycle strategies (see Table 4.1). The lower part of Figure 4.5 shows that four of the five life-cycle discussed strategies are more expensive than the fifty-fifty stock-bond buy-and-hold strategy. These strategies are: the moderate, the conservative, the naive one, and the 100-x investment rule. Only the aggressive strategy can cost more than the fifty-fifty stock-bond strategy, for the middle and the long term contracts. For some long term contracts, its cost even exceedes the cost of the 75%-25% investment portfolio.

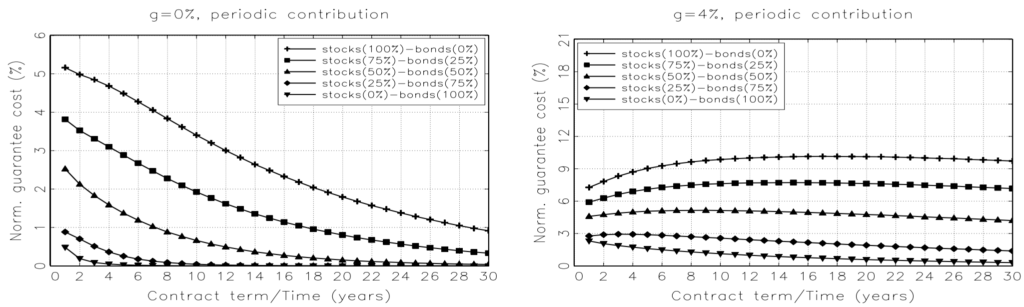
Finally, let us turn our attention to the zero-bond strategy. For the risk-averse guarantee provider, this is clearly the best strategy, as it is costless for all possible configurations of guarantee levels and contract durations. However, it is an open question if this strategy is attractive for clients. It is possible that the price they pay for the protection is an unsatisfactory expected profit. This issue will be discussed in Section 4.7.2.8.

4.7.2.5 Impact of time on the guarantee cost

This Section addresses the impact of time on the (normalized) guarantee cost. First, we discuss buy-and-hold strategies, then life-cycle strategies, and, finally, the zero-bond strategy. The left panel in Figure 4.6 shows the impact of time on the normalized cost of a 0% guarantee. For all buy-and-hold investment strategies, the cost function is decreasing (all curves except the pure stock strategy are concave) and converges toward zero. The lower the stock proportion in the backed portfolio, the faster the cost decreases to zero. For portfolios with a stock proportion up to 50%, the cost reaches zero between the 7th (pure bond) and the 27th year (fifty-fifty stock-bond). Neither of the remaining strategies reach zero before the 30th year. An additional simulation, however, has shown that they converge towards zero afterwards.

The right panel in Figure 4.6 displays the impact of time on the normalized cost of a 4% guarantee. The cost function is concave and decreasing for strategies with a maximum stock proportion of 20%. However, it does not reach zero, even in the 30th year. For portfolios with a bond proportion of more than 20%, the cost function is increasing until a certain maximum, and then slowly decreases afterwards. The decrease is slower the higher the stock

Figure 4.6: Impact of time and the contract term on the normalized guarantee cost (buy-and-hold strategies)



Note:
This figure depicts the impact of time and the contract term on the normalized guarantee cost. The cost is computed for the MS(1-2) model under the Esscher probability measure. The solid line with pluses represents the pure stock investment strategy, the solid line with squares – the 75%-25% stock-bond, the solid line with triangles – the 50%-50% stock-bond, the solid line with diamonds – the 25%-75% stock-bond, the solid line with inverse triangles – the pure bond investment strategy, respectively. In all cases contributions of €1200 are paid up-front annually. The left column represents the low level guarantee ($g = 0\%$ p.a.) and the right row the high level guarantee ($g = 4\%$ p.a.), respectively.

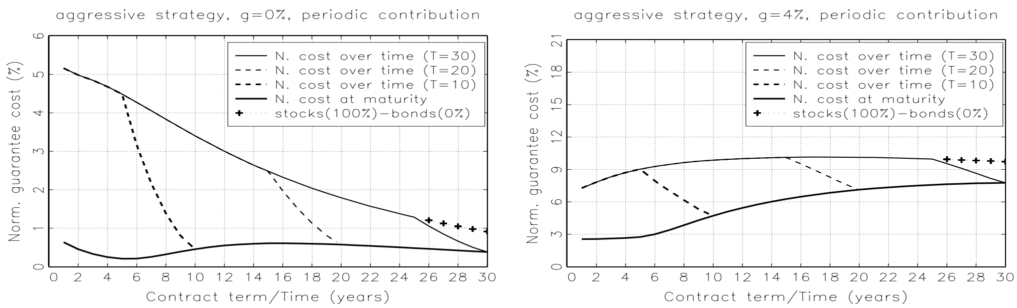
proportion is. For example, for a 20%-80% stock-bond portfolio, the maximum of 2.52% is reached after 3 years. Then the normalized cost decreases to 0.95% after 30 years. For a pure stock portfolio, the maximum of 10.15% is reached after 17 years. Then the normalized cost decreases to 9.72% after 30 years.

The convergence of the cost toward zero also holds true for a high level guarantee of 4%, however, it is much slower than for low level guarantees.

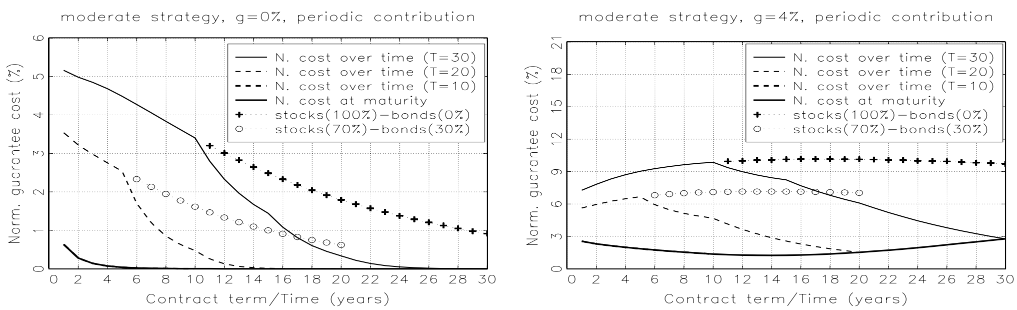
Figure 4.7 shows the impact of time on the life-cycle strategies in three examples: a 10-year contract (thick dashed line), a 20-year contract (thin dashed line), and a 30-year contract (thin solid line). As stated in Section 4.7.2.4, the behavior of the aggressive strategy is different from the behavior of the remaining life-cycle strategies. It will therefore be discussed separately. The remaining life-cycle strategies will be discussed using the example of the

Figure 4.7: Impact of time and the contract term on the guarantee cost (life-cycle strategies)

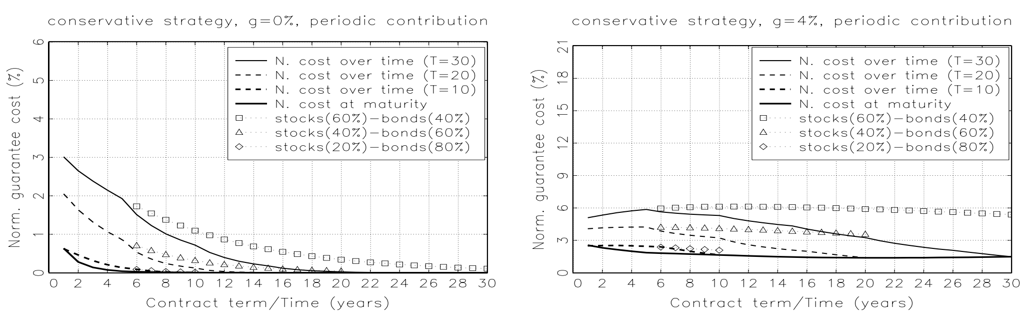
Aggressive strategy



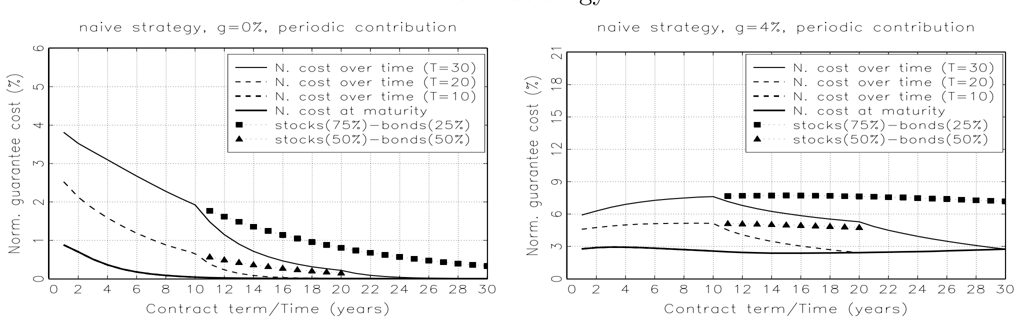
Moderate strategy



Conservative strategy



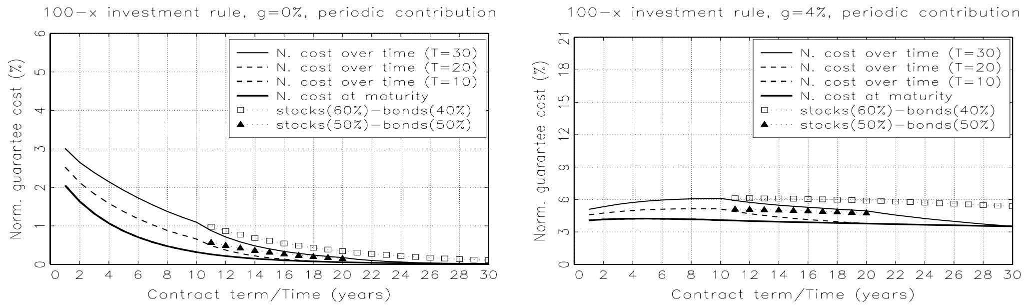
Naive strategy



continued on the next page

continued from the previous page

100-x investment rule



Note:
This figure depicts the impact of time and the contract term on the normalized guarantee cost. The cost is computed for the MS(1-2) model under the Esscher probability measure. The bold dashed line shows how the normalized cost changes through time for a 10-year contract, the thin dashed line for a 20-year contract, and the thin solid line for a 30-year contract, respectively. The bold solid line shows the normalized contract at contract expiration. The bold pluses depict the 100%-0%, the solid squares – the 75%-25%, the empty circles – the 70%-30%, the empty squares – the 60%-40%, the solid triangles – the 50%-50%, the empty triangles – the 40%-60%, and the empty diamonds – the 20%-80% stock-bond buy-and-hold portfolio, respectively. In all cases, contributions of €1200 are paid up-front annually. The left column represents the low level guarantee ($g = 0\%$ p.a.), and the right row the high level guarantee ($g = 4\%$ p.a.), respectively.

moderate strategy.

Let us discuss the 0% guarantee backed by the moderate strategy. The guarantee cost decreases for all three of the discussed contracts. The speed with which the cost decreases grows at each shifting date (e.g. after the 10th and 20th year for the 30-year contract). The cost for all three contracts reaches zero before contract expiration, i.e. in the 9th, the 17th, the 28th year for the 10-, 20-, and 30-year contract, respectively. This means that if the shift from a more risky toward a less risky portfolio occurred one or two years later, the client would still recive a contract with a costless guarantee, however, with a higher expected profit. Thus, the contract provider could slightly optimize the moderate strategy to better satisfy the needs of the client, i.e. of safety and of a higher expected profit.

Furthermore, it is possible to quantify the cost reduction of the moderate strategy. As the 30-year contract invests in a pure stock portfolio, we can compare the cost of this portfolio to the cost of the moderate strategy. The difference in the normalized cost of both investment portfolios equals 0.91%. For the 20-year contract, the difference between the initial portfolio (i.e. 70%-30% stock-bond portfolio) and the moderate strategy equals 0.62%. As the 10-year contract invests in the 10%-90% stock-bond portfolio for the whole investment period, no cost reduction can be found.

Let us discuss the 4% guarantee backed by the moderate strategy. The normalized cost of the 10-year contract is always decreasing. The normalized cost of the 20- and 30-year contract increases until the first shifting date and then decreases. Thereby, at each further shifting date, the speed of cost reduction increases. In contrast to the 0% guarantee, the cost at the expiration date does not reach zero for either contract. However, the cost reduction compared to the initial portfolio is much higher than in the case of the 0% guarantee. For the 20-year contract, the normalized cost at expiration sinks from 7.03% (70%-30% stock-bond portfolio) to 1.53% (moderate strategy), and for the 30-year contract from 9.72% (pure bond portfolio) to 2.78% (moderate strategy). Since the 10-year contract only invests in the 10%-90% stock-bond portfolio, no cost reduction occurs.

The behavior of the aggressive strategy is slightly different. Although the cost of the 0% guarantee decreases along with time and although the normalized cost of the 4% guarantee increases until the first asset shifting date, and then decreases, it fails to obtain zero for either of the discussed contracts. Another difference is that the cost reduction is greater for short- than for long-term contracts.

The cost of the zero-bond strategy is insensitive to changes in time, as it equals zero, due to the strategy construction.

4.7.2.6 Impact of the contract term on the guarantee cost

In the case of buy-and-hold investment strategies, the impact of the contract term on the guarantee cost is exactly the same as the impact of time. Therefore, this discussion will be omitted.

The impact of the contract term on the normalized guarantee cost of life-cycle strategies is shown in Figure 4.7 (thick solid line). As the aggressive strategy behaves differently from other life-cycle strategies, it is discussed separately. Other life-cycle strategies will be discussed using the example of the moderate strategy.

In the case of the aggressive strategy, the normalized cost of the 0% guarantee is relatively stable around the 0.45% mark. There are two factors which influence this behavior. First, the guarantee cost associated with the initial portfolio (here the pure stock portfolio) sinks as the contract term increases. Thus, this factor has a greater influence on the long-term contracts. The second factor is the shifting of assets from more risky into less risky ones. The influence of this factor increases the shorter the contract term is.

In the case of the aggressive strategy, the normalized cost of the 4% guarantee increases as the contract term grows. The reason for this is that the guarantee cost associated with the initial portfolio is relativ high. Since the risk-reducing influence of asset shifts weakens as the contract term grows, it is not able to predominate the influence of the first factor. To reduce the cost of long-term contracts, it would be necessary to start the asset shifting earlier than in the last five years or to start with a less risky initial portfolio than the pure stock one (see, for example, the moderate strategy).

The normalized cost of the 0% guarantee associated with the moderate strategy is influenced by the same two factors: (1) the sinking cost of the initial portfolio and (2) the asset shifting. The difference to the agressive strategy, however, is that the normalized cost is a decreasing function that converges to zero for a 9-year contract. This is due to the earlier asset shifting compared to the aggressive portfolio.

The normalized cost of the 4% guarantee associated with the moderate strategy is a concave function of the contract term, with the minimum being the 14-year contract. This is due to the strategy construction: First, the shorter the contract term is, the less risky the initial portfolio. Thus, it does not predominate the shifting effect. Second, the asset shifting begins prior to the last five contract years – thus, it has a stronger impact on the guarantee costs than in the case of the aggressive strategy.

We stated above that the normalized cost function in the moderate investment strategy with the 4% guarantee has a minimum. It illustrates that for a given contract term, the desired guarantee cost can be achieved. The guarantee provider should hence optimize the investment strategy by choosing the optimal initial portfolio and the optimal shifting design.

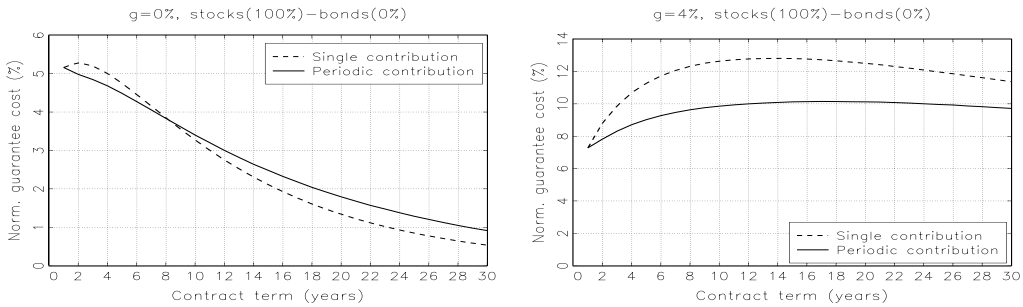
Finally, we will discuss the zero-bond strategy. Its cost is insensitive to changes in the contract term, as it equals zero due to the strategy construction.

4.7.2.7 Impact of the contribution payment scheme on the guarantee cost

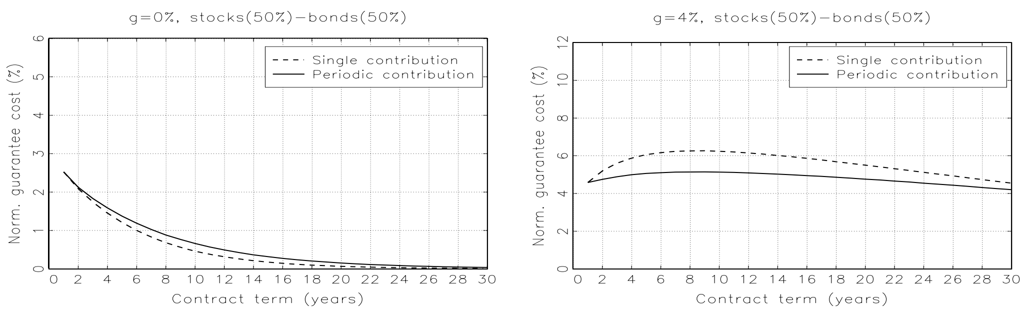
This section compares the normalized guarantee cost for two alternative contribution payment schemes. The single premium and the periodic premium (€1200 up-front annually). For the sake of comparability, the single premium is chosen to be the net present value of the aggregated periodic payments.

Figure 4.8: Impact of the contribution payment scheme on the normalized guarantee cost (buy-and-hold strategies)

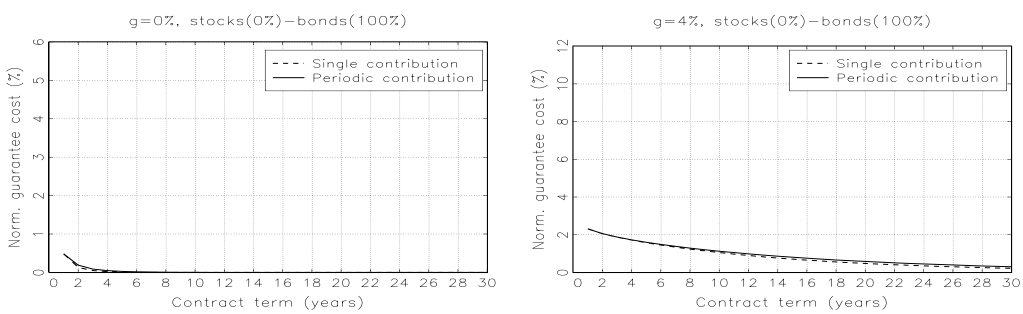
(a) Pure stock portfolio (b)



(c) 50%-50% stock-bond portfolio (d)



(e) Pure bond portfolio (f)



Note:

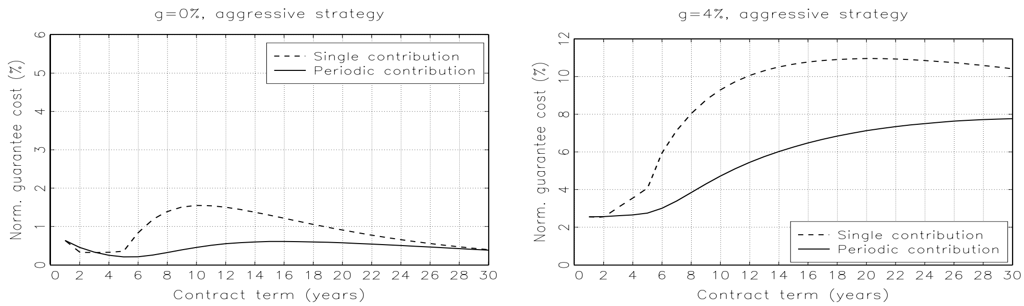
This figure depicts the impact of the payment scheme on the normalized guarantee cost using the example of buy-and-hold strategies, and the MS(1-2) model under the Esscher probability measure. The solid line represents the periodic payment scheme (€1200 up-front annually), and the dashed line the single premium case. The single premium equals the net present value of periodic contributions. The left column shows contracts with a low level guarantee ($g = 0\%$ p.a.), and the right column contracts with a high level guarantee ($g = 4\%$ p.a.).

For buy-and-hold strategies, both curves have the same shape (see Figure 4.8). The normalized cost for the periodic payment scheme is higher than for the single payment scheme when the cost function is concave (see Figure 4.8, Panels (c), (e), and (f)). The opposite holds true if the cost function is convex. In this case, contracts with a single premium have a higher price than contracts with a periodic premium (see Figure 4.8, Panels (b) and (d)). This rule also holds true for cost functions with changing convexity. In such a case, both cost functions cross at their inflection points (see Figure 4.8, Panel (a)). Furthermore, the discrepancy between both cost functions is higher, the higher the stock proportion in the investment portfolio (compare the top and the bottom row in Figure 4.8) and/or the higher the guarantee level (compare the left and the right column of Figure 4.8). In other words, the cost for the single contribution is higher than the cost for the periodic contribution contract if the risk of achieving the guarantee is high (i.e. for contracts with a high stock proportion and a high guarantee level). In all other cases, the contracts with periodic contributions are more expensive.

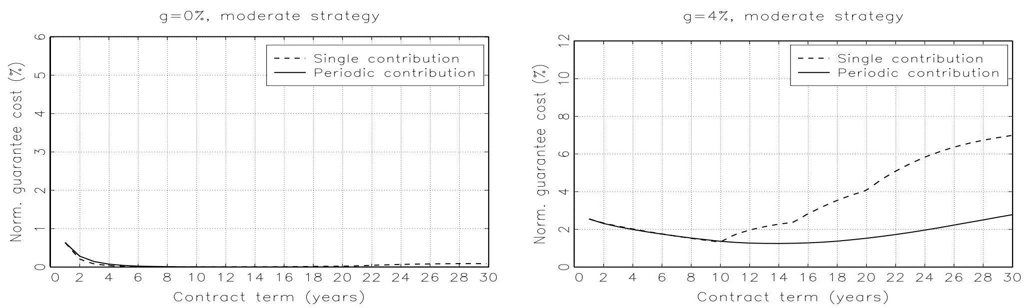
Figure 4.9 compares the normalized cost of the single and periodic payment scheme for life-cycle investment strategies. In this case, cost functions for the single and periodic payment do not necessarily have the same shape. The cost function for periodic payment is smooth, and for single contributions it may have kinks. The kinks always occur when the investment design changes. Consider, for example, the 4% guarantee backed by the naive strategy. The kinks occur in contracts with terms of 10 and 20 years. The reason for this that all contracts with a term of up to 10 years only invest in the 25%-75% stock-bond portfolio. All contracts with a term between 11 and 20 years invest in the fifty-fifty stock-bond portfolio at the beginning of the contract, and then the portfolio manager shifts to the 25%-75% stock-bond

Figure 4.9: Impact of the contribution payment scheme on the normalized guarantee cost (life-cycle strategies)

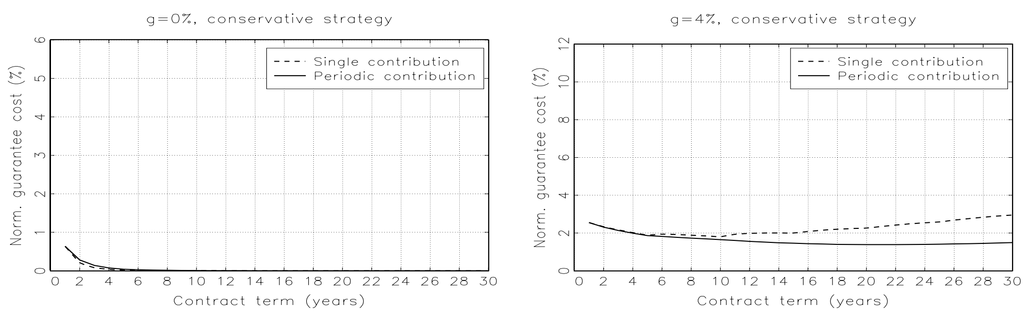
Aggressive strategy



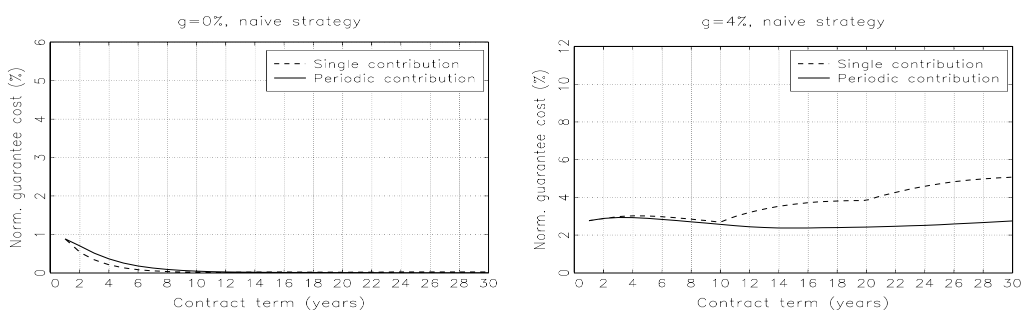
Moderate strategy



Conservative strategy



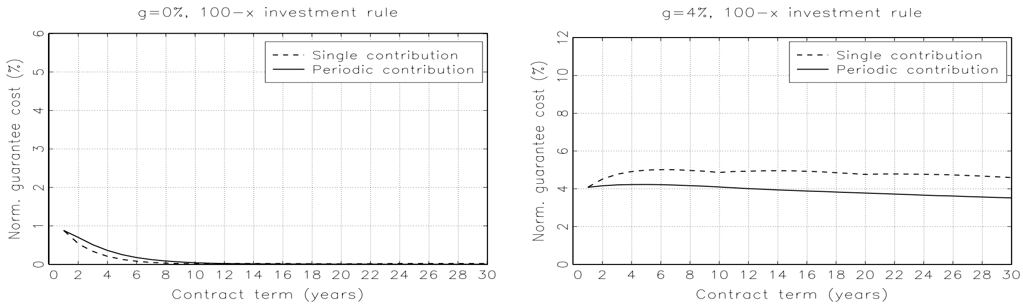
Naive strategy



continued on the next page

continued from the previous page

100-x investment rule



Note:

This figure depicts the impact of the payment scheme on the normalized guarantee cost using the example of life-cycle strategies, and the MS(1-2) model under the Esscher probability measure. The solid line represents the periodic payment scheme (€1200 up-font annually) and the dashed line the single premium case. The single premium is equal to the net present value of periodic contributions. The top row depicts the naive strategy, and the bottom row – the 100-x investment rule, respectively. The left column represents the low level guarantee ($g = 0\%$ p.a.) and the right row the high level guarantee ($g = 4\%$ p.a.), respectively.

portfolio towards the end of the contract. All contracts with a term between 21 and 30 years invest in the 75%-25%, then in 50%-50%, and then in the 25%-75% portfolio. These three classes of contracts therefore have three different investment designs.

In a further discussion, we divide the investment strategies into two groups, which disclose a similar behavior. The first group consists of the aggressive strategy (with both the low and the high level guarantee), and all remaining life-cycle strategies (with the high level guarantee). The second group consists of the conservative and naive strategy and the 100-x investment rule (with the low level guarantee). The moderate strategy with a low level guarantee has a unique behavior and will thus be discussed separately. In the first group, the single payment scheme leads to higher cost than in the periodic payment scheme. There are three exceptions to this rule: the moderate strategy (for all contracts with terms of up to 10 years), the conservative strategy (with terms of up to 5 years), and the aggressive strategy (with terms of up to 2 years) the high level guarantee has the same cost for both payment schemes. For all contracts from the first group, the single contribution curve has kinks. This shows that contracts with a high guarantee level and a single contribution are more sensitive to changes in the portfolio design than contracts with periodic contributions. Finally, no contract from the first group has a zero cost guarantee.

Contracts from the second group have a higher cost if the contribution payment occurs periodically and the contract term is low. For the remaining durations, the cost in both payment schemes equals zero.

Another very interesting case is that of the moderate investment strategy with a low guarantee level. The cost for the single payment contract for short term contracts is lower than the cost for a periodic contribution payment. For middle term contracts, the cost of both payment schemes equals zero. For long term contracts, the cost in the case of the periodic payment remains zero, but in the case of a single premium, it rises slightly above zero.

The cost of the zero-bond strategy is insensitive to changes in the payment scheme, as its cost equals zero due to the strategy construction.

4.7.2.8 Interrelation between guarantee cost and expected profit

Sections 4.7.2.3-4.7.2.7 addressed the impact of diverse factors (guarantee level, investment strategy, time, contract term, and contribution payment scheme) on the guarantee cost. All of these aspects are important from the perspective of a financial company considering to sell a guarantee to its customers or managing the guarantee risk. On the contrary, this section considers the guarantee from the customer's point of view. A future pensioner is interested in three aspects when buying a guarantee: the maximization of the protection level (level of the guarantee), the reduction of the cost of the protection, and the maximization of profit. As we discuss several investment strategies, we would like to see which one of them is most suitable for a client and most likely to satisfy his goals. To compare the profit potential of different investment strategies, we use the normalized expected profit.

Definition 4.53 (Normalized expected profit) *Let the stochastic process* $(\phi)_{t\in[t_0,T]}$ *be an investment strategy with a portfolio value* $(S_t)_{t\in[T_0,T]}$*, and an investment horizon* T*. Let* C_{t_n} *be a contribution paid at time* t_n *(*$t_0 < t_n < T$*) and let* r *be a riskless interest rate, then*

$$\widetilde{\Pi}_{t_0} = \frac{e^{-r(T-t_0)}\mathbb{E}[S_T]}{\sum_{t_n=0}^{T-1} C_{t_n} e^{-r(t_n-t_0)}} - 1. \tag{4.33}$$

is a normalized expected profit of this strategy.

Table 4.4 ranks chosen investment guarantees from the least to the most costly. If two strategies have the same cost, better ranking is given to the one with the higher expected profit. The aim of Table 4.4 is to help the client to decide which product to buy. We thereby assume that he knows his investment horizon (10, 20 or 30 years) and how much risk protection he needs (guarantee of 0% or 4%). Furthermore, we assume that he wants to maximize his normalized profit by a given normalized guarantee cost of 0% (costless guarantee), 1%, 2.5%, 5%, 7.5% or 10%.

In the following, we will consider the example of the 30-year investment horizon, the remaining examples can be seen in Table 4.4. A client interested in a 0% guarantee and accepting only costless investment strategies would purchase a contract with the zero-bond investment strategy, which has the normalized expected profit of 86.66%. This example shows how important the investment strategy is from the client's point of view: some strategies

Table 4.4: Ranking of investment strategies (normalized guarantee cost vs. normalized expected profit)

g=0%, T=10			g=4%, T=10		
Strategy	**Norm. cost**	**Norm. profit**	**Strategy**	**Norm. cost**	**Norm. profit**
Moderate*	0.00%	8.27%	**Zero-bond**	0.00%	2.08%
Zero-bond	0.00%	7.34%	**100% bond**	1.12%	6.84%
100% bond	0.00%	6.84%	**Moderate***	1.36%	8.27%
Conservative	0.01%	8.54%	**Conservative**	1.64%	8.54%
Naive**	0.04%	10.15%	**Naive****	2.57%	10.15%
100-x rule***	0.32%	12.74%	**20%-80% stock-bond**	2.08%	9.17%
Aggressive	0.45%	13.32%	**100-x rule*****	4.09%	12.74%
50%-50% stock-bond	0.66%	13.63%	**Aggressive**	4.71%	13.32%
80%-20% stock-bond	2.20%	18.22%	**70%-30% stock-bond**	7.10%	17.17%
100% stock	3.40%	22.20%	**100% stock**	9.86%	22.20%
g=0%, T=20			g=4%, T=20		
Strategy	**Norm. cost**	**Norm. profit**	**Strategy**	**Norm. cost**	**Norm. profit**
Zero-bond	0.00%	33.98%	**Zero-bond**	0.00%	11.08%
25%-75% stock-bond	0.00%	27.18%	**100% bond**	0.58%	18.57%
Moderate	0.00%	26.38%	**10%-90% stock-bond**	0.75%	22.20%
Conservative	0.00%	25.60%	**Conservative**	1.38%	25.60%
100% bond	0.00%	18.75%	**Moderate**	1.53%	26.38%
Naive	0.01%	29.82%	**30%-70% stock-bond**	2.41%	29.94%
40%-60% stock-bond	0.04%	34.00%	**Naive**	2.42%	29.82%
100-x rule	0.05%	34.71%	**100-x rule**	3.77%	34.71%
Aggressive	0.57%	48.58%	**50%-50% stock-bond**	4.76%	36.63%
80%-20% stock-bond	0.97%	50.00%	**70%-30% stock-bond**	7.03%	46.68%
100% stock	1.79%	62.14%	**Aggressive**	7.13%	48.58%
			100% stock	10.12%	62.14%
g=0%, T=30			g=4%, T=30		
Strategy	**Norm. cost**	**Norm. profit**	**Strategy**	**Norm. cost**	**Norm. profit**
Zero-bond	0.00%	86.66%	**Zero-bond**	0.00%	32.20%
Moderate	0.00%	63.66%	**100% bond**	0.30%	33.44%
Naive	0.00%	63.17%	**20%-80% stock-bond**	0.95%	44.96%
30%-70% stock-bond	0.00%	55.07%	**Conservative**	1.49%	52.64%
Conservative	0.00%	52.64%	**30%-70% stock-bond**	1.85%	55.07%
100% bond	0.00%	33.44%	**Naive**	2.75%	63.17%
100-x rule	0.01%	67.71%	**Moderate**	2.78%	63.66%
50%-50% stock-bond	0.04%	68.51%	**100-x rule**	3.52%	67.71%
60%-40% stock-bond	0.11%	80.03%	**50%-50% stock-bond**	4.20%	68.51%
Aggressive	0.38%	104.53%	**75%-25% stock-bond**	7.17%	95.91%
100% stock	0.91%	123.74%	**Aggressive**	7.76%	104.53%
			100% stock	9.72%	123.74%

Note:
This table depicts the ranking of investment strategies according to the normalized guarantee cost (under the Esscher measure) and the normalized expected return for the MS(1-2) model and periodic contributions (€1200 up-front annually). The top part represents the a 10-year contract, the middle part the 20-year contract, and the bottom part the 30-year contract. The left section depicts the results for the low level guarantee (g = 0% p.a.), and the right section for the high level guarantee (g = 4% p.a.), respectively.
* – for contracts with a 10-year term, the moderate life-cycle strategy invests in the 10%-90% stock-bond portfolio. ** – for contracts with a 10-year term, the naive life-cycle strategy invests in the 25%-75% stock-bond portfolio. *** – for contracts with 10-year term, the 100-x life-cycle strategy invests in the 40%-60% stock-bond portfolio.

can lead to the same guarantee cost, but to very different expected portfolio wealth. For instance, the pure bond strategy and the zero-bond strategy are both costless but have a normalized expected profit of 33.44% and 86.66%, respectively. Furthermore, the zero-bond strategy is better than the 60%-40% stock-bond buy-and-hold portfolio, which has a higher normalized guarantee cost (0.11%) but a lower normalized profit (80.03%).

Further, note that if the same client accepts a normalized guarantee cost of 0.91%, he can purchase a pure stock portfolio with the normalized expected profit of 123.74%, which is significantly higher than the profit of the zero-bond strategy. This example demonstrates one further regularity. The guarantee cost of a given investment strategy is twofold: the explicit cost the customer has to pay (put price) and the opportunity cost (lower expected profit in comparison to the reference strategy, e.g the pure stock investment strategy). Thus, instead of buying the zero-bond strategy it would be advisable to the client to pay the normalized price of 0.91% and to buy the product with a pure stock strategy. The advantage of this strategy would lie in a much higher normalized expected profit, which would increase from 86.66% to 123.74%.

Assume now that the client maintains the accepted cost level but raises the protection level from 0% to 4%. If his objective was to buy a costless guarantee, he would still buy the zero-bond strategy. His normalized expected profit, however, would decrease from 86.66% (for a 0% guarantee) to 32.20% (for a 4% guarantee). The reason for this is that the portfolio manager has to invest a higher asset proportion into zero-bonds in order to meet the higher guarantee level. If the client was ready to accept the normalized cost of up to 1%, he would invest in the 20%-80% stock-bond portfolio. If we compare this option to a pure stock portfolio with a guarantee of 0%, we can state that the cost remains at the same level (0.95% in comparison to 0.91% in the previous example). However, the normalized expected cost is 44.96%, which is significantly lower than the 123.74% from the previous example. This example shows, once again, that the choice of the guarantee level is not only associated with the explicit guarantee cost, but also with the opportunity cost of expected profit.

Finally, we focus our attention on the zero-bond strategy. For the 0% guarantee, the attractiveness of the zero-bond strategy increases with an increasing contract term. For a 10-year contract, the zero-bond strategy is only more attractive than one other strategy (the pure bond portfolio). For the 20-year contract, it is more attractive than eight strategies (the conservative, naive, and moderate strategy and all buy-and-hold strategies a the stock proportion of up to 30%). For the 30-year contract, it is more attractive than thirteen strategies (all buy-and-hold strategies with a stock portfolio of up to 50%, and all life-cycle strategies with the excption of the aggressive one). This is due to the fact that with an increasing contract term, less capital is needed to assure that the guarantee is fulfilled (i.e. investment in the zero-bond) and more capital can be invested in risky assets (i.e. equities).

In the 4% guarantee case, the zero-bond strategy is the only costless one, regardless of the investment horizon. However, it is also the one with the lowest normalized expected profit. This makes it less attractive in the eyes of the client. Instead, he should accept a higher explicit guarantee cost, which would significantly lower the opportunity cost (i.e. it would lead to an increase of the expected profit).

4.8 Conclusion

This chapter analyzes seven factors which influence the cost of an investment guarantee: the initial state (i.e., the market phase at the inception of the contract), the model governing the stochastic behavior of the investment portfolio, the guarantee level, the investment strategy, time, the contract term, and the contribution payment scheme.

We compare the cost of the guarantee within three option pricing models: the Black-Scholes, the Bollen-Hardy model, and the Webb model. The cost is the highest in the Black-Scholes and the lowest in the Webb approach. The Bollen-Hardy price lies in between the above two. This result may seem surprising at first glance. However, one has to keep in mind that the GBM, which is used in the Black-Scholes model, does not differentiate between the low and the high volatility phase (i.e. low and high market risk phase). A glance at the average regime duration (see Tables B.1-B.13 in Appendix B) or smoothed probabilities (see Figures D.1-D.39 in Appendix D), however, shows that the stable market phases are longer than the high volatility phases. As the GBM does not differentiate between the low and the high risk market phase, it only provides an average volatility. Furthermore, it does not account for the fact that the high risk market phases only occur rarely, thus leading to an overestimation of their impact on the guarantee cost. Thus, we can state that not taking into account the stochastic volatility risk leads to an overpricing of the guarantee. This effect is lower, *ceteris paribus*, (1) the lower the contract term and/or (2) the lower the stock proportion in the investment strategy. The effect of neglecting the switching risk (i.e., uncertainty if the market is in the stable or in the turbulent phase) leads to more crucial overpricement of the guarantee. Furthermore, this effect grows as (1) the contract term and/or (2) the stock proportion in the strategy increase.

This is important in the view of the fact that the retirement products we discuss have a long-term nature.

The guarantee cost under the Markov switching model is very sensitive to the probability of the initial state for short-term contracts. This effect is less significant for middle- and long-term contracts. The guarantee cost is higher when the product is sold in the high volatility market phase. Unfortunately, the market state cannot be observed. However, smoothed probabilities are very good proxies for the initial state probabilities. Additionally, there are two ways the risk-averse agent can manage the risk associated with the uncertainty about the initial state. First, he can restrict himself to only selling guarantees with contract terms above 10 years. This does not pose a problem, as retirement saving products are generally middle- and long-term products, which are predominantly sold on the market anyway. Second, the upper bound of the guarantee cost is always associated with a high volatility regime at the contract inception. The guarantee provider can therefore assume this to be the case and use the results as the conservative guarantee price.

Another cost factor we studied was the contribution payment scheme. We found that the impact of this factor varies strongly with respect to other cost factors. Generally, if the cost is high, then the single contribution scheme yields a higher guarantee cost than the periodic scheme. If the cost is low, the opposite holds true.

Three further factors which influence the guarantee cost, namely the guarantee level, the stock proportion in the investment strategy, and time/contract term are tightly connected with each other. Accordingly, manipulating one of the above three factors can achieve the same amount of the cost reduction. Thus, they should always be considered together. The impact of the guarantee level and the stock proportion is always the same: the higher the guarantee level and/or the higher the stock proportion, the higher the cost of the guarantee. Furthermore, guarantee levels significantly lower than the risk-free interest rate (e.g. $g = -2\%, 0\%$ and 2% p.a.) have a different impact on the guarantee cost and result in a different shape of the cost curve than guarantee levels closer to the risk-free interest rate (e.g. $g = 4\%$ p.a.).

In contrast, the impact of time/contract term on the guarantee cost differs from that of the guarantee level and the stock proportion in the investment portfolio. In the case of low level guarantees and in the case of high level guarantees with a low stock proportion the guarantee cost decreases as time/contract term increases. On the contrary, for high level guarantees with a high stock proportion, the cost function of time/contract term has an inverted U shape.

In conclusion, the cost of low level guarantees can be reduced to an acceptable level for middle and high contract terms, regardless of the stock proportion in the investment portfolio. The cost of high level guarantees is acceptable only if the stock proportion is low. Since the guarantee cost is rather high for high stock proportions, clients may not purchase such contracts. Instead of selling guarantees backed by investment strategies with a high stock proportion, the guarantee provider would be advised to sell guarantees backed by a life-cycle strategy. The idea of a life-cycle strategy is to invest in more risky assets at the contract inception and to reduce the stock proportion as the contract nears its expiration date. This approach enables the guarantee provider to construct a product which well meets the expectations of his client. This means that, for a given contract term, guarantee level, and guarantee cost, the guarantee provider can find a life-cycle strategy which maximizes the expected profit of the individual pension account.

Apart from the buy-and-hold and life-cycle strategies, the zero-bond strategy also constitutes a very interesting option for the guarantee provider. The main idea is to invest the portion of the contribution needed to fulfill the guarantee in a risk-free zero-bond, and the remaining capital in stocks. This leads to a costless product. This strategy is very interesting for the guarantee provider, as it does not require any solvency capital (under the assumption that the seller of the zero-bond cannot default). A very interesting result is that, for low level guarantees, the expected profit of the zero-bond strategy outperforms the expected profit of several buy-and-hold and life-cycle strategies with a positive cost. This does not hold true for high level guarantees when the zero-bond strategy is the only costless one. Thus, the guarantee provider has the choice between selling low level guarantees backed with a zero-bond and selling those backed with one of the life-cycle strategies with a positive – but still acceptable – cost and an expected profit which is higher than the expected profit of the zero-bond strategy.

If the guarantee provider would like to sell high level guarantees backed with a zero-bond strategy, he should be aware of the fact that, while being a costless product, it also yields a very low expected profit. Thus, it is doubtful whether buying the individual pension product with high level guarantee backed by the zero-bond strategy really is in the best interest of the client. The guarantee provider would thus be advised to back high level guarantees with a life-cycle strategy, instead.

Chapter 5

Risk analysis and solvency requirements

5.1 Risk definition

In a financial context, risk can be understood as an *ex ante* unknown quantity of the invested portfolio value at the end of the investment horizon. Or, in a broader sense, as the *ex ante* unknown distribution of this portfolio value. Alternatively, one can speak of the *ex ante* unknown return from the investment or of its distribution as these two concepts (value and return) are very closely related with each other.

Fishburn (1984, p. 397) characterizes risk as being *based in part on outcome preferences and targets. (...) risk increases as bad outcomes become more probable, and as probable bad outcomes get worse. While favorable outcomes are not associated with risk by themselves, their presence in a distribution that has positive probability for bad outcomes might decrease the risk of the distribution.*

According to that, this thesis defines risk as Albrecht's (2004, p. 1493) risk of the first type, i.e., *as the magnitude of deviations from a target.* To quantify this risk we will use several risk measures, which are defined as follows.

Definition 5.1 (Risk measure) *Let* $(\Omega, \mathcal{F}, \mathcal{P})$ *be a probability space and let* $\mathscr{Y}$ *be a family of random variables on this probability space, then functional* $\mathcal{R} : \mathscr{Y} \to \mathbb{R}$ *is called a risk measure (Fischer 2003, p. 136).*

This means that there is a rule (risk measure) which allows us to assign a real number to each random variable (in our case a risky asset) from family $\mathscr{Y}$. This real number describes the risk level (or "riskiness") of a particular risky asset which enables us to compare the risk level of different risk strategies.

This Chapter analyzes the risk and solvency requirements associated with investment guarantees. It is constructed as follows. Section 5.2 discusses several risk measures and their suitability to measure the risk of an investment guarantee. It proposes using lower partial moments and conditional lower partial moments for this purpose. Section 5.3 discusses the design of the study which is similar to that from the previous Chapter. Section 5.4 analyzes the risk of an investment guarantee from the point of view of the guarantee provider. It discusses the impact of several risk factors on the risk of an investment guarantee. Section 5.5 extends the discussion by adding the point of view of the solvency supervising authority and Section 5.6 adds that of the potential buyer of the guarantee. It proposes to use the mean excess loss to quantify the required solvency capital. Section 5.7 concludes the results of this Chapter.

5.2 Risk measure

5.2.1 Dispersion risk measures

Markowitz (1952) was the first to quantify financial risk. In his groundbreaking *Portfolio selection* paper he measured the risk by means of standard deviation. Alternatively, one can also use variance to quantify financial risk as Sharpe (1964) did in his Capital Asset Pricing Model. Standard deviation and variance can be defined as follows:

Definition 5.2 (Standard deviation and variance) *Let X be a random variable with the probability function F and the expected value μ, then*

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 dF(x)$$

is called variance of X. Additionally, $\sigma = \sqrt{\sigma^2}$ is called standard deviation of X.

Ever since Markowitz (1952) and Sharpe (1964), both these dispersion risk measures have commonly been used in financial literature. Their advantage lies in their understandability and computational straightforwardness. However, they have the disadvantage of equally weighting the chance (positive deviation from the mean) and the risk (negative deviation from the mean). This is unproblematic in the case of symmetric distributions, e.g., the normal distribution. Yet, in Section 3.1 it was shown that the returns of financial time series are not symmetrical. In order to deal with this drawback, one can use semi-dispersion measures:

Definition 5.3 (Semi-standard deviation and semi-variance) *Let X be a random variable with the probability function F and the expected value μ,* *then*

$$\sigma_s^2 = \int_{-\infty}^{\mu} (x-\mu)^2 dF(x)$$

is called semi-variance of X. Additionally, $\sigma_s = \sqrt{\sigma_s^2}$ is called semi-standard deviation of X.

Contrary to the standard deviation and variance, semi-dispersion risk measures only account for a negative deviation from the mean. Thus, they correspond better to our concept of risk than standard deviation and variance. They were, e.g., used by Hogan and Warren (1974) who proposed an equilibrium pricing model with semi-variance as a risk measure.

5.2.2 Quantile risk measures

Another approach is to use risk measures based on quantiles.

Definition 5.4 (Quantile) *Let X be a random variable with distribution function F, and let $0 \leq \alpha \leq 1$ be a constant, then Q_α such that*

$$F(Q_\alpha(X)) = \alpha$$

is called α quantile of X.

A commonly used quantile risk measure is the value at risk (VaR), introduced by JP Morgan (1996) in their RiskMetrics concept and recommended by the Basel Committee on Banking Supervision (2001) to quantify risk in the banking system.

Definition 5.5 (Value at risk (VaR)) *Let X be a random variable and let $0 \leq \alpha \leq 1$, then a $(1-\alpha)$-quantile of X, namely*

$$F(VaR_\alpha(X)) = 1 - \alpha = F(Q_{1-\alpha}(X)).$$

is a value at risk of X at the confidence level α. Thereby, Q denotes a quantile and F denotes the distribution function of X.

Remark 5.6 *Note that Definition 5.5 is the equivalent of*

$$\Pr(X > VaR_\alpha(X)) = \alpha$$

(Albrecht 2004, p. 1498).

Value at risk can be interpreted as the maximal loss with a given probability $100(1-\alpha)\%$. The confidence level α can be interpreted as the risk aversion cofficient. The lower the probabilty α, the more risk averse is the decision maker.

Value at risk is very controversial as a risk measure. Szegö (2002, p. 1261) points out several drawbacks of VaR: it does not measure losses exceeding VaR; it may yield contradictory results for different confidence levels; it is not sub-additive, which means that diversification may increase the risk; it is not convex, which implies that it is problematic to use it for optimization problems; furthermore it has many local maxima, which may lead to unstable risk rankings. Acerbi and Tasche (2002) and Szegö (2002), among others, discuss several other quantile risk measures dealing with these drawbacks.

5.2.3 Shortfall risk measures

Both above mentioned risk measure families cannot be used to quantify the risk of falling below a target portfolio wealth (or a target return). *Shortfall risk measures* address precisely this problem. They are based on the idea that any given financial random variable can be divided into a risk and a chance component, according to the given target $\bar{z}$ (e.g. guaranteed portfolio value):

$$X = \underbrace{-\max(\bar{z} - X; 0)}_{\text{risk}} + \bar{z} + \underbrace{\max(X - \bar{z}; 0)}_{\text{chance}}, \tag{5.1}$$

where $-\max(\bar{z} - X; 0)$ represents the risk and $\max(X - \bar{z}; 0)$ the chance potential, respectively. Define now the loss function, such that $L = \max(\bar{z} - x; 0)$, then the shortfall risk measure can be defined as follows.

Definition 5.7 (Shortfall risk measure) *Let X be a random variable with the distribution function F, let $\bar{z}$ be a target, and let $L : \mathbb{R} \to \mathbb{R}_+ \cup \{0\}$ be a continuous, monotonously decreasing function, then*

$$SRM_{\bar{z}}(X) = \int_{-\infty}^{\bar{z}} L(\bar{z} - x) dF(x)$$

is called the shortfall risk measure. The function L is called the loss function (Albrecht 1994, p. 92).

Remark 5.8 *Very often in literature, the loss function is defined as $L(x) = x^n$ with $n \geq 0$. Accordingly, the shortfall risk measure $SRM_{\bar{z}}(X)$ is quantified by $LPM_{\bar{z}}^n(X)$, i.e. the lower partial moment of the n-th order and the target $\bar{z}$ (Albrecht and Klett 2004, p. 3).*

The most often used shortfall risk is the lower partial moment.

Definition 5.9 (Lower partial moment) *Let X be a random variable with distribution function F, let $n \geq 0$ and $\bar{z} \in \mathbb{R}$ be constants, then*

$$LPM_{\bar{z}}^n(X) = \int_{-\infty}^{\bar{z}} (\bar{z} - x)^n dF(x)$$

is called the lower partial moment[1] of X of the order n and the target $\bar{z}$ (Albrecht and Klett 2004, p. 3).

[1] Some authors define the lower partial moment as $\overline{LPM}_{\bar{z}}^n(X) = \sqrt[n]{\int_{-\infty}^{\bar{z}} (\bar{z} - x)^n dF(x)}$, see Albrecht (2004, p. 1497).

Example 5.10 *The constant n can be interpreted as the risk aversion measure of the financial agent. It can take on any non-negative value, however, $n \in \mathbb{N}$ leads to standard cases. E.g., for $n = 0$, it is a shortfall probability*

$$LPM_{\bar{z}}^{0}(X) = \int_{-\infty}^{\bar{z}} dF(x) = F(\bar{z}) =: \mathcal{P}_{\bar{z}}^{\mathbb{S}}(X), \tag{5.2}$$

which can be interpreted as the probability of falling below the target. For $n = 1$, it is a shortfall expected value

$$LPM_{\bar{z}}^{1}(X) = \int_{-\infty}^{\bar{z}} (\bar{z} - x) dF(x) =: \mathbb{E}_{\bar{z}}^{\mathbb{S}}[X], \tag{5.3}$$

which can be interpreted as the expected severity of loss. Other examples of lower partial moments are: the shortfall variance, for $n = 2$,

$$LPM_{\bar{z}}^{2}(X) = \int_{-\infty}^{\bar{z}} (\bar{z} - x)^2 dF(x) =: \mathbb{V}ar_{\bar{z}}^{\mathbb{S}}[X] \tag{5.4}$$

and the shortfall standard deviation

$$\overline{LPM}_{\bar{z}}^{2}(X) = \sqrt{\int_{-\infty}^{\bar{z}} (\bar{z} - x)^2 dF(x)} =: \sigma_{\bar{z}}^{\mathbb{S}}[X]. \tag{5.5}$$

In analogy to the shortfall risk measure, a worst case measure can be constructed. It is based on conditional lower partial moments.

Definition 5.11 (Conditional lower partial moment) *Let X be a random variable with distribution function F, let $n \geq 0$ and $\bar{z} \in \mathbb{R}$ be constants, then*

$$CLPM_{\bar{z}}^{n}(X) = \mathbb{E}[\max(\bar{z} - X; 0)^n | X \leq \bar{z}] = \mathbb{E}[(\bar{z} - X)^n | X \leq \bar{z}]$$

is called a conditional lower partial moment of X of the order n and the target $\bar{z}$ (Albrecht and Klett 2004, p. 4).

Remark 5.12 *Note that the partial moment equals the product of the conditional lower partial moment and the shortfall probability.*

$$\begin{aligned}LPM_{\bar{z}}^{n}(X) &= \int_{-\infty}^{\bar{z}} (\bar{z}-x)^n dF(x) = F(\bar{z}) \int_{-\infty}^{\infty} (\bar{z}-x)^n dF(x) \\ &= \mathcal{P}_{\bar{z}}^{\mathbb{S}}(X) CLPM_{\bar{z}}^{n}(X)\end{aligned}$$

(Albrecht and Klett 2004, p. 5).

Example 5.13 *In analogy to the case of lower partial moments, n is a risk aversion coefficient which can take on all non-negative values. The best known example of a conditional lower partial moment is the mean excess loss (MEL), for $n = 1$,*

$$CLPM_{\bar{z}}^{1}(X) = \mathbb{E}[\bar{z}-X|X \leq \bar{z}] = \frac{1}{F(\bar{z})} \int_{-\infty}^{\bar{z}} (\bar{z}-x) dF(x) =: MEL_{\bar{z}}[X], \tag{5.6}$$

which can be interpreted as the expected loss when the loss occurs. Other examples are the conditional shortfall variance, for $n = 2$,

$$CLPM_{\bar{z}}^{2}(X) = \mathbb{E}[(\bar{z}-X)^2|X \leq \bar{z}] = \frac{1}{F(\bar{z})} \int_{-\infty}^{\bar{z}} (\bar{z}-x)^2 dF(x) =: \mathbb{V}ar_{\bar{z}}^{\mathbb{CS}}[X] \tag{5.7}$$

and the conditional shortfall standard deviation

$$\overline{CLPM}_{\bar{z}}^{2}(X) = \sqrt{\mathbb{E}[(\bar{z}-X)^2|X \leq \bar{z}]} = \sqrt{\frac{1}{F(\bar{z})} \int_{-\infty}^{\bar{z}} (\bar{z}-x)^2 dF(x)} =: \sigma_{\bar{z}}^{\mathbb{CS}}[X]. \tag{5.8}$$

In the context of investment guarantees embedded in personal pension plans, the risk is defined as the possibility that the value of the investment portfolio which backs the guarantee falls below the guaranteed amount (or, equivalently, that the realized return from the backing portfolio is lower than the guaranteed return). Thus, from all of the above discussed risk measures, the lower partial moments and the conditional partial moments are the best to quantify this risk, as they address the investment target explicitly. It is useful to analyze several of them rather than selecting only one. The shortfall probability can be interpreted as the probability of bankruptcy of the guarantee provider (under the assumption that he does not have any reserves). The shortfall expected value quantifies the expected average loss of the guarantee provider. The shortfall variance (standard deviation) measures the variance (standard deviation) of the loss.

A very interesting probability risk measure is the mean excess loss, since it exhibits a worst case measure: it discloses how much on average the guarantee provider additionally has to pay to the guarantee buyer if the value of the guaranteed portfolio falls below the guaranteed amount. Therefore, in our opinion, this measure is very well suited to quantify the solvency requirements (or, equivalently, the reserves the guarantee provider should hold in order to avoid his bankruptcy). The last risk measure we discuss in this dissertation is the conditional shortfall variance (standard deviation), which quantifies the variance (standard deviation) of the loss when the shortfall occurs.

5.3 Design of the study

Before we present the results of the study on the risk associated with the return guarantee, the simulation design should be explained. The simulation is based on the one described in Section 4.7.1.3. However, there are three major differences: First, the stochastic variables are simulated under the real world probability measure $\mathcal{P}$. Second, the risk measures are simulated with 20.000.000 iterations. The rise of the iteration number is necessary to achieve the convergence of several risk measures. In the case of the shortfall probability, the shortfall expected value, and the shortfall standard deviation, this goal was attained. In the case of the mean excess loss and the conditional shortfall standard deviation, this only proved successful for investment strategies with a high and a middle stock proportion. For strategies with a low stock engagement, the results are unstable. However, it is not clear how high the number of iterations should be for the Monte Carlo method to deliver robust numerical results. As the numerical approach with 20 million iterations was very time consuming, we decided not to increase the iteration number. Thus, for the mean excess loss and the conditional shortfall standard deviation, the results will be discussed only with regard to investment strategies with a high stock proportion. We believe, however, that results for other cases are qualitatively similar, even though we do not receive any numerical results.

Third, as different contracts have different cash-flows and different guarantee values, we need to standardize the risk measures to render results comparable. Let $LPM_g^n(V_T)$ and $CLPM_g^n(V_T)$ be the lower partial moments and the conditional lower partial moments of portfolio value V at the time point T and the target portfolio value G_T defined in equation (4.30) (guaranteed value) with the associated guarantee level g. Accordingly, the *normalized shortfall expected value* is defined as

$$\widetilde{LPM}_g^1(V_T) = \frac{e^{-r(T-t_0)} LPM_g^1(V_T)}{\sum_{t_n=0}^{T-1} C_{t_n} e^{-r(t_n-t_0)}}, \tag{5.9}$$

the *normalized shortfall standard deviation* as

$$\widetilde{\overline{LPM}}{}_g^2(V_T) = \frac{e^{-r(T-t_0)} \overline{LPM_g^2(V_T)}}{\sum_{t_n=0}^{T-1} C_{t_n} e^{-r(t_n-t_0)}}, \tag{5.10}$$

the *normalized mean excess loss* as

$$\widetilde{MEL}(V_T) = \widetilde{CLPM}_g^1(V_T) = \frac{e^{-r(T-t_0)} CLPM_g^1(V_T)}{\sum_{t_n=0}^{T-1} C_{t_n} e^{-r(t_n-t_0)}}, \tag{5.11}$$

and the *normalized conditional shortfall standard deviation* as

$$\widetilde{\overline{CLPM_g^2}}(V_T) = \frac{e^{-r(T-t_0)}\overline{CLPM_g^2(V_T)}}{\sum_{t_n=0}^{T-1} C_{t_n} e^{-r(t_n-t_0)}}, \tag{5.12}$$

with C_{t_n} contribution payed in time $t_n \in [0, T]$.

5.4 Risk analysis

5.4.1 Impact of the initial state on the guarantee risk

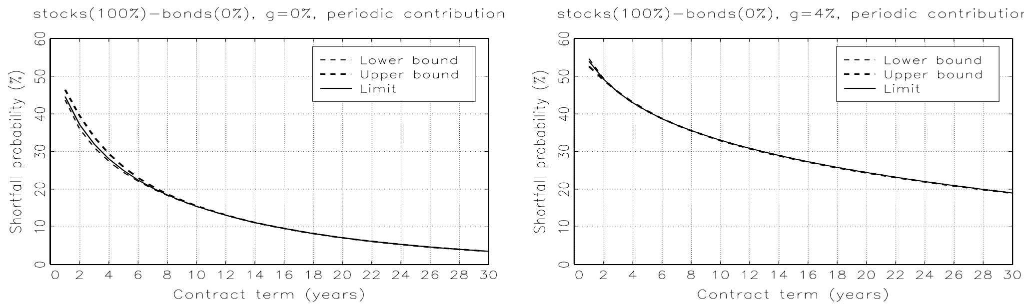
Figures 5.1-5.2 show the impact of the initial state on the risk measures. The thin dashed line represents risk measures for the agent who knows with certainty that the market is in the low volatility state at the beginning of the contract ($\Pr[Z_{t_0} = 1] = 1$), and the thick dashed line depicts risk measures for the agent who knows with certainty that the market is in the high volatility regime at the contract inception ($\Pr[Z_{t_0} = 2] = 1$). The solid line represents risk measures for the agent who does not know which state the process was in at the beginning of the contract. Therefore, he assumes that the process was in the low volatility regime with its ergodic probability (see equation (2.69)). I.e., he assumes: $\Pr[Z_{t_0} = 1] = \pi_1$ and $\Pr[Z_{t_0} = 2] = \pi_2$.

The first two cases represent the lower and the upper bound of the risk measure in the Markov switching model. From Remark 2.34 we know that the limit transition probabilities of the homogeneous Markov chain (the state variable Z_{t_n}) is the ergodic probability, as time goes to infinity. Thus, the case of the uninformed agent is, in fact, the limit of the risk measure.

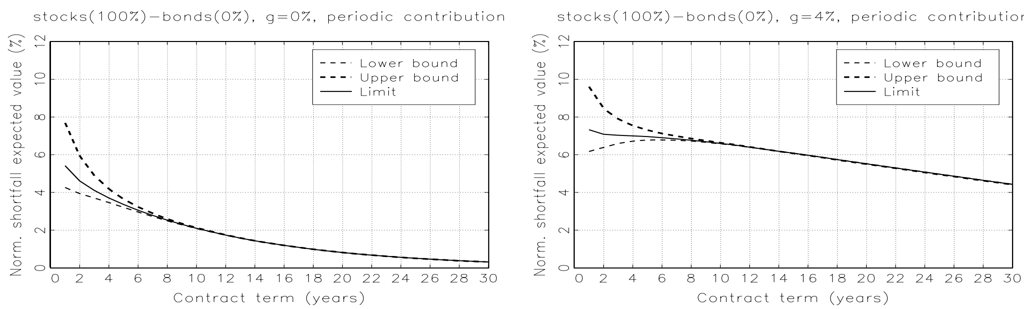
The impact of the initial state on the risk measures is similar to the impact on the guarantee cost. The exception is the shortfall probability, which will be described later. For other risk measures, the spread between the lower and upper bound is greater, *ceteris paribus*, the shorter the contract term,

Figure 5.1: Impact of the initial state on risk measures (periodic contribution)

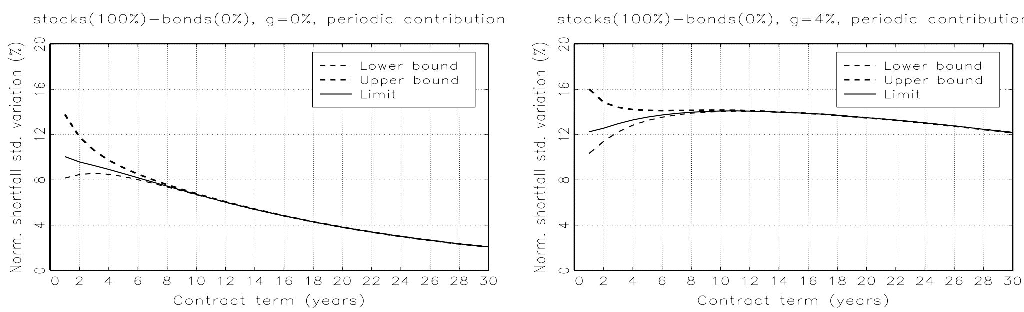
Shortfall probability



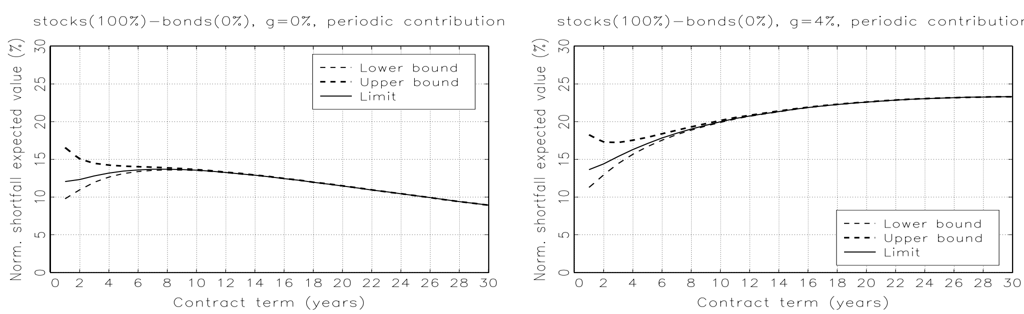
Normalized shortfall expected value



Normalized shortfall standard deviation



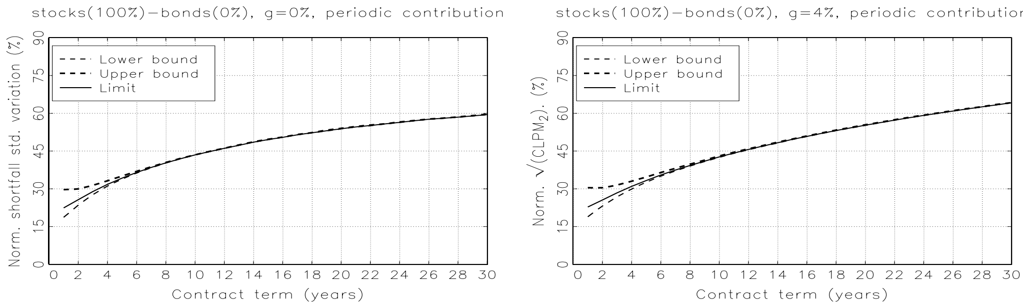
Normalized mean excess loss



continued on the next page

continued from the previous page

Normalized conditional shortfall standard deviation



Note:

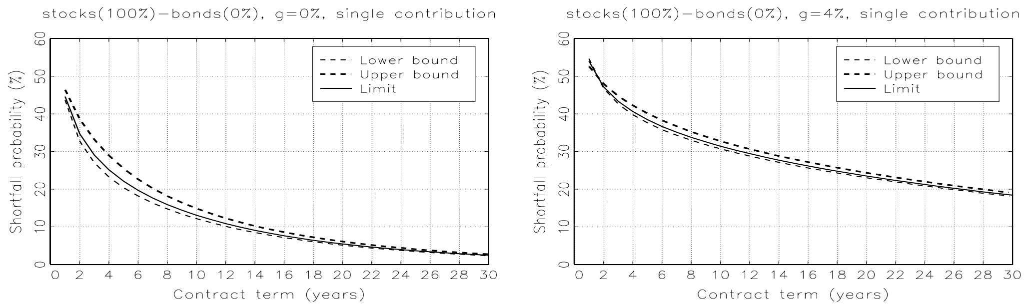
This figure depicts the impact of the initial state on the guarantee shortfall risk by means of the pure stock portfolio with periodic contributions (€1200 up-front annually) and the MS(1-2) model. The thin dashed line represents the lower bound ($\Pr[Z_{t_0} = 1] = 1$) and the bold dashed line the upper bound ($\Pr[Z_{t_0} = 2] = 1$) of the risk measure, respectively. The solid line represents the limit of the risk measure ($\Pr[Z_{t_0} = 1] = \pi_1$). The left column displays the low level guarantee ($g = 0\%$ p.a.) and the right column the high level guarantee ($g = 4\%$ p.a.), respectively.

the higher the guarantee level, and/or the higher the stock proportion in the investment strategy. Furthermore, contracts with single contributions are more sensitive to the initial state than contracts with a periodic payment scheme. The detailed discussion can thus be omitted, as it is similar to Section 4.7.2.1. In this section we will restrict our discussion to the behavior of the shortfall probability which is different.

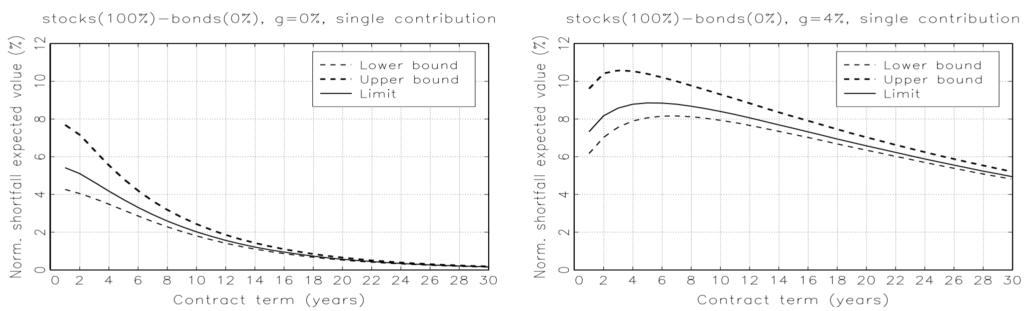
In the case of short term contracts, the impact of the initial state on the shortfall probability is greater, *ceteris paribus*, the lower the stock proportion in the investment strategy and/or the lower the guarantee level. At first glance, this is contra-intuitive, as one would assume that the higher both risk factors, the higher the risk, and the higher the sensitivity to the initial state. In fact, both of these risk factors increase the probability of shortfall. For instance, a one-year contract with a 0% guarantee backed with a pure stock contract has a shortfall probability of 43.66% in the low volatility initial state and 46.42% in the high volatility initial state. Therefore, the shortfall

Figure 5.2: Impact of the initial state on risk measures (single contribution)

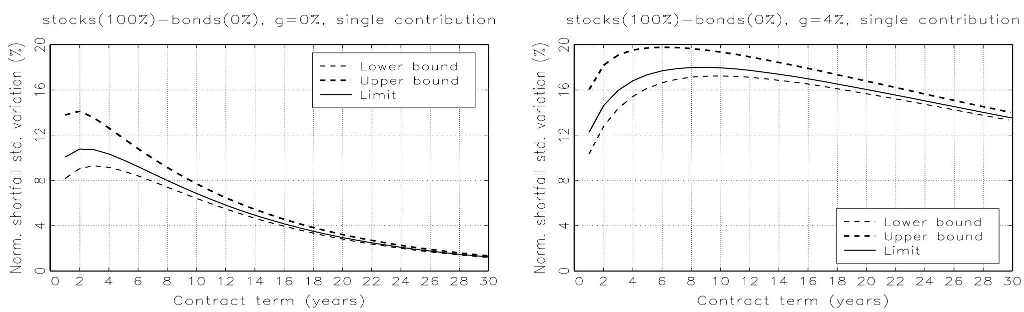
Shortfall probability



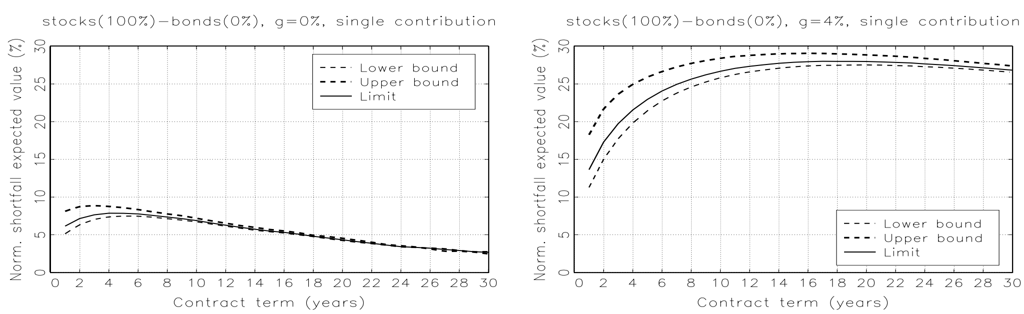
Normalized shortfall expected value



Normalized shortfall standard deviation



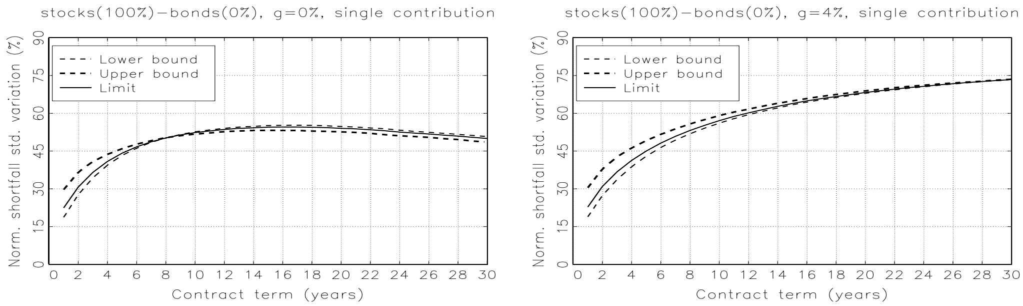
Normalized mean excess loss



continued on the next page

continued from the previous page

Normalized conditional shortfall standard deviation



Note:
This figure depicts the impact of the initial state on the guarantee shortfall risk by means of the pure stock portfolio with a single contribution (equal to the net present value of periodic contributions) and the MS(1-2) model. The thin dashed line represents the lower bound ($\Pr[Z_{t_0} = 1] = 1$) and the bold dashed line the upper bound ($\Pr[Z_{t_0} = 2] = 1$) of the risk measure, respectively. The solid line represents the limit of the risk measure ($\Pr[Z_{t_0} = 1] = \pi_1$). The left column displays the low level guarantee ($g = 0\%$ p.a.) and the right column the high level guarantee ($g = 4\%$ p.a.), respectively.

probability is very high independnetly on the initial state. For a pure bond portfolio, the shortfall probability equals 13.06% for the low volatility initial state and 25.47% for the high volatility initial state. In this case, the initial state has a significant influence on the shortfall probability, which explains why the sensitivity of the shortfall probability to the initial state is higher when the risk is low.

The impact of the remaining risk factors is the same as in the case of other risk measures, i.e. the spread between the upper and the lower bound converges to zero (or equivalently, the shortfall probability converges to the limit) faster, the lower the guarantee level and/or the lower the stock proportion in the investment portfolio. It also converges faster for the periodic than for the single contribution scheme.

We conclude that the choice of the initial state is crucial for risk measures, especially for short time contracts. Unfortunately, the probability of the initial state cannot be observed. In Section 4.7.2.1 we discuss five ways to deal with this issue. Hereafter, we will use the smoothed probabilities from the EM algorithm (see Table 4.2) as a proxy for initial state probabilities.

5.4.2 Impact of the stochastic model on the guarantee risk

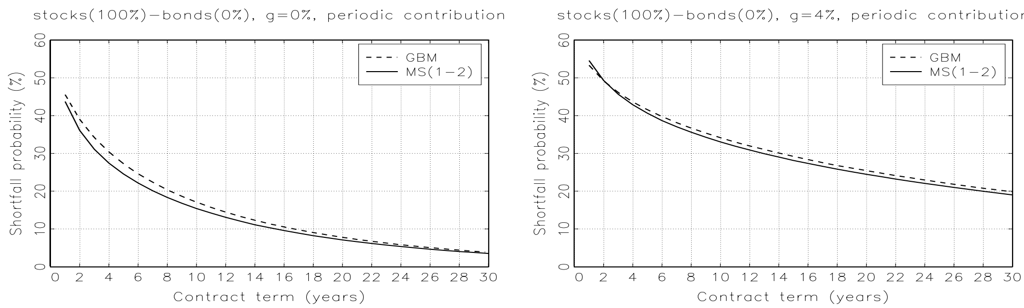
Figure 5.3 shows the impact of the stochastic model on shortfall measures. The dashed line represents the result for the geometric Brownian motion, and the solid line the result for the geometric Brownian motion with Markov switching (model MS(1-2)). The shortfall probability and the normalized shortfall expected value is higher in the GBM model. As long as these risk measures approach zero, the results of both models converge to each other.

The normalized shortfall standard deviation is higher in the GBM model for contracts with a high and middle stock proportion in the investment strategy. However, the difference between both models becomes smaller with growing contract duration. For models with a low stock proportion in the investment strategy, the shortfall standard deviation is lower in the case of short-time contracts. The shortfall standard deviation converges to zero faster in the GBM model. Thus, for the middle term contracts, this risk measure is higher in the Markov switching model. This phenomenon is not observed for contracts with a high and middle stock proportion. In these cases, however, the shortfall standard deviation does not converge to zero – even for the 30-year contract. However, it can be assumed that this phenomenon occurs later, since the difference between both models diminishes as the contract duration increases.

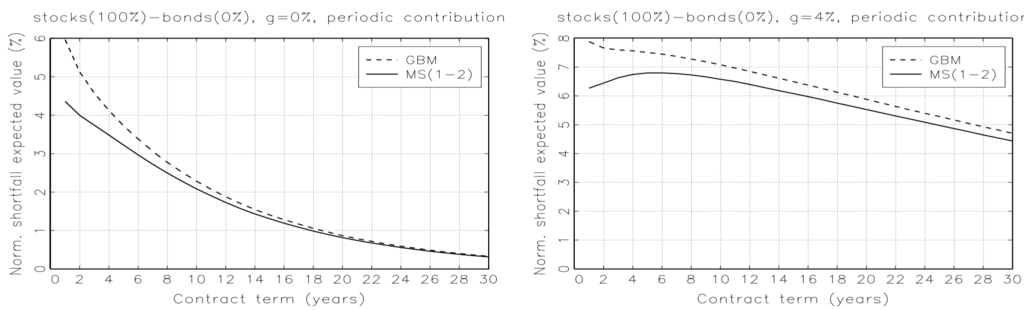
Both conditional risk measures, the normalized mean excess loss and the normalized conditional shortfall standard deviation, behave similarly. For short term contracts, they are higher in the GBM. In such a case, both

Figure 5.3: Impact of the stochastic model on risk measures

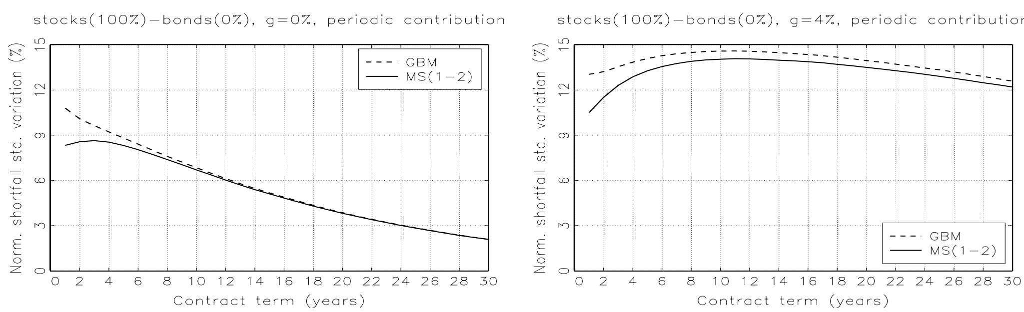
Shortfall probability



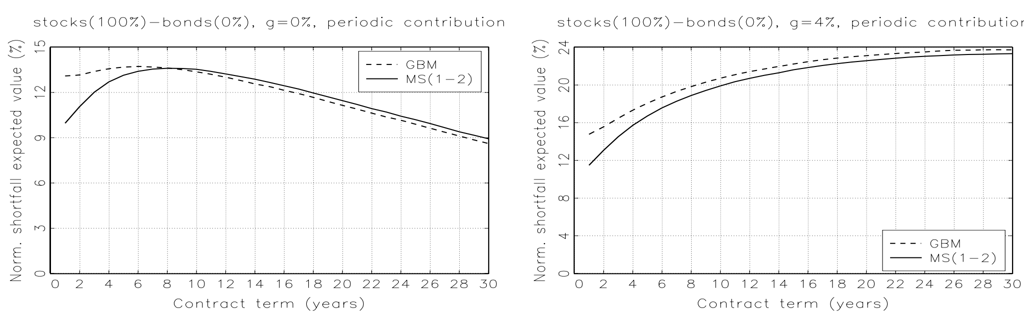
Normalized shortfall expected value



Normalized shortfall standard deviation



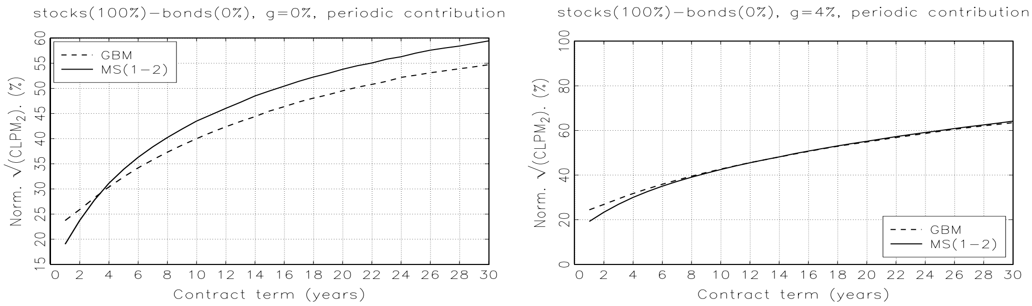
Normalized mean excess loss



continued on the next page

continued from the previous page

Normalized conditional shortfall standard deviation



Note:
This figure depicts the impact of the stochastic model on the guarantee shortfall risk using the example of the pure stock strategy with periodic contributions (€1200 up-front annually). The dashed line represents the GBM, and the solid line the MS(1-2) model. The left column depicts contracts with a low level guarantee ($g = 0\%$ p.a.), and the right column contracts with a high level guarantee ($g = 4\%$ p.a.), respectively.

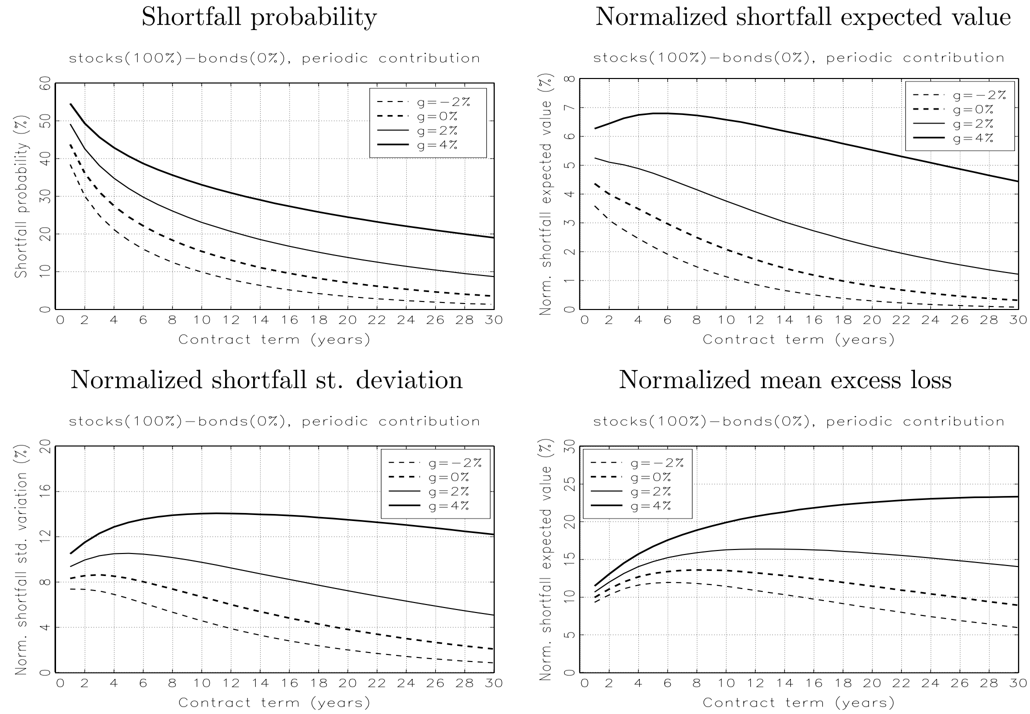
curves cross, and the conditional risk measures are higher in the Markov switching model. An exception is the normalized MEL for contracts with a high guarantee level and a high stock proportion. In these cases, conditional risk measures are greater in the GBM for all contract terms between 1 and 30 years. However, the spread between both models decreases, and it can thus be assumed that both curves will cross each other for contracts with higher durations.

5.4.3 Impact of the guarantee level on the guarantee risk

This section discusses the impact of the guarantee level on the guarantee risk. The most self-evident finding is that the risk increases with the increase of the guarantee level. This holds true for all risk measures (see Figure 5.4[2]). In

[2]Figure 5.4 shows results for the pure stock portfolio. Results for the fifty-fifty stock-bond portfolio and the pure bond strategy can be found in Figures K.1-K.2 in Appendix K.

Figure 5.4: Impact of the guarantee level on risk measures

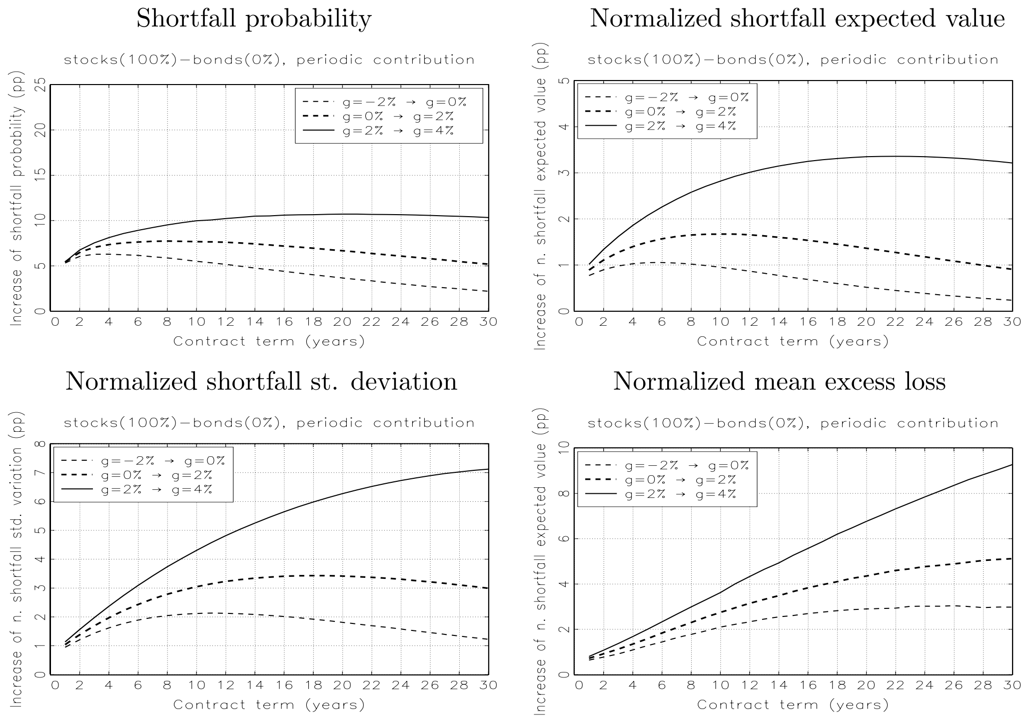


Note:
This figure depicts the impact of the guarantee level on the guarantee shortfall risk using the example of the pure stock strategy with periodic contributions (€1200 up-front annually). All risk measures are computed for the MS(1-2) model. The thin dashed line represents the guarantee of -2% p.a., the bold dashed line – 0% p.a., the thin solid line – 2% p.a., and the bold solid line – 4% p.a., respectively.

most cases, the shape of the risk curve is the same for all studied guarantee levels. Some exceptions occur for the normalized shortfall standard deviation and the normalized mean excess loss.

For portfolios with a stock proportion between 80% and 100%, the normalized shortfall expected value is a decreasing function of the contract term for low guarantee levels (i.e. $g = -2\%, 0\%, 2\%$). In the case of the high level guarantee, it is an increasing function for the short-term contracts. Consequently, it reaches a maximum and starts to fall. For other investment portfolios, the normalized shortfall expected value is a decreasing function of

Figure 5.5: Sensitivity of risk measures to changes in the guarantee level (pure stock portfolio)



Note:
This figure depicts the sensitivity of risk measures to changes of the guarantee level using the example of the pure stock strategy with periodic contributions (€1200 up-front annually). The sensitivity is measured in per cent points (pp). All risk strategies are computed for the MS(1-2) model. The thin dashed line represents the guarantee increase from -2% to 0% p.a., the bold dashed line – from 0% to 2% p.a., and the solid line – from 2% to 4% p.a., respectively.

the contract term, regardless of the guarantee level.

The the case of the normalized shortfall, standard deviation is slightly different. The shape of this curve is similar for all guarantee levels if the stock is either low or high. In the first case, the curve decreases, and in the second case, it increases to a maximum and decreases afterwards. For investment portfolios with a stock proportion between 20% and 90%, the normalized shortfall standard deviation is a decreasing function if the guarantee level is equal to -2% p.a.; and it has the inverted U shape if the guarantee level equals 4% p.a. The other two guarantees (i.e. $g = 0\%$ or 2% p.a.) behave as any one of the above two cases.

Figure 5.5[3] shows how the risk reacts if the guarantee level increases by 2 per cent points: from -2% to 0% (thin dashed line), from 0% to 2% (thick dashed line), and from 2% to 4% (solid line). The figure shows that the risk reacts overproportinally to the change in the guarantee level. In particular, the increase is the lowest if the guarantee level rises from -2% to 0%; and the highest if the guarantee level rises from 2% to 4%. The sensitivity of the risk to the guarantee level is much smaller in the two remaining cases (i.e., the increase from -2% to 0% and from 0% to 2%).

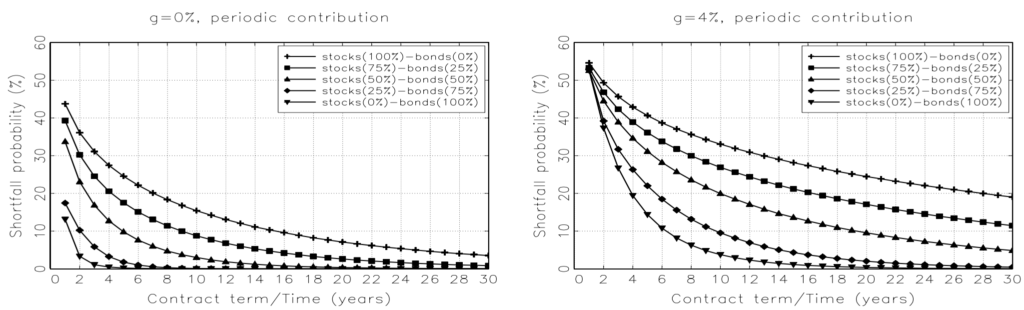
5.4.4 Impact of the investment strategy on the guarantee risk

This section discusses the impact of the investment strategy on the shortfall risk. One would intuitively assume that the risk associated with buy-and-hold strategies should decrease along with the decrease of the stock proportion in the portfolio. This holds true regardless of which risk measure is used to quantify the risk (see Figure 5.6). The risk of the short-term buy-and-hold strategies is always greater than zero. However, for some strategies, it can drop to zero when the contract term is higher. This is the case for investment strategies with a low level guarantee (i.e. $g = 0\%$ p.a.), and a stock proportion between 0% and 40% (the lower the stock proportion, the earlier the guarantee becomes riskless). For the 4% guarantee, only the normalized shortfall expected value allows to declare the 10%-90% stock-

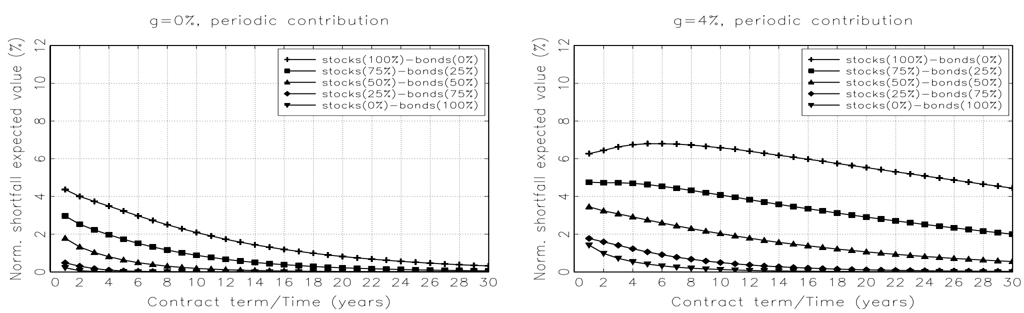
[3]Figure 5.4 shows the results for the pure stock portfolio. Results for the fifty-fifty stock-bond portfolio and the pure bond strategy can be found in Figures K.4-K.5 in Appendix K.

Figure 5.6: Impact of the investment strategy on the guarantee risk (buy-and-hold strategies)

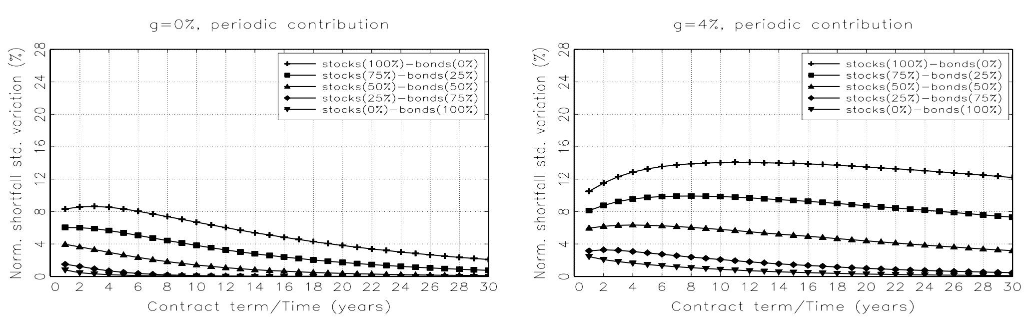
Shortfall probability



Normalized shortfall expected value



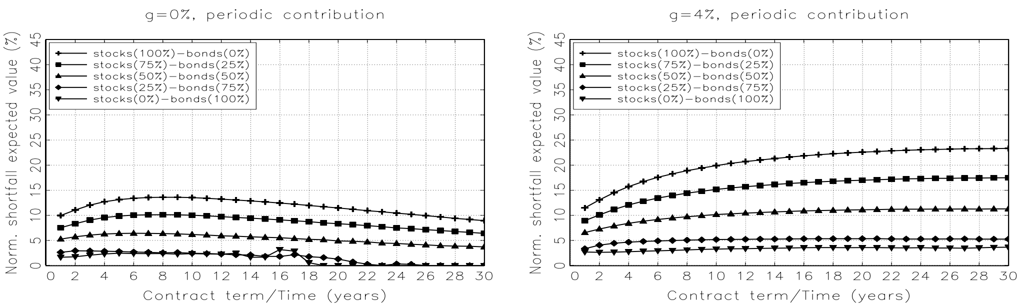
Normalized shortfall standard deviation



continued on the next page

continued from the previous page

Normalized mean excess loss



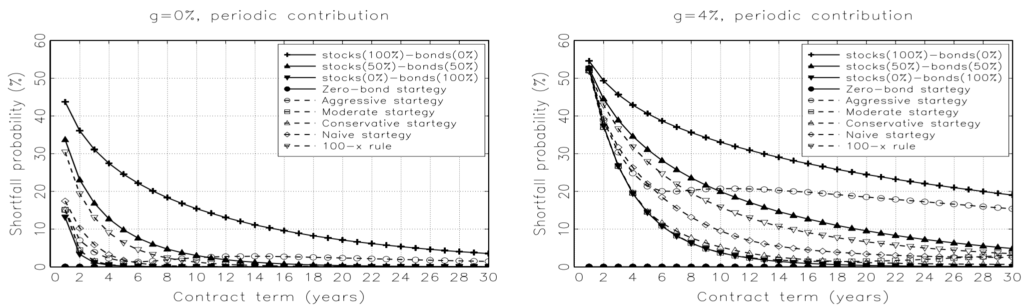
Note:
This figure depicts the impact of the investment strategy on the guarantee shortfall risk. All risk measures are computed for the MS(1-2) model. The solid line with pluses represents the pure stock investment strategy, the solid line with squares – the 75%-25% stock-bond, the solid line with triangles – the 50%-50% stock-bond, the solid line with diamonds – the 25%-75% stock-bond, the solid line with inverse triangles – the pure bond strategy, respectively. In all cases, contributions of €1200 are paid up-front annually. The left column represents the low level guarantee ($g = 0\%$ p.a.) and the right row the high level guarantee ($g = 4\%$ p.a.), respectively.

bond and the pure bond strategies as riskless for certain long-term contracts. The shortfall probability and shortfall standard deviation have a positive value, even for the pure bond strategy with a 30-year investment horizon. They are, however, close to zero. On the countrary, the normalized MEL is significantly higher than zero for all investment strategies. The less risky strategy, namely the pure bond one, has a normalized MEL between 2.62% and 3.66%.

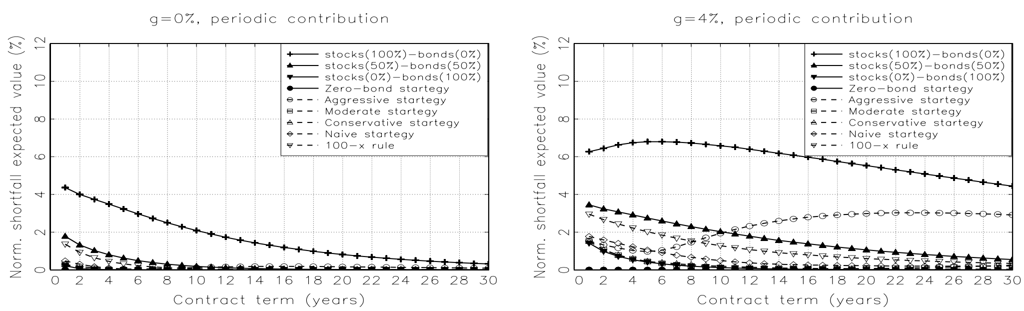
The normalized MEL of the 0% guarantee seems to be higher for the pure bond strategy than for the 25%-75% stock-bond strategy, e.g., for the 16-year contract. This is not due to the higher risk of the pure bond strategy, but due to the fact that the Monte Carlo simulation with 20 millions iterations does not provide any stable results. This conclusion can be derived from the irregularity of the pure-bond-curve.

Figure 5.7: Impact of the investment strategy on the guarantee risk (all strategies)

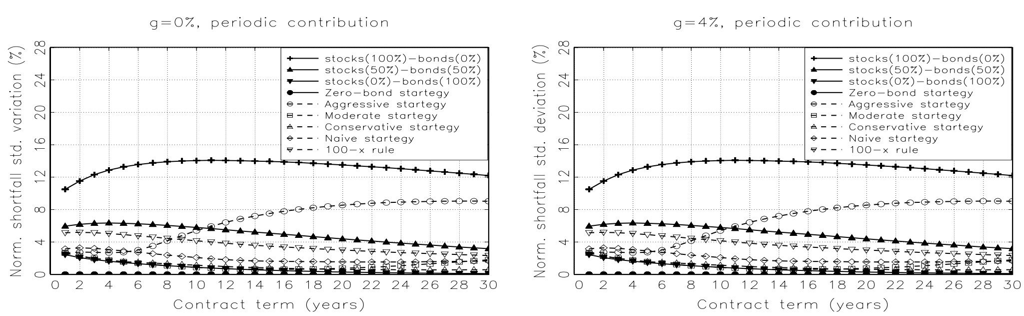
Shortfall probability



Normalized shortfall expected value



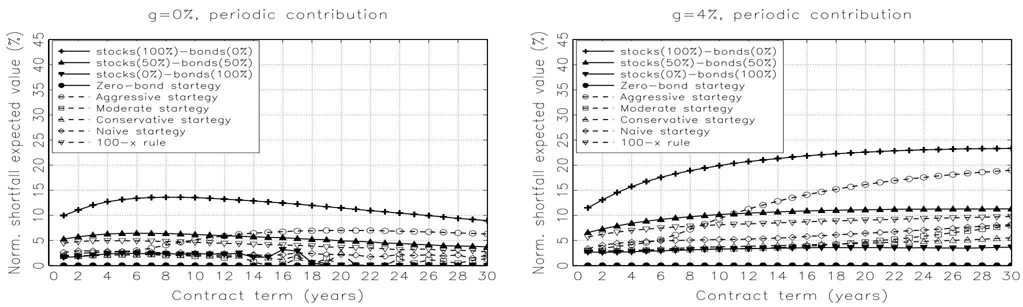
Normalized shortfall standard deviation



continued on the next page

continued from the previous page

Normalized mean excess loss



Note:
This figure depicts the impact of the investment strategy on the guarantee shortfall risk. All risk measures are computed for the MS(1-2) model. The solid line with pluses represents the pure stock investment strategy, the solid line with triangles – the 50%-50% stock-bond, the solid line with inverse triangles – the pure bond, the solid line with circles – the zero-bond, the dashed line with circles – the aggressive, the dashed line with squares – the moderate, the dashed line with triangles – the conservative, the dashed line with diamonds – the naive, and the dashed line with inverse triangles – the 100-x investment strategy, respectively. In all cases, contributions of €1200 are paid up-front annually. The left column represents the low level guarantee ($g = 0\%$ p.a.) and the right row the high level guarantee ($g = 4\%$ p.a.), respectively.

Persuant to the above discussion, we can state that in order to reduce the shortfall risk, the guarantee provider can reduce the stock proportion in the buy-and-hold strategies. Another solution would be to use one of the life-cycle strategies defined in Table 4.1. Figure 5.7 shows the impact of life-cycle investment strategies on the guarantee risk. The results are the same for all risk measures.

The least risky of the life-cycle strategies is the conservative one. This is followed by the moderate, the naive, the 100-x, and the aggressive strategy. There are some exceptions for the middle-term contracts which yield a slightly higher risk for the conservative strategy than for the moderate one.

The risk of the conservative strategy is always lower than the risk of the 25%-75% buy-and-hold strategy, independent of the guarantee level or contract term. The risk of the conservative and the moderate investment strategy is almost always lower than the risk of the 25%-75% stock-boond strategy. Only long-term contracts with a 4% guarantee have a lower risk when backed by the 25%-75% stock-bond portfolio than when backed by the conservative or moderate strategy. The naive strategy has the same risk as the 25%-75% stock-bond portfolio for short-term contracts due to the strategy construction. For the middle and long-term contracts, the risk for this strategy lies between the risk of the 25%-75% and the fifty-fifty stock-bond strategy, independent of the guarantee level. The risk of the 100-x investment rule always lies between the risk of the 25%-75% and the 50%-50% buy-and-hold stratgy. The aggressive strategy is the most risky one among the life-cycle strategies. For the short-term contracts, its risk is below that of the 25%-75% buy-and-hold strategy. However, for middle-term contracts, its risk exceeds the risk of the fifty-fifty, and for very long-term contracts, it even exceeds the risk of the 75%-25% buy-and-hold strategy.

Lastly, we turn our attention to the zero-bond strategy, which is riskless due to its construction for all possible configurations of the guarantee level and the contract duration. From the point of view of the guarantee provider, it is clearly the best strategy, as it does not require any solvency capital. However, it remains an open question if this strategy is attractive for his clients, since the price for the risk reduction might be a significantly lower expected profit. This issue will be discussed in Section 5.6.

5.4.5 Impact of time on the guarantee risk

This Section discusses the impact of time on the guarantee shortfall risk. First, buy-and-hold strategies will be discussed. We will then proceed by examining life-cycle strategies and lastly, we focus our attention on the zero-bond strategy.

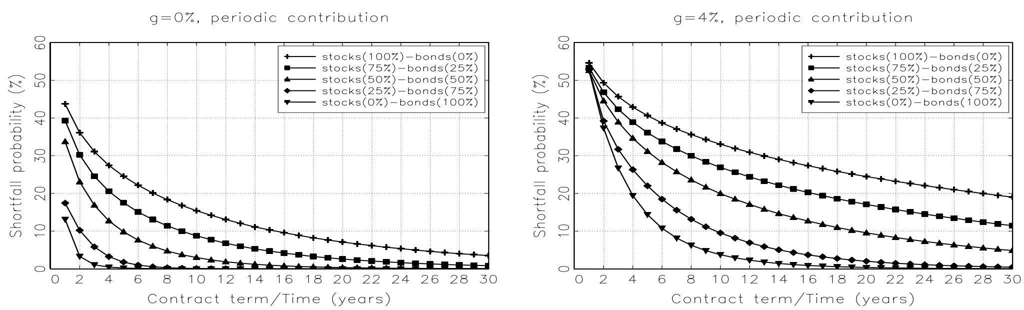
We will begin by discussing the impact of time on the 0% guarantee backed by buy-and-hold strategies (see left column of Figure 5.8). The shortfall probability and the normalized shortfall expected value are convex decreasing functions of time. The same holds true for the normalized shortfall standard deviation if the stock proportion in the backing portfolio is not higher than 50%. If the stock proportion is between 60% and 75%, the normalized shortfall standard deviation is a decreasing function of time, however, it is not convex any more. For the remaining buy-and-hold portfolios, this risk measure increases to a certain maximum and decreases thereafter. This maximum occurs in the 2nd or 3rd year. The normalized mean excess loss increases until it reaches a peak and decreases thereafter. The maximum occurs between the 2nd and the 8th year. The Monte Carlo simulation with 20.000.000 iterations does not provide any stable results for the normalized conditional shortfall standard deviation.

Generally speaking, contracts with a low and middle stock proportion have a moderate risk, regardless of the risk measure we use. Furthermore, as time increases, they become riskless. On the contrary, the risk of contracts with a high stock proportion is significantly high in the short time. As time passes, the risk of these contracts becomes moderate.

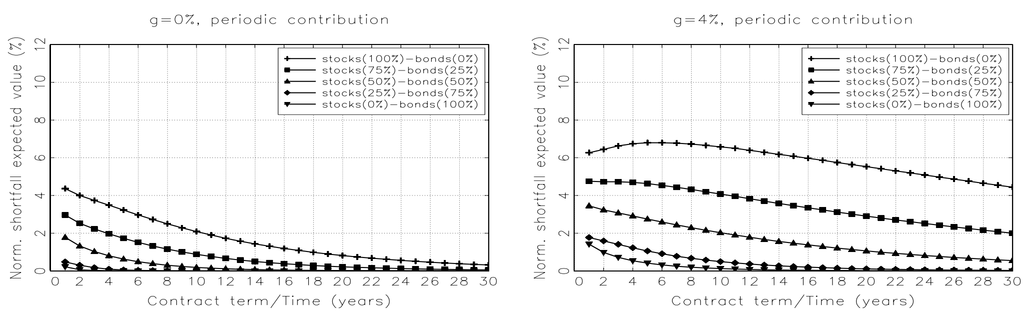
The impact of time on the risk of the 4% guarantee backed by the buy-and-hold strategies is slightly different (see right column of Figure 5.8). Only the shortfall probability is a convex decreasing function of time for all buy-and-hold strategies. The normalized shortfall expected value and the normalized shortfall standard deviation are decreasing functions of time if the stock proportion does not exceed 75% and 10%, respectively. In case of remaining portfolios, these risk functions increase to a certain maximum and decrease

Figure 5.8: Impact of time and the contract term on the guarantee shortfall risk (buy-and-hold strategies)

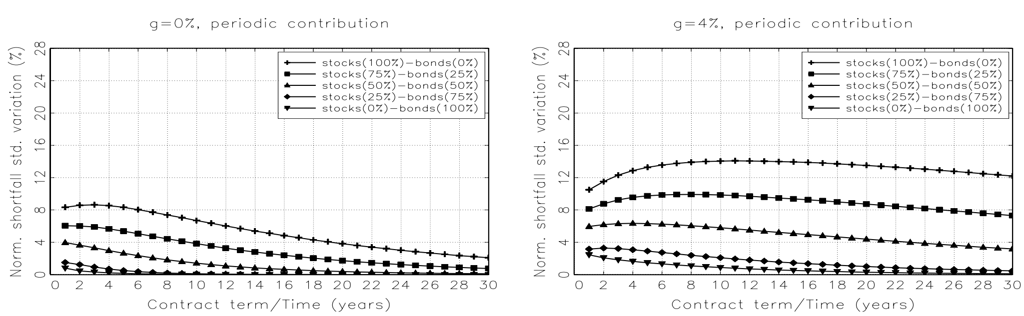
Shortfall probability



Normalized shortfall expected value



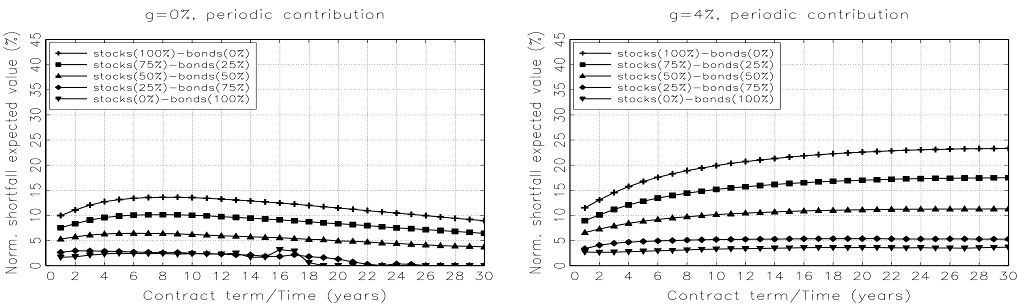
Normalized shortfall standard deviation



continued on the next page

continued from the previous page

Normalized mean excess loss



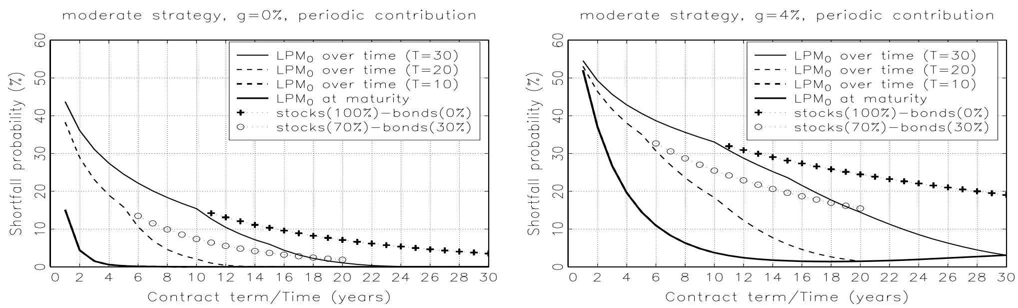
Note:
This figure depicts the impact of time and the contract term on the guarantee shortfall risk. All risk measures are computed for the MS(1-2) model. The solid line with pluses represents the pure stock investment strategy, the solid line with squares – the 75%-25% stock-bond, the solid line with triangles – the 50%-50% stock-bond, the solid line with diamonds – the 25%-75% stock-bond, the solid line with inverse triangles – the pure bond investment strategy, respectively. In all cases, contributions of €1200 are paid up-front annually. The left column represents the low level guarantee ($g = 0\%$ p.a.) and the right row the high level guarantee ($g = 4\%$ p.a.), respectively.

afterwards. The normalized shortfall expected value reaches its maximum between the 4th (80%-20% stock bond portfolio) and the 6th (pure stock portfolio) year, and the normalized shortfall standard deviation reaches its maximum between the 2nd (20%-80% stock bond portfolio) and the 11th (pure stock portfolio) year. The normalized mean excess loss is an increasing function of time in the interval between 1 and 30 years. The Monte Carlo simulation with 20.000.000 iterations only provides stable results for the normalized conditional shortfall standard deviation for portfolios with a high stock proportion. In these cases, this risk measure is an increasing function of time.

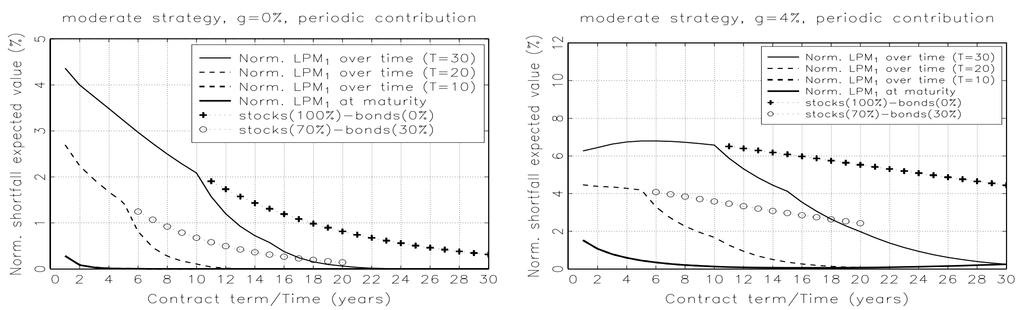
Generally speaking, contracts with low stock portfolios have a moderate risk when the risk is measured with the normalized shortfall expected value, the normalized shortfall standard deviation, or the normalized MEL. On the

Figure 5.9: Impact of time and the contract term on the guarantee shortfall risk (moderate strategy)

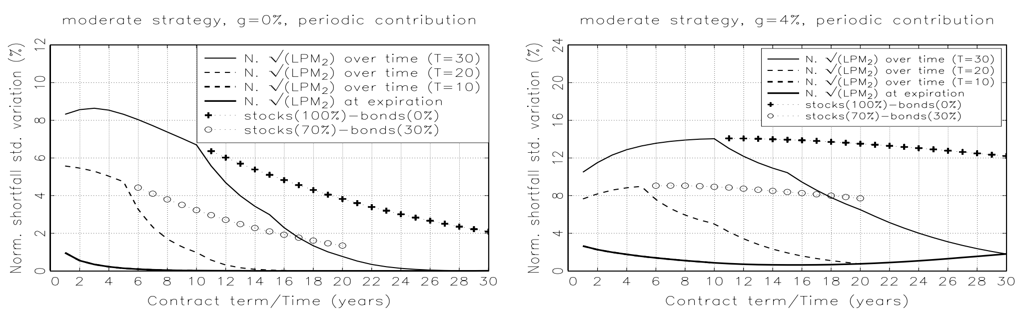
Shortfall probability



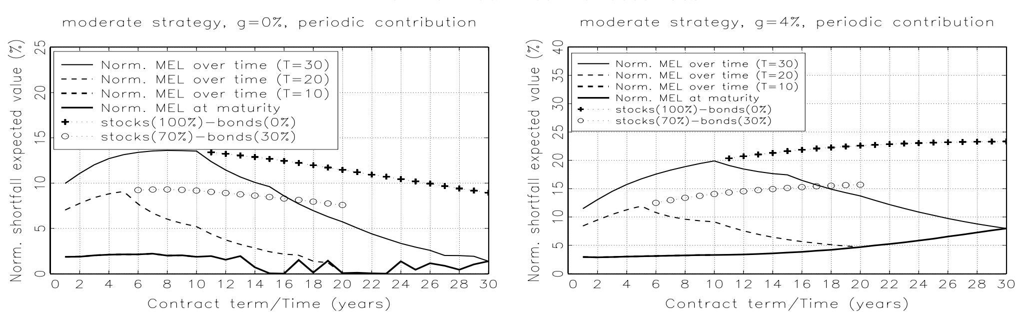
Normalized shortfall expected value



Normalized shortfall standard deviation



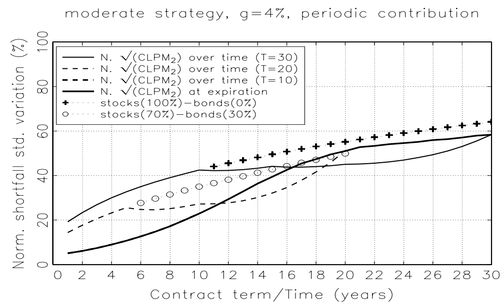
Normalized mean excess loss



continued on the next page

continued from the previous page

Normalized conditional shortfall standard deviation



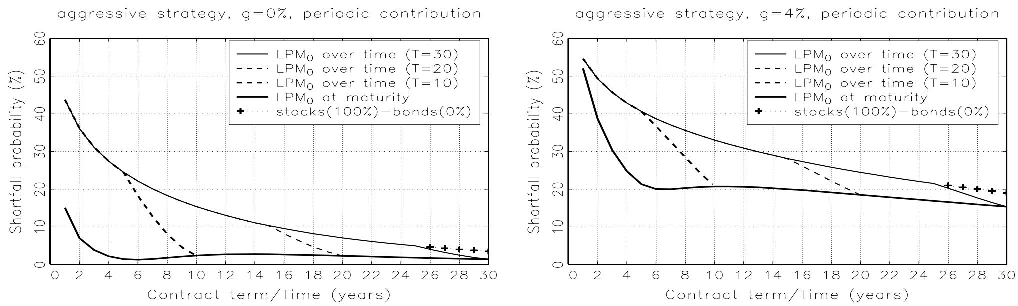
Note:
This figure depicts the impact of time and the contract term on the guarantee shortfall risk using the example of the moderate strategy with periodic contributions (€1200 are paid up-front annually). All risk measures are computed for the MS(1-2) model. The bold dashed line shows how the risk measure changes over time for a 10-year contract, the thin dashed line for a 20-year contract, and the thin solid line for a 30-year contract, respectively. The bold solid line shows the normalized contract at contract expiration. The bold pluses depict the 100%-0% and the empty circles – the 70%-30% stock-bond buy-and-hold portfolio, respectively. The left column represents the low level guarantee ($g = 0\%$ p.a.), and the right row the high level guarantee ($g = 4\%$ p.a.), respectively.

contrary, risk measured by the shortfall probability for the short time is very high. As time increases, the risk measured with the shortfall probability, normalized shortfall expected value, and the normalized shortfall standard deviation becomes lower or even vanishes completely. Only the normalized MEL increases along time. This means that for short time, the shortfall occurs relatively often. However, in these cases, the realized loss is moderate. As time increases, the probability of shortfall decreases, but solvency capital increases slightly. On the other hand, the risk of contracts with a high and a middle stock proportion is high. As time increases, the risk remains on a high level regardless of the risk measure used.

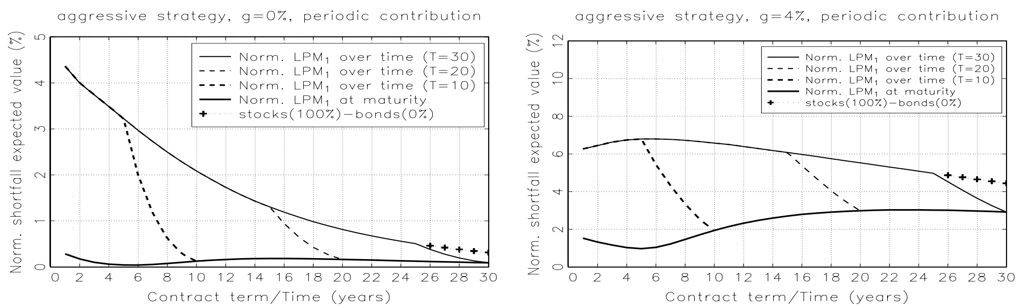
The impact of time on the life-cycle strategies will be discussed in the example of the moderate strategy (see Figure 5.9). These results are rep-

Figure 5.10: Impact of time and the contract term on the guarantee shortfall risk (aggressive strategy)

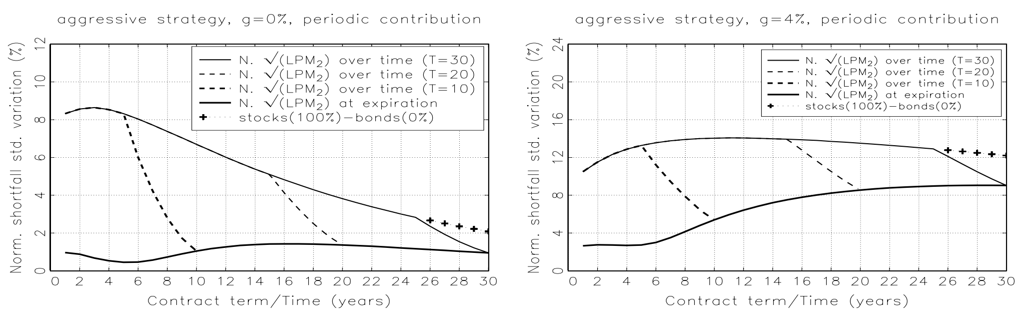
Shortfall probability



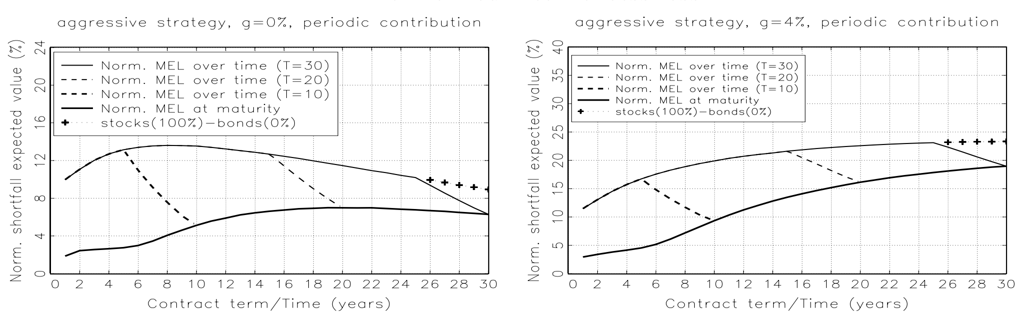
Normalized shortfall expected value



Normalized shortfall standard deviation



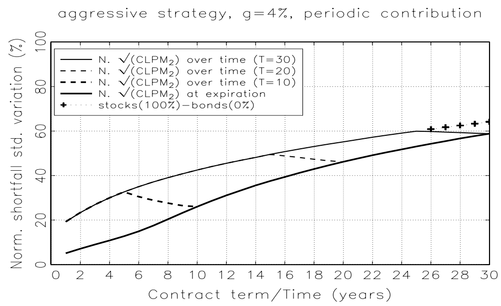
Normalized mean excess loss



continued on the next page

continued from the previous page

Normalized conditional shortfall standard deviation



Note:
This figure depicts the impact of time and the contract term on the guarantee shortfall risk using the example of the aggressive strategy with periodic contributions (€1200 are paid up-front annually). All risk measures are computed for the MS(1-2) model. The bold dashed line shows how the normalized cost changes over time for a 10-year contract, the thin dashed line for a 20-year contract, and the thin solid line for a 30-year contract, respectively. The bold solid line shows the normalized contract at contract expiration. The bold pluses depict the pure stock buy-and-hold portfolio, respectively. The left column represents the low level guarantee ($g = 0\%$ p.a.), and the right row the high level guarantee ($g = 4\%$ p.a.), respectively.

resentative for all but the aggressive strategy (see Figure 5.10), which will be discussed separately. Results for all life-cycle strategies can bee seen in Figures L.1-L.5 in Appendix L.

Figure 5.9 shows the impact of time on the moderate strategy for three examples: a 10-year contract (thick dashed line), a 20-year contract (thin dashed line), and a 30-year contract (thin solid line). At the beginning of the contract, the moderate strategy has the same risk as its initial portfolio, i.e. the 10%-90% stock-bond portfolio (not depicted on the figure), 70%-30% stock bond portfolio (empty circles), and the pure stock portfolio (thick pluses) for the 10-year, 20-year, and 30-year contract, respectively. Beginning at the first shifting date, the risk starts to decrease, regardless of whether the risk of the initial portfolio grows or falls. On each further shifting date, the speed of the risk reduction increases. The longer the contract term, the higher the risk reduction in comparison to the risk of the initial portfolio. These results remain the same, regardless of the risk measure used to quantify the shortfall risk.

Figure 5.10 shows the results for the aggressive strategy. In this case, the risk is also equal to the risk of the initial portfolio (i.e the pure stock portfolio), and it decreases after the shifting date – independent of whether the risk measure of the initial portfolio grows or declines. However, the risk reduction is higher, the shorter the contract term. This is due to the strategy construction: the shorter the contract term the proportionally longer the contract invests in the less risky pure bond portfolio.

The shortfall risk of the zero-bond strategy is insensitive to the change of time, as it is equal to zero due to the strategy construction.

5.4.6 Impact of the contract term on the guarantee risk

In the case of buy-and-hold investment strategies, the impact of the contract term on the guarantee risk is exactly the same as the impact of time. Therefore, we omit this discussion, and only discuss the life-cycle strategies and the zero-bond strategy.

The impact of the contract term on the life-cycle strategy will be discussed in the example of the moderate strategy, which is representative for all but the aggressive strategy. The aggressive strategy will thus be discussed separately. (Remaining investment strategies can be found in Figures L.1-L.5 in Appendix L). The normalized conditional shortfall standard deviation will be discussed only for the 4% guarantee backed with the aggressive strategy, as it is the only case in which the Monta Carlo simulation produced stable results.

The results for the moderate strategy are depicted in Figure 5.9 (thick solid line). The risk of the 0% guarantee is low for short-term contracts. The risk measured with the shortfall probability, the normalized shortfall expected value, and the noramlized shortfall standard deviation decreases very rapidly towards zero and remains at this level. The risk measured with the normalized MEL is stable at the level of ca. 2% for contract terms between 1 and 14. Thereafter the curve becoms irregular, because the shortfall occurs too rarely. Thus, the Monte Carlo simulation with 20 milions runs is not able to give a stable numerical solution.

The shortfall risk the 4% guarantee backed with the moderate strategy is a convex U-shaped function of the contract term. Several risk measures have a minimum for different contract terms. The shortfall probability has its minimum for the 19-year contract at the level of 14.03%, the normalized shortfall expected value for the 17-year contract at the level of 0.06%, the normalized shortfall standard deviation for the 18-year contract at the level of 0.64%, and the normalized MEL for the 2-year contract at the level of 0.02.88%, respectively. This shows that the shortfall risk can be optimized by a suitable choice of a contract term and a suitable design of the investment strategy (by choosing an optimal initial portfolio and an optimal shifting design). These results are similar to those of the guarantee cost, see Section 4.7.2.6.

Figure 5.10 shows the impact of the contract term on the shortfall risk of the aggressive strategy (thick solid line). We begin with a discussion of the 0% guarantee. The shortfall probability equals 15.11% for the one year contract, it then decreases rapidly and for contracts with a duration greater than 5 years, it remains stable at a level of ca. 2%. The normalized shortfall expected value and the normalized shortfall standard deviation are stable at a level of 0.1% and 1%, respectively, for all contract terms between 1 and 30 years. The normalized MEL has an inverted U-shape. It increases from 1.86% (one year contract) to 6.99% (19 years contract), and then starts decreasing to 6.29% (30-year contract).

Curves of the shortfall risk measures for a 4% guarantee backed by the aggressive strategy have the same shape as for a 0% guarantee. However, the risk is higher than in the case of the 0% guarantee. The shortfall probability starts at a very high level of 52.05% (one year contract), then sinks to 21.34% (5-year contract) and remains near the 20% mark with a slightly decreasing tendency thereafter. The remainig risk measures behave similarly. They are stable for contracts with maturities between 1 and 5 years and increase thereafter. The normalized expected value increases from 1.52% to 2.92%, the normalized shortfall standard deviation from 2.64% to 9.04%, the normalized MEL from 2.93% to 18.97%, and the normalized shortfall standard deviation from 5.08% to 58.83%, between the 1-year and 30-year contract, respectively. It can be seen that the slope is higher (a) for higher partial moments and (b) for conditional rather than for unconditional risk measures.

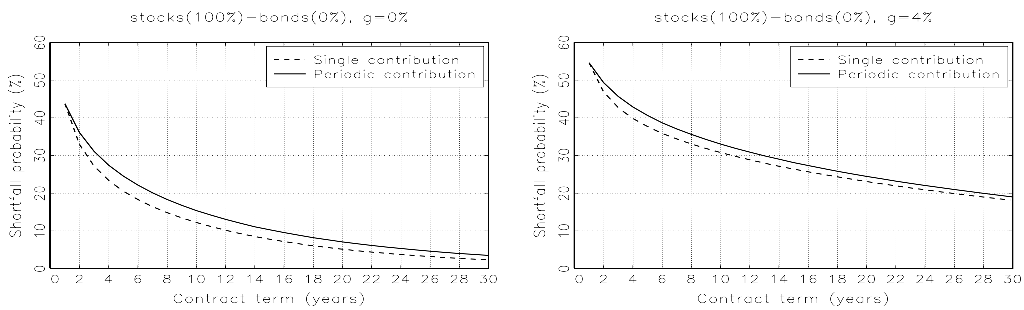
The shortfall risk of the zero-bond strategy is insensitive to the change of contract term, as it equals zero due to the strategy construction.

5.4.7 Impact of the contribution payment scheme on the guarantee risk

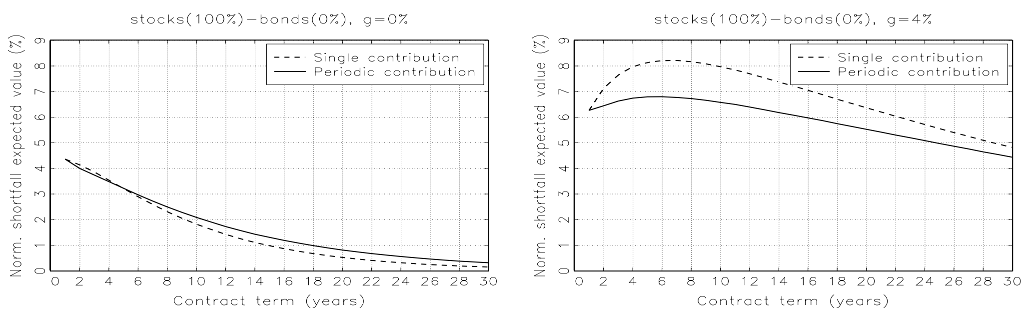
This Section compares the guarantee risk for two alternative contribution payment schemes: the single premium and the periodic premium (€1200 up-front yearly). For the sake of comparability, the single premium is chosen to take on the net present value of aggregated periodic payments. The

Figure 5.11: Impact of the contribution payment scheme on risk measures (pure stock portfolio)

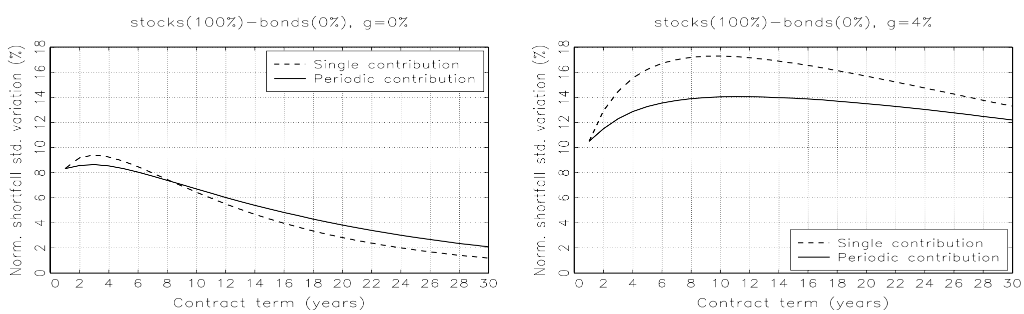
Shortfall probability



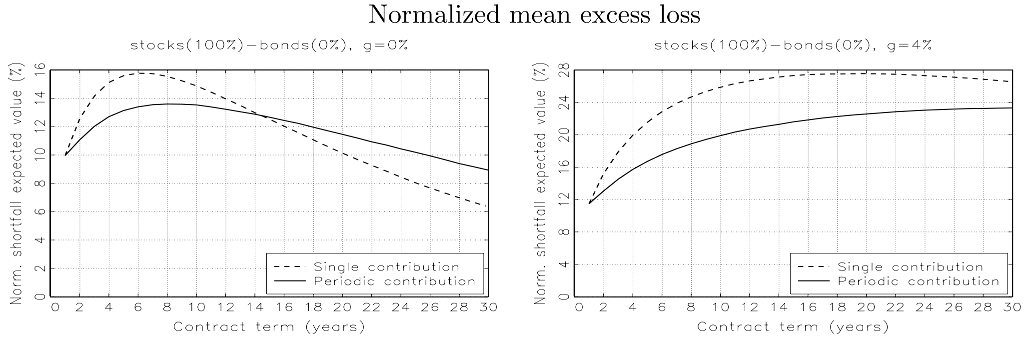
Normalized shortfall expected value



Normalized shortfall standard deviation



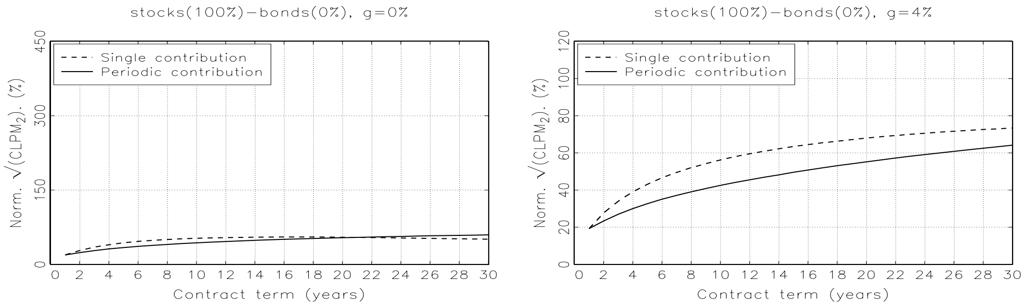
Normalized mean excess loss



continued on the next page

continued from the previous page

Normalized conditional shortfall standard deviation



Note:

This figure depicts the impact of the payment scheme on the guarantee shortfall risk using the example of the pure stock strategy and the MS(1-2) model. The solid line represents the periodic payment scheme (€1200 up-front annually), and the dashed line the single premium case. The single premium equals the net present value of periodic contributions. The left column shows contracts with a low level guarantee ($g = 0\%$ p.a.), and the right column contracts with a high level guarantee ($g = 4\%$ p.a.), respectively.

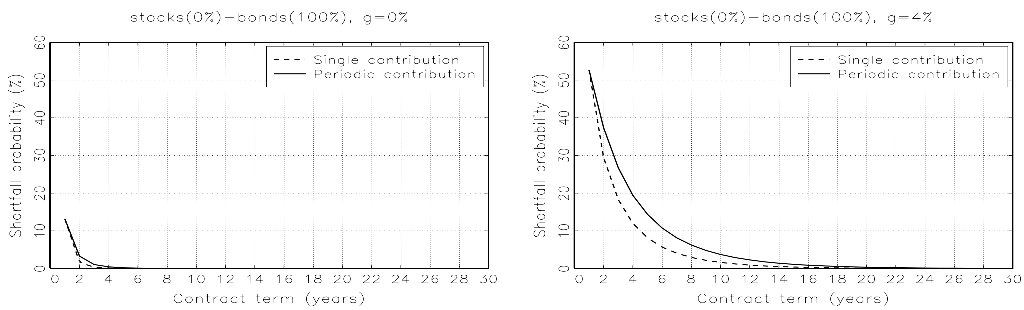
subsequent paragraph begins by discussing the buy-and-hold strategies. We will proceed by examining the life-cycle strategies and lastly, the zero-bond strategy.

Figures 5.11 and 5.12 show the impact of the contribution payment scheme using the example of a pure stock and a pure bond portfolio, respectively. The shortfall probability is higher for contracts with periodic contributions independent of the stock engagement and guarantee level.

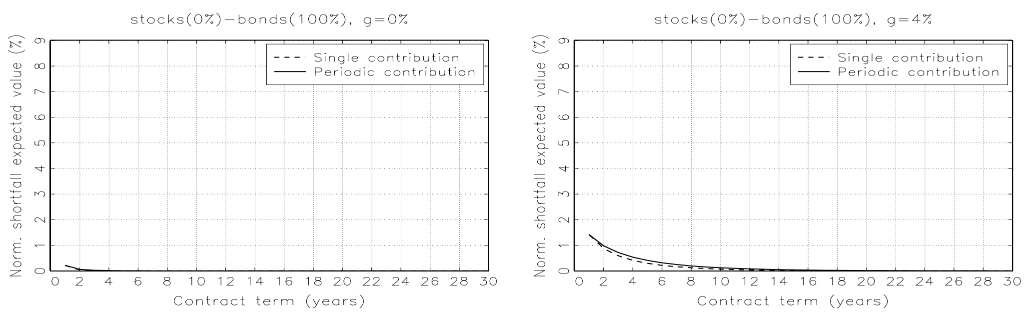
The normalized shortfall expected value, the normalized shortfall standard deviation, and the normalized MEL react in the same way with respect to change in the contribution payment. In the case of the 0% guarantee, the shortfall risk is higher for the periodic contribution payment when using these three risk measures. The only exceptions are short-term contracts with a very high stock proportion (see Figure 5.11). In the case of the 4% guarantee, the shortfall risk is higher for contracts with single premium backed by buy-and-hold portfolios with a very high stock proportion (see Figure

Figure 5.12: Impact of the contribution payment scheme on risk measures (pure bond portfolio)

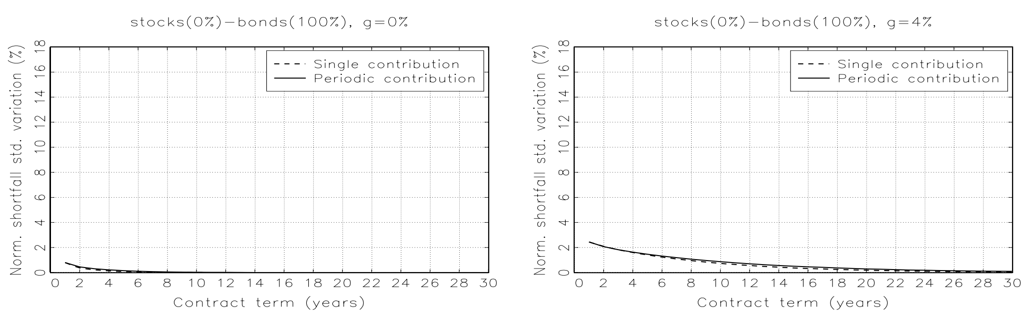
Shortfall probability



Normalized shortfall expected value



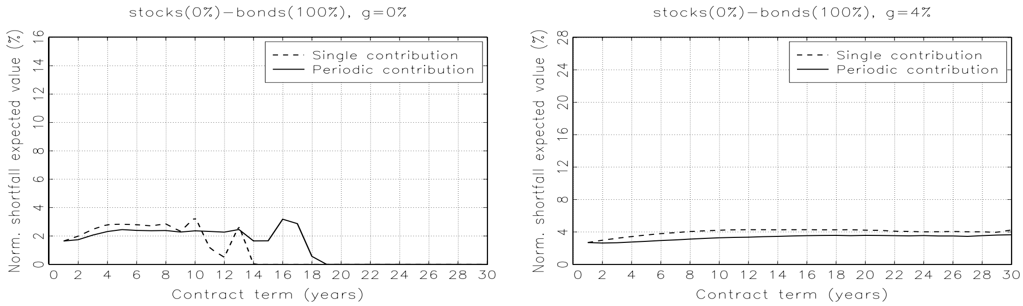
Normalized shortfall standard deviation



continued on the next page

continued from the previous page

Normalized mean excess loss



Note:
This figure depicts the impact of the payment scheme on the guarantee shortfall risk using the example of the pure bond strategy and the MS(1-2) model. The solid line represents the periodic payment scheme (€1200 up-front annually), and the dashed line the single premium case. The single premium equals the net present value of periodic contributions. The left column shows contracts with a low level guarantee ($g = 0\%$ p.a.), and the right column contracts with a high level guarantee ($g = 4\%$ p.a.), respectively.

5.11). If the bond proportion is high, contracts with periodic premiums have a slightly higher shortfall risk (see Figure 5.12). For all other buy-and-hold portfolios, the risk measure curves cross each other. For short-term contracts, the shortfall risk is higher for the single premium payment scheme. For long-term contracts, the shortfall risk is higher in the periodic contribution payment scheme. The lower the stock proportion in the portfolio, the sooner both curves cross each other.

The normalized conditional shortfall standard deviation behaves differently from the three risk measures discussed above. In the case of the 0% guarantee, the normalized conditional shortfall standard deviation is higher for short-term contracts with a single premium. As the contract term grows, the discrepancy between both payment schemes declines. For middle-term contracts, both curves cross each other. For long-term contracts, the single contribution scheme yields a lower shortfall risk. In the case of the 4% guarantee, the risk is higher in the single premium scheme, independent of the respective stock engagement and the contract term (for cases which yield stable numeric results).

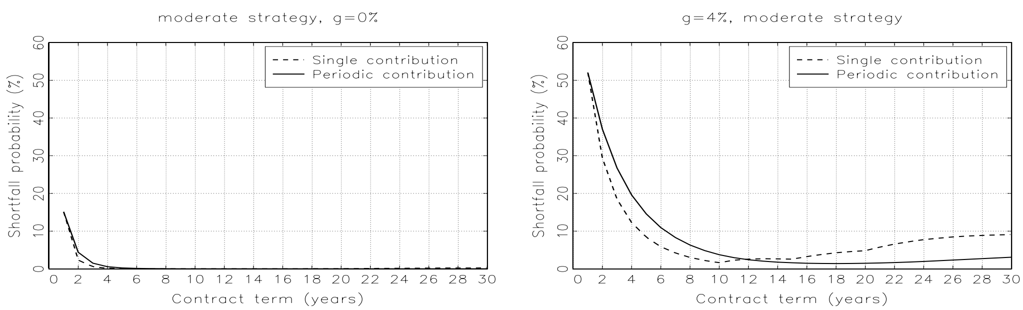
The impact of the contribution payment scheme on the life-cycle strategy will be discussed using the example of the moderate strategy (see Figure 5.13) which is representative for all but the aggressive strategy. The aggressive strategy (see Figure 5.14) will be discussed separately. (Remaining investment strategies can be found in Figures M.1-M.5 in Appendix M).

First, we turn our attention to the moderate strategy (see Figure 5.13). In this case the shortfall probability, the normalized shortfall expected value, and the normalized shortfall standard deviation behave similarly. Considering the 0% guarantee, the shortfall risk is higher for a periodic contribution scheme when the contract term is short. For the middle contract term, the shortfall risk sinks to zero independent of the payment scheme. As the contract term grows, the risk in the periodic contribution scheme remains at the zero level, but the risk in the single contribution scheme becomes slightly positive. Only the shortfall expected value remains at the zero level – even in the single contribution scheme. Now we consider the the 4% guarantee level. For short-term contracts, the shortfall risk is higher in the periodic contribution scheme. Approximately at the 10-year mark of the contract term both curves cross each other and then the single contribution scheme yields the higher risk. The discrepancy is higher, the longer the contract term. It is worth mentioning that the shortfall risk of the 4% guarantee has a minimum in both payment schemes. However, the minimum of the single contribution function occurs for lower contract terms than the minimum of the periodic payment scheme.

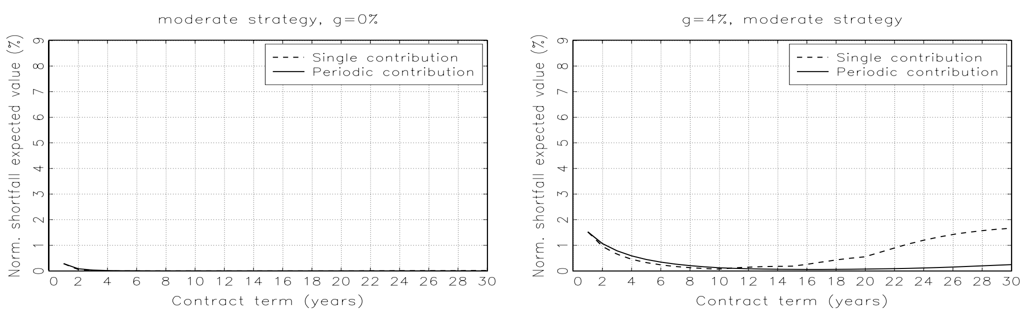
The normalized MEL is higher for the single payment scheme, regardless of the guarantee level. The normalized conditional shortfall standard devia-

Figure 5.13: Impact of the contribution payment scheme on the guarantee shortfall risk (moderate strategy)

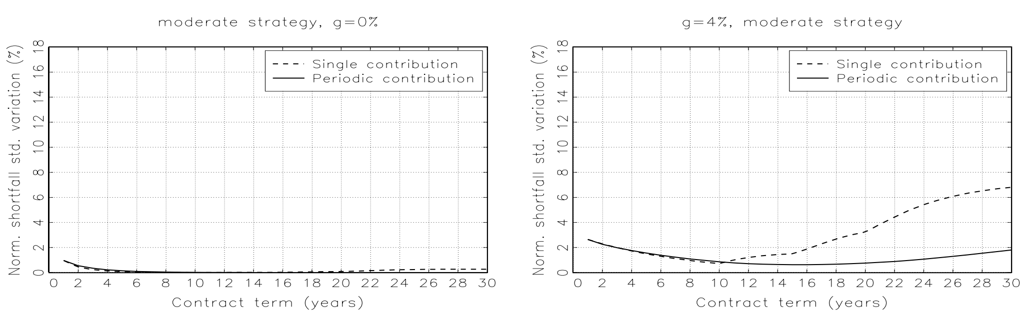
Shortfall probability



Normalized shortfall expected value



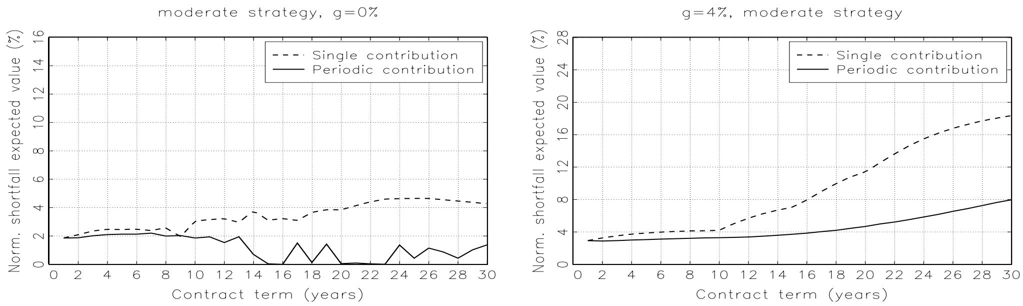
Normalized shortfall standard deviation



continued on the next page

continued from the previous page

Normalized mean excess loss



Note:

This figure depicts the impact of the payment scheme on the guarantee shortfall risk using the example of the moderate strategy and the MS(1-2) model. The solid line represents the periodic payment scheme (€1200 up-front annually), and the dashed line the single premium case. The single premium equals the net present value of periodic contributions. The left column shows contracts with a low level guarantee ($g = 0\%$ p.a.), and the right column contracts with a high level guarantee ($g = 4\%$ p.a.), respectively.

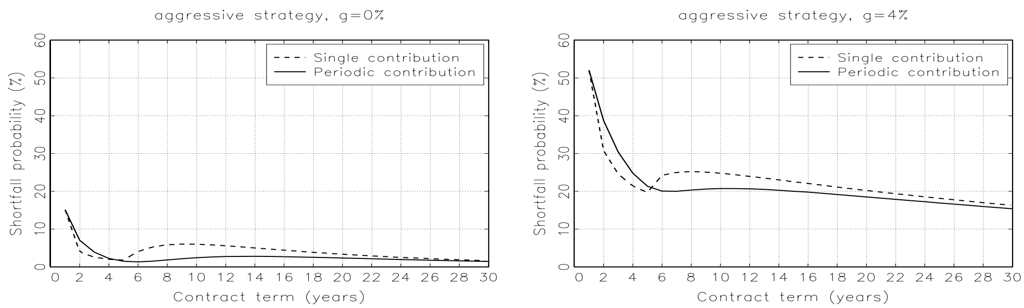
tion does not yield any stable numeric results for the moderate strategy.

Figure 5.14 shows the results for the aggressive strategy. All of the discussed risk measures react similarly to the change of the payment scheme regardless of the guarantee level. For short-term contracts, the shortfall risk is lower in the single contribution scheme. Nevertheless, both curves cross each other at the latest in the 5-year contract. After that, the shortfall risk is higher in the single payment scheme. In contrast to the moderate strategy, the gap between both payment schemes does not widen with a growing contract term. The discrepancy is the highest for about the 10-year contract and then begins to decline. In the case of the normalized MEL, the scissors even close in the case of a 0% guarantee, and for a contract term higher or equal to 28, the risk is higher for the periodic contribution scheme.

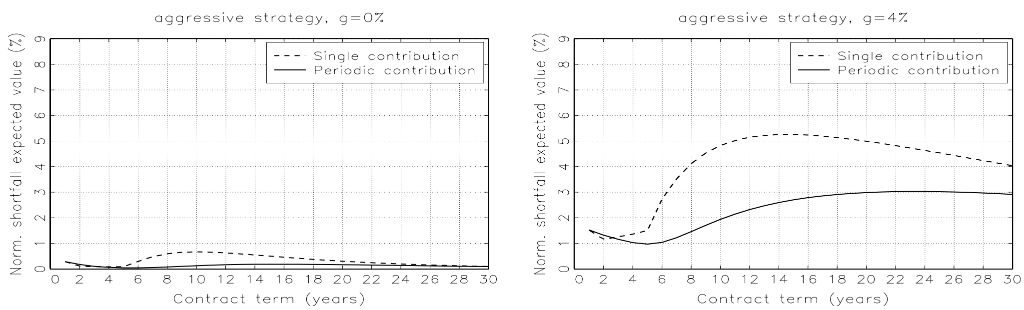
The shortfall risk of the zero-bond strategy is insensitive to a change of the contribution payment scheme as it equals zero due to the strategy construction.

Figure 5.14: Impact of the contribution payment scheme on the guarantee shortfall risk (aggressive strategy)

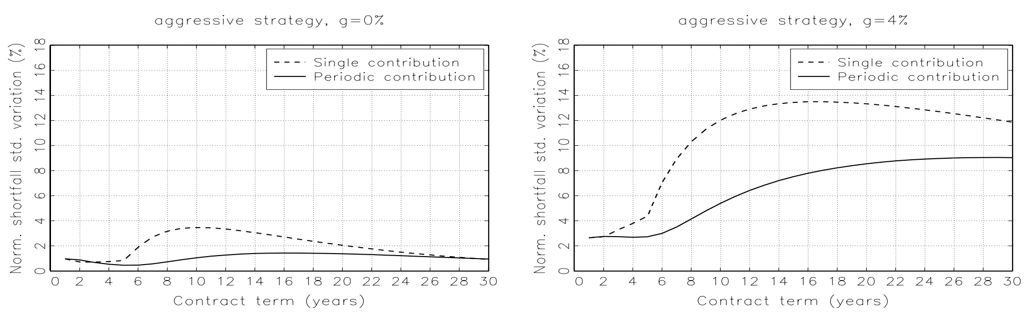
Shortfall probability



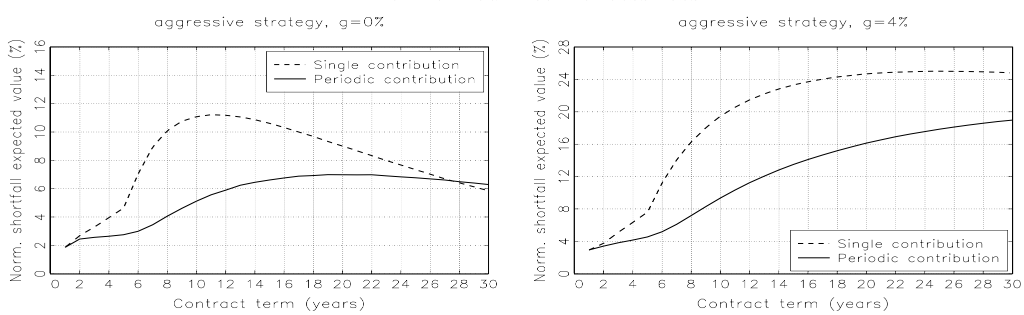
Normalized shortfall expected value



Normalized shortfall standard deviation



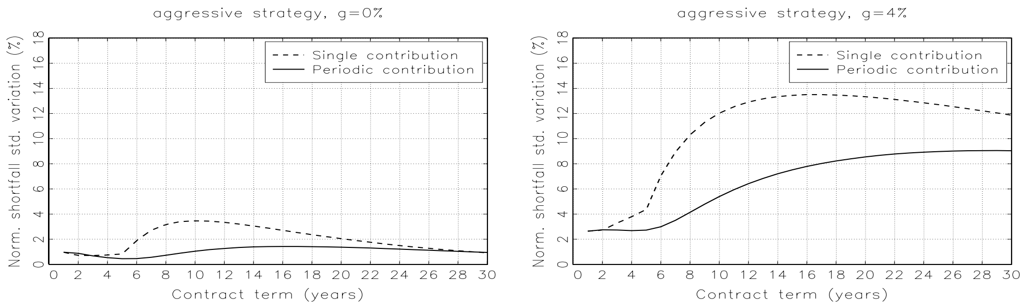
Normalized mean excess loss



continued on the next page

continued from the previous page

Normalized conditional shortfall standard deviation



Note:

This figure depicts the impact of the payment scheme on the guarantee shortfall risk using the example of the aggressive strategy and the MS(1-2) model. The solid line represents the periodic payment scheme (€1200 up-front annually), and the dashed line the single premium case. The single premium equals the net present value of periodic contributions. The left column shows contracts with a low level guarantee ($g = 0\%$ p.a.), and the right column contracts with a high level guarantee ($g = 4\%$ p.a.), respectively.

5.5 Solvency requirements

Section 5.4 addresses the impact of diverse factors (guarantee level, investment strategy, time, contract term, and contribution payment scheme) on the guarantee risk. All of these aspects are important from the perspective of the financial company managing the guarantee risk or considering to sell a guarantee to its customers.

However, for the solvency supervising authority, a very important issue concerns the question of how to quantify solvency requirements. As we have mentioned already at the end of Section 5.2.3, the mean excess loss is a very good measure of solvency requirements as it quantifies the average loss when loss occurs. In other words, it quantifies how much capital on average the guarantee provider should collect *ex ante* to compensate the loss resulting from falling of the portfolio value below the guaranteed amount when this scenario would realize. For this reason the MEL can be considered as a worst-case measure and therefore it is interesting from the point of view of

Table 5.1: Ranking of investment strategies ($g = 0\%$)

Strategy	shortf. prob.	norm. SEV	norm. SSD	norm. MEL	norm. cost	norm. profit
g=0%, T=10						
Zero-bond	0.00%	0.00%	0.00%	**0.00%**	0.00%	**7.34%**
Moderate*	0.00%	0.00%	0.01%	**1.86%**	0.00%	**8.27%**
Conservative	0.01%	0.00%	0.03%	**2.16%**	0.01%	**8.54%**
100% bond	0.01%	0.00%	0.02%	**2.36%**	0.00%	**6.84%**
Naive**	0.09%	0.00%	0.10%	**2.46%**	0.04%	**10.15%**
100-x rule***	1.23%	0.06%	0.68%	**4.62%**	0.32%	**12.74%**
Aggressive	2.41%	0.12%	1.04%	**5.12%**	0.45%	**13.32%**
50%-50% stock-bond	2.99%	0.18%	1.41%	**6.18%**	0.66%	**13.63%**
75%-25% stock-bond	8.76%	0.88%	3.82%	**10.03%**	1.92%	**18.20%**
100% stock	15.40%	2.08%	6.70%	**13.54%**	3.40%	**22.20%**
g=0%, T=20						
Strategy	**shortf. prob.**	**norm. SEV**	**norm. SSD**	**norm. MEL**	**norm. cost**	**norm. profit**
Zero-bond	0.00%	0.00%	0.00%	**0.00%**	0.00%	**33.98%**
100% bond	0.00%	0.00%	0.00%	**0.00%**	0.00%	**18.57%**
Moderate	0.00%	0.00%	0.00%	**0.05%**	0.00%	**26.38%**
25%-75% stock-bond	0.00%	0.00%	0.00%	**1.28%**	0.00%	**27.18%**
Naive	0.00%	0.00%	0.01%	**2.22%**	0.01%	**29.82%**
Conservative	0.00%	0.00%	0.00%	**2.52%**	0.00%	**25.60%**
100-x rule	0.07%	0.00%	0.13%	**3.58%**	0.05%	**34.71%**
50%-50% stock-bond	0.33%	0.02%	0.37%	**4.91%**	0.15%	**36.63%**
Aggressive	2.33%	0.16%	1.36%	**6.98%**	0.57%	**48.58%**
75%-25% stock-bond	2.59%	0.22%	1.72%	**8.34%**	0.80%	**49.68%**
100% stock	7.09%	0.81%	3.82%	**11.46%**	1.79%	**62.14%**
g=0%, T=30						
Strategy	**shortf. prob.**	**norm. SEV**	**norm. SSD**	**norm. MEL**	**norm. cost**	**norm. profit**
Zero-bond	0.00%	0.00%	0.00%	**0.00%**	0.00%	**86.66%**
Conservative	0.00%	0.00%	0.00%	**0.00%**	0.00%	**52.64%**
25%-75% stock-bond	0.00%	0.00%	0.00%	**0.00%**	0.00%	**49.75%**
100% bond	0.00%	0.00%	0.00%	**0.00%**	0.00%	**33.44%**
Moderate	0.00%	0.00%	0.00%	**1.37%**	0.00%	**63.66%**
Naive	0.00%	0.00%	0.01%	**2.04%**	0.00%	**63.17%**
100-x rule	0.01%	0.00%	0.03%	**2.70%**	0.01%	**67.71%**
50%-50% stock-bond	0.04%	0.00%	0.10%	**3.72%**	0.04%	**68.51%**
Aggressive	1.41%	0.09%	0.95%	**6.29%**	0.38%	**104.53%**
75%-25% stock-bond	0.83%	0.05%	0.75%	**6.40%**	0.33%	**95.91%**
100% stock	3.51%	0.31%	2.08%	**8.94%**	0.91%	**123.74%**

Note:
This table ranks investment strategies according to the normalized solvency capital (fifth column) and the normalized expected return (seventh column) for the low level guarantee ($g = 0\%$ p.a.) with periodic contributions (€1200 up-front annually). All results are computed for the MS(1-2) model. The top section represents the 10-year contract, the middle section the 20-year contract, and the bottom section the 30-year contract. SEV denotes the shortfall expected value, SSD the shortfall standard deviation, and MEL the mean excess loss.
* – for contracts with a 10-year term, the moderate life-cycle strategy invests in the 10%-90% stock-bond portfolio. ** – for contracts with a 10-year term, the naive life-cycle strategy invests in the 25%-75% stock-bond portfolio. *** – for contracts with a 10-year term, the 100-x life-cycle strategy invests in the 40%-60% stock-bond portfolio.

the solvency supervising authority.

Tables 5.1 and 5.2 rank the chosen investment guarantees from the lowest to the highest solvency capital (i.e. normalized MEL) in the case of the 0% and the 4% guarantee respectively. Please note that in the case of the low level guarantee with middle and high investment horizon, there are several investment strategies which require no solvency capital (or reserves). For example, for the 30-year contract, these are the zero-bond strategy, the conservative strategy and all buy-and-hold strategies with a stock proportion of up to 25%. The shorter the investment horizon, the less the number of investment guarantees which do not require any solvency capital. For example, for the 10-year contract only the zero-bond strategy requires no solvency capital. Similarly, in the case of the 4% guarantee, the zero-bond strategy is the only one which does not require any solvency capital independent of the contract term.

Tables 5.1 and 5.2 show additionally that the supervising authority should rather require from the guarantee provider (1) to hold solvency capital according to the risk exposure of investment guarantees, rather than (2) to hold a constant solvency capital rate independent of the risk management measure the guarantee provider employs. The second solution would punish the guarantee providers who pursue a conservative risk management policy. It would also give an incentive to retail products which are near this solvency rate, in order to avoid of "wasting" of the solvency capital. This would lead to an increase of the risk affinity of guarantee providers which may not be intended by the supervising authority.

In our opinion, the best solution is one which joins both of the above mentioned possibilities. The supervising authority should let the guarantee provider hold the solvency capital according to the risk he is exposed to. How-

Table 5.2: Ranking of investment strategies ($g = 4\%$)

g=4%, T=10						
Strategy	**shortf. prob.**	**norm. SEV**	**norm. SSD**	**norm. MEL**	**norm. cost**	**norm. profit**
Zero-bond	0.00%	0.00%	0.00%	**0.00%**	0.00%	**2.08%**
100% bond	3.74%	0.12%	0.87%	**3.27%**	1.12%	**6.84%**
Moderate*	3.74%	0.12%	0.86%	**3.29%**	1.36%	**8.27%**
Conservative	5.23%	0.19%	1.14%	**3.72%**	1.64%	**8.54%**
Naive**	9.51%	0.49%	2.08%	**5.18%**	2.57%	**10.15%**
100-x rule***	15.81%	1.28%	4.15%	**8.10%**	4.09%	**12.74%**
Aggressive	20.73%	1.94%	5.39%	**9.35%**	4.71%	**13.32%**
50%-50% stock-bond	19.94%	2.03%	5.79%	**10.16%**	5.13%	**13.63%**
75%-25% stock-bond	26.90%	4.08%	9.85%	**15.15%**	7.62%	**18.20%**
100% stock	33.05%	6.58%	14.05%	**19.90%**	9.86%	**22.20%**
g=4%, T=20						
Strategy	**shortf. prob.**	**norm. SEV**	**norm. SSD**	**norm. MEL**	**norm. cost**	**norm. profit**
Zero-bond	0.00%	0.00%	0.00%	**0.00%**	0.00%	**11.08%**
100% bond	0.36%	0.01%	0.29%	**3.57%**	0.58%	**18.57%**
Conservative	1.13%	0.05%	0.63%	**4.42%**	1.38%	**25.60%**
Moderate	1.48%	0.07%	0.75%	**4.68%**	1.53%	**26.38%**
25%-75% stock-bond	2.00%	0.11%	1.00%	**5.35%**	1.91%	**27.18%**
Naive	3.38%	0.22%	1.54%	**6.42%**	2.42%	**29.82%**
100-x rule	6.64%	0.60%	3.01%	**9.05%**	3.77%	**34.71%**
50%-50% stock-bond	9.54%	1.06%	4.38%	**11.10%**	4.76%	**36.63%**
Aggressive	18.51%	2.99%	8.55%	**16.13%**	7.13%	**48.58%**
75%-25% stock-bond	17.08%	2.90%	8.72%	**17.00%**	7.63%	**49.68%**
100% stock	24.48%	5.53%	13.50%	**22.58%**	10.12%	**62.14%**
g=4%, T=30						
Strategy	**shortf. prob.**	**norm. SEV**	**norm. SSD**	**norm. MEL**	**norm. cost**	**norm. profit**
Zero-bond	0.00%	0.00%	0.00%	**0.00%**	0.00%	**32.20%**
100% bond	0.04%	0.00%	0.10%	**3.66%**	0.30%	**33.44%**
25%-75% stock-bond	0.44%	0.02%	0.46%	**5.27%**	1.38%	**49.75%**
Conservative	0.62%	0.03%	0.57%	**5.53%**	1.49%	**52.64%**
Moderate	3.08%	0.25%	1.80%	**7.97%**	2.78%	**63.66%**
Naive	2.46%	0.20%	1.66%	**8.16%**	2.75%	**63.17%**
100-x rule	3.47%	0.34%	2.34%	**9.76%**	3.52%	**67.71%**
50%-50% stock-bond	4.85%	0.55%	3.17%	**11.28%**	4.20%	**68.51%**
75%-25% stock-bond	11.44%	2.00%	7.31%	**17.49%**	7.17%	**95.91%**
Aggressive	15.37%	2.92%	9.04%	**18.97%**	7.76%	**104.53%**
100% stock	19.02%	4.44%	12.20%	**23.33%**	9.72%	**123.74%**

Note:
This table ranks investment strategies according to the normalized solvency capital (fifth column) and the normalized expected return (seventh column) for the high level guarantee ($g = 4\%$ p.a.) with periodic contributions (€1200 up-front annually). All results are computed for the MS(1-2) model. The top section represents the 10-year contract, the middle section the 20-year contract, and the bottom section the 30-year contract. SEV denotes the shortfall expected value, SSD the shortfall standard deviation, and MEL the mean excess loss.
* – for contracts with a 10-year term, the moderate life-cycle strategy invests in the 10%-90% stock-bond portfolio. ** – for contracts with a 10-year term, the naive life-cycle strategy invests in the 25%-75% stock-bond portfolio. *** – for contracts with a 10-year term, the 100-x life-cycle strategy invests in the 40%-60% stock-bond portfolio.

ever, it should forbid selling guarantees which require the solvency capital above some arbitrary upper boundary. This would not punish a conservative risk policy and prevent taking too much risk into books.

Having arrive at the end of this Section, it is worth noting that the value at risk is not a suitable solvency measure. Table 5.1 shows that in the case of the 0% guarantee with a 30-year contract term, the 75%-25% stock-bond strategy has the shortfall probability of 0.83% and the normalized shortfall expected value of 0.05%. This corresponds to the value at risk of 0.05% at the confidence level of 99.2%. However, if the shortfall would occur, the guarantee provider suffers an average loss of 6.40% (see normalized MEL). This example shows that the value at risk can lead to the underfunding of the risk position of the guarantee provider and thus jeopardize his existence. One could argue that the guarantor could then try to acquire the capital on the market in order to cover the loss. Neverless, this scenario is more likely to occur in times of financial market distress, when the guarantee provider could have difficulties to aquiring the necessary capital.

5.6 Interrelation between solvency requirements and expected profit

The previous Section has shown that several investment strategies do not require any solvency capital. Thus, as one possible strategy, the guarantee seller can choose contracts which do not require any solvency capital and offer these to clients. However, a future pensioner is interested in three aspects when buying a guarantee: the protection level (level of the guarantee), the reduction of the cost of protection, and the maximization of profit. Thus, it might happen that an investment strategy that does not require any solvency capital would be unsatisfactory from the client's point of view. For instance, its expected profit would be too low. As we discuss several investment strategies, we would like to see which one of these is most suitable for a client and most likely to satisfy his goals and still be financeable in terms of solvency capital, for the guarantee provider. To compare the profit potential of different investment strategies, we use the normalized expected profit as defined in Definition 4.53.

Tables 5.1 and 5.2 rank chosen investment guarantees from the lowest to the highest solvency capital (i.e. normalized MEL). If two strategies have the same solvency capital, the better ranking is given to the one with the higher expected profit. We thereby assume that the client knows his investment horizon (10, 20 or 30 years) and how much risk protection he needs (guarantee of 0% or 4%). Besides the solvency capital, Tables 5.1 and 5.2 show the remaining risk measures (i.e., shortfall probability, normalized shortfall expected value, and normalized shortfall standard deviation[4]) and normalized cost computed in previous Chapter.[5]

In the following we will discuss the 0% guarantee at the 30-year investment horizon example. Results for the 20-year horizon are similar. Results for the 10-year horizon are similar to results for the 4% guarantee, which will be discussed later in this section. Six investment strategies (the zero-bond, the conservative, and buy-and-hold strategies with up to 25% stock proportion) do not require any solvency capital. From all of them the zero-bond investment strategy has the highest normalized profit of 86.66%. In com-

[4]Normalized conditional shortfall standard deviation was postponed as it does not yield results for all studied investment strategies within the 20 million Monte Carlo runs.

[5]All risk measures are based on the 20 million simulations, and the normalized cost and the normalized expected profit are based on 1 million simulations in order to provide a convergence of results.

parison, the pure bond strategy has a normalized expected profit of 33.44%. This shows that it is possible to choose an investment strategy that does not require any solvency reserves. Moreover, it does matter which of these strategies the guarantee provider chooses, as an unsuitable strategy can significantly lower the profit the customer can expect.

Surprisingly, a higher normalized expected profit yields the most conservative investment strategy: the zero-bond strategy.

Another interesting result is that several investment strategies (the moderate, the naive, the 100-x rule, and the 50%-50% stock-bond strategy) require a positive solvency capital. However, these yield a lower expected return than the zero-bond strategy. For example, the 50%-50% stock-bond buy-and-hold strategy requires 3.72% of the net present value of contributions as solvency capital and has a normalized expected value of 68.51%. First, far riskier strategies yield higher expected return than the zero-bond strategy. For instance, the aggressive strategy requires a solvency capital of 6.29% and yields a normalized expected profit of 104.53%.

Last, we discuss the case of the 4% guarantee using the example of a 30-year time horizon (see Table 5.2). Results for other investment horizons, e.g. for 10 and 20 years are similar. The zero-bond strategy is the only one which does not require any solvency capital. However, it yields a normalized profit of only 32.20% which is rather unsatisfactory for the investment horizon of 30 years. This low expected profit is a result of the high guarantee level which requires that a high portion of capital has to be invested in a riskless zero-bond in order to fulfill the guarantee of 4%. This lead us to the conclusion that a higher protection (4% p.a. instead of 0% p.a.) results in an opportunity cost of a lower expected profit. As one of the goals of pension saving is the maximization of consumption in old age, this investment strategy is rather unfavorable for the customer. The increase of the expected profit is equivalent to the increase of the solvency capital. For instance, the pure bond buy-and-hold strategy yields the normalized expected profit of 33.44% (which is only 1.24 per cent points higher in comparison to the zero-bond strategy) but it requires a solvency capital of 3.66%.

5.7 Conclusion

This Chapter analyzes seven factors which influence the risk of an investment guarantee: the initial state (i.e., the market phase at the inception of the contract), the model governing the stochastic behavior of the investment portfolio, the guarantee level, the investment strategy, time, the contract term, and the contribution payment scheme. We have discussed five risk measures: the shortfall probability, the shortfall expected value, the shortfall standard deviation, the mean excess loss, and the conditional shortfall standard deviation. The most interesting one is the MEL, as it is a worst case risk measure. Thus, we propose using it as a quantification of the solvency capital the guarantee provider should accumulate.

We maintain that the GBM model overestimates the risk associated with the investment guarantee with the comparison to the GBM with Markov switching. The GBM does not differentiate between the low and high volatility phase (i.e. low and high market risk phase). A glance at the average regime duration (see Tables B.1-B.13 in Appendix B) or smoothed probabilities (see Figures D.1-D.39 in Appendix D), however, shows that the stable market phases are longer than the high volatility phases. As the GBM does not differentiate between the low and the high risk market phase, it only provides an average volatility. Furthermore, it does not account for the fact that the high risk market phases occur seldomly and thus lead to an overestimation of their impact on the shortfall risk.

The Markov switching model is very sensitive to the probability of the initial state for short-term contracts. This effect is less significant for middle- and long-term contracts. The risk is higher when the product is sold in the high volatility market phase. Unfortunately, the market state cannot be observed. However, smoothed probabilities are very good proxies for the initial state probabilities. Additionally, there are two ways the risk-averse agent can manage the risk associated with the uncertainty about the initial state. First, he can restrict himself to selling only guarantees with contract terms above 10 years. This does not pose a problem, as retirement saving products are generally middle and long-term products, which are predominantly sold on the market anyway. Second, the upper bounds of all risk measures discussed in this thesis are always associated with the high volatility regime at the contract inception. The guarantee provider can therefore assume this to be the case and use the results as the conservative risk measure.

Another risk factor we studied was the contribution payment scheme. We found out that the impact of this factor varies strongly with respect to other risk factors, e.g. the investment strategy underlying the guarantee, the guarantee level, time/contract term, and the risk measure used.

Three further risk factors, namely the guarantee level, the stock proportion in the investment strategy, and time/contract term are tightly connected with each other. Accordingly, manipulating one of the above three factors can achieve a risk reduction of the same amount. Thus, they should always be considered together. The impact of the guarantee level and the stock proportion is always the same: the higher the guarantee level and/or the higher the stock proportion, the higher the risk of the guarantee, regardless of which risk measure is used. Furthermore, guarantee levels significantly lower than the risk-free interest rate (e.g. $g = -2\%, 0\%$ and 2% p.a.) have a different impact on the risk level and result in a different shape of the risk measure curve than guarantee levels closer to the risk-free interest rate (e.g. $g = 4\%$ p.a.).

On the contrary, the impact of time/contract term is different for different risk measures. The shortfall probability decreases if time/contract term increases regardless of the guarantee level and stock proportion. In most cases, the normalized shortfall expected value decreases as time/contract term increases. The exception are high level guarantees with a high stock proportion. In these cases, the normalized shortfall expected value function has an inverted U shape. The normalized shortfall standard deviation is a decreasing function of time/contract term when the stock proportion is low. For a high stock proportion it increases to a maximum and then decreases. This holds true for both the low and the high guarantee levels. In the case of low level guarantees, the normalized MEL is a decreasing function of time/contract term if the stock proportion is low, and an inverted U shape function if the stock proportion is high. For the high level guarantees, the normalized MEL is an increasing function of time/contract term. The normalized conditional shortfall standard deviation can be only computed for investment strategies with a high stock proportion. In this case, it is an increasing function of time/contract term.

These results can be summed up in the following manner. Risk measures we use here are more sensitive to changes in time/contract term, the higher the order of the lower partial or conditional lower partial moment is. Furthermore, conditional lower partial moments are more sensitive with respect to time/contract term than lower partial moments. In addition, it is vital that the risk manager takes into account that all risk measures used here can be interpreted differently. The most informative ones seem to be the probability of shortfall and the MEL, which can be interpreted as the required solvency capital.

In conclusion, the risk of low level guarantees can be reduced to an acceptable level for middle and high contract terms, regardless of the stock proportion in the investment portfolio. The risk of high level guarantees is only acceptable for middle and high contract terms only if the stock proportion is low. Since the risk is rather high for a high stock proportion, such products should not be offered to clients as it could significantly jeopardize the existence of the guarantee provider, which is evidently not in the interest of guarantee buyers. Instead of selling guarantees backed by investment strategies with a high stock proportion, the guarantee provider can sell guarantees backed by a life-cycle strategy. The idea of a life-cycle strategy is to invest in more risky assets at the contract inception and to reduce the stock proportion as the contract nears its expiration date. This approach enables the guarantee provider to construct a product which fits the expectations of his client. This means that, for a given contract term, guarantee level, risk level, and/or solvency capital one can find a life-cycle strategy which maximizes the expected profit of the individual pension account.

Apart from the buy-and-hold and life-cycle strategies, the zero-bond strategy also constitutes a very interesting option for the risk manager. The main idea is to invest the portion of the contribution needed to fulfill the guarantee in a risk-free zero-bond and the remaining capital in stocks. This leads to a risk-free product. This strategy is very interesting for the guarantee provider, as it does not require any solvency capital (under the assumption that the seller of the zero-bond cannot default). A very interesting result is that, for low level guarantees, the expected profit of the zero-bond strategy outperforms the expected profit of several buy-and-hold and life-cycle strategies which have a positive risk (and positive solvency capital requirement). This does not hold true for high level guarantees when the zero-bond strategy is the only riskless (and solvency capital free) one. Thus, the guarantee provider has the choice between selling low level guarantees backed with a zero-bond and those backed with one of the life-cycle strategies with a positive – but still acceptable – risk (solvency capital) and an expected profit which is higher than the expected profit of the zero-bond strategy. If the guarantee provider would like to sell high level guarantees backed with a zero-bond strategy, he should be aware of the fact that, while being a risk-free product, it yields a very low expected profit. Thus, it is doubtful that it would be in the interest of the client who buys an individual pension product. The guarantee provider should instead back high level guarantees with a life-cycle strategy. However, he should be aware that these are very risky and thus require a high solvency capital.

Last but not least, we discussed the solvency capital requirements for investment guarantees embedded in personal pension plans. We have shown in an example that the value at risk which is used in the banking industry to quantify solvency requirements, can lead to an underestimation of the solvency capital and thus should not be used in this context. Instead, we proposed applying the mean excess loss for this purpose. In our opinion, the solvency supervising authority should allow the guarantee provider to hold solvency capital according to his risk exposure and set a maximal allowed risk position in order to prevent too risky behavior of guarantee providers.

Chapter 6

Main results and further research

6.1 Main results

This thesis discusses how to price, how to measure risk, and how to quantify solvency capital for investment guarantees embedded in individual pension products. The first main contribution of this thesis is to implement a model with a stochastic volatility of the return rate of risky assets backing the guarantee. This is all the more important since the saving process of a pension product lasts for several decades. Furthermore, it is unreasonable to assume that the risky asset backing the guarantee follows the GBM. We decided to use the Markov regime model among several other stochastic volatility models as it has a very appealing intuitive interpretation. This model takes into account the stylized fact that the financial market reveals two phases: the bull and the bear market phase.

Since the Markov switching model we use is not as well-known as the GBM, we discussed how to estimate and test its parameters. Thus, the second main contribution of this thesis is a discussion on the suitability of the MS model for the German financial market (stock and interest rate). Furthermore, to the best of our knowledge, we are the first to implement suitable statistical tests in order to test the null hypothesis of the MS model against the arithmetic Brownian motion/Vasiček model in the context of financial markets. Other authors only use the AIC and BSC or the likelihood ratio test (see, e.g., Hardy (2001)). Particularly, the third one is unsuitable for testing this null hypothesis, as Markov models violate certain crucial assumptions of the likelihood ratio test. Instead, we implemented tests proposed by Hamilton (1996) and Garcia (1998).

We found that the GBM with Markov switching better describes the stochasticity of Germany stocks than the GBM. Moreover, the Vasiček process with Markov switching better describes the stochasticity of the German interest rates.[1] Our results are very robust as we used several tests designed especially for MS models, all of which provided similar results.

The price one has to pay for using MS models is an estimation and testing procedure, as well as an option pricing theory which are more complex than those of the GBM model. However, we should take into consideration that pension saving products have contract durations of many years and that the differences especially of the guarantee cost within both approaches are significant. In our opinion, this additional effort is therefore worth making.

Third, the usage of the MS model implies the incompleteness of the financial market. This affects option pricing since several martingale risk measures are possible in the arbitrage-free market. There is a common consensus that

[1]We use GBM with Markov switching for the bond portfolio, as there is no option pricing theory for the Vasiček process with Markov switching. This simplification, however, should not influence the results to a noteworthy degree.

the market "chooses" the "right" martingale measure. However, guarantees discussed in this thesis are not traded on the market, therefore, their prices cannot be observed. Accordingly, the guarantee provider has to make a suitable choice concerning the equivalent probability measure based, among others, on his risk aversion. We decided to opt for the Esscher measure which is well-known in actuarial science. Reasons for this choice are fourfold: (1) The process under the Esscher martingale measure $\mathcal{Q}$ remains in the same class of models as the process under the real-word probability measure $\mathcal{P}$. (2) The solution reduces to the well-known Black and Scholes (1973) formula in the case of one switching regime (i.e. $K = 1$). (3) The Esscher transform approach is conform with maximizing the expected utility with the constant risk aversion utility function $u(x) = \frac{x^{\gamma}}{\gamma}$ $(0 < \gamma < 1)$. (4) The Esscher probability measure allows us to price the uncertainty of whether the market is in a stable or in a turbulent phase.

We compared these results with those of the Black-Scholes and Bollen-Hardy model. The comparison has shown that the difference between the Back-Scholes and the Esscher model can be explained by two factors: the stochastic volatility and the unobservable state variable. The latter factor has a stronger impact.

Fourth, we proposed measuring the risk of the guarantee with several lower partial and conditional lower partial moments. In our opinion, they are better able to quantify the risk of the discussed guarantees than dispersion or quantile risk measures (especially the value at risk widely used in the banking industry). The reason for this is that they quantify the risk which is defined as an unfavorable deviation from a target (e.g. guaranteed portfolio wealth). By this means, they take into account the intuitive difference between "risk" and "chance".

Fifth, we discussed the solvency capital requirements for investment guarantees embedded in personal pension plans. We have shown in an example that the value at risk, which is used in the banking industry to quantify solvency requirements, can lead to an underestimation of the solvency capital and thus should not be used in this context. Instead, we proposed applying the mean excess loss for this purpose. In our opinion, the solvency supervising authority should allow the guarantee provider to hold solvency capital according to his risk exposure and set a maximal allowed risk position in order to prevent too risky behavior of guarantee providers.

Sixth, we found that the GBM overestimates the cost and risk of the guarantee in comparison to the GBM with Markov switching. This result is rather suprising, since we added an additional source of uncertainty to the model: the market state variable, which describes whether the market is in a stable or in a turbulent phase. However, a descriptive analysis of the MS model shows that stable phases (i.e. phases with a low volatility σ_1) last significantly longer on average than in turbulent phases (i.e. phases with a high volatility σ_2). On the other hand, the GBM only has one market phase with volatility σ_{GBM} such that $\sigma_1 < \sigma_{GBM} < \sigma_2$. Thus, it inevitable overestimates the influence of the turbulent market phase on the outcomes, which also explains why the GBM provides biased results.

Seventh, we discussed several factors which influence the cost and the risk of an investment guarantee. From the point of view of the risk manager, three of these factors are of particular importance: the guarantee level, the stock proportion, and time/contract term. Making a suitable choice among these factors enables one to control the risk associated with the guarantee.

Eighth, we discussed several investment strategies the guarantee provider can use. We showed that life-cycle strategies are more suitable than buy-and-hold strategies to control the cost and risk of the guarantee and/or increase its expected profit. A suitable management of these variables can give the guarantee provider a competitive advantage, since the purchaser of an individual pension account is interested in both a high safety for his savings and in a maximization of the expected profit. The life-cycle strategies we discussed reduce the volatility of the investment portfolio as the contract nears its date of expiration, regardless of the market phase. An interesting topic for further research would be studying investment strategies which take the market phase into account. This would imply investing more risky during the bull market phase and less risky during the bear market phase. We decided against implementing such strategies, as such a program code would be too time-consuming in GAUSS. We suppose, however, that it would be feasible in other programing languages such as C++.

Ninth, we discussed the zero-bond strategy. The idea of this strategy is to invest a portion of the contribution needed to fulfill the guarantee in a riskless zero-bond, in order to provide a costless/riskless guarantee. The remaining capital is to be invested in stocks in order to maximize the expected profit. It turned out that this strategy performs very well for low level guarantees. With respect to the expected profit, it even outperforms several buy-and-hold and life-cycle strategies with positive cost, positive risk, or positive solvency capital. For the high level guarantees there exists only one costless (riskless, and solvency capital free) strategy and it could theoretically be implemented by a risk averse guarantee provider. However, the expected profit of the zero-bond strategy decreases along with an increase of the guarantee level. For instance, a 4% guarantee backed with a zero-bond strategy is completely uninteresting with respect to the expected profit. Therefore, it is doubtful if such a product could be successful in the market. Instead, the guarantee provider should concentrate on low level guarantees or use a more risky investment strategy. In our opinion, the first solution is more advisable as high level guarantees with a satisfactory expected profit are expensive, risky, and require a high solvency capital. Lastly, we should discuss some practical constraints of the zero-bond strategy. It assumes that the portfolio manager can purchase a riskless zero-bond with every desired maturity. However, such zero-bonds migth not be available in the market, particularly for very long maturities.

6.2 Further research

In our model we have made several assumptions. Relaxing them would provide an opportunity for further research. Below, we will discuss several of them, some of which we are working on already. First, we have studied deterministic guarantees. As we mentioned in Chapter 1, there are several guaranteed pension products all over the world containing stochastic guarantees, e.g., a stochastic benchmark portfolio or inflation. Such guarantees could be priced as the Margrabe (1978) option to exchange one risky asset for another.

Second, we omitted the mortality risk, which could easily be implemented using the maturity tables. We intentionally discarded this solution as it would merely constitute a sum of option prices (risk measures or solvency requirements) weighted by the mortality probabilities, since the market and biometric risks are independent from each other. While being able to quantify the mortality risk, however, this approach does not contribute to understanding the stochasticity of this type of risk. In our opinion, it would be much more interesting to use one of modern stochastic mortality models, e.g. Dahl

(2004) or Cox and Lin (2005).

Third, we do not take the credit risk into account. Adding it to the model would provide some interesting results. At this point, one could use the approach proposed by Schönbucher (2000).

Fourth, the model could be extended by the stochastic risk-free interest rate, e.g., the Vasiček (1977), the Cox, Ingersoll, and Ross (1985), or the Heath, Jarrow, and Morton (1992) model could be implemented.

Fifth, the choice of the Esscher martingale measure does not allow for hedging. As this is a very important issue for a manager of option risk, a stochastic volatility model which allows to construct a hedging strategy would be very interesting. At this point, e.g., the mean-variance approach, which was introduced by Föllmer and Sondermann (1986) and Föllmer and Schweizer (1988) could be used.

Sixth, we assumed the guarantee to be an European-option-type. An extension of the model with a surrender American-option-typed guarantee would be interesting. For instance, one could use the Longstaff and Schwartz (2001) least-squares model.

Finally, it would be interesting to study the performance of the guarantees we discussed in this thesis within a stochastic volatility model. In this case, performance measures such as the Sortino, the Omega, or the Psi ratio could be implemented, see Keating and Shadwick (2002a, 2002b) and Sortino and Satchell (2001).

Appendix A

Proofs for Chapter 3

A.1 Assumptions for the Hansen test

Assumption A.1

$$\sup_{\boldsymbol{\alpha}^H \in A} \sqrt{T} \| D(\boldsymbol{\alpha}^H) \| = O_p(1), \tag{A.1}$$

where $D(\boldsymbol{\alpha}^H) = \widehat{\boldsymbol{\gamma}}_1(\boldsymbol{\alpha}^H) - \boldsymbol{\gamma}_1(\boldsymbol{\alpha}^H)$.

Assumption A.2

$$\sup_{\boldsymbol{\alpha}^H \in A,\ \boldsymbol{\gamma}_1 \in \Gamma} \| M_T(\boldsymbol{\alpha}^H, \boldsymbol{\gamma}_1) \| = O_p(T), \tag{A.2}$$

with

$$M_T(\boldsymbol{\alpha}^H, \boldsymbol{\gamma}_1) = \frac{\partial^2}{\partial \boldsymbol{\gamma}_1 \boldsymbol{\gamma}_1} L_T(\boldsymbol{\alpha}^H, \boldsymbol{\gamma}_1).$$

Assumption A.3 *Assume now that* $Q_T^*(\boldsymbol{\alpha}^H)$ *satisfies an empirical process law:*

$$Q_T^*(\boldsymbol{\alpha}^H) \rightarrow_p \frac{Q(\boldsymbol{\alpha}^H)}{\sqrt{V(\boldsymbol{\alpha}^H)}} := Q^*(\boldsymbol{\alpha}^H), \tag{A.3}$$

where $Q^*(\boldsymbol{\alpha}^H)$ *is a zero mean normal with the covariance function*

$$K^*(\boldsymbol{\alpha}_1^H, \boldsymbol{\alpha}_2^H) = \frac{K^*(\boldsymbol{\alpha}_1^H, \boldsymbol{\alpha}_2^H)}{\sqrt{V(\boldsymbol{\alpha}_1^H) V(\boldsymbol{\alpha}_2^H)}}\ .$$

Assumption A.4 *Let*

1. A^G *and* Δ *are compact.*

2. $Q(\boldsymbol{\alpha}^G, \boldsymbol{\delta}) = \lim_{T\to\infty} \mathbb{E}Q_T(\boldsymbol{\alpha}^G, \boldsymbol{\delta})$ *is continuous in* $(\boldsymbol{\alpha}^G, \boldsymbol{\delta})$ *uniformly over* $A^G \times \Delta$.

3. $Q_T(\boldsymbol{\alpha}^G, \boldsymbol{\delta}) \to_p Q(\boldsymbol{\alpha}^G, \boldsymbol{\delta})$ *for all* $(\boldsymbol{\alpha}^G, \boldsymbol{\delta}) \in A^G \times \Delta$.

4. $Q_T(\boldsymbol{\alpha}^G, \boldsymbol{\delta}) - Q(\boldsymbol{\alpha}^G, \boldsymbol{\delta})$ *is stochastically equi-continuous in* $(\boldsymbol{\alpha}^G, \boldsymbol{\delta})$ *over* $A^G \times \Delta$.

5. *For all* $\boldsymbol{\delta} \in \Delta$, $Q(\boldsymbol{\alpha}^G, \boldsymbol{\delta})$ *is uniquely maximized over* $\boldsymbol{\alpha}^G \in A^G$ *at* $\boldsymbol{\alpha}_0^G$.

Assumption A.5 *For* $\boldsymbol{\alpha}^G \in A_0^G = \{\boldsymbol{\alpha}^G \in A^G : \ h(\boldsymbol{\alpha}^G) = 0\}$, $Q_T(\boldsymbol{\alpha}^G, \boldsymbol{\delta})$ *does not depend upon* $\boldsymbol{\delta}$.

Assumption A.6 *Let*

1. $M(\boldsymbol{\alpha}^G, \boldsymbol{\delta}) = \lim_{T\to\infty} \mathbb{E}[M_T(\boldsymbol{\alpha}^G, \boldsymbol{\delta})]$ *and* $V(\boldsymbol{\alpha}^G, \boldsymbol{\delta}) = \lim_{T\to\infty} \mathbb{E}[S_T^c(\boldsymbol{\alpha}^G, \boldsymbol{\delta})\ S_T^c(\boldsymbol{\alpha}^G, \boldsymbol{\delta})']$ *are continuous in* $(\boldsymbol{\alpha}^G, \boldsymbol{\delta})$ *uniformly over* $A_\epsilon^G \times \Delta$, *where* $\boldsymbol{\alpha}_\epsilon^G$ *is some neighbourhood of* $\boldsymbol{\alpha}_0^G$.

2. $[M_T(\boldsymbol{\alpha}^G, \boldsymbol{\delta}), V_T(\boldsymbol{\alpha}^G, \boldsymbol{\delta})] \to_p [M(\boldsymbol{\alpha}^G, \boldsymbol{\delta}), V(\boldsymbol{\alpha}^G, \boldsymbol{\delta})]$ *for all* $(\boldsymbol{\alpha}^G, \boldsymbol{\delta}) \in A_\epsilon^G \times \Delta$.

3. $M_T(\boldsymbol{\alpha}^G, \boldsymbol{\delta}) - M(\boldsymbol{\alpha}^G, \boldsymbol{\delta})$ *and* $V_T(\boldsymbol{\alpha}^G, \boldsymbol{\delta}) - V(\boldsymbol{\alpha}^G, \boldsymbol{\delta})$ *are stochastically equi-continuous in* $(\boldsymbol{\alpha}^G, \boldsymbol{\delta})$ *over* $A_\epsilon^G \times \Delta$.

4. $M(\boldsymbol{\delta}) = M(\boldsymbol{\alpha}_0^G, \boldsymbol{\delta})$ *and* $V(\boldsymbol{\delta}) = V(\boldsymbol{\alpha}_0^G, \boldsymbol{\delta})$ *are positive definite uniformly over* $\boldsymbol{\delta} \in \Delta$.

5. $\sqrt{T} S_T^c(\boldsymbol{\alpha}_0^G, \boldsymbol{\delta}) \xrightarrow{p} S^c(\boldsymbol{\delta})$ *on* $\boldsymbol{\delta} \in \Delta$, *where* $S^c(\cdot)$ *is a zero-mean Gaussian process with the covariance function* $K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2) = \lim_{T\to\infty} T\mathbb{E}[S_T^c(\boldsymbol{\alpha}_0^G, \boldsymbol{\delta}_1), S_T^c(\boldsymbol{\alpha}_0^G, \boldsymbol{\delta}_2)']$,

where $\xrightarrow{p}$ denotes a weak convergence of probability measures with respect to the uniform metric.

A.2 Proof of Lemma 2

Proof of Lemma 3.3 from Section 3.6.2.

Knowing the log-likelihood function for observation y_t

$$l_t(y_t|\mathscr{Y}_T;\boldsymbol{\theta}^*) = \sum_{Z_t(\boldsymbol{\delta})=1}^{2} \left\{ -\frac{1}{2}\ln[2\pi] - \frac{1}{2}\ln\left[\left(\sigma_1 + \sum_{j=2}^{K}\sigma_j^*\mathbb{I}_{[Z_t=j]}\right)^2\right] \right.$$
$$\left. - \frac{\varepsilon_t^2}{2\left(\sigma_1 + \sum_{j=2}^{K}\sigma_j^*\mathbb{I}_{[Z_t=j]}\right)^2} \right\} p_t, \tag{A.4}$$

with the increment

$$\varepsilon_t = y_t - (\mu_1 + \sum_{j=2}^{K}\mu_j^*\mathbb{I}_{[Z_t=j]} + \sum_{i=1}^{r}\left(\phi_{i(1)} + \sum_{j=2}^{K}\phi_{i(j)}^*\mathbb{I}_{[Z_t=i]}\right) y_{t-i})$$

and the smoothed probability

$$p_t = \Pr[Z_t(\boldsymbol{\delta}) = z_t(\boldsymbol{\delta})|\mathscr{Y}_T;\boldsymbol{\alpha}^G,\boldsymbol{\delta}]$$

computing derivatives $\frac{\partial l_t\left(y_t|\mathscr{Y}_T;\boldsymbol{\theta}^*\right)}{\partial \boldsymbol{\alpha}^G}$ becomes fairly straightforward

$$\frac{\partial l_t(y_t|\mathscr{Y}_T;\boldsymbol{\theta}^*)}{\partial \mu_1} = \frac{\varepsilon_t}{(\sigma_1 + \sigma_2^*\mathbb{I}_{[Z_t(\boldsymbol{\delta})=2]})^2} p_t \tag{A.5}$$

$$\frac{\partial l_t(y_t|\mathscr{Y}_T;\boldsymbol{\theta}^*)}{\partial \mu_2^*} = \frac{\varepsilon_t\mathbb{I}_{[Z_t(\boldsymbol{\delta})=2]}}{(\sigma_1 + \sigma_2^*\mathbb{I}_{[Z_t(\boldsymbol{\delta})=2]})^2} p_t \tag{A.6}$$

$$\frac{\partial l_t(y_t|\mathscr{Y}_T;\boldsymbol{\theta}^*)}{\partial \phi_{i(1)}} = \frac{y_{t-i}\varepsilon_t}{(\sigma_1 + \sigma_2^*\mathbb{I}_{[Z_t(\boldsymbol{\delta})=2]})^2} p_t, \qquad i = 1,\dots,r \tag{A.7}$$

$$\frac{\partial l_t(y_t|\mathscr{Y}_T;\boldsymbol{\theta}^*)}{\partial \phi_{i(2)}^*} = \frac{y_{t-i}\varepsilon_t\mathbb{I}_{[Z_t(\boldsymbol{\delta})=2]}}{(\sigma_1 + \sigma_2^*\mathbb{I}_{[Z_t(\boldsymbol{\delta})=2]})^2} p_t, \qquad i = 1,\dots,r \tag{A.8}$$

$$\frac{\partial l_t(y_t|\mathscr{Y}_T;\boldsymbol{\theta}^*)}{\partial \sigma_1^2} = \frac{1}{2\sigma_1(\sigma_1 + \sigma_2^*\mathbb{I}_{[Z_t(\boldsymbol{\delta})=2]})} \left(\frac{\varepsilon_t^2}{(\sigma_1 + \sigma_2^*\mathbb{I}_{[Z_t(\boldsymbol{\delta})=2]})^2} - 1\right) p_t \tag{A.9}$$

$$\frac{\partial l_t(y_t|\mathscr{Y}_T;\boldsymbol{\theta}^*)}{\partial \sigma_2^{*2}} = \frac{\mathbb{I}_{[Z_t(\boldsymbol{\delta})=2]}}{2\sigma_2^*(\sigma_1+\sigma_2^*\mathbb{I}_{[Z_t(\boldsymbol{\delta})=2]})}\left(\frac{\varepsilon_t^2}{(\sigma_1+\sigma_2^*\mathbb{I}_{[Z_t(\boldsymbol{\delta})=2]})^2}-1\right)p_t. \quad \text{(A.10)}$$

Now, it remains to use the average log-likelihood of the full sample which is given by

$$\overline{\mathscr{L}}_T(\boldsymbol{\theta}^*) = \frac{1}{T}\sum_{t=1}^{T}\sum_{Z_t(\boldsymbol{\delta})=1}^{K} l_t(y_t|\mathscr{Y}_T;\boldsymbol{\theta}^*) \quad \text{(A.11)}$$

and its score

$$\frac{\partial \overline{\mathscr{L}}_T(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\alpha}^G} = \frac{1}{T}\sum_{t=1}^{T}\sum_{Z_t(\boldsymbol{\delta})=1}^{K}\frac{\partial l_t(y_t|\mathscr{Y}_T;\boldsymbol{\theta}^*)}{\partial \boldsymbol{\alpha}^G}. \quad \text{(A.12)}$$

Subsequently, the scores (3.41)-(3.46) follow. ■

Proof of Lemma 3.4 from Section 3.6.2.

In this proof the elements of the $K(\boldsymbol{\delta}_1,\boldsymbol{\delta}_2)$ matrix are computed.[1] At the beginning, one should bear in mind that the matrix $K(\boldsymbol{\delta}_1,\boldsymbol{\delta}_2)$ is the covariance process of the test statistic under the null hypothesis. Thus it is useful to note that the variance of the Markov switching process equals

$$\left(\sigma_1+\sigma_2^*\mathbb{I}_{[Z_t(\boldsymbol{\delta})=2]}\right)^2 = \sigma_1^2 \quad \text{(A.13)}$$

regardless of the value of the nuisance parameter vector $\boldsymbol{\delta}$.

Then, according to equations (3.38), (3.41) and (A.13) the (μ_1,μ_1) element of the covariance matrix can be computed as

$$K(\boldsymbol{\delta}_1,\boldsymbol{\delta}_2)_{\mu_1,\mu_1} = \lim_{T\to\infty} T\mathbb{E}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{Z_t(\boldsymbol{\delta}_1)=1}^{2}\sum_{Z_s(\boldsymbol{\delta}_2)=1}^{2}\frac{1}{T^2}\frac{\varepsilon_t}{\sigma_1^2}\frac{\varepsilon_s}{\sigma_1^2}p_t p_s, \quad \text{(A.14)}$$

where

$$p_t = \Pr[Z_t(\boldsymbol{\delta}_1)=z_t(\boldsymbol{\delta}_1)|\mathscr{Y}_T;\boldsymbol{\alpha}_0^G,\boldsymbol{\delta}_1]$$

$$p_s = \Pr[Z_s(\boldsymbol{\delta}_2)=z_s(\boldsymbol{\delta}_2)|\mathscr{Y}_T,z_s(\boldsymbol{\delta}_1);\boldsymbol{\alpha}_0^G,\boldsymbol{\delta}_1].$$

Note that p_s is dependent on $z_s(\boldsymbol{\delta}_1)$. This results from the fact that filters based on $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$ are not independent because they are both derived from the same series $\mathscr{Y}_t$.

[1] The proof is analogical to the proof of Garcia (1998, Appendix 2, p. 781-785).

As the sums of the products of the probabilities p_t and p_s, respectively, are equal to unity, one has

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_1,\mu_1} = \lim_{T\to\infty} \mathbb{E}T \sum_{t=1}^{T} \sum_{s=1}^{T} \frac{1}{T^2} \frac{\varepsilon_t \varepsilon_s}{\sigma_1^4}. \tag{A.15}$$

Since the increments ε are serially independent and under the null hypothesis $\mathbb{E}[\epsilon_t^2] = \sigma_1^2$, it follows that

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_1,\mu_1} = \lim_{T\to\infty} T \sum_{t=1}^{T} \frac{1}{T^2} \frac{\mathbb{E}[\varepsilon_t^2]}{\sigma_1^4} = \lim_{T\to\infty} \frac{1}{\sigma_1^2} = \frac{1}{\sigma_1^2}. \tag{A.16}$$

Now, compute the (μ_1, μ_2^*) element of the covariance matrix. From equations (3.38), (3.41) and (3.42) it follows that

$$\begin{aligned} K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_1,\mu_2^*} = \lim_{T\to\infty} T\mathbb{E} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{Z_t(\boldsymbol{\delta}_1)=1}^{2} \sum_{Z_s(\boldsymbol{\delta}_2)=1}^{2} \\ \frac{1}{T^2} \frac{\varepsilon_t}{(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_t(\boldsymbol{\delta_1})=2]})^2} \frac{\varepsilon_s \mathbb{I}_{[Z_s(\boldsymbol{\delta_2})=2]}}{(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_s(\boldsymbol{\delta_2})=2]})^2} p_t p_s \end{aligned} \tag{A.17}$$

From the fact that p_t sum to unity and ε are serially independent, it follows that

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_1,\mu_2^*} = \lim_{T\to\infty} T\mathbb{E} \sum_{s=1}^{T} \sum_{Z_s(\boldsymbol{\delta}_2)=1}^{2} \frac{1}{T^2} \frac{\varepsilon_s^2 \mathbb{I}_{[Z_s(\boldsymbol{\delta_2})=2]}}{(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_s(\boldsymbol{\delta_2})=2]})^4} p_s. \tag{A.18}$$

As ε and $z(\boldsymbol{\delta})$ are independent, therefore this equation can be rewritten as

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_1,\mu_2^*} = \lim_{T\to\infty} T\mathbb{E} \left[\sum_{s=1}^{T} \frac{1}{T^2} \frac{\varepsilon_s^2}{(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_s(\boldsymbol{\delta_2})=2]})^4} \mathbb{E}\left[\mathbb{I}_{[Z_s(\boldsymbol{\delta_2})=2]} | \mathscr{Y}_T, \boldsymbol{\alpha}_0^G, \boldsymbol{\delta}_2 \right] \right]. \tag{A.19}$$

Then apply the law of iterated expectations

$$\mathbb{E}\left[\mathbb{E}[\mathbb{I}_{[Z_s(\boldsymbol{\delta_2})=2]} | \mathscr{Y}_T, \boldsymbol{\alpha}_0^G, \boldsymbol{\delta}_2] \right] = \mathbb{E}[\mathbb{I}_{[Z_s(\boldsymbol{\delta_2})=2]}] \tag{A.20}$$

and note that $\mathbb{E}[\mathbb{I}_{[Z_s(\boldsymbol{\delta_2})=2]}] = \pi_2(\boldsymbol{\delta}_2)$ and use the fact that under the null hypothesis $\mathbb{E}[\varepsilon_s] = \sigma_1^2$ and (A.13), which provides the following equation

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_1,\mu_2^*} = \lim_{T\to\infty} \frac{\pi_2(\boldsymbol{\delta}_2)}{\sigma_1^2} = \frac{\pi_2(\boldsymbol{\delta}_2)}{\sigma_1^2}. \tag{A.21}$$

Analogically the (μ_2^*, μ_1) element of the covariance matrix equals

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_2^*, \mu_1} = \frac{\pi_2(\boldsymbol{\delta}_1)}{\sigma_1^2}. \tag{A.22}$$

To compute the (μ_2^*, μ_2^*) element of the covariance matrix one has to use equations (3.38) and (3.42), from which follows that

$$\begin{aligned} K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_2^*, \mu_2^*} = &\lim_{T\to\infty} T\mathbb{E} \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{Z_t(\boldsymbol{\delta}_1)=1}^{2}\sum_{Z_s(\boldsymbol{\delta}_2)=1}^{2} \\ &\frac{1}{T^2}\frac{\varepsilon_t \mathbb{I}_{[Z_t(\boldsymbol{\delta_1})=2]}}{(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_t(\boldsymbol{\delta_1})=2]})^2}\frac{\varepsilon_s \mathbb{I}_{[Z_s(\boldsymbol{\delta_2})=2]}}{(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_s(\boldsymbol{\delta_2})=2]})^2} p_t p_s \end{aligned} \tag{A.23}$$

From independence between ε and $z(\boldsymbol{\delta})$, it follows that

$$\begin{aligned} K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_2^*, \mu_2^*} = \lim_{T\to\infty} T\mathbb{E} \Bigg[&\sum_{t=1}^{T}\sum_{s=1}^{T} \frac{1}{T^2}\mathbb{E}[\mathbb{I}_{[Z_t(\boldsymbol{\delta_1})=2]}\mathbb{I}_{[Z_s(\boldsymbol{\delta_2})=2]} | \mathscr{Y}_t, \boldsymbol{\alpha}_0^G, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2] \\ &\frac{\varepsilon_s \varepsilon_t}{(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_t(\boldsymbol{\delta_1})=2]})^2 (\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_s(\boldsymbol{\delta_2})=2]})^2} \Bigg]. \end{aligned} \tag{A.24}$$

Now, apply the law of iterated expectations (A.20) and note that $\mathbb{E}[\mathbb{I}_{[Z_t(\boldsymbol{\delta_1})=2]}\mathbb{I}_{[Z_s(\boldsymbol{\delta_2})=2]}] = \min[\pi_2(\boldsymbol{\delta}_1), \pi_2(\boldsymbol{\delta}_2)]$. Additionally, use the serial independence of ε and (A.13), which results in

$$\begin{aligned} K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_2^*, \mu_2^*} &= \lim_{T\to\infty} T \sum_{t=1}^{T} \frac{1}{T^2} \min[\pi_2(\boldsymbol{\delta}_1), \pi_2(\boldsymbol{\delta}_2)] \frac{\mathbb{E}[\varepsilon_t^2]}{\sigma_1^4} \\ &= \lim_{T\to\infty} \min[\pi_2(\boldsymbol{\delta}_1), \pi_2(\boldsymbol{\delta}_2)] \frac{1}{\sigma_1^2} = \min[\pi_2(\boldsymbol{\delta}_1), \pi_2(\boldsymbol{\delta}_2)] \frac{1}{\sigma_1^2}. \end{aligned} \tag{A.25}$$

To find the $(\mu_1, \phi_{i(1)})$ elements of the covariance matrix, let us prove the two following lemmas.

Lemma A.7

$$\lim_{T\to\infty} \sum_{t=1}^{T} \frac{1}{T} y_{t-i} = \mu_1 \tag{A.26}$$

Proof. The left-hand side of the equation (A.26) can be rewritten as

$$\lim_{T\to\infty}\sum_{t=1}^{T}\frac{1}{T}y_{t-i} = \underbrace{\lim_{T\to\infty}\sum_{t=1}^{T}\frac{1}{T}y_t}_{(1)} + \underbrace{\lim_{T\to\infty}\frac{1}{T}\sum_{j=1}^{i}(y_{1-j}-y_{T+1-j})}_{(2)} \tag{A.27}$$

Note that term (1) converges to the expected value of $\mathscr{Y}_T$ which under the null hypothesis equals μ_1 and term (2) tends to zero as $i \ll T$. Thus, the right-hand side of (A.27) is equal to μ_1 and equation (A.26) is proven. ■

Lemma A.8

$$\lim_{T\to\infty}\sum_{t=1}^{T}\frac{1}{T}y_{t-i}y_{t-j} = cov[y_{t-i},y_{t-j}] + 2\mu_1^2 \tag{A.28}$$

Proof. The left-hand side of equation (A.28) can be rewritten as

$$\begin{aligned}\lim_{T\to\infty}\sum_{t=1}^{T}\frac{1}{T}y_{t-i}y_{t-j} &= \underbrace{\lim_{T\to\infty}\sum_{t=1}^{T}\frac{1}{T}(y_{t-i}-\mu_1)(y_{t-j}-\mu_1)}_{(1)}\\ &+\mu_1\underbrace{\lim_{T\to\infty}\frac{1}{T}y_{t-i}}_{(2)}+\mu_1\underbrace{\lim_{T\to\infty}\frac{1}{T}y_{t-j}}_{(3)}-\mu_1^2\underbrace{\lim_{T\to\infty}\frac{1}{T}}_{(4)}\end{aligned} \tag{A.29}$$

Note that term (1) converges to the (i,j) element of the auto-covariance matrix, as under the null hypothesis the expected value of $\mathscr{Y}_T$ equals μ_1. According to Lemma A.7, terms (2) and (3) converge to μ_1. Furthermore, term (4) goes to zero. Therefore the right-hand side of the equation (A.29) converges to $\mathrm{cov}[y_{t-i},y_{t-j}] + 2\mu_1^2$ as postulated in equation (A.28). ■

Now, use equations (3.38), (3.41) and (3.43)

$$\begin{aligned}K(\boldsymbol{\delta}_1,\boldsymbol{\delta}_2)_{\mu_1,\phi_{i(1)}} &= \lim_{T\to\infty}T\mathbb{E}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{Z_t(\boldsymbol{\delta}_1)=1}^{2}\sum_{Z_s(\boldsymbol{\delta}_2)=1}^{2}\\ &\frac{1}{T^2}\frac{\varepsilon_t}{(\sigma_1+\sigma_2^*\mathbb{I}_{[Z_t(\boldsymbol{\delta_1})=2]})^2}\frac{\varepsilon_s y_{s-i}}{(\sigma_1+\sigma_2^*\mathbb{I}_{[Z_s(\boldsymbol{\delta_2})=2]})^2}p_t p_s.\end{aligned} \tag{A.30}$$

As the probabilities p_t and p_s sum to unity, respectively, the ε are serially independent, $\mathbb{E}[\varepsilon_t^2] = \sigma_1^2$ and (A.13), thus it follows that

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_1, \phi_{i(1)}} = \frac{1}{\sigma_1^2} \lim_{T \to \infty} \left(\frac{1}{T} \sum_{s=1}^{T} y_{s-i} \right). \tag{A.31}$$

From Lemma A.7 it follows that

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_1, \phi_{i(1)}} = \frac{\mu_1}{\sigma_1^2}. \tag{A.32}$$

Analogically, the $(\phi_{i(1)}, \mu_1)$ element of the covariance matrix equals

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\phi_{i(1)}, \mu_1} = \frac{\mu_1}{\sigma_1^2}. \tag{A.33}$$

The $(\mu_1, \phi^*_{i(2)})$ element of the covariance matrix can be computed from equations (3.38), (3.41) and (3.44)

$$\begin{aligned} K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_1, \phi^*_{i(2)}} = \lim_{T \to \infty} T \mathbb{E} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{Z_t(\boldsymbol{\delta}_1)=1}^{2} \sum_{Z_s(\boldsymbol{\delta}_2)=1}^{2} \\ \frac{1}{T^2} \frac{\varepsilon_t}{(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_t(\boldsymbol{\delta}_1)=2]})^2} \frac{y_{s-i} \varepsilon_s \mathbb{I}_{[Z_s(\boldsymbol{\delta})=2]}}{(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]})^2} p_t p_s. \end{aligned} \tag{A.34}$$

Since p_t sum to unity and ε are serially independent, thus

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_1, \phi^*_{i(2)}} = \lim_{T \to \infty} T \mathbb{E} \sum_{s=1}^{T} \sum_{Z_s(\boldsymbol{\delta}_2)=1}^{2} \frac{y_{s-i}}{T^2} \frac{\varepsilon_s^2 \mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]}}{(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]})^4} p_s. \tag{A.35}$$

From the independence between ε and $z(\boldsymbol{\delta})$ it follows that

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_1, \phi^*_{i(2)}} = \lim_{T \to \infty} T \mathbb{E} \left[\sum_{s=1}^{T} \frac{1}{T^2} \frac{y_{s-i} \varepsilon_s^2}{(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]})^4} \mathbb{E} \left[\mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]} | \mathscr{Y}_T, \boldsymbol{\alpha}_0^G, \boldsymbol{\delta}_2 \right] \right]. \tag{A.36}$$

Apply then, the law of iterated expectations (A.20) and note that $\mathbb{E}[\mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]}] = \pi_2(\boldsymbol{\delta}_2)$. In addition, apply the fact that under the null hypothesis $\mathbb{E}[\varepsilon_s^2] = \sigma_1^2$ and (A.13)

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_1, \phi^*_{i(2)}} = \frac{\pi_2(\boldsymbol{\delta}_2)}{\sigma_1^2} \lim_{T \to \infty} \frac{1}{T} y_{s-i}. \tag{A.37}$$

From Lemma A.7 one concludes that

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_1, \phi^*_{i(2)}} = \frac{\mu_1 \pi_2(\boldsymbol{\delta}_2)}{\sigma_1^2}. \tag{A.38}$$

Analogically, the $(\phi^*_{i(2)}, \mu_1)$ element of the covariance matrix equals

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\phi^*_{i(2)}, \mu_1} = \frac{\mu_1 \pi_2(\boldsymbol{\delta}_1)}{\sigma_1^2}, \tag{A.39}$$

the $(\mu^*_2, \phi_{i(1)})$ element of the covariance matrix equals

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu^*_2, \phi_{i(1)}} = \frac{\mu_1 \pi_2(\boldsymbol{\delta}_1)}{\sigma_1^2} \tag{A.40}$$

and the $(\phi_{i(1)}, \mu^*_2)$ element of the covariance equals

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\phi_{i(1)}, \mu^*_2} = \frac{\mu_1 \pi_2(\boldsymbol{\delta}_2)}{\sigma_1^2}. \tag{A.41}$$

The $(\mu^*_2, \phi^*_{i(2)})$ element of the covariance matrix can be computed from equations (3.38), (3.42) and (3.44)

$$\begin{aligned} K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu^*_2, \phi^*_{i(2)}} = & \lim_{T \to \infty} T \mathbb{E} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{Z_t(\boldsymbol{\delta}_1)=1}^{2} \sum_{Z_s(\boldsymbol{\delta}_2)=1}^{2} \\ & \frac{1}{T^2} \frac{\varepsilon_t \mathbb{I}_{[Z_t(\boldsymbol{\delta}_1)=2]}}{(\sigma_1 + \sigma^*_2 \mathbb{I}_{[Z_t(\boldsymbol{\delta}_1)=2]})^2} \frac{y_{s-i} \varepsilon_s \mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]}}{(\sigma_1 + \sigma^*_2 \mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]})^2} p_t p_s. \end{aligned} \tag{A.42}$$

Now, use the independence between ε and $z(\boldsymbol{\delta})$ which enables to rewrite this equation as

$$\begin{aligned} K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu^*_2, \phi^*_{i(2)}} = \lim_{T \to \infty} T \mathbb{E} \Bigg[& \sum_{t=1}^{T} \sum_{s=1}^{T} \frac{1}{T^2} \mathbb{E}[\mathbb{I}_{[Z_t(\boldsymbol{\delta}_1)=2]} \mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]} | \mathscr{Y}_t, \boldsymbol{\alpha}^G_0, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2] \\ & \frac{y_{s-i} \varepsilon_s \varepsilon_t}{(\sigma_1 + \sigma^*_2 \mathbb{I}_{[Z_t(\boldsymbol{\delta}_1)=2]})^2 (\sigma_1 + \sigma^*_2 \mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]})^2} \Bigg]. \end{aligned} \tag{A.43}$$

Then apply the law of iterated expectations

$$\mathbb{E}\left[\mathbb{E}[\mathbb{I}_{[Z_t(\boldsymbol{\delta_1})=2]} \mathbb{I}_{[Z_s(\boldsymbol{\delta_2})=2]} | \mathscr{Y}_T, \boldsymbol{\alpha}^G_0, \boldsymbol{\delta}_2]\right] = \mathbb{E}[\mathbb{I}_{[Z_t(\boldsymbol{\delta_1})=2]} \mathbb{I}_{[Z_s(\boldsymbol{\delta_2})=2]}] \tag{A.44}$$

and note that $\mathbb{E}[\mathbb{I}_{[Z_t(\boldsymbol{\delta_1})=2]}\mathbb{I}_{[Z_s(\boldsymbol{\delta_2})=2]}] = \min[\pi_2(\boldsymbol{\delta}_1), \pi_2(\boldsymbol{\delta}_2)]$. Furthermore, apply the serial independence of ε, the fact that under the null hypothesis $\mathbb{E}[\varepsilon_s^2] = \sigma_1^2$ and (A.13)

$$\begin{aligned} K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_2^*, \phi_{i(2)}^*} &= \lim_{T\to\infty} T \sum_{s=1}^{T} \frac{1}{T^2} \min[\pi_2(\boldsymbol{\delta}_1), \pi_2(\boldsymbol{\delta}_2)] \frac{y_{s-i}\mathbb{E}[\varepsilon_s^2]}{\sigma_1^4} \\ &= \min[\pi_2(\boldsymbol{\delta}_1), \pi_2(\boldsymbol{\delta}_2)] \frac{1}{\sigma_1^2} \lim_{T\to\infty} \sum_{s=1}^{T} \frac{y_{s-i}}{T}. \end{aligned} \tag{A.45}$$

Applying Lemma A.7 one has

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_2^*, \phi_{i(2)}^*} = \min[\pi_2(\boldsymbol{\delta}_1), \pi_2(\boldsymbol{\delta}_2)] \frac{\mu_1}{\sigma_1^2}. \tag{A.46}$$

Analogically, the $(\phi_{i(2)}^*, \mu_2^*)$ element of the covariance matrix equals

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\phi_{i(2)}^*, \mu_2^*} = \min[\pi_2(\boldsymbol{\delta}_1), \pi_2(\boldsymbol{\delta}_2)] \frac{\mu_1}{\sigma_1^2}. \tag{A.47}$$

Then, according to equations (3.38), (3.43) and (A.13) the $(\phi_{i(1)}, \phi_{j(1)})$ element of the covariance matrix can be computed

$$\begin{aligned} K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\phi_{i(1)}, \phi_{j(1)}} = \lim_{T\to\infty} T\mathbb{E} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{Z_t(\boldsymbol{\delta}_1)=1}^{2} \sum_{Z_s(\boldsymbol{\delta}_2)=1}^{2} \\ \frac{1}{T^2} \frac{y_{t-i}\varepsilon_t}{\sigma_1^2} \frac{y_{s-j}\varepsilon_s}{\sigma_1^2} p_t p_s. \end{aligned} \tag{A.48}$$

From the fact that the probabilities sum to unity, the increments ε are serially independent and under the null hypothesis $\mathbb{E}[\varepsilon^2] = \sigma_1^2$, and it follows that

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\phi_{i(1)}, \phi_{j(1)}} = \frac{1}{\sigma_1^2} \lim_{T\to\infty} \sum_{t=1}^{T} \frac{1}{T} y_{t-i} y_{s-j}. \tag{A.49}$$

From Lemma A.8 one can conclude that

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\phi_{i(1)}, \phi_{j(1)}} = \frac{1}{\sigma_1^2} \left(\text{cov}[y_{t-i}, y_{t-j}] + 2\mu_1^2 \right). \tag{A.50}$$

The $(\phi_{i(1)}, \phi^*_{j(2)})$ element of the covariance matrix can be derived from equations (3.38), (3.43) and (3.44)

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\phi_{i(1)},\phi^*_{j(2)}} = \lim_{T\to\infty} T\mathbb{E} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{Z_t(\boldsymbol{\delta}_1)=1}^{2} \sum_{Z_s(\boldsymbol{\delta}_2)=1}^{2} \frac{1}{T^2} \frac{y_{t-i}\varepsilon_t}{(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_t(\boldsymbol{\delta}_1)=2]})^2} \frac{y_{s-j}\varepsilon_s \mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]}}{(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]})^2} p_t p_s. \tag{A.51}$$

As the probabilities p_t sum to unity and increments ε are serially independent

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\phi_{i(1)},\phi^*_{j(2)}} = \lim_{T\to\infty} T\mathbb{E} \sum_{s=1}^{T} \sum_{Z_s(\boldsymbol{\delta}_2)=1}^{2} \frac{1}{T^2} y_{s-i} y_{s-j} \frac{\varepsilon_s^2 \mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]}}{(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]})^4} p_s. \tag{A.52}$$

Now, use the independence between ε and $z(\boldsymbol{\delta})$ to write

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\phi_{i(1)},\phi^*_{j(2)}} = \lim_{T\to\infty} \mathbb{E} \left[\sum_{s=1}^{T} \frac{1}{T} \frac{y_{s-i} y_{s-j} \varepsilon_s^2}{(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]})^4} \mathbb{E}[\mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]} | \mathscr{Y}_T, \boldsymbol{\alpha_0}^G, \boldsymbol{\delta}_2] \right]. \tag{A.53}$$

Then apply the law of iterated expectations (A.20) and note that $\mathbb{E}[\mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]}] = \pi_2(\boldsymbol{\delta}_2)$. Furthermore, apply the fact that under the null hypothesis $\mathbb{E}[\varepsilon_s^2] = \sigma_1^2$ and (A.13)

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\phi_{i(1)},\phi^*_{j(2)}} = \frac{\pi_2(\boldsymbol{\delta}_2)}{\sigma_1^2} \lim_{T\to\infty} \frac{1}{T} \sum_{s=1}^{T} y_{s-i} y_{s-j}. \tag{A.54}$$

From Lemma A.8 it follows that

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\phi_{i(1)},\phi^*_{j(2)}} = \frac{\pi_2(\boldsymbol{\delta}_2)}{\sigma_1^2} (\mathrm{cov}[y_{s-i}, y_{s-j}] + 2\mu_1^2). \tag{A.55}$$

Analogically, the $(\phi^*_{i(2)}, \phi_{j(1)})$ element of the covariance matrix equals

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\phi^*_{i(2)},\phi_{j(1)}} = \frac{\pi_2(\boldsymbol{\delta}_1)}{\sigma_1^2} (\mathrm{cov}[y_{t-i}, y_{t-j}] + 2\mu_1^2). \tag{A.56}$$

Now, according to equations (3.38) and (3.44) compute the $(\phi^*_{i(2)}, \phi^*_{j(2)})$ element of the covariance matrix

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\phi^*_{i(2)},\phi^*_{j(2)}} = \lim_{T\to\infty} T\mathbb{E} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{Z_t(\boldsymbol{\delta}_1)=1}^{2} \sum_{Z_s(\boldsymbol{\delta}_2)=1}^{2} \frac{1}{T^2} \frac{y_{t-i}\varepsilon_t \mathbb{I}_{[Z_t(\boldsymbol{\delta}_1)=2]}}{(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_t(\boldsymbol{\delta}_1)=2]})^2} \frac{y_{s-j}\varepsilon_s \mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]}}{(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]})^2} p_t p_s. \tag{A.57}$$

Using the independence between ε and $z(\boldsymbol{\delta})$ this can be rewritten as

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\phi^*_{i(2)},\phi^*_{j(2)}} = \lim_{T\to\infty} T\mathbb{E} \left[\sum_{t=1}^{T} \sum_{s=1}^{T} \frac{1}{T^2} \frac{y_{s-i} y_{s-j} \varepsilon_t \varepsilon_s}{(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_t(\boldsymbol{\delta}_1)=2]})^2 (\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]})^2} \mathbb{E}[\mathbb{I}_{[Z_t(\boldsymbol{\delta}_1)=2]} \mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]} | \mathscr{Y}_T, \boldsymbol{\alpha_0}^G, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2] \right]. \tag{A.58}$$

Now, apply the law of iterated expectations (A.44) and note that $\mathbb{E}[\mathbb{I}_{[Z_t(\boldsymbol{\delta}_1)=2]} \mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]}] = \min[\pi_2(\boldsymbol{\delta}_1), \pi_2(\boldsymbol{\delta}_2)]$. In addition, apply the serial independence of increments ε, the fact that under the null hypothesis $\mathbb{E}[\varepsilon^2] = \sigma_1^2$ and (A.13), then

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\phi^*_{i(2)},\phi^*_{j(2)}} = \frac{\min[\pi_2(\boldsymbol{\delta}_1), \pi_2(\boldsymbol{\delta}_2)]}{\sigma_1^2} \lim_{T\to\infty} \frac{1}{T} \sum_{s=1}^{T} y_{s-i} y_{s-j}. \tag{A.59}$$

From Lemma A.8 it follows that

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\phi^*_{i(2)},\phi^*_{j(2)}} = \frac{\min[\pi_2(\boldsymbol{\delta}_1), \pi_2(\boldsymbol{\delta}_2)]}{\sigma_1^2} (\mathrm{cov}[y_{s-i}, y_{s-j}] + 2\mu_1^2). \tag{A.60}$$

The (σ_1^2, σ_1^2) element of the covariance matrix can be computed from equations (3.38), (3.45) and (A.13)

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\sigma_1^2,\sigma_1^2} = \lim_{T\to\infty} T\mathbb{E} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{Z_t(\boldsymbol{\delta}_1)=1}^{2} \sum_{Z_s(\boldsymbol{\delta}_2)=1}^{2} \frac{1}{T^2} \frac{1}{2\sigma_1^2} \frac{1}{2\sigma_1^2} \left(\frac{\varepsilon_t^2}{\sigma_1^2} - 1 \right) \left(\frac{\varepsilon_s^2}{\sigma_1^2} - 1 \right) p_t p_s. \tag{A.61}$$

As probabilities p_t and p_s sum to unity, respectively, it follows that

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\sigma_1^2,\sigma_1^2} = \lim_{T\to\infty} T\mathbb{E} \sum_{t=1}^{T} \sum_{s=1}^{T} \frac{1}{T^2} \frac{1}{4\sigma_1^4} \left(\frac{\varepsilon_t^2}{\sigma_1^2} - 1 \right) \left(\frac{\varepsilon_s^2}{\sigma_1^2} - 1 \right). \tag{A.62}$$

Since the increments ε are serially independent, thus

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\sigma_1^2,\sigma_1^2} = \lim_{T\to\infty} T\mathbb{E}\sum_{t=1}^{T} \frac{1}{T^2}\frac{1}{4\sigma_1^4}\left(\frac{\varepsilon_t^4}{\sigma_1^4} - 2\frac{\varepsilon_t^2}{\sigma_1^2} + 1\right). \tag{A.63}$$

Now, note that under the null hypothesis $\mathbb{E}(\varepsilon_t^4) = 3\sigma_1^4$ and $\mathbb{E}(\varepsilon_t^2) = \sigma_1^2$, from which one can conclude that

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\sigma_1^2,\sigma_1^2} = \lim_{T\to\infty} \frac{1}{2\sigma_1^4} = \frac{1}{2\sigma_1^4}. \tag{A.64}$$

According to equations (3.38), (3.45) and (3.46) compute the $(\sigma_1^2, \sigma_2^{*2})$ element of the covariance matrix

$$\begin{aligned} K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\sigma_1^2,\sigma_2^{*2}} = \lim_{T\to\infty} T\mathbb{E}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{Z_t(\boldsymbol{\delta}_1)=1}^{2}\sum_{Z_s(\boldsymbol{\delta}_2)=1}^{2} \\ \frac{1}{T^2}\frac{1}{2\sigma_1(\sigma_1+\sigma_2^*\mathbb{I}_{[Z_t(\boldsymbol{\delta}_1)=2]})}\frac{\mathbb{I}_{[Z_t(\boldsymbol{\delta}_1)=2]}}{2\sigma_2^*(\sigma_1+\sigma_2^*\mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]})} \\ \left(\frac{\varepsilon_t^2}{(\sigma_1+\sigma_2^*\mathbb{I}_{[Z_t(\boldsymbol{\delta}_1)=2]})^2} - 1\right)\left(\frac{\varepsilon_s^2}{(\sigma_1+\sigma_2^*\mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]})^2} - 1\right)p_t p_s. \end{aligned} \tag{A.65}$$

Since the probabilities p_t sum to unity and the increments ε are serially independent, thus

$$\begin{aligned} K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\sigma_1^2,\sigma_2^{*2}} = \lim_{T\to\infty} T\mathbb{E}\sum_{s=1}^{T}\sum_{Z_s(\boldsymbol{\delta}_2)=1}^{2} \\ \frac{1}{T^2}\frac{\mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]}}{4\sigma_1\sigma_2^*(\sigma_1+\sigma_2^*\mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]})^2}\left(\frac{\varepsilon_s^2}{(\sigma_1+\sigma_2^*\mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]})^2} - 1\right)^2 p_s. \end{aligned} \tag{A.66}$$

After applying the independence between ε and $z(\boldsymbol{\delta})$ the equation can be rewritten as

$$\begin{aligned} K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\sigma_1^2,\sigma_2^{*2}} = \lim_{T\to\infty} T\mathbb{E}\sum_{s=1}^{T}\Bigg[\frac{1}{T^2}\frac{1}{4\sigma_1\sigma_2^*(\sigma_1+\sigma_2^*\mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]})^2} \\ \left(\frac{\varepsilon_s^2}{(\sigma_1+\sigma_2^*\mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]})^2} - 1\right)^2 \mathbb{E}\left[Z_s(\boldsymbol{\delta}_2)|\mathscr{Y}_T, \boldsymbol{\alpha}_0^G, \boldsymbol{\delta}_2\right]\Bigg]. \end{aligned} \tag{A.67}$$

Then apply the law of iterated expectations (A.20) and note that $\mathbb{E}[Z_s(\boldsymbol{\delta}_2)] = \pi_2(\boldsymbol{\delta}_2)$ and that under the null hypothesis (A.13)

$$\begin{aligned} K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\sigma_1^2, \sigma_2^{*2}} &= \lim_{T\to\infty} T\mathbb{E} \sum_{s=1}^{T} \frac{1}{T^2} \frac{1}{4\sigma_1^3 \sigma_2^*} \left(\frac{\varepsilon_s^2}{\sigma_1^2} - 1 \right)^2 \pi_2(\boldsymbol{\delta}_2) \\ &= \lim_{T\to\infty} T\mathbb{E} \sum_{t=1}^{T} \frac{1}{T^2} \frac{1}{4\sigma_1^3 \sigma_2^*} \left(\frac{\varepsilon_s^4}{\sigma_1^4} - 2\frac{\varepsilon_s^2}{\sigma_1^2} + 1 \right) \pi_2(\boldsymbol{\delta}_2). \end{aligned} \tag{A.68}$$

Since $\mathbb{E}(\varepsilon_t^4) = 3\sigma_1^4$ and $\mathbb{E}(\varepsilon_t^2) = \sigma_1^2$ under the null hypothesis, the equation simplifies to

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\sigma_1^2, \sigma_2^{*2}} = \lim_{T\to\infty} \frac{\pi_2(\boldsymbol{\delta}_2)}{2\sigma_1^3 \sigma_2^*} = \frac{\pi_2(\boldsymbol{\delta}_2)}{2\sigma_1^3 \sigma_2^*}. \tag{A.69}$$

Analogically, the $(\sigma_2^{*2}, \sigma_1^2)$ element of the covariance matrix equals

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\sigma_2^{*2}, \sigma_1^2} = \frac{\pi_2(\boldsymbol{\delta}_1)}{2\sigma_1^3 \sigma_2^*} \tag{A.70}$$

The $(\sigma_2^{*2}, \sigma_2^{*2})$ element of the covariance matrix can be computed from equations (3.38) and (3.46)

$$\begin{aligned} K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\sigma_2^{*2}, \sigma_2^{*2}} = \lim_{T\to\infty} T\mathbb{E} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{Z_t(\boldsymbol{\delta}_1)=1}^{2} \sum_{Z_s(\boldsymbol{\delta}_2)=1}^{2} \\ \frac{1}{T^2} \frac{\mathbb{I}_{[Z_t(\boldsymbol{\delta}_1)=2]}}{2\sigma_2^*(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_t(\boldsymbol{\delta}_1)=2]})} \frac{\mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]}}{2\sigma_2^*(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]})} \\ \left(\frac{\varepsilon_t^2}{(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_t(\boldsymbol{\delta}_1)=2]})^2} - 1 \right) \left(\frac{\varepsilon_s^2}{(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]})^2} - 1 \right) p_t p_s. \end{aligned} \tag{A.71}$$

Applying the independence of ε and $z(\boldsymbol{\delta})$ the equation reduces to

$$\begin{aligned} K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\sigma_2^{*2}, \sigma_2^{*2}} = \lim_{T\to\infty} T\mathbb{E} \Bigg[\sum_{t=1}^{T} \sum_{s=1}^{T} \frac{1}{T^2} \frac{\mathbb{E}[\mathbb{I}_{[Z_t(\boldsymbol{\delta}_1)=2]} \mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]}(\boldsymbol{\delta}_2) | \mathscr{Y}_T, \boldsymbol{\alpha}_0^G, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2]}{4\sigma_2^{*2}(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_t(\boldsymbol{\delta}_1)=2]})(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]})} \\ \left(\frac{\varepsilon_t^2}{(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_t(\boldsymbol{\delta}_1)=2]})^2} - 1 \right) \left(\frac{\varepsilon_s^2}{(\sigma_1 + \sigma_2^* \mathbb{I}_{[Z_s(\boldsymbol{\delta}_2)=2]})^2} - 1 \right) \Bigg]. \end{aligned} \tag{A.72}$$

Then apply the law of iterated expectations (A.44) and note that $\mathbb{E}[Z_t(\boldsymbol{\delta}_1)Z_s(\boldsymbol{\delta}_2)] = \min[\pi_2(\boldsymbol{\delta}_1), \pi_2(\boldsymbol{\delta}_2)]$. In addition, use the serial independence of ε and (A.13), which provides that

$$\begin{aligned} K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\sigma_2^{*2},\sigma_2^{*2}} &= \lim_{T\to\infty} T\mathbb{E}\sum_{s=1}^{T} \frac{1}{T^2}\frac{1}{4\sigma_1^2\sigma_2^{*2}}\left(\frac{\varepsilon_s^2}{\sigma_1^2}-1\right)^2 \min[\pi_2(\boldsymbol{\delta}_1), \pi_2(\boldsymbol{\delta}_2)] \\ &= \lim_{T\to\infty} T\mathbb{E}\sum_{t=1}^{T} \frac{1}{T^2}\frac{1}{4\sigma_1^2\sigma_2^{*2}}\left(\frac{\varepsilon_s^4}{\sigma_1^4}-2\frac{\varepsilon_s^2}{\sigma_1^2}+1\right) \min[\pi_2(\boldsymbol{\delta}_1), \pi_2(\boldsymbol{\delta}_2)]. \end{aligned} \tag{A.73}$$

Finally, recall that under the null hypothesis $\mathbb{E}(\varepsilon_t^4) = 3\sigma_1^4$ and $\mathbb{E}(\varepsilon_t^2) = \sigma_1^2$, thus

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\sigma_2^{*2},\sigma_2^{*2}} = \lim_{T\to\infty} \frac{\min[\pi_2(\boldsymbol{\delta}_1), \pi_2(\boldsymbol{\delta}_2)]}{2\sigma_1^2\sigma_2^{*2}} = \frac{\min[\pi_2(\boldsymbol{\delta}_1), \pi_2(\boldsymbol{\delta}_2)]}{2\sigma_1^2\sigma_2^{*2}}. \tag{A.74}$$

The (μ_1, σ_1^2) element of the covariance matrix results from equations (3.38), (3.41), (3.45) and (A.13)

$$\begin{aligned} K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_1,\sigma_1^2} = \lim_{T\to\infty} T\mathbb{E}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{Z_t(\boldsymbol{\delta}_1)=1}^{2}\sum_{Z_s(\boldsymbol{\delta}_2)=1}^{2} \\ \frac{1}{T^2}\frac{1}{2\sigma_1^2}\left(\frac{\varepsilon_t^2}{\sigma_1^2}-1\right)\frac{\varepsilon_t}{\sigma_1^2}p_t p_s. \end{aligned} \tag{A.75}$$

Given the fact that the products of the probabilities p_t and p_s equal unity, respectively, and the increments ε are serially independent,

$$\begin{aligned} K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_1,\sigma_1^2} &= \lim_{T\to\infty} T\sum_{t=1}^{T}\frac{1}{T^2}\mathbb{E}\left[\frac{\varepsilon_t}{2\sigma_1^4}\left(\frac{\varepsilon_t^2}{\sigma_1^2}-1\right)\right] \\ &= \lim_{T\to\infty} T\sum_{t=1}^{T}\frac{1}{T^2}\mathbb{E}\left[\frac{\varepsilon_t^3}{2\sigma_1^6}-\frac{\varepsilon_t}{2\sigma_1^4}\right]. \end{aligned} \tag{A.76}$$

Since ε_t is normally distributed, $\mathbb{E}(\varepsilon_t^3) = 0$ and $\mathbb{E}(\varepsilon_t) = 0$, thus

$$K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)_{\mu_1,\sigma_1^2} = 0. \tag{A.77}$$

Analogically, all elements of the covariance matrix dependent on μ_2, $\phi_{i(1)}$ or $\phi_{i(2)}$ on the one side, and dependent on σ_1^2 or σ_2^{*2} on the other side are equal zero. Collecting all elements of the covariance matrix $K(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)$ computed above in the matrix (3.47) ends the proof. $\blacksquare$

Appendix B

MS estimation (1.1975-12.2004) – parameters

Table B.1: MS parameter for DAX30 (1.1975-12.2004)

	MS(1-1)		MS(1-2)		MS(2-1)		MS(2-2)	
	GBM	**AR(1)**	**GBM**	**AR(1)**	**GBM**	**AR(1)**	**GBM**	**AR(1)**
μ_1	0.0066**	0.0060*	0.0066**	0.0060**	0.0122***	0.0126***	0.0095***	0.0097***
	(0.0031)	(0.0031)	(0.0025)	(0.0026)	(0.0028)	(0.0028)	(0.0029)	(0.0029)
	[2.0938]	[1.9215]	[2.5770]	[2.3619]	[4.4398]	[4.5301]	[3.3274]	[3.3989]
μ_2	—	—	—	—	-0.1742***	-0.1901***	0.0012	0.0004
					(0.0200)	(0.0289)	(0.0082)	(0.0081)
					[-8.6987]	[-6.5835]	[0.1432]	[0.0443]
$\phi_{1(1)}$	—	0.0364	—	0.0364	—	-0.0861*	—	-0.0316
		(0.0531)		(0.0559)		(0.0504)		(0.0706)
		[0.6849]		[0.6509]		[-1.7102]		[-0.4472]
$\phi_{1(2)}$	—	—	—	—	—	-0.2382	—	0.0603
						(0.2695)		(0.0942)
						[-0.8836]		[0.6405]
σ_1^2	0.0035**	0.0035*	0.0015***	0.0015***	0.0025***	0.0025***	0.0015***	0.0015***
	(0.0003)	(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
	[13.4160]	[13.3972]	[8.4579]	[9.1259]	[12.7124]	[12.9028]	[8.2691]	[8.9727]
σ_2^2	—	—	0.0070***	0.0071***	—	—	0.0072***	0.0072***
			(0.0010)	(0.0010)			(0.0010)	(0.0010)
			[7.1092]	[7.1432]			[6.9915]	[7.0178]
p_{11}	—	—	0.9797***	0.9810***	0.9788***	0.9816***	0.9807***	0.9813***
			(0.0121)	(0.0121)	(0.0088)	(0.0080)	(0.0118)	(0.0119)
			[81.1105]	[81.1804]	[111.7417]	[122.6466]	[82.9419]	[82.4070]
p_{22}	—	—	0.9601***	0.9646***	0.3274**	0.4090**	0.9606***	0.9630***
			(0.0238)	(0.0207)	(0.1574)	(0.1930)	(0.0239)	(0.0224)
			[40.3044]	[46.5386]	[2.0807]	[2.1189]	[40.1487]	[42.9104]
p_{21}	—	—	0.0203	0.0190	0.0212	0.0184	0.0193	0.0187
p_{12}	—	—	0.0399	0.0354	0.6726	0.5910	0.0394	0.0370
π_1	—	—	0.6632	0.6509	0.9695	0.9699	0.6716	0.6641
π_2	—	—	0.3368	0.3491	0.0305	0.0301	0.3284	0.3359
D_1	—	—	49.34	52.70	47.20	54.43	51.84	53.46
D_2	—	—	25.06	28.27	1.49	1.69	25.35	27.04

Note:
MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance. GBM denotes no auto-regression in the mean equation(s) and AR(1) an auto-regression of the first order in each mean equation. $[\cdot]$ denotes standard error, $(\cdot)$ and t test statistic. *, ** and *** denote significance at the 10%, 5% and 1% confidence level, respectively.

Table B.2: MS parameter for REXP(10%)-DAX30(90%) (1.1975-12.2004)

	MS(1-1)		MS(1-2)		MS(2-1)		MS(2-2)	
	GBM	**AR(1)**	**GBM**	**AR(1)**	**GBM**	**AR(1)**	**GBM**	**AR(1)**
μ_1	0.0067**	0.0062**	0.0067***	0.0062***	0.0117***	0.0121***	0.0092***	0.0094***
	(0.0028)	(0.0028)	(0.0023)	(0.0023)	(0.0025)	(0.0025)	(0.0026)	(0.0026)
	[2.3659]	[2.1843]	[2.8826]	[2.6567]	[4.6842]	[4.7784]	[3.5301]	[3.6252]
μ_2	—	—	—	—	-0.1536***	-0.1671***	0.0021	0.0012
					(0.0183)	(0.0266)	(0.0072)	(0.0072)
					[-8.3865]	[-6.2812]	[0.2865]	[0.1655]
$\phi_{1(1)}$	—	0.0334	—	0.0334	—	-0.0862*	—	-0.0302
		(0.0527)		(0.0547)		(0.0507)		(0.0660)
		[0.6349]		[0.6113]		[-1.7008]		[-0.4579]
$\phi_{1(2)}$	—	—	—	—	—	-0.2305	—	0.0564
						(0.2864)		(0.0980)
						[-0.8049]		[0.5757]
σ_1^2	0.0029**	0.0028**	0.0012***	0.0012***	0.0021***	0.0020***	0.0013***	0.0013***
	(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0001)
	[13.4158]	[13.3972]	[8.4194]	[9.0986]	[12.6769]	[12.8788]	[8.2401]	[8.9816]
σ_2^2	—	—	0.0056***	0.0057***			0.0057***	0.0057***
			(0.0008)	(0.0008)			(0.0008)	(0.0008)
			[7.0914]	[7.1251]			[6.9959]	[7.0257]
p_{11}	—	—	0.9794***	0.9808***	0.9789***	0.9817***	0.9804***	0.9813***
			(0.0123)	(0.0123)	(0.0088)	(0.0081)	(0.0120)	(0.0120)
			[79.6277]	[79.8796]	[110.7029]	[121.1451]	[81.5221]	[82.0565]
p_{22}	—	—	0.9595***	0.9644***	0.3275**	0.4094**	0.9600***	0.9631***
			(0.0243)	(0.0209)	(0.1591)	(0.1981)	(0.0243)	(0.0223)
			[39.5183]	[46.0476]	[2.0580]	[2.0668]	[39.4451]	[43.1252]
p_{21}	—	—	0.0206	0.0192	0.0211	0.0183	0.0196	0.0187
p_{12}	—	—	0.0405	0.0356	0.6725	0.5906	0.0400	0.0369
π_1	—	—	0.6625	0.6499	0.9695	0.9699	0.6710	0.6630
π_2	—	—	0.3375	0.3501	0.0305	0.0301	0.3290	0.3370
D_1	—	—	48.52	52.18	47.29	54.55	50.96	53.34
D_2	—	—	24.71	28.10	1.49	1.69	24.99	27.11

Note:
MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance. GBM denotes no auto-regression in the mean equation(s) and AR(1) an auto-regression of the first order in each mean equation. [$\cdot$] denotes standard error, ($\cdot$) and t test statistic. *, ** and *** denote significance at the 10%, 5% and 1% confidence level, respectively.

Table B.3: MS parameter for REXP(20%)-DAX30(80%) (1.1975-12.2004)

	MS(1-1)		**MS(1-2)**		**MS(2-1)**		**MS(2-2)**	
	GBM	**AR(1)**	**GBM**	**AR(1)**	**GBM**	**AR(1)**	**GBM**	**AR(1)**
μ_1	0.0067***	0.0063**	0.0067***	0.0063***	0.0111***	0.0115***	0.0089***	0.0092***
	(0.0025)	(0.0025)	(0.0021)	(0.0021)	(0.0022)	(0.0023)	(0.0024)	(0.0024)
	[2.6905]	[2.4953]	[3.2389]	[2.9973]	[4.9653]	[5.0811]	[3.7730]	[3.8904]
μ_2	—	—	—	—	-0.1335***	-0.1444***	0.0028	0.0019
					(0.0166)	(0.0247)	(0.0064)	(0.0064)
					[-8.0318]	[-5.8591]	[0.4375]	[0.2957]
$\phi_{1(1)}$	—	0.0303	—	0.0303	—	-0.0854	—	-0.0280
		(0.0526)		(0.0537)		(0.0513)		(0.0659)
		[0.5756]		[0.5639]		[-1.6666]		[-0.4247]
$\phi_{1(2)}$	—	—	—	—	—	-0.2165	—	0.0517
						(0.3155)		(0.0972)
						[-0.6862]		[0.5320]
σ_1^2	0.0023***	0.0022**	$9.95\cdot10^{-4}$***	0.0010***	0.0016***	0.0016***	0.0010***	0.0010***
	(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
	[13.4157]	[13.3970]	[8.3632]	[9.0605]	[12.6272]	[12.8294]	[8.1994]	[8.9782]
σ_2^2	—	—	0.0044***	0.0044***	—	—	0.0045***	0.0045***
			(0.0006)	(0.0006)			(0.0006)	(0.0006)
			[7.0454]	[7.0867]			[6.9681]	[7.0096]
p_{11}	—	—	0.9788***	0.9805***	0.9789***	0.9816***	0.9800***	0.9811***
			(0.0127)	(0.0126)	(0.0090)	(0.0083)	(0.0123)	0.0121
			[76.7676]	[77.8729]	[109.3144]	[118.1393]	[79.7952]	[80.9836]
p_{22}	—	—	0.9585***	0.9640***	0.3266**	0.4067**	0.9592***	0.9630***
			(0.0252)	(0.0214)	(0.1600)	(0.2058)	(0.0250)	(0.0225)
			[37.9889]	[45.0672]	[2.0413]	[1.9761]	[38.4068]	[42.8231]
p_{21}	—	—	0.0212	0.0195	0.0211	0.0184	0.0200	0.0189
p_{12}	—	—	0.0415	0.0360	0.6734	0.5933	0.0408	0.0370
π_1	—	—	0.6616	0.6491	0.9696	0.9700	0.6710	0.6623
π_2	—	—	0.3384	0.3509	0.0304	0.0300	0.3290	0.3377
D_1	—	—	47.09	51.36	47.37	54.46	50.04	52.96
D_2	—	—	24.08	27.77	1.48	1.69	24.54	27.00

Note:
MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance. GBM denotes no auto-regression in the mean equation(s) and AR(1) an auto-regression of the first order in each mean equation. [$\cdot$] denotes standard error, ($\cdot$) and t test statistic. *, ** and *** denote significance at the 10%, 5% and 1% confidence level, respectively.

Table B.4: MS parameter for REXP(25%)-DAX30(75%) (1.1975-12.2004)

	MS(1-1)		MS(1-2)		MS(2-1)		MS(2-2)	
	GBM	**AR(1)**	**GBM**	**AR(1)**	**GBM**	**AR(1)**	**GBM**	**AR(1)**
μ_1	0.0068***	0.0063***	0.0068***	0.0063***	0.0108***	0.0112***	0.0088***	0.0090***
	(0.0024)	(0.0024)	(0.0020)	(0.0020)	(0.0021)	(0.0021)	(0.0022)	(0.0023)
	[2.8765]	[2.6743]	[3.4367]	[3.1905]	[5.1220]	[5.2379]	[3.9164]	[4.0086]
μ_2	—	—	—	—	-0.1237***	-0.1332***	0.0031	0.0022
					(0.0158)	(0.0228)	(0.0060)	(0.0060)
					[-7.8318]	[-5.8384]	[0.5111]	[0.3653]
$\phi_{1(1)}$	—	0.0286	—	0.0286	—	-0.0846	—	-0.0263
		(0.0523)		(0.0497)		(0.0518)		(0.0701)
		[0.5477]		[0.5766]		[-1.6334]		[-0.3750]
$\phi_{1(2)}$	—	—	—	—	—	-0.2053	—	0.0490
						(0.3029)		(0.1004)
						[-0.6777]		[0.4879]
σ_1^2	0.0020***	0.0020***	$8.91 \cdot 10^{-4}$***	$9.04 \cdot 10^{-4}$***	0.0015***	0.0014***	$9.14 \cdot 10^{-4}$***	$9.18 \cdot 10^{-4}$***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
	[13.4156]	[13.3969]	[8.3361]	[9.0351]	[12.5922]	[12.8167]	[8.1756]	[8.9676]
σ_2^2	—	—	0.0039***	0.0039***	—	—	0.0039***	0.0039***
			(0.0006)	(0.0005)			(0.0006)	(0.0006)
			[7.0048]	[7.0548]			[6.9356]	[6.9839]
p_{11}	—	—	0.9785***	0.9803***	0.9789***	0.9816***	0.9799***	0.9810***
			(0.0130)	(0.0128)	(0.0090)	(0.0084)	(0.0124)	(0.0123)
			[75.5296]	[76.4072]	[108.3796]	[117.3435]	[78.9043]	[80.0795]
p_{22}	—	—	0.9579***	0.9636***	0.3256**	0.4034*	0.9588***	0.9627***
			(0.0258)	(0.0218)	(0.1606)	(0.2054)	(0.0254)	(0.0227)
			[37.1791]	[44.2581]	[2.0272]	[1.9643]	[37.7283]	[42.3450]
p_{21}	—	—	0.0215	0.0197	0.0211	0.0184	0.0201	0.0190
p_{12}	—	—	0.0421	0.0364	0.6744	0.5966	0.0412	0.0373
π_1	—	—	0.6617	0.6487	0.9697	0.9700	0.6715	0.6623
π_2	—	—	0.3383	0.3513	0.0303	0.0300	0.3285	0.3377
D_1	—	—	46.52	50.75	47.40	54.25	49.64	52.64
D_2	—	—	23.78	27.49	1.48	1.68	24.28	26.84

Note:
MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance. GBM denotes no auto-regression in the mean equation(s) and AR(1) an auto-regression of the first order in each mean equation. [·] denotes standard error, (·) and t test statistic. *, ** and *** denote significance at the 10%, 5% and 1% confidence level, respectively.

Table B.5: MS parameter for REXP(30%)-DAX30(70%) (1.1975-12.2004)

	MS(1-1)		MS(1-2)		MS(2-1)		MS(2-2)	
	GBM	**AR(1)**	**GBM**	**AR(1)**	**GBM**	**AR(1)**	**GBM**	**AR(1)**
μ_1	0.0068***	0.0064***	0.0068***	0.0064***	0.0105***	0.0109***	0.0087***	0.0089***
	(0.0022)	(0.0022)	(0.0019)	(0.0019)	(0.0020)	(0.0020)	(0.0021)	(0.0021)
	[3.0840]	[2.8722]	[3.6584]	[3.3911]	[5.3016]	[5.4244]	[4.0672]	[4.2199]
μ_2	—	—	—	—	-0.1140***	-0.1219***	0.0033	0.0025
					(0.0150)	(0.0220)	(0.0056)	(0.0056)
					[-7.6170]	[-5.5384]	[0.5933]	[0.4393]
$\phi_{1(1)}$	—	0.0270	—	0.0270	—	-0.0832	—	-0.0241
		(0.0529)		(0.0535)		(0.0512)		(0.0621)
		[0.5099]		[0.5042]		[-1.6268]		[-0.3880]
$\phi_{1(2)}$	—	—	—	—	—	-0.1900	—	0.0459
						(0.3179)		(0.0886)
						[-0.5977]		[0.5182]
σ_1^2	0.0017***	0.0017***	$7.93 \cdot 10^{-4}$***	$8.04 \cdot 10^{-4}$***	0.0013***	0.0013***	$8.13 \cdot 10^{-4}$***	$8.17 \cdot 10^{-4}$***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
	[13.4155]	[13.3968]	[8.3009]	[9.0015]	[12.5461]	[12.7690]	[8.1400]	[8.9537]
σ_2^2	—	—	0.0033***	0.0033***	—	—	0.0034***	0.0034***
			(0.0005)	(0.0005)			(0.0005)	(0.0005)
			[6.9482]	[7.0107]			[6.8855]	[6.9504]
p_{11}	—	—	0.9781***	0.9800***	0.9789***	0.9814***	0.9795***	0.9808***
			(0.0133)	(0.0131)	(0.0091)	(0.0086)	(0.0127)	(0.0125)
			[73.6263]	[74.5589]	[107.1942]	[114.6109]	[76.8612]	[78.6885]
p_{22}	—	—	0.9572***	0.9631***	0.3240**	0.3982**	0.9580***	0.9624***
			(0.0266)	(0.0223)	(0.1623)	(0.2018)	(0.0263)	(0.0231)
			[36.0200]	[43.1822]	[1.9970]	[1.9727]	[36.4614]	[41.6268]
p_{21}	—	—	0.0219	0.0200	0.0211	0.0186	0.0205	0.0192
p_{12}	—	—	0.0428	0.0369	0.6760	0.6018	0.0420	0.0376
π_1	—	—	0.6617	0.6484	0.9697	0.9701	0.6719	0.6621
π_2	—	—	0.3383	0.3516	0.0303	0.0299	0.3281	0.3379
D_1	—	—	45.67	50.00	47.40	53.87	48.70	52.07
D_2	—	—	23.35	27.12	1.48	1.66	23.78	26.57

Note:
MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance. GBM denotes no auto-regression in the mean equation(s) and AR(1) an auto-regression of the first order in each mean equation. [$\cdot$] denotes standard error, ($\cdot$) and t test statistic. *, ** and *** denote significance at the 10%, 5% and 1% confidence level, respectively.

Table B.6: MS parameter for REXP(40%)-DAX30(60%) (1.1975-12.2004)

	MS(1-1)		MS(1-2)		MS(2-1)		MS(2-2)	
	GBM	**AR(1)**	**GBM**	**AR(1)**	**GBM**	**AR(1)**	**GBM**	**AR(1)**
μ_1	0.0068^{***}	0.0064^{***}	0.0068^{***}	0.0064^{***}	0.0099^{***}	0.0103^{***}	0.0084^{***}	0.0086^{***}
	(0.0019)	(0.0019)	(0.0016)	(0.0017)	(0.0017)	(0.0018)	(0.0019)	(0.0019)
	[3.5759]	[3.3387]	[4.1548]	[3.8632]	[5.7021]	[5.8006]	[4.4221]	[4.5664]
μ_2	—	—	—	—	-0.0948^{***}	-0.0996^{***}	0.0038	0.0029
					(0.0133)	(0.0187)	(0.0050)	(0.0049)
					[-7.1083]	[-5.3219]	[0.7583]	[0.6000]
$\phi_{1(1)}$	—	0.0236	—	0.0236	—	-0.0785	—	-0.0172
		(0.0531)		(0.0529)		(0.0524)		(0.0554)
		[0.4445]		[0.4467		[-1.4982]		[-0.3108]
$\phi_{1(2)}$	—	—	—	—	—	-0.1442	—	0.0381
						(0.3031)		(0.0951)
						[-0.4756]		[0.4007]
σ_1^2	0.0013^{***}	0.0013^{***}	$6.16\cdot10^{-4***}$	$6.25\cdot10^{-4***}$	$9.75\cdot10^{-4***}$	$9.62\cdot10^{-4***}$	$6.34\cdot10^{-4***}$	$6.37\cdot10^{-4***}$
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
	[13.4152]	[13.3965]	[8.1642]	[8.8849]	[12.3997]	[12.6194]	[8.0171]	[8.8844]
σ_2^2	—	—	0.0024^{***}	0.0024^{***}	—	—	0.0025^{***}	0.0025^{***}
			(0.0004)	(0.0004)			(0.0004)	(0.0004)
			[6.7620]	[6.8579]			[6.6919]	[6.8196]
p_{11}	—	—	0.9765^{***}	0.9789^{***}	0.9789^{***}	0.9809^{***}	0.9783^{***}	0.9801^{***}
			(0.0147)	(0.0143)	(0.0094)	(0.0091)	(0.0139)	(0.0133)
			[66.5756]	[68.5280]	[103.6212]	[108.0096]	[70.5803]	[73.9496]
p_{22}	—	—	0.9543^{***}	0.9613^{***}	0.3184^{*}	0.3801^{*}	0.9552^{***}	0.9609^{***}
			(0.0296)	(0.0244)	(0.1657)	(0.1995)	(0.0294)	(0.0247)
			[32.2125]	[39.4636]	[1.9209]	[1.9052]	[32.5336]	[38.8371]
p_{21}	—	—	0.0235	0.0211	0.0211	0.0191	0.0217	0.0199
p_{12}	—	—	0.0457	0.0387	0.6816	0.6199	0.0448	0.0391
π_1	—	—	0.6602	0.6479	0.9699	0.9701	0.6734	0.6625
π_2	—	—	0.3398	0.3521	0.0301	0.0299	0.3266	0.3375
D_1	—	—	42.56	47.49	47.31	52.32	46.03	50.21
D_2	—	—	21.90	25.81	1.47	1.61	22.32	25.58

Note:
MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance. GBM denotes no auto-regression in the mean equation(s) and AR(1) an auto-regression of the first order in each mean equation. [$\cdot$] denotes standard error, ($\cdot$) and t test statistic. *, ** and *** denote significance at the 10%, 5% and 1% confidence level, respectively.

Table B.7: MS parameter for REXP(50%)-DAX30(50%) (1.1975-12.2004)

	MS(1-1)		MS(1-2)		MS(2-1)		MS(2-2)	
	GBM	**AR(1)**	**GBM**	**AR(1)**	**GBM**	**AR(1)**	**GBM**	**AR(1)**
μ_1	0.0067***	0.0064***	0.0067***	0.0064***	0.0093***	0.0097***	0.0082***	0.0083***
	(0.0016)	(0.0016)	(0.0014)	(0.0014)	(0.0015)	(0.0015)	(0.0017)	(0.0017)
	[4.2157]	[3.9332]	[4.7595]	[4.4359]	[6.1740]	[6.2312]	[4.8173]	[4.9933]
μ_2	—	—	—	—	-0.0763***	-0.0779***	0.0040	0.0032
					(0.0118)	(0.0150)	(0.0043)	(0.0042)
					[-6.4467]	[-5.2033]	[0.9218]	[0.7677]
$\phi_{1(1)}$	—	0.0207	—	0.0207	—	-0.0697	—	-0.0050
		(0.0539)		(0.0469)		(0.0518)		(0.0517)
		[0.3841]		[0.4416]		[-1.3451]		[-0.0975]
$\phi_{1(2)}$	—	—	—	—	—	-0.0833	—	0.0267
						(0.2627)		(0.0965)
						[-0.3170]		[0.2763]
σ_1^2	$9.23{\cdot}10^{-4}$***	$9.13{\cdot}10^{-4}$***	$4.66{\cdot}10^{-4}$***	$4.72{\cdot}10^{-4}$***	$7.12{\cdot}10^{-4}$***	$6.99{\cdot}10^{-4}$***	$4.80{\cdot}10^{-4}$***	$4.82{\cdot}10^{-4}$***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
	[13.4146]	[13.3960]	[7.6893]	[8.4663]	[12.1339]	[12.3046]	[7.6008]	[8.6049]
σ_2^2	—	—	0.0017***	0.0017***	—	—	0.0017***	0.0017***
			(0.0003)	(0.0003)			(0.0003)	(0.0003)
			[6.1562]	[6.4983]			[6.1532]	[6.5056]
p_{11}	—	—	0.9722***	0.9760***	0.9788***	0.9798***	0.9752***	0.9780***
			(0.0187)	(0.0177)	(0.0100)	(0.0100)	(0.0172)	(0.0157)
			[51.9707]	[55.2607]	[97.4724]	[98.3979]	[56.7792]	[62.1544]
p_{22}	—	—	0.9457***	0.9560***	0.3067*	0.3505*	0.9482***	0.9566***
			(0.0400)	(0.0305)	(0.1689)	(0.1882)	(0.0388)	(0.0299)
			[23.6550]	[31.3198]	[1.8165]	[1.8622]	[24.4437]	[32.0014]
p_{21}	—	—	0.0278	0.0240	0.0212	0.0202	0.0248	0.0220
p_{12}	—	—	0.0543	0.0440	0.6933	0.6495	0.0518	0.0434
π_1	—	—	0.6615	0.6475	0.9703	0.9699	0.6758	0.6633
π_2	—	—	0.3385	0.3525	0.0297	0.0301	0.3242	0.3367
D_1	—	—	35.95	41.71	47.08	49.58	40.28	45.37
D_2	—	—	18.40	22.71	1.44	1.54	19.32	23.03

Note:
MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance. GBM denotes no auto-regression in the mean equation(s) and AR(1) an auto-regression of the first order in each mean equation. [·] denotes standard error, (·) and t test statistic. *, ** and *** denote significance at the 10%, 5% and 1% confidence level, respectively.

Table B.8: MS parameter for REXP(60%)-DAX30(40%) (1.1975-12.2004)

	MS(1-1)		MS(1-2)		MS(2-1)		MS(2-2)	
	GBM	**AR(1)**	**GBM**	**AR(1)**	**GBM**	**AR(1)**	**GBM**	**AR(1)**
μ_1	0.0067^{***}	0.0064^{***}	0.0067^{***}	0.0064^{***}	0.0086^{***}	0.0089^{***}	0.0080^{***}	0.0081^{***}
	(0.0013)	(0.0014)	(0.0012)	(0.0013)	(0.0013)	(0.0013)	(0.0016)	(0.0016)
	[5.0804]	[4.7178]	[5.5627]	[4.9534]	[6.7072]	[6.6595]	[4.9729]	[5.1094]
μ_2	—	—	—	—	-0.0587^{***}	-0.0582^{***}	0.0043	0.0033
					(0.0107)	(0.0125)	(0.0041)	(0.0037)
					[-5.4814]	[-4.6448]	[1.0394]	[0.8758]
$\phi_{1(1)}$	—	0.0196	—	0.0196	—	-0.0543	—	0.0178
		(0.0520)		(0.0703)		(0.0579)		(0.0958)
		[0.3766]		[0.2784]		[-0.9373]		[0.1856]
$\phi_{1(2)}$	—	—	—	—	—	-0.0241	—	0.0058
						(0.2924)		(0.1701)
						[-0.0825]		[0.0343]
σ_1^2	$6.22 \cdot 10^{-4***}$	$6.15 \cdot 10^{-4***}$	$3.22 \cdot 10^{-4***}$	$3.34 \cdot 10^{-4***}$	$4.97 \cdot 10^{-4***}$	$4.85 \cdot 10^{-4***}$	$3.38 \cdot 10^{-4***}$	$3.45 \cdot 10^{-4***}$
	$(0.46 \cdot 10^{-4})$	$(0.46 \cdot 10^{-4})$	(0.0001)	(0.0001)	$(0.42 \cdot 10^{-4})$	$(0.41 \cdot 10^{-4})$	(0.0001)	(0.0001)
	[13.4138]	[13.3951]	[5.9653]	[6.4301]	[11.7477]	[11.8097]	[5.3480]	[6.8033]
σ_2^2	—	—	0.0011^{***}	0.0011^{***}	—	—	0.0011^{***}	0.0011^{***}
			(0.0002)	(0.0002)			(0.0003)	(0.0002)
			[5.1934]	[5.3855]			[4.0915]	[5.1102]
p_{11}	—	—	0.9541^{***}	0.9614^{***}	0.9789^{***}	0.9786^{***}	0.9588^{***}	0.9658^{***}
			(0.0281)	(0.0297)	(0.0112)	(0.0118)	(0.0294)	(0.0279)
			[33.9678]	[32.4224]	[87.6017]	[83.0597]	[32.5674]	[34.6155]
p_{22}	—	—	0.9218^{***}	0.9321^{***}	0.2836	0.3080^{*}	0.9196^{***}	0.9344^{***}
			(0.0559)	(0.0505)	(0.1728)	(0.1826)	(0.0844)	(0.0582)
			[16.4913]	[18.4575]	[1.6410]	[1.6870]	[10.8954]	[16.0433]
p_{21}	—	—	0.0459	0.0386	0.0211	0.0214	0.0412	0.0342
p_{12}	—	—	0.0782	0.0679	0.7164	0.6920	0.0804	0.0656
π_1	—	—	0.6301	0.6375	0.9714	0.9700	0.6614	0.6571
π_2	—	—	0.3699	0.3625	0.0286	0.0300	0.3386	0.3429
D_1	—	—	21.78	25.89	47.39	46.68	24.29	29.21
D_2	—	—	12.79	14.72	1.40	1.45	12.44	15.24

Note:
MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance. GBM denotes no auto-regression in the mean equation(s) and AR(1) an auto-regression of the first order in each mean equation. [·] denotes standard error, (·) and t test statistic. *, ** and *** denote significance at the 10%, 5% and 1% confidence level, respectively.

Table B.9: MS parameter for REXP(70%)-DAX30(30%) (1.1975-12.2004)

	MS(1-1)		MS(1-2)		MS(2-1)		MS(2-2)	
	GBM	**AR(1)**	**GBM**	**AR(1)**	**GBM**	**AR(1)**	**GBM**	**AR(1)**
μ_1	0.0066^{***}	0.0063^{***}	0.0066^{***}	0.0063^{***}	0.0079^{***}	0.0080^{***}	0.0085^{***}	0.0083^{***}
	(0.0010)	(0.0011)	(0.0010)	(0.0011)	(0.0011)	(0.0012)	(0.0017)	(0.0018)
	[6.2945]	[5.7463]	[6.6349]	[5.8669]	[7.2762]	[6.8589]	[5.0853]	[4.6648]
μ_2	—	—	—	—	-0.0423^{***}	-0.0402^{***}	0.0044	0.0035
					(0.0107)	(0.0114)	(0.0028)	(0.0028)
					[-3.9623]	[-3.5144]	[1.5798]	[1.2448]
$\phi_{1(1)}$	—	0.0242	—	0.0242	—	-0.0247	—	0.0628
		(0.0520)		(0.0583)		(0.0499)		(0.1009)
		[0.4643]		[0.4146]		[-0.4962]		[0.6223]
$\phi_{1(2)}$	—	—	—	—	—	0.0623	—	-0.0247
						(0.2619)		(0.2039)
						[0.2380]		[-0.1211]
σ_1^2	$3.93\cdot10^{-4***}$	$3.87\cdot10^{-4***}$	$1.94\cdot10^{-4***}$	$2.02\cdot10^{-4***}$	$3.29\cdot10^{-4***}$	$3.19\cdot10^{-4***}$	$2.08\cdot10^{-4***}$	2.15^{***}
	$(0.29\cdot10^{-4})$	$(0.29\cdot10^{-4})$	$(0.43\cdot10^{-4})$	$(0.45\cdot10^{-4})$	$(0.29\cdot10^{-4})$	$(0.29\cdot10^{-4})$	$(0.45\cdot10^{-4})$	$(0.39\cdot10^{-4})$
	[13.4123]	[13.3936]	[4.5424]	[4.5056]	[11.3259]	[11.0809]	[4.6424]	[5.5332]
σ_2^2	—	—	$5.84\cdot10^{-4***}$	$5.78\cdot10^{-4***}$	—	—	$5.92\cdot10^{-4***}$	$5.87\cdot10^{-4***}$
			(0.0001)	(0.0001)			(0.0001)	(0.0001)
			[5.1696]	[5.0009]			[5.5757]	[5.5776]
p_{11}	—	—	0.9297^{***}	0.9350^{***}	0.9794^{***}	0.9763^{***}	0.9431^{***}	0.9467^{***}
			(0.0362)	(0.0374)	(0.0139)	(0.0170)	(0.0328)	(0.0355)
			[25.6517]	[24.9820]	[70.5791]	[57.4994]	[28.7783]	[26.6821]
p_{22}	—	—	0.9281^{***}	0.9308^{***}	0.2347	0.2288	0.9315^{***}	0.9318^{***}
			(0.0587)	(0.0552)	(0.1783)	(0.1806)	(0.0556)	(0.0585)
			[15.8066]	[16.8467]	[1.3160]	[1.2669]	[16.7682]	[15.9196]
p_{21}	—	—	0.0703	0.0650	0.0206	0.0237	0.0569	0.0533
p_{12}	—	—	0.0719	0.0692	0.7653	0.7712	0.0685	0.0682
π_1	—	—	0.5056	0.5159	0.9738	0.9702	0.5463	0.5613
π_2	—	—	0.4944	0.4841	0.0262	0.0298	0.4537	0.4387
D_1	—	—	14.23	15.39	48.48	42.22	17.58	18.77
D_2	—	—	13.92	14.44	1.31	1.30	14.60	14.67

Note:
MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance. GBM denotes no auto-regression in the mean equation(s) and AR(1) an auto-regression of the first order in each mean equation. [·] denotes standard error, (·) and t test statistic. *, ** and *** denote significance at the 10%, 5% and 1% confidence level, respectively.

Table B.10: MS parameter for REXP(75%)-DAX30(25%) (1.1975-12.2004)

	MS(1-1)		MS(1-2)		MS(2-1)		MS(2-2)	
	GBM	**AR(1)**	**GBM**	**AR(1)**	**GBM**	**AR(1)**	**GBM**	**AR(1)**
μ_1	0.0065^{***}	0.0062^{***}	0.0065^{***}	0.0062^{***}	0.0076^{***}	0.0089^{***}	0.0045^{***}	0.0029^{*}
	(0.0009)	(0.0010)	(0.0009)	(0.0009)	(0.0010)	(0.0018)	(0.0016)	(0.0017)
	[7.0792]	[6.3423]	[7.6299]	[6.8148]	[7.2795]	[5.0413]	[2.8804]	[1.7195]
μ_2	—	—	—	—	-0.0333^{***}	-0.0003	0.0067^{***}	0.0064^{***}
					(0.0112)	(0.0037)	(0.0010)	(0.0011)
					[-2.9885]	[-0.0908]	[6.6109]	[6.0073]
$\phi_{1(1)}$	—	0.0317	—	0.0317	—	-0.1996^{*}	—	0.3091
		(0.0521)		(0.0508)		(0.1106)		(0.2184)
		[0.6086]		[0.6236]		[-1.8054]		[1.4155]
$\phi_{1(2)}$	—	—	—	—	—	0.5620^{***}	—	0.0242
						(0.1887)		(0.0504)
						[2.9786]		[0.4803]
σ_1^2	$3.05\cdot10^{-4***}$	$2.99\cdot10^{-4***}$	$0.50\cdot10^{-4***}$	$0.49\cdot10^{-4***}$	$2.60\cdot10^{-4***}$	$2.58\cdot10^{-4***}$	$0.46\cdot10^{-4***}$	$0.45\cdot10^{-4***}$
	$(0.23\cdot10^{-4})$	$(0.22\cdot10^{-4})$	$(0.19\cdot10^{-4})$	$(0.18\cdot10^{-4})$	$(0.25\cdot10^{-4})$	$(0.23\cdot10^{-4})$	$(0.15\cdot10^{-4})$	$(0.15\cdot10^{-4})$
	[13.4111]	[13.3923]	[2.6244]	[2.6978]	[10.4622]	[11.1935]	[3.0635]	[3.0281]
σ_2^2	—	—	$3.32\cdot10^{-4***}$	$3.25\cdot10^{-4***}$	—	—	$3.31\cdot10^{-4***}$	$3.26\cdot10^{-4***}$
			$(0.28\cdot10^{-4})$	$(0.27\cdot10^{-4})$			$(0.027\cdot10^{-4})$	$(0.27\cdot10^{-4})$
			[11.9811]	[12.0611]			[12.1169]	[12.0674]
p_{11}	—	—	0.9215^{***}	0.9239^{***}	0.9775^{***}	0.5602^{**}	0.9204^{***}	0.9264^{***}
			(0.0750)	(0.0723)	(0.0195)	(0.2619)	(0.0761)	(0.0701)
			[12.2807]	[12.7811]	[50.0512]	[2.1392]	[12.0973]	[13.2214]
p_{22}	—	—	0.9893^{***}	0.9897^{***}	0.1961	0.1471	0.9896^{***}	0.9893^{***}
			(0.0105)	(0.0099)	(0.1768)	(0.2692)	(0.0103)	(0.0105)
			[93.9417]	[100.0337]	[1.1090]	[0.5464]	[96.4384]	[94.2113]
p_{21}	—	—	0.0785	0.0761	0.0225	0.4398	0.0796	0.0736
p_{12}	—	—	0.0107	0.0103	0.8039	0.8529	0.0104	0.0107
π_1	—	—	0.1202	0.1197	0.9728	0.6598	0.1156	0.1270
π_2	—	—	0.8798	0.8803	0.0272	0.3402	0.8844	0.8730
D_1	—	—	12.740	13.149	44.501	2.274	12.560	13.591
D_2	—	—	93.278	96.668	1.244	1.172	96.094	93.429

Note:
MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance. GBM denotes no auto-regression in the mean equation(s) and AR(1) an auto-regression of the first order in each mean equation. [·] denotes standard error, (·) and t test statistic. *, ** and *** denote significance at the 10%, 5% and 1% confidence level, respectively.

Table B.11: MS parameter for REXP(80%)-DAX30(20%) (1.1975-12.2004)

	MS(1-1)		MS(1-2)		MS(2-1)		MS(2-2)	
	GBM	**AR(1)**	**GBM**	**AR(1)**	**GBM**	**AR(1)**	**GBM**	**AR(1)**
μ_1	0.0064^{***}	0.0060^{***}	0.0064^{***}	0.0060^{***}	0.0076^{***}	0.0083^{***}	0.0048^{***}	0.0027^{*}
	(0.0008)	(0.0009)	(0.0007)	(0.0008)	(0.0012)	(0.0013)	(0.0015)	(0.0016)
	[7.9948]	[6.9568]	[8.7030]	[7.4603]	[6.5004]	[6.4601]	[3.2073]	[1.7066]
μ_2	—	—	—	—	-0.0235^{**}	-0.0007	0.0066^{***}	0.0063^{***}
					(0.0106)	(0.0035)	(0.0009)	(0.0010)
					[-2.2169]	[-0.2117]	[7.3225]	[6.5171]
$\phi_{1(1)}$	—	0.0461	—	0.0461	—	-0.1663^{*}	—	0.3852^{**}
		(0.0522)		(0.0531)		(0.0930)		(0.1613)
		[0.8828]		[0.8683]		[-1.7881]		[2.3875]
$\phi_{1(2)}$	—	—	—	—	—	0.6619^{***}	—	0.0329
						(0.2047)		(0.0517)
						[3.2337]		[0.6371]
σ_1^2	$2.33 \cdot 10^{-4***}$	$2.28 \cdot 10^{-4***}$	$0.44 \cdot 10^{-4***}$	$0.43 \cdot 10^{-4***}$	$1.99 \cdot 10^{-4***}$	$1.95 \cdot 10^{-4***}$	$0.43 \cdot 10^{-4***}$	$0.41 \cdot 10^{-4***}$
	$(0.17 \cdot 10^{-4})$	$(0.17 \cdot 10^{-4})$	$(0.15 \cdot 10^{-4})$	$(0.14 \cdot 10^{-4})$	$(0.23 \cdot 10^{-4})$	$(0.17 \cdot 10^{-4})$	$(0.14 \cdot 10^{-4})$	$(0.13 \cdot 10^{-4})$
	[13.4095]	[13.3907]	[3.0277]	[3.1345]	[8.5165]	[11.4438]	[3.0365]	[3.1668]
σ_2^2	—	—	$2.55 \cdot 10^{-4***}$	$2.49 \cdot 10^{-4***}$	—	—	$2.56 \cdot 10^{-4***}$	$2.51 \cdot 10^{-4***}$
			$(0.21 \cdot 10^{-4})$	$(0.21 \cdot 10^{-4})$			$(0.21 \cdot 10^{-4})$	$(0.21 \cdot 10^{-4})$
			[12.0891]	[12.1562]			[11.9101]	[11.8797]
p_{11}	—	—	0.9364^{***}	0.9394^{***}	0.9673^{***}	0.5917^{***}	0.9310^{***}	0.9327^{***}
			(0.0600)	(0.0567)	(0.0367)	(0.2139)	(0.0670)	(0.0663)
			[15.6132]	[16.5557]	[26.3405]	[2.7666]	[13.9049]	[14.0776]
p_{22}	—	—	0.9901^{***}	0.9905^{***}	0.1611	0.0374	0.9894^{***}	0.9887^{***}
			(0.0090)	(0.0084)	(0.1701)	(0.2054)	(0.0104)	(0.0113)
			[110.0704]	[118.4324]	[0.9471]	[0.1823]	[94.7167]	[87.5118]
p_{21}	—	—	0.0636	0.0606	0.0327	0.4083	0.0690	0.0673
p_{12}	—	—	0.0099	0.0095	0.8389	0.9626	0.0106	0.0113
π_1	—	—	0.1344	0.1354	0.9625	0.7021	0.1333	0.1434
π_2	—	—	0.8656	0.8646	0.0375	0.2979	0.8667	0.8566
D_1	—	—	15.7122	16.5109	30.5910	2.4490	14.4949	14.8511
D_2	—	—	101.1635	105.4021	1.1920	1.0389	94.2090	88.7184

Note:
MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance. GBM denotes no auto-regression in the mean equation(s) and AR(1) an auto-regression of the first order in each mean equation. [$\cdot$] denotes standard error, ($\cdot$) and t test statistic. *, ** and *** denote significance at the 10%, 5% and 1% confidence level, respectively.

Table B.12: MS parameter for REXP(90%)-DAX30(10%) (1.1975-12.2004)

	MS(1-1)		MS(1-2)		MS(2-1)		MS(2-2)	
	GBM	**AR(1)**	**GBM**	**AR(1)**	**GBM**	**AR(1)**	**GBM**	**AR(1)**
μ_1	0.0063^{***}	0.0055^{***}	0.0063^{***}	0.0055^{***}	0.0073^{***}	0.0080^{***}	0.0068^{***}	0.0059^{***}
	(0.0006)	(0.0007)	(0.0006)	(0.0007)	(0.0008)	(0.0010)	(0.0008)	(0.0009)
	[9.9460]	[7.8287]	[10.4366]	[8.0988]	[9.0828]	[7.9363]	[8.8634]	[6.6865]
μ_2	—	—	—	—	-0.0184^{**}	-0.0003	0.0033	0.0006
					(0.0072)	(0.0021)	(0.0042)	(0.0033)
					[-2.5476]	[-0.1231]	[0.7854]	[0.1760]
$\phi_{1(1)}$	—	0.1100^{**}	—	0.1100^{**}	—	-0.1852^{**}	—	0.2231^{***}
		(0.0518)		(0.0524)		(0.0922)		(0.0792)
		[2.1239]		[2.0980]		[-2.0086]		[2.8153]
$\phi_{1(2)}$	—	—	—	—	—	0.7539^{***}	—	-0.2439
						(0.1557)		(0.1827)
						[4.8410]		[-1.3348]
σ_1^2	$1.43\cdot10^{-4***}$	$1.38\cdot10^{-4***}$	$1.05\cdot10^{-4***}$	$1.03\cdot10^{-4***}$	$1.17\cdot10^{-4***}$	$1.09\cdot10^{-4***}$	$1.08\cdot10^{-4***}$	$0.94\cdot10^{-4***}$
	$(0.11\cdot10^{-4})$	$(0.10\cdot10^{-4})$	$(0.12\cdot10^{-4})$	$(0.10\cdot10^{-4})$	$(0.12\cdot10^{-4})$	$(0.10\cdot10^{-4})$	$(0.13\cdot10^{-4})$	$(0.15\cdot10^{-4})$
	[13.4050]	[13.3860]	[9.1297]	[9.7866]	[9.6435]	[10.7219]	[8.4352]	[6.3452]
σ_2^2	—	—	$3.13\cdot10^{-4***}$	$2.99\cdot10^{-4***}$	—	—	$3.13\cdot10^{-4***}$	$2.52\cdot10^{-4***}$
			(0.0001)	(0.0001)			(0.0001)	(0.0001)
			[4.1971]	[4.1231]			[3.7640]	[3.6257]
p_{11}	—	—	0.9887^{***}	0.9898^{***}	0.9663^{***}	0.5036^{***}	0.9882^{***}	0.9440^{***}
			(0.0086)	(0.0088)	(0.0264)	(0.1505)	(0.0099)	(0.0482)
			[114.5919]	[113.1124]	[36.5398]	[3.3466]	[99.8033]	[19.5718]
p_{22}	—	—	0.9370^{***}	0.9531^{***}	0.1998	0.1724	0.9301^{***}	0.7638^{***}
			(0.0479)	(0.0338)	(0.1761)	(0.1863)	(0.0597)	(0.1162)
			[19.5472]	[28.1983]	[1.1341]	[0.9256]	[15.5728]	[6.5735]
p_{21}	—	—	0.0113	0.0102	0.0337	0.4964	0.0118	0.0560
p_{12}	—	—	0.0630	0.0469	0.8002	0.8276	0.0699	0.2362
π_1	—	—	0.8474	0.8217	0.9596	0.6251	0.8554	0.8084
π_2	—	—	0.1526	0.1783	0.0404	0.3749	0.1446	0.1916
D_1	—	—	88.17	98.29	29.67	2.01	84.67	17.86
D_2	—	—	15.88	21.33	1.25	1.21	14.31	4.23

Note:
MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance. GBM denotes no auto-regression in the mean equation(s) and AR(1) an auto-regression of the first order in each mean equation. [·] denotes standard error, (·) and t test statistic. *, ** and *** denote significance at the 10%, 5% and 1% confidence level, respectively.

Table B.13: MS parameter for REXP (1.1975-12.2004)

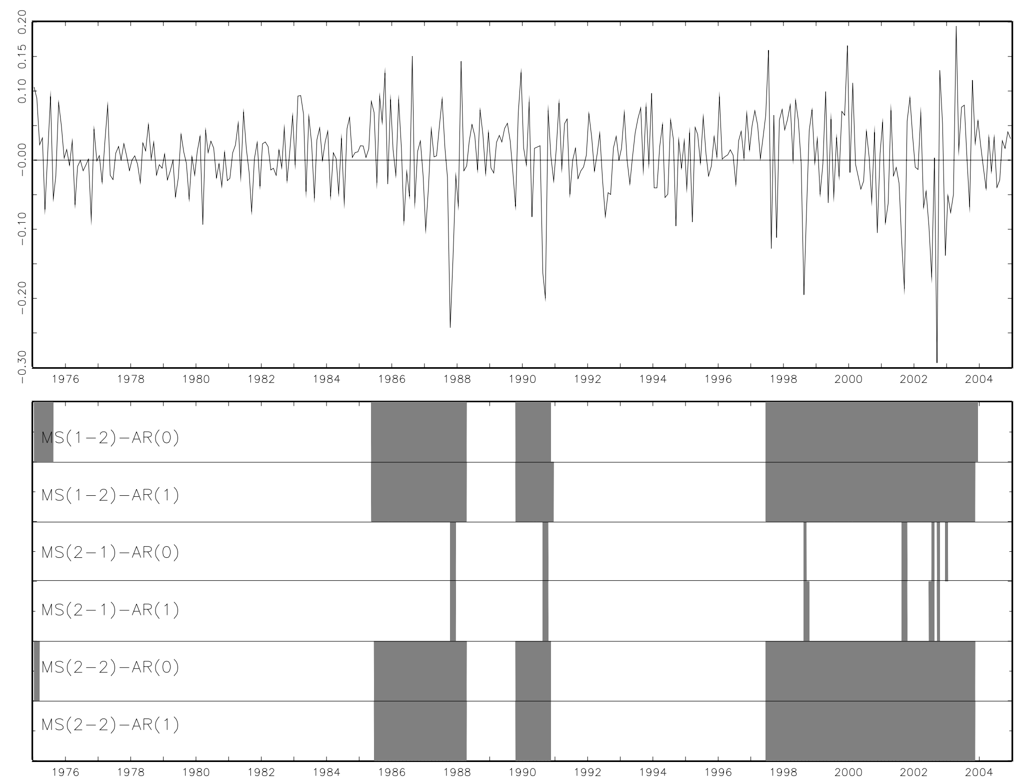
	MS(1-1)		**MS(1-2)**		**MS(2-1)**		**MS(2-2)**	
	GBM	**AR(1)**	**GBM**	**AR(1)**	**GBM**	**AR(1)**	**GBM**	**AR(1)**
μ_1	0.0061^{***}	0.0048^{***}	0.0061^{***}	0.0048^{***}	0.0069^{***}	0.0077^{***}	0.0062^{***}	0.0049^{***}
	(0.0006)	(0.0006)	(0.0005)	(0.0006)	(0.0007)	(0.0010)	(0.0006)	(0.0007)
	[10.4700]	[7.4175]	[11.1744]	[7.8653]	[10.0594]	[7.6620]	[10.3349]	[7.1244]
μ_2	—	—	—	—	-0.0213^{***}	-0.0020	0.0052	-0.0038
					(0.0081)	(0.0029)	(0.0034)	(0.0050)
					[-2.6412]	[-0.6875]	[1.5363]	[-0.7621]
$\phi_{1(1)}$	—	0.1997^{***}	—	0.1997^{***}	—	-0.0610	—	0.2756^{***}
		(0.0512)		(0.0514)		(0.1155)		(0.0569)
		[3.9036]		[3.8836]		[-0.5286]		[4.8479]
$\phi_{1(2)}$	—	—	—	—	—	0.8212^{***}	—	-0.3569
						(0.1798)		(0.2556)
						[4.5665]		[-1.3965]
σ_1^2	$1.21 \cdot 10^{-4***}$	$1.14 \cdot 10^{-4***}$	$0.90 \cdot 10^{-4***}$	$0.84 \cdot 10^{-4***}$	$0.98 \cdot 10^{-4***}$	$0.90 \cdot 10^{-4***}$	$0.90 \cdot 10^{-4***}$	$0.83 \cdot 10^{-4***}$
	$(0.09 \cdot 10^{-4})$	$(0.08 \cdot 10^{-4})$	$(0.08 \cdot 10^{-4})$	$(0.08 \cdot 10^{-4})$	$(0.10 \cdot 10^{-4})$	$(0.08 \cdot 10^{-4})$	$(0.09 \cdot 10^{-4})$	$(0.10 \cdot 10^{-4})$
	[13.4029]	[13.3835]	[10.8067]	[10.8810]	[10.0776]	[11.3808]	[10.3429]	[8.6508]
σ_2^2	—	—	$3.47 \cdot 10^{-4***}$	$3.40 \cdot 10^{-4***}$	—	—	$3.44 \cdot 10^{-4***}$	$3.15 \cdot 10^{-4**}$
			(0.0001)	(0.0001)			(0.0001)	(0.0001)
			[3.9464]	[3.8216]			[3.8576]	[2.3034]
p_{11}	—	—	0.9947^{***}	0.9946^{***}	0.9759^{***}	0.5052^{**}	0.9941^{***}	0.9756^{***}
			(0.0057)	(0.0061)	(0.0208)	(0.2433)	(0.0066)	(0.0233)
			[173.9853]	[164.0404]	[46.8879]	[2.0763]	[151.3679]	[41.8672]
p_{22}	—	—	0.9543^{***}	0.9583^{***}	0.1989	0.1692	0.9513^{***}	0.7252^{***}
			(0.0426)	(0.0394)	(0.1645)	(0.1648)	(0.0459)	(0.1349)
			[22.3826]	[24.3296]	[1.2089]	[1.0267]	[20.7237]	[5.3750]
p_{21}	—	—	0.0053	0.0054	0.0241	0.4948	0.0059	0.0244
p_{12}	—	—	0.0457	0.0417	0.8011	0.8308	0.0487	0.2748
π_1	—	—	0.8960	0.8852	0.9708	0.6267	0.8919	0.9185
π_2	—	—	0.1040	0.1148	0.0292	0.3733	0.1081	0.0815
D_1	—	—	188.57	184.91	41.52	2.02	169.47	41.02
D_2	—	—	21.89	23.99	1.25	1.20	20.54	3.64

Note:
MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance. GBM denotes no auto-regression in the mean equation(s) and AR(1) an auto-regression of the first order in each mean equation. [·] denotes standard error, (·) and t test statistic. *, ** and *** denote significance at the 10%, 5% and 1% confidence level, respectively.

Appendix C

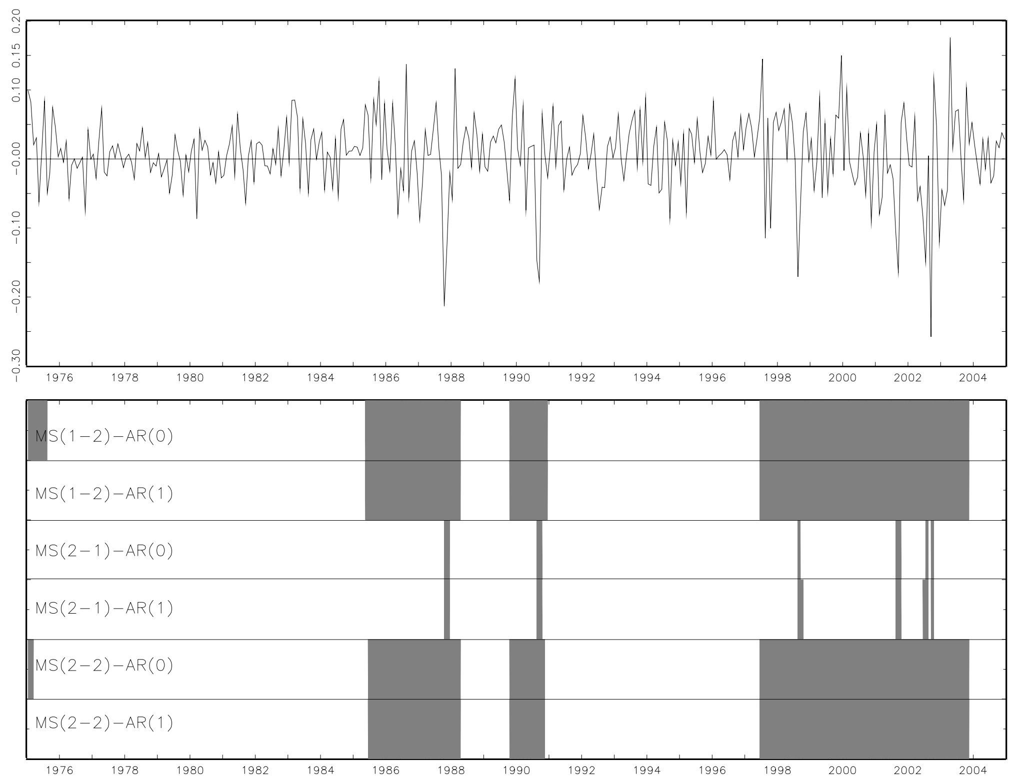
States of Markov switching models (1.1975-12.2004)

Figure C.1: States in the Markov switching models: DAX30



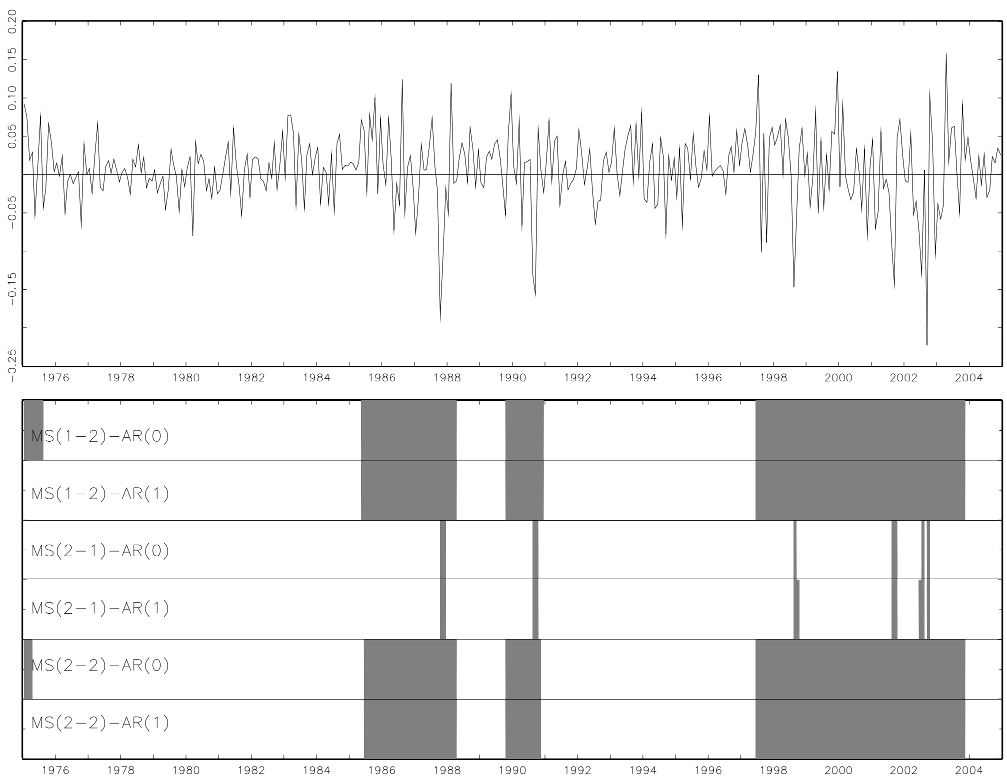
Note:
The top panel shows log-returns of the pure DAX30 portfolio, the bottom panel shows states in the Markov switching models. For the MS(1-2), MS(1-2)-AR(1), MS(2-2) and MS(2-2)-AR(1), the grey area represents the high volatility regime and for MS(2-1) and MS(2-1)-AR(1), the grey area represents the low mean regime (i.e the smoothed probability $\Pr[Z_t = 2|\mathscr{Y}_T] > 0.5$).

Figure C.2: States in the Markov switching models: REXP(10%)-DAX30(90%)



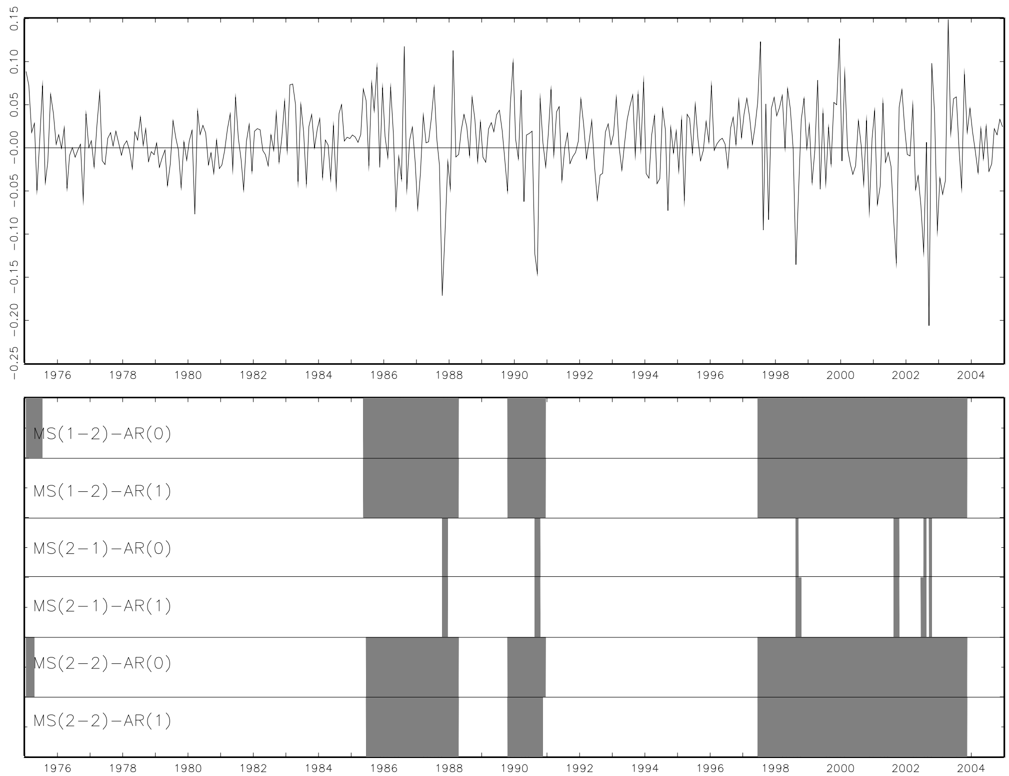
Note:
The top panel shows log-returns of the REXP(10%)-DAX30(90%) portfolio, the bottom panel shows states in the Markov switching models. For the MS(1-2), MS(1-2)-AR(1), MS(2-2) and MS(2-2)-AR(1), the grey area represents the high volatility regime and for MS(2-1) and MS(2-1)-AR(1), the grey area represents the low mean regime (i.e the smoothed probability $\Pr[Z_t = 2|\mathscr{Y}_T] > 0.5$).

Figure C.3: States in the Markov switching models: REXP(20%)-DAX30(80%)



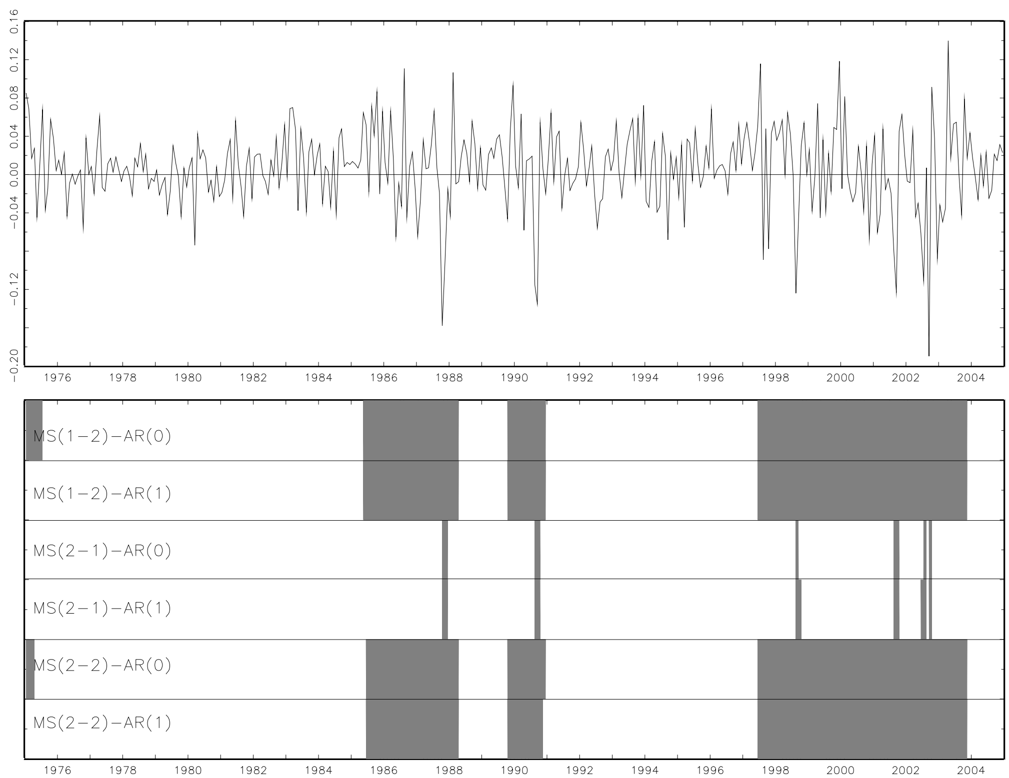
Note:
The top panel shows log-returns of the REXP(20%)-DAX30(80%) portfolio, the bottom panel shows states in the Markov switching models. For the MS(1-2), MS(1-2)-AR(1), MS(2-2) and MS(2-2)-AR(1), the grey area represents the high volatility regime and for MS(2-1) and MS(2-1)-AR(1), the grey area represents the low mean regime (i.e the smoothed probability $\Pr[Z_t = 2|\mathscr{Y}_T] > 0.5$).

Figure C.4: States in the Markov switching models: REXP(25%)-DAX30(75%)



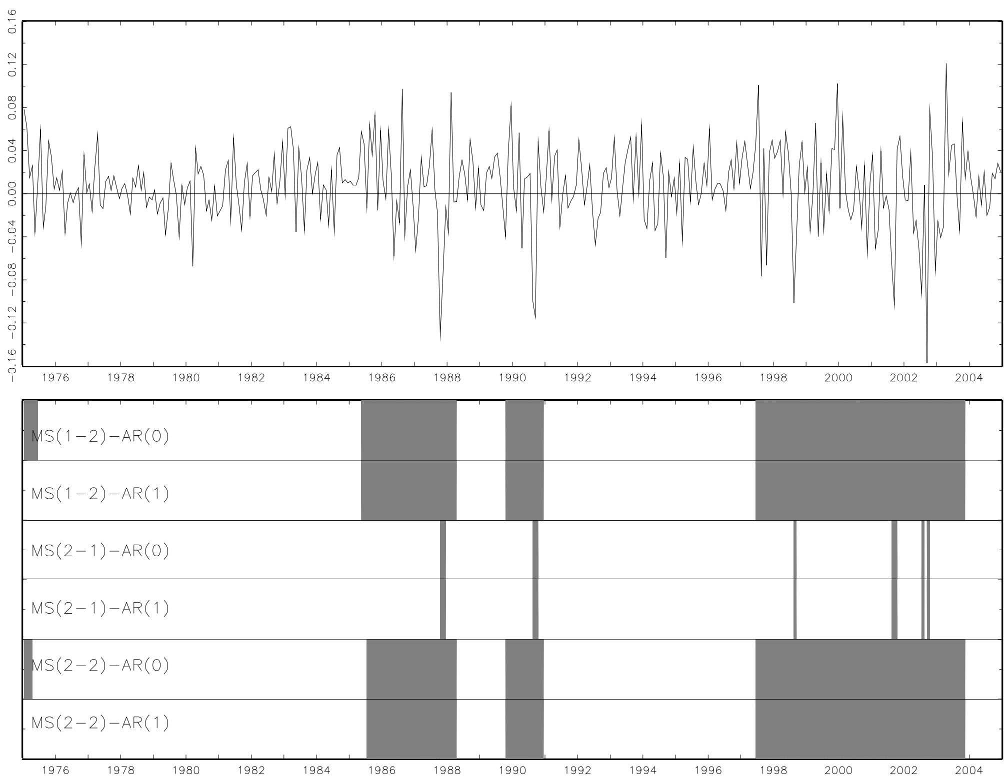
Note:
The top panel shows log-returns of the REXP(25%)-DAX30(75%) portfolio, the bottom panel shows states in the Markov switching models. For the MS(1-2), MS(1-2)-AR(1), MS(2-2) and MS(2-2)-AR(1), the grey area represents the high volatility regime and for MS(2-1) and MS(2-1)-AR(1), the grey area represents the low mean regime (i.e the smoothed probability $\Pr[Z_t = 2|\mathscr{Y}_T] > 0.5$).

Figure C.5: States in the Markov switching models: REXP(30%)-DAX30(70%)



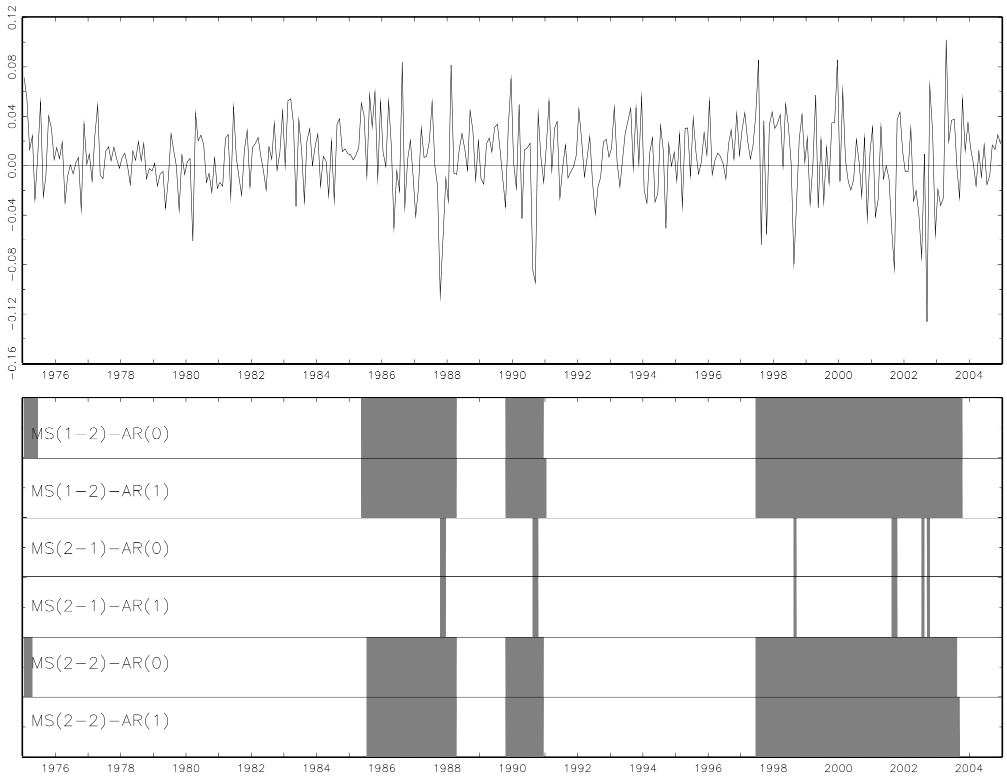
Note:
The top panel shows log-returns of the REXP(30%)-DAX30(70%) portfolio, the bottom panel shows states in the Markov switching models. For the MS(1-2), MS(1-2)-AR(1), MS(2-2) and MS(2-2)-AR(1), the grey area represents the high volatility regime and for MS(2-1) and MS(2-1)-AR(1), the grey area represents the low mean regime (i.e the smoothed probability $\Pr[Z_t = 2|\mathscr{Y}_T] > 0.5$).

Figure C.6: States in the Markov switching models: REXP(40%)-DAX30(60%)



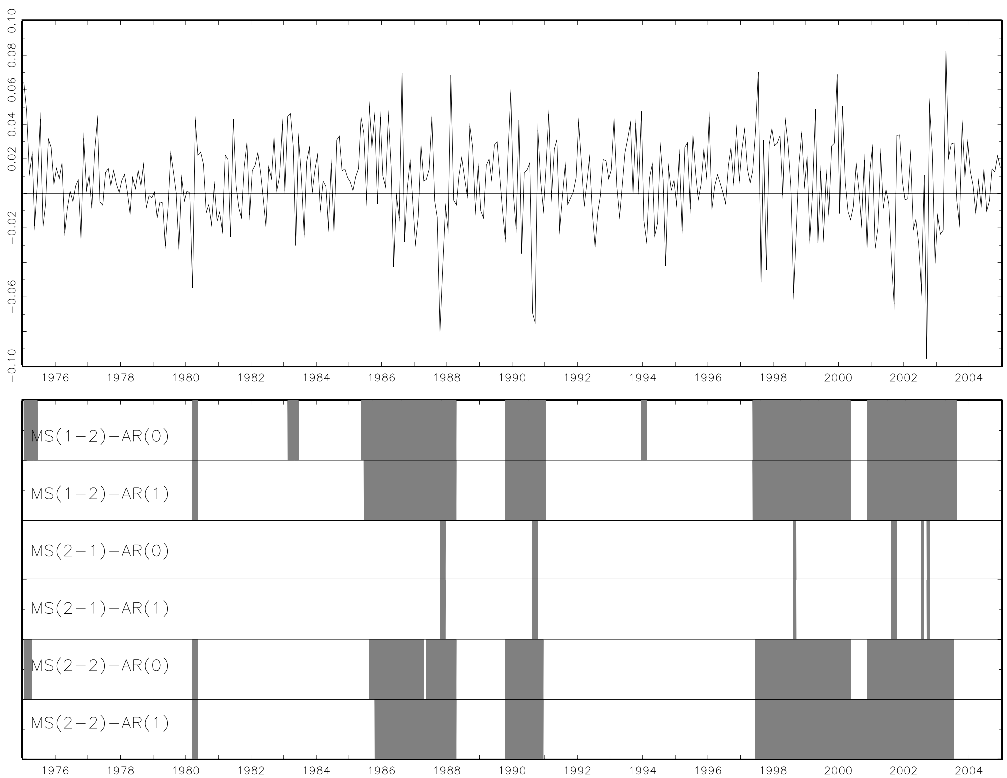
Note:
The top panel shows log-returns of the REXP(40%)-DAX30(60%) portfolio, the bottom panel shows states in the Markov switching models. For the MS(1-2), MS(1-2)-AR(1), MS(2-2) and MS(2-2)-AR(1), the grey area represents the high volatility regime and for MS(2-1) and MS(2-1)-AR(1), the grey area represents the low mean regime (i.e the smoothed probability $\Pr[Z_t = 2|\mathscr{Y}_T] > 0.5$).

Figure C.7: States in the Markov switching models: REXP(50%)-DAX30(50%)



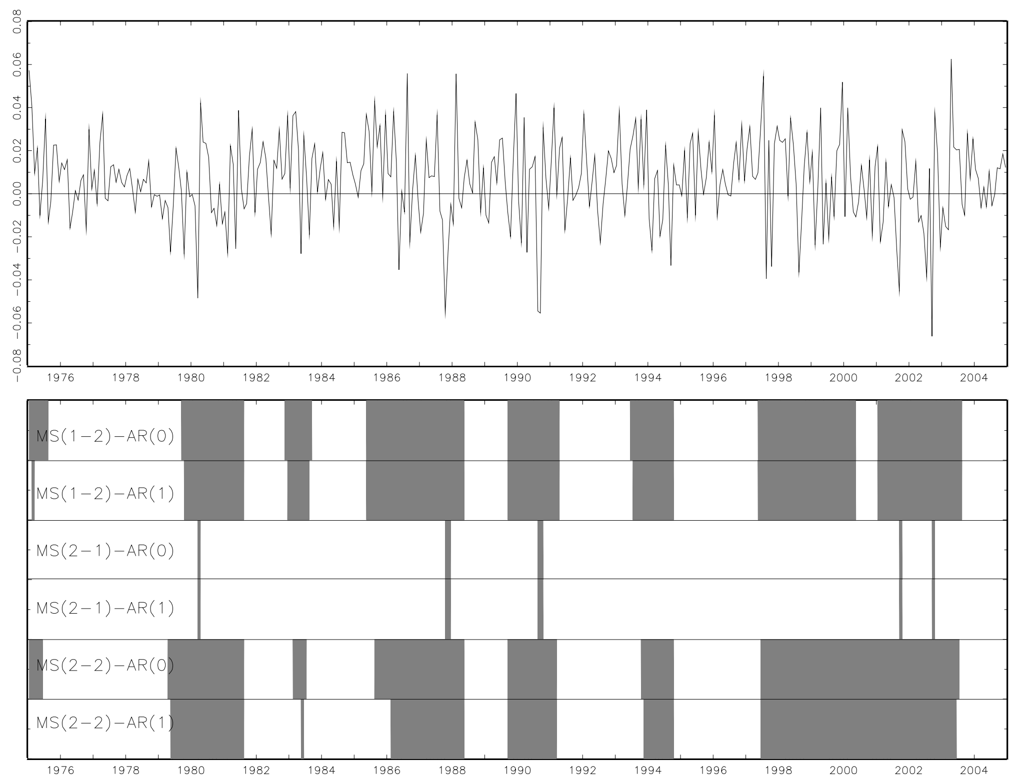
Note:
The top panel shows log-returns of the REXP(50%)-DAX30(50%) portfolio, the bottom panel shows states in the Markov switching models. For the MS(1-2), MS(1-2)-AR(1), MS(2-2) and MS(2-2)-AR(1), the grey area represents the high volatility regime and for MS(2-1) and MS(2-1)-AR(1), the grey area represents the low mean regime (i.e the smoothed probability $\Pr[Z_t = 2|\mathscr{Y}_T] > 0.5$).

Figure C.8: States in the Markov switching models: REXP(60%)-DAX30(40%)



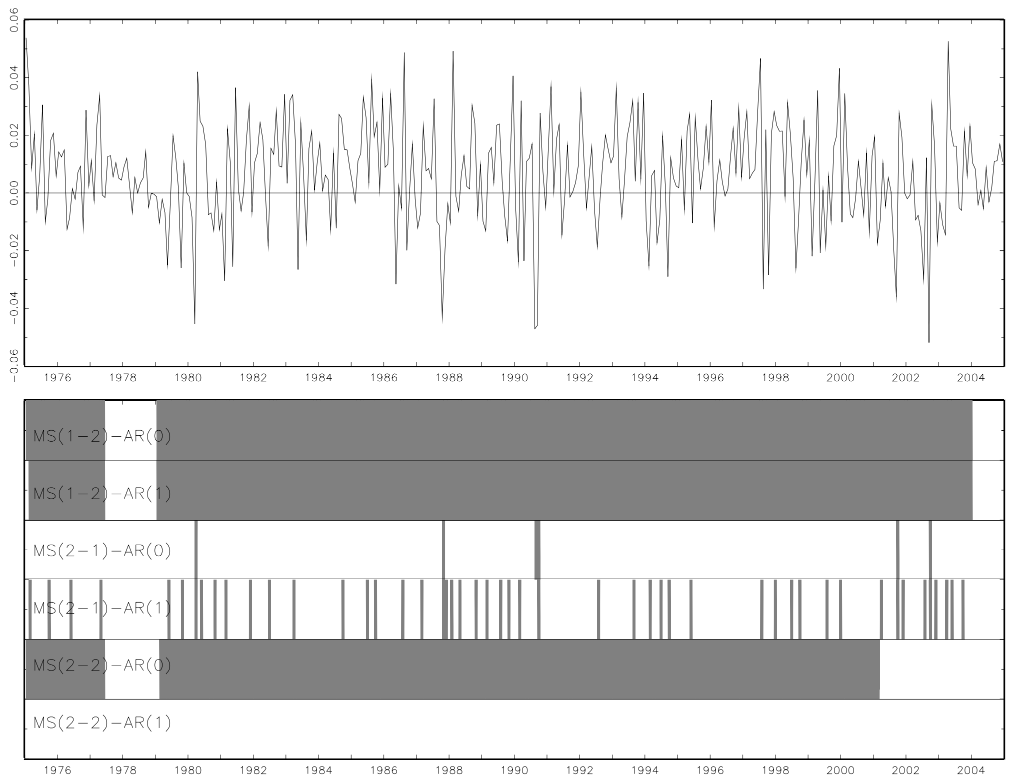
Note:
The top panel shows log-returns of the REXP(60%)-DAX30(40%) portfolio, the bottom panel shows states in the Markov switching models. For the MS(1-2), MS(1-2)-AR(1), MS(2-2) and MS(2-2)-AR(1), the grey area represents the high volatility regime and for MS(2-1) and MS(2-1)-AR(1), the grey area represents the low mean regime (i.e the smoothed probability $\Pr[Z_t = 2|\mathscr{Y}_T] > 0.5$).

Figure C.9: States in the Markov switching models: REXP(70%)-DAX30(30%)



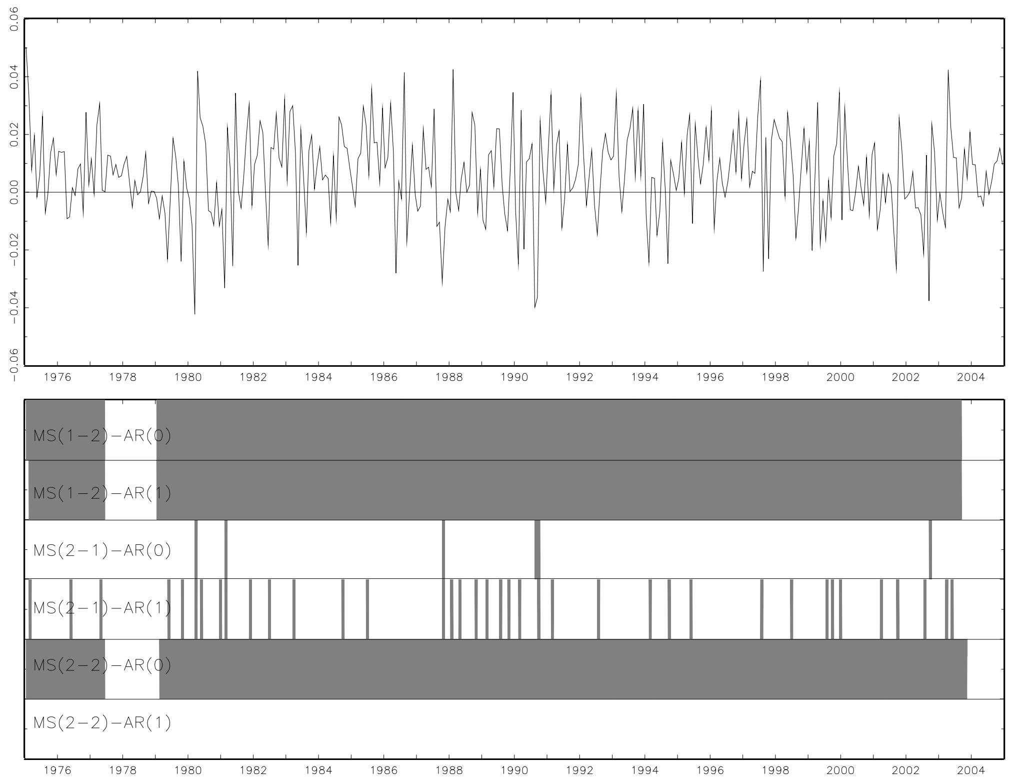
Note:
The top panel shows log-returns of the REXP(70%)-DAX30(30%) portfolio, the bottom panel shows states in the Markov switching models. For the MS(1-2), MS(1-2)-AR(1), MS(2-2) and MS(2-2)-AR(1), the grey area represents the high volatility regime and for MS(2-1) and MS(2-1)-AR(1), the grey area represents the low mean regime (i.e the smoothed probability $\Pr[Z_t = 2|\mathscr{Y}_T] > 0.5$).

Figure C.10: States in the Markov switching models: REXP(75%)-DAX30(25%)



Note:
The top panel shows log-returns of the REXP(75%)-DAX30(25%) portfolio, the bottom panel shows states in the Markov switching models. For the MS(1-2), MS(1-2)-AR(1), MS(2-2) and MS(2-2)-AR(1), the grey area represents the high volatility regime and for MS(2-1) and MS(2-1)-AR(1), the grey area represents the low mean regime (i.e the smoothed probability $\Pr[Z_t = 2|\mathscr{Y}_T] > 0.5$).

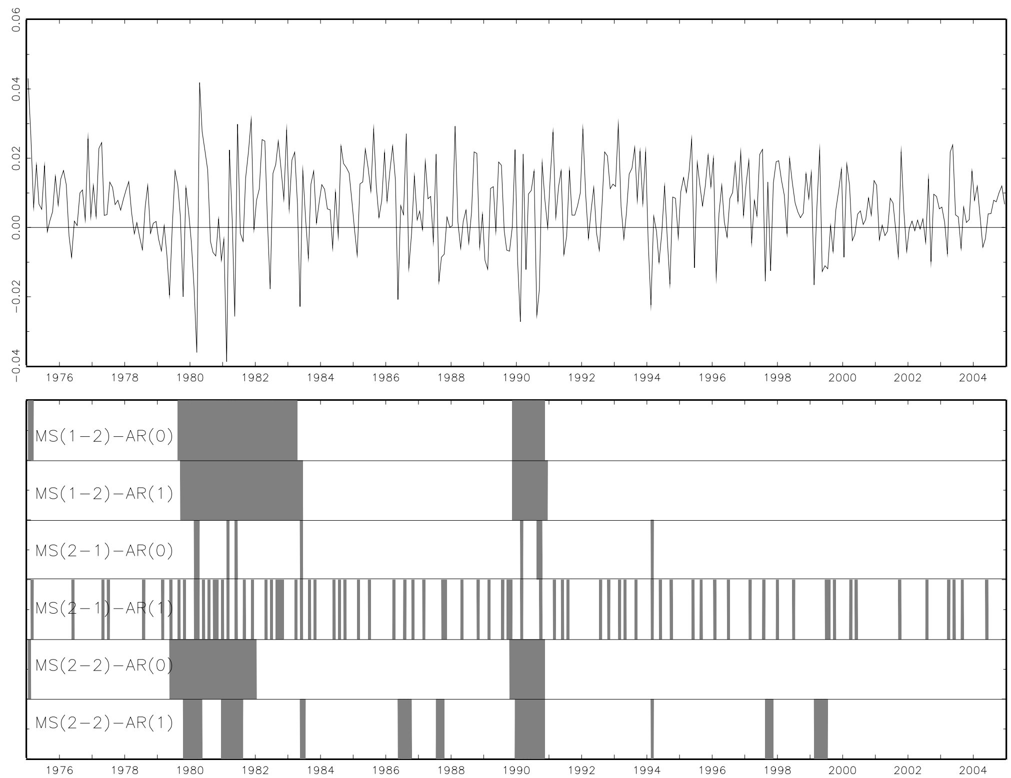
Figure C.11: States in the Markov switching models: REXP(80%)-DAX30(20%)



Note:

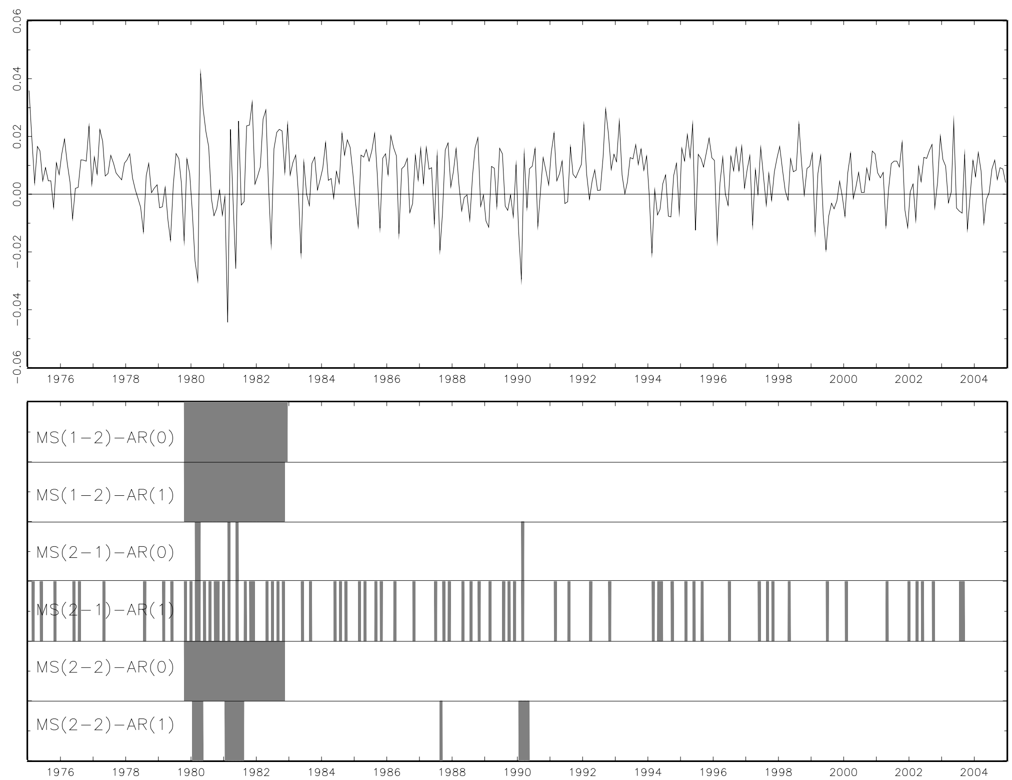
The top panel shows log-returns of the REXP(80%)-DAX30(20%) portfolio, the bottom panel shows states in the Markov switching models. For the MS(1-2), MS(1-2)-AR(1), MS(2-2) and MS(2-2)-AR(1), the grey area represents the high volatility regime and for MS(2-1) and MS(2-1)-AR(1), the grey area represents the low mean regime (i.e the smoothed probability $\Pr[Z_t = 2|\mathscr{Y}_T] > 0.5$).

Figure C.12: States in the Markov switching models: REXP(90%)-DAX30(10%)



Note:
The top panel shows log-returns of the REXP(90%)-DAX30(10%) portfolio, the bottom panel shows states in the Markov switching models. For the MS(1-2), MS(1-2)-AR(1), MS(2-2) and MS(2-2)-AR(1), the grey area represents the high volatility regime and for MS(2-1) and MS(2-1)-AR(1), the grey area represents the low mean regime (i.e the smoothed probability $\Pr[Z_t = 2|\mathscr{Y}_T] > 0.5$).

Figure C.13: States in the Markov switching models: REXP

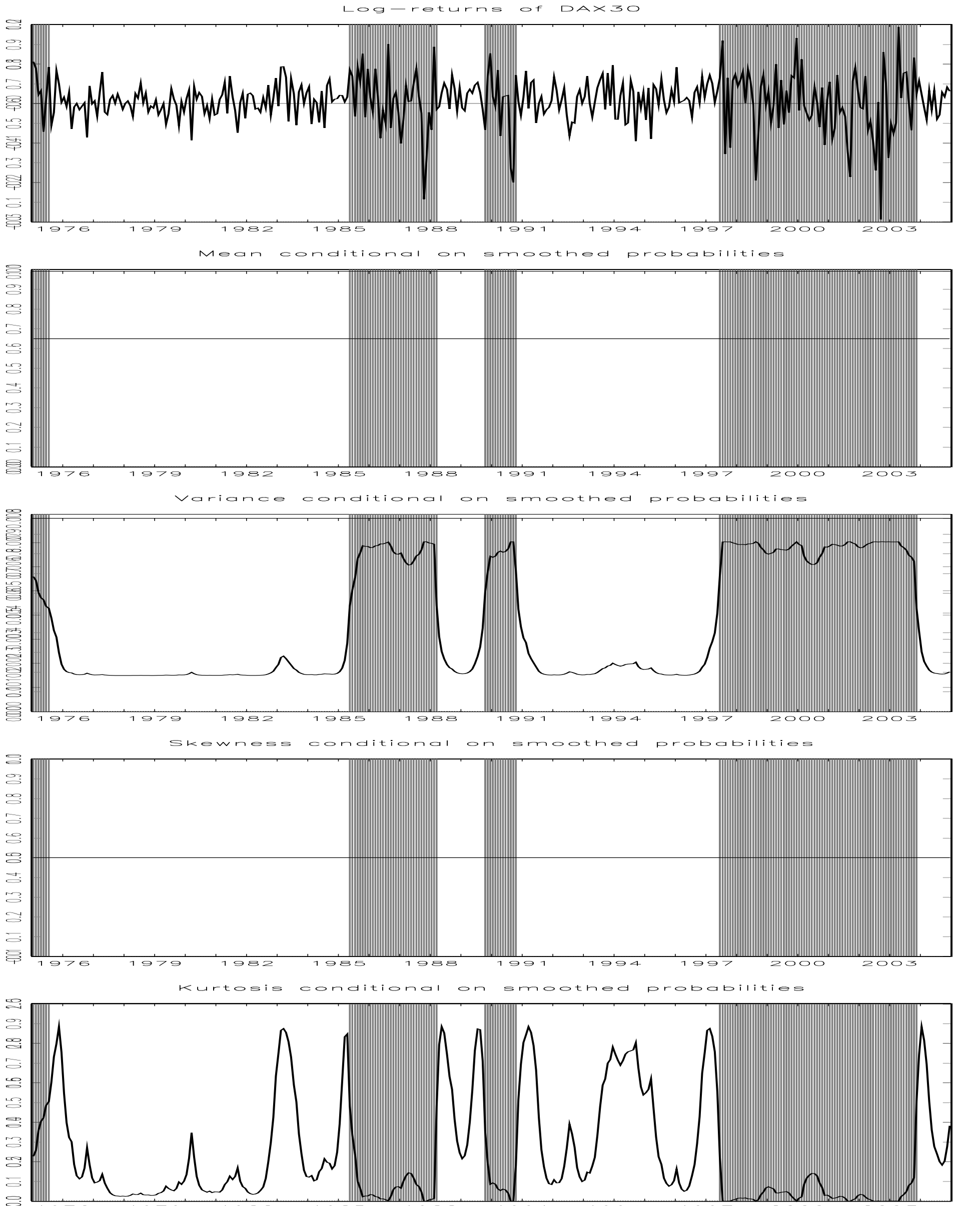


Note:
The top panel shows log-returns of the pure REXP portfolio, the bottom panel shows states in the Markov switching models. For the MS(1-2), MS(1-2)-AR(1), MS(2-2) and MS(2-2)-AR(1), the grey area represents the high volatility regime and for MS(2-1) and MS(2-1)-AR(1), the grey area represents the low mean regime (i.e the smoothed probability $\Pr[Z_t = 2|\mathscr{Y}_T] > 0.5$).

Appendix D

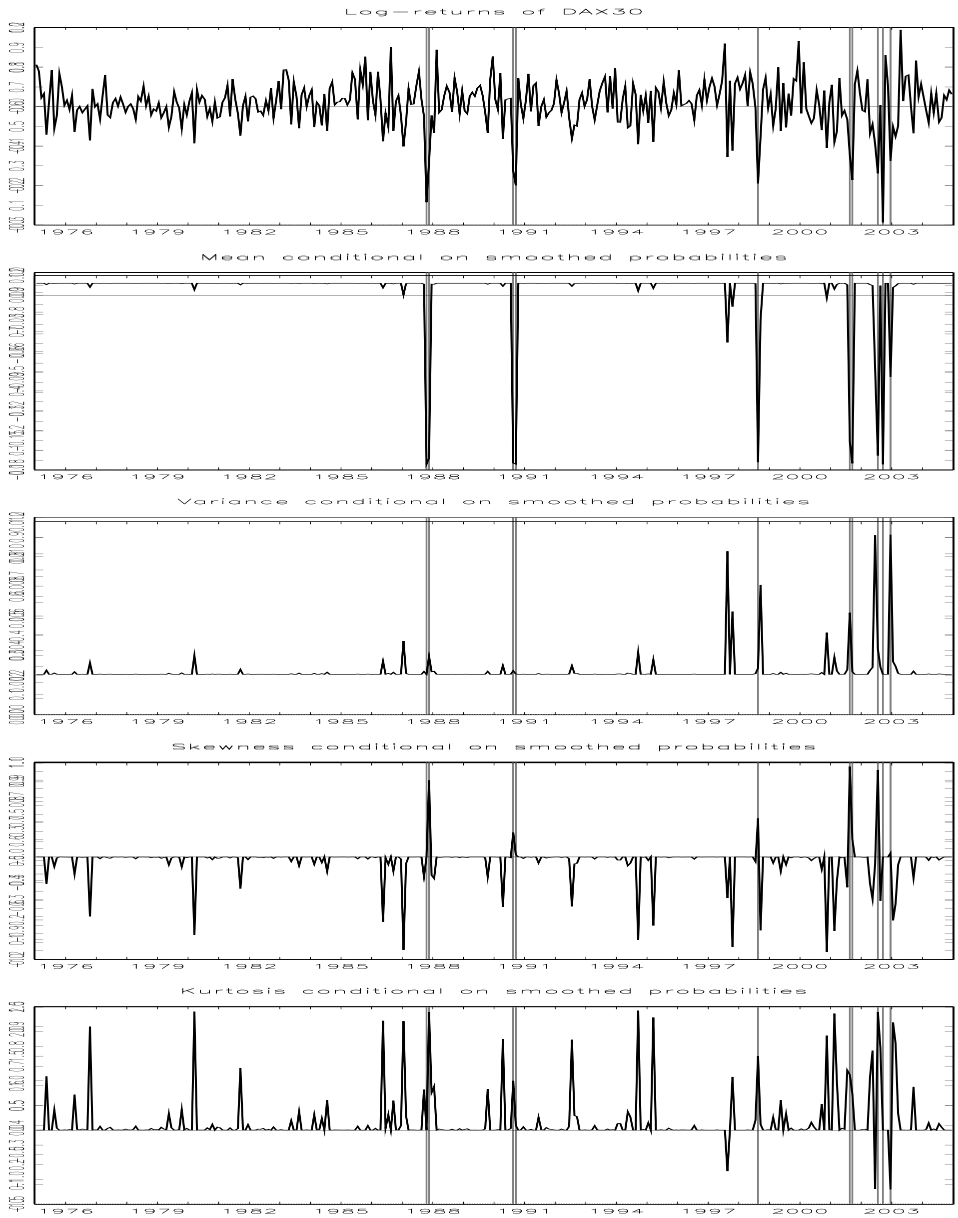
Conditional moments (1.1975-12.2004) – graphs

Figure D.1: Conditional moments: DAX30 - MS(1-2)-AR(0)



Note:
The top panel shows the log-returns of the pure DAX30 portfolio. The remaining panels indicate the moments for the MS(1-2) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high volatility regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

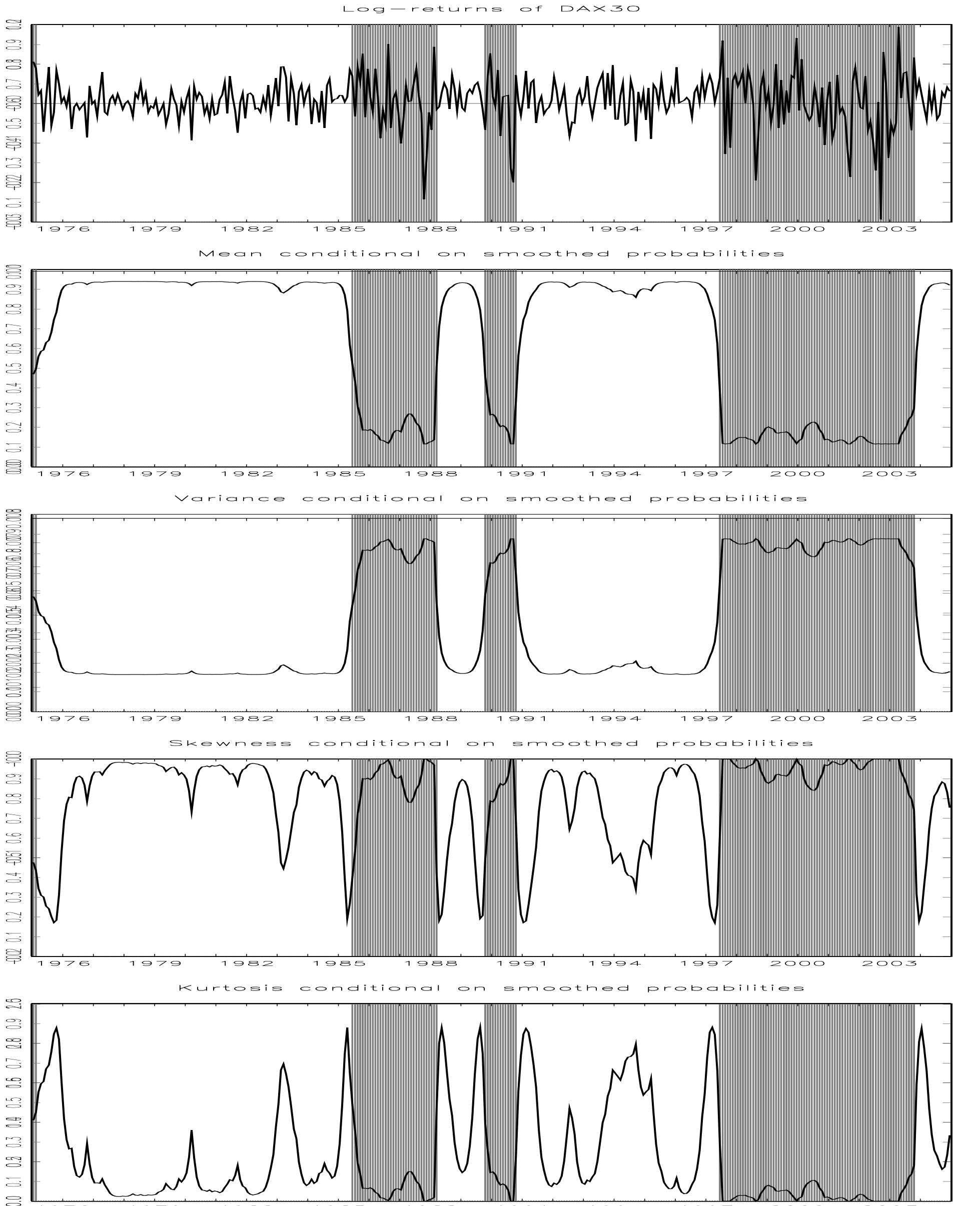
Figure D.2: Conditional moments: DAX30 - MS(2-1)-AR(0)



Note:

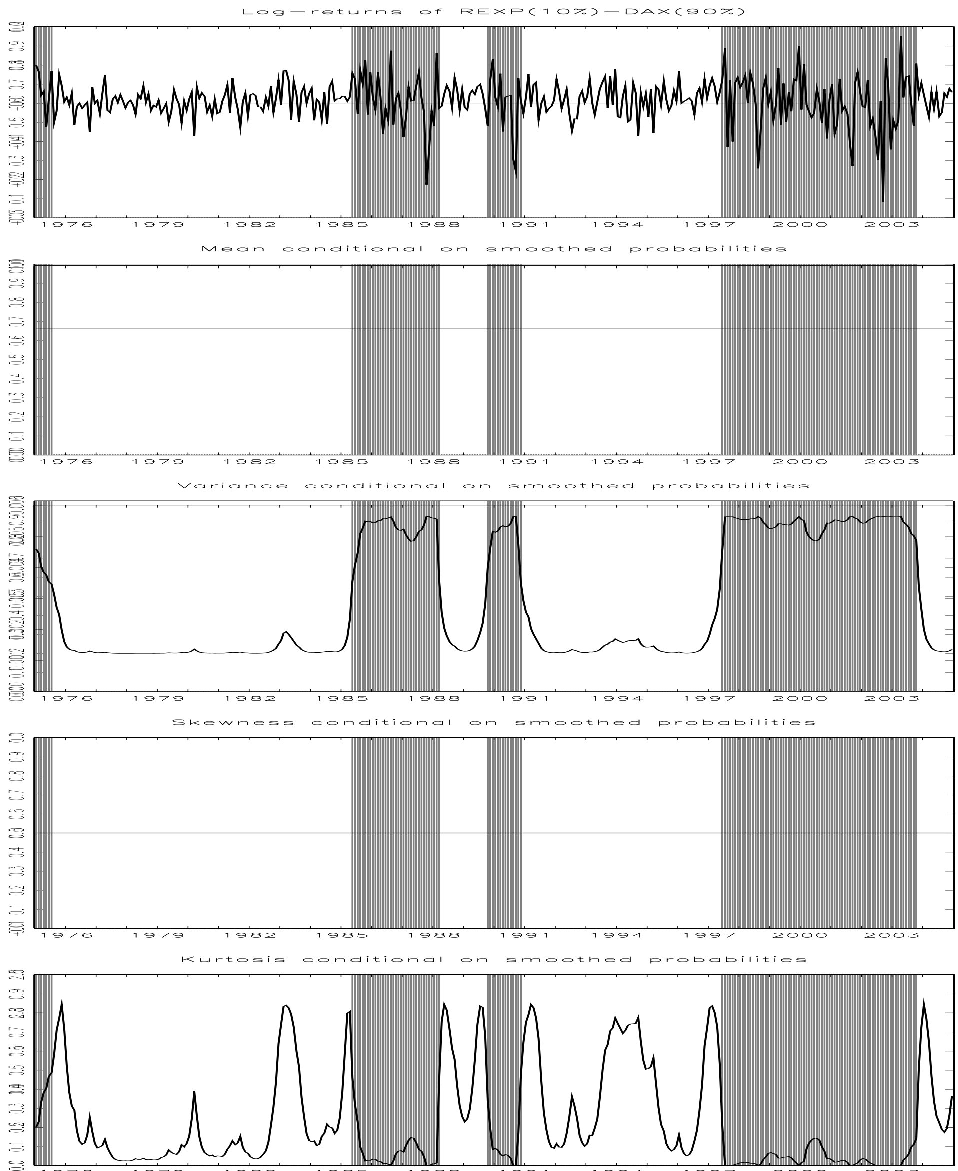
The top panel shows the log-returns of the pure DAX30 portfolio. The remaining panels indicate the moments for the MS(2-1) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high mean regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.3: Conditional moments: DAX30 - MS(2-2)-AR(0)



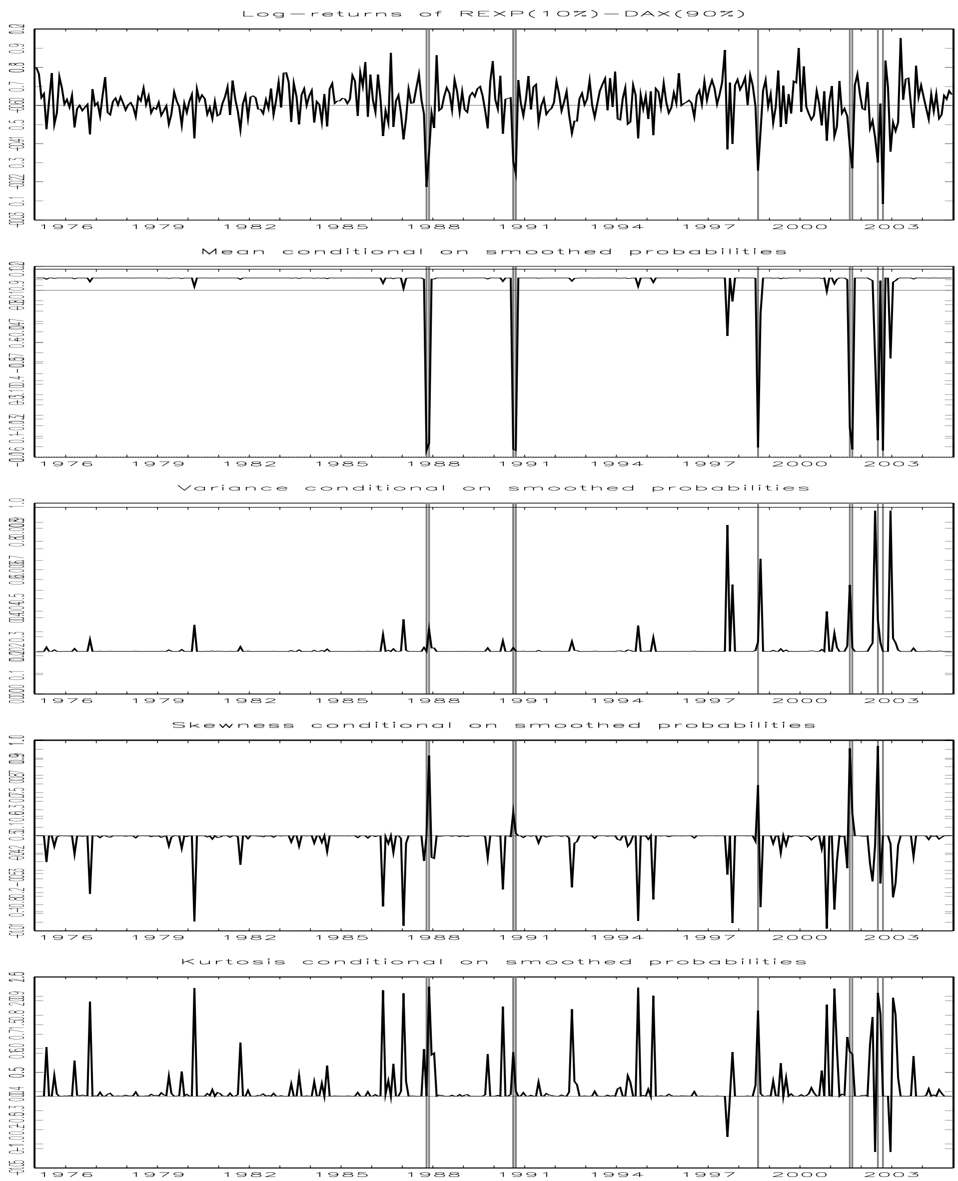
Note:
The top panel shows the log-returns of the pure DAX30 portfolio. The remaining panels indicate the moments for the MS(2-2) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high volatility regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.4: Conditional moments: REXP(10%)-DAX30(90%) - MS(1-2)-AR(0)



Note:
The top panel shows the log-returns of the REXP(10%)-DAX30(90%) portfolio. The remaining panels indicate the moments for the MS(1-2) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high volatility regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

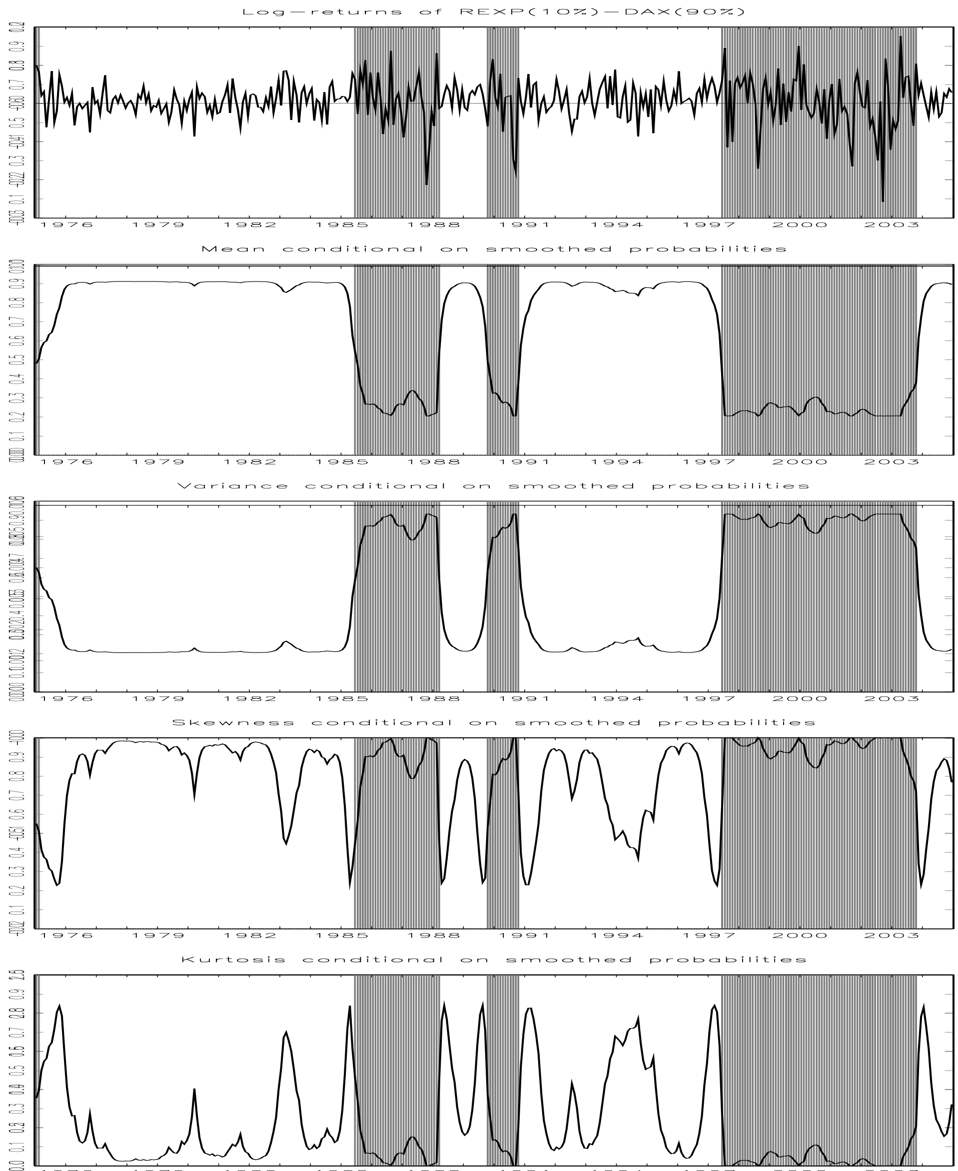
Figure D.5: Conditional moments: REXP(10%)-DAX30(90%) - MS(2-1)-AR(0)



Note:

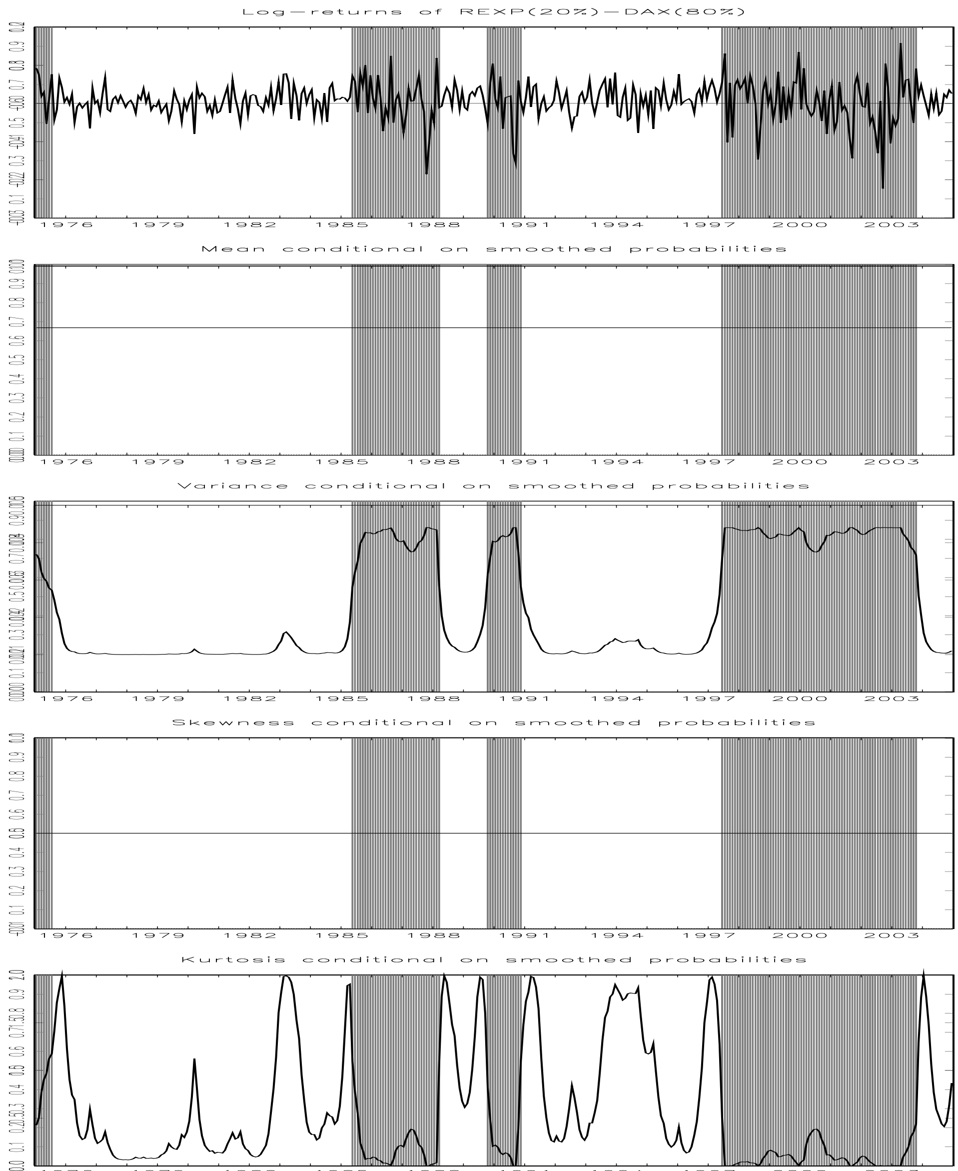
The top panel shows the log-returns of the REXP(10%)-DAX30(90%) portfolio. The remaining panels indicate the moments for the MS(2-1) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high mean regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.6: Conditional moments: REXP(10%)-DAX30(90%) - MS(2-2)-AR(0)



Note:
The top panel shows the log-returns of the REXP(10%)-DAX30(90%) portfolio. The remaining panels indicate the moments for the MS(2-2) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high volatility regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

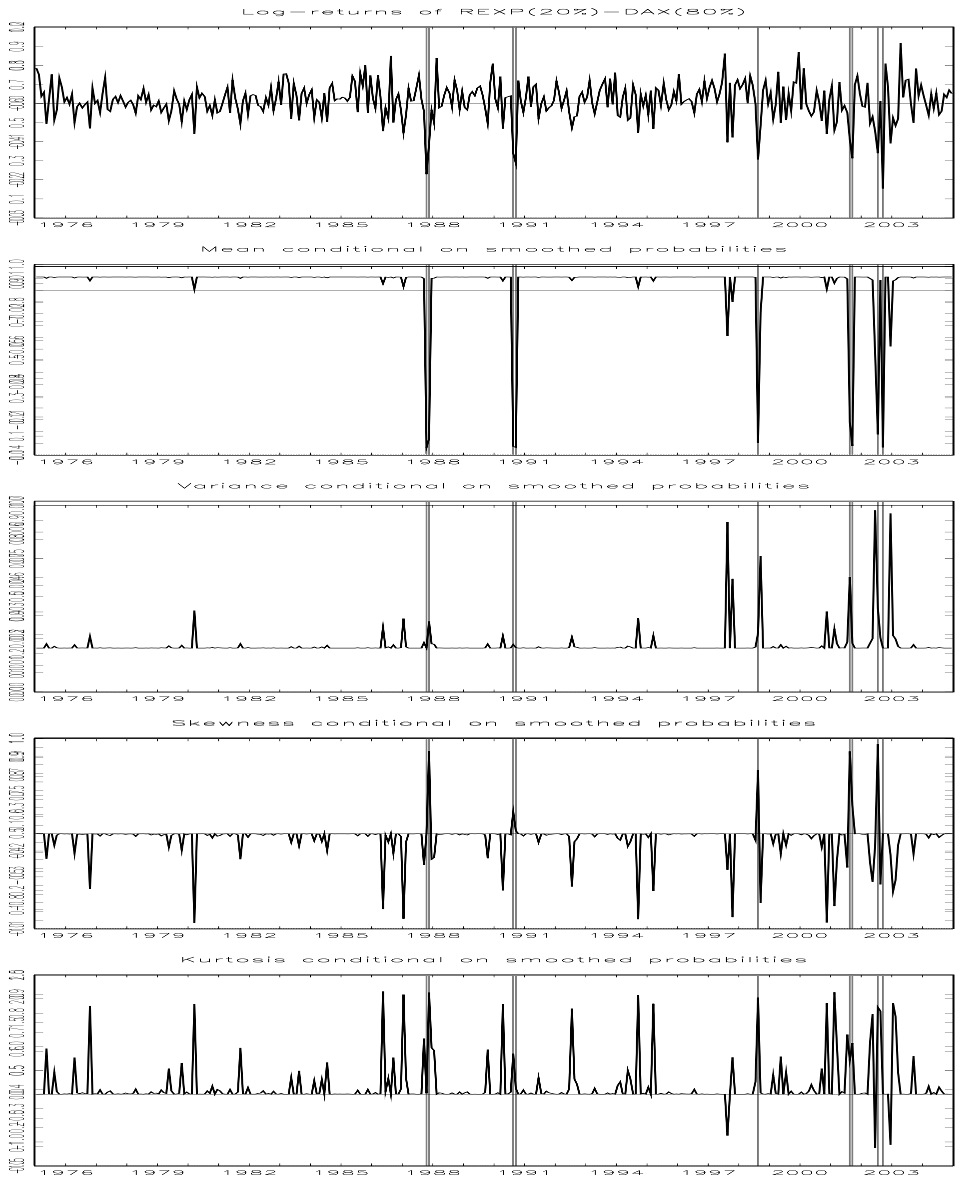
Figure D.7: Conditional moments: REXP(20%)-DAX30(80%) - MS(1-2)-AR(0)



Note:

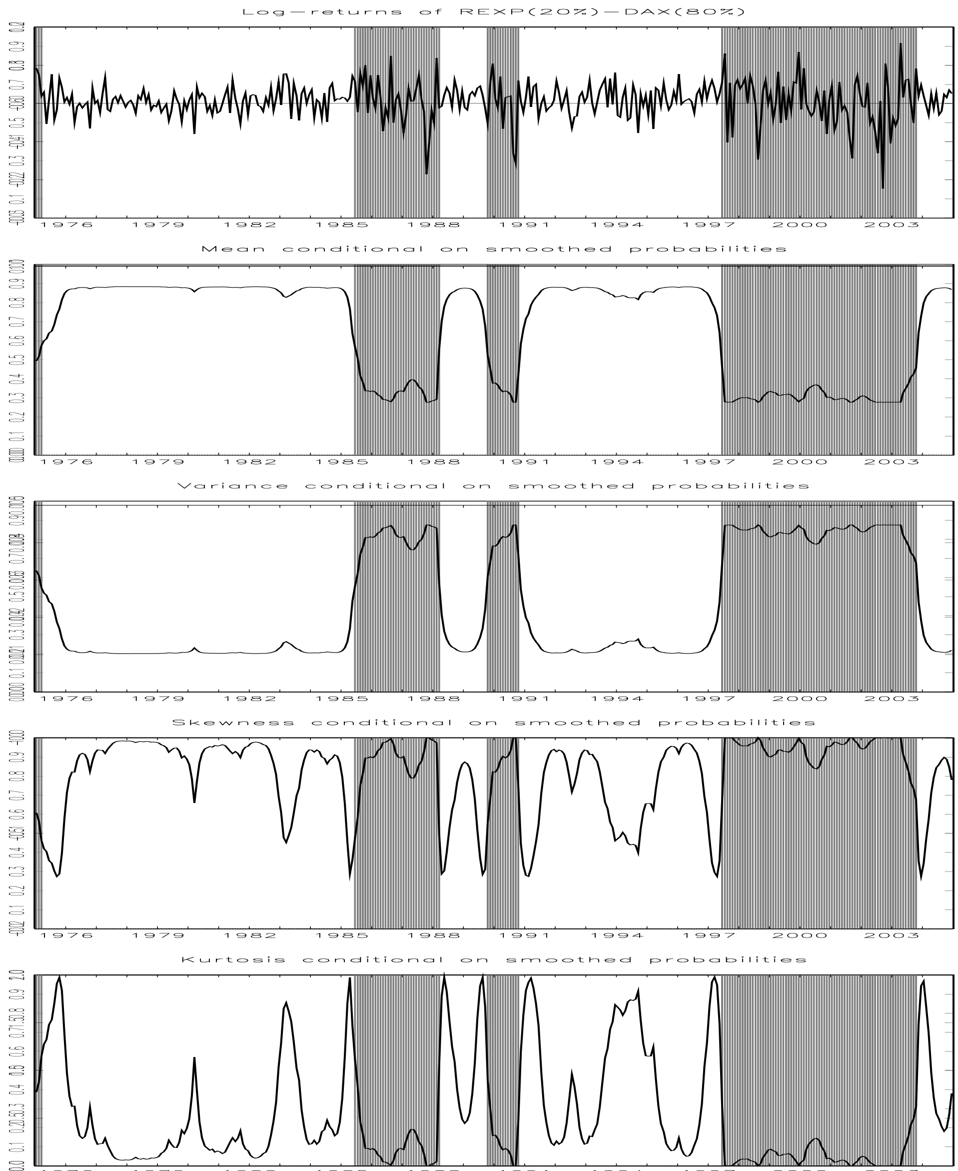
The top panel shows the log-returns of the REXP(20%)-DAX30(80%) portfolio. The remaining panels indicate the moments for the MS(1-2) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high volatility regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.8: Conditional moments: REXP(20%)-DAX30(80%) - MS(2-1)-AR(0)



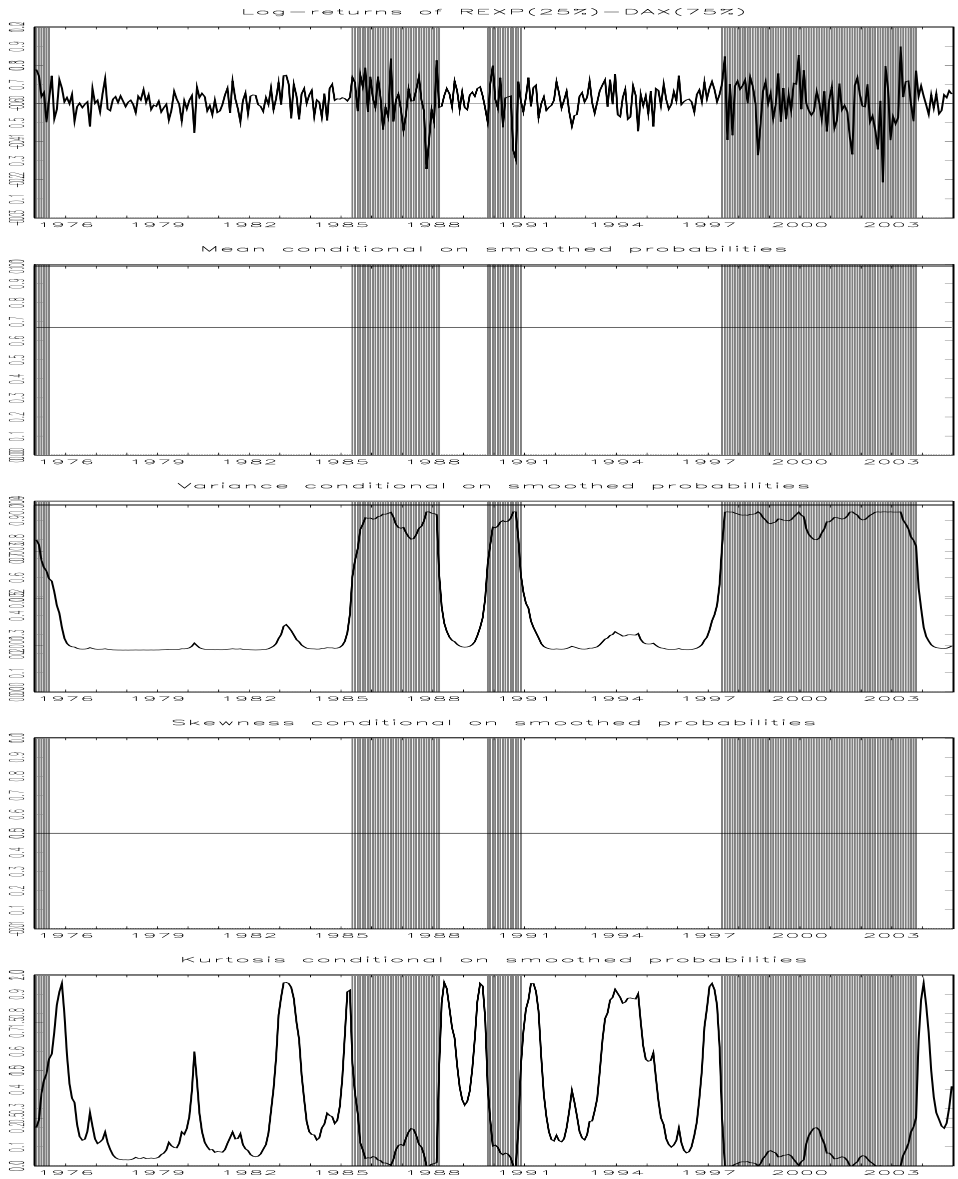
Note:
The top panel shows the log-returns of the REXP(20%)-DAX30(80%) portfolio. The remaining panels indicate the moments for the MS(2-1) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high mean regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.9: Conditional moments: REXP(20%)-DAX30(80%) - MS(2-2)-AR(0)



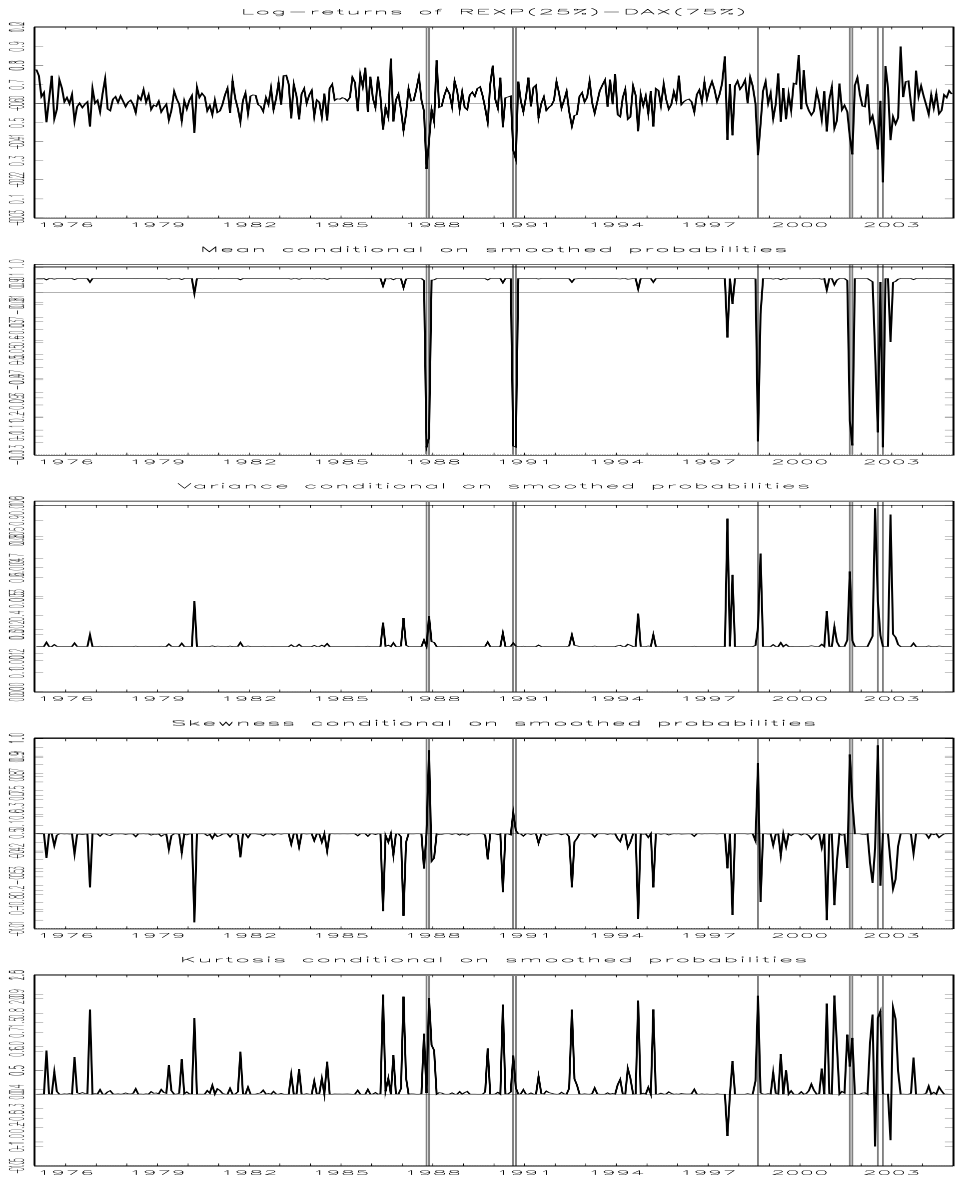
Note:
The top panel shows the log-returns of the REXP(20%)-DAX30(80%) portfolio. The remaining panels indicate the moments for the MS(2-2) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high volatility regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.10: Conditional moments: REXP(25%)-DAX30(75%) - MS(1-2)-AR(0)



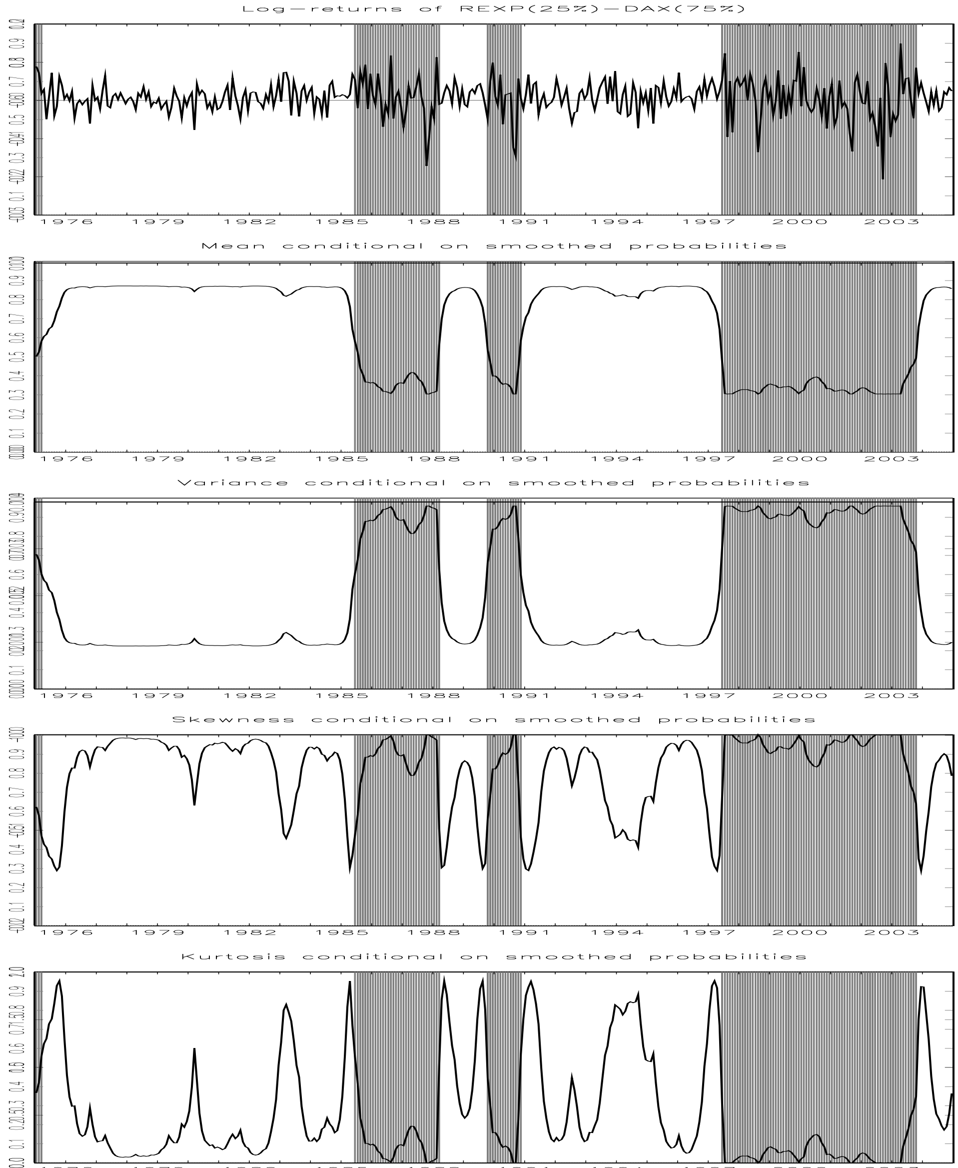
Note:
The top panel shows the log-returns of the REXP(25%)-DAX30(75%) portfolio. The remaining panels indicate the moments for the MS(1-2) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high volatility regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.11: Conditional moments: REXP(25%)-DAX30(75%) - MS(2-1)-AR(0)



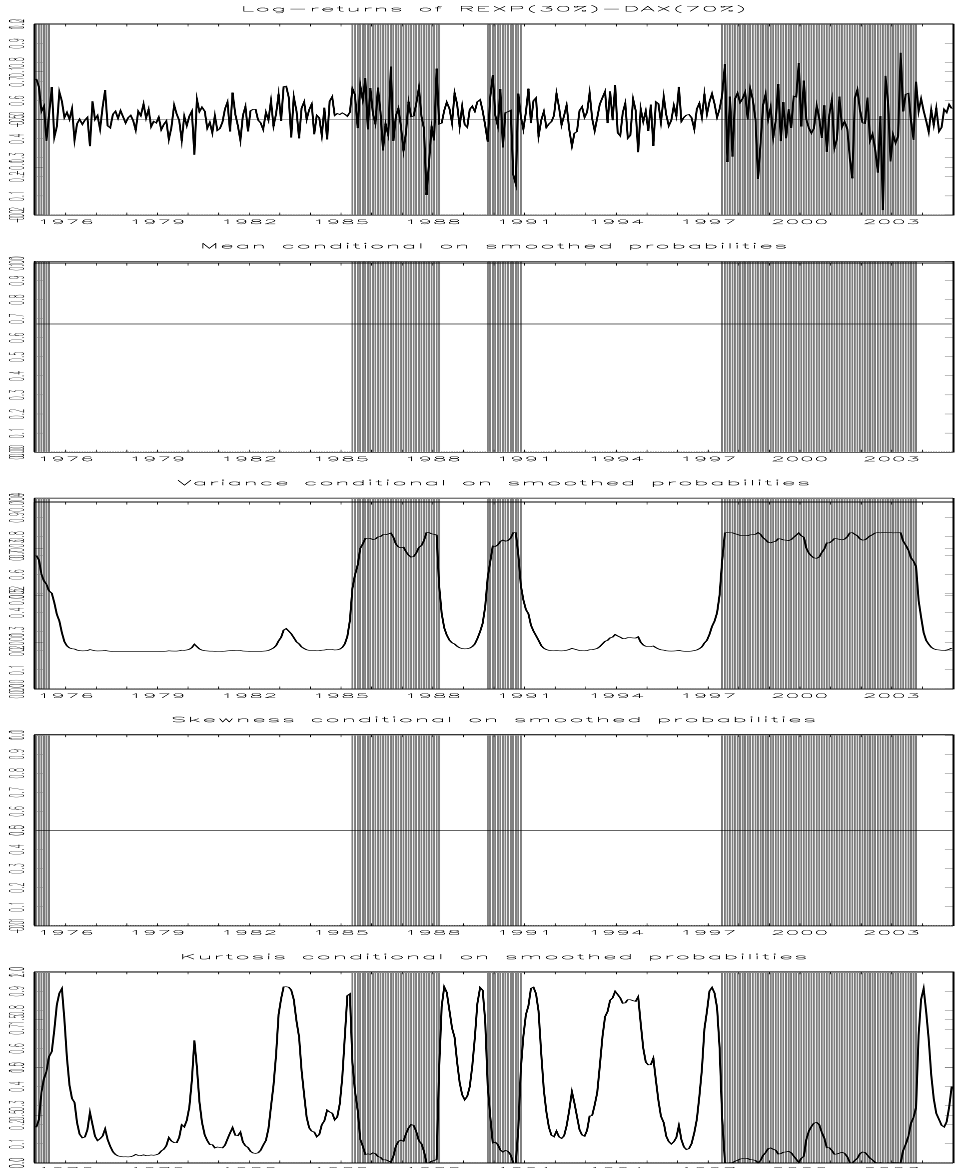
Note:
The top panel shows the log-returns of the REXP(25%)-DAX30(75%) portfolio. The remaining panels indicate the moments for the MS(2-1) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high mean regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.12: Conditional moments: REXP(25%)-DAX30(75%) - MS(2-2)-AR(0)



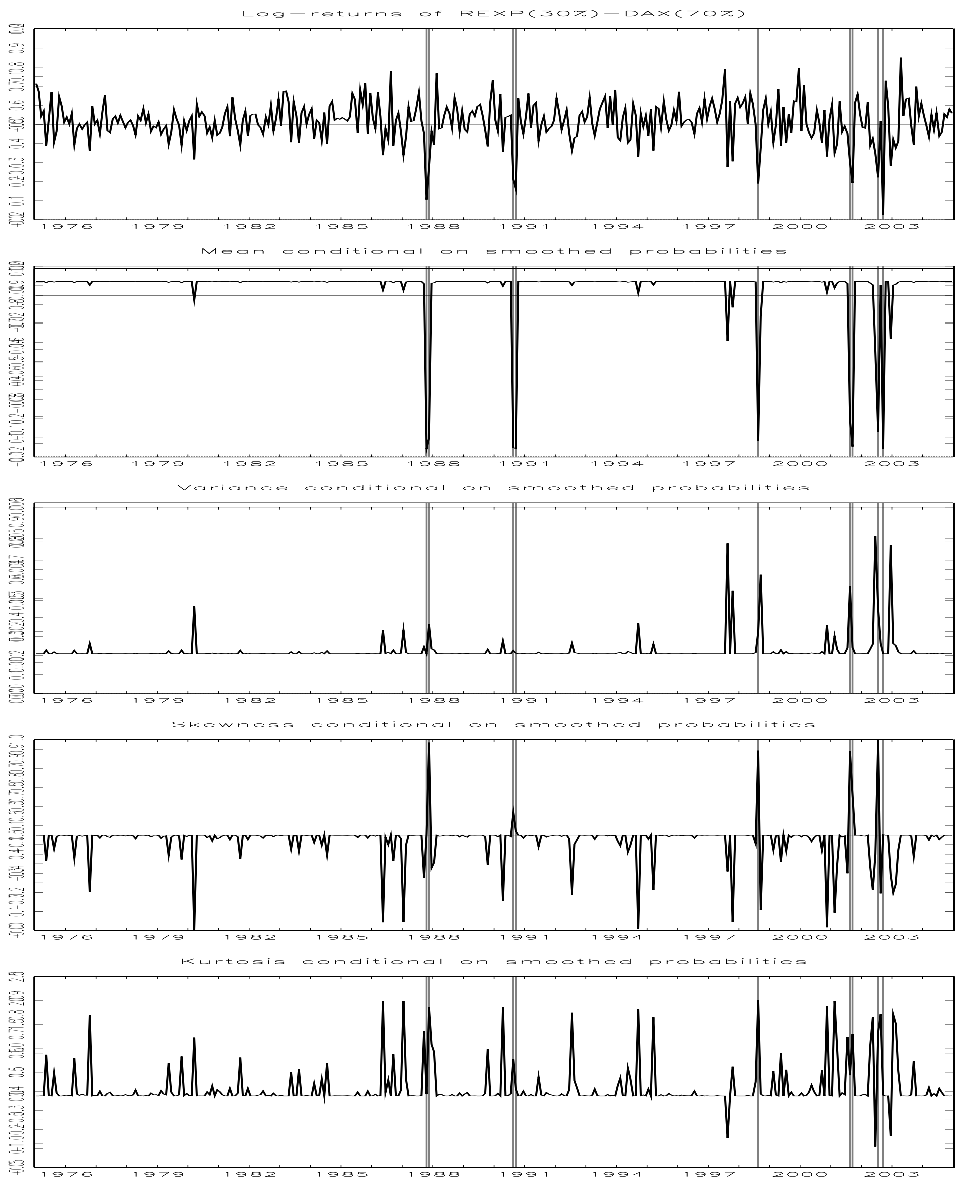
Note:
The top panel shows the log-returns of the REXP(25%)-DAX30(75%) portfolio. The remaining panels indicate the moments for the MS(2-2) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high volatility regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.13: Conditional moments: REXP(30%)-DAX30(70%) - MS(1-2)-AR(0)



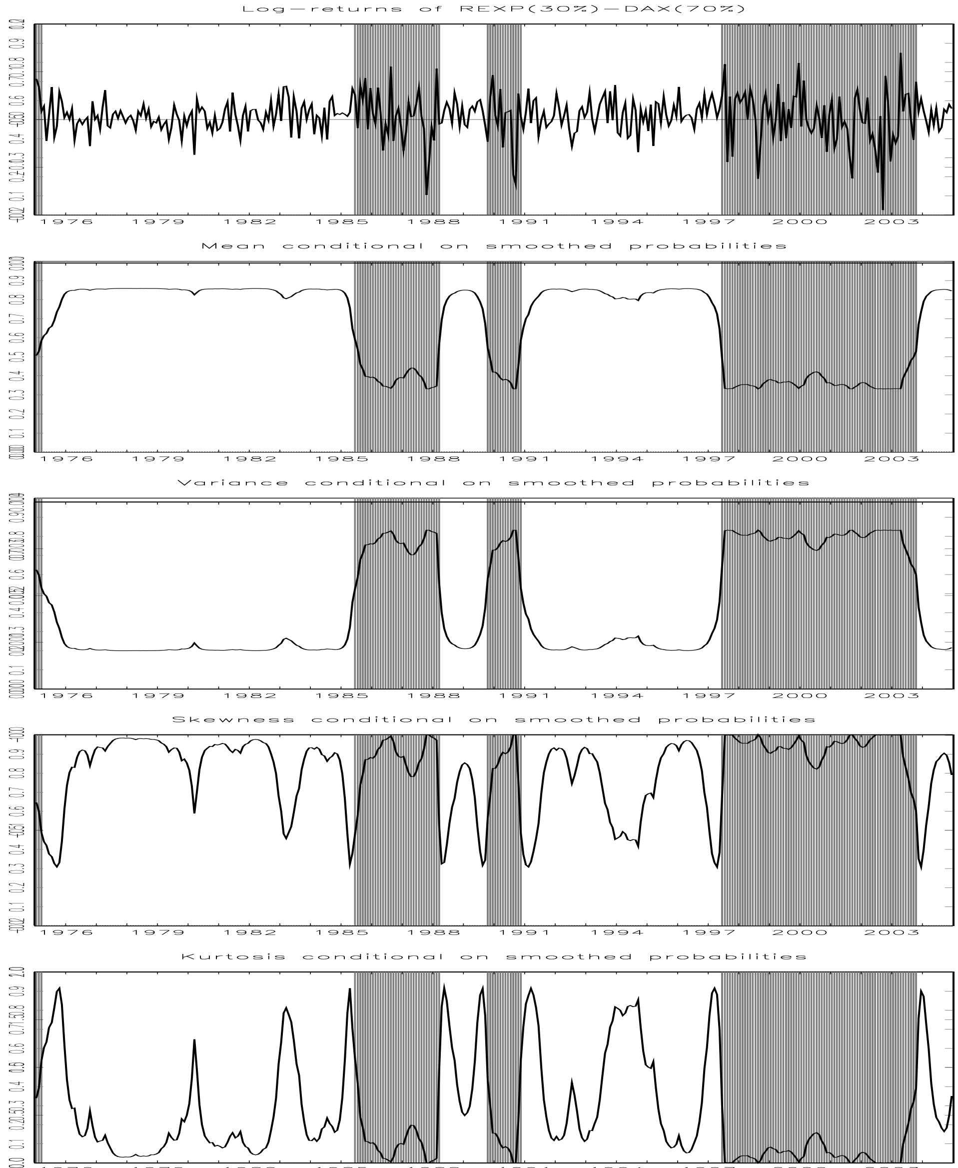
Note:
The top panel shows the log-returns of the REXP(30%)-DAX30(70%) portfolio. The remaining panels indicate the moments for the MS(1-2) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high volatility regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.14: Conditional moments: REXP(30%)-DAX30(70%) - MS(2-1)-AR(0)



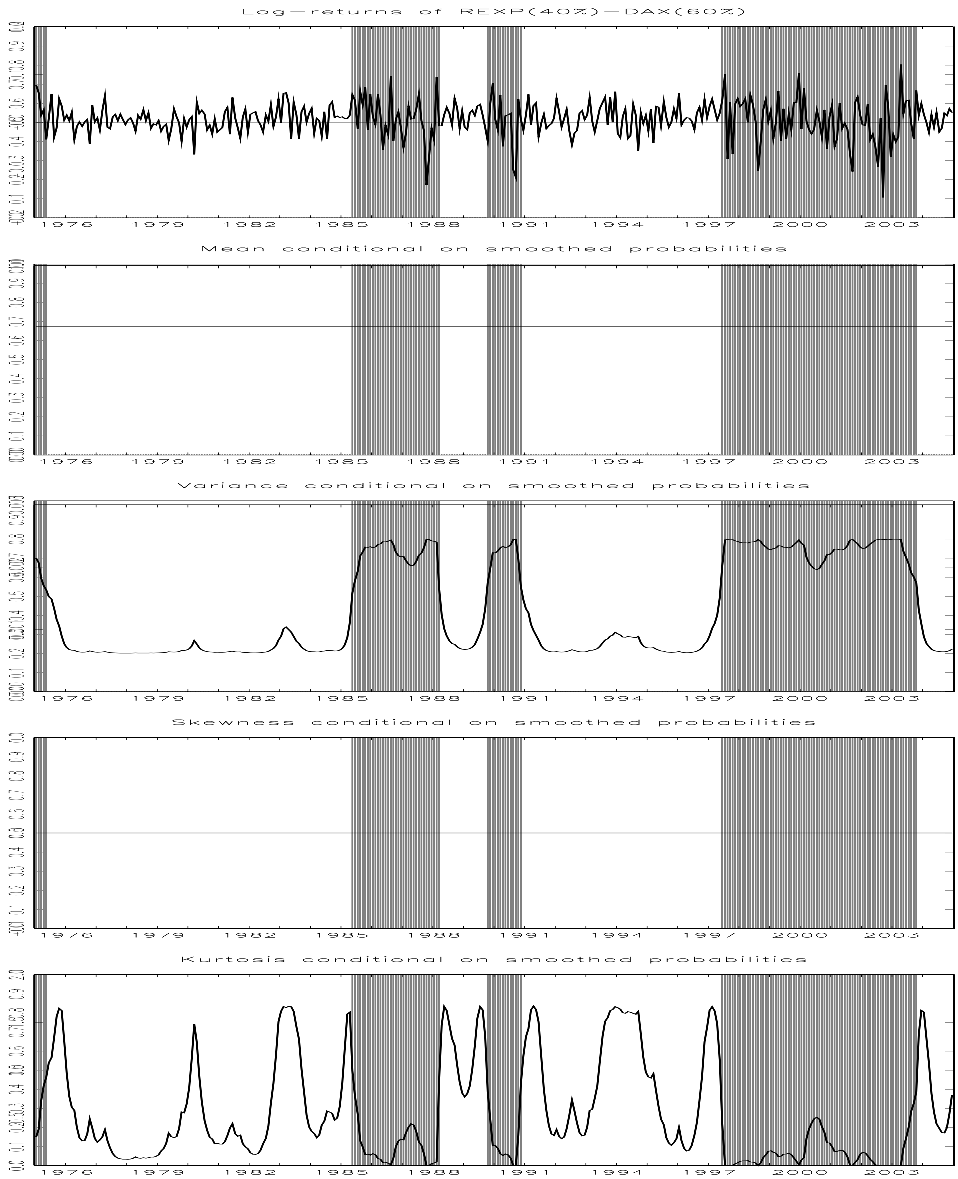
Note:
The top panel shows the log-returns of the REXP(30%)-DAX30(70%) portfolio. The remaining panels indicate the moments for the MS(2-1) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high mean regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.15: Conditional moments: REXP(30%)-DAX30(70%) - MS(2-2)-AR(0)



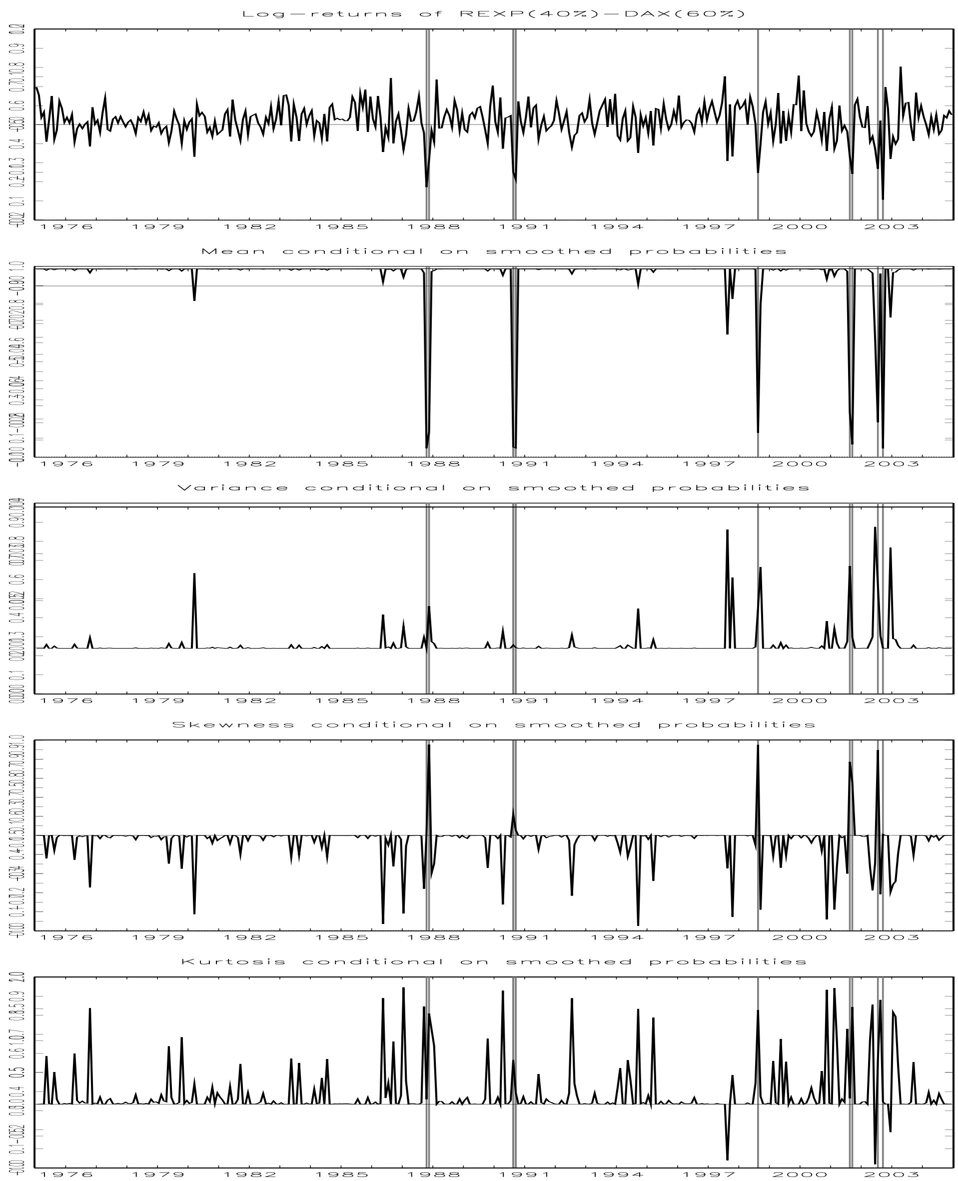
Note:
The top panel shows the log-returns of the REXP(30%)-DAX30(70%) portfolio. The remaining panels indicate the moments for the MS(2-2) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high volatility regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.16: Conditional moments: REXP(40%)-DAX30(60%) - MS(1-2)-AR(0)



Note:
The top panel shows the log-returns of the REXP(40%)-DAX30(60%) portfolio. The remaining panels indicate the moments for the MS(1-2) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high volatility regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

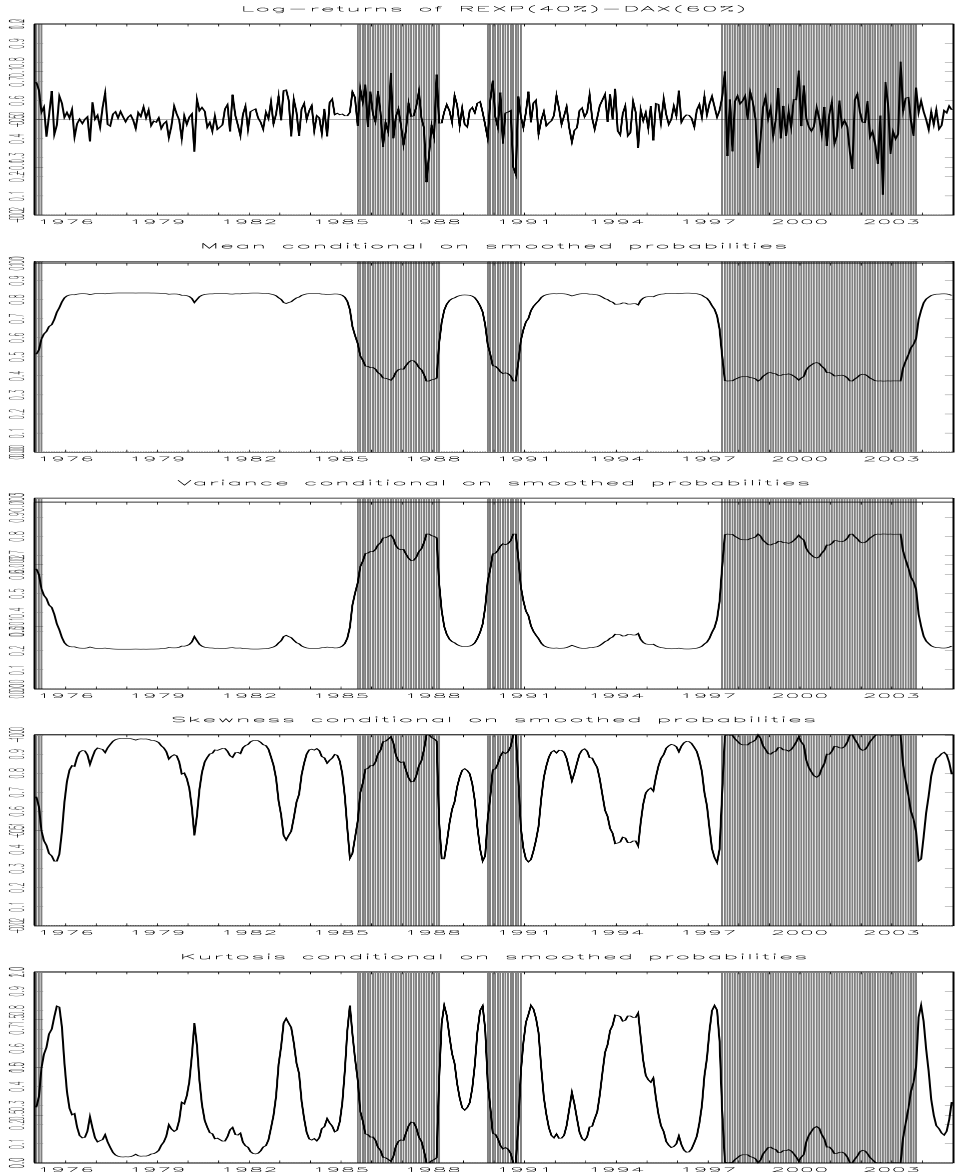
Figure D.17: Conditional moments: REXP(40%)-DAX30(60%) - MS(2-1)-AR(0)



Note:

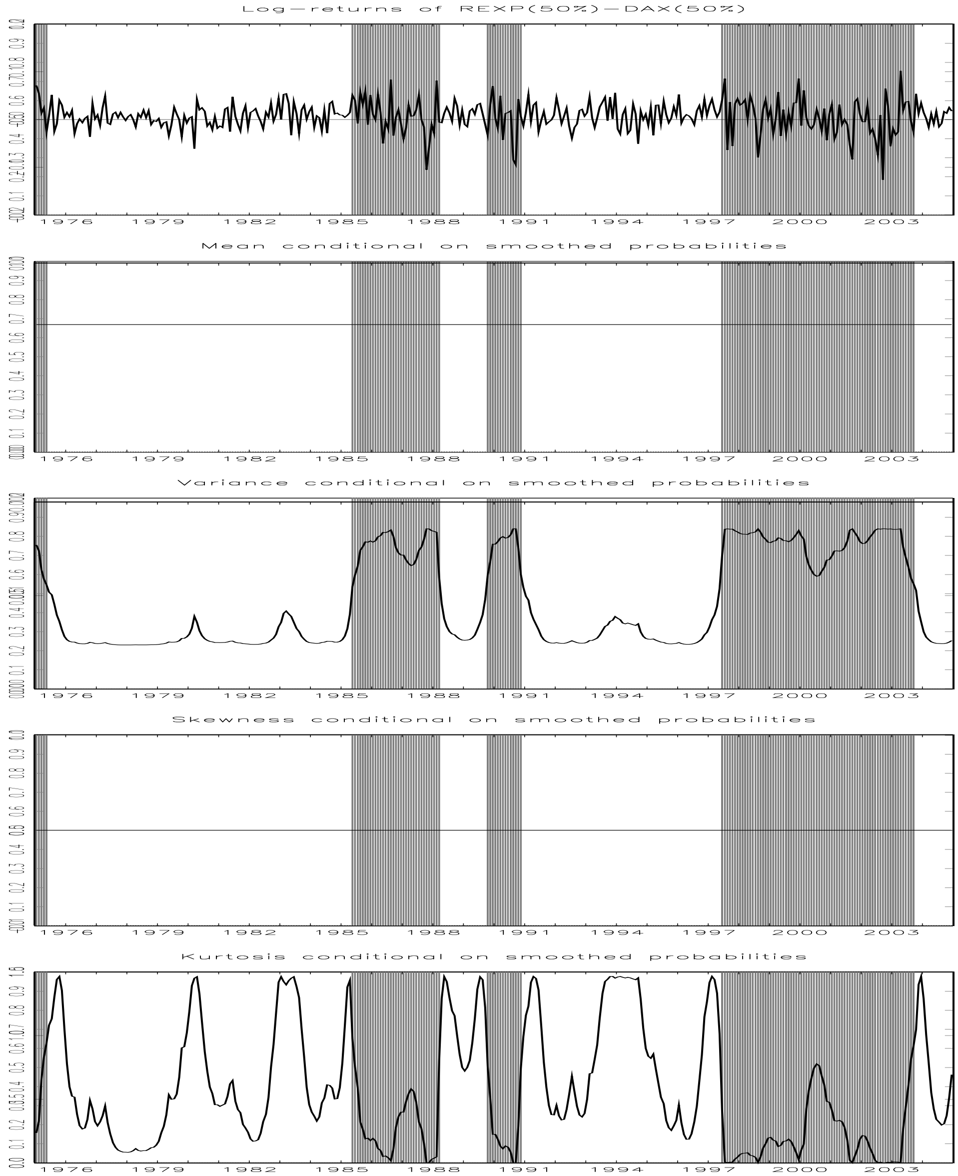
The top panel shows the log-returns of the REXP(40%)-DAX30(60%) portfolio. The remaining panels indicate the moments for the MS(2-1) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high mean regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.18: Conditional moments: REXP(40%)-DAX30(60%) - MS(2-2)-AR(0)



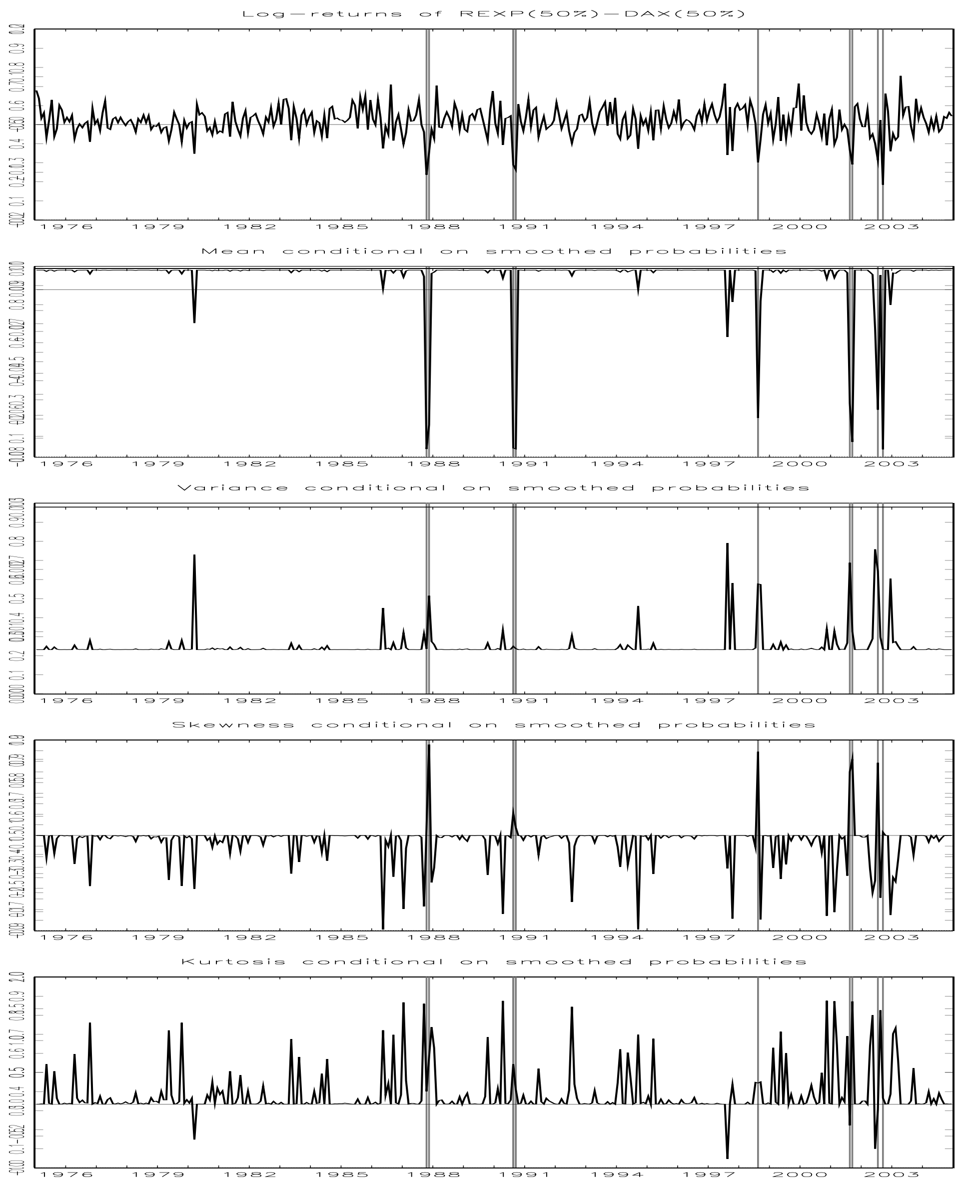
Note:
The top panel shows the log-returns of the REXP(40%)-DAX30(60%) portfolio. The remaining panels indicate the moments for the MS(2-2) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high volatility regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.19: Conditional moments: REXP(50%)-DAX30(50%) - MS(1-2)-AR(0)



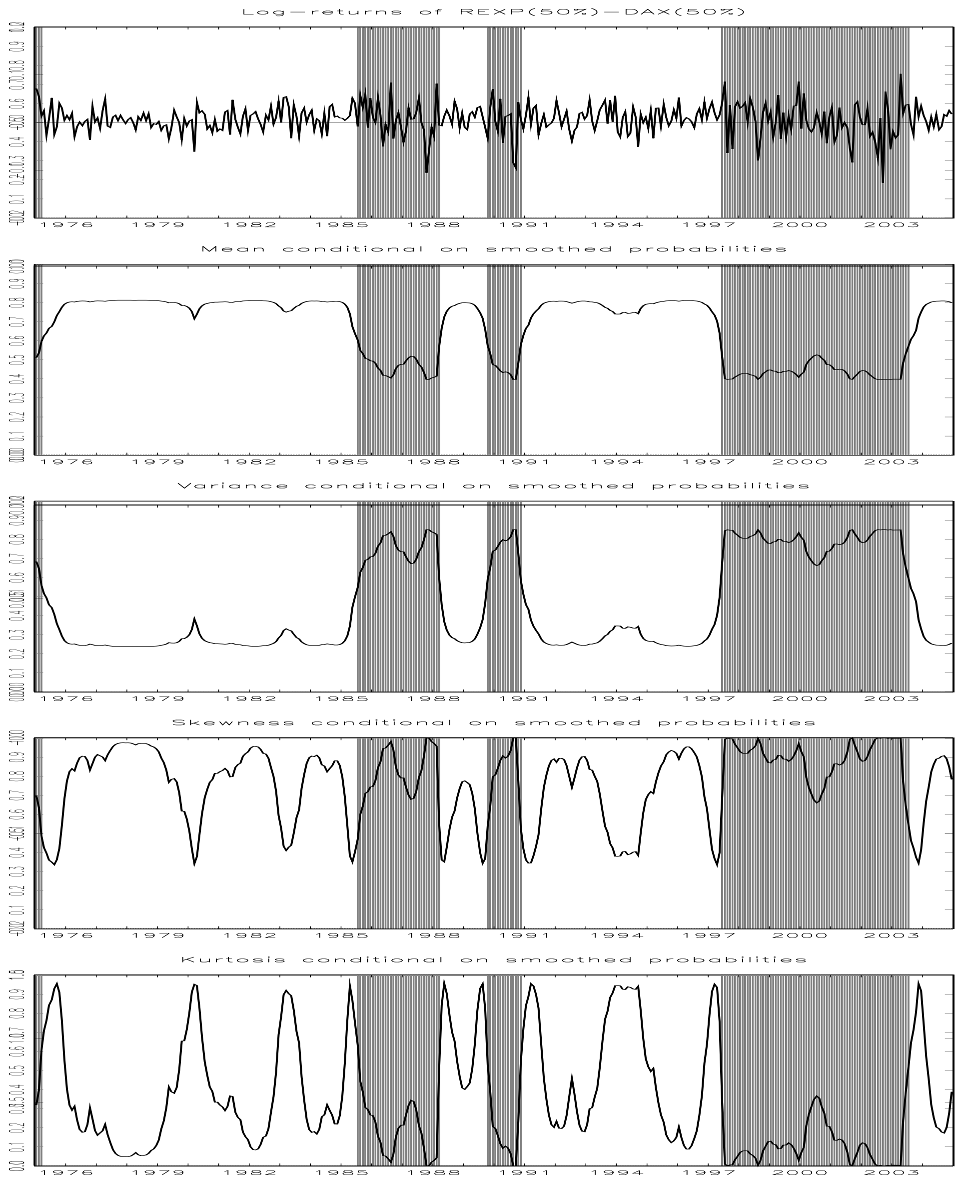
Note:
The top panel shows the log-returns of the REXP(50%)-DAX30(50%) portfolio. The remaining panels indicate the moments for the MS(1-2) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high volatility regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.20: Conditional moments: REXP(50%)-DAX30(50%) - MS(2-1)-AR(0)



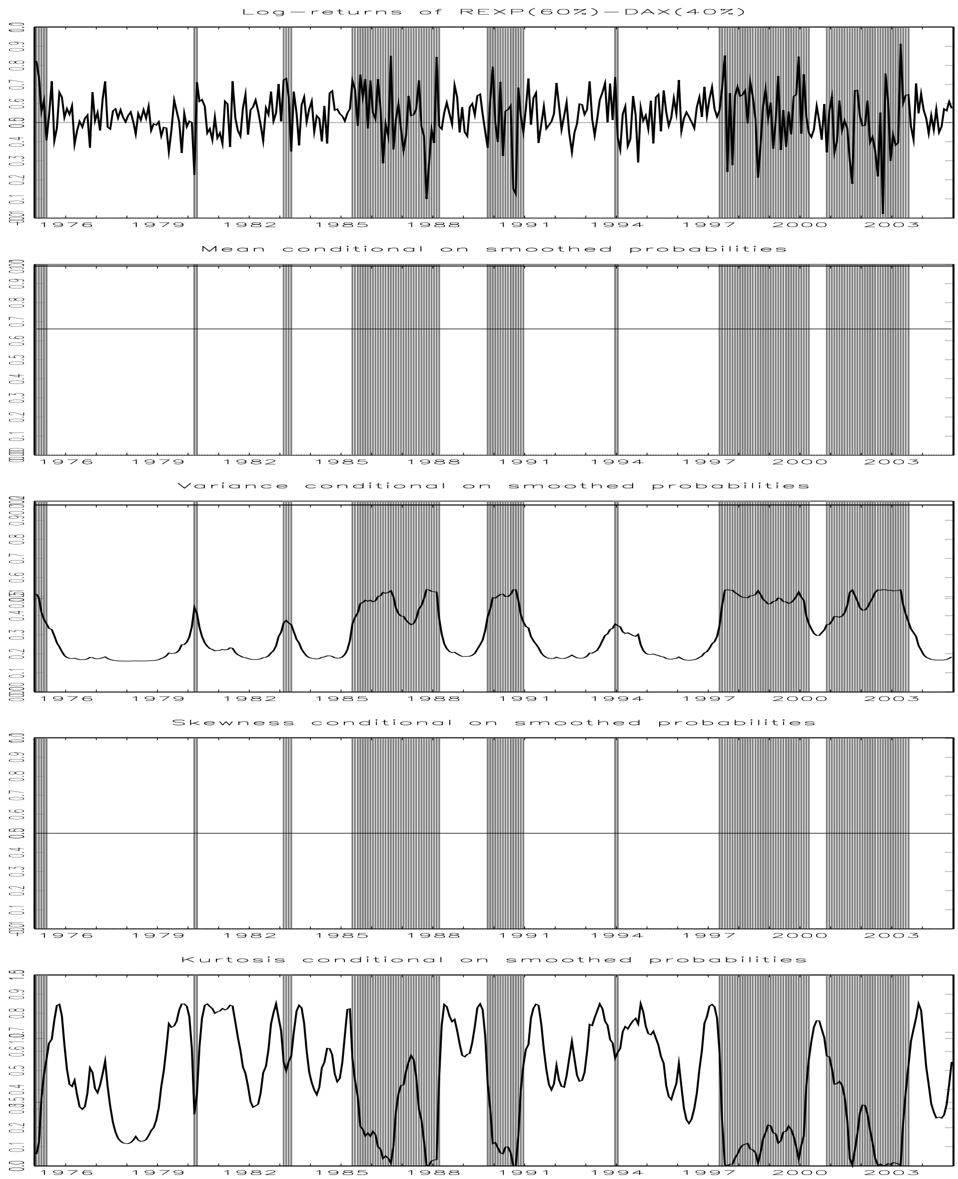
Note:
The top panel shows the log-returns of the REXP(50%)-DAX30(50%) portfolio. The remaining panels indicate the moments for the MS(2-1) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high mean regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.21: Conditional moments: REXP(50%)-DAX30(50%) - MS(2-2)-AR(0)



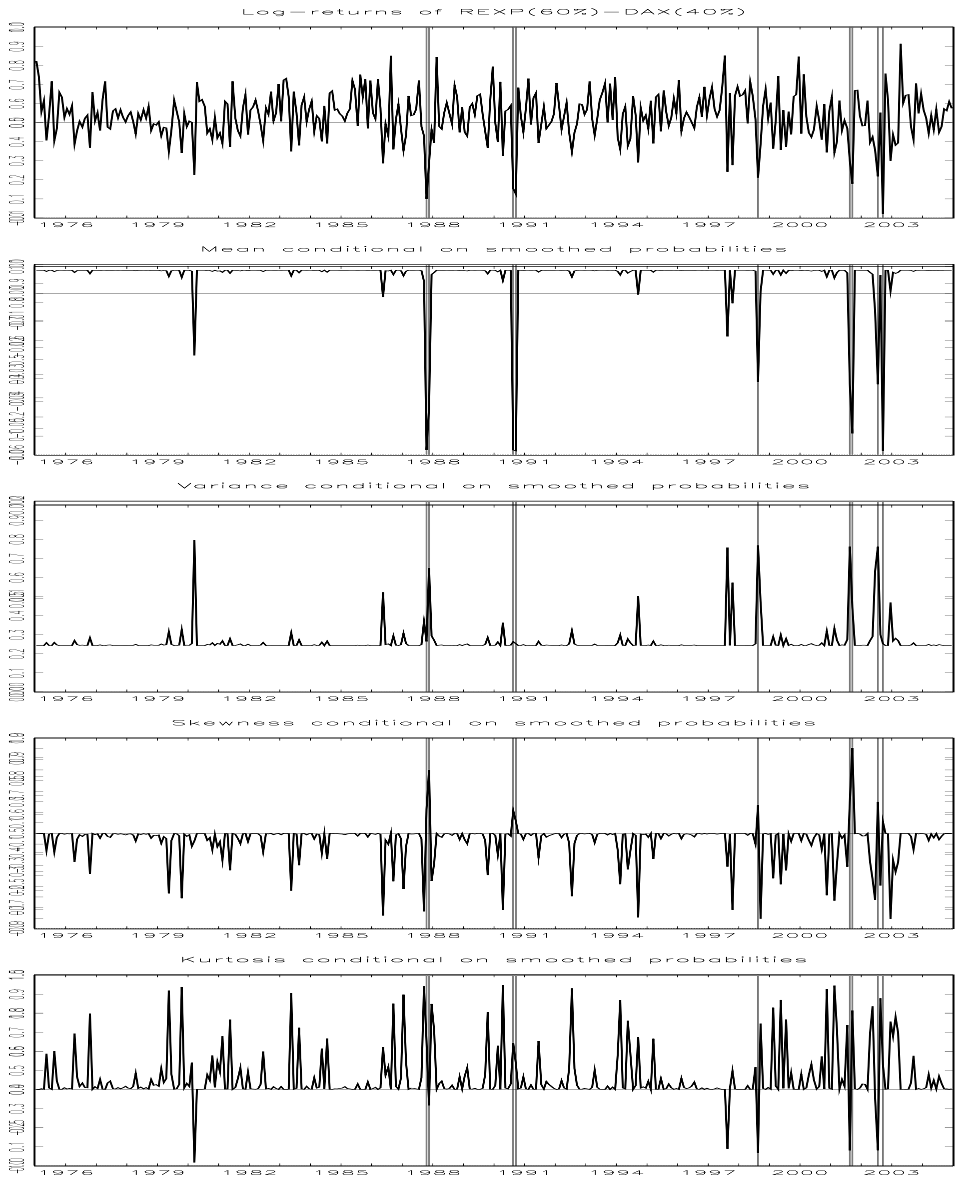
Note:
The top panel shows the log-returns of the REXP(50%)-DAX30(50%) portfolio. The remaining panels indicate the moments for the MS(2-2) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high volatility regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.22: Conditional moments: REXP(60%)-DAX30(40%) - MS(1-2)-AR(0)



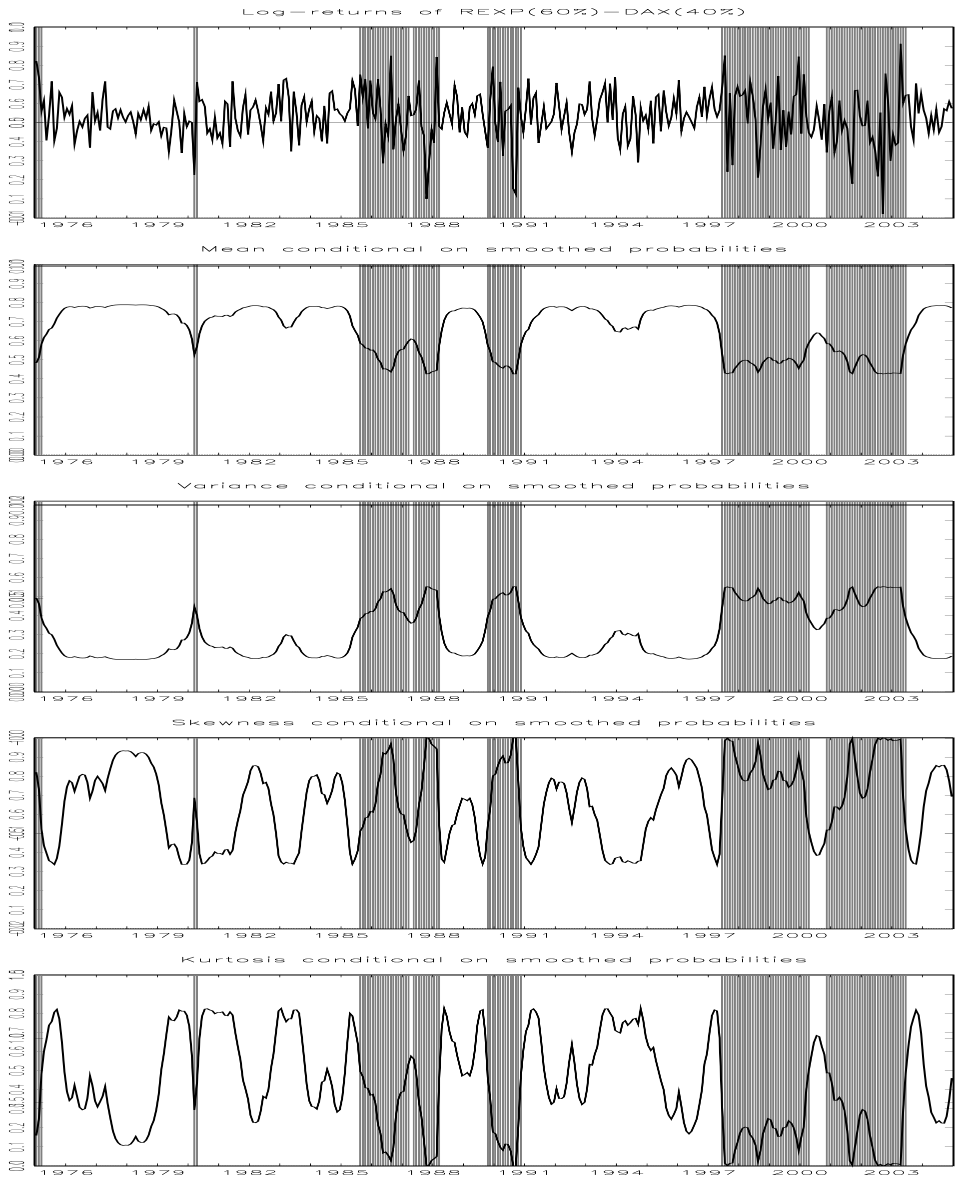
Note:
The top panel shows the log-returns of the REXP(60%)-DAX30(40%) portfolio. The remaining panels indicate the moments for the MS(1-2) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high volatility regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.23: Conditional moments: REXP(60%)-DAX30(40%) - MS(2-1)-AR(0)



Note:
The top panel shows the log-returns of the REXP(60%)-DAX30(40%) portfolio. The remaining panels indicate the moments for the MS(2-1) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high mean regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

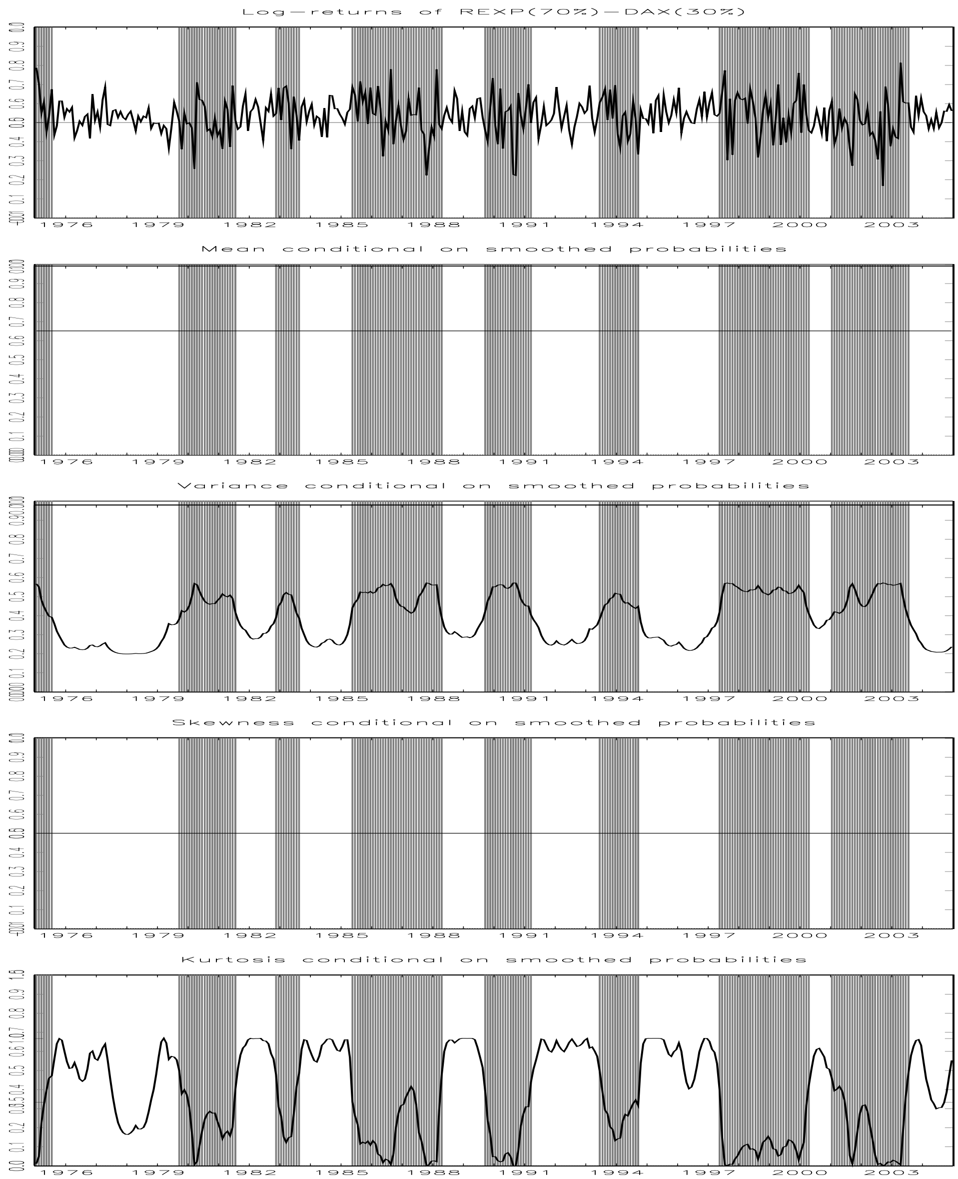
Figure D.24: Conditional moments: REXP(60%)-DAX30(40%) - MS(2-2)-AR(0)



Note:

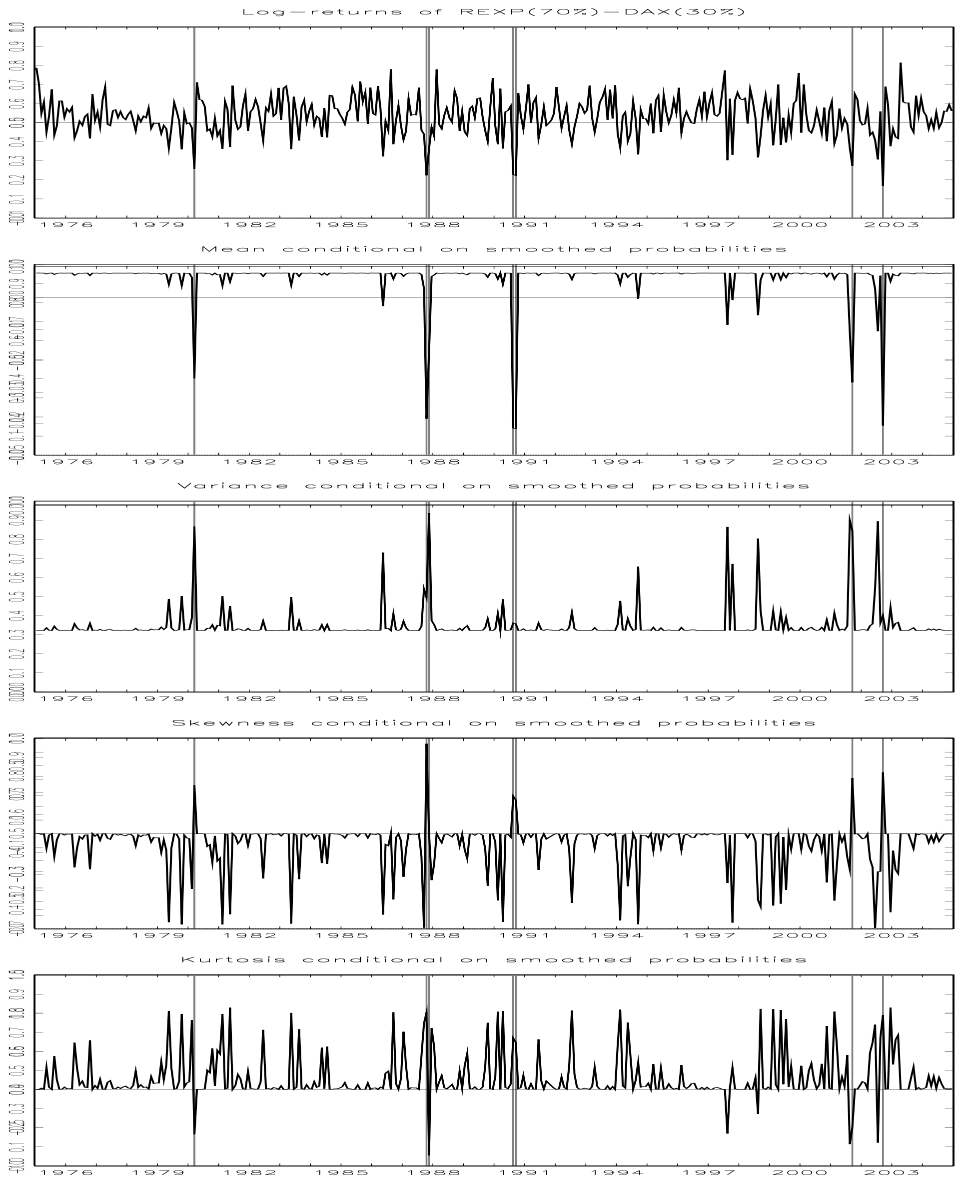
The top panel shows the log-returns of the REXP(60%)-DAX30(40%) portfolio. The remaining panels indicate the moments for the MS(2-2) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high volatility regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.25: Conditional moments: REXP(70%)-DAX30(30%) - MS(1-2)-AR(0)



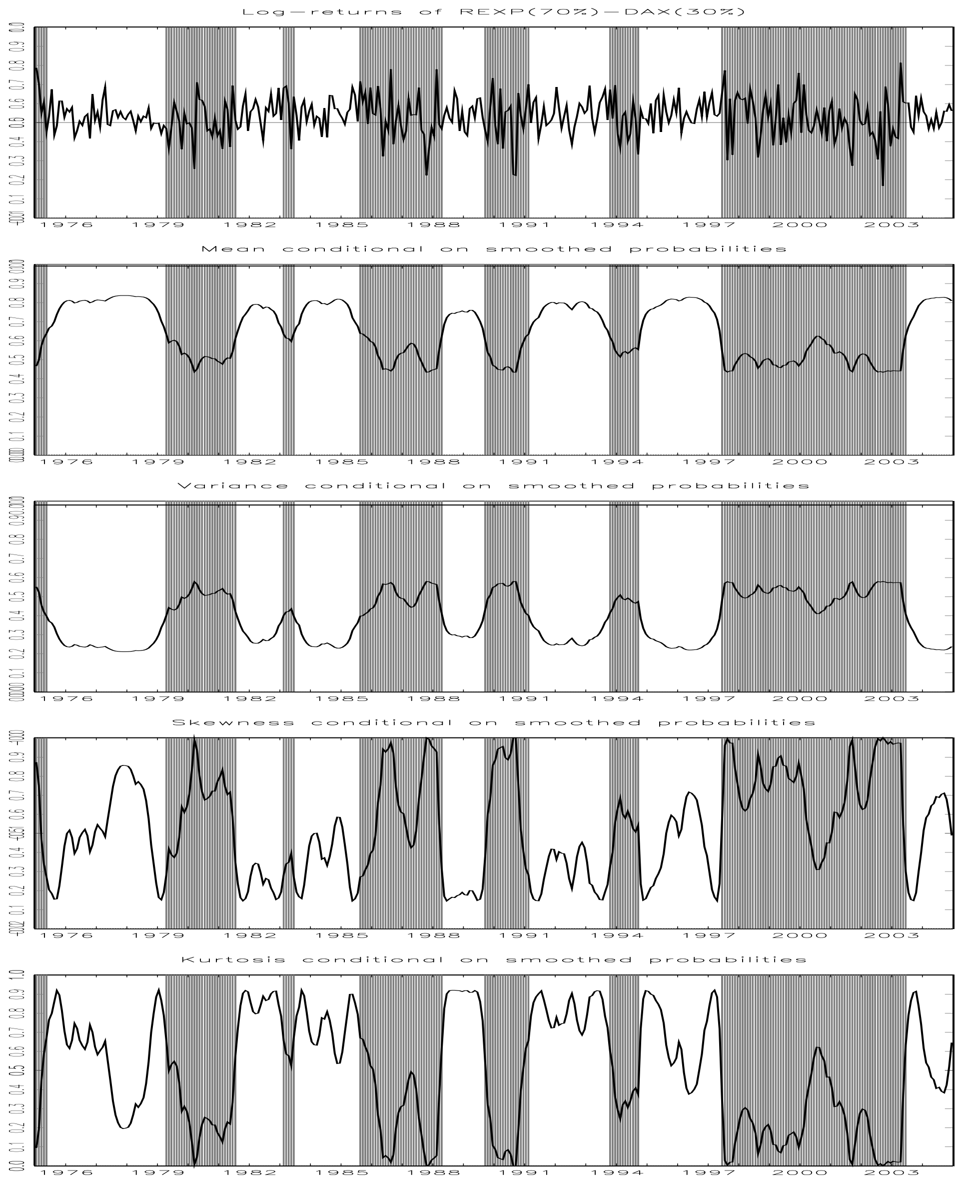
Note:
The top panel shows the log-returns of the REXP(70%)-DAX30(30%) portfolio. The remaining panels indicate the moments for the MS(1-2) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high volatility regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.26: Conditional moments: REXP(70%)-DAX30(30%) - MS(2-1)-AR(0)



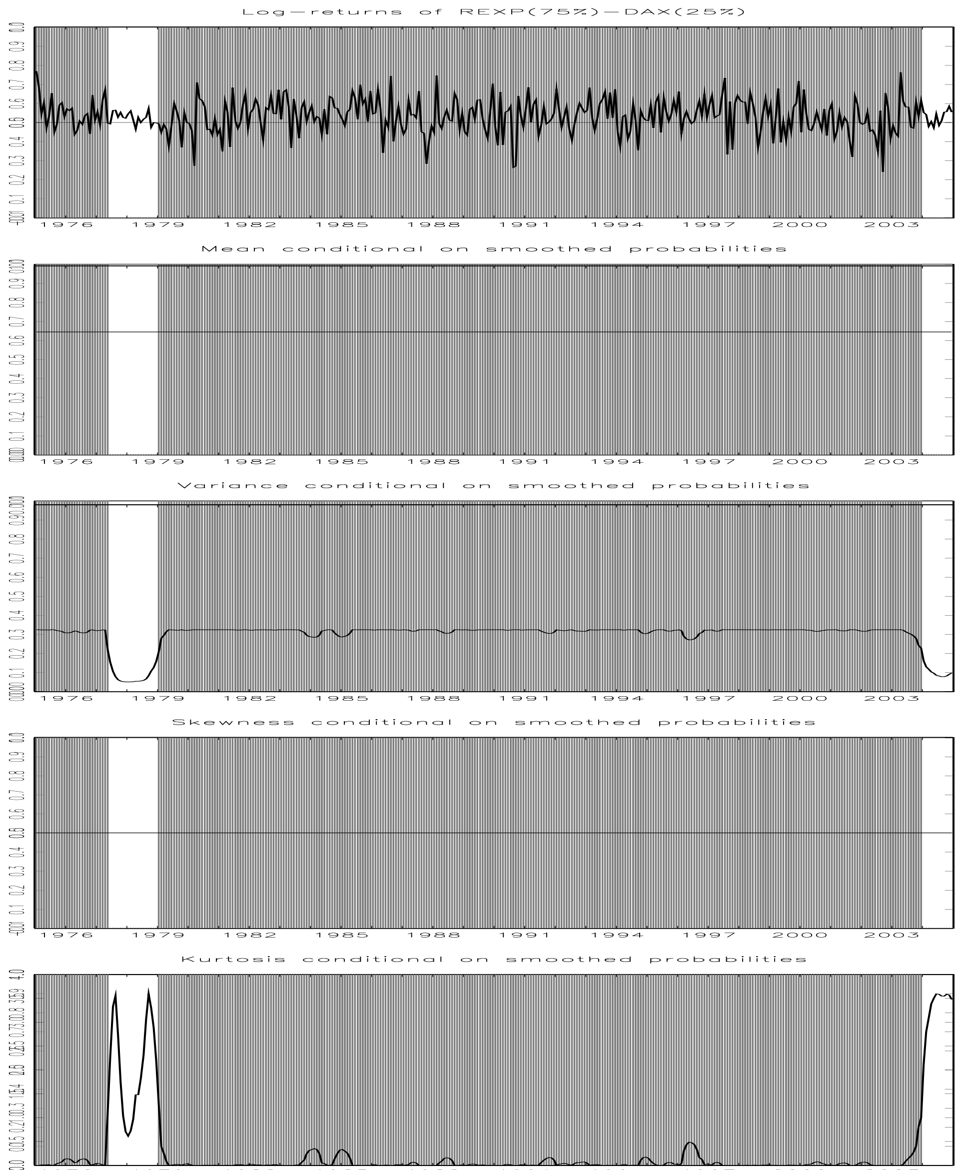
Note:
The top panel shows the log-returns of the REXP(70%)-DAX30(30%) portfolio. The remaining panels indicate the moments for the MS(2-1) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high mean regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.27: Conditional moments: REXP(70%)-DAX30(30%) - MS(2-2)-AR(0)



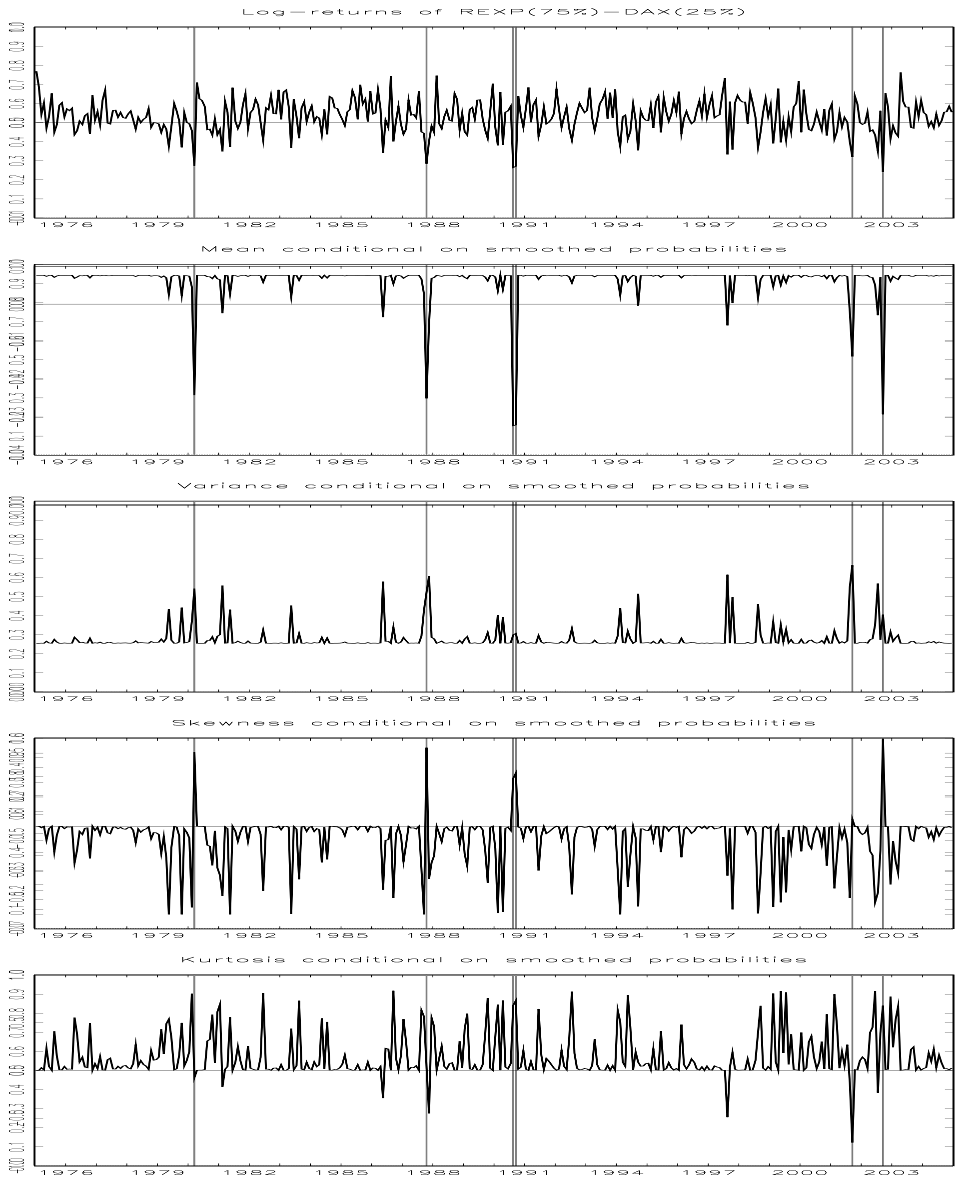
Note:
The top panel shows the log-returns of the REXP(70%)-DAX30(30%) portfolio. The remaining panels indicate the moments for the MS(2-2) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high volatility regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.28: Conditional moments: REXP(75%)-DAX30(25%) - MS(1-2)-AR(0)



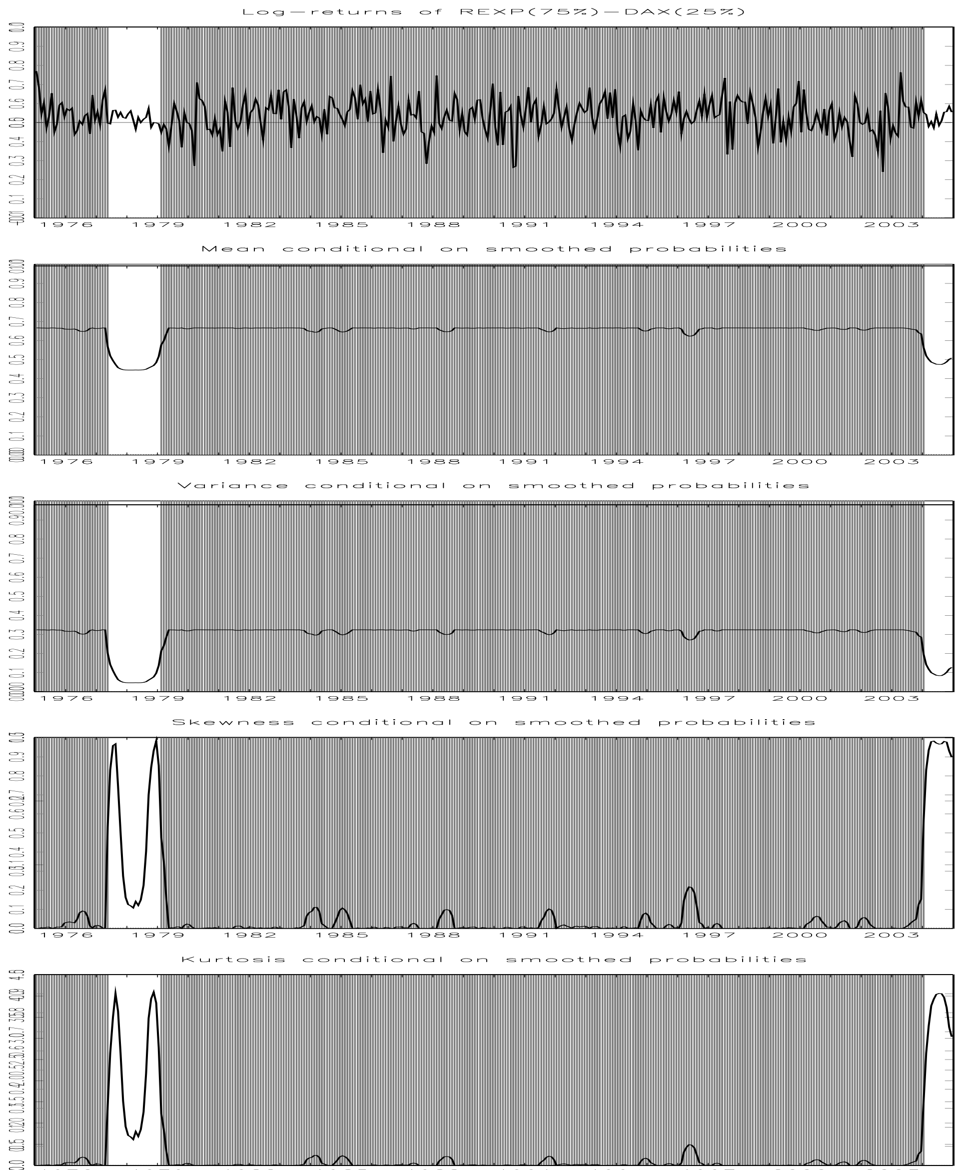
Note:
The top panel shows the log-returns of the REXP(75%)-DAX30(25%) portfolio. The remaining panels indicate the moments for the MS(1-2) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high volatility regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.29: Conditional moments: REXP(75%)-DAX30(25%) - MS(2-1)-AR(0)



Note:
The top panel shows the log-returns of the REXP(75%)-DAX30(25%) portfolio. The remaining panels indicate the moments for the MS(2-1) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high mean regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

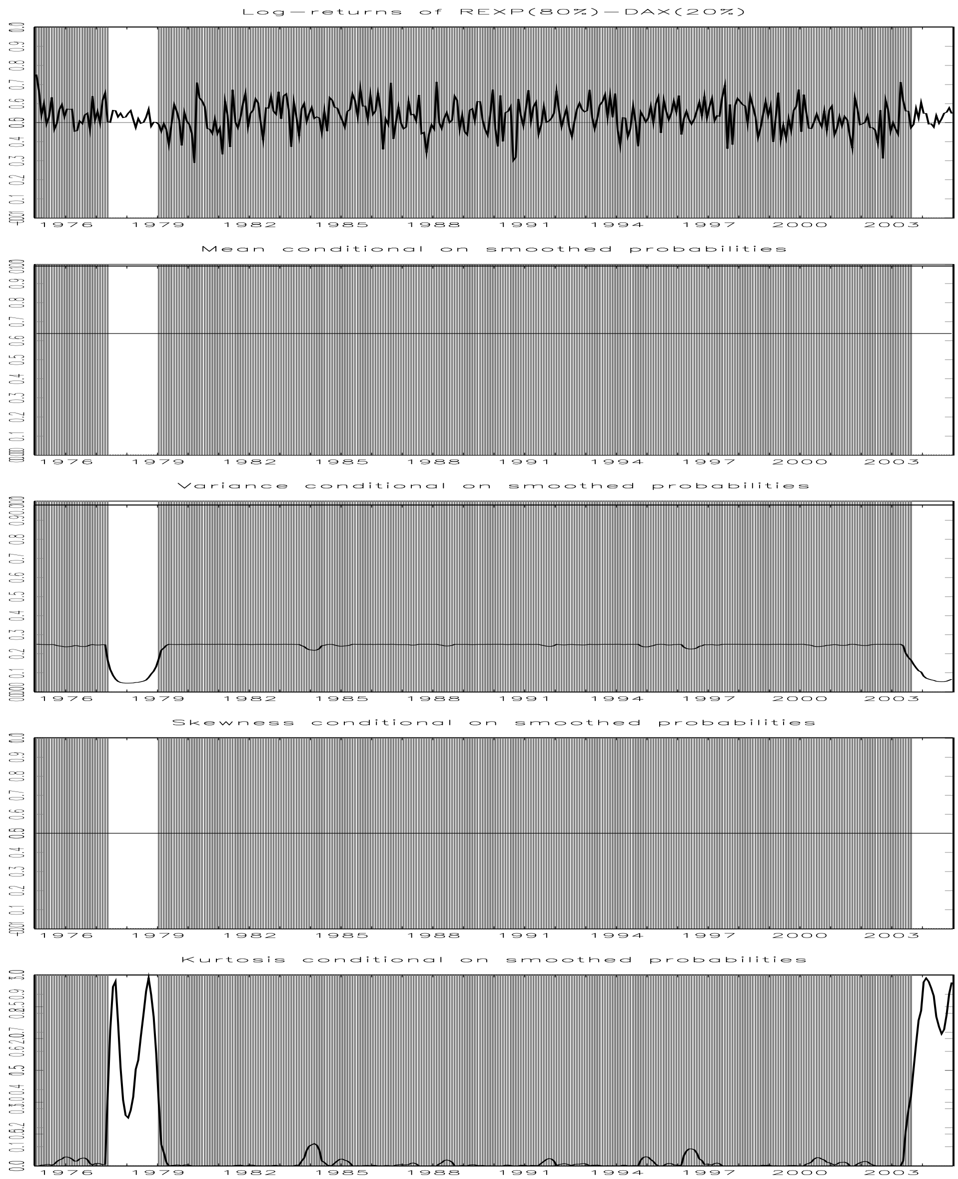
Figure D.30: Conditional moments: REXP(75%)-DAX30(25%) - MS(2-2)-AR(0)



Note:

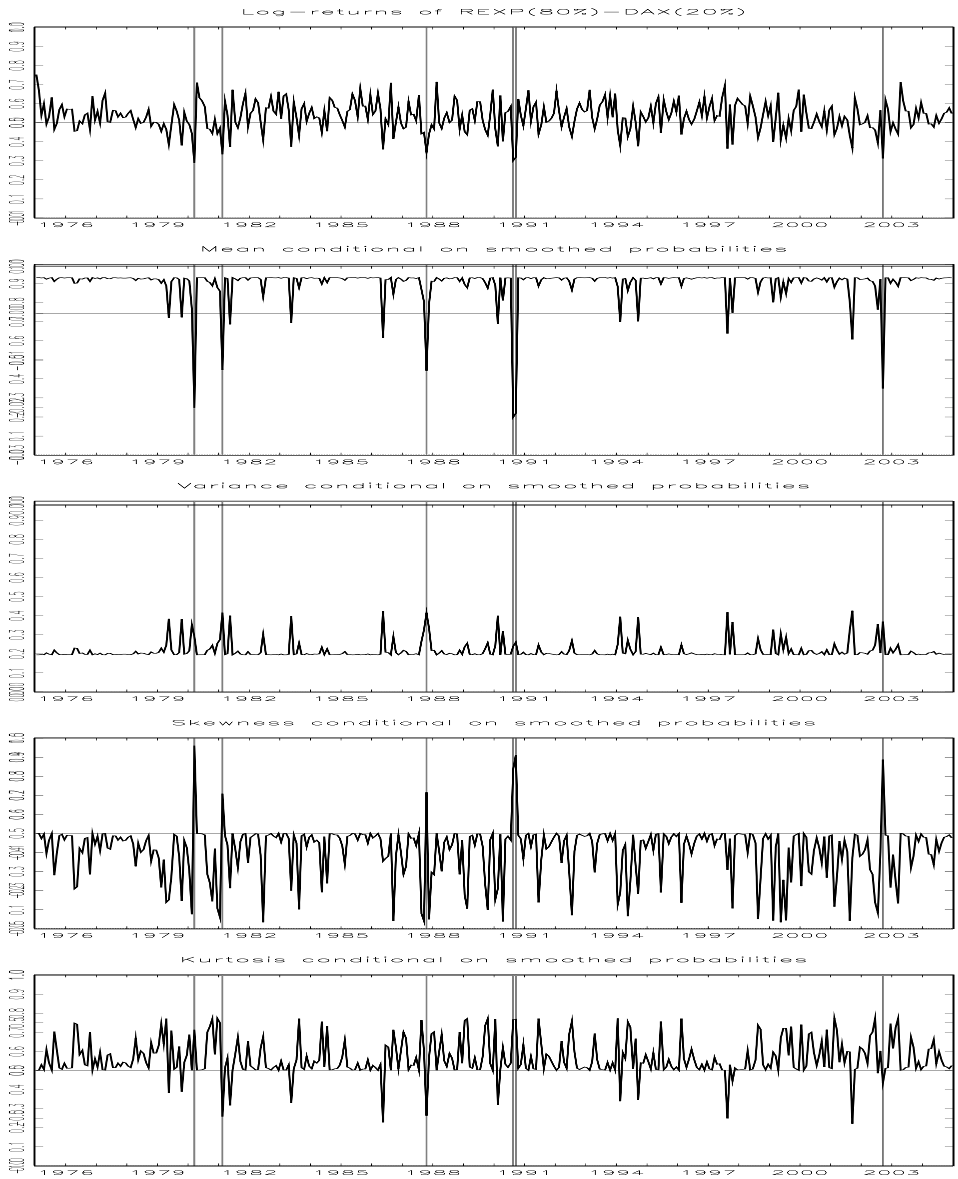
The top panel shows the log-returns of the REXP(75%)-DAX30(25%) portfolio. The remaining panels indicate the moments for the MS(2-2) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high volatility regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.31: Conditional moments: REXP(80%)-DAX30(20%) - MS(1-2)-AR(0)



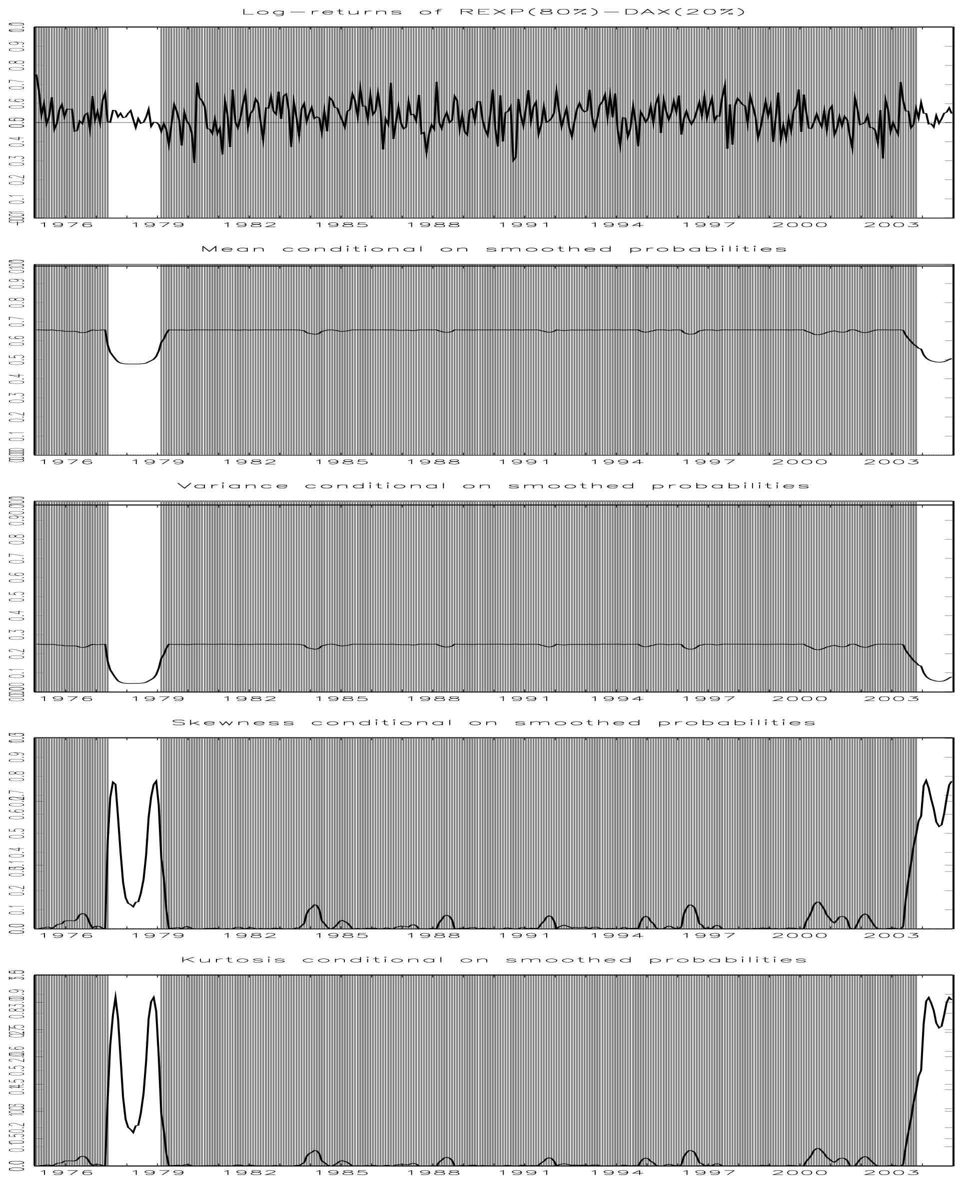
Note:
The top panel shows the log-returns of the REXP(80%)-DAX30(20%) portfolio. The remaining panels indicate the moments for the MS(1-2) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high volatility regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.32: Conditional moments: REXP(80%)-DAX30(20%) - MS(2-1)-AR(0)



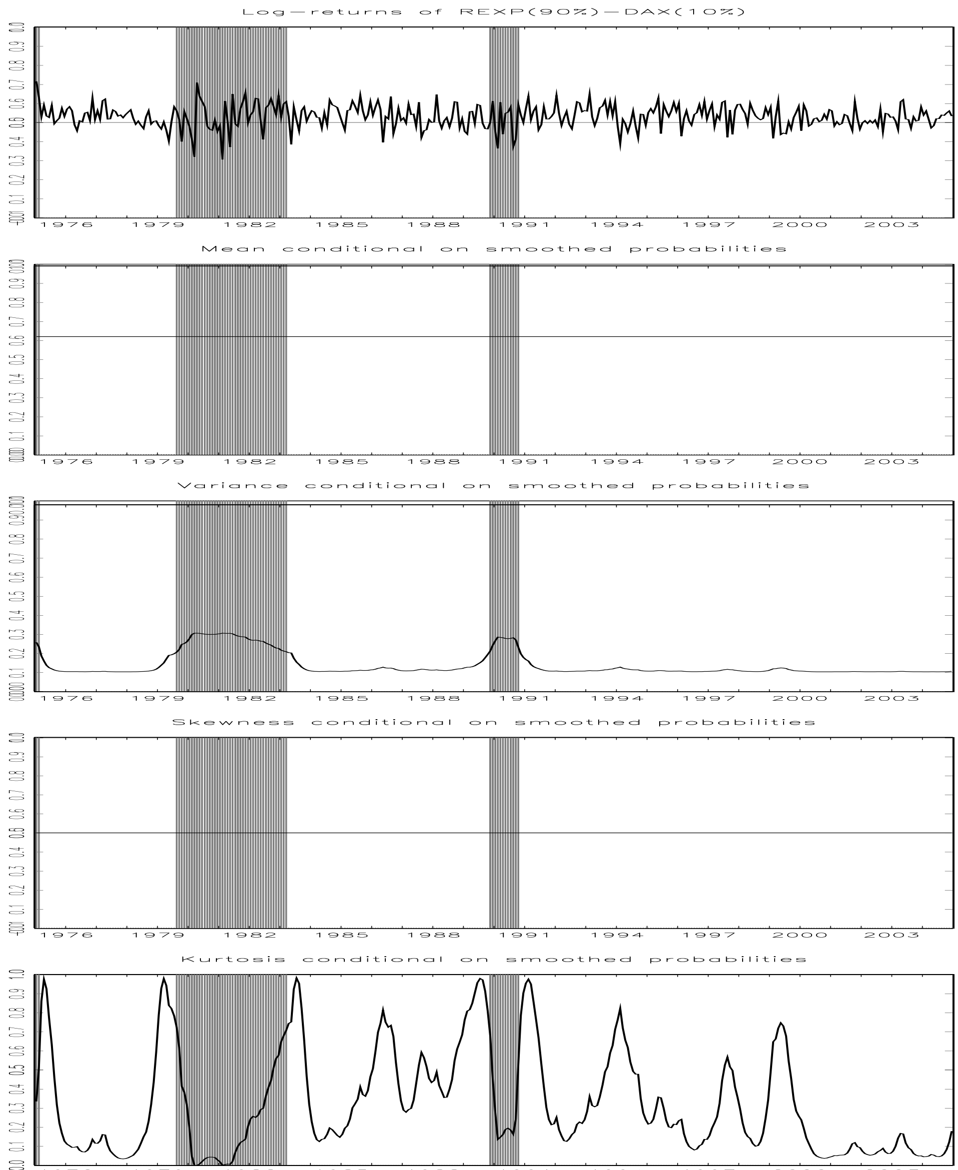
Note:
The top panel shows the log-returns of the REXP(80%)-DAX30(20%) portfolio. The remaining panels indicate the moments for the MS(2-1) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high mean regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.33: Conditional moments: REXP(80%)-DAX30(20%) - MS(2-2)-AR(0)



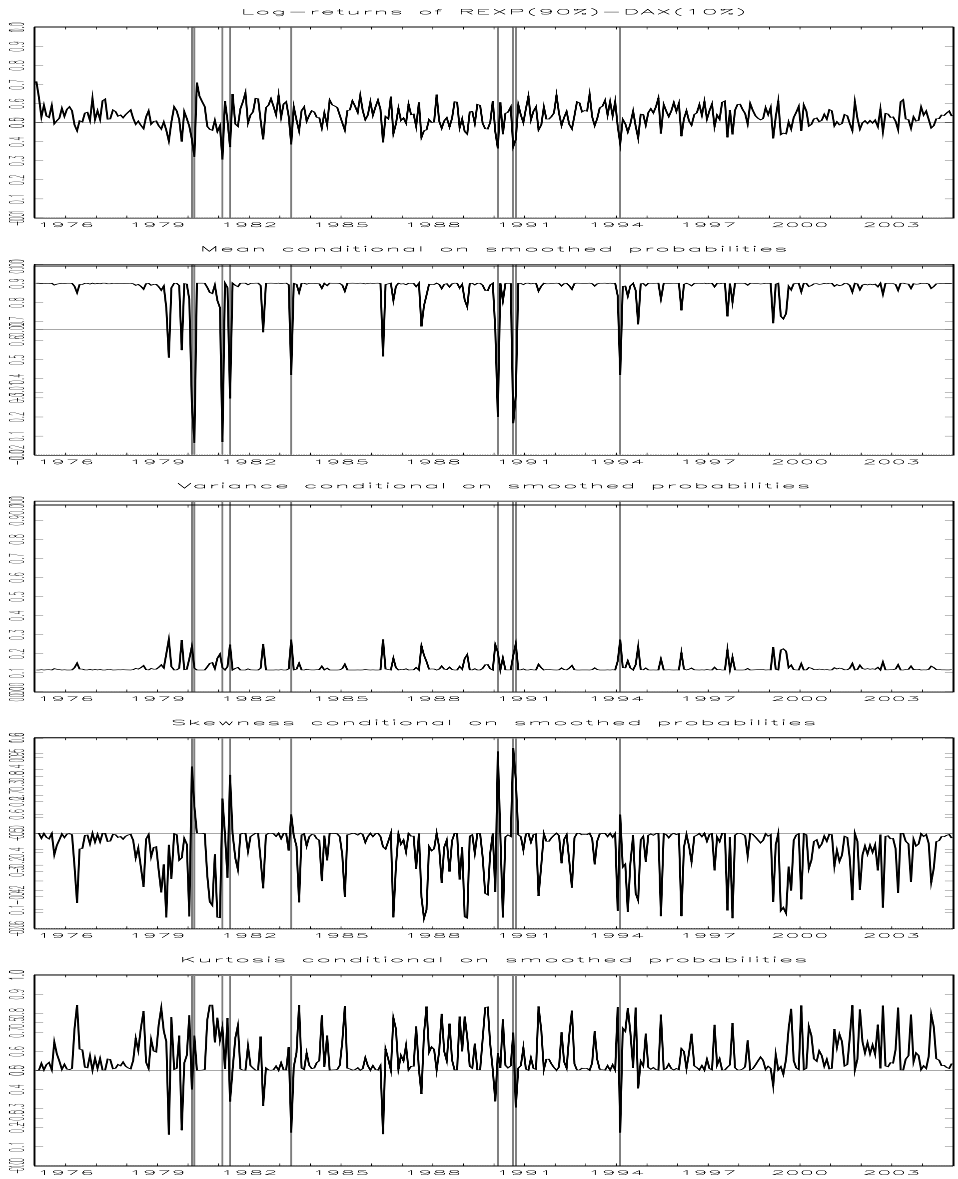
Note:
The top panel shows the log-returns of the REXP(80%)-DAX30(20%) portfolio. The remaining panels indicate the moments for the MS(2-2) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high volatility regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.34: Conditional moments: REXP(90%)-DAX30(10%) - MS(1-2)-AR(0)



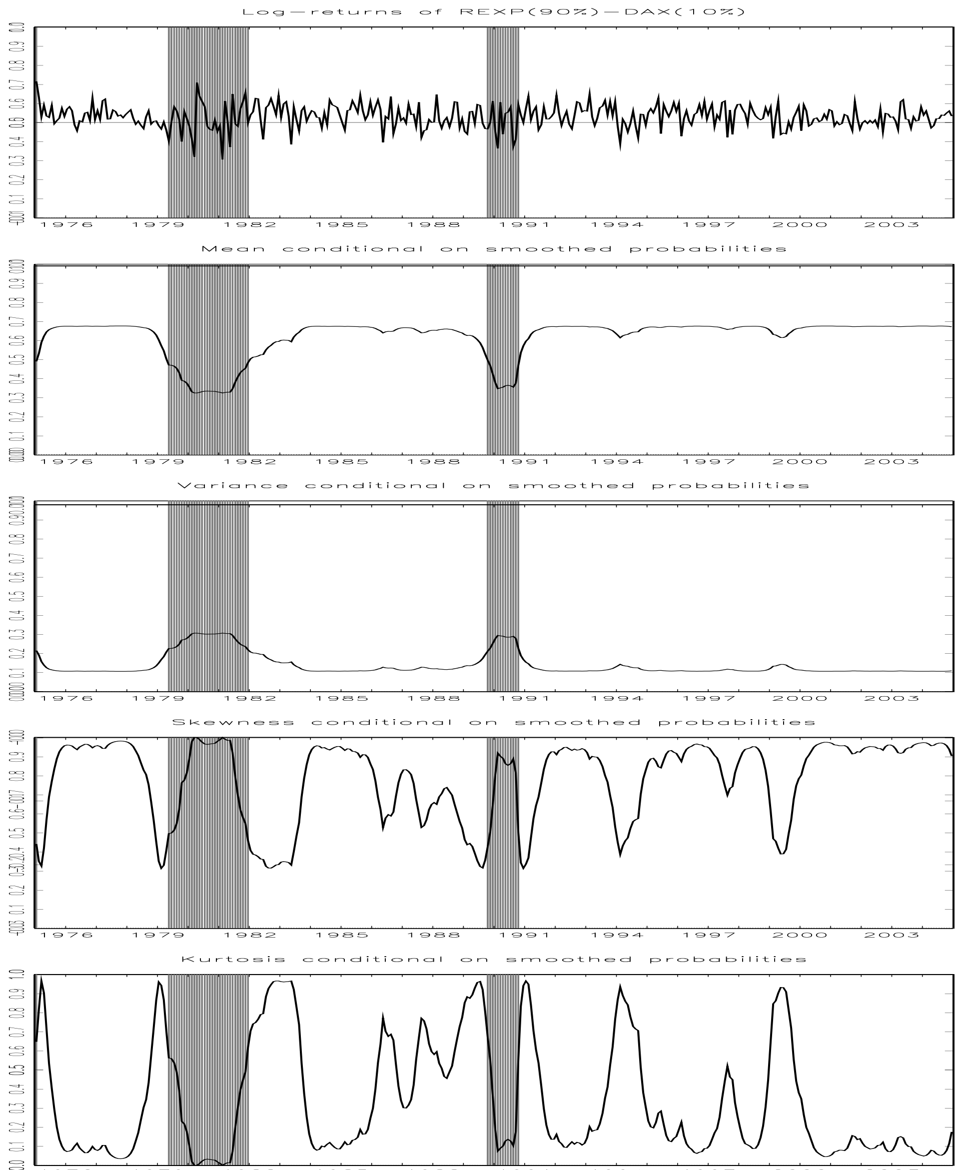
Note:
The top panel shows the log-returns of the REXP(90%)-DAX30(10%) portfolio. The remaining panels indicate the moments for the MS(1-2) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high volatility regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.35: Conditional moments: REXP(90%)-DAX30(10%) - MS(2-1)-AR(0)



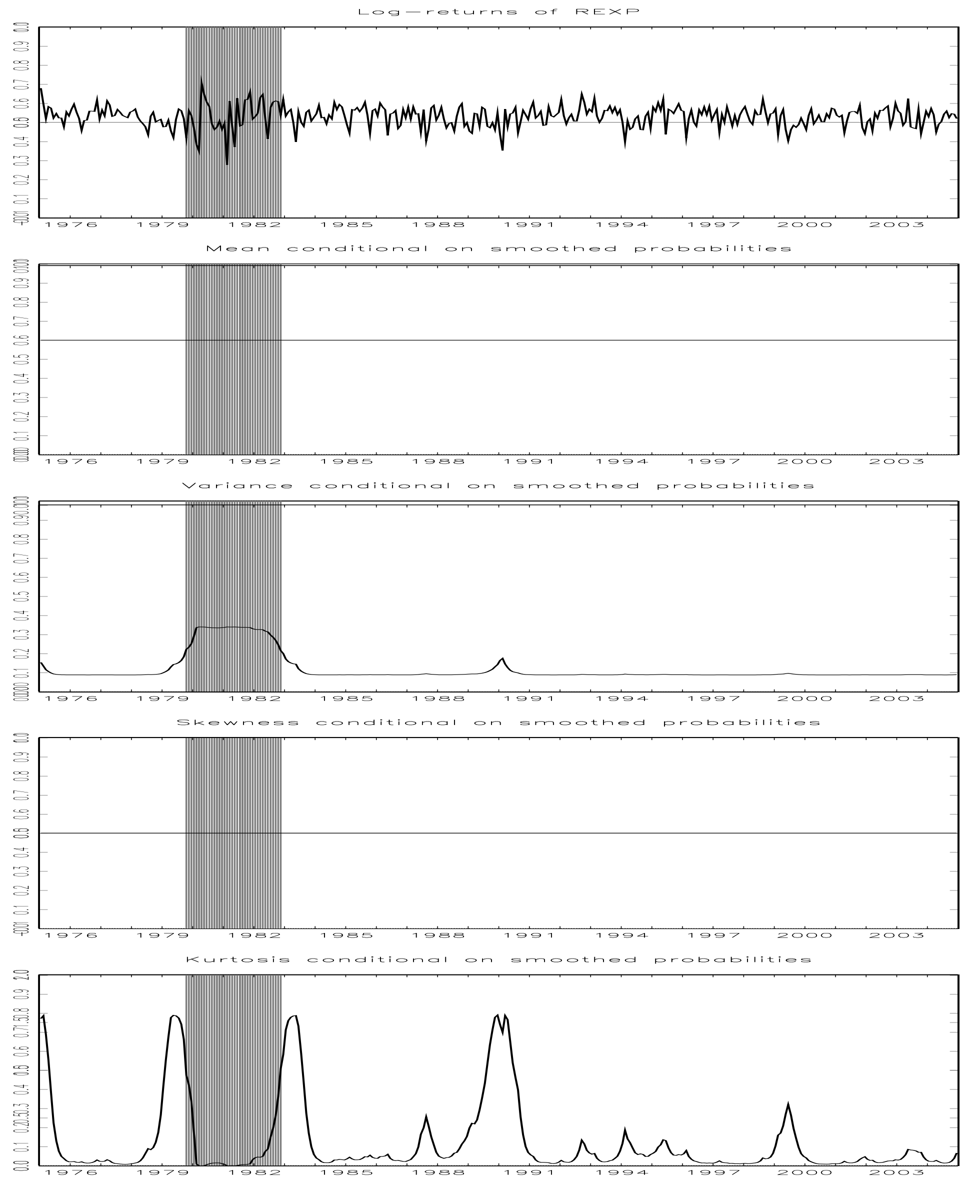
Note:
The top panel shows the log-returns of the REXP(90%)-DAX30(10%) portfolio. The remaining panels indicate the moments for the MS(2-1) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high mean regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.36: Conditional moments: REXP(90%)-DAX30(10%) - MS(2-2)-AR(0)



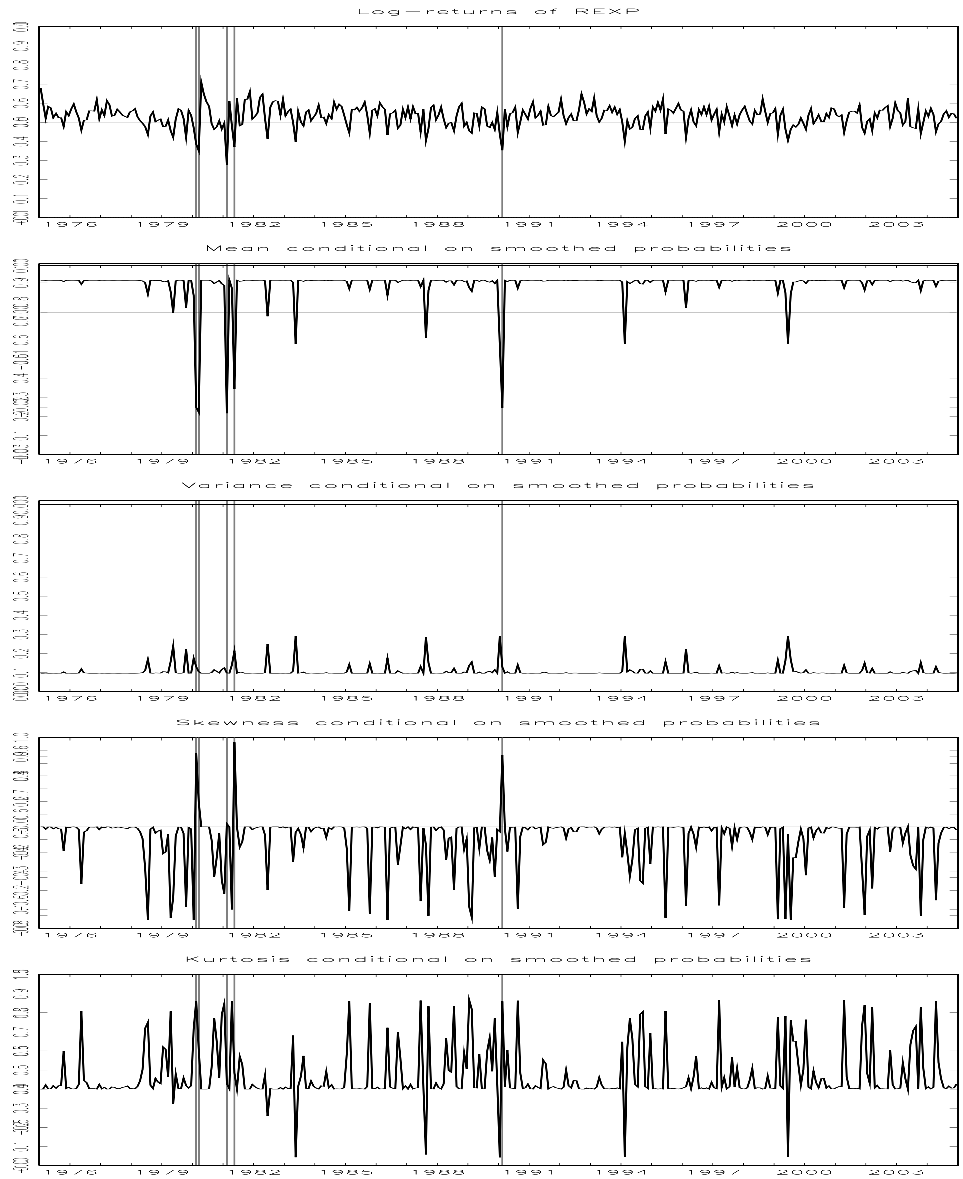
Note:
The top panel shows the log-returns of the REXP(90%)-DAX30(10%) portfolio. The remaining panels indicate the moments for the MS(2-2) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high volatility regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.37: Conditional moments: REXP - MS(1-2)-AR(0)



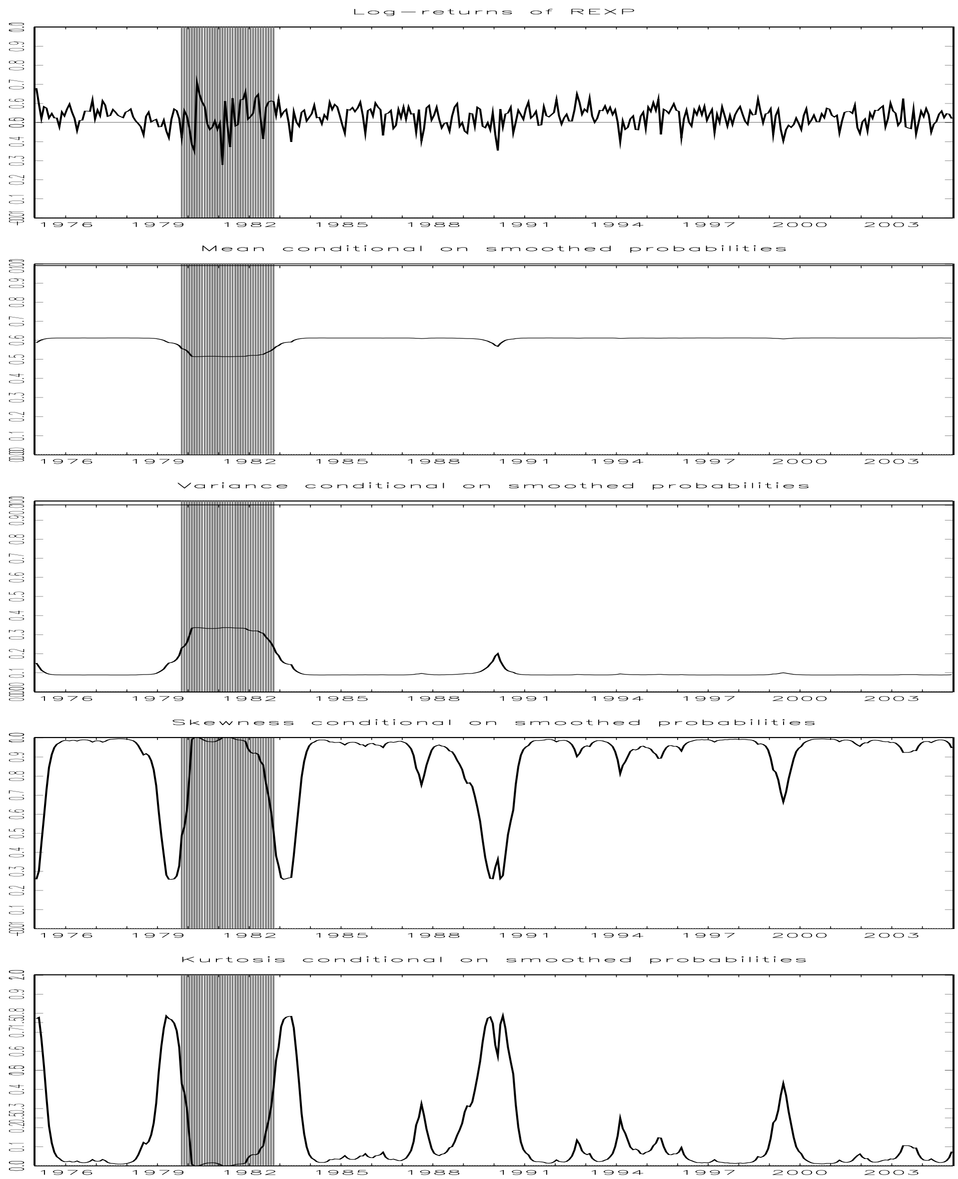
Note:
The top panel shows the log-returns of the pure REXP portfolio. The remaining panels indicate the moments for the MS(1-2) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high volatility regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.38: Conditional moments: REXP - MS(2-1)-AR(0)



Note:
The top panel shows the log-returns of the pure REXP portfolio. The remaining panels indicate the moments for the MS(2-1) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high mean regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

Figure D.39: Conditional moments: REXP - MS(2-2)-AR(0)

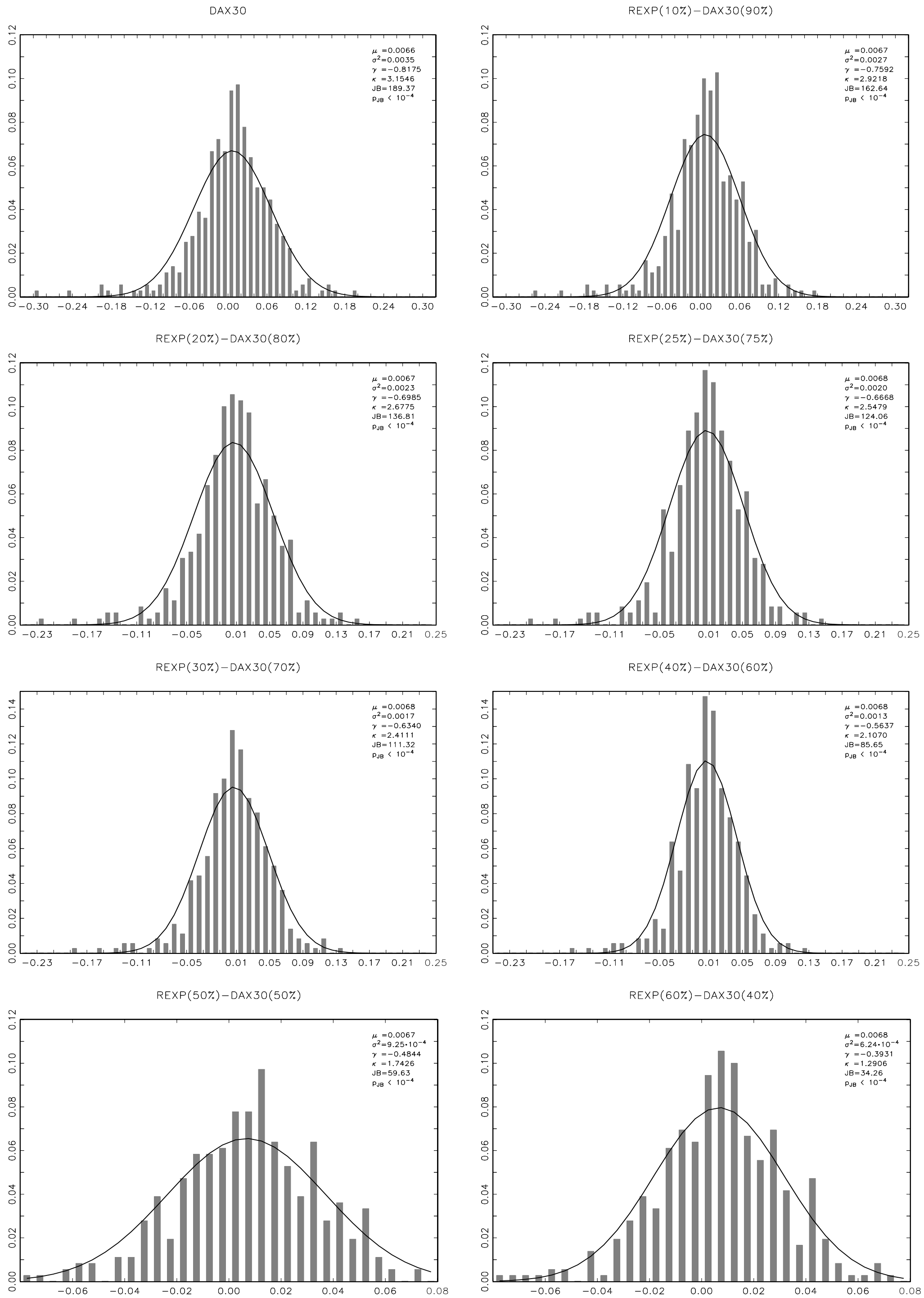


Note:
The top panel shows the log-returns of the pure REXP portfolio. The remaining panels indicate the moments for the MS(2-2) model conditional on the smoothed probabilities, computed with equations (2.57)-(2.60) with π_j replaced by smoothed probabilities $\Pr[Z_{t_n} = j|\mathscr{Y}_T]$. The second panel depicts the conditional expectation, the third panel shows the conditional variance, the fourth panel displays the conditional skewness, and the bottom panel shows the conditional excess kurtosis. The grey area represents the high volatility regime (i.e the smoothed probability $\Pr[Z_{t_n} = 2|\mathscr{Y}_T] > 0.5$).

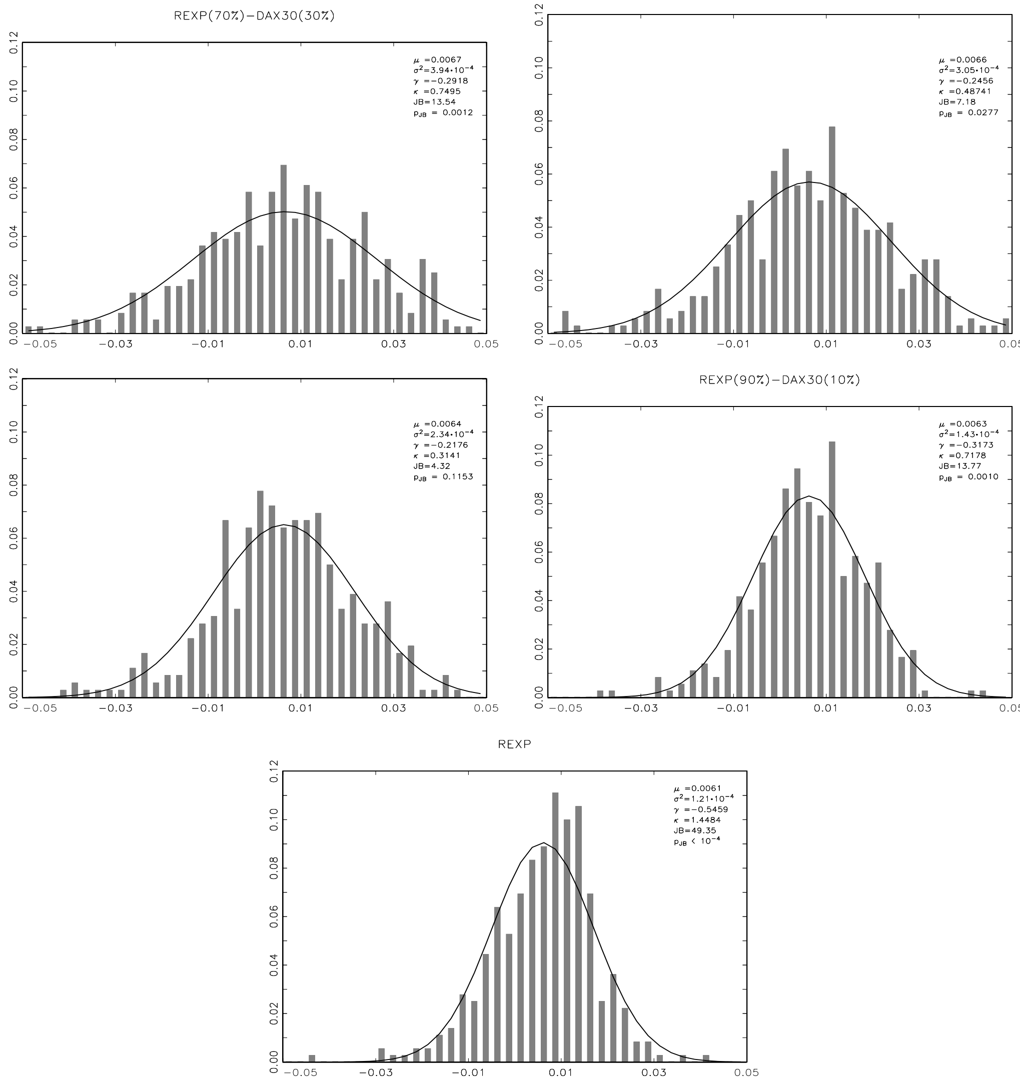
Appendix E

Histogram of the log-returns (1.1975-12.2004)

Figure E.1: Histograms of the log-returns of the REXP-DAX30 portfolios



Continued on the next page.



Note:

The diagrams depict the histograms of the log-returns of the DAX30/REXP mixed portfolios (grey bars). The values of the first four moments of these distributions (μ for mean, σ^2 for variance, γ for skewness, and κ for the excess-kurtosis) and the Jarque-Bera test statistic (JB) with the associated p value (p_{JB}) are located in the top right corner. The solid line represents the density of the normal distribution with the same mean (μ) and variance (σ^2) as the empirical distribution of the log-returns. Note that for estimated portfolios the histograms are plotted for different intervals.

Appendix F

Information Criterion Tests (1.1975-12.2004)

Table F.1: Information Criterion Tests: DAX30 (1.1975-12.2004)

Model	$\mathscr{L}(\boldsymbol{\theta})$	**AIC**	**SBC**
GBM	505.3850	503.3850	499.4989
AR(1)	505.1181	502.1181	496.2889
MS(1-2)-AR(0)	539.1954	**534.1954***	**524.4802***
MS(1-2)-AR(1)	538.9984	532.9984	521.3401
MS(2-1)-AR(0)	530.8818	525.8818	516.1665
MS(2-1)-AR(1)	532.1312	525.1312	511.5298
MS(2-2)-AR(0)	539.8721	533.8721	522.2138
MS(2-2)-AR(1)	540.3052	532.3052	516.7608
ARCH(1)-AR(0)	511.4766	508.4766	502.6474
ARCH(1)-AR(1)	510.8315	506.8315	499.0593
ARCH(2)-AR(0)	522.0284	518.0284	510.2562
ARCH(2)-AR(1)	521.4442	516.4442	506.7289
ARCH(3)-AR(0)	526.0642	521.0642	511.3489
ARCH(3)-AR(1)	525.5402	519.5402	507.8819
GARCH(1,1)-AR(0)	529.8907	525.8907	518.1185*
GARCH(1,1)-AR(1)	529.4322	524.4322	514.7169
E-ARCH(1)-AR(0)	513.2663	509.2663	501.4941
E-ARCH(1)-AR(1)	512.9044	507.9044	498.1892
E-ARCH(2)-AR(0)	519.8657	513.8657	502.2074
E-ARCH(2)-AR(1)	519.9709	512.9709	499.3695
E-ARCH(3)-AR(0)	524.5419	516.5419	500.9975
E-ARCH(3)-AR(1)	524.3925	515.3925	497.9050
E-GARCH(1,1)-AR(0)	531.0400	526.0400*	516.3248
E-GARCH(1,1)-AR(1)	530.6973	524.6973	513.0390
T-ARCH(1)-AR(0)	513.0173	509.0173	501.2451
T-ARCH(1)-AR(1)	512.6031	507.6031	497.8878
T-ARCH(2)-AR(0)	523.8126	518.8126	509.0973
T-ARCH(2)-AR(1)	523.4650	517.4650	505.8067
T-ARCH(3)-AR(0)	526.8153	520.8153	509.1569
T-ARCH(3)-AR(1)	526.4255	519.4255	505.8241
T-GARCH(1,1)-AR(0)	530.4080	525.4080	515.6927
T-GARCH(1,1)-AR(1)	530.0564	524.0564	512.3981

Note:
GBM - Geometric Brownian motion, AR(p) denotes an auto-regressive model of the p-th order, MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance, (G)ARCH(p,q) - (Generalised) Auto-regression Conditional Heteroscedasticity with q GARCH processes and p ARCH processes, E-(G)ARCH - exponential (G)ARCH, T-(G)ARCH - threshold (G)ARCH, $\mathscr{L}(\boldsymbol{\theta})$ - log-likelihood, AIC - Akaike Information Criterion $\left[\text{AIC}=\mathscr{L}(\boldsymbol{\theta}) - k\right]$, SBC - Schwarz Bayesian Information Criterion $\left[\text{SBC}=\mathscr{L}(\boldsymbol{\theta}) - 0.5k\ln(T)\right]$, k - number of parameters, T - number of observations. Bold - model with the best information statistic. * - the best statistics in the MS or (G)ARCH model class, respectively.

Table F.2: Information Criterion Tests: REXP(10%)-DAX30(90%) (1.1975-12.2004)

Model	$\mathscr{L}(\boldsymbol{\theta})$	**AIC**	**SBC**
GBM	543.5887	541.5887	537.7026
AR(1)	543.2778	540.2778	534.4486
MS(1-2)-AR(0)	575.5732	**570.5732***	**560.8579***
MS(1-2)-AR(1)	575.3713	569.3713	557.7130
MS(2-1)-AR(0)	567.3510	562.3510	552.6357
MS(2-1)-AR(1)	568.5750	561.5750	547.9736
MS(2-2)-AR(0)	576.1791	570.1791	558.5208
MS(2-2)-AR(1)	576.5748	568.5748	553.0304
ARCH(1)-AR(0)	549.5258	546.5258	540.6967
ARCH(1)-AR(1)	548.8451	544.8451	537.0729
ARCH(2)-AR(0)	559.3363	555.3363	547.5641
ARCH(2)-AR(1)	558.6935	553.6935	543.9782
ARCH(3)-AR(0)	563.1767	558.1767	548.4615
ARCH(3)-AR(1)	562.5998	556.5998	544.9415
GARCH(1,1)-AR(0)	566.7938	562.7938*	555.0216*
GARCH(1,1)-AR(1)	566.2973	561.2973	551.5821
E-ARCH(1)-AR(0)	551.2057	547.2057	539.4335
E-ARCH(1)-AR(1)	550.8272	545.8272	536.1120
E-ARCH(2)-AR(0)	557.1632	551.1632	539.5049
E-ARCH(2)-AR(1)	557.1811	550.1811	536.5797
E-ARCH(3)-AR(0)	561.6330	553.6330	538.0885
E-ARCH(3)-AR(1)	561.4255	552.4255	534.9380
E-GARCH(1,1)-AR(0)	567.4892	562.4892	552.7739
E-GARCH(1,1)-AR(1)	567.1041	561.1041	549.4457
T-ARCH(1)-AR(0)	551.0046	547.0046	539.2324
T-ARCH(1)-AR(1)	550.5849	545.5849	535.8697
T-ARCH(2)-AR(0)	561.0884	556.0884	546.3731
T-ARCH(2)-AR(1)	560.7067	554.7067	543.0484
T-ARCH(3)-AR(0)	563.9340	557.9340	546.2757
T-ARCH(3)-AR(1)	563.5006	556.5006	542.8993
T-GARCH(1,1)-AR(0)	567.1168	562.1168	552.4016
T-GARCH(1,1)-AR(1)	566.7122	560.7122	549.0539

Note:
GBM - Geometric Brownian motion, AR(p) denotes an auto-regressive model of the p-th order, MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance, (G)ARCH(p,q) - (Generalised) Auto-regression Conditional Heteroscedasticity with q GARCH processes and p ARCH processes, E-(G)ARCH - exponential (G)ARCH, T-(G)ARCH - threshold (G)ARCH, $\mathscr{L}(\boldsymbol{\theta})$ - log-likelihood, AIC - Akaike Information Criterion $\left[\text{AIC}=\mathscr{L}(\boldsymbol{\theta}) - k\right]$, SBC - Schwarz Bayesian Information Criterion $\left[\text{SBC}=\mathscr{L}(\boldsymbol{\theta}) - 0.5k\ln(T)\right]$, k - number of parameters, T - number of observations. Bold - model with the best information statistic. * - the best statistics in the MS or (G)ARCH model class, respectively.

Table F.3: Information Criterion Tests: REXP(20%)-DAX30(80%) (1.1975-12.2004)

Model	$\mathscr{L}(\boldsymbol{\theta})$	**AIC**	**SBC**
GBM	585.7735	583.7735	579.8874
AR(1)	585.4329	582.4329	576.6037
MS(1-2)-AR(0)	615.5740	**610.5740***	**600.8588***
MS(1-2)-AR(1)	615.3795	609.3795	597.7212
MS(2-1)-AR(0)	607.6496	602.6496	592.9343
MS(2-1)-AR(1)	608.8376	601.8376	588.2362
MS(2-2)-AR(0)	616.1200	610.1200	598.4617
MS(2-2)-AR(1)	616.4807	608.4807	592.9363
ARCH(1)-AR(0)	591.5056	588.5056	582.6764
ARCH(1)-AR(1)	590.7971	586.7971	579.0249
ARCH(2)-AR(0)	600.4899	596.4899	588.7177
ARCH(2)-AR(1)	599.7873	594.7873	585.0720
ARCH(3)-AR(0)	604.0787	599.0787	589.3634
ARCH(3)-AR(1)	603.4471	597.4471	585.7888
GARCH(1,1)-AR(0)	607.4099	603.4099*	595.6377*
GARCH(1,1)-AR(1)	606.8781	601.8781	592.1628
E-ARCH(1)-AR(0)	593.0388	589.0388	581.2666
E-ARCH(1)-AR(1)	592.6514	587.6514	577.9361
E-ARCH(2)-AR(0)	598.2876	592.2876	580.6293
E-ARCH(2)-AR(1)	598.2330	591.2330	577.6316
E-ARCH(3)-AR(0)	602.4986	594.4986	578.9542
E-ARCH(3)-AR(1)	602.2370	593.2370	575.7496
E-GARCH(1,1)-AR(0)	607.5735	602.5735	592.8583
E-GARCH(1,1)-AR(1)	607.1185	601.1185	589.4602
T-ARCH(1)-AR(0)	592.9157	588.9157	581.1435
T-ARCH(1)-AR(1)	592.5029	587.5029	577.7877
T-ARCH(2)-AR(0)	602.2085	597.2085	587.4933
T-ARCH(2)-AR(1)	601.7954	595.7954	584.1371
T-ARCH(3)-AR(0)	604.8545	598.8545	587.1962
T-ARCH(3)-AR(1)	604.3792	597.3792	583.7779
T-GARCH(1,1)-AR(0)	607.5611	602.5611	592.8458
T-GARCH(1,1)-AR(1)	607.0997	601.0997	589.4414

Note:
GBM - Geometric Brownian motion, AR(p) denotes an auto-regressive model of the p-th order, MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance, (G)ARCH(p,q) - (Generalised) Auto-regression Conditional Heteroscedasticity with q GARCH processes and p ARCH processes, E-(G)ARCH - exponential (G)ARCH, T-(G)ARCH - threshold (G)ARCH, $\mathscr{L}(\boldsymbol{\theta})$ - log-likelihood, AIC - Akaike Information Criterion $\left[\text{AIC}=\mathscr{L}(\boldsymbol{\theta}) - k\right]$, SBC - Schwarz Bayesian Information Criterion $\left[\text{SBC}=\mathscr{L}(\boldsymbol{\theta}) - 0.5k\ln(T)\right]$, k - number of parameters, T - number of observations. Bold - model with the best information statistic. * - the best statistics in the MS or (G)ARCH model class, respectively.

Table F.4: Information Criterion Tests: REXP(25%)-DAX30(75%) (1.1975-12.2004)

Model	$\mathscr{L}(\boldsymbol{\theta})$	AIC	SBC
GBM	608.6370	606.6370	602.7509
AR(1)	608.2896	605.2896	599.4604
MS(1-2)-AR(0)	637.1615	**632.1615***	**622.4462***
MS(1-2)-AR(1)	636.9782	630.9782	619.3199
MS(2-1)-AR(0)	629.4895	624.4895	614.7742
MS(2-1)-AR(1)	630.6530	623.6530	610.0517
MS(2-2)-AR(0)	637.6832	631.6832	620.0249
MS(2-2)-AR(1)	638.0285	630.0285	614.4841
ARCH(1)-AR(0)	614.2417	611.2417	605.4125
ARCH(1)-AR(1)	613.5238	609.5238	601.7516
ARCH(2)-AR(0)	622.7696	618.7696	610.9974
ARCH(2)-AR(1)	622.0374	617.0374	607.3222
ARCH(3)-AR(0)	626.2048	621.2048	611.4895
ARCH(3)-AR(1)	625.5455	619.5455	607.8872
GARCH(1,1)-AR(0)	629.3608	625.360*	617.5886*
GARCH(1,1)-AR(1)	628.8128	623.8128	614.0976
E-ARCH(1)-AR(0)	615.6853	611.6853	603.9131
E-ARCH(1)-AR(1)	615.3024	610.3024	600.5872
E-ARCH(2)-AR(0)	620.5718	614.5718	602.9134
E-ARCH(2)-AR(1)	620.4793	613.4793	599.8780
E-ARCH(3)-AR(0)	624.5955	616.5955	601.0511
E-ARCH(3)-AR(1)	624.2556	615.2556	597.7681
E-GARCH(1,1)-AR(0)	629.2497	624.2497	614.5345
E-GARCH(1,1)-AR(1)	628.7615	622.7615	611.1031
T-ARCH(1)-AR(0)	615.6133	611.6133	603.8411
T-ARCH(1)-AR(1)	615.2107	610.2107	600.4955
T-ARCH(2)-AR(0)	624.4692	619.4692	609.7540
T-ARCH(2)-AR(1)	624.0421	618.0421	606.3838
T-ARCH(3)-AR(0)	626.9950	620.9950	609.3367
T-ARCH(3)-AR(1)	626.5001	619.5001	605.8988
T-GARCH(1,1)-AR(0)	629.4428	624.4428	614.7275
T-GARCH(1,1)-AR(1)	628.9513	622.9513	611.2930

Note:
GBM - Geometric Brownian motion, AR(p) denotes an auto-regressive model of the p-th order, MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance, (G)ARCH(p,q) - (Generalised) Auto-regression Conditional Heteroscedasticity with q GARCH processes and p ARCH processes, E-(G)ARCH - exponential (G)ARCH, T-(G)ARCH - threshold (G)ARCH, $\mathscr{L}(\boldsymbol{\theta})$ - log-likelihood, AIC - Akaike Information Criterion $\left[\text{AIC}=\mathscr{L}(\boldsymbol{\theta}) - k\right]$, SBC - Schwarz Bayesian Information Criterion $\left[\text{SBC}=\mathscr{L}(\boldsymbol{\theta}) - 0.5k\ln(T)\right]$, k - number of parameters, T - number of observations. Bold - model with the best information statistic. * - the best statistics in the MS or (G)ARCH model class, respectively.

Table F.5: Information Criterion Tests: REXP(30%)-DAX30(70%) (1.1975-12.2004)

Model	$\mathscr{L}(\boldsymbol{\theta})$	**AIC**	**SBC**
GBM	632.8643	630.8643	626.9782
AR(1)	632.5173	629.5173	623.6882
MS(1-2)-AR(0)	659.9572	**654.9572***	**645.2420***
MS(1-2)-AR(1)	659.7898	653.7898	642.1315
MS(2-1)-AR(0)	652.6225	647.6225	637.9072
MS(2-1)-AR(1)	653.7556	646.7556	633.1542
MS(2-2)-AR(0)	660.4584	654.4584	642.8001
MS(2-2)-AR(1)	660.7890	652.7890	637.2446
ARCH(1)-AR(0)	638.3207	635.3207	629.4915
ARCH(1)-AR(1)	637.5980	633.5980	625.8258
ARCH(2)-AR(0)	646.3558	642.3558	634.5836
ARCH(2)-AR(1)	645.5949	640.5949	630.8796
ARCH(3)-AR(0)	649.6143	644.6143	634.8991
ARCH(3)-AR(1)	648.9278	642.9278	631.2695
GARCH(1,1)-AR(0)	652.5720	648.5720*	640.7998*
GARCH(1,1)-AR(1)	652.0094	647.0094	637.2941
E-ARCH(1)-AR(0)	639.6556	635.6556	627.8834
E-ARCH(1)-AR(1)	639.2873	634.2873	624.5720
E-ARCH(2)-AR(0)	644.1798	638.1798	626.5215
E-ARCH(2)-AR(1)	644.0385	637.0385	623.4371
E-ARCH(3)-AR(0)	647.9908	639.9908	624.4464
E-ARCH(3)-AR(1)	647.6933	638.6933	621.2058
E-GARCH(1,1)-AR(0)	652.2033	647.2033	637.4881
E-GARCH(1,1)-AR(1)	651.7719	645.7719	634.1136
T-ARCH(1)-AR(0)	639.6499	635.6499	627.8777
T-ARCH(1)-AR(1)	639.2634	634.2634	624.5481
T-ARCH(2)-AR(0)	648.0331	643.0331	633.3178
T-ARCH(2)-AR(1)	647.5932	641.5932	629.9349
T-ARCH(3)-AR(0)	650.4225	644.4225	632.7641
T-ARCH(3)-AR(1)	649.9095	642.9095	629.3081
T-GARCH(1,1)-AR(0)	652.6025	647.6025	637.8872
T-GARCH(1,1)-AR(1)	652.0794	646.0794	634.4211

Note:
GBM - Geometric Brownian motion, AR(p) denotes an auto-regressive model of the p-th order, MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance, (G)ARCH(p,q) - (Generalised) Auto-regression Conditional Heteroscedasticity with q GARCH processes and p ARCH processes, E-(G)ARCH - exponential (G)ARCH, T-(G)ARCH - threshold (G)ARCH, $\mathscr{L}(\boldsymbol{\theta})$ - log-likelihood, AIC - Akaike Information Criterion $\left[\text{AIC}=\mathscr{L}(\boldsymbol{\theta}) - k\right]$, SBC - Schwarz Bayesian Information Criterion $\left[\text{SBC}=\mathscr{L}(\boldsymbol{\theta}) - 0.5k\ln(T)\right]$, k - number of parameters, T - number of observations. Bold - model with the best information statistic. * - the best statistics in the MS or (G)ARCH model class, respectively.

Table F.6: Information Criterion Tests: REXP(40%)-DAX30(60%) (1.1975-12.2004)

Model	$\mathscr{L}(\boldsymbol{\theta})$	**AIC**	**SBC**
GBM	686.0921	684.0921	680.2060
AR(1)	685.7774	682.7774	676.9482
MS(1-2)-AR(0)	709.7270	**704.7270***	**695.0118***
MS(1-2)-AR(1)	709.6024	703.6024	691.9441
MS(2-1)-AR(0)	703.3806	698.3806	688.6653
MS(2-1)-AR(1)	704.4328	697.4328	683.8314
MS(2-2)-AR(0)	710.1928	704.1928	692.5345
MS(2-2)-AR(1)	710.5015	702.5015	686.9571
ARCH(1)-AR(0)	691.1746	688.1746	682.3455
ARCH(1)-AR(1)	690.4620	686.4620	678.6898
ARCH(2)-AR(0)	698.0841	694.0841	686.3119
ARCH(2)-AR(1)	697.2734	692.2734	682.5581
ARCH(3)-AR(0)	700.9072	695.9072	686.1920
ARCH(3)-AR(1)	700.1700	694.1700	682.5117
GARCH(1,1)-AR(0)	703.3973	699.3973*	691.6251*
GARCH(1,1)-AR(1)	702.8121	697.8121	688.0969
E-ARCH(1)-AR(0)	692.3283	688.3283	680.5561
E-ARCH(1)-AR(1)	692.0058	687.0058	677.2906
E-ARCH(2)-AR(0)	696.3620	690.3620	678.7037
E-ARCH(2)-AR(1)	696.0833	689.0833	675.4820
E-ARCH(3)-AR(0)	699.5672	691.5672	676.0228
E-ARCH(3)-AR(1)	699.3367	690.3367	672.8492
E-GARCH(1,1)-AR(0)	702.8994	697.8994	688.1842
E-GARCH(1,1)-AR(1)	702.4073	696.4073	684.7490
T-ARCH(1)-AR(0)	692.3972	688.3972	680.6250
T-ARCH(1)-AR(1)	692.0649	687.0649	677.3496
T-ARCH(2)-AR(0)	699.6942	694.6942	684.9790
T-ARCH(2)-AR(1)	699.2363	693.2363	681.5780
T-ARCH(3)-AR(0)	701.7568	695.7568	684.0985
T-ARCH(3)-AR(1)	701.2157	694.2157	680.6143
T-GARCH(1,1)-AR(0)	703.4039	698.4039	688.6886
T-GARCH(1,1)-AR(1)	702.8124	696.8124	685.1541

Note:
GBM - Geometric Brownian motion, AR(p) denotes an auto-regressive model of the p-th order, MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance, (G)ARCH(p,q) - (Generalised) Auto-regression Conditional Heteroscedasticity with q GARCH processes and p ARCH processes, E-(G)ARCH - exponential (G)ARCH, T-(G)ARCH - threshold (G)ARCH, $\mathscr{L}(\boldsymbol{\theta})$ - log-likelihood, AIC - Akaike Information Criterion $\left[\text{AIC}=\mathscr{L}(\boldsymbol{\theta}) - k\right]$, SBC - Schwarz Bayesian Information Criterion $\left[\text{SBC}=\mathscr{L}(\boldsymbol{\theta}) - 0.5k\ln(T)\right]$, k - number of parameters, T - number of observations. Bold - model with the best information statistic. * - the best statistics in the MS or (G)ARCH model class, respectively.

Table F.7: Information Criterion Tests: REXP(50%)-DAX30(50%) (1.1975-12.2004)

Model	$\mathscr{L}(\boldsymbol{\theta})$	**AIC**	**SBC**
GBM	747.0932	745.0932	741.2071
AR(1)	746.8760	743.8760	738.0469
MS(1-2)-AR(0)	766.2067	**761.2067***	**751.4915***
MS(1-2)-AR(1)	766.1446	760.1446	748.4862
MS(2-1)-AR(0)	761.3885	756.3885	746.6733
MS(2-1)-AR(1)	762.3362	755.3362	741.7348
MS(2-2)-AR(0)	766.6560	760.6560	748.9977
MS(2-2)-AR(1)	766.9586	758.9586	743.4142
ARCH(1)-AR(0)	751.6724	748.6724	742.8432
ARCH(1)-AR(1)	751.0151	747.0151	739.2429
ARCH(2)-AR(0)	757.2090	753.2090	745.4368
ARCH(2)-AR(1)	756.3753	751.3753	741.6600
ARCH(3)-AR(0)	759.4674	754.4674	744.7521
ARCH(3)-AR(1)	758.6964	752.6964	741.0381
GARCH(1,1)-AR(0)	761.4098	757.4098*	749.6376*
GARCH(1,1)-AR(1)	760.8167	755.8167	746.1015
E-ARCH(1)-AR(0)	752.5376	748.5376	740.7654
E-ARCH(1)-AR(1)	752.3341	747.3341	737.6189
E-ARCH(2)-AR(0)	755.9808	749.9808	738.3225
E-ARCH(2)-AR(1)	755.8381	748.8381	735.2367
E-ARCH(3)-AR(0)	758.5124	750.5124	734.9680
E-ARCH(3)-AR(1)	758.3807	749.3807	731.8933
E-GARCH(1,1)-AR(0)	761.0593	756.0593	746.3440
E-GARCH(1,1)-AR(1)	760.4096	754.4096	742.7513
T-ARCH(1)-AR(0)	752.7308	748.7308	740.9586
T-ARCH(1)-AR(1)	752.5002	747.5002	737.7849
T-ARCH(2)-AR(0)	758.6795	753.6795	743.9642
T-ARCH(2)-AR(1)	758.2226	752.2226	740.5643
T-ARCH(3)-AR(0)	760.3404	754.3404	742.6820
T-ARCH(3)-AR(1)	759.7906	752.7906	739.1892
T-GARCH(1,1)-AR(0)	761.5563	756.5563	746.8411
T-GARCH(1,1)-AR(1)	760.8911	754.8911	743.2328

Note:
GBM - Geometric Brownian motion, AR(p) denotes an auto-regressive model of the p-th order, MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance, (G)ARCH(p,q) - (Generalised) Auto-regression Conditional Heteroscedasticity with q GARCH processes and p ARCH processes, E-(G)ARCH - exponential (G)ARCH, T-(G)ARCH - threshold (G)ARCH, $\mathscr{L}(\boldsymbol{\theta})$ - log-likelihood, AIC - Akaike Information Criterion $\left[\text{AIC}=\mathscr{L}(\boldsymbol{\theta}) - k\right]$, SBC - Schwarz Bayesian Information Criterion $\left[\text{SBC}=\mathscr{L}(\boldsymbol{\theta}) - 0.5k\ln(T)\right]$, k - number of parameters, T - number of observations. Bold - model with the best information statistic. * - the best statistics in the MS or (G)ARCH model class, respectively.

Table F.8: Information Criterion Tests: REXP(60%)-DAX30(40%) (1.1975-12.2004)

Model	$\mathscr{L}(\boldsymbol{\theta})$	**AIC**	**SBC**
GBM	817.9412	815.9412	812.0551
AR(1)	817.9352	814.9352	809.1061
MS(1-2)-AR(0)	831.4834	**826.4834***	816.7681*
MS(1-2)-AR(1)	831.3772	825.3772	813.7189
MS(2-1)-AR(0)	828.5282	823.5282	813.8129
MS(2-1)-AR(1)	829.3607	822.3607	808.7593
MS(2-2)-AR(0)	831.8985	825.8985	814.2402
MS(2-2)-AR(1)	832.1894	824.1894	808.6450
ARCH(1)-AR(0)	821.8803	818.8803	813.0511
ARCH(1)-AR(1)	821.3667	817.3667	809.5945
ARCH(2)-AR(0)	825.7276	821.7276	813.9554
ARCH(2)-AR(1)	824.9485	819.9485	810.2332
ARCH(3)-AR(0)	827.2926	822.2926	812.5774
ARCH(3)-AR(1)	826.5410	820.5410	808.8826
GARCH(1,1)-AR(0)	828.6796	824.6796*	**816.9074***
GARCH(1,1)-AR(1)	828.1091	823.1091	813.3938
E-ARCH(1)-AR(0)	822.4937	818.4937	810.7215
E-ARCH(1)-AR(1)	822.5206	817.5206	807.8053
E-ARCH(2)-AR(0)	825.2233	819.2233	807.5650
E-ARCH(2)-AR(1)	825.3152	818.3152	804.7138
E-ARCH(3)-AR(0)	826.9675	818.9675	803.4231
E-ARCH(3)-AR(1)	827.0512	818.0512	800.5637
E-GARCH(1,1)-AR(0)	828.9814	823.9814	814.2661
E-GARCH(1,1)-AR(1)	828.2141	822.2141	810.5558
T-ARCH(1)-AR(0)	822.6968	818.6968	810.9246
T-ARCH(1)-AR(1)	822.6525	817.6525	807.9373
T-ARCH(2)-AR(0)	826.9516	821.9516	812.2363
T-ARCH(2)-AR(1)	826.5499	820.5499	808.8916
T-ARCH(3)-AR(0)	828.1449	822.1449	810.4866
T-ARCH(3)-AR(1)	827.6377	820.6377	807.0363
T-GARCH(1,1)-AR(0)	829.1975	824.1975	814.4822
T-GARCH(1,1)-AR(1)	828.4623	822.4623	810.8040

Note:
GBM - Geometric Brownian motion, AR(p) denotes an auto-regressive model of the p-th order, MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance, (G)ARCH(p,q) - (Generalised) Auto-regression Conditional Heteroscedasticity with q GARCH processes and p ARCH processes, E-(G)ARCH - exponential (G)ARCH, T-(G)ARCH - threshold (G)ARCH, $\mathscr{L}(\boldsymbol{\theta})$ - log-likelihood, AIC - Akaike Information Criterion $\left[\text{AIC}=\mathscr{L}(\boldsymbol{\theta}) - k\right]$, SBC - Schwarz Bayesian Information Criterion $\left[\text{SBC}=\mathscr{L}(\boldsymbol{\theta}) - 0.5k\ln(T)\right]$, k - number of parameters, T - number of observations. Bold - model with the best information statistic. * - the best statistics in the MS or (G)ARCH model class, respectively.

Table F.9: Information Criterion Tests: REXP(70%)-DAX30(30%) (1.1975-12.2004)

Model	$\mathscr{L}(\boldsymbol{\theta})$	**AIC**	**SBC**
GBM	900.6899	898.6899	894.8038*
AR(1)	901.1033	898.1033	892.2741
MS(1-2)-AR(0)	908.6233	903.6233*	893.9080*
MS(1-2)-AR(1)	908.6912	902.6912	891.0329
MS(2-1)-AR(0)	906.9026	901.9026	892.1873
MS(2-1)-AR(1)	907.7328	900.7328	887.1314
MS(2-2)-AR(0)	909.3142	903.3142	891.6559
MS(2-2)-AR(1)	910.1357	902.1357	886.5913
ARCH(1)-AR(0)	903.9931	900.9931	895.1640
ARCH(1)-AR(1)	903.8112	899.8112	892.0390
ARCH(2)-AR(0)	905.9051	901.9051	894.1329
ARCH(2)-AR(1)	905.4121	900.4121	890.6968
ARCH(3)-AR(0)	906.7322	901.7322	892.0169
ARCH(3)-AR(1)	906.1795	900.1795	888.5212
GARCH(1,1)-AR(0)	907.7665	903.7665	**895.9943**
GARCH(1,1)-AR(1)	907.3124	902.3124	892.5972
E-ARCH(1)-AR(0)	904.2318	900.2318	892.4596
E-ARCH(1)-AR(1)	904.7370	899.7370	890.0218
E-ARCH(2)-AR(0)	905.8549	899.8549	888.1966
E-ARCH(2)-AR(1)	906.4341	899.4341	885.8327
E-ARCH(3)-AR(0)	906.8083	898.8083	883.2639
E-ARCH(3)-AR(1)	907.3730	898.3730	880.8856
E-GARCH(1,1)-AR(0)	908.6020	903.6020	893.8868
E-GARCH(1,1)-AR(1)	907.9011	901.9011	890.2428
T-ARCH(1)-AR(0)	904.4911	900.4911	892.7189
T-ARCH(1)-AR(1)	904.8161	899.8161	890.1009
T-ARCH(2)-AR(0)	906.7210	901.7210	892.0057
T-ARCH(2)-AR(1)	906.5925	900.5925	888.9342
T-ARCH(3)-AR(0)	907.4473	901.4473	889.7890
T-ARCH(3)-AR(1)	907.1656	900.1656	886.5643
T-GARCH(1,1)-AR(0)	908.7736	**903.7736**	894.0583
T-GARCH(1,1)-AR(1)	908.0695	902.0695	890.4112

Note:
GBM - Geometric Brownian motion, AR(p) denotes an auto-regressive model of the p-th order, MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance, (G)ARCH(p,q) - (Generalised) Auto-regression Conditional Heteroscedasticity with q GARCH processes and p ARCH processes, E-(G)ARCH - exponential (G)ARCH, T-(G)ARCH - threshold (G)ARCH, $\mathscr{L}(\boldsymbol{\theta})$ - log-likelihood, AIC - Akaike Information Criterion $\left[\text{AIC}=\mathscr{L}(\boldsymbol{\theta}) - k\right]$, SBC - Schwarz Bayesian Information Criterion $\left[\text{SBC}=\mathscr{L}(\boldsymbol{\theta}) - 0.5k\ln(T)\right]$, k - number of parameters, T - number of observations. Bold - model with the best information statistic. * - the best statistics in the MS or (G)ARCH model class, respectively.

Table F.10: Information Criterion Tests: REXP(75%)-DAX30(25%) (1.1975-12.2004)

Model	$\mathscr{L}(\boldsymbol{\theta})$	**AIC**	**SBC**
GBM	946.5840	944.5840	940.6979*
AR(1)	947.3466	944.3466	938.5174
MS(1-2)-AR(0)	952.6548	947.6548*	937.9395*
MS(1-2)-AR(1)	953.3999	947.3999	935.7416
MS(2-1)-AR(0)	950.7026	945.7026	935.9873
MS(2-1)-AR(1)	952.4809	945.4809	931.8796
MS(2-2)-AR(0)	953.4877	947.4877	935.8294
MS(2-2)-AR(1)	954.8684	946.8684	931.3239
ARCH(1)-AR(0)	949.7645	946.7645	**940.9354***
ARCH(1)-AR(1)	949.8958	945.8958	938.1236
ARCH(2)-AR(0)	950.7991	946.7991	939.0269
ARCH(2)-AR(1)	950.6632	945.6632	935.9479
ARCH(3)-AR(0)	951.3268	946.3268	936.6116
ARCH(3)-AR(1)	951.0719	945.0719	933.4136
GARCH(1,1)-AR(0)	952.3619	948.3619	940.5897
GARCH(1,1)-AR(1)	952.0943	947.0943	937.3790
E-ARCH(1)-AR(0)	949.7166	945.7166	937.9444
E-ARCH(1)-AR(1)	950.6749	945.6749	935.9596
E-ARCH(2)-AR(0)	950.7251	944.7251	933.0668
E-ARCH(2)-AR(1)	951.7432	944.7432	931.1418
E-ARCH(3)-AR(0)	951.3144	943.3144	927.7700
E-ARCH(3)-AR(1)	952.3218	943.3218	925.8344
E-GARCH(1,1)-AR(0)	953.3516	948.3516	938.6364
E-GARCH(1,1)-AR(1)	952.8574	946.8574	935.1991
T-ARCH(1)-AR(0)	950.1006	946.1006	938.3284
T-ARCH(1)-AR(1)	950.7590	945.7590	936.0437
T-ARCH(2)-AR(0)	951.3480	946.3480	936.6327
T-ARCH(2)-AR(1)	951.6025	945.6025	933.9442
T-ARCH(3)-AR(0)	951.8891	945.8891	934.2307
T-ARCH(3)-AR(1)	951.9398	944.9398	931.3384
T-GARCH(1,1)-AR(0)	953.4637	**948.4637***	938.7484
T-GARCH(1,1)-AR(1)	952.9431	946.9431	935.2848

Note:
GBM - Geometric Brownian motion, AR(p) denotes an auto-regressive model of the p-th order, MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance, (G)ARCH(p,q) - (Generalised) Auto-regression Conditional Heteroscedasticity with q GARCH processes and p ARCH processes, E-(G)ARCH - exponential (G)ARCH, T-(G)ARCH - threshold (G)ARCH, $\mathscr{L}(\boldsymbol{\theta})$ - log-likelihood, AIC - Akaike Information Criterion $\left[\text{AIC}=\mathscr{L}(\boldsymbol{\theta}) - k\right]$, SBC - Schwarz Bayesian Information Criterion $\left[\text{SBC}=\mathscr{L}(\boldsymbol{\theta}) - 0.5k\ln(T)\right]$, k - number of parameters, T - number of observations. Bold - model with the best information statistic. * - the best statistics in the MS or (G)ARCH model class, respectively.

Table F.11: Information Criterion Tests: REXP(80%)-DAX30(20%) (1.1975-12.2004)

Model	$\mathscr{L}(\boldsymbol{\theta})$	**AIC**	**SBC**
GBM	994.4461	992.4461	988.5600*
AR(1)	995.7300	992.7300	986.9008
MS(1-2)-AR(0)	1000.0601	995.0601	985.3448*
MS(1-2)-AR(1)	1001.4135	995.4135	983.7552
MS(2-1)-AR(0)	997.2255	992.2255	982.5103
MS(2-1)-AR(1)	1001.5554	994.5554	980.9541
MS(2-2)-AR(0)	1000.7053	994.7053	983.0470
MS(2-2)-AR(1)	1003.5288	995.5288*	979.9844
ARCH(1)-AR(0)	997.9415	994.9415	**989.1124***
ARCH(1)-AR(1)	998.6077	994.6077	986.8355
ARCH(2)-AR(0)	998.3823	994.3823	986.6101
ARCH(2)-AR(1)	998.8935	993.8935	984.1783
ARCH(3)-AR(0)	998.7810	993.7810	984.0657
ARCH(3)-AR(1)	999.1178	993.1178	981.4595
GARCH(1,1)-AR(0)	999.9388	995.9388	988.1666
GARCH(1,1)-AR(1)	1000.1872	995.1872	985.4719
E-ARCH(1)-AR(0)	997.5094	993.5094	985.7372
E-ARCH(1)-AR(1)	999.1807	994.1807	984.4654
E-ARCH(2)-AR(0)	998.1307	992.1307	980.4724
E-ARCH(2)-AR(1)	999.8759	992.8759	979.2745
E-ARCH(3)-AR(0)	998.4948	990.4948	974.9504
E-ARCH(3)-AR(1)	1000.2132	991.2132	973.7257
E-GARCH(1,1)-AR(0)	1000.9413	**995.9413***	986.2260
E-GARCH(1,1)-AR(1)	1000.9072	994.9072	983.2489
T-ARCH(1)-AR(0)	998.1512	994.1512	986.3790
T-ARCH(1)-AR(1)	999.3781	994.3781	984.6628
T-ARCH(2)-AR(0)	998.6895	993.6895	983.9743
T-ARCH(2)-AR(1)	999.6659	993.6659	982.0076
T-ARCH(3)-AR(0)	999.1640	993.1640	981.5056
T-ARCH(3)-AR(1)	999.8805	992.8805	979.2791
T-GARCH(1,1)-AR(0)	1000.9211	995.9211	986.2058
T-GARCH(1,1)-AR(1)	1000.9223	994.9223	983.2640

Note:
GBM - Geometric Brownian motion, AR(p) denotes an auto-regressive model of the p-th order, MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance, (G)ARCH(p,q) - (Generalised) Auto-regression Conditional Heteroscedasticity with q GARCH processes and p ARCH processes, E-(G)ARCH - exponential (G)ARCH, T-(G)ARCH - threshold (G)ARCH, $\mathscr{L}(\boldsymbol{\theta})$ - log-likelihood, AIC - Akaike Information Criterion $\left[\text{AIC}=\mathscr{L}(\boldsymbol{\theta}) - k\right]$, SBC - Schwarz Bayesian Information Criterion $\left[\text{SBC}=\mathscr{L}(\boldsymbol{\theta}) - 0.5k\ln(T)\right]$, k - number of parameters, T - number of observations. Bold - model with the best information statistic. * - the best statistics in the MS or (G)ARCH model class, respectively.

Table F.12: Information Criterion Tests: REXP(90%)-DAX30(10%) (1.1975-12.2004)

Model	$\mathscr{L}(\boldsymbol{\theta})$	**AIC**	**SBC**
GBM	1082.8462	1080.8462	1076.9601
AR(1)	1086.3876	1083.3876	1077.5584
MS(1-2)-AR(0)	1092.3177	1087.3177	1077.6025
MS(1-2)-AR(1)	1096.5412	1090.5412*	1078.8828*
MS(2-1)-AR(0)	1087.6867	1082.6867	1072.9715
MS(2-1)-AR(1)	1095.8523	1088.8523	1075.2509
MS(2-2)-AR(0)	1092.7810	1086.7810	1075.1227
MS(2-2)-AR(1)	1097.6935	1089.6935	1074.1491
ARCH(1)-AR(0)	1090.0045	1087.0045	1081.1754
ARCH(1)-AR(1)	1093.7291	1089.7291	1081.9569
ARCH(2)-AR(0)	1090.4338	1086.4338	1078.6616
ARCH(2)-AR(1)	1094.3255	1089.3255	1079.6102
ARCH(3)-AR(0)	1091.7058	1086.7058	1076.9905
ARCH(3)-AR(1)	1094.8356	1088.8356	1077.1772
GARCH(1,1)-AR(0)	1094.8277	1090.8277	1083.0555
GARCH(1,1)-AR(1)	1099.2014	**1094.2014***	**1084.4861***
E-ARCH(1)-AR(0)	1088.7425	1084.7425	1076.9703
E-ARCH(1)-AR(1)	1094.0499	1089.0499	1079.3346
E-ARCH(2)-AR(0)	1090.4625	1084.4625	1072.8041
E-ARCH(2)-AR(1)	1095.9958	1088.9958	1075.3944
E-ARCH(3)-AR(0)	1091.0672	1083.0672	1067.5228
E-ARCH(3)-AR(1)	1096.1260	1087.1260	1069.6385
E-GARCH(1,1)-AR(0)	1095.9996	1090.9996	1081.2844
E-GARCH(1,1)-AR(1)	1099.4253	1093.4253	1081.7669
T-ARCH(1)-AR(0)	1090.1117	1086.1117	1078.3395
T-ARCH(1)-AR(1)	1094.5388	1089.5388	1079.8236
T-ARCH(2)-AR(0)	1090.5360	1085.5360	1075.8208
T-ARCH(2)-AR(1)	1094.9809	1088.9809	1077.3226
T-ARCH(3)-AR(0)	1091.9059	1085.9059	1074.2476
T-ARCH(3)-AR(1)	1095.5048	1088.5048	1074.9034
T-GARCH(1,1)-AR(0)	1095.0432	1090.0432	1080.3279
T-GARCH(1,1)-AR(1)	1099.4961	1093.4961	1081.8378

Note:
GBM - Geometric Brownian motion, AR(p) denotes an auto-regressive model of the p-th order, MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance, (G)ARCH(p,q) - (Generalised) Auto-regression Conditional Heteroscedasticity with q GARCH processes and p ARCH processes, E-(G)ARCH - exponential (G)ARCH, T-(G)ARCH - threshold (G)ARCH, $\mathscr{L}(\boldsymbol{\theta})$ - log-likelihood, AIC - Akaike Information Criterion $\left[\text{AIC}=\mathscr{L}(\boldsymbol{\theta}) - k\right]$, SBC - Schwarz Bayesian Information Criterion $\left[\text{SBC}=\mathscr{L}(\boldsymbol{\theta}) - 0.5k\ln(T)\right]$, k - number of parameters, T - number of observations. Bold - model with the best information statistic. * - the best statistics in the MS or (G)ARCH model class, respectively.

Table F.13: Information Criterion Tests: REXP (1.1975-12.2004)

Model	$\mathscr{L}(\boldsymbol{\theta})$	**AIC**	**SBC**
GBM	1113.4151	1111.4151	1107.5290
AR(1)	1121.0048	1118.0048	1112.1756
MS(1-2)-AR(0)	1129.5333	1124.5333	1114.8181
MS(1-2)-AR(1)	1138.8598	**1132.8598***	1121.2015*
MS(2-1)-AR(0)	1122.3571	1117.3571	1107.6418
MS(2-1)-AR(1)	1132.4950	1125.4950	1111.8936
MS(2-2)-AR(0)	1129.5800	1123.5800	1111.9217
MS(2-2)-AR(1)	1139.0877	1131.0877	1115.5433
ARCH(1)-AR(0)	1120.9730	1117.9730	1112.1439
ARCH(1)-AR(1)	1130.4496	1126.4496	1118.6774
ARCH(2)-AR(0)	1121.3755	1117.3755	1109.6033
ARCH(2)-AR(1)	1131.2224	1126.2224	1116.5071
ARCH(3)-AR(0)	1122.3499	1117.3499	1107.6346
ARCH(3)-AR(1)	1131.3569	1125.3569	1113.6986
GARCH(1,1)-AR(0)	1125.9038	1121.9038	1114.1316
GARCH(1,1)-AR(1)	1136.9318	1131.9318*	**1122.2166***
E-ARCH(1)-AR(0)	1120.7666	1116.7666	1108.9944
E-ARCH(1)-AR(1)	1130.2170	1125.2170	1115.5017
E-ARCH(2)-AR(0)	1121.4488	1115.4488	1103.7904
E-ARCH(2)-AR(1)	1131.3639	1124.3639	1110.7625
E-ARCH(3)-AR(0)	1123.1402	1115.1402	1099.5958
E-ARCH(3)-AR(1)	1131.7382	1122.7382	1105.2507
E-GARCH(1,1)-AR(0)	1126.8686	1121.8686	1112.1533
E-GARCH(1,1)-AR(1)	1137.4573	1131.4573	1119.7990
T-ARCH(1)-AR(0)	1121.3916	1117.3916	1109.6194
T-ARCH(1)-AR(1)	1132.3429	1127.3429	1117.6276
T-ARCH(2)-AR(0)	1121.9218	1116.9218	1107.2066
T-ARCH(2)-AR(1)	1133.0920	1127.0920	1115.4337
T-ARCH(3)-AR(0)	1122.8179	1116.8179	1105.1596
T-ARCH(3)-AR(1)	1133.1365	1126.1365	1112.5351
T-GARCH(1,1)-AR(0)	1126.5432	1121.5432	1111.8280
T-GARCH(1,1)-AR(1)	1137.7694	1131.7694	1120.1110

Note:
GBM - Geometric Brownian motion, AR(p) denotes an auto-regressive model of the p-th order, MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance, (G)ARCH(p,q) - (Generalised) Auto-regression Conditional Heteroscedasticity with q GARCH processes and p ARCH processes, E-(G)ARCH - exponential (G)ARCH, T-(G)ARCH - threshold (G)ARCH, $\mathscr{L}(\boldsymbol{\theta})$ - log-likelihood, AIC - Akaike Information Criterion $\left[\text{AIC}=\mathscr{L}(\boldsymbol{\theta})-k\right]$, SBC - Schwarz Bayesian Information Criterion $\left[\text{SBC}=\mathscr{L}(\boldsymbol{\theta})-0.5k\ln(T)\right]$, k - number of parameters, T - number of observations. Bold - model with the best information statistic. * - the best statistics in the MS or (G)ARCH model class, respectively.

Appendix G

Wald and likelihood ratio test (1.1975-12.2004)

Table G.1: Wald and LR tests: DAX30

	MS(1-2)		MS(2-1)		MS(2-2)	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
WT$\left(\mathrm{H}_0^{(2b)}\right)$	923.15***	1143.56***	3.74*	4.03**	937.36***	1017.73***
WT$\left(\mathrm{H}_0^{(3)}\right)$	—	—	88.25***	49.84***	0.84	1.10
WT$\left(\mathrm{H}_0^{(4a)}\right)$	32.63***	32.16***	—	—	31.66***	31.38***
WT$\left(\mathrm{H}_0^{(5)}\right)$	—	—	—	0.33	—	0.55
LR$\left(\mathrm{H}_0^{(2b)}\right)$	25.99***	26.52***	8.76***	10.80***	21.74***	18.64***
LR$\left(\mathrm{H}_0^{(4b)}\right)$	—	—	—	—	17.98***	16.35***
LR$\left(\mathrm{H}_0^{(6)}\right)$	—	—	—	—	1.35	2.61

Table G.2: Wald and LR tests: REXP(90%)-DAX30(90%)

	MS(1-2)		MS(2-1)		MS(2-2)	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
WT$\left(\mathrm{H}_0^{(2b)}\right)$	881.41***	1108.86***	3.66*	3.83*	900.15***	1022.84***
WT$\left(\mathrm{H}_0^{(3)}\right)$	—	—	83.02***	45.85***	0.78	1.08
WT$\left(\mathrm{H}_0^{(4a)}\right)$	31.96***	31.47***	—	—	31.18***	30.87***
WT$\left(\mathrm{H}_0^{(5)}\right)$	—	—	—	0.26	—	0.49
LR$\left(\mathrm{H}_0^{(2b)}\right)$	25.44***	25.90***	8.54***	10.52***	21.57***	18.45***
LR$\left(\mathrm{H}_0^{(4b)}\right)$	—	—	—	—	17.66***	16.00***
LR$\left(\mathrm{H}_0^{(6)}\right)$	—	—	—	—	1.21	2.41

Note:
MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance. WT - Wald test, LR - likelihood ratio test. *, **, *** - the null hypothesis can be rejected at the 10%, the 5% and the 1% confidence level, respectively.

$\mathrm{H}_0^{(2b)}$: $p_{11} = 1 - p_{22}$ — $\mathrm{H}_1^{(2b)}$: $p_{11} \neq 1 - p_{22}$

$\mathrm{H}_0^{(3)}$: $\mu_1 = \mu_2$, $\sigma_1^2 \neq \sigma_2^2$, p_{11}, p_{22} - not specified, — $\mathrm{H}_1^{(3)}$: $\mu_1 \neq \mu_2$

$\mathrm{H}_0^{(4a)}$: $\mu_1 \neq \mu_2$, $\sigma_1^2 = \sigma_2^2$, p_{11}, p_{22} - not specified, — $\mathrm{H}_1^{(4a)}$: $\sigma_1^2 \neq \sigma_2^2$

$\mathrm{H}_0^{(4b)}$: $\mu_1 \neq \mu_2$, $\phi_{1(1)} \neq \phi_{1(2)}$, $\sigma_1^2 = \sigma_2^2$, $p_{11} \neq 1 - p_{22}$, — $\mathrm{H}_1^{(4b)}$: $\sigma_1^2 \neq \sigma_2^2$

$\mathrm{H}_0^{(5)}$: $\phi_{1(1)} = \phi_{1(2)}$, p_{11}, p_{22} - not specified, — $\mathrm{H}_1^{(5)}$: $\phi_{1(1)} \neq \phi_{1(2)}$

$\mathrm{H}_0^{(6)}$: $\mu_1 = \mu_2$, $\phi_{1(1)} = \phi_{1(2)}$, $\sigma_1^2 \neq \sigma_2^2$, $p_{11} \neq 1 - p_{22}$, — $\mathrm{H}_1^{(6)}$: $\mu_1 \neq \mu_2$, $\phi_{1(1)} \neq \phi_{1(2)}$

Table G.3: Wald and LR tests: REXP(20%)-DAX30(80%)

	MS(1-2)		**MS(2-1)**		**MS(2-2)**	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
WT$\left(\mathrm{H}_0^{(2b)}\right)$	804.42***	1048.45***	3.60*	3.49*	848.57***	1000.91***
WT$\left(\mathrm{H}_0^{(3)}\right)$	—	—	77.30***	40.50***	0.72	1.05
WT$\left(\mathrm{H}_0^{(4a)}\right)$	30.96***	30.47***	—	—	30.29***	30.03***
WT$\left(\mathrm{H}_0^{(5)}\right)$	—	—	—	0.18	—	0.43
LR$\left(\mathrm{H}_0^{(2b)}\right)$	24.48***	24.89***	8.23***	10.09***	21.00***	17.90***
LR$\left(\mathrm{H}_0^{(4b)}\right)$	—	—	—	—	16.94***	15.29***
LR$\left(\mathrm{H}_0^{(6)}\right)$	—	—	—	—	1.09	2.20

Table G.4: Wald and LR tests: REXP(25%)-DAX30(75%)

	MS(1-2)		**MS(2-1)**		**MS(2-2)**	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
WT$\left(\mathrm{H}_0^{(2b)}\right)$	767.11***	1002.53***	3.54*	3.44*	817.24***	974.09***
WT$\left(\mathrm{H}_0^{(3)}\right)$	—	—	74.14***	40.68***	0.71	1.06
WT$\left(\mathrm{H}_0^{(4a)}\right)$	30.23***	29.81***	—	—	29.61***	29.40***
WT$\left(\mathrm{H}_0^{(5)}\right)$	—	—	—	0.17	—	0.31
LR$\left(\mathrm{H}_0^{(2b)}\right)$	23.81***	24.19***	8.03***	9.80***	20.54***	17.44***
LR$\left(\mathrm{H}_0^{(4b)}\right)$	—	—	—	—	16.39***	14.75***
LR$\left(\mathrm{H}_0^{(6)}\right)$	—	—	—	—	1.04	2.10

Note:
MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance. WT - Wald test, LR - likelihood ratio test. *, **, *** - the null hypothesis can be rejected at the 10%, the 5% and the 1% confidence level, respectively.

$\mathrm{H}_0^{(2b)}$: $p_{11} = 1 - p_{22}$ — $\mathrm{H}_1^{(2b)}$: $p_{11} \neq 1 - p_{22}$

$\mathrm{H}_0^{(3)}$: $\mu_1 = \mu_2$, $\sigma_1^2 \neq \sigma_2^2$, p_{11}, p_{22} - not specified, — $\mathrm{H}_1^{(3)}$: $\mu_1 \neq \mu_2$

$\mathrm{H}_0^{(4a)}$: $\mu_1 \neq \mu_2$, $\sigma_1^2 = \sigma_2^2$, p_{11}, p_{22} - not specified, — $\mathrm{H}_1^{(4a)}$: $\sigma_1^2 \neq \sigma_2^2$

$\mathrm{H}_0^{(4b)}$: $\mu_1 \neq \mu_2$, $\phi_{1(1)} \neq \phi_{1(2)}$, $\sigma_1^2 = \sigma_2^2$, $p_{11} \neq 1 - p_{22}$, — $\mathrm{H}_1^{(4b)}$: $\sigma_1^2 \neq \sigma_2^2$

$\mathrm{H}_0^{(5)}$: $\phi_{1(1)} = \phi_{1(2)}$, p_{11}, p_{22} - not specified, — $\mathrm{H}_1^{(5)}$: $\phi_{1(1)} \neq \phi_{1(2)}$

$\mathrm{H}_0^{(6)}$: $\mu_1 = \mu_2$, $\phi_{1(1)} = \phi_{1(2)}$, $\sigma_1^2 \neq \sigma_2^2$, $p_{11} \neq 1 - p_{22}$, — $\mathrm{H}_1^{(6)}$: $\mu_1 \neq \mu_2$, $\phi_{1(1)} \neq \phi_{1(2)}$

Table G.5: Wald and LR tests: REXP(30%)-DAX30(70%)

	MS(1-2)		**MS(2-1)**		**MS(2-2)**	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
WT$\left(\mathrm{H}_0^{(2b)}\right)$	715.00***	944.61***	3.43*	3.46*	758.47***	934.88***
WT$\left(\mathrm{H}_0^{(3)}\right)$	—	—	70.85***	36.96***	0.69	1.05
WT$\left(\mathrm{H}_0^{(4a)}\right)$	29.34***	29.00***	—	—	28.78***	28.66***
WT$\left(\mathrm{H}_0^{(5)}\right)$	—	—	—	0.12	—	0.43
LR$\left(\mathrm{H}_0^{(2b)}\right)$	22.96***	23.32***	7.79***	9.44***	19.91***	16.83***
LR$\left(\mathrm{H}_0^{(4b)}\right)$	—	—	—	—	15.67***	14.07***
LR$\left(\mathrm{H}_0^{(6)}\right)$	—	—	—	—	1.00	2.00

Table G.6: Wald and LR tests: REXP(40%)-DAX30(60%)

	MS(1-2)		**MS(2-1)**		**MS(2-2)**	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
WT$\left(\mathrm{H}_0^{(2b)}\right)$	555.66***	765.58***	3.16*	3.20*	593.23***	797.03***
WT$\left(\mathrm{H}_0^{(3)}\right)$	—	—	63.29***	35.14***	0.68	1.07
WT$\left(\mathrm{H}_0^{(4a)}\right)$	26.87***	26.69***	—	—	26.23***	26.49***
WT$\left(\mathrm{H}_0^{(5)}\right)$	—	—	—	0.0496	—	0.25
LR$\left(\mathrm{H}_0^{(2b)}\right)$	20.64***	20.96***	7.16***	8.47***	18.03***	15.06***
LR$\left(\mathrm{H}_0^{(4b)}\right)$	—	—	—	—	13.62***	12.14***
LR$\left(\mathrm{H}_0^{(6)}\right)$	—	—	—	—	0.93	1.80

Note:
MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance. WT - Wald test, LR - likelihood ratio test. *, **, *** - the null hypothesis can be rejected at the 10%, the 5% and the 1% confidence level, respectively.

$\mathrm{H}_0^{(2b)}$: $p_{11} = 1 - p_{22}$ | $\mathrm{H}_1^{(2b)}$: $p_{11} \neq 1 - p_{22}$
$\mathrm{H}_0^{(3)}$: $\mu_1 = \mu_2$, $\sigma_1^2 \neq \sigma_2^2$, p_{11}, p_{22} - not specified, | $\mathrm{H}_1^{(3)}$: $\mu_1 \neq \mu_2$
$\mathrm{H}_0^{(4a)}$: $\mu_1 \neq \mu_2$, $\sigma_1^2 = \sigma_2^2$, p_{11}, p_{22} - not specified, | $\mathrm{H}_1^{(4a)}$: $\sigma_1^2 \neq \sigma_2^2$
$\mathrm{H}_0^{(4b)}$: $\mu_1 \neq \mu_2$, $\phi_{1(1)} \neq \phi_{1(2)}$, $\sigma_1^2 = \sigma_2^2$, $p_{11} \neq 1 - p_{22}$, | $\mathrm{H}_1^{(4b)}$: $\sigma_1^2 \neq \sigma_2^2$
$\mathrm{H}_0^{(5)}$: $\phi_{1(1)} = \phi_{1(2)}$, p_{11}, p_{22} - not specified, | $\mathrm{H}_1^{(5)}$: $\phi_{1(1)} \neq \phi_{1(2)}$
$\mathrm{H}_0^{(6)}$: $\mu_1 = \mu_2$, $\phi_{1(1)} = \phi_{1(2)}$, $\sigma_1^2 \neq \sigma_2^2$, $p_{11} \neq 1 - p_{22}$, | $\mathrm{H}_1^{(6)}$: $\mu_1 \neq \mu_2$, $\phi_{1(1)} \neq \phi_{1(2)}$

Table G.7: Wald and LR tests: REXP(50%)-DAX30(50%)

	MS(1-2)		**MS(2-1)**		**MS(2-2)**	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
WT$\left(\mathrm{H}_0^{(2b)}\right)$	289.67***	454.82***	2.80*	3.00*	324.88***	516.33***
WT$\left(\mathrm{H}_0^{(3)}\right)$	—	—	53.97***	35.23***	0.69	1.13
WT$\left(\mathrm{H}_0^{(4a)}\right)$	21.53***	22.75***	—	—	21.30***	22.78***
WT$\left(\mathrm{H}_0^{(5)}\right)$	—	—	—	0.0027	—	0.0775
LR$\left(\mathrm{H}_0^{(2b)}\right)$	17.19***	17.48***	6.23**	7.03***	15.07***	12.32***
LR$\left(\mathrm{H}_0^{(4b)}\right)$	—	—	—	—	10.53***	9.24***
LR$\left(\mathrm{H}_0^{(6)}\right)$	—	—	—	—	0.90	1.63

Table G.8: Wald and LR tests: REXP(60%)-DAX30(40%)

	MS(1-2)		**MS(2-1)**		**MS(2-2)**	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
WT$\left(\mathrm{H}_0^{(2b)}\right)$	134.64***	147.64***	2.25	2.38	68.33***	124.75***
WT$\left(\mathrm{H}_0^{(3)}\right)$	—	—	41.14***	29.60***	0.55	1.19
WT$\left(\mathrm{H}_0^{(4a)}\right)$	16.13***	15.94***	—	—	10.39***	14.08***
WT$\left(\mathrm{H}_0^{(5)}\right)$	—	—	—	0.0121	—	0.0022
LR$\left(\mathrm{H}_0^{(2b)}\right)$	12.88***	12.84***	4.87**	5.00**	11.17***	8.55***
LR$\left(\mathrm{H}_0^{(4b)}\right)$	—	—	—	—	6.74***	5.66**
LR$\left(\mathrm{H}_0^{(6)}\right)$	—	—	—	—	0.83	1.62

Note:
MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance. WT - Wald test, LR - likelihood ratio test. *, **, *** - the null hypothesis can be rejected at the 10%, the 5% and the 1% confidence level, respectively.

$\mathrm{H}_0^{(2b)}$: $p_{11} = 1 - p_{22}$ — $\mathrm{H}_1^{(2b)}$: $p_{11} \neq 1 - p_{22}$
$\mathrm{H}_0^{(3)}$: $\mu_1 = \mu_2$, $\sigma_1^2 \neq \sigma_2^2$, p_{11}, p_{22} - not specified, — $\mathrm{H}_1^{(3)}$: $\mu_1 \neq \mu_2$
$\mathrm{H}_0^{(4a)}$: $\mu_1 \neq \mu_2$, $\sigma_1^2 = \sigma_2^2$, p_{11}, p_{22} - not specified, — $\mathrm{H}_1^{(4a)}$: $\sigma_1^2 \neq \sigma_2^2$
$\mathrm{H}_0^{(4b)}$: $\mu_1 \neq \mu_2$, $\phi_{1(1)} \neq \phi_{1(2)}$, $\sigma_1^2 = \sigma_2^2$, $p_{11} \neq 1 - p_{22}$, — $\mathrm{H}_1^{(4b)}$: $\sigma_1^2 \neq \sigma_2^2$
$\mathrm{H}_0^{(5)}$: $\phi_{1(1)} = \phi_{1(2)}$, p_{11}, p_{22} - not specified, — $\mathrm{H}_1^{(5)}$: $\phi_{1(1)} \neq \phi_{1(2)}$
$\mathrm{H}_0^{(6)}$: $\mu_1 = \mu_2$, $\phi_{1(1)} = \phi_{1(2)}$, $\sigma_1^2 \neq \sigma_2^2$, $p_{11} \neq 1 - p_{22}$, — $\mathrm{H}_1^{(6)}$: $\mu_1 \neq \mu_2$, $\phi_{1(1)} \neq \phi_{1(2)}$

Table G.9: Wald and LR tests: REXP(70%)-DAX30(30%)

	MS(1-2)		MS(2-1)		MS(2-2)	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
WT$\left(\mathrm{H}_0^{(2b)}\right)$	110.34***	124.12***	1.39	1.22	127.97***	108.02***
WT$\left(\mathrm{H}_0^{(3)}\right)$	—	—	23.34***	19.02***	1.18	1.69
WT$\left(\mathrm{H}_0^{(4a)}\right)$	16.98***	15.15***	—	—	16.88***	15.09***
WT$\left(\mathrm{H}_0^{(5)}\right)$	—	—	—	0.11	—	0.10
LR$\left(\mathrm{H}_0^{(2b)}\right)$	9.28***	9.00***	3.03*	2.40	6.14**	6.12**
LR$\left(\mathrm{H}_0^{(4b)}\right)$	—	—	—	—	4.82**	4.81**
LR$\left(\mathrm{H}_0^{(6)}\right)$	—	—	—	—	1.38	2.89

Table G.10: Wald and LR tests: REXP(75%)-DAX30(25%)

	MS(1-2)		MS(2-1)		MS(2-2)	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
WT$\left(\mathrm{H}_0^{(2b)}\right)$	122.98***	134.66***	0.91	0.65	120.31***	140.44***
WT$\left(\mathrm{H}_0^{(3)}\right)$	—	—	14.51***	4.45**	1.37	2.88*
WT$\left(\mathrm{H}_0^{(4a)}\right)$	80.36***	81.78***	—	—	90.57***	90.56***
WT$\left(\mathrm{H}_0^{(5)}\right)$	—	—	—	19.61***	—	1.55
LR$\left(\mathrm{H}_0^{(2b)}\right)$	8.64***	9.31***	1.88	0.79	2.63	5.38**
LR$\left(\mathrm{H}_0^{(4b)}\right)$	—	—	—	—	5.57**	4.77**
LR$\left(\mathrm{H}_0^{(6)}\right)$	—	—	—	—	1.67	2.94

Note:
MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance. WT - Wald test, LR - likelihood ratio test. *, **, *** - the null hypothesis can be rejected at the 10%, the 5% and the 1% confidence level, respectively.

$\mathrm{H}_0^{(2b)}$: $p_{11} = 1 - p_{22}$	$\mathrm{H}_1^{(2b)}$: $p_{11} \neq 1 - p_{22}$
$\mathrm{H}_0^{(3)}$: $\mu_1 = \mu_2$, $\sigma_1^2 \neq \sigma_2^2$, p_{11}, p_{22} - not specified,	$\mathrm{H}_1^{(3)}$: $\mu_1 \neq \mu_2$
$\mathrm{H}_0^{(4a)}$: $\mu_1 \neq \mu_2$, $\sigma_1^2 = \sigma_2^2$, p_{11}, p_{22} - not specified,	$\mathrm{H}_1^{(4a)}$: $\sigma_1^2 \neq \sigma_2^2$
$\mathrm{H}_0^{(4b)}$: $\mu_1 \neq \mu_2$, $\phi_{1(1)} \neq \phi_{1(2)}$, $\sigma_1^2 = \sigma_2^2$, $p_{11} \neq 1 - p_{22}$,	$\mathrm{H}_1^{(4b)}$: $\sigma_1^2 \neq \sigma_2^2$
$\mathrm{H}_0^{(5)}$: $\phi_{1(1)} = \phi_{1(2)}$, p_{11}, p_{22} - not specified,	$\mathrm{H}_1^{(5)}$: $\phi_{1(1)} \neq \phi_{1(2)}$
$\mathrm{H}_0^{(6)}$: $\mu_1 = \mu_2$, $\phi_{1(1)} = \phi_{1(2)}$, $\sigma_1^2 \neq \sigma_2^2$, $p_{11} \neq 1 - p_{22}$,	$\mathrm{H}_1^{(6)}$: $\mu_1 \neq \mu_2$, $\phi_{1(1)} \neq \phi_{1(2)}$

Table G.11: Wald and LR tests: REXP(80%)-DAX30(20%)

	MS(1-2)		**MS(2-1)**		**MS(2-2)**	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
WT$\left(H_0^{(2b)}\right)$	199.66***	227.24***	0.51	2.14	154.36***	153.59***
WT$\left(H_0^{(3)}\right)$	—	—	9.79***	5.51**	0.99	3.63*
WT$\left(H_0^{(4a)}\right)$	76.62***	78.04***	—	—	79.55***	82.40***
WT$\left(H_0^{(5)}\right)$	—	—	—	19.79***	—	4.07
LR$\left(H_0^{(2b)}\right)$	67.10***	10.14***	0.90	1.76	3.73*	5.03***
LR$\left(H_0^{(4b)}\right)$	—	—	—	—	6.96***	3.95***
LR$\left(H_0^{(6)}\right)$	—	—	—	—	1.29	4.23

Table G.12: Wald and LR tests: REXP(90%)-DAX30(10%)

	MS(1-2)		**MS(2-1)**		**MS(2-2)**	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
WT$\left(H_0^{(2b)}\right)$	321.97***	615.70***	0.90	2.77*	207.86***	23.07***
WT$\left(H_0^{(3)}\right)$	—	—	14.21***	12.14***	0.60	2.30
WT$\left(H_0^{(4a)}\right)$	8.28***	7.83***	—	—	6.54**	6.24**
WT$\left(H_0^{(5)}\right)$	—	—	—	37.70***	—	5.69**
LR$\left(H_0^{(2b)}\right)$	13.69***	16.68***	1.23	2.50	7.20***	6.11**
LR$\left(H_0^{(4b)}\right)$	—	—	—	—	10.19***	3.68*
LR$\left(H_0^{(6)}\right)$	—	—	—	—	0.93	2.30

Note:
MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance. WT - Wald test, LR - likelihood ratio test. *, **, *** - the null hypothesis can be rejected at the 10%, the 5% and the 1% confidence level, respectively.

$H_0^{(2b)}$: $p_{11} = 1 - p_{22}$ — $H_1^{(2b)}$: $p_{11} \neq 1 - p_{22}$

$H_0^{(3)}$: $\mu_1 = \mu_2$, $\sigma_1^2 \neq \sigma_2^2$, p_{11}, p_{22} - not specified, — $H_1^{(3)}$: $\mu_1 \neq \mu_2$

$H_0^{(4a)}$: $\mu_1 \neq \mu_2$, $\sigma_1^2 = \sigma_2^2$, p_{11}, p_{22} - not specified, — $H_1^{(4a)}$: $\sigma_1^2 \neq \sigma_2^2$

$H_0^{(4b)}$: $\mu_1 \neq \mu_2$, $\phi_{1(1)} \neq \phi_{1(2)}$, $\sigma_1^2 = \sigma_2^2$, $p_{11} \neq 1 - p_{22}$, — $H_1^{(4b)}$: $\sigma_1^2 \neq \sigma_2^2$

$H_0^{(5)}$: $\phi_{1(1)} = \phi_{1(2)}$, p_{11}, p_{22} - not specified, — $H_1^{(5)}$: $\phi_{1(1)} \neq \phi_{1(2)}$

$H_0^{(6)}$: $\mu_1 = \mu_2$, $\phi_{1(1)} = \phi_{1(2)}$, $\sigma_1^2 \neq \sigma_2^2$, $p_{11} \neq 1 - p_{22}$, — $H_1^{(6)}$: $\mu_1 \neq \mu_2$, $\phi_{1(1)} \neq \phi_{1(2)}$

Table G.13: Wald and LR tests: REXP

	MS(1-2)		MS(2-1)		MS(2-2)	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
WT$\left(\mathrm{H}_0^{(2b)}\right)$	436.51***	499.71***	1.09	2.51	364.78***	22.38***
WT$\left(\mathrm{H}_0^{(3)}\right)$	—	—	13.36***	11.46***	0.08	2.98*
WT$\left(\mathrm{H}_0^{(4a)}\right)$	8.65***	8.44***	—	—	8.32***	3.08*
WT$\left(\mathrm{H}_0^{(5)}\right)$	—	—	—	33.80***	—	6.14**
LR$\left(\mathrm{H}_0^{(2b)}\right)$	19.63***	22.53***	2.20	1.75	10.80***	8.66***
LR$\left(\mathrm{H}_0^{(4b)}\right)$	—	—	—	—	14.45***	13.19***
LR$\left(\mathrm{H}_0^{(6)}\right)$	—	—	—	—	0.09	0.46

Note:
MS(m-s) denotes a Markov switching model with m mean equations and s regimes for the variance. WT - Wald test, LR - likelihood ratio test. *, **, *** - the null hypothesis can be rejected at the 10%, the 5% and the 1% confidence level, respectively.

$\mathrm{H}_0^{(2b)}$: $p_{11} = 1 - p_{22}$ — $\mathrm{H}_1^{(2b)}$: $p_{11} \neq 1 - p_{22}$

$\mathrm{H}_0^{(3)}$: $\mu_1 = \mu_2$, $\sigma_1^2 \neq \sigma_2^2$, p_{11}, p_{22} - not specified, — $\mathrm{H}_1^{(3)}$: $\mu_1 \neq \mu_2$

$\mathrm{H}_0^{(4a)}$: $\mu_1 \neq \mu_2$, $\sigma_1^2 = \sigma_2^2$, p_{11}, p_{22} - not specified, — $\mathrm{H}_1^{(4a)}$: $\sigma_1^2 \neq \sigma_2^2$

$\mathrm{H}_0^{(4b)}$: $\mu_1 \neq \mu_2$, $\phi_{1(1)} \neq \phi_{1(2)}$, $\sigma_1^2 = \sigma_2^2$, $p_{11} \neq 1 - p_{22}$, — $\mathrm{H}_1^{(4b)}$: $\sigma_1^2 \neq \sigma_2^2$

$\mathrm{H}_0^{(5)}$: $\phi_{1(1)} = \phi_{1(2)}$, p_{11}, p_{22} - not specified, — $\mathrm{H}_1^{(5)}$: $\phi_{1(1)} \neq \phi_{1(2)}$

$\mathrm{H}_0^{(6)}$: $\mu_1 = \mu_2$, $\phi_{1(1)} = \phi_{1(2)}$, $\sigma_1^2 \neq \sigma_2^2$, $p_{11} \neq 1 - p_{22}$, — $\mathrm{H}_1^{(6)}$: $\mu_1 \neq \mu_2$, $\phi_{1(1)} \neq \phi_{1(2)}$

Appendix H

Newey-Tauchen-White test (1.1975-12.2004)

Table H.1: Newey-Tauchen-White test: DAX30

	MS(1-2)		**MS(2-1)**		**MS(2-2)**	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
$H_0^{(7)}$	0.0385***	0.1245***	2.6824***	0.7404***	0.5320***	0.4048***
$H_0^{(8)}$	0.2494***	0.2691***	1.2467***	1.4783***	0.0802***	0.0682***
$H_0^{(9)}$	1.5982***	3.5223***	0.0001***	0.0004***	2.6309***	3.9217***

Table H.2: Newey-Tauchen-White test: REXP 10% DAX30 90%

	MS(1-2)		**MS(2-1)**		**MS(2-2)**	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
$H_0^{(7)}$	0.0506***	0.0908***	2.5655***	0.8089***	0.3944***	0.3868***
$H_0^{(8)}$	0.1238***	0.1472***	1.0878***	1.3169***	0.0436***	0.0485***
$H_0^{(9)}$	1.7578***	3.9651***	0.0001***	0.0003***	3.2040***	4.8181

Table H.3: Newey-Tauchen-White test: REXP 20% DAX30 80%

	MS(1-2)		**MS(2-1)**		**MS(2-2)**	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
$H_0^{(7)}$	0.0589***	0.0702***	2.0102***	1.0355***	0.5587***	0.7242***
$H_0^{(8)}$	0.0673***	0.0694***	1.3934***	1.1404***	0.0799***	0.1608***
$H_0^{(9)}$	1.7616***	4.1819***	0.0000***	0.0002***	2.6404***	4.2862***

Table H.4: Newey-Tauchen-White test: REXP 25% DAX30 75%

	MS(1-2)		**MS(2-1)**		**MS(2-2)**	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
$H_0^{(7)}$	0.1129***	0.0263***	3.1287***	1.3197***	0.3398***	0.5532***
$H_0^{(8)}$	0.1331***	0.0552***	0.6830***	0.9238***	0.3896***	0.5991***
$H_0^{(9)}$	1.5622***	3.8840***	0.0001***	0.0002***	2.8789***	4.7363

Note:
MS(m-s)-AR(p) denotes a Markov switching model with m mean equations, s regimes for the variance and p auto-regression lags in each mean equation. *, **, *** - the null hypothesis can be rejected at the 10%, the 5% and the 1% confidence level, respectively.

$H_0^{(7)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{\mu}_j)\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{\mu}_i)\right] = 0$ $H_1^{(7)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{\mu}_j)\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{\mu}_i)\right] \neq 0$ (i,j=1,...,K)

$H_0^{(8)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{\sigma}_j^2)\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{\sigma}_i^2)\right] = 0$ $H_1^{(8)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{\sigma}_j^2)\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{\sigma}_i^2)\right] \neq 0$ (i,j=1,...,K)

$H_0^{(9)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{p}_{jj})\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{p}_{ii})\right] = 0$ $H_1^{(9)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{p}_{jj})\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{p}_{ii})\right] \neq 0$ (i,j=1,...,K)

Table H.5: Newey-Tauchen-White test: REXP 30% DAX30 70%

	MS(1-2)		MS(2-1)		MS(2-2)	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
$H_0^{(7)}$	0.1140***	0.0236***	3.7029***	1.3446***	1.3728***	2.0674***
$H_0^{(8)}$	0.5985***	0.8335***	0.6045***	0.6354***	2.5051***	2.2552***
$H_0^{(9)}$	1.9194***	4.3325***	0.0000***	0.0001***	3.1280***	4.7664

Table H.6: Newey-Tauchen-White test: REXP 40% DAX30 60%

	MS(1-2)		MS(2-1)		MS(2-2)	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
$H_0^{(7)}$	0.2412***	0.0000***	3.8495***	2.9442***	1.1010***	1.8956***
$H_0^{(8)}$	3.3005***	3.0942***	1.5099***	1.2272***	2.7466***	2.2147***
$H_0^{(9)}$	2.4685***	4.7443	0.0000***	0.0000***	1.5288***	2.8541***

Table H.7: Newey-Tauchen-White test: REXP 50% DAX30 50%

	MS(1-2)		MS(2-1)		MS(2-2)	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
$H_0^{(7)}$	0.5671***	0.0863***	2.1997***	1.7095***	0.7707***	1.4316***
$H_0^{(8)}$	3.5214***	3.0547***	1.7089***	0.9912***	3.6182***	2.9731***
$H_0^{(9)}$	2.0993***	3.6666***	0.0000***	0.0000***	1.0599***	1.9203***

Table H.8: Newey-Tauchen-White test: REXP 60% DAX30 40%

	MS(1-2)		MS(2-1)		MS(2-2)	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
$H_0^{(7)}$	1.5262***	0.5206***	0.4421***	0.1869***	1.8566***	1.8112***
$H_0^{(8)}$	3.5318***	3.1192***	1.8435***	0.8816***	3.3500***	2.6457***
$H_0^{(9)}$	2.7483***	3.5277***	0.0000***	0.0000***	0.8615***	0.6491***

Note:
MS(m-s)-AR(p) denotes a Markov switching model with m mean equations, s regimes for the variance and p auto-regression lags in each mean equation. *, **, *** - the null hypothesis can be rejected at the 10%, the 5% and the 1% confidence level, respectively.

$H_0^{(7)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{\mu}_j)\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{\mu}_i)\right] = 0$ $H_1^{(7)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{\mu}_j)\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{\mu}_i)\right] \neq 0$ (i,j=1,...,K)
$H_0^{(8)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{\sigma}_j^2)\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{\sigma}_i^2)\right] = 0$ $H_1^{(8)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{\sigma}_j^2)\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{\sigma}_i^2)\right] \neq 0$ (i,j=1,...,K)
$H_0^{(9)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{p}_{jj})\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{p}_{ii})\right] = 0$ $H_1^{(9)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{p}_{jj})\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{p}_{ii})\right] \neq 0$ (i,j=1,...,K)

Table H.9: Newey-Tauchen-White test: REXP 70% DAX30 30%

	MS(1-2)		MS(2-1)		MS(2-2)	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
$H_0^{(7)}$	5.2262*	3.1539	1.3988***	1.6960***	2.3853***	1.6922***
$H_0^{(8)}$	2.5963***	2.2208***	1.2558***	0.1098***	3.3305***	2.6419***
$H_0^{(9)}$	2.6925***	3.0857***	0.0000***	0.0000***	0.8577***	0.2747***

Table H.10: Newey-Tauchen-White test: REXP 75% DAX30 25%

	MS(1-2)		MS(2-1)		MS(2-2)	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
$H_0^{(7)}$	2.3163***	1.7864***	6.9399*	1.8300***	0.7654***	0.3508***
$H_0^{(8)}$	2.9348***	2.1521***	1.2195***	2.6644***	2.8147***	2.3179***
$H_0^{(9)}$	2.1356***	2.2766***	0.0000***	0.0000***	0.5837***	1.2836***

Table H.11: Newey-Tauchen-White test: REXP 80% DAX30 20%

	MS(1-2)		MS(2-1)		MS(2-2)	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
$H_0^{(7)}$	3.9834*	3.2329	8.0251*	2.4802***	2.0132***	0.7650***
$H_0^{(8)}$	2.5076***	1.5242***	0.7389***	3.5001	3.2705***	2.0356***
$H_0^{(9)}$	1.1283***	1.1823***	0.0000***	0.0000***	3.2306***	0.9423***

Table H.12: Newey-Tauchen-White test: REXP 90% DAX30 10%

	MS(1-2)		MS(2-1)		MS(2-2)	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
$H_0^{(7)}$	8.1041**	2.0137***	20.4211**	4.0373***	6.3384*	7.3723*
$H_0^{(8)}$	2.8795***	3.7998***	1.1200***	3.5334	4.6619	7.7374*
$H_0^{(9)}$	4.9265	5.0124	0.0000***	0.0000***	1.6992***	2.7630***

Note:
MS(m-s)-AR(p) denotes a Markov switching model with m mean equations, s regimes for the variance and p auto-regression lags in each mean equation. *, **, *** - the null hypothesis can be rejected at the 10%, the 5% and the 1% confidence level, respectively.

$H_0^{(7)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{\mu}_j)\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{\mu}_i)\right] = 0$ $H_1^{(7)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{\mu}_j)\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{\mu}_i)\right] \neq 0$ (i,j=1,...,K)

$H_0^{(8)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{\sigma}_j^2)\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{\sigma}_i^2)\right] = 0$ $H_1^{(8)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{\sigma}_j^2)\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{\sigma}_i^2)\right] \neq 0$ (i,j=1,...,K)

$H_0^{(9)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{p}_{jj})\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{p}_{ii})\right] = 0$ $H_1^{(9)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{p}_{jj})\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{p}_{ii})\right] \neq 0$ (i,j=1,...,K)

Table H.13: Newey-Tauchen-White test: REXP

	MS(1-2)		**MS(2-1)**		**MS(2-2)**	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
$H_0^{(7)}$	20.1919**	1.7086***	25.3082**	10.0650**	13.5856**	12.1780**
$H_0^{(8)}$	3.1807***	1.9153***	0.6140***	7.4031**	3.9949***	3.4779***
$H_0^{(9)}$	2.6856***	2.6608***	0.0000***	0.0000***	2.7332***	1.9048***

Note:
MS(m-s)-AR(p) denotes a Markov switching model with m mean equations, s regimes for the variance and p auto-regression lags in each mean equation. *, **, *** - the null hypothesis can be rejected at the 10%, the 5% and the 1% confidence level, respectively.

$H_0^{(7)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{\mu}_j)\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{\mu}_i)\right] = 0$ $H_1^{(7)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{\mu}_j)\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{\mu}_i)\right] \neq 0$ (i,j=1,...,K)
$H_0^{(8)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{\sigma}_j^2)\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{\sigma}_i^2)\right] = 0$ $H_1^{(8)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{\sigma}_j^2)\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{\sigma}_i^2)\right] \neq 0$ (i,j=1,...,K)
$H_0^{(9)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{p}_{jj})\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{p}_{ii})\right] = 0$ $H_1^{(9)}$: $\mathbb{E}\left[\boldsymbol{h}_t^c(\widehat{p}_{jj})\,\middle|\boldsymbol{h}_{t-1}^c(\widehat{p}_{ii})\right] \neq 0$ (i,j=1,...,K)

Appendix I

Lagrange multiplier test (1.1975-12.2004)

Table I.1: Lagrange multiplier test: DAX30

	MS(1-2)		MS(2-1)		MS(2-2)	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
$H_0^{(10)}$	0.5134***	3.1364	1.5889***	32.0739**	0.0469***	7.3589**
$H_0^{(11)}$	1.3375***	3.3922	2.4954***	32.6655**	0.5255***	7.4295**

Table I.2: Lagrange multiplier test: REXP 10% DAX30 90%

	MS(1-2)		MS(2-1)		MS(2-2)	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
$H_0^{(10)}$	0.6146***	2.9073	1.4475***	31.6912**	0.3752***	7.4362**
$H_0^{(11)}$	1.4965***	3.1855	2.1163***	32.0458**	0.8734***	7.4735**

Table I.3: Lagrange multiplier test: REXP 20% DAX30 80%

	MS(1-2)		MS(2-1)		MS(2-2)	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
$H_0^{(10)}$	0.7113***	2.9885	1.2838***	27.5140**	0.3679***	6.9847**
$H_0^{(11)}$	1.5682***	3.1392	3.2836	28.1928**	1.2182***	7.0760**

Table I.4: Lagrange multiplier test: REXP 25% DAX30 75%

	MS(1-2)		MS(2-1)		MS(2-2)	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
$H_0^{(10)}$	0.5003***	2.5266***	2.9217	26.8444**	1.7288***	8.0665**
$H_0^{(11)}$	1.2676***	2.7711	4.2705*	27.6564**	3.8086	8.8296**

Table I.5: Lagrange multiplier test: REXP 30% DAX30 70%

	MS(1-2)		MS(2-1)		MS(2-2)	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
$H_0^{(10)}$	1.7347***	4.2598*	3.7162	25.5256**	2.1742***	8.0737**
$H_0^{(11)}$	4.5544*	6.3333*	4.3131*	26.2148**	4.2882*	8.7864**

Note:
MS(m-s)-AR(p) denotes a Markov switching model with m mean equations, s regimes for the variance and p auto-regression lags in each mean equation. *, **, *** - the null hypothesis can be rejected at the 10%, the 5% and the 1% confidence level, respectively.

$H_0^{(10)}$: $\rho = 0$ $\quad H_1^{(10)}$: $(y_t|\mathscr{X}_t, z_t, z_{t-1}; \boldsymbol{\theta}, \rho) \sim N(\{\boldsymbol{x}_t'\boldsymbol{\beta}_{z_t} + \rho(y_{t-1} - \boldsymbol{x}_{t-1}'\boldsymbol{\beta}_{z_{t-1}})\}, \sigma^2_{z_t})$

$H_0^{(11)}$: $\xi = 0$ $\quad H_1^{(11)}$: $(y_t|\mathscr{X}_t, z_t, z_{t-1}; \boldsymbol{\theta}, \xi) \sim N\left(\boldsymbol{x}_t'\boldsymbol{\beta}_{z_t}, \sigma^2_{z_t}\left[1 + \frac{\xi(y_{t-1} - \boldsymbol{x}_{t-1}'\boldsymbol{\beta}_{z_{t-1}})^2}{\sigma^2_{z_{t-1}}}\right]\right)$

Table I.6: Lagrange multiplier test: REXP 40% DAX30 60%

	MS(1-2)		MS(2-1)		MS(2-2)	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
$H_0^{(10)}$	1.7930***	4.4297*	4.1814*	22.8426**	1.4301***	6.6310*
$H_0^{(11)}$	4.4790*	6.1316*	4.6978*	23.3392**	2.9664	7.0271**

Table I.7: Lagrange multiplier test: REXP 50% DAX30 50%

	MS(1-2)		MS(2-1)		MS(2-2)	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
$H_0^{(10)}$	1.4459***	4.0189*	3.4487	20.4251**	0.9632***	6.0025*
$H_0^{(11)}$	4.3124*	5.7061*	4.1420*	20.7475**	2.5985***	6.5242*

Table I.8: Lagrange multiplier test: REXP 60% DAX30 40%

	MS(1-2)		MS(2-1)		MS(2-2)	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
$H_0^{(10)}$	1.3026***	3.3304	2.5108***	14.4731**	0.8554***	5.9498*
$H_0^{(11)}$	4.2470*	5.0723*	3.9228*	14.9722**	2.0995***	6.1672*

Table I.9: Lagrange multiplier test: REXP 70% DAX30 30%

	MS(1-2)		MS(2-1)		MS(2-2)	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
$H_0^{(10)}$	2.2077***	3.3280	1.3547***	6.8935**	1.4149***	8.9650**
$H_0^{(11)}$	4.8967*	4.9808*	2.6618***	7.0226**	3.1584	8.3499**

Table I.10: Lagrange multiplier test: REXP 75% DAX30 25%

	MS(1-2)		MS(2-1)		MS(2-2)	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
$H_0^{(10)}$	2.2907***	4.6571*	1.0178***	32.6253**	2.2687***	2.9427
$H_0^{(11)}$	4.8184*	8.0783**	2.3351***	32.1797**	5.0304*	5.9998*

Note:
MS(m-s)-AR(p) denotes a Markov switching model with m mean equations, s regimes for the variance and p auto-regression lags in each mean equation. *, **, *** - the null hypothesis can be rejected at the 10%, the 5% and the 1% confidence level, respectively.

$H_0^{(10)}$: $\rho = 0$ $\quad H_1^{(10)}$: $(y_t|\mathscr{X}_t, z_t, z_{t-1}; \boldsymbol{\theta}, \rho) \sim N(\{\boldsymbol{x}_t'\boldsymbol{\beta}_{z_t} + \rho(y_{t-1} - \boldsymbol{x}_{t-1}'\boldsymbol{\beta}_{z_{t-1}})\}, \sigma^2_{z_t})$

$H_0^{(11)}$: $\xi = 0$ $\quad H_1^{(11)}$: $(y_t|\mathscr{X}_t, z_t, z_{t-1}; \boldsymbol{\theta}, \xi) \sim N\left(\boldsymbol{x}_t'\boldsymbol{\beta}_{z_t}, \sigma^2_{z_t}\left[1 + \frac{\xi(y_{t-1} - \boldsymbol{x}_{t-1}'\boldsymbol{\beta}_{z_{t-1}})^2}{\sigma^2_{z_{t-1}}}\right]\right)$

Table I.11: Lagrange multiplier test: REXP 80% DAX30 20%

	MS(1-2)		MS(2-1)		MS(2-2)	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
$H_0^{(10)}$	3.4497	3.9394*	1.7673***	34.1679**	3.5869	2.4138***
$H_0^{(11)}$	4.6954*	5.8412*	1.9581***	33.5088**	4.4118*	3.7996

Table I.12: Lagrange multiplier test: REXP 90% DAX30 10%

	MS(1-2)		MS(2-1)		MS(2-2)	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
$H_0^{(10)}$	7.1439**	1.1552***	7.0739**	41.0724**	6.1799*	11.0327**
$H_0^{(11)}$	3.0185	0.9793***	3.4857	41.1626**	3.1917	11.7223**

Table I.13: Lagrange multiplier test: REXP

	MS(1-2)		MS(2-1)		MS(2-2)	
	AR(0)	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
$H_0^{(10)}$	16.6099**	2.1669***	13.9641**	40.9100**	64.0846**	23.8996**
$H_0^{(11)}$	1.9282***	2.6027***	2.6903***	49.0283**	62.8667**	22.4693**

Note:
MS(m-s)-AR(p) denotes a Markov switching model with m mean equations, s regimes for the variance and p auto-regression lags in each mean equation. *, **, *** - the null hypothesis can be rejected at the 10%, the 5% and the 1% confidence level, respectively.

$H_0^{(10)}$: $\rho = 0$ $\quad H_1^{(10)}$: $(y_t|\mathscr{X}_t, z_t, z_{t-1}; \boldsymbol{\theta}, \rho) \sim N(\{\boldsymbol{x}_t'\boldsymbol{\beta}_{z_t} + \rho(y_{t-1} - \boldsymbol{x}_{t-1}'\boldsymbol{\beta}_{z_{t-1}})\}, \sigma^2_{z_t})$

$H_0^{(11)}$: $\xi = 0$ $\quad H_1^{(11)}$: $(y_t|\mathscr{X}_t, z_t, z_{t-1}; \boldsymbol{\theta}, \xi) \sim N\left(\boldsymbol{x}_t'\boldsymbol{\beta}_{z_t}, \sigma^2_{z_t}\left[1 + \frac{\xi(y_{t-1} - \boldsymbol{x}_{t-1}'\boldsymbol{\beta}_{z_{t-1}})^2}{\sigma^2_{z_{t-1}}}\right]\right)$

Appendix J

Garcia test (1.1975-12.2004)

Table J.1: Distribution of Garcia's SupC statistics: DAX30 (1.1975-12.2004)

Quantile	**MS(1-2)**		**MS(2-1)**		**MS(2-2)**	
(%)	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
1	1.95	1.87	1.85	3.49	3.48	4.93
5	2.54	2.49	2.47	4.39	4.34	5.97
10	2.95	2.91	2.91	4.99	4.91	6.67
15	3.29	3.25	3.27	5.37	5.35	7.23
20	3.58	3.54	3.58	5.74	5.73	7.64
25	3.86	3.82	3.88	6.08	6.06	8.02
30	4.13	4.09	4.13	6.41	6.39	8.38
35	4.39	4.36	4.40	6.73	6.71	8.73
40	4.66	4.63	4.65	7.04	7.01	9.08
45	4.96	4.91	4.92	7.38	7.34	9.44
50	5.24	5.20	5.19	7.70	7.69	9.83
55	5.55	5.51	5.51	8.07	8.05	10.23
60	5.85	5.83	5.85	8.41	8.40	10.67
65	6.23	6.22	6.20	8.83	8.80	11.12
70	6.63	6.62	6.59	9.30	9.26	11.54
75	7.09	7.05	7.07	9.84	9.74	12.10
80	7.66	7.59	7.64	10.45	10.28	12.76
85	8.35	8.36	8.33	11.27	10.94	13.60
90	9.31	9.25	9.22	12.35	12.03	14.69
95	10.98	10.67	10.79	14.01	13.64	16.41
99	14.39	13.97	14.66	17.65	17.18	20.48

Note:
Simulated distributions ware conducted for specifications from table B.1 and 10,000 simulations. MS(m-s)-AR(p) denotes a Markov switching model with m mean equations, s regimes for the variance and p auto-regression lags in each mean equation.

Table J.2: Distribution of Garcia's SupC statistics: REXP(10%)-DAX30(90%) (1.1975-12.2004)

Quantile	**MS(1-2)**		**MS(2-1)**		**MS(2-2)**	
(%)	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
1	1.88	1.90	1.91	3.55	3.45	4.85
5	2.55	2.53	2.57	4.41	4.36	6.00
10	2.96	2.95	2.97	4.98	4.94	6.67
15	3.29	3.26	3.31	5.39	5.38	7.19
20	3.56	3.55	3.61	5.77	5.74	7.62
25	3.84	3.83	3.88	6.12	6.13	8.03
30	4.08	4.11	4.17	6.40	6.44	8.40
35	4.33	4.37	4.43	6.72	6.77	8.76
40	4.60	4.63	4.69	7.04	7.10	9.12
45	4.89	4.91	4.97	7.35	7.41	9.47
50	5.19	5.19	5.25	7.67	7.76	9.83
55	5.51	5.52	5.55	8.02	8.12	10.21
60	5.82	5.86	5.87	8.36	8.49	10.64
65	6.20	6.20	6.23	8.80	8.91	11.12
70	6.59	6.58	6.64	9.21	9.33	11.57
75	7.05	7.01	7.11	9.72	9.88	12.15
80	7.57	7.57	7.66	10.36	10.47	12.82
85	8.29	8.28	8.36	11.11	11.26	13.60
90	9.22	9.16	9.27	12.07	12.25	14.64
95	10.74	10.64	10.76	13.65	14.05	16.47
99	14.55	14.12	14.39	17.58	17.81	20.59

Note:
Simulated distributions ware conducted for specifications from table B.2 and 10,000 simulations. MS(m-s)-AR(p) denotes a Markov switching model with m mean equations, s regimes for the variance and p auto-regression lags in each mean equation.

Table J.3: Distribution of Garcia's SupC statistics: REXP(20%)-DAX30(80%) (1.1975-12.2004)

Quantile	**MS(1-2)**		**MS(2-1)**		**MS(2-2)**	
(%)	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
1	1.91	1.86	1.94	3.50	3.48	4.91
5	2.56	2.53	2.55	4.36	4.35	6.01
10	3.00	2.98	2.96	4.97	4.95	6.66
15	3.33	3.30	3.29	5.39	5.40	7.16
20	3.61	3.60	3.60	5.76	5.75	7.59
25	3.86	3.87	3.87	6.07	6.10	7.95
30	4.13	4.12	4.11	6.39	6.41	8.34
35	4.41	4.37	4.37	6.71	6.73	8.69
40	4.67	4.64	4.65	7.03	7.06	9.04
45	4.95	4.90	4.92	7.38	7.39	9.44
50	5.24	5.17	5.21	7.71	7.75	9.78
55	5.53	5.48	5.49	8.07	8.09	10.19
60	5.84	5.80	5.82	8.45	8.46	10.59
65	6.19	6.16	6.14	8.83	8.88	11.02
70	6.64	6.57	6.55	9.26	9.31	11.50
75	7.08	7.02	6.99	9.80	9.82	12.02
80	7.60	7.57	7.56	10.46	10.41	12.69
85	8.32	8.21	8.25	11.17	11.20	13.45
90	9.26	9.18	9.12	12.23	12.20	14.56
95	10.84	10.76	10.60	13.91	13.82	16.28
99	14.09	14.08	14.00	17.65	17.43	19.99

Note:
Simulated distributions ware conducted for specifications from table B.3 and 10,000 simulations. MS(m-s)-AR(p) denotes a Markov switching model with m mean equations, s regimes for the variance and p auto-regression lags in each mean equation.

Table J.4: Distribution of Garcia's SupC statistics: REXP(25%)-DAX30(75%) (1.1975-12.2004)

Quantile	**MS(1-2)**		**MS(2-1)**		**MS(2-2)**	
(%)	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
1	1.87	1.90	1.91	3.54	3.43	4.97
5	2.52	2.51	2.51	4.40	4.39	5.98
10	2.95	2.96	2.95	4.94	4.96	6.65
15	3.29	3.30	3.28	5.37	5.37	7.18
20	3.56	3.60	3.58	5.74	5.76	7.61
25	3.86	3.88	3.87	6.10	6.09	8.03
30	4.11	4.14	4.14	6.42	6.43	8.39
35	4.35	4.41	4.41	6.74	6.76	8.78
40	4.62	4.66	4.66	7.06	7.09	9.13
45	4.87	4.92	4.94	7.40	7.45	9.47
50	5.16	5.23	5.22	7.71	7.80	9.85
55	5.47	5.54	5.51	8.08	8.13	10.25
60	5.81	5.85	5.87	8.45	8.52	10.65
65	6.17	6.19	6.22	8.89	8.89	11.08
70	6.56	6.56	6.63	9.35	9.34	11.59
75	7.02	7.02	7.08	9.86	9.88	12.10
80	7.57	7.57	7.62	10.46	10.51	12.73
85	8.23	8.25	8.31	11.24	11.26	13.53
90	9.16	9.15	9.29	12.26	12.29	14.59
95	10.76	10.70	10.77	14.03	13.93	16.19
99	14.31	14.48	14.31	17.70	17.51	19.87

Note:
Simulated distributions ware conducted for specifications from table B.4 and 10,000 simulations. MS(m-s)-AR(p) denotes a Markov switching model with m mean equations, s regimes for the variance and p auto-regression lags in each mean equation.

Table J.5: Distribution of Garcia's SupC statistics: REXP(30%)-DAX30(70%) (1.1975-12.2004)

Quantile	**MS(1-2)**		**MS(2-1)**		**MS(2-2)**	
(%)	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
1	1.90	1.90	1.90	3.50	3.54	4.94
5	2.53	2.54	2.50	4.41	4.40	6.01
10	2.94	2.97	2.95	4.95	4.95	6.67
15	3.30	3.30	3.27	5.37	5.39	7.20
20	3.58	3.59	3.58	5.75	5.78	7.65
25	3.87	3.86	3.84	6.12	6.15	8.03
30	4.12	4.12	4.12	6.43	6.47	8.42
35	4.37	4.39	4.36	6.77	6.81	8.76
40	4.63	4.63	4.63	7.13	7.13	9.10
45	4.89	4.92	4.88	7.47	7.48	9.48
50	5.17	5.21	5.16	7.83	7.82	9.88
55	5.44	5.52	5.47	8.18	8.16	10.27
60	5.75	5.83	5.80	8.54	8.52	10.69
65	6.12	6.16	6.16	8.94	8.94	11.13
70	6.55	6.56	6.54	9.40	9.38	11.62
75	7.00	7.03	6.95	9.94	9.89	12.16
80	7.51	7.61	7.47	10.59	10.49	12.82
85	8.18	8.24	8.16	11.34	11.26	13.57
90	9.07	9.15	9.10	12.37	12.30	14.71
95	10.65	10.75	10.60	14.08	14.08	16.55
99	14.26	14.09	13.97	17.69	17.67	20.28

Note:
Simulated distributions ware conducted for specifications from table B.5 and 10,000 simulations. MS(m-s)-AR(p) denotes a Markov switching model with m mean equations, s regimes for the variance and p auto-regression lags in each mean equation.

Table J.6: Distribution of Garcia's SupC statistics: REXP(40%)-DAX30(60%) (1.1975-12.2004)

Quantile	**MS(1-2)**		**MS(2-1)**		**MS(2-2)**	
(%)	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
1	1.91	1.97	1.91	3.41	3.51	4.89
5	2.51	2.54	2.51	4.36	4.41	6.01
10	2.95	2.94	2.95	4.93	5.01	6.69
15	3.30	3.28	3.28	5.37	5.46	7.18
20	3.58	3.55	3.58	5.77	5.82	7.59
25	3.86	3.82	3.87	6.10	6.16	7.99
30	4.14	4.12	4.14	6.45	6.51	8.36
35	4.38	4.37	4.39	6.79	6.83	8.69
40	4.65	4.63	4.65	7.09	7.13	9.04
45	4.91	4.91	4.93	7.43	7.45	9.41
50	5.19	5.21	5.21	7.77	7.79	9.79
55	5.53	5.50	5.52	8.14	8.15	10.19
60	5.87	5.81	5.86	8.49	8.53	10.60
65	6.23	6.15	6.20	8.89	8.94	11.04
70	6.64	6.54	6.58	9.32	9.39	11.55
75	7.10	6.98	7.00	9.81	9.85	12.11
80	7.64	7.58	7.54	10.44	10.42	12.76
85	8.29	8.23	8.27	11.21	11.13	13.56
90	9.23	9.15	9.23	12.21	12.10	14.66
95	10.70	10.66	10.77	13.94	13.82	16.58
99	14.37	14.14	13.96	17.57	17.36	20.16

Note:
Simulated distributions ware conducted for specifications from table B.6 and 10,000 simulations. MS(m-s)-AR(p) denotes a Markov switching model with m mean equations, s regimes for the variance and p auto-regression lags in each mean equation.

Table J.7: Distribution of Garcia's SupC statistics: REXP(50%)-DAX30(50%) (1.1975-12.2004)

Quantile	**MS(1-2)**		**MS(2-1)**		**MS(2-2)**	
(%)	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
1	1.92	1.95	1.88	3.53	3.47	4.93
5	2.53	2.54	2.54	4.42	4.36	6.07
10	2.97	2.99	2.95	4.98	4.92	6.75
15	3.28	3.33	3.28	5.41	5.36	7.24
20	3.55	3.60	3.56	5.81	5.74	7.65
25	3.83	3.88	3.85	6.14	6.11	8.06
30	4.11	4.15	4.12	6.46	6.45	8.41
35	4.39	4.41	4.39	6.78	6.76	8.72
40	4.65	4.67	4.65	7.13	7.08	9.08
45	4.93	4.94	4.92	7.47	7.41	9.46
50	5.22	5.22	5.22	7.80	7.74	9.82
55	5.52	5.54	5.53	8.15	8.08	10.18
60	5.84	5.86	5.87	8.53	8.44	10.60
65	6.20	6.19	6.24	8.90	8.83	11.04
70	6.60	6.56	6.63	9.34	9.29	11.56
75	7.09	6.99	7.06	9.87	9.80	12.12
80	7.63	7.57	7.59	10.47	10.35	12.77
85	8.30	8.23	8.23	11.21	11.11	13.53
90	9.23	9.15	9.13	12.22	12.20	14.67
95	10.68	10.69	10.67	13.74	13.81	16.59
99	14.29	14.28	14.28	17.61	17.18	20.56

Note:
Simulated distributions ware conducted for specifications from table B.7 and 10,000 simulations. MS(m-s)-AR(p) denotes a Markov switching model with m mean equations, s regimes for the variance and p auto-regression lags in each mean equation.

Table J.8: Distribution of Garcia's SupC statistics: REXP(60%)-DAX30(40%) (1.1975-12.2004)

Quantile	**MS(1-2)**		**MS(2-1)**		**MS(2-2)**	
(%)	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
1	1.89	1.92	1.94	3.43	3.60	5.00
5	2.50	2.52	2.55	4.39	4.38	6.03
10	2.94	2.93	2.99	4.95	4.96	6.72
15	3.27	3.28	3.32	5.37	5.38	7.21
20	3.58	3.58	3.63	5.76	5.74	7.62
25	3.87	3.89	3.91	6.09	6.08	8.01
30	4.15	4.15	4.18	6.44	6.40	8.39
35	4.43	4.40	4.42	6.77	6.73	8.76
40	4.68	4.67	4.67	7.11	7.04	9.11
45	4.95	4.95	4.93	7.47	7.36	9.47
50	5.25	5.24	5.21	7.79	7.70	9.84
55	5.53	5.54	5.51	8.13	8.08	10.23
60	5.85	5.86	5.84	8.54	8.43	10.62
65	6.20	6.22	6.18	8.94	8.87	11.07
70	6.61	6.63	6.57	9.37	9.33	11.53
75	7.10	7.09	7.05	9.85	9.85	12.07
80	7.61	7.64	7.62	10.47	10.45	12.71
85	8.27	8.31	8.28	11.22	11.18	13.55
90	9.20	9.13	9.15	12.20	12.22	14.65
95	10.79	10.74	10.77	13.86	13.71	16.29
99	14.17	14.18	14.35	17.48	17.35	20.31

Note:
Simulated distributions ware conducted for specifications from table B.8 and 10,000 simulations. MS(m-s)-AR(p) denotes a Markov switching model with m mean equations, s regimes for the variance and p auto-regression lags in each mean equation.

Table J.9: Distribution of Garcia's SupC statistics: REXP(70%)-DAX30(30%) (1.1975-12.2004)

Quantile	**MS(1-2)**		**MS(2-1)**		**MS(2-2)**	
(%)	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
1	1.90	1.88	1.91	3.53	3.54	4.90
5	2.52	2.53	2.55	4.39	4.37	5.99
10	2.94	2.97	2.98	5.00	4.96	6.68
15	3.29	3.30	3.31	5.42	5.36	7.19
20	3.58	3.59	3.61	5.78	5.76	7.65
25	3.85	3.84	3.89	6.11	6.12	8.05
30	4.13	4.11	4.15	6.44	6.43	8.42
35	4.39	4.37	4.42	6.77	6.75	8.79
40	4.68	4.63	4.70	7.11	7.08	9.15
45	4.95	4.92	4.99	7.45	7.40	9.51
50	5.25	5.19	5.27	7.79	7.77	9.87
55	5.52	5.47	5.58	8.16	8.10	10.27
60	5.85	5.79	5.87	8.53	8.50	10.69
65	6.22	6.12	6.22	8.95	8.91	11.13
70	6.62	6.52	6.61	9.37	9.31	11.61
75	7.09	6.99	7.05	9.89	9.82	12.17
80	7.64	7.49	7.56	10.50	10.42	12.80
85	8.30	8.18	8.24	11.23	11.20	13.59
90	9.25	9.13	9.21	12.23	12.17	14.70
95	10.75	10.68	10.71	13.98	13.76	16.44
99	14.58	14.38	13.90	17.73	17.12	20.41

Note:
Simulated distributions ware conducted for specifications from table B.9 and 10,000 simulations. MS(m-s)-AR(p) denotes a Markov switching model with m mean equations, s regimes for the variance and p auto-regression lags in each mean equation.

Table J.10: Distribution of Garcia's SupC statistics: REXP(75%)-DAX30(25%) (1.1975-12.2004)

Quantile	**MS(1-2)**		**MS(2-1)**		**MS(2-2)**	
(%)	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
1	1.89	1.87	1.91	3.53	3.41	4.87
5	2.53	2.53	2.54	4.38	4.32	6.02
10	2.95	2.95	2.97	4.93	4.92	6.69
15	3.29	3.28	3.31	5.36	5.37	7.19
20	3.59	3.58	3.60	5.71	5.74	7.62
25	3.88	3.83	3.89	6.07	6.08	7.99
30	4.17	4.08	4.15	6.40	6.42	8.33
35	4.43	4.36	4.43	6.70	6.73	8.71
40	4.70	4.62	4.70	7.01	7.06	9.08
45	4.95	4.89	4.99	7.33	7.40	9.46
50	5.24	5.19	5.27	7.68	7.74	9.84
55	5.54	5.51	5.58	8.08	8.10	10.21
60	5.85	5.82	5.90	8.46	8.44	10.61
65	6.23	6.18	6.23	8.84	8.86	11.06
70	6.64	6.60	6.61	9.31	9.28	11.53
75	7.10	7.05	7.06	9.77	9.77	12.10
80	7.61	7.60	7.58	10.39	10.41	12.76
85	8.28	8.30	8.26	11.15	11.13	13.55
90	9.26	9.22	9.15	12.19	12.12	14.56
95	10.71	10.75	10.68	13.93	13.75	16.43
99	14.16	14.13	13.77	17.54	17.61	19.78

Note:
Simulated distributions ware conducted for specifications from table B.10 and 10,000 simulations. MS(m-s)-AR(p) denotes a Markov switching model with m mean equations, s regimes for the variance and p auto-regression lags in each mean equation.

Table J.11: Distribution of Garcia's SupC statistics: REXP(80%)-DAX30(20%) (1.1975-12.2004)

Quantile	**MS(1-2)**		**MS(2-1)**		**MS(2-2)**	
(%)	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
1	1.90	1.90	1.86	3.52	3.47	4.88
5	2.52	2.53	2.53	4.39	4.35	5.99
10	2.97	2.96	2.94	4.99	4.94	6.65
15	3.30	3.31	3.28	5.43	5.39	7.14
20	3.60	3.59	3.56	5.79	5.76	7.57
25	3.86	3.85	3.84	6.15	6.11	7.98
30	4.12	4.10	4.11	6.44	6.44	8.34
35	4.38	4.37	4.36	6.75	6.76	8.70
40	4.66	4.63	4.63	7.06	7.11	9.06
45	4.93	4.91	4.91	7.38	7.41	9.42
50	5.22	5.20	5.19	7.71	7.77	9.79
55	5.51	5.50	5.51	8.06	8.15	10.15
60	5.83	5.82	5.82	8.46	8.52	10.57
65	6.20	6.20	6.14	8.84	8.94	11.01
70	6.59	6.58	6.54	9.28	9.37	11.50
75	7.09	7.03	6.98	9.76	9.89	12.07
80	7.63	7.55	7.50	10.39	10.45	12.71
85	8.35	8.24	8.18	11.14	11.22	13.54
90	9.24	9.19	9.12	12.17	12.21	14.67
95	10.76	10.73	10.71	13.93	13.80	16.49
99	13.96	14.11	13.98	17.63	17.71	20.51

Note:
Simulated distributions ware conducted for specifications from table B.11 and 10,000 simulations. MS(m-s)-AR(p) denotes a Markov switching model with m mean equations, s regimes for the variance and p auto-regression lags in each mean equation.

Table J.12: Distribution of Garcia's SupC statistics: REXP(90%)-DAX30(10%) (1.1975-12.2004)

Quantile	**MS(1-2)**		**MS(2-1)**		**MS(2-2)**	
(%)	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
1	1.89	1.94	1.86	3.45	3.58	4.95
5	2.53	2.55	2.53	4.36	4.38	5.98
10	2.94	2.97	2.95	4.94	4.95	6.69
15	3.28	3.31	3.32	5.42	5.39	7.18
20	3.56	3.59	3.60	5.78	5.76	7.66
25	3.84	3.88	3.87	6.11	6.11	8.04
30	4.10	4.14	4.13	6.46	6.46	8.42
35	4.36	4.39	4.40	6.79	6.78	8.82
40	4.62	4.63	4.66	7.10	7.11	9.18
45	4.89	4.91	4.93	7.43	7.40	9.54
50	5.17	5.19	5.21	7.76	7.70	9.91
55	5.47	5.49	5.52	8.13	8.05	10.31
60	5.80	5.82	5.87	8.50	8.44	10.75
65	6.14	6.14	6.21	8.88	8.87	11.17
70	6.56	6.54	6.61	9.30	9.32	11.64
75	7.00	6.97	7.05	9.83	9.86	12.17
80	7.53	7.46	7.59	10.47	10.44	12.85
85	8.21	8.16	8.26	11.23	11.22	13.70
90	9.13	9.07	9.16	12.25	12.27	14.80
95	10.61	10.59	10.76	13.85	13.98	16.58
99	14.01	14.05	13.97	17.85	17.38	20.00

Note:
Simulated distributions ware conducted for specifications from table B.12 and 10,000 simulations. MS(m-s)-AR(p) denotes a Markov switching model with m mean equations, s regimes for the variance and p auto-regression lags in each mean equation.

Table J.13: Distribution of Garcia's SupC statistics: REXP (1.1975-12.2004)

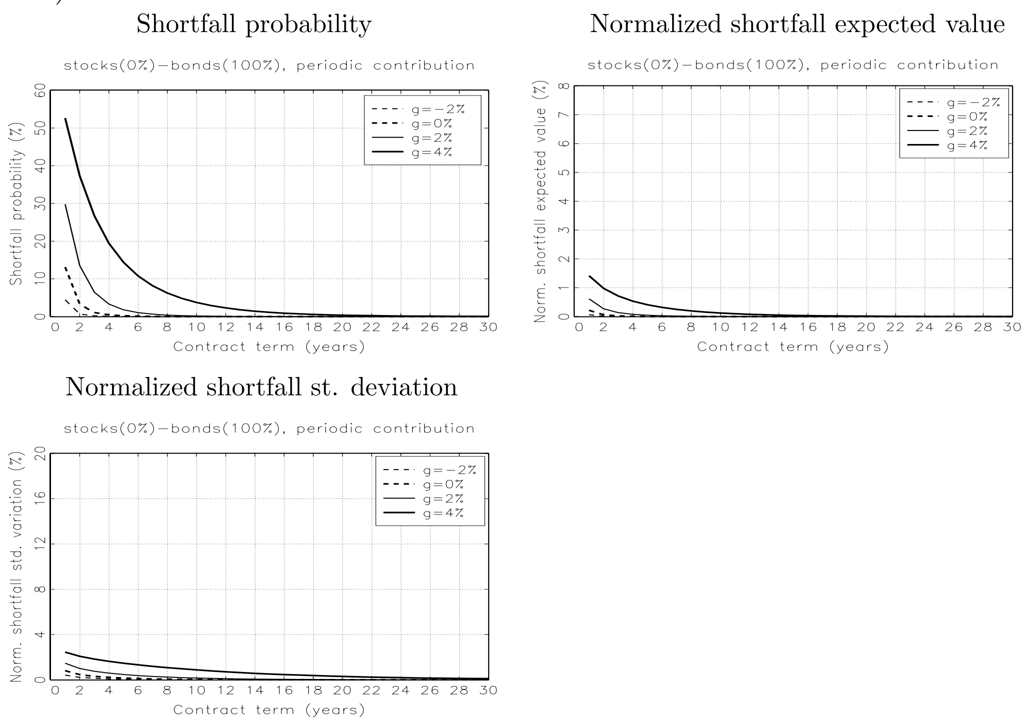
Quantile	**MS(1-2)**		**MS(2-1)**		**MS(2-2)**	
(%)	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**	**AR(0)**	**AR(1)**
1	1.85	1.91	1.88	3.48	3.45	4.90
5	2.47	2.56	2.54	4.42	4.37	6.04
10	2.93	2.97	2.96	4.99	4.96	6.68
15	3.25	3.32	3.27	5.41	5.41	7.18
20	3.54	3.59	3.58	5.77	5.78	7.60
25	3.82	3.87	3.84	6.15	6.12	7.99
30	4.09	4.15	4.10	6.50	6.46	8.38
35	4.38	4.41	4.39	6.80	6.77	8.74
40	4.64	4.68	4.65	7.11	7.09	9.08
45	4.93	4.95	4.91	7.42	7.45	9.46
50	5.22	5.24	5.19	7.79	7.77	9.80
55	5.53	5.54	5.47	8.17	8.14	10.22
60	5.85	5.85	5.79	8.53	8.49	10.65
65	6.22	6.17	6.18	8.95	8.92	11.14
70	6.58	6.56	6.57	9.41	9.36	11.62
75	7.05	7.03	7.02	9.96	9.89	12.19
80	7.58	7.56	7.55	10.58	10.48	12.88
85	8.28	8.28	8.18	11.30	11.27	13.66
90	9.28	9.17	9.13	12.30	12.26	14.73
95	10.80	10.59	10.63	13.99	13.80	16.38
99	14.20	14.29	14.17	18.09	17.51	19.93

Note:
Simulated distributions ware conducted for specifications from table B.13 and 10,000 simulations. MS(m-s)-AR(p) denotes a Markov switching model with m mean equations, s regimes for the variance and p auto-regression lags in each mean equation.

Appendix K

Impact of the guarantee level on the guarantee risk

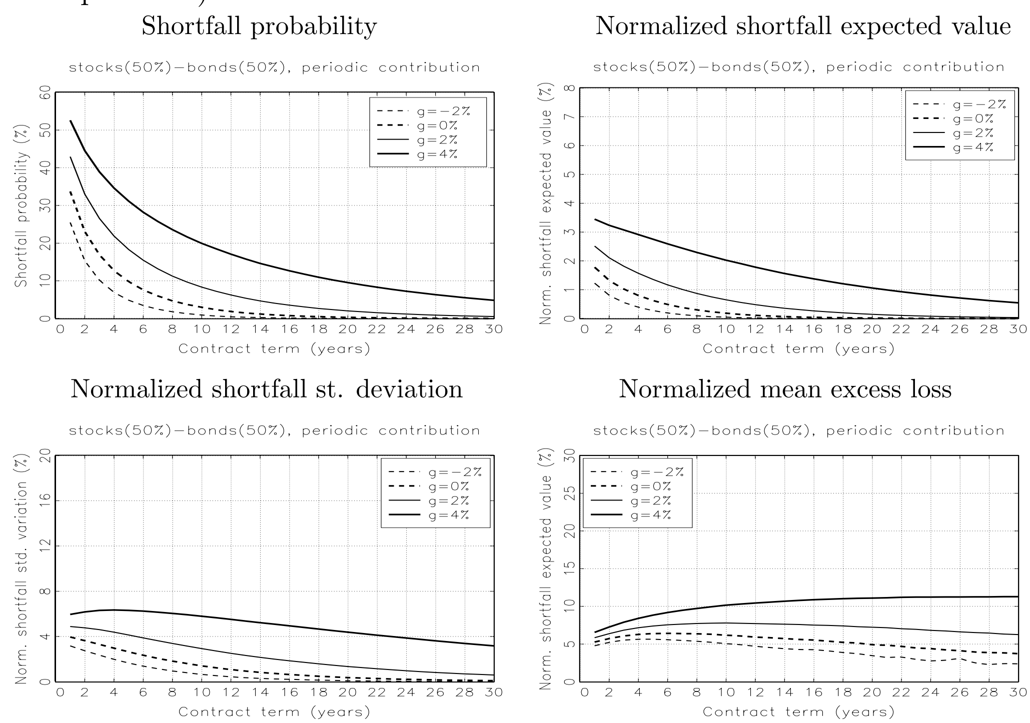
Figure K.1: Impact of the guarantee level on risk measures (pure bond portfolio)



Note:

The figure depicts the impact of the guarantee level on the normalized guarantee cost using the example of the pure stock (to left panel), the fifty-fifty stock-bond (top right panel), and the pure bond portfolio (bottom panel). The cost is computed for the MS(1-2) model under the Esscher probability measure. The thin dashed line represents the guarantee of -2% p.a., the thick dashed line – 0% p.a., the thin solid line – 2% p.a., and the thick solid line – 4% p.a., respectively.

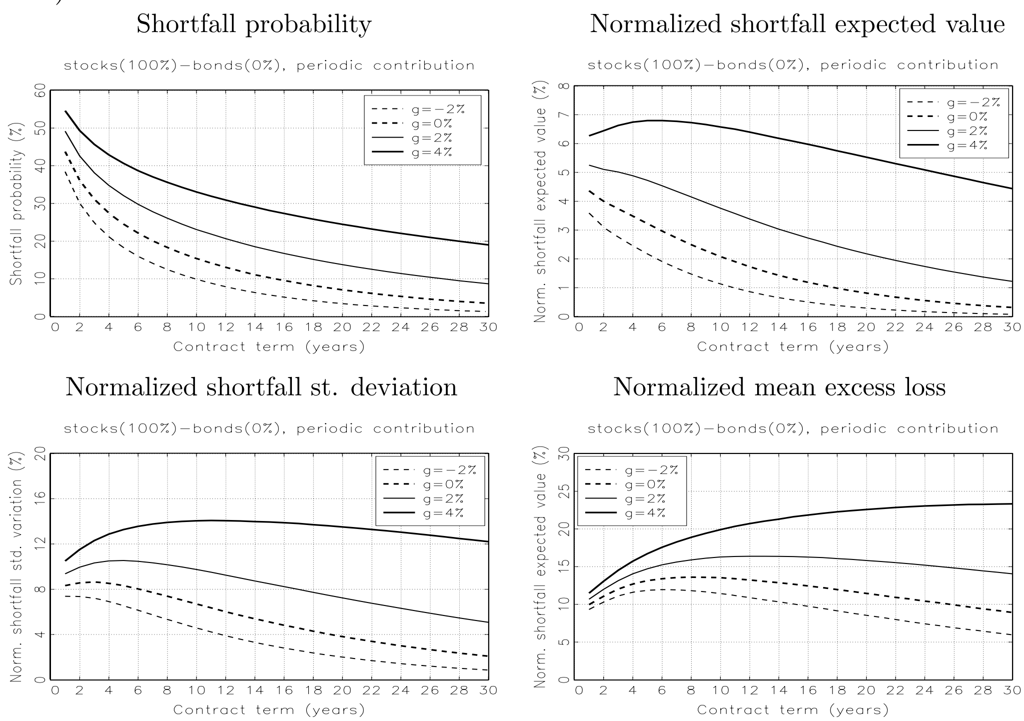
Figure K.2: Impact of the guarantee level on risk measures (50%-50% stock-bond portfolio)



Note:

The figure depicts the impact of the guarantee level on the normalized guarantee cost using the example of the pure stock (to left panel), the fifty-fifty stock-bond (top right panel), and the pure bond portfolio (bottom panel). The cost is computed for the MS(1-2) model under the Esscher probability measure. The thin dashed line represents the guarantee of -2% p.a., the thick dashed line – 0% p.a., the thin solid line – 2% p.a., and the thick solid line – 4% p.a., respectively.

Figure K.3: Impact of the guarantee level on risk measures (pure stock portfolio)

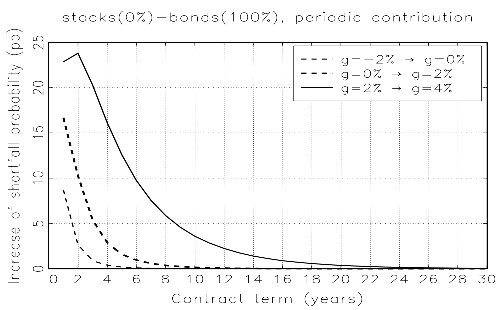


Note:

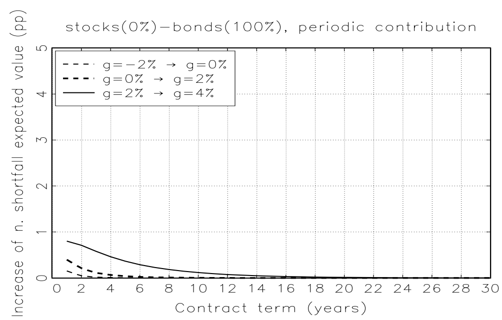
The figure depicts the impact of the guarantee level on the normalized guarantee cost using the example of the pure stock (to left panel), the fifty-fifty stock-bond (top right panel), and the pure bond portfolio (bottom panel). The cost is computed for the MS(1-2) model under the Esscher probability measure. The thin dashed line represents the guarantee of -2% p.a., the thick dashed line – 0% p.a., the thin solid line – 2% p.a., and the thick solid line – 4% p.a., respectively.

Figure K.4: Sensitivity of risk measures to changes in the guarantee level (pure bond portfolio)

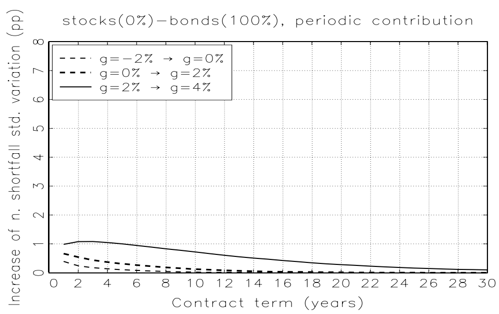
Shortfall probability



Normalized shortfall expected value

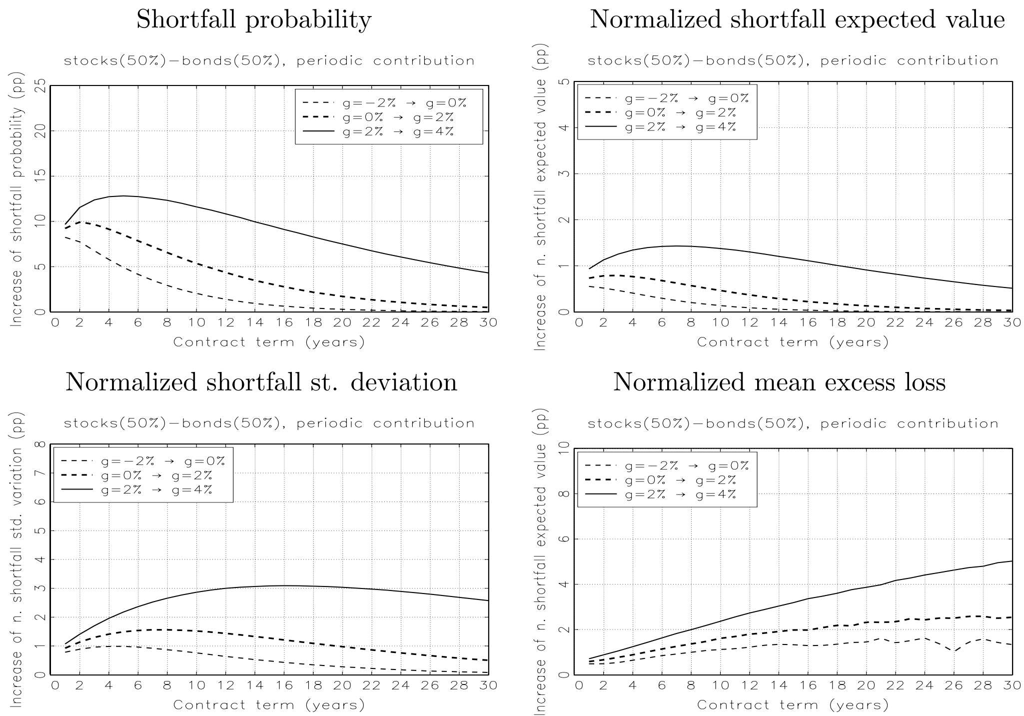


Normalized shortfall st. deviation



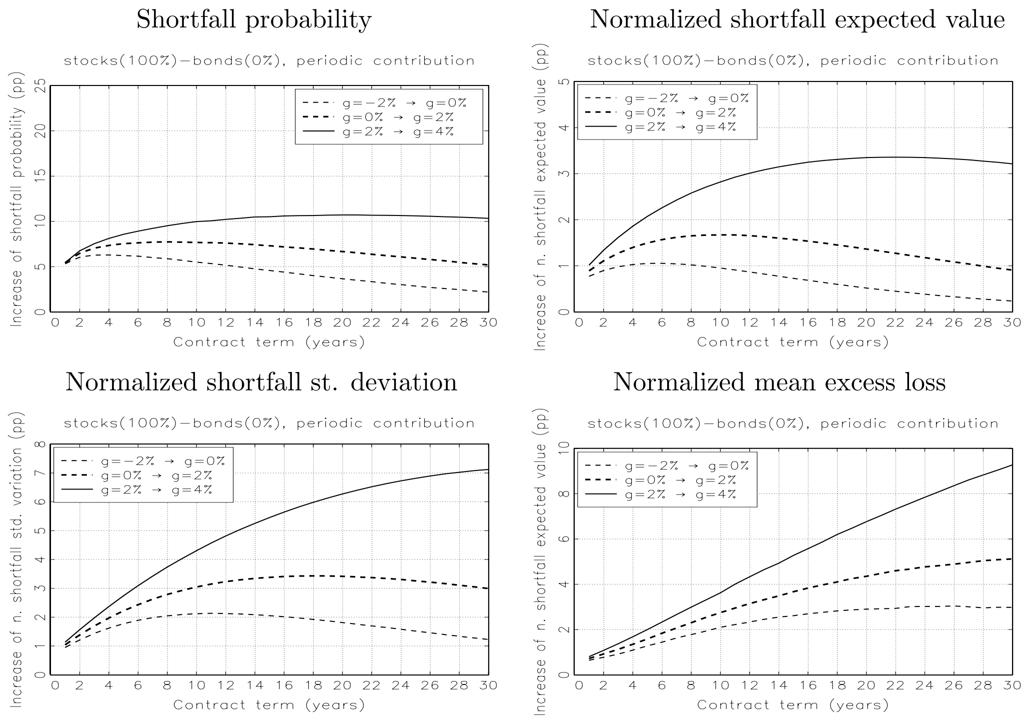
Note:
The figure depicts the sensitivity of the normalized cost to the change of the guarantee level using the example of the pure stock (to left panel), the fifty-fifty stock-bond (top right panel), and the pure bond portfolio (bottom panel). The cost is computed for the MS(1-2) model under the Esscher probability measure. The thin dashed line represents the guarantee increase from -2% to 0% p.a., the thick dashed line – from 0% to 2% p.a., and the solid line – from 2% to 4% p.a., respectively.

Figure K.5: Sensitivity of risk measures to changes in the guarantee level (50%-50% stock-bond portfolio)



Note:
The figure depicts the sensitivity of the normalized cost to the change of the guarantee level using the example of the pure stock (to left panel), the fifty-fifty stock-bond (top right panel), and the pure bond portfolio (bottom panel). The cost is computed for the MS(1-2) model under the Esscher probability measure. The thin dashed line represents the guarantee increase from -2% to 0% p.a., the thick dashed line – from 0% to 2% p.a., and the solid line – from 2% to 4% p.a., respectively.

Figure K.6: Sensitivity of risk measures to changes in the guarantee level (pure stock portfolio)



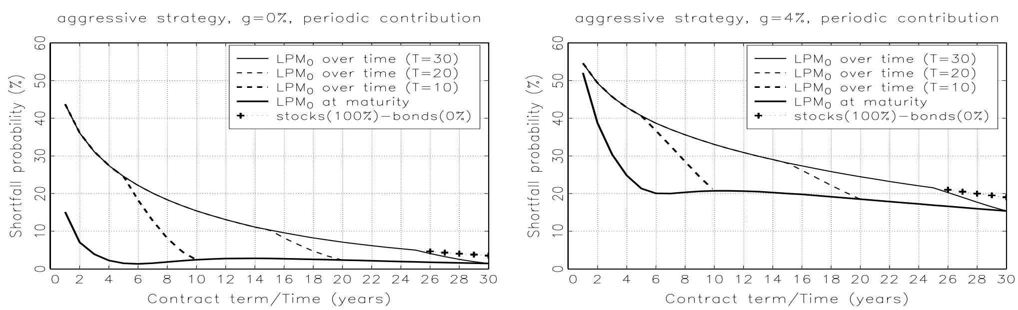
Note:
The figure depicts the sensitivity of the normalized cost to the change of the guarantee level using the example of the pure stock (to left panel), the fifty-fifty stock-bond (top right panel), and the pure bond portfolio (bottom panel). The cost is computed for the MS(1-2) model under the Esscher probability measure. The thin dashed line represents the guarantee increase from -2% to 0% p.a., the thick dashed line – from 0% to 2% p.a., and the solid line – from 2% to 4% p.a., respectively.

Appendix L

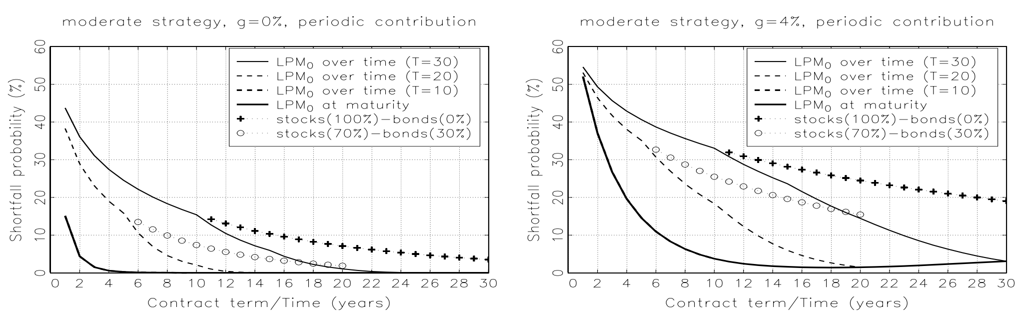
Impact of time and the contract term on the guarantee shortfall risk (life-cycle strategies)

Figure L.1: Impact of time and the contract term on the shortfall probability (life-cycle strategies)

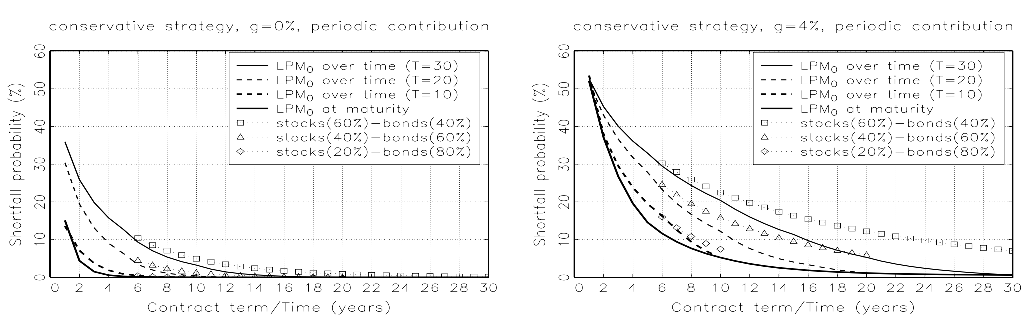
Aggressive strategy



Moderate strategy



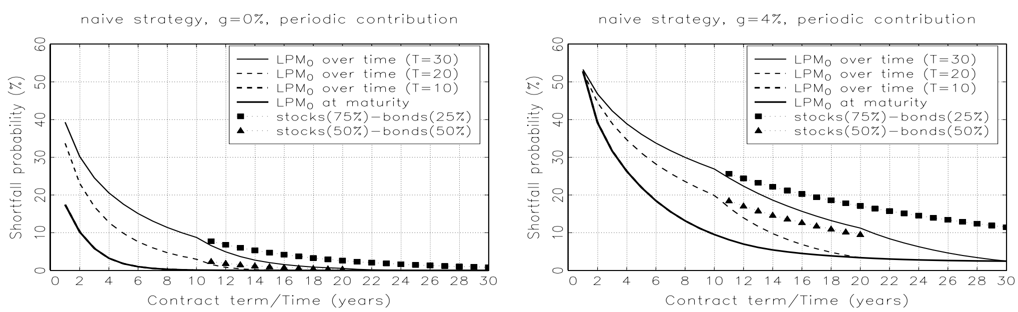
Conservative strategy



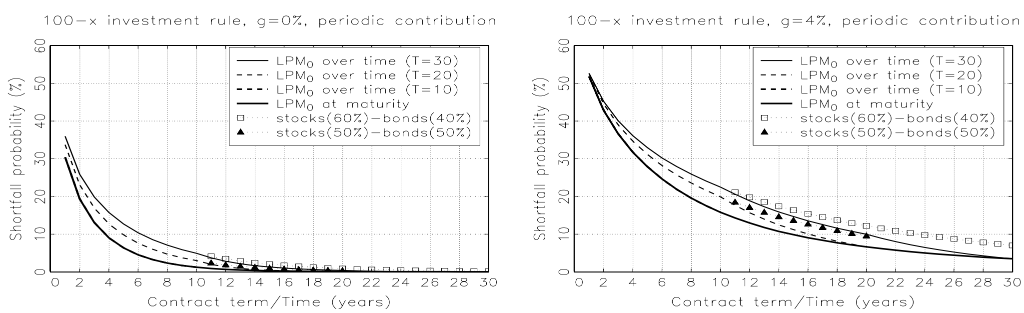
continued on the next page

continued from the previous page

Naive strategy



100-x investment rule

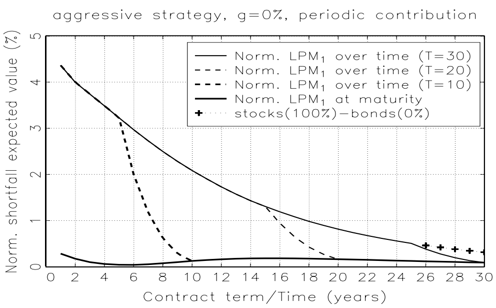


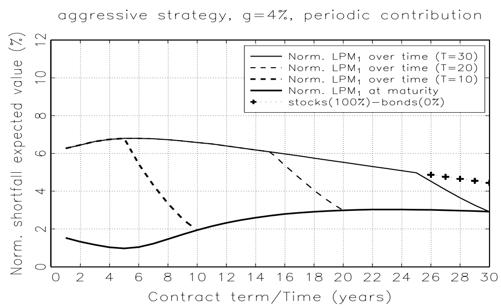
Note:

The figure depicts the impact of time and contract term on the normalized guarantee cost. The cost is computed for the MS(1-2) model under the Esscher probability measure. The thick dashed line shows how the normalized cost changes through time for a 10-year contract, the thin dashed line – for a 20-year contract, and the thin solid line – for a 30-year contract, respectively. The thick solid line shows the normalized contract at contract maturity. The left column represents the 0% guarantee, and the right row the 4% guarantee.

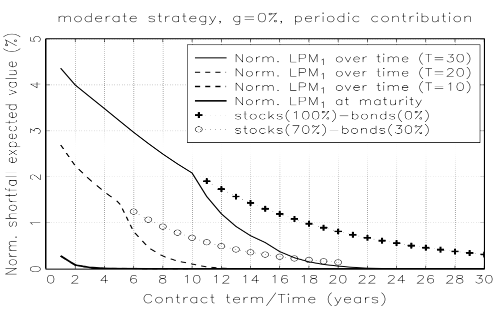
Figure L.2: Impact of time and the contract term on the normalized shortfall expectation (life-cycle strategies)

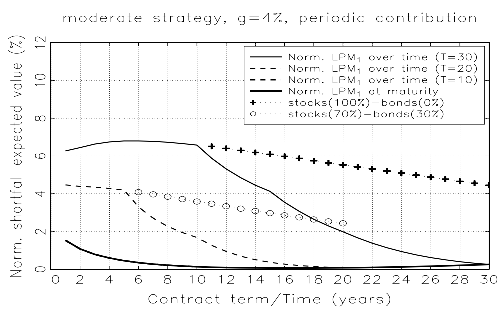
Aggressive strategy



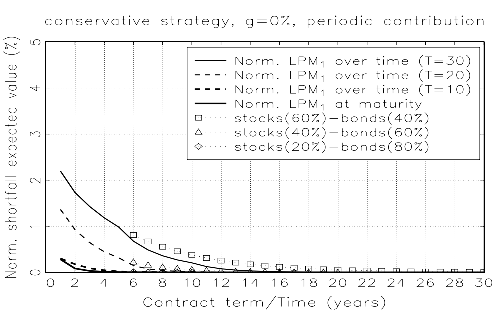


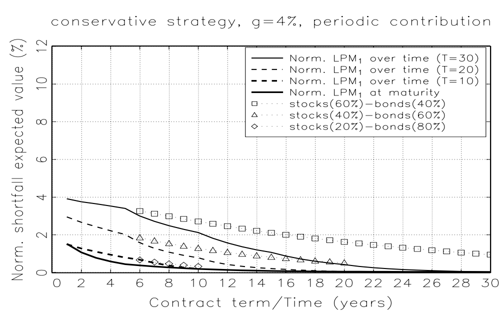
Moderate strategy





Conservative strategy

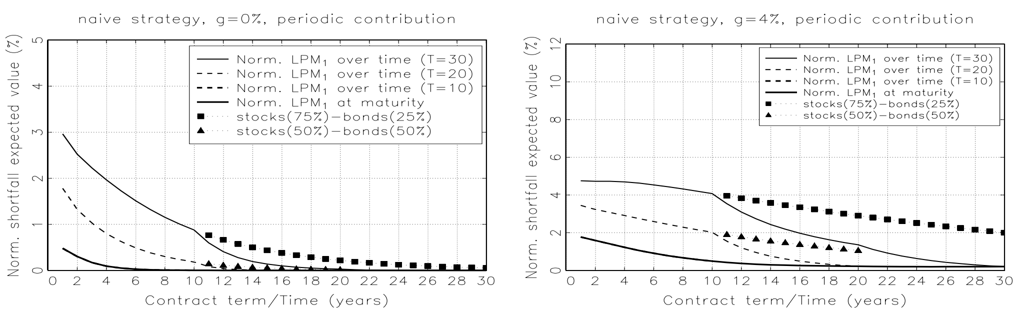




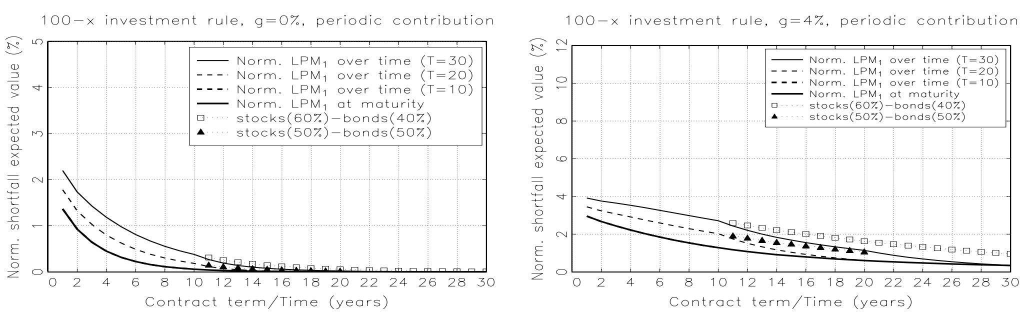
continued on the next page

continued from the previous page

Naive strategy



100-x investment rule

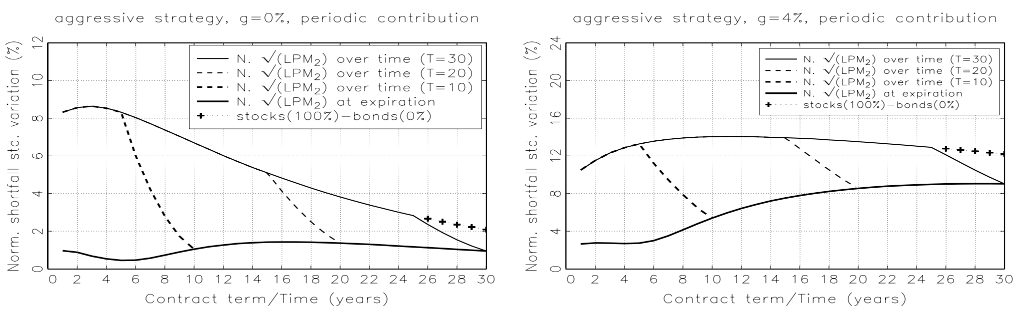


Note:

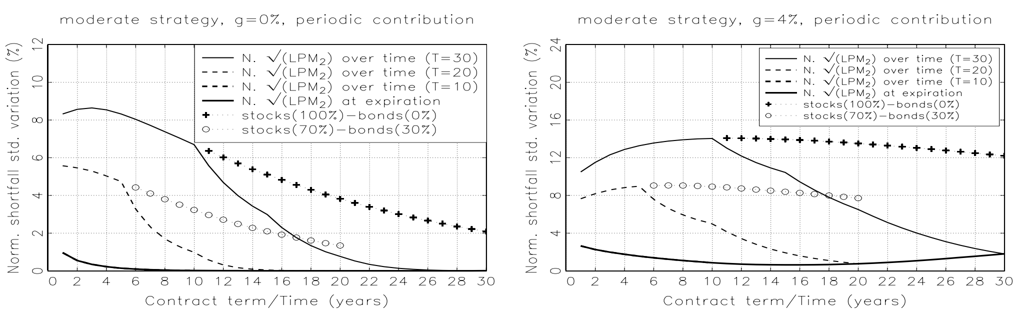
The figure depicts the impact of time and contract term on the normalized guarantee cost. The cost is computed for the MS(1-2) model under the Esscher probability measure. The thick dashed line shows how the normalized cost changes through time for a 10-year contract, the thin dashed line – for a 20-year contract, and the thin solid line – for a 30-year contract, respectively. The thick solid line shows the normalized contract at contract maturity. The left column represents the 0% guarantee, and the right row the 4% guarantee.

Figure L.3: Impact of time and the contract term on the normalized shortfall standard deviation (life-cycle strategies)

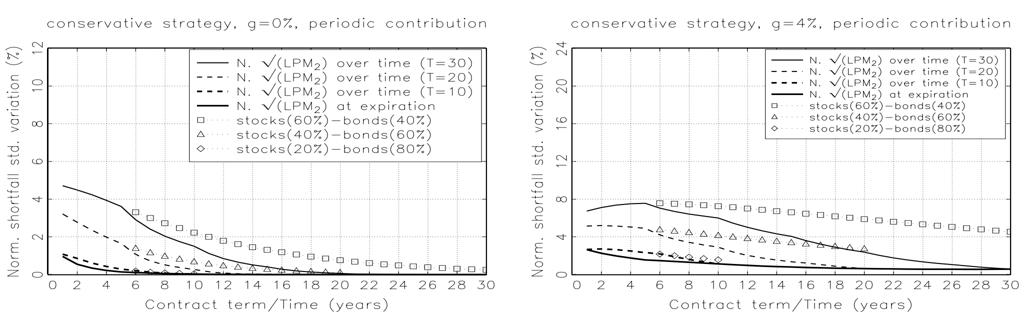
Aggressive strategy



Moderate strategy



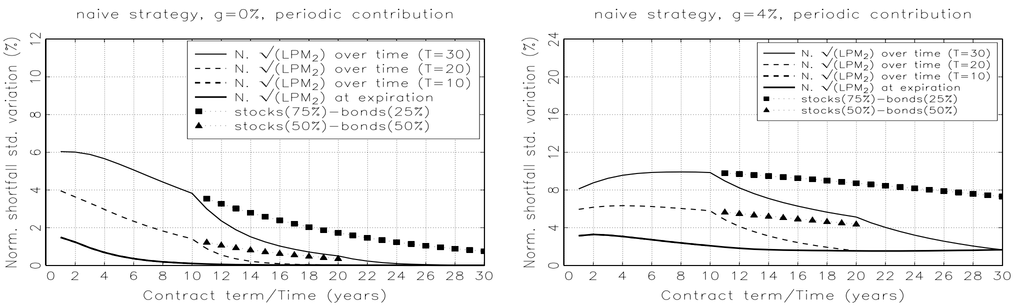
Conservative strategy



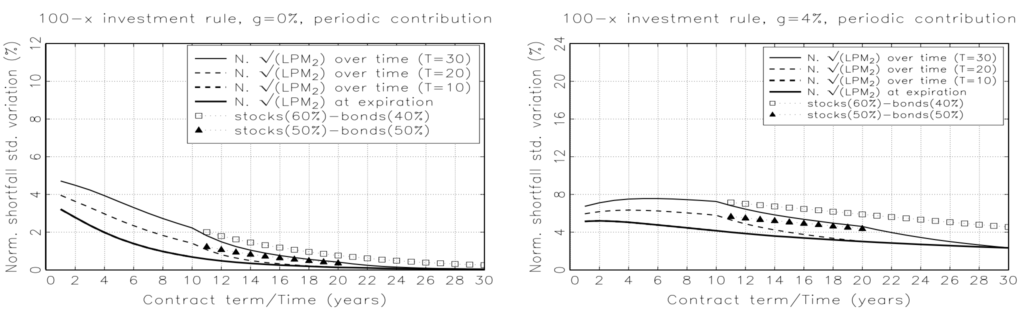
continued on the next page

continued from the previous page

Naive strategy



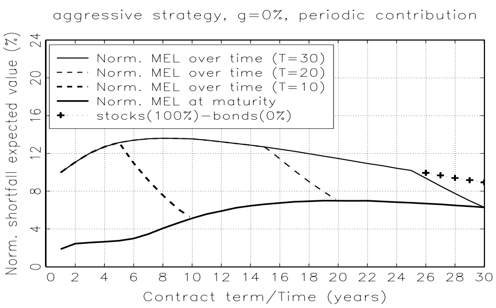
100-x investment rule

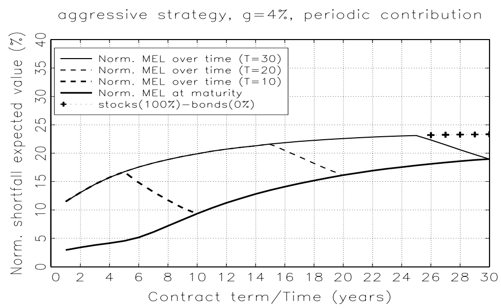


Note:
The figure depicts the impact of time and contract term on the normalized guarantee cost. The cost is computed for the MS(1-2) model under the Esscher probability measure. The thick dashed line shows how the normalized cost changes through time for a 10-year contract, the thin dashed line – for a 20-year contract, and the thin solid line – for a 30-year contract, respectively. The thick solid line shows the normalized contract at contract maturity. The left column represents the 0% guarantee, and the right row the 4% guarantee.

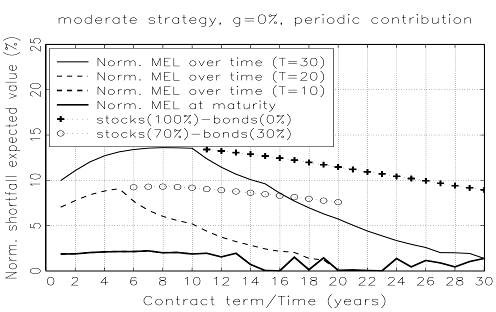
Figure L.4: Impact of time and the contract term on the normalized mean excess loss (life-cycle strategies)

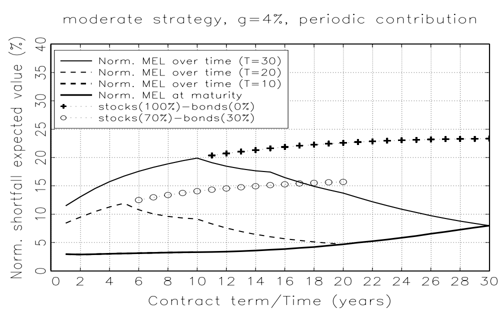
Aggressive strategy



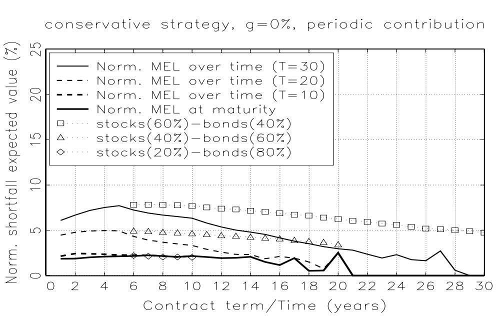


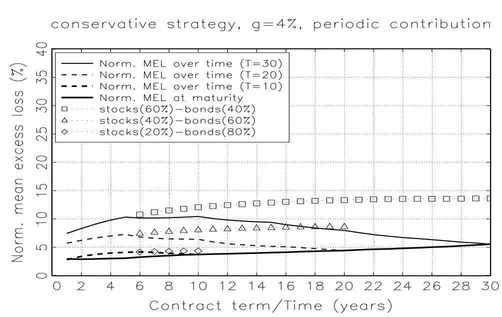
Moderate strategy





Conservative strategy

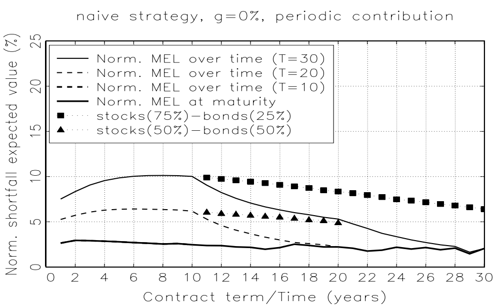


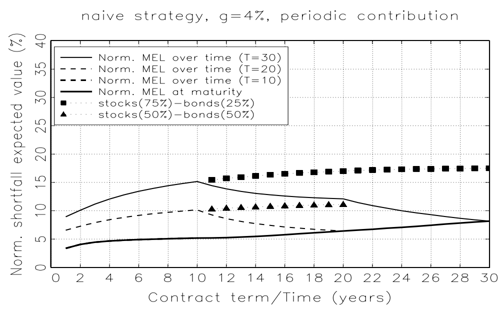


continued on the next page

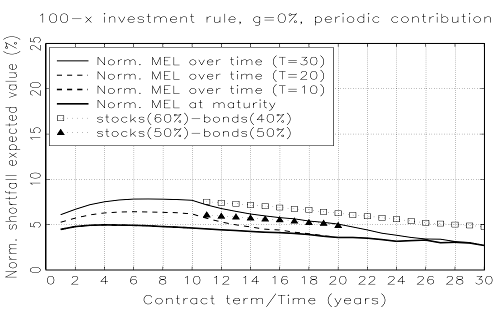
continued from the previous page

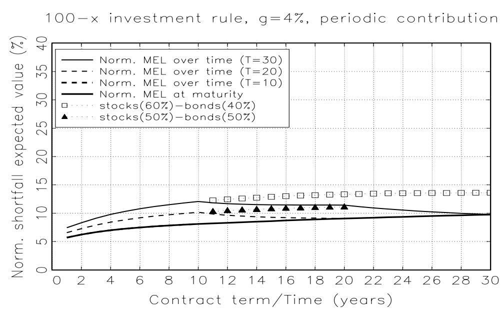
Naive strategy





100-x investment rule

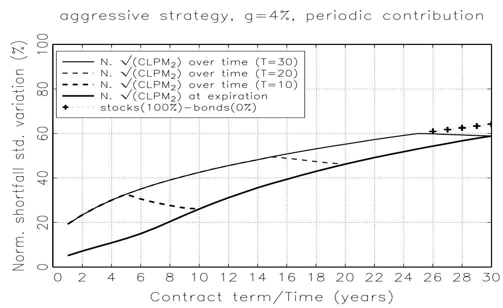




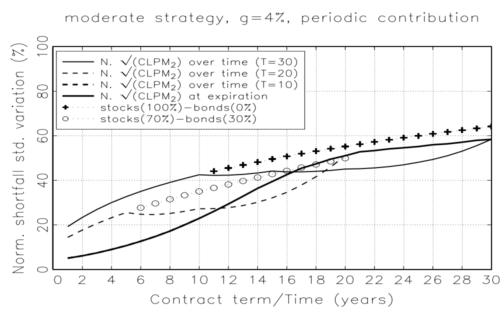
Note:
The figure depicts the impact of time and contract term on the normalized guarantee cost. The cost is computed for the MS(1-2) model under the Esscher probability measure. The thick dashed line shows how the normalized cost changes through time for a 10-year contract, the thin dashed line – for a 20-year contract, and the thin solid line – for a 30-year contract, respectively. The thick solid line shows the normalized contract at contract maturity. The left column represents the 0% guarantee, and the right row the 4% guarantee.

Figure L.5: Impact of time and the contract term on the normalized conditional shortfall standard deviation (life-cycle strategies)

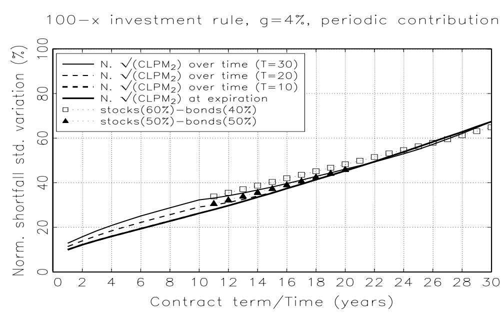
Aggressive strategy



Moderate strategy



100-x investment rule



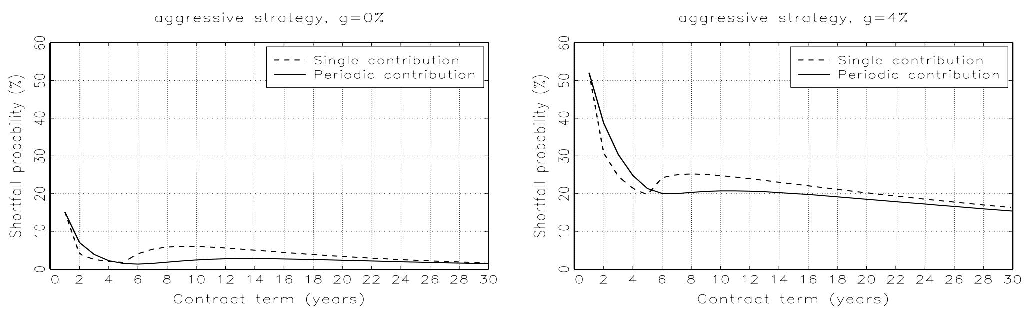
Note:
The figure depicts the impact of time and contract term on the normalized guarantee cost. The cost is computed for the MS(1-2) model under the Esscher probability measure. The thick dashed line shows how the normalized cost changes through time for a 10-year contract, the thin dashed line – for a 20-year contract, and the thin solid line – for a 30-year contract, respectively. The thick solid line shows the normalized contract at contract maturity. The left column represents the 0% guarantee, and the right row the 4% guarantee.

Appendix M

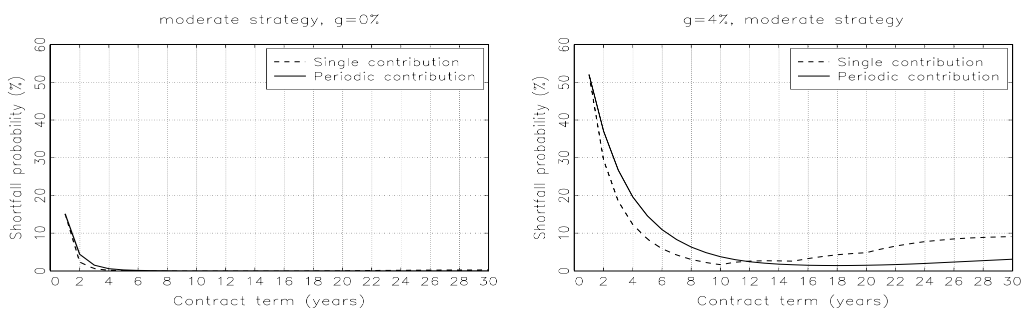
Impact of the contribution payment scheme on the guarantee shortfall risk (life-cycle strategies)

Figure M.1: Impact of the contribution payment scheme on the shortfall probability (life-cycle strategies)

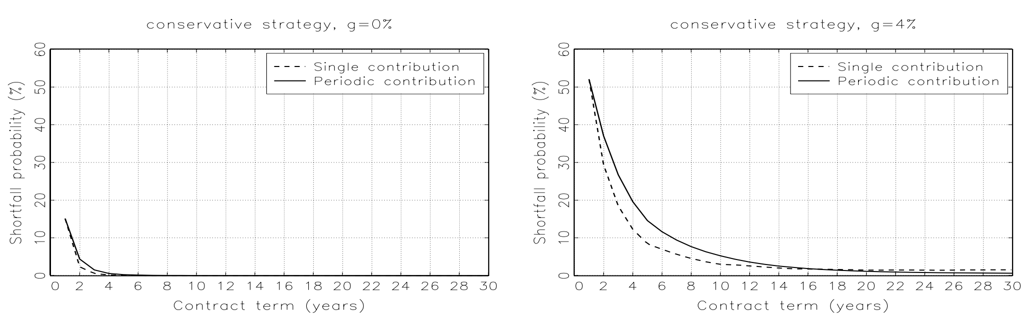
Aggressive strategy



Moderate strategy



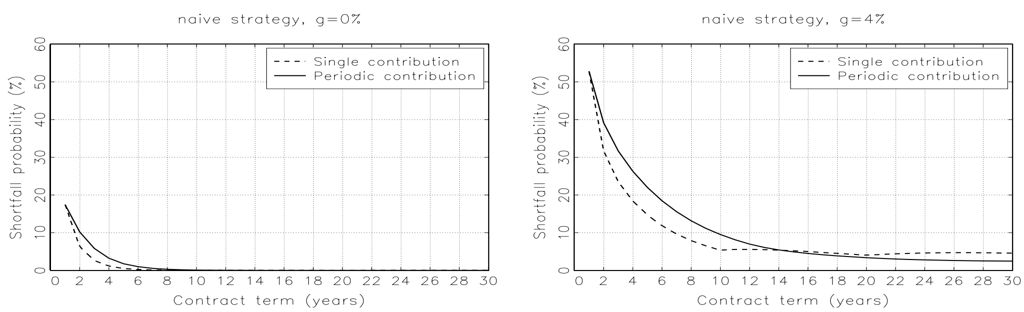
Conservative strategy



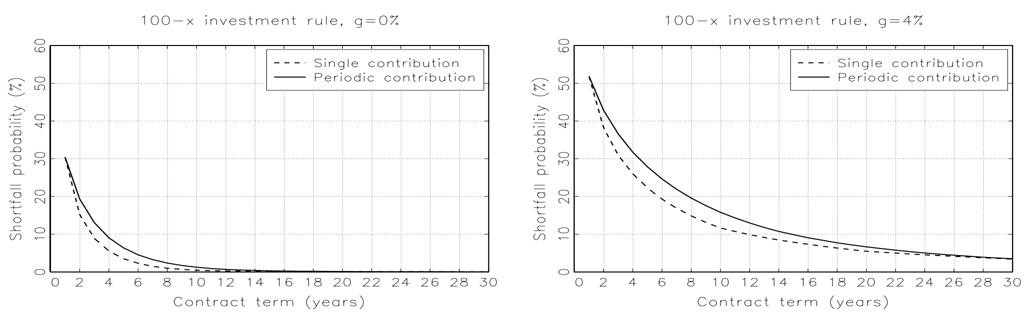
continued on the next page

continued from the previous page

Naive strategy



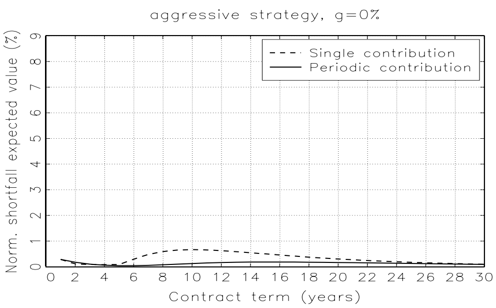
100-x investment rule

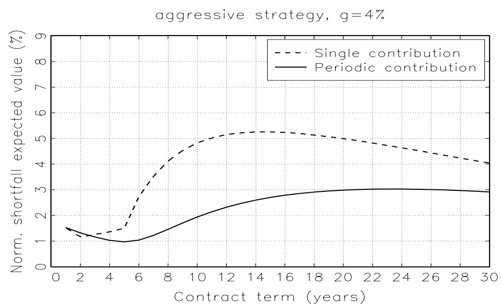


Note:
The figure depicts the impact of the payment scheme on the normalized guarantee cost using the example of life-cycle strategies. The solid line represents the periodic payment scheme (€1200 up-font yearly) and the dashed line the single premium case. The single premium is equal to the net present value of periodic contributions. The left column shows contracts with a 0% guarantee and the right column contracts with a 4% guarantee. The x-axis shows the time/contract maturity in years and the y-axis – the normalized guarantee cost in per cent.

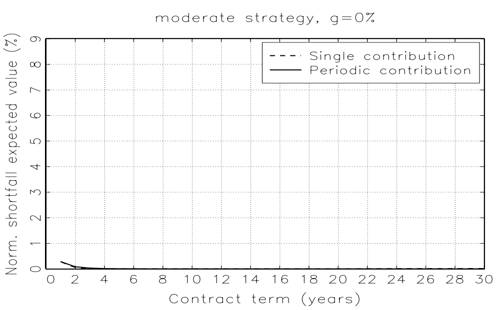
Figure M.2: Impact of the contribution payment scheme on the normalized shortfall expectation (life-cycle strategies)

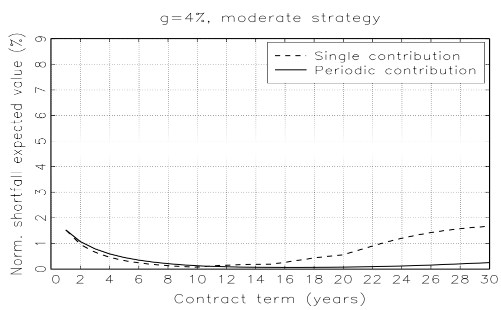
Aggressive strategy



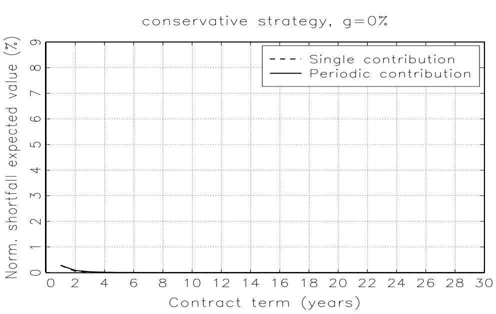


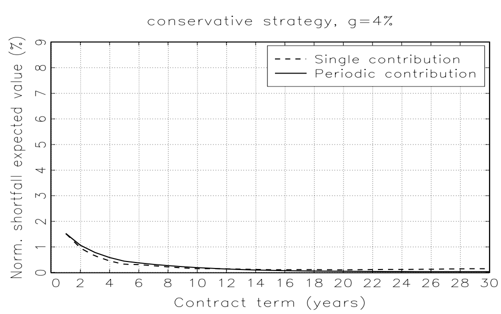
Moderate strategy





Conservative strategy

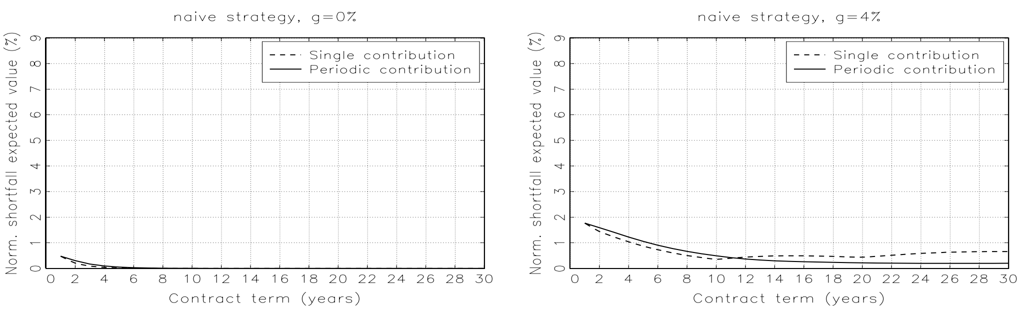




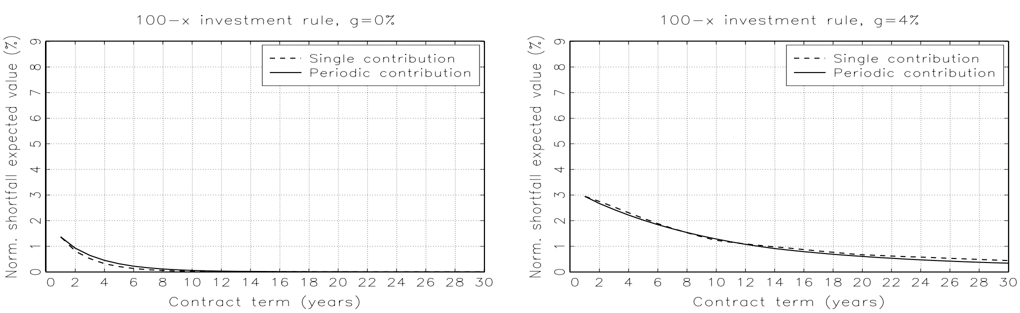
continued on the next page

continued from the previous page

Naive strategy



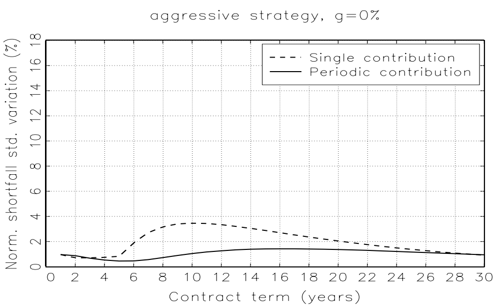
100-x investment rule

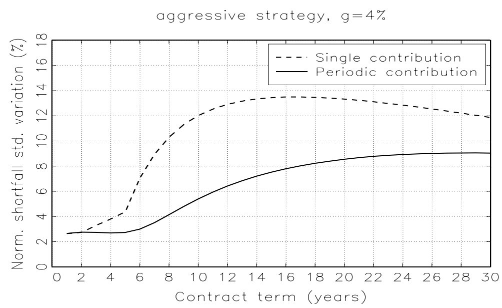


Note:
The figure depicts the impact of the payment scheme on the normalized guarantee cost using the example of life-cycle strategies. The solid line represents the periodic payment scheme (€1200 up-font yearly) and the dashed line the single premium case. The single premium is equal to the net present value of periodic contributions. The left column shows contracts with a 0% guarantee and the right column contracts with a 4% guarantee. The x-axis shows the time/contract maturity in years and the y-axis – the normalized guarantee cost in per cent.

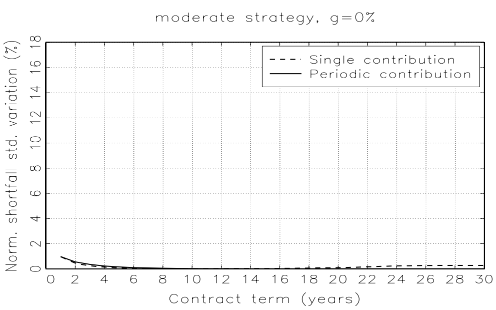
Figure M.3: Impact of the contribution payment scheme on the normalized shortfall standard deviation (life-cycle strategies)

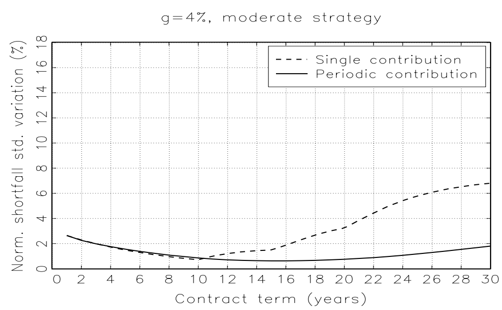
Aggressive strategy



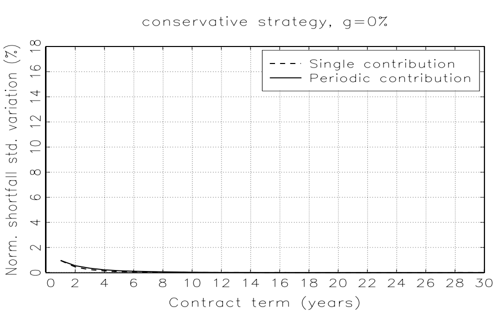


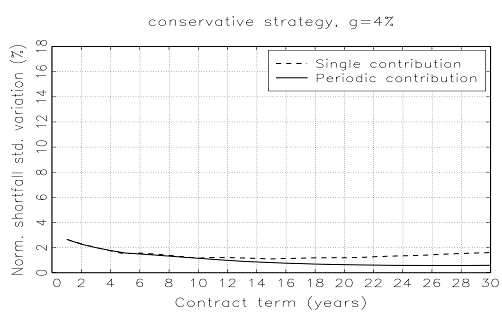
Moderate strategy





Conservative strategy

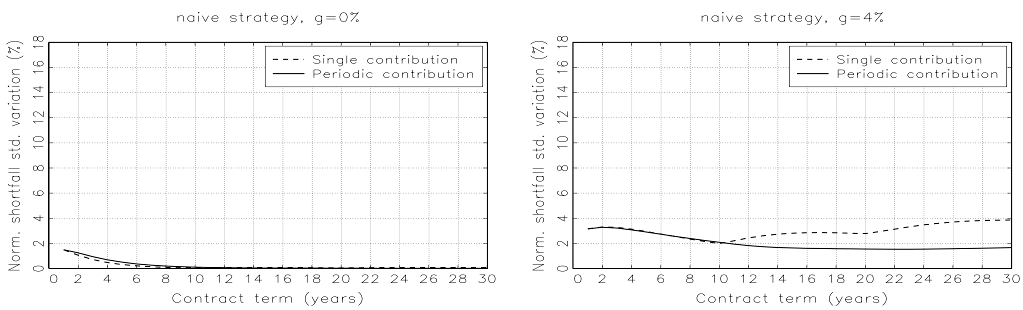




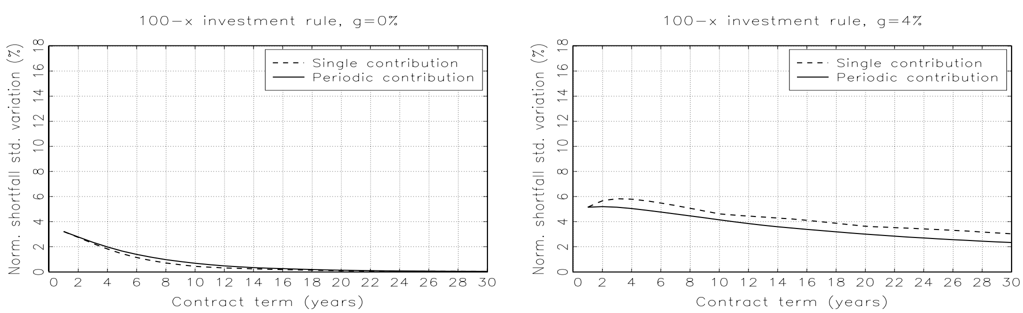
continued on the next page

continued from the previous page

Naive strategy



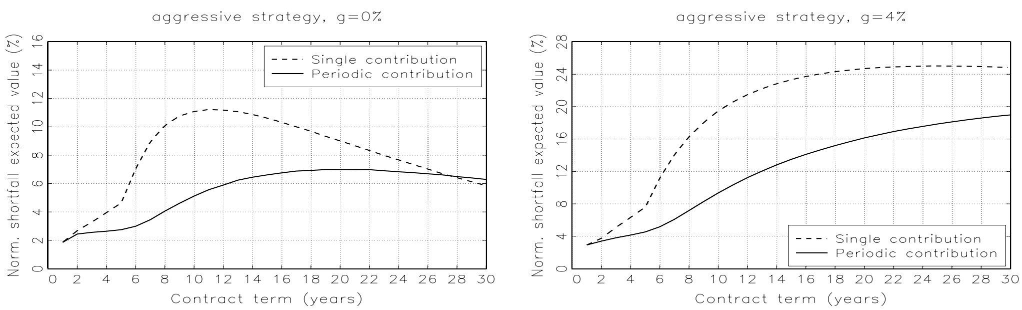
100-x investment rule



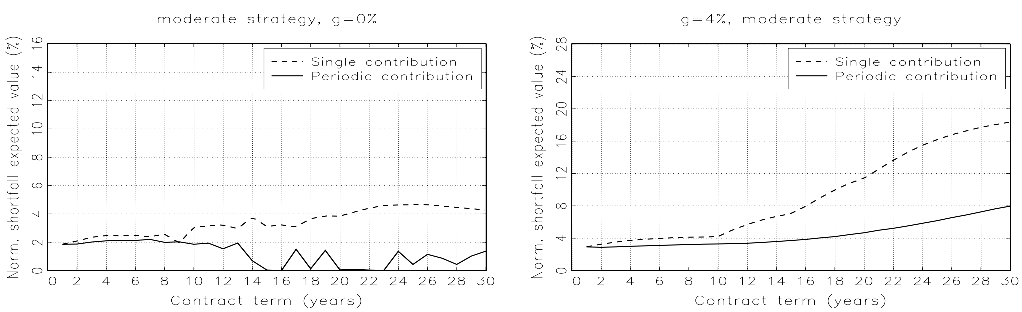
Note:
The figure depicts the impact of the payment scheme on the normalized guarantee cost using the example of life-cycle strategies. The solid line represents the periodic payment scheme (€1200 up-font yearly) and the dashed line the single premium case. The single premium is equal to the net present value of periodic contributions. The left column shows contracts with a 0% guarantee and the right column contracts with a 4% guarantee. The x-axis shows the time/contract maturity in years and the y-axis – the normalized guarantee cost in per cent.

Figure M.4: Impact of the contribution payment scheme on the normalized mean excess loss (life-cycle strategies)

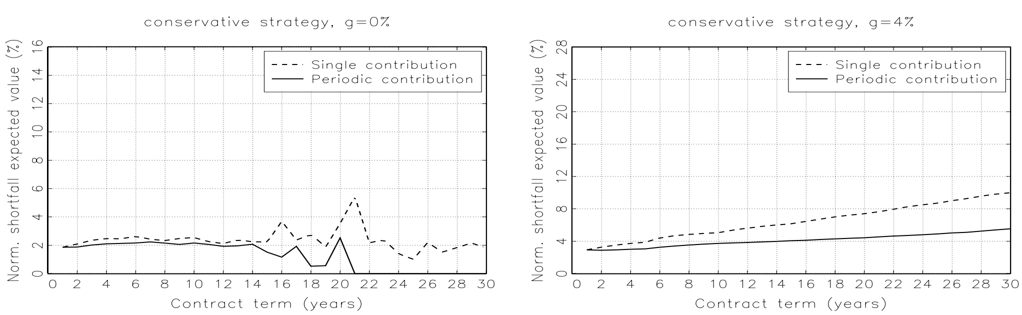
Aggressive strategy



Moderate strategy



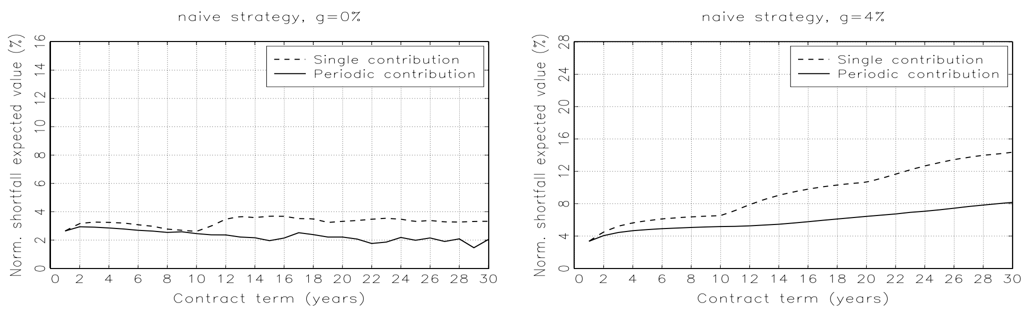
Conservative strategy



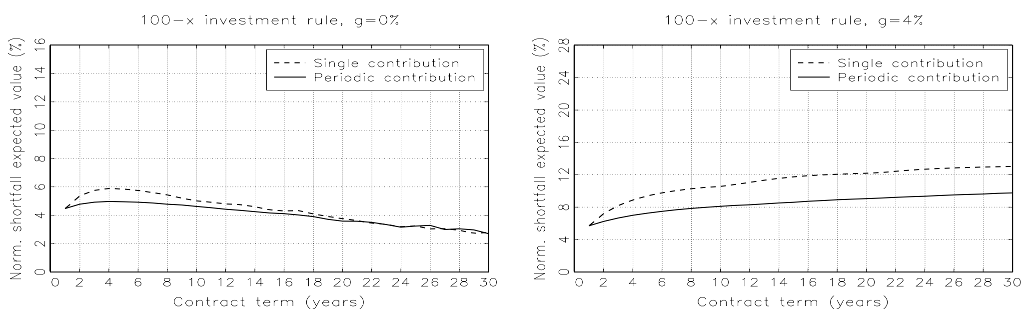
continued on the next page

continued from the previous page

Naive strategy



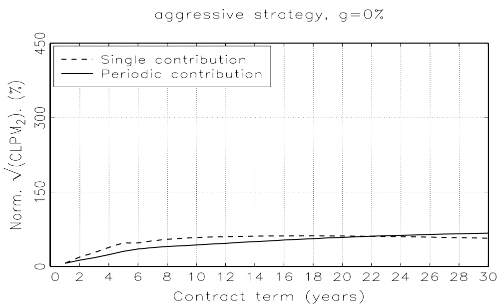
100-x investment rule

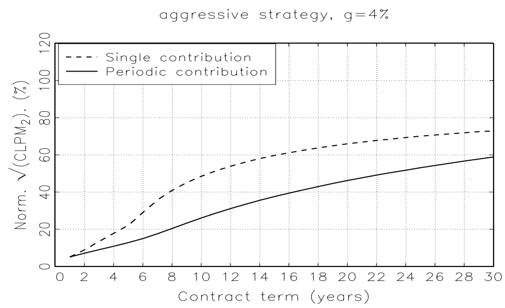


Note:
The figure depicts the impact of the payment scheme on the normalized guarantee cost using the example of life-cycle strategies. The solid line represents the periodic payment scheme (€1200 up-font yearly) and the dashed line the single premium case. The single premium is equal to the net present value of periodic contributions. The left column shows contracts with a 0% guarantee and the right column contracts with a 4% guarantee. The x-axis shows the time/contract maturity in years and the y-axis – the normalized guarantee cost in per cent.

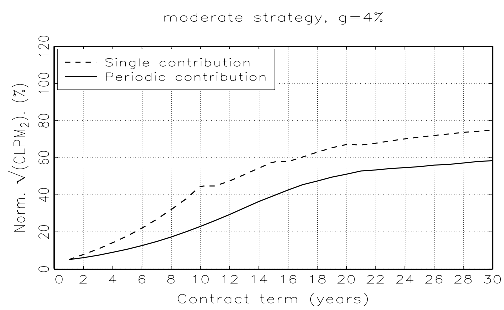
Figure M.5: Impact of the contribution payment scheme on the normalized conditional shortfall standard deviation (life-cycle strategies)

Aggressive strategy

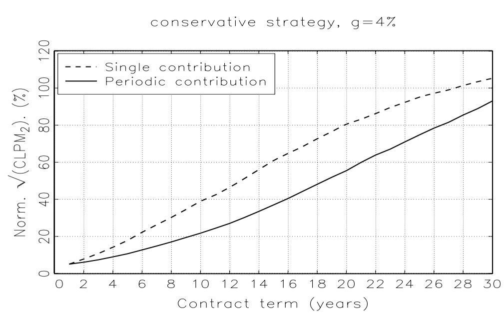




Moderate strategy



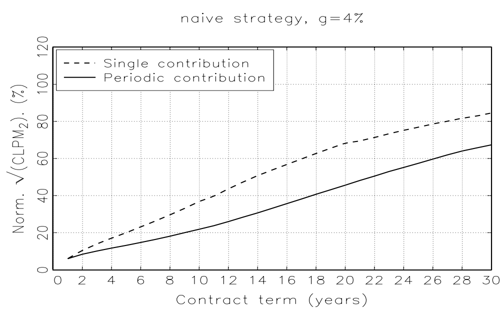
Conservative strategy



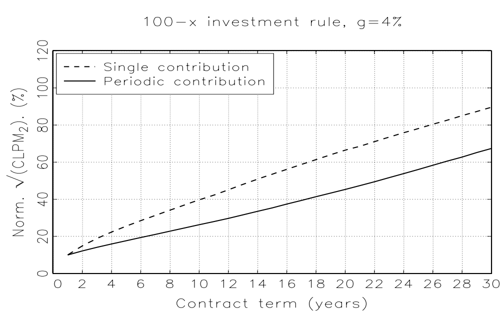
continued on the next page

continued from the previous page

Naive strategy



100-x investment rule



Note:
The figure depicts the impact of the payment scheme on the normalized guarantee cost using the example of life-cycle strategies. The solid line represents the periodic payment scheme (€1200 up-font yearly) and the dashed line the single premium case. The single premium is equal to the net present value of periodic contributions. The left column shows contracts with a 0% guarantee and the right column contracts with a 4% guarantee. The x-axis shows the time/contract maturity in years and the y-axis – the normalized guarantee cost in per cent.

Bibliography

Acerbi, C. and R. Tasche (2002). On the coherence of expected shortfall. *Journal of Banking and Finance 26*, 1487–1503.

Ahrens, R. (1998). Prognose von Zinsvolatilitäten mit Regime-Switching-Modellen: Eine empirische Analyse des Euro-DM-Geldmarktes. *Kradit und Kapital* (Heft 3), 370–399.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csake (Eds.), *Proceeding of the Second International Symposium on Information Theory*, Budapest, pp. 267–281. Akademiai Kiado.

Albrecht, P. (1994). Shortfall returns and shortfall risk. In *Proceedings of the 4th International AFIR Colloquium*, Volume 1, pp. 87–110.

Albrecht, P. (2004). Risk measures. In J. L. Teugles and B. Sundt (Eds.), *Encyclopedia of Actuarial Science*, Volume 3, pp. 1493–1501. New York: John Wiley and Sons.

Albrecht, P. and T. Klett (2004). Referenzpunktbezogene risikoadjustierte Performancemaße: Theoretische Grundlagen. Mannheimer Manuskripte zu Risikotheorie, Portfolio Management und Versicherungswirtschaft.

Albrecht, P. and R. Maurer (2008). *Investment– und Risikomanagement.*

Modelle, Methoden, Anwendungen (3 ed.). Stuttgart: Schäfer-Poeschel Verlag.

Alizadeh, A. and N. Namikos (2004). A Markov regime switching approach for hedging stock indices. *The Journal of Futures Markets* (7), 649–674.

Anderson, W. J. (1991). *Continuous-Time Markov Chains. An Application-Oriented Approach*. Springer series in Statistics. Probability and its Applications. New York at. al.: Springer-Verlag.

Andrews, D. W. and W. Ploberger (1994, November). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica 62*(6), 1383–1414.

Arnold, L. (1973). *Stochastische Differentialgleichungen. Theorie und Anwendung*. München Wien: R. Oldenbourg Verlag.

Bachelier, L. (1900). Théorie de la spéculation. *Annales Scientifiques de École Normale Supérieure 17*(3), 21–86.

Bacinello, A. R. (1997). Pricing contingent-claims concerning the Italian pension plan. In *Proceedings of the 7th AFIR International Colloquium*, Volume 1, pp. 149–165. Working Paper, University of Trieste.

Bacinello, A. R. (2000). Valuation of contingent-claims characterising particular pension schemes. *Insurance: Mathematics and Economics 27*, 171–188.

Barndorf-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory*. Chichester: Wiley.

Basel Committee on Banking Supervision (2001, January). *New Basel Capital Accord*. Basel: Bank for International Settlements.

Berndt, E., B. Hall, R. Hall, and J. Hausman (1974). Estimation and inference in nonlinear structural models. *Annals of Social Measurement 3*, 653–665.

Bingham, N. H. and R. Kiesel (2004). *Risk-Neutral Valuation. Pricing and Hedging of Financial Derivatives* (2 ed.). Springer Finance. London: Springer.

Björk, T. (2004). *Arbitrage Theory in Continuous Time* (2 ed.). Oxford: Oxford University Press.

Black, F. (1976). Studies in stock price volatility changes. In *Proceedings of the 1976 Meeting of the Business and Economic Statistics Section*, pp. 177–181. American Statistical Associtaion.

Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *Journal of Political Economy 81*, 637–654.

Bollen, N. P. (1998, Fall). Valuing options in regime-switching models. *The Journal of Derivatives 6*(1), 38–49.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*, 307–327.

Boyle, P. P. and E. S. Schwartz (1977, December). Equilibrium prices of guarantees under equity-linked contracts. *Journal of Risk and Insurance 44*(4), 639–660.

Breiman, L. (1968). *Probability*. Menlo Park: Addison-Wesley.

Brennan, M. J. and E. S. Schwartz (1976). The pricing of equity-linked life insurance policies with an asset value guarantee. *Journal of Financial Economics 3*, 195–213.

Brennan, M. J. and E. S. Schwartz (1979a). Alternative investment strategies for the issurers of equity linked life insurance with an asset value guarantee. *Journal of Business 52*, 63–93.

Brennan, M. J. and E. S. Schwartz (1979b). A continuous time approach to the pricing of bonds. *Journal of Banking and Finance 3*, 133–155.

Brennan, M. J. and E. S. Schwartz (1980). Analyzing convertible bonds. *Journal of Financial and Quantitative Analysis 15*(4), 907–929.

Brennan, M. J. and E. S. Schwartz (1982). An equilibrium model of bond pricing and a test of market efficiency. *Journal of Financial and Quantitative Analysis 3*, 301–329.

Brigo, D. and F. Mercurio (2006). *Interest Rate Models - Theory and Practice. With Smile, Inflation and Credit* (2 ed.). Berlin Heidelberg: Springer.

Cai, J. (1994, July). A Markov model of switching-regime ARCH. *Journal of Business and Economic Statisitcs 12*(3), 309–316.

Chan, K., G. A. Karolyi, F. A. Longstaff, and A. B. Sanders (1992, July). An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance XLVII*(3), 1209–1227. Papers and Proceedings of the Fifty-Second Annual Meeting of the American Finance Association, New Orleans, Louisianna, January 3-5, 1992.

Cootner, P. N. (1964). *The Random Character of Stock Market Prices.* Risk Books.

Cox, D. and H. Miller (1965). *The Theory of Stochastic Process.* London et. al.: Chapman & Hall.

Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985). A theory of the term structure of interest rates. *Econometrica 53*, 385–407.

Cox, S. and Y. Lin (2005). Securitization of mortality risk in life annuities.

Journal of Risk and Insurance 72, 227–252.

Crawford, G. W. and M. C. Fratantoni (2003). Assessing the forecasting performance of regime-switching, ARIMA and GARCH models of house prices. *Real Estate Economics 31*(2), 223–243.

Dahl, M. (2004). Stochastic mortality in life insurance: Market reserves and mortality linked insurance contracts. *Insurance: Mathematics and Economics 35*, 113–136.

Dahlquist, M. and S. F. Gray (2000). Regime-switching and interest rates in the european monetary system. *Journal of International Economics*, 399–419.

Davies, R. B. (1977, August). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika 64*(2), 247–254.

Davies, R. B. (1987, March). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika 74*(1), 33–43.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum Likelihood for Incomplete Data via EM Algorithm. *Journal of the Royal Statistical Society*, 149–192.

Einstein, A. (1905). *Eine neue Bestimmung der Moleküldimensionen*. Ph. D. thesis, Universität Zürich.

Elliott, R. J. and P. E. Kopp (2005). *Mathematics of Financial Markets* (2 ed.). Springer Finance. New York: Springer.

Engel, C. and J. D. Hamilton (1990, September). Long swings in the dollar: Are they in the data and do markets know it? *The American Economic Review* (4), 689–713.

Engle, R. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica 50*, 987–1008.

Esscher, F. (1932). On the probability function in the collective theory of risk. *Skandinavisk Aktuarietidskrift 15*, 175–195.

Feller, W. (1943). On a general class of contagious distributions. *Annals of Mathematical Statistics 14*, 389–400.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications* (2 ed.), Volume 2. New York: Willey.

Fischer, K. P. (1998). A discrete martingale model of pension fund guarantees in Colombia: Pricing and market effects. Centre de recherche en conomie et finance apliques (CRFA), Universit Laval, Qubec (Canada), Working Paper98-02.

Fischer, K. P. (1999). Pricing pension fund guarantees: A discrete martingale approach. *Canadian Journal of Administrative Sciences 16*(3), 256–266. Centre de recherche en conomie et finance apliques (CRFA), Universit Laval, Qubec (Canada), Working Paper98-02.

Fischer, T. (2003). Risk capital allocation by coherent risk measures based on one-sided moments. *Insurance: Mathematics and Economics 32*(1), 135–146.

Fishburn, P. C. (1984, April). Foundation of risk measurement. I. Risk as probable loss. *Management Science 30*(4), 396–406.

Föllmer, H. and M. Schweizer (1988). Hedging by sequential regression: An introduction to the mathematics of option trading. *ASTIN Bulletin 18*, 147–160.

Föllmer, H. and D. Sondermann (1986). Hedging of non-redundant contingent claims. In W. Hildenbrand and A-Mas-Cloell (Eds.), *Contributions* *to Mathematical Economics*, pp. 205–223.

Fong, H. and O. A. Vasiček (1991, Summer). Fixed income-volatility management. *Journal of Portfolio Management*, 41–46.

French, K. and R. Roll (1986). Stock return variances: Te arrival of information and the reaction of traders. *Journal of Financial Economics 17*, 5–26.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer Series in Statistics. New York: Springer.

Garcia, R. (1998, August). Asymptotic null distribution of the likelihood ratio test in Markov switching models. *International Economic Review 39*(9), 763–788.

Gerber, H. U. and E. S. W. Shiu (1994a, April 20-22). Martingale approach to pricing perpetual American options. In *Proceedings of the 4th International AFIR Colloquium*, Volume 2, Orlando, pp. 659–689.

Gerber, H. U. and E. S. W. Shiu (1994b). Option pricing by Esscher transforms. *Transactions of Society of Actuaries 46*, 99–191.

Gerber, H. U. and E. S. W. Shiu (1996). Actuarial bridges to dynamic hedging and option pricing. *Insurance: Mathematics and Economics 18*, 183–218.

Glosten, L., R. Jagannathan, and D. E. Runkle (1993, December). Relationship between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance 48*(5), 1779–1801.

Goldfeld, S. M. and R. E. Quandt (1972). *Nonlinear Methods in Econometrics*. Amsterdam: North-Holland Publishing.

Goldfeld, S. M. and R. E. Quandt (1973, March). A Markov model for switching regressions. *Journal of Econometrics*, 3–16.

Gourieroux, C. and J. Jasiak (2001). *Financial Econometrics. Problems, Models and Methods*. Princeton and Oxford: Princeton University Press.

Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 27–62.

Gründl, H., B. Nietert, and H. Schmeiser (2004). Zur Zusage der nominalen Kapitalerhaltung bei investmentfondsbasierten Riester-Produkten: Einige überlegungen aus finanztheoretischer Sicht. *Zeitschrift für Betriebswirtschaft 74*(2), 119–137.

Haas, M. (2004). *Dynamic Mixture Models for Financial Time Series*. Berlin: Pro BUSINESS.

Haas, M., S. Mittnik, and M. S. Paolella (2004). A new approach to markov-switching GARCH models. *Journal of Financial Econometrics 2*(4), 493–530.

Hamilton, J. D. (1988). Rational-expectations econometric analysis of changes in regime. *Journal of Economic Dynamics and Control*, 385–423.

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 357–384.

Hamilton, J. D. (1990). Analysis of Time Series Subject to Changes in Regime. *Journal of Econometrics*, 39–70.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Hamilton, J. D. (1996). Specification testing in Markov-switching time-series models. *Journal of Econometrics 70*, 127–157.

Hamilton, J. D. and G. Lin (1996). Stock market volatility and business cycle. *Journal of Applied Econometrics* (5), 573–593. Special Issue: Economic Forecasting, Sept.-Oct. 1996.

Hamilton, J. D. and R. Susmel (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, 307–333.

Hansen, B. (1992, December). The likelihood ratio test under nonstandard conditions: Testing the Markov switching model of GNP. *Journal of Applied Econometrics 7*, S61–S82. Supplement: Special Issue on Nonlinear Dynamics and Econometrics.

Hansen, B. E. (1996a, March-April). Erratum: The likelihood ratio test under nonstandard conditions: Testing the Markov switching model of GNP. *Journal of Applied Econometrics 11*(2), 195–198.

Hansen, B. E. (1996b, March). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica 64*(2), 413–430.

Hardy, M. R. (2001). A regime-switching model of long-term stock returns. *North American Actuarial Journal 5*(2), 41–53.

Hardy, M. R. (2003). *Investment Guarantees. Modeling and Risk Management for Equity-Linked Life Insurance*. Hoboken: John Willey & Sons.

Harrison, J. M. and S. R. Pliska (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications 11*, 215–260.

Heath, D., R. Jarrow, and A. Morton (1992). Bond pricing and the term structure of interest rates; a new methodology for contingent claims valuation. *Econometrica 60*(1), 77–105.

Hogan, W. W. and J. M. Warren (1974, January). Toward the development of an equilibrium capital-mark model based on semivariance. *The Journal of Financial and Quantitative Analysis 9*(1), 1–11.

Hull, J. and A. White (1990). Pricing interest-rate-derivative securities. *Review of Financial Studies 3*, 573–592.

Hull, J. C. (2006). *Options, futures, and other derivatives* (6 ed.). Upper Saddle River: Pearson Prentice Hall.

Itô, K. (1951). On stochastic differential equations. *Memoirs of the American Mathematical Society 4*, 1–51.

Jarque, C. M. and A. K. Bera (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters 6*(3), 255–259.

JP Morgan (1996, December). *Risk Metrics - Technical Document* (4 ed.). New York: JP Morgan.

Keating, C. and W. F. Shadwick (2002a). *An Introduction to Omega*. Finance Developement Centre.

Keating, C. and W. F. Shadwick (2002b). A universal performance measure. *Journal of Performance Measurement 6*, 59–84.

Kijima, M. (1997). *Markov Processes for Stochastic Modeling*. Stochasic Modeling Series. London at. al.: Chapman & Hall.

Kim, C.-J. (1993, July). Unobserved-component time series models with Markov-switching heteroscedasticity: Changes in regime and the link between inflation rates and inflation uncertainity. *Journal of Business and Economic Statisitcs 11*(3), 341–349.

Kim, C.-J. (1994). Dynamic Linear Models with Markov-Switching. *Journal of Econometrics*, 1–22.

Kim, C.-J. and C. R. Nelson (1999). *State-Space Models with Regime Switching. Classical and Gibbs-Sampling Approaches with Applications.* Cambridge, MA and London: The MIT Press.

Kling, A., J. Russ, and H. Schmeiser (2006). Analysis of embedded options in individual pension schemes in Germany. *Geneva Risk and Insurance 31*, 46–60.

Kraus, A. and M. Smith (1993, June). A simple multifactor term structure model. *Journal of Fixed Income*, 19–23.

Lachance, M.-E. and O. S. Mitchell (2003). Understanding individual account guarantees. In O. S. Mitchell and K. Smetters (Eds.), *The Pension Challenge. Risk Transfers and Retirement Income Security*, Chapter 8, pp. 159–186. Oxford: Oxford University Press.

Lam, P.-s. (1990). The Hamilton Model with a General Autoregressive Component: Estimation and Comparison with other Models of Economic Time Series. *Journal of Monetary Economics*, 409–432.

Lee, L.-F. and A. Chesher (1986). Specification testing when score test statistics are identically zero. *Journal of Econometrics*, 121–149.

Longstaff, F. A. and E. S. Schwartz (1992). Interest rate volatility and the term structure - a two-factor general equilibrium model. *Journal of Finance 47*, 1259–1282.

Longstaff, F. A. and E. S. Schwartz (2001). Valuing American options by simulation: a simple least-squares approach. *Review of Financial Studies 14*, 113–147.

Mandelbrot, B. B. (1963, October). The variation of certain speculative prices. *Journal of Buisness 36*, 394–419.

Mandelbrot, B. B. and R. L. Hudson (2004). *The (mis)Behavior of Markets. A Fractal View of Risk, Ruin, and Reward*. New York: Basic Book.

Margrabe, W. (1978). The value of an option to exchange one asset for another. *Journal of Finance 33*, 177–186.

Markowitz, H. M. (1952). Portfolio selection. *Journal of Finance 7*, 77–91.

Maurer, R. and C. Schlag (2003). Money-back guarantees in individual pension accounts: Evidence from the German pension reform. In O. S. Mitchell and K. Smetters (Eds.), *The Pension Challenge. Risk Transfers and Retirement Income Security*, Chapter 8, pp. 187–213. Oxford University Press.

Merton, R. C. (1973, Spring). Theory of rational option pricing. *Bell Journal of Economic and Management Science 4*(1), 141–183.

Musiela, M. and M. Rutkowski (2007). *Martingale Methods in Financial Modelling* (2 ed.). Number 36 in Stochastic Modelling and Applied Probability. Berlin and Heidelberg: Springer.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica 59*, 347–370.

Newey, W. K. (1985, September). Maximum likelihood specification testing and conditional moment tests. *Econometrica 53*, 1047–1070.

Norris, J. (1997). *Markov Chains*. Cambridge Series in Statistical and

Probabilistic Mathematics. Cambridge: Cambridge University Press.

Øksendal, B. (2003). *Stochastic Differential Equations. An Introduction with Application* (6 ed.). Universitext. Berlin: Springer.

Osborne, M. (1959, March-April). Brownian motion in the stock market. *Operations Research 7*(2), 145–173.

Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Phylosophical Transactions of the Royal Society of London A 185*, 71–110.

Pennacchi, G. (1999). The value of guarantees on pension fund returns. *The Journal of Risk and Insurance 66*(2), 219–237.

Piaskowski, W. (2005, September 6.-9.). Pricing and risk analysis of maturity guarantees embedded in the investment retirement products - Markov switching approach. In *Proceedings of the 15th International AFIR Colloquium*, Zurich.

Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association* (284 (Dec. 1985)), 873–880.

Rolski, T., H. Schmidli, V. Schmidt, and J. Teugels (1999). *Stochastic Processes for Insurance and Finance*. Wiley Series in Probability and Stochastics. Chichester et. al.: John Wiley & Sons.

Rossi, P., G. Allenby, and R. McCulloch (2005). *Bayesian Statistics and Marketing*. Chichester: Wiley.

Rydén, T., T. Teräsvirta, and S. Åsbrink (1998, May-June). Stylized facts of daily return and the hidden Markov model. *Journal of Applied Econometrics 13*(3), 217–244.

Schaefer, S. and E. Schwartz (1984). A two-factor model of the term-structure - an approximate analytical solution. *Journal of Financial and Quantitative Analysis 19*, 413–424.

Schönbucher, P. J. (2000). *Credit Risk Modeling and Credit Derivatives.* Ph. D. thesis, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics 6*(2), 461–464.

Shaked, M. (1980). On mixtures from exponential familes. *Journal of the Royal Statistical Society, Ser. B 42*, 192–198.

Sharpe, W. F. (1964, September). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance 19*(3), 425–442.

Shreve, S. E. (2004). *Stochastic Calculus for Finance II. Continuous-Time Models.* Springer Finance. New York: Springer.

Sin, Y. (2002, 3-5 July). Minimum pension guarantees. Introductory report. Moscow, Russian Federation. International Social Security Association. Seminar for Social Security Actuaries and Statisticians: Actuarial Aspects of Pension Reforms.

Sola, M. and A. Timmermann (1994, March). Fitting the moments: A comparison of ARCH and regime switching modells for daily stock returns. Discussion Paper. London: Centre for Economic Forecasting (London Business School).

Sortino, F. A. and S. Satchell (Eds.) (2001). *Managing Downside Risk in Financial Markets.* Axford: Butterworth-Heinemann.

Spiegelhalter, D., K. Abrams, and J. Myles (2003). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation.* Chichester: Wiley.

Szegö, G. (2002). Measures of risk. *Journal of Banking and Finance 26*(7), 1253–1272.

Tauchen, G. (1985, October/November). Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics 30*, 415–443.

Turner, C. M., R. Startz, and C. R. Nelson (1989). A Markov model of heteroskedasticity, risk, and learning in the stock market. *Journal of Financial Econometrics*, 3–22.

Turner, J. A. and D. M. Rajnes (2003). Retirement guarantees in voluntary defined contribution plans. In O. Mitchell and K. Smetters (Eds.), *The Pension Challenge. Risk Transfers and Retirement Income Security*, Chapter 12, pp. 251–267. Oxford: Oxford University Press.

Vasiček, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics 5*, 177–188.

Watson, M. W. and R. F. Engle (1985, May). Testing for regression coefficient stability with stationary AR(1) alternative. *The Review of Economics and Statistics* (2), 341–346.

Webb, M. A. (2003, April). *A Switching Black-Scholes Model and Option Pricing*. Ph. D. thesis, University of Adelaide, Department of Mathematics.

White, H. (1987). Specification testing in dynamic models. In T. F. Bewley (Ed.), *Advance in Econometrics. Fifth World Congress*, pp. 1–58. Cambridge: Cambridge University Press.

Yang, H. and G. Yin (2004). Ruin probability for a model under Markovian switching regime. In T. L. Lai, H. Yang, and S. Yung (Eds.), *Probability, Finance and Insurance*, pp. 206–217. Singapore: World Scientific.

Yin, G., Y. Liu, and H. Yang (2006). Bounds of ruin probability for regime-switching models using time scale separation. *Scandinavian Actuarial Journal*, 111–127.

Zakoïan, J. (1994). Threshold heteroskedastic functions. *Journal of Economic Dynamics and Control 18*, 931–955.