GARCH Models with Long Memory and Nonparametric Specifications

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
der Universität Mannheim

vorgelegt von

Christian Conrad

Juni 2006

Dekan:	Prof. Konrad Stahl, Ph.D.
Referent:	Prof. Dr. Enno Mammen
Korreferent:	Prof. Dr. Menelaos Karanasos
Tag der mündlichen Prüfung:	28. Juli 2006

Acknowledgements

First of all, I am very grateful to my supervisor, Enno Mammen, for his inspiring and encouraging advice. I very much benefited from numerous insightful discussions with him. He offered an excellent and exciting research environment at his Chair.

I am greatly indebted to my second advisor, Menelaos Karanasos. He stimulated my interest in long memory time series analysis while he was supervising my master thesis at the University of York and provided many valuable research ideas and suggestions.

Many thanks are due to my colleagues at the "Lehrstuhl für Statistik" who actively commented on my research: Berthold R. Haag, Stefan Hoderlein, Melanie Schienle and Kyusang Yu. Especially, I enjoyed the fruitful collaboration with Berthold. The interaction with him was not only productive but also a great pleasure. Moreover, I appreciate the helpful comments other academic researchers gave to my work. I would like to thank Karim M. Abadir, Julian Clough, James Davidson, Marc Endres, Liudas Giraitis, Leslie Godfrey, Oliver Grimm, Uwe Hassler, Marika Karanassou, Oliver Linton, Alexander Moutchnik and Peter Phillips.

I heavily benefited from the financial support by the Deutsche Forschungsgemeinschaft (DFG) and from the stimulating atmosphere within the Ph.D. Program of the graduate school CDSEM at the University of Mannheim.

Finally and most of all, I would like to thank my girlfriend, Oxana, and my parents who always supported and encouraged me during the last four years. They gave me the mental support, optimism, and retreat that enabled me to successfully finish my dissertation.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Some General Remarks on ARCH

Who could imagine 20 years ago, the flowering of research and applications that would develop around the ARCH model? It certainly was not an instant success. Years went by before anyone except my students and I wrote a paper on autoregressive conditional heteroscedasticity (ARCH).

Robert F. Engle (2002)

The above quote stems from an article by Robert F. Engle on "New frontiers for ARCH models" in 2002. At that time the ARCH model had become a story of success. One year later Robert F. Engle was awarded the Nobel Prize in economics. In his Nobel Lecture Engle describes the ARCH model as a logical consequence of the work of former Nobel Laureates. The empirical implementation of the Markowitz (1952) and Tobin (1958) theory on portfolio optimization, Sharpe's (1964) Capital Asset Pricing Model (CAPM) and the Black and Scholes (1973) and Merton (1973b) option pricing theory requires estimates of assets' *volatilities* and *co-volatilities*. Such estimates should reflect the stylized facts observed in almost every economic and financial time series: "unconditional distributions tend to be leptokurtic, variances change over time and large (small) changes tend to be followed by large (small) changes of either sign" (Bera and

Higgins, 1993, p. 306).

The Engle (1982) ARCH model not only replicates those stylized facts but also provides a theory of dynamic volatilities explaining "the apparent changes in the volatility of economic time series by a specific type of nonlinear dependence rather than by exogenous structural change of the variance" (Bera and Higgins, 1993, p. 315). Following the publication of the ARCH model numerous modifications and refinements led to the development of various ARCH-type models. The most well known modification, the generalized ARCH (GARCH) model, was suggested by Bollerslev (1986) – a graduate student of Engle. This specification allows for a more parsimonious parametrization of the conditional variance in comparison to the ARCH model, similar to the generalization of the autoregressive (AR) to the autoregressive moving average (ARMA) process. Some other influential models are Nelson's (1991) exponential GARCH (EGARCH), the GJR model of Glosten et al. (1993), the asymmetric power GARCH (APGARCH) of Ding et al. (1993), Engle and Lee's (1993) component GARCH (CGARCH) and the threshold ARCH (TARCH) of Zakoian (1994). However, these models are just a small section from the universe of existing ARCH specifications. In a recent review article Degiannakis and Xekalaki (2004) present more than thirty variants of the original ARCH specification. Of course, many of these models have multivariate extensions. For a review article on multivariate GARCH see Bauwens et al. (2006).

Apart from the development of new specifications which capture more and more features of the observed data, there has been intensive research in identifying the theoretical properties of those competing models. A summary of recent theoretical results on GARCH models can be found in Li et al. (2002). We just refer to some of the well known articles. Nelson (1990) and Bougerol and Picard (1992) established conditions for the stationarity and ergodicity of the GARCH process. Lee and Hansen (1994) as well as Lumsdaine (1996) proved the consistency and asymptotic normality of the quasi-maximum likelihood estimator for the GARCH$(1,1)$. Ling and McAleer (2002a,b) derived conditions for the existence of moments in the GARCH(p,q) and He and Teräsvirta (1999a,b) and Karanasos (1999) obtained formulas for the theoretical autocorrelation function in the GARCH(p, q) model.

Despite all these modifications and refinements, it is the simple GARCH($1, 1$) specification which is still most often used in financial applications. As pointed out by Engle (2004, p. 330) "it is remarkable that one model can be used to describe the volatility dynamics of almost any financial return series". This is probably also the main reason why the model has become so popular among practitioners. Today GARCH predicted volatilities are widely used for the pricing of financial derivatives, portfolio selection, and measuring and managing investment risk. Two comprehensive reviews of GARCH models and their applications in economics and finance are provided by Bollerslev et al. (1992) and Bera and Higgins (1993).

In the following we explain the motivation behind two popular classes of GARCH models in more detail. Those two types of GARCH models will be the focus of the subsequent chapters of this thesis.

1.1.1 Long Memory GARCH Models

Long memory models were introduced into the econometrics literature by Granger (1980), Granger and Joyeux (1980) and Hosking (1981). While short memory times series models are characterized by rapidly decaying autocovariances, it is the central feature of long memory models that their autocovariances decay slowly and are not summable. The class of fractionally integrated autoregressive moving average (ARFIMA) models relaxed the "knife-edge distinction between I(0) and I(1) processes" as imposed by the stationary ARMA model and the nonstationary integrated ARMA (ARIMA) model which were apparently too restrictive to match with the observed features of the data (Baillie et al., 1996a, p. 4). While the ARMA and the ARIMA model essentially assume a known degree of memory, namely the order of integration (zero or one) which reduces the series to short memory, the ARFIMA model allows for a fractional order of integration and estimates this order from the data. Initially long memory models were used as a tool to capture the apparent high degree of persistence in the levels of many macroeconomic variables. Around the mid-1990s research interest shifted to models which could allow for long memory in the conditional second moment of a time series. Studies by Taylor (1986), Ding et al. (1993), Ding and Granger (1996) among others had revealed that there is significant evidence of long memory in the empirical autocorrelations of nonlinear transformations such as the absolute or squared observations of many financial times series. This evidence conflicted with the standard Bollerslev (1986) GARCH model which implies exponentially decaying autocorrelations of the squared innovations as well as with the Engle and Bollerslev (1986) integrated GARCH (IGARCH) model which is characterized by complete persistence of shocks to the conditional variance and hence volatility forecasts which increase linearly with the time horizon. In analogy to the extension of the ARMA and the ARIMA model to the ARFIMA model, Baillie et al. (1996a) introduced the fractionally integrated GARCH (FIGARCH) model. The FIGARCH model allows for fractional orders of integration between zero and one, and implies hyperbolically decaying impulse response weights. Endowed with this additional flexibility the FIGARCH proved to be successful in modelling the long-run features in the volatility of many time series such as asset returns (Bollerslev and Mikkelsen, 1996), exchange rates (Tse, 1998) and inflation rates (Baillie et al. 2002). Although the FIGARCH model shares many of its properties with the ARFIMA model, the analogy is not complete. One of the drawbacks of the FIGARCH model is that its unconditional variance does not exist and so the innovation process is not covariance stationary. An alternative specification, the long memory GARCH (LMGARCH) model was proposed by Karanasos et al. (2004b).[1] In contrast to the FIGARCH model, the LMGARCH combines the properties of long memory and covariance stationarity.

For an up-to-date overview of other GARCH specifications allowing for long memory and of theoretical findings on long memory GARCH processes see Giraitis et al. (2005). Excellent surveys on long memory in economic and financial time series are provided by Baillie (1996) and Henry and Zaffaroni (2003).

[1] In the following we use the term long memory GARCH to refer either to the whole class of GARCH models which obey long memory or to the specific LMGARCH model of Karanasos et al. (2004b).

1.1.2 GARCH-in-Mean Models

Many economic theories predict a relationship between the level of a macroeconomic or financial variable and its conditional second moment. A typical example from finance is the approximate linear relationship between the conditional expected excess return and the conditional variance of the market portfolio implied by Merton's (1973a) Intertemporal CAPM (ICAPM). Examples from macroeconomics are the relationships between inflation and output growth and their uncertainties.

The GARCH-in-Mean (GARCH-M) model by Engle et al. (1987) was developed to explicitly capture the effect of the conditional variance – modelled by the usual GARCH equation – on the conditional mean. The model specifies the conditional mean as a monotonic function of the conditional variance. In this way the model incorporates what is usually referred to as a risk premium. Most commonly, the functional form of the risk premium is assumed to be linear or logarithmic in the conditional variance or standard deviation. In some cases such a choice can be justified by economic theory while in other cases it is simply a matter of convenience.

While previous studies on the existence of risk premia focused on constant risk premia, it is the advantage of the GARCH-M that it allows to "test for and estimate a time varying risk premium" (Bera and Higgins, 1993, p. 347). The GARCH-M model helped to establish significant relations between the conditional first and second moments of stock returns (French et al., 1987), output growth (Caporale and McKiernan, 1996) and inflation rates (Grier and Perry, 2000) etc.

Recently, Linton and Perron (2003) argued that the functional form of the risk premium commonly assumed is much too restrictive. There is no general reason to believe that the risk premium is linear or logarithmic in the conditional variance or standard deviation. Therefore, they proposed a semiparametric GARCH-M model with parametric conditional second but nonparametric conditional first moment. In this model the shape of the risk premium is estimated by nonparametric smoothing methods. The attractiveness of using nonparametric regression techniques in this context is given by fact that they do not require assumptions about the functional form of the risk premium apart from certain smoothness conditions. Therefore, the complexity of the model will be determined completely by the data, i.e. "one lets the data speak" and avoids the subjectivity in selecting a specific parametric model.

1.2 Outline of the Thesis

This section outlines the structure of the thesis and briefly summarizes the subsequent chapters and their main results. The thesis addresses two major topics which have recently received considerable attention in the financial econometrics literature: (i) long memory GARCH models and (ii) GARCH-M models with nonparametric specifications.

The following chapters represent a collection of five research articles. Each chapter is self-contained and can be read independently. The thesis is organized in three parts. Part I (Chapters 2 and 3) deals with theoretical aspects of long memory GARCH models, while Part II (Chapters 4 and 5) is concerned with empirical applications of those models. Part III (Chapter 6) is devoted to a specification test for the parametric GARCH-M model.

Chapter 2 is the joint work with my colleague Berthold R. Haag and was published in the *Journal of Financial Econometrics*. Chapters 3 to 5 have been written in collaboration with my second supervisor Prof. Dr. Menelaos Karanasos. The corresponding articles were published in *Economics Letters*, *Japan and the World Economy* and *Studies in Nonlinear Dynamics & Econometrics*. The last chapter of the thesis was written jointly with my first supervisor Prof. Dr. Enno Mammen. In the following we briefly describe the content of each chapter.

Chapter 2 is concerned with the FIGARCH(p, d, q) model of Baillie et al. (1996a) introduced in Section 1.1.1. Although this model has been intensively used in empirical applications, several theoretical properties of the model have not yet been fully understood. An import aspect in specifying a valid FIGARCH model is that the parameters of the process have to be chosen such that the nonnegativity of the conditional variance is guarantied. This problem is not specific to the FIGARCH model, but applies to the GARCH model as well. Nelson and Cao (1992) identified that the problem can be approached by investigating the so-called ARCH(∞) representation of the process. This representation expresses the conditional variance as an infinite sum of weighted lagged squared residuals. The weights in this sum – usually referred to as ARCH(∞) coefficients – are functions of the parameters of the underlying GARCH process. For the process to be well defined in the sense that the conditional variance is nonnegative almost surely for all points in time, it must be ensured that all ARCH(∞) coefficients are nonnegative. Nelson and Cao (1992) derived necessary and sufficient conditions for the parameters of the GARCH(p, q) model with $p \leq 2$ and sufficient conditions for the general model. No such conditions were available for the FIGARCH model, apart from a sufficient condition for the FIGARCH($1, d, 1$) provided by Bollerslev and Mikkelsen (1996). In Chapter 2 we extend the results of Nelson and Cao (1992) to the FIGARCH model, i.e. we derive conditions on the parameters of the FIGARCH(p, d, q) process which are necessary and sufficient for $p \leq 2$ and sufficient for $p > 2$ to guarantee the nonnegativity of all the ARCH(∞) coefficients. The availability of such conditions is of great importance for any researcher estimating FIGARCH models and in particular when using the parameter estimates to construct volatility forecasts for option pricing or value at risk computations. We illustrate this by an empirical application of the FIGARCH($1, d, 1$) model to Japanese Yen vs. US Dollar exchange rate data. A graphical representation of the necessary and sufficient set for the $(1, d, 1)$ model illustrates that our results dramatically enlarge the feasible parameter set compared to the set given by the sufficient conditions of Bollerslev and Mikkelsen (1996). Moreover, our results reveal two remarkable properties of the FIGARCH model which contrast sharply with the GARCH model: first, even if all parameters are nonnegative the conditional variance can become negative and, second, even if all parameters are negative (apart from the fractional differencing parameter d) the conditional variance can almost surely be nonnegative. These two observations highlight the importance of our results, because they imply that – independent of the sign of the estimated parameters – nonnegativity conditions should always be verified in the FIGARCH model.

In Chapter 3 we turn to the LMGARCH(p, d, q) process of Karanasos et al. (2004b). The persistence of economic shocks is usually measured by looking at the long-run effect of an innovation on the level of the series. Similarly, impulse response functions can be used to measure the persistence of shocks to the conditional variance. Baillie et al. (1996) derived an explicit expression for the impulse response function of the LMGARCH$(1, d, 0)$. In Chapter 3 we extend their results by deriving convenient representations for the impulse response function of the general LMGARCH(p, d, q) model. As special cases the corresponding impulse response functions of the GARCH(p, q) and the IGARCH(p, q) model can be obtained by restricting d to zero or one. We then use our results to compare the persistence of shocks to the conditional variance in these three GARCH specifications. As an empirical illustration we estimate several LMGARCH specifications on a long time series of Deutschmark vs. US Dollar exchange rate returns. The empirical example demonstrates the practical implications of our results. The impulse response functions can be used to distinguish between short and long memory specifications and to compare the persistence implied by alternative LMGARCH specifications.

Chapters 4 and 5 are concerned with the empirical application of dual long memory models, i.e. models which allow for long memory in the conditional mean and the conditional variance. In a first step, ARFIMA-FIGARCH models are applied to analyze the dynamics of European and international inflation data. In a second step, Granger methods are used to test several hypotheses concerning the causal relationship between inflation, nominal uncertainty and output growth. In this respect, the two chapters closely follow the original motivation of the ARCH model, since in his Nobel Lecture Engle (2004, p. 327) recalls that he "was looking for a model that could assess the validity of a conjecture of Friedman (1977) that the unpredictability of inflation was a primary cause of business cycles".

In Chapter 4 we use parametric models of long memory in both the conditional mean and the conditional variance of inflation and monthly data for the USA, Japan and the

UK for the period 1962–2000 to examine the relationship between inflation and inflation uncertainty. For the USA and the UK we provide evidence of long memory in the first and second conditional moment of the inflation rate, for Japan in the second conditional moment only. Using the impulse response functions derived in Chapter 3, we illustrate the importance of taking into account long memory in the second conditional moment by comparing the effect of shocks to the conditional variance of the US inflation rate in GARCH, IGARCH and FIGARCH specifications. In all countries, inflation significantly raises inflation uncertainty as predicted by Friedman. Increased nominal uncertainty affects inflation in Japan and the UK but not in the same manner, while no effect is found for the USA. The results from Japan support the Cukierman and Meltzer (1986) hypothesis, i.e. higher inflation uncertainty causes higher average inflation rates. In the UK uncertainty surrounding the future inflation appears to have a mixed impact on inflation.

In Chapter 5 we analyze the inflation dynamics of nine countries belonging to the European Monetary Union and of the UK. We first estimate the two main parameters driving the degree of persistence in inflation and its uncertainty using a dual long memory process. For all ten European inflation rates we detect the property of persistence in both their first and their second conditional moments. Then we investigate the possible existence of heterogeneity in inflation dynamics across Euro area countries and examine the link between nominal uncertainty and macroeconomic performance measured by the inflation and output growth rates. Strong evidence is provided for the hypothesis that increased inflation raises nominal uncertainty in all countries. However, we find that uncertainty surrounding future inflation has a mixed impact on output growth. This result brings out an important asymmetry in the transmission mechanism of monetary policy in Europe in addition to the difference in the economic sizes of the countries. We also investigate whether one can find a correlation between central bank independence and inflation policy. Our conclusion is that the most independent central banks are in countries where inflation falls in response to increased uncertainty.

Chapter 6 deals with the GARCH-M model of Engle et al. (1987). In this model the risk premium is usually assumed to be either a linear or logarithmic function of the conditional variance or conditional standard deviation (see Section 1.1.2). As in Linton and Perron (2003), we question whether this narrow class of functions is appropriate to all fields of applications of the GARCH-M model. As an alternative to the parametric specification of the risk premium we suggest estimating the shape of risk premium by nonparametric smoothing techniques. While Linton and Perron (2003) only visually compare the parametric and the nonparametric regression fits, we go a step further and test whether the two curves are significantly different from each other. Therefore, we propose a specification test for the functional form of the conditional mean in the GARCH-M model. The test statistic is based on the L_2-distance between a parametric estimate of the mean function and a nonparametric estimate. Since the conditional variance is unobservable a nonparametric fit of the mean function is not readily available in this setting. We suggest a nonparametric estimate obtained via an iterative estimation procedure which employs a fitted conditional variance series as a regressor replacing the unobserved conditional variance. Although the asymptotic distribution of the test statistic is shown to be normal, we suggest approximating the distribution by bootstrap resampling. Monte-Carlo simulations show that the bootstrap approximates the distribution of the test statistic under the null hypothesis reasonably well in finite samples. Under the alternative, the test statistic reveals good power properties. The availability of such a test is of great importance since many economic theories suggest relations between macroeconomic or financial variables and their conditional second moments. The suggested test procedure provides a formal framework for testing such theories. We illustrate the usefulness of the method by testing the linear risk-return relation predicted by the ICAPM. Using monthly as well as daily return data on the CRSP we cannot reject the hypothesis that the market excess return is a linear function of its conditional variance with positive slope parameter.

Part I

Long Memory GARCH Models: Theoretical Results

Chapter 2

Inequality Constraints in the Fractionally Integrated GARCH Model

2.1 Introduction

The empirical relevance of long memory conditional heteroscedasticity, which was initially addressed in the work of Ding et al. (1993) and Ding and Granger (1996), has emerged in a variety of studies of economic and financial time series. By now it is a widely accepted stylized fact that the empirical autocorrelation functions (ACFs) of the squared or absolute values of many macro and financial variables are characterized by a very slow decay indicating long memory and persistence.

The linear ARCH (LARCH) by Robinson (1991) was the first model permitting for long memory in the conditional variance. Subsequently, many researchers have proposed extensions of GARCH-type models which can produce long memory behavior. The

---

This chapter was published as: Conrad, C., and B. R. Haag (2006). "Inequality constraints in the fractionally integrated GARCH model." *Journal of Financial Econometrics* 4, 413–449.

fractionally integrated GARCH (FIGARCH) by Baillie et al. (1996a) can definitely be considered as the most established model among those. It proved to be suitable to handle the typical data features in many empirical applications (see, for example, Bollerslev and Mikkelsen, 1996, Beine and Laurent, 2003, Conrad and Karanasos, 2005a,b [see Chapter 4 and 5]). Alternative specifications were suggested by Davidson (2004), Giraitis et al. (2004), Karanasos et al. (2004b) and Zaffaroni (2004). Recent research has been aimed at a better understanding of the properties of these well established models, for instance Karanasos et al. (2004b) derive convenient representations for the ACF of the squared values of long memory GARCH (LMGARCH) processes, while in a related study Conrad and Karanasos (2006) [see Chapter 3] derive expressions for the impulse response function (IRF) of the LMGARCH model. For an up-to-date overview of theoretical findings on long memory GARCH processes see Giraitis et al. (2005). Finally, Baillie (1996) and Henry and Zaffaroni (2003) provide excellent surveys of major econometric work on long memory processes and their applications in economics and finance.

As in the Bollerslev (1986) GARCH model conditions on the parameters of the FIGARCH model have to be imposed to ensure the nonnegativity of the conditional variance. Originally, Bollerslev (1986) imposed conditions on the parameters of the GARCH(p,q) model which were sufficient to ensure the nonnegativity of the conditional variance, but these conditions simply required the nonnegativity of all parameters in the conditional variance specification.

Nelson and Cao (1992) showed that the restrictions imposed by Bollerslev (1986) can be substantially relaxed. By investigating the ARCH(∞) representation of the process they derived necessary and sufficient conditions for the GARCH(p,q) model with $p = 1$ or 2 and sufficient conditions for the general model. In particular, some of the parameters are allowed to have a negative sign. This is important since empirical findings (see Nelson and Cao, 1992, and the references therein) suggest that for many financial time series typically the parameter associated with the second lag of the squared innovation in the GARCH specification has a negative sign. The Bollerslev (1986) conditions rule out this case and thereby unnecessarily limit the flexibility of the model. This is nicely illustrated by He and Teräsvirta (1999c) who showed that for the GARCH(p, q) model with $\max\{p, q\} = 2$ these weaker conditions imply richer shapes of the ACF of the squared residuals.

An easy way to guarantee the nonnegativity of the conditional variance in the GARCH(p, q) with $p \leq 2$ is therefore firstly to estimate the unrestricted model, and then to validate the Nelson and Cao (1992) conditions only in case that there are parameter estimates with a negative sign. By now the Nelson and Cao (1992) conditions are implemented in econometric packages such as the financial analysis package for GAUSS, PcGive, S-Plus, Rats and G@RCH.

To validate whether a set of parameters suffices for the nonnegativity of the conditional variance in the FIGARCH(p, d, q) is substantially more difficult. In contrast to the GARCH model, it is possible that (i) the conditional variance becomes negative although all the parameters are positive, and (ii) the conditional variance is nonnegative a.s. (almost surely) for all t although all the parameters are negative (apart from d). These two observations imply that – independent of the sign of the estimated parameters – nonnegativity conditions should always be verified.

Bollerslev and Mikkelsen (1996) provide sufficient conditions for the parameters of the FIGARCH($1, d, 1$) model. These conditions are validated in programs such as the G@RCH package for Ox developed by Laurent and Peters (2002). Since these conditions are only sufficient there exist parameter values for which the conditions are violated, but still the conditional variance will be nonnegative almost surely. No conditions (not even sufficient) are available for higher order models. Bollerslev and Mikkelsen (1996), p. 159:

> *Of course, for the FIGARCH(p, d, q) model to be well-defined and the conditional variance positive almost surely for all t, all the coefficients in the infinite ARCH representation must be nonnegative.*

In this chapter we derive necessary and sufficient conditions for the FIGARCH(p, d, q) model of orders up to $p = 2$ and sufficient conditions for the general (p, d, q) model, which reduce an infinite number of inequalities to a finite number. Once the parameters are estimated one can easily validate these conditions. The results for the $(1, d, 1)$ specification which is used most often in empirical applications are discussed in detail. We illustrate graphically how the necessary and sufficient conditions dramatically enlarge the feasible parameter set compared to the set given by the sufficient conditions provided by Bollerslev and Mikkelsen (1996). For models of higher order ($p \geq 3$) we derive sufficient constraints which require only mild conditions on the parameters of the process. However, in practical applications one will rarely have to make use of a specification with $p > 2$. We provide an efficient algorithm for computing the coefficients in the ARCH(∞) representation which can be used if the sufficient conditions are violated. Plotting the sequence of coefficients indicates whether the conditional variance can become negative or not.

The availability of these inequality constraints is of importance for any researcher estimating FIGARCH models and in particular when utilizing parameter estimates to obtain volatility forecasts which are then employed for e.g. long term option pricing or value at risk computations.

An empirical example illustrates the importance of our results. For Japanese Yen vs. US Dollar exchange rate data we estimate a FIGARCH$(1, d, 1)$ model using the G@RCH package for Ox. The parameter estimates clearly fail to satisfy the Bollerslev and Mikkelsen (1996) conditions, which would lead any researcher relying on these conditions to reject the model. The set of parameters does however satisfy the necessary and sufficient conditions derived in this chapter, and hence guarantees the nonnegativity of the conditional variance.

We should mention that the conditions derived in this chapter also apply to the LM-GARCH since the coefficients in the ARCH(∞) representations of the FIGARCH and the LMGARCH coincide. Moreover, the results directly extend to the multivariate constant correlation FIGARCH and to the fractionally integrated autoregressive conditional duration (FIACD) model proposed by Jasiak (1998) which requires the nonnegativity of the conditional duration time.

Chapter 2 is organized as follows. Section 2.2 sets out the model of interest, assumptions and notation. In Section 2.3 we derive the necessary and sufficient conditions for the nonnegativity of the conditional variance in the FIGARCH(p, d, q) process. Section 2.4 discusses the empirical example. In the conclusions we suggest future developments. All proofs are deferred to the appendix.

2.2 The Fractionally Integrated GARCH Model

Following Robinson (1991) and Zaffaroni (2004) we define an ARCH(∞) process $\{\varepsilon_t, t \in \mathbb{Z}\}$ by the equations

$$\varepsilon_t = Z_t \sqrt{h_t}, \tag{2.1}$$

where $\{Z_t,\ t \in \mathbb{Z}\}$ is a sequence of independent and identically distributed random variables with $\mathbf{E}(Z_t) = 0$, $\sigma_Z^2 = \mathbf{E}(Z_t^2) < \infty$, and

$$h_t = \widetilde{\omega} + \sum_{i=1}^{\infty} \psi_i \varepsilon_{t-i}^2. \tag{2.2}$$

The parameter $\sigma_Z^2 \in \mathbb{R}^+$ was introduced Zaffaroni (2004) and relaxes the assumption that $\mathbf{E}(Z_t^2) = 1$ which is common in the GARCH literature. A major issue in specifying a valid ARCH(∞) process is to guarantee the nonnegativity of the conditional variance a.s. for all t. For this to hold it must be assumed that $\widetilde{\omega} \geq 0$ and $\psi_i \geq 0$ for all $i \geq 1$.[1] Now, define $v_t = \varepsilon_t^2 - \sigma_Z^2 h_t$ which is, by construction, a martingale-difference sequence with respect to the filtration generated by $\{\varepsilon_s, s \leqslant t\}$. Let $\Psi(L) = \sum_{i=1}^{\infty} \psi_i L^i$ with L being the lag operator, then ε_t^2 can be represented as

$$[1 - \sigma_Z^2 \Psi(L)]\varepsilon_t^2 = \sigma_Z^2 \widetilde{\omega} + v_t. \tag{2.3}$$

From equation (2.1) we have $\mathbf{E}[\varepsilon_t] = 0$, $\mathbf{Cov}(\varepsilon_t, \varepsilon_{t-j}) = 0$ for $j \geq 1$ and by equation (2.3) $\mathbf{E}[\varepsilon_t^2] = (\sigma_Z^2 \widetilde{\omega})/(1 - \sigma_Z^2 \Psi(1))$. Therefore, it follows that the covariance stationarity of ε_t in the ARCH(∞) model requires $\widetilde{\omega}$, $\sigma_Z^2 \in \mathbb{R}^+$ and $\sigma_Z^2 \Psi(1) < 1$.

[1]Requiring that $\psi_i \geq 0$ for all i implies that $\mathbf{P}(h_t < \widetilde{\omega}) = 0$. Hence $\widetilde{\omega}$ is the lower bound for the conditional variance. Note, that the same statement holds for the GARCH(p, q) (see Nelson and Cao, 1992).

In the following we explain how the Baillie et al. (1996a) FIGARCH and the Karanasos et al. (2004b) LMGARCH relate to the ARCH(∞) model given by equations (2.1) and (2.2) under specific assumptions on $\widetilde{\omega}$, σ_Z^2 and for certain finite parameterizations of $\Psi(L)$.

Baillie et al. (1996a) introduce the FIGARCH(p,d,q) model by assuming $\sigma_Z^2 = 1$ and defining ε_t^2 via the well known 'ARMA in squares' representation[2]

$$(1-L)^d \Phi(L)\varepsilon_t^2 = \omega + B(L)v_t, \tag{2.4}$$

for some $\omega \in \mathbb{R}^+$, $0 \leq d \leq 1$ and lag polynomials $\Phi(L)$, $B(L)$ defined as

$$\Phi(L) = 1 - \sum\nolimits_{i=1}^{q} \phi_i L^i \quad \text{and} \quad B(L) = 1 - \sum\nolimits_{i=1}^{p} \beta_i L^i.$$

The FIGARCH can be interpreted as a special case of equation (2.2) with

$$\widetilde{\omega} = \omega / B(1) \quad \text{and} \quad \Psi(L) = 1 - \frac{(1-L)^d \Phi(L)}{B(L)}.$$

For any $0 < d < 1$ the ψ_i coefficients will be characterized by a slow hyperbolic decay implying persistent impulse response weights (see Chapter 3). However, the Baillie et al. (1996a) specification with $0 < d < 1$ and $\sigma_Z^2 = 1$ is not compatible with the covariance stationarity of the ε_t, since in this case we have $\Psi(1) = 1$ and the above

[2] Baillie and Mikkelsen (1996a) alternatively proposed the fractionally integrated exponential GARCH (FIEGARCH) which specifies the logarithm of the conditional variance as a fractionally integrated process. This formulation allows to model the so-called leverage effect and nests the Nelson (1991) EGARCH as a special case when $d = 0$. Moreover, the conditional variance of the FIEGARCH is positive by construction and so no constraints on the parameters are required. A discussion of the moment and memory properties of the FIEGARCH can be found in Giraitis et al. (2005), p. 18. Despite the nice properties of the FIEGARCH, it is evident that the FIGARCH model is much more popular in empirical applications. One reason might be that the leverage effect is primarily a short run phenomenon. Therefore FIGARCH and FIEGARCH perform very similar in modelling the long-run features of e.g. stock market volatility. However, the FIEGARCH often encounters convergence problems in the estimation procedure due to the fact that the current conditional variance is a highly non-linear function of lagged conditional variances. Moreover, to our knowledge no distribution theory for the maximum likelihood estimator has been established even for the EGARCH with $d = 0$.

covariance stationarity condition is violated. For $0 < d < 1$ it is possible to obtain the covariance stationarity of the ε_t by assuming $\sigma_Z^2 < 1$ but as shown by Zaffaroni (2004), Theorem 2 and Remark 2.1, this implies absolute summability of the autocorrelation function (ACF) of the ε_t^2, ruling out long memory in ε_t^2.

The FIGARCH nests the Bollerslev (1986) GARCH model for $d = 0$. Then the condition $\sigma_Z^2 \Psi(1) < 1$ reduces to well known covariance stationarity condition for ε_t stated in Bollerslev (1986): $\sum_{j=1}^q \phi_j < 1$. This specification implies exponentially decaying coefficients ψ_i which lead to an absolutely summable exponentially decaying ACF of ε_t^2 and hence to a short memory process. On the other hand, the IGARCH model is obtained under the restriction $d = 1$. Then $\Psi(1) = 1$ and the model is again not covariance stationary.

A model which is closely related to the FIGARCH was suggested by Karanasos et al. (2004b). They define the LMGARCH(p, d, q) also by assuming $\sigma_Z^2 = 1$ but model the squared residuals in terms of deviations from $\omega \in \mathbb{R}^+$, i.e. by the equation

$$(1 - L)^d \Phi(L)(\varepsilon_t^2 - \omega) = B(L)v_t. \tag{2.5}$$

This small modification makes the LMGARCH being analogously defined to the ARFIMA model for the mean, and has important implications for the properties of ε_t. Equations (2.1) and (2.5) imply $\mathbf{E}[\varepsilon_t] = 0$, $\mathbf{Cov}(\varepsilon_t, \varepsilon_{t-j}) = 0$ for $j \geq 1$ and $\mathbf{E}[\varepsilon_t^2] = \omega < \infty$. This means that the LMGARCH specifies a covariance stationary ε_t process, although $\sigma_Z^2 = 1$ and $\Psi(1) = 1$ for any $0 < d < 1$. Moreover, the LMGARCH specification implies that the autocorrelations $\{\rho_m(\varepsilon_t^2), m = 1, 2, \ldots\}$ satisfy $\rho_m(\varepsilon_t^2) = O(m^{2d-1})$. Hence, provided that the fourth moment of the ε_t is finite, ε_t^2 exhibits long memory for all $0 < d < 0.5$, in the sense that the series $\sum_{m=0}^\infty |\rho_m(\varepsilon_t^2)|$ is properly divergent (see Karanasos et al., 2004b). In summary, the advantage of the LMGARCH compared to the FIGARCH model is that it combines the covariance stationarity of the ε_t with the long memory in the ε_t^2. The question whether the LMGARCH and/or the FIGARCH are strictly stationary or not is still open at present, see Giraitis et al. (2005), p. 11. The LMGARCH leads to an ARCH(∞) representation with $\widetilde{\omega} = 0$

and $\Psi(L) = 1 - (1-L)^d \Phi(L)/B(L)$.

Hence, both models obey ARCH(∞) coefficients generated from the expansion of

$$\Psi(L) = 1 - \frac{(1-L)^d \Phi(L)}{B(L)} = \sum_{i=1}^{\infty} \psi_i L^i.$$

In the following section we derive conditions on the parameters $(\beta_1, \ldots, \beta_p, d, \phi_1, \ldots, \phi_q)$ which guarantee that $\psi_i \geq 0$ for all $i \geq 1$. Since the ARCH(∞) coefficients are the same for the FIGARCH and the LMGARCH, our results hold for both models. Moreover, even if $\sigma_Z^2 \neq 1$ this will not affect the coefficients in the ARCH(∞) representation and so our results hold for an even broader class of ARCH(∞) models than FIGARCH and LMGARCH. Because of the predominant role played by the FIGARCH in the literature on empirical applications we state all the results in the following section in terms of this model.

Before we present our results we state further assumptions and introduce some more notation which we utilize in the proofs of all theorems. We assume that the inverse roots λ_i, $i = 1, \ldots, p$, of the polynomial $B(L)$ are real and $0 \neq |\lambda_i| < 1$ for $i = 1, \ldots, p$.[3] Additionally we assume that the roots of $\Phi(L)$ lie outside the unit circle and $\Phi(L)$ and $B(L)$ have no common roots.[4] The assumptions on the roots of $\Phi(L)$ and $B(L)$ imply that $\Phi(1) > 0$ and $B(1) > 0$.

The fractional differencing operator $(1-L)^d$ is most conveniently expressed in terms of the hypergeometric function $H(\cdot)$

$$(1-L)^d = H(-d, 1; 1; L) = \sum_{j=0}^{\infty} g_j L^j,$$

where the coefficients g_j are given by

$$g_j = f_j \cdot g_{j-1} = \prod_{i=1}^{j} f_i \qquad \text{with} \qquad f_j = \frac{j-1-d}{j} \quad \text{for } j = 1, 2, \ldots$$

[3]Our analysis does not cover complex roots in $B(L)$. Since we could not find any article in which a FIGARCH model was estimated with complex roots we expect the empirical relevance of this case to be rather small. However, most of the recursions we derive also hold for complex roots and so in principle it is possible the extend our results in this direction.

[4]If $\Phi(L)$ and $B(L)$ have common roots the FIGARCH process reduces to a model of lower order.

and $g_0 = 1$. Note, that $f_1 = -d < 0$, $f_2 = (1-d)/2 > 0$ and $f_j > 0$ for all $j > 2$ and hence $g_j < 0$ for all $j \geq 1$. It is easy to see that $f_j < f_{j+1}$ and $f_j \to 1$ as $j \to \infty$.

Furthermore, for $i > q \geq 0$ we define $F_i = -\sum_{l=0}^{q} \phi_l \prod_{j=l}^{q-1} f_{i-j}$ with $\phi_0 = -1$ and $\prod_{j=0}^{-1} = 1$, $F_i < F_{i+1}$ and $F_i \to 1 - \phi_1 - \ldots - \phi_q > 0$ as $i \to \infty$.

Let $(\lambda_{(1)}, \lambda_{(2)}, \ldots, \lambda_{(p)})$ be an ordering of the roots λ_i and define $\Lambda_r = \sum_{i=1}^{r} \lambda_{(i)}$, $r \leq p$. Hence, it follows that

$$F_i^{(r)} = \Lambda_r F_{i-1} + F_i f_{i-q} \to (\Lambda_r + 1)(1 - \phi_1 - \ldots - \phi_q) \tag{2.6}$$

and the limit is positive provided that $\Lambda_r > -1$.

2.3 Inequality Constraints for FIGARCH(p, d, q)

In this section we will derive the inequality constraints which are necessary and sufficient for the nonnegativity of the conditional variance in the FIGARCH(p, d, q) model with $p \leq 2$ and sufficient conditions for the general model. The inequality constraints provided in Bollerslev and Mikkelsen (1996) for the FIGARCH$(1, d, 1)$ are substantially relaxed. As a special case ($d = 0$) the results of Nelson and Cao (1992) can be obtained.

As mentioned before, the nonnegativity of the conditional variance requires that all ψ_i coefficients in the ARCH(∞) representation are nonnegative. In general this would mean imposing infinitely many inequality constraints on the ψ_i. By investigating the sequence for the different models we find that the infinite number of restrictions reduces to a finite number. This means that it suffices to check the nonnegativity of $\psi_1, \ldots, \psi_k$ to guarantee the nonnegativity of the conditional variance. To relate the ψ_i sequence to the parameters of the process we have to find convenient representations of the coefficients as functions of the parameters.

2.3.1 FIGARCH(1, d, q)

We begin with deriving the inequality constraints for the FIGARCH$(1, d, q)$ process. Then we discuss the empirically important examples of the $(1, d, 1)$, $(0, d, 1)$ and $(1, d, 0)$

model in detail.

Theorem 2.1. *The conditional variance of the FIGARCH($1, d, q$) is nonnegative a.s. iff*

Case 1: $0 < \beta_1 < 1$

1. $\psi_1, \ldots, \psi_{q-1} \geq 0$ *and*

2. *either* $\psi_q \geq 0$ *and* $F_{q+1} \geq 0$ *or for* $k > q + 1$ *with* $F_{k-1} < 0 \leq F_k$ *it holds that* $\psi_{k-1} \geq 0$.

Case 2: $-1 < \beta_1 < 0$

1. $\psi_1, \ldots, \psi_{q-1} \geq 0$ *and*

2. *either* $\psi_q \geq 0$, $\psi_{q+1} \geq 0$ *and* $F^{(1)}_{q+2} \geq 0$ *or for* $k > q + 2$ *with* $F^{(1)}_{k-1} < 0 \leq F^{(1)}_k$ *it holds that* $\psi_{k-1} \geq 0$ *and* $\psi_{k-2} \geq 0$.

In the proof of Theorem 2.1 we obtain an easily computable recursion for the ψ_i coefficients which can be used in practice to validate the requirements of the theorem for a given set of parameter estimates. It is clear that in the FIGARCH($1, d, q$) it suffices to check $q + 1$ conditions if $\beta_1 > 0$ and $q + 2$ conditions if $\beta_1 < 0$ to ensure the nonnegativity of the conditional variance for all t.

Because the FIGARCH($1, d, 1$) is definitely the most often used specification in empirical applications we intensively discuss the derivation of the corresponding inequalities and their interpretation. The ARCH(∞) representation of the FIGARCH($1, d, 1$) leads to the following recursions (see proof of Theorem 2.1) for the corresponding ψ_i coefficients:

$$\psi_1 = d + \phi_1 - \beta_1 \tag{2.7}$$

$$\psi_i = \beta_1 \psi_{i-1} + (f_i - \phi_1)(-g_{i-1}) \text{ for all } i \geq 2, \text{ and alternatively,} \tag{2.8}$$

$$\psi_i = \beta_1^2 \psi_{i-2} + [\beta_1(f_{i-1} - \phi_1) + (f_i - \phi_1) f_{i-1}](-g_{i-2}) \text{ for all } i \geq 3 \tag{2.9}$$

Corollary 2.1. *The conditional variance of the FIGARCH($1, d, 1$) is nonnegative a.s. iff*

Case 1: $0 < \beta_1 < 1$

either $\psi_1 \geq 0$ *and* $\phi_1 \leq f_2$ *or for* $k > 2$ *with* $f_{k-1} < \phi_1 \leq f_k$ *it holds that* $\psi_{k-1} \geq 0$.

Case 2: $-1 < \beta_1 < 0$

either $\psi_1 \geq 0$, $\psi_2 \geq 0$ *and* $\phi_1 \leq f_2(\beta_1 + f_3)/(\beta_1 + f_2)$ *or for* $k > 3$ *with* $f_{k-2}(\beta_1 + f_{k-1})/(\beta_1 + f_{k-2}) < \phi_1 \leq f_{k-1}(\beta_1 + f_k)/(\beta_1 + f_{k-1})$ *it holds that* $\psi_{k-1} \geq 0$ *and* $\psi_{k-2} \geq 0$.

This corollary can be derived from the recursions by the following considerations. First, note that $-g_i > 0$ for $i \geq 1$. The proof uses then the fact that $F_i = f_i - \phi_1$ and $F_i^{(1)} = \beta_1(f_{i-1} - \phi_1) + (f_i - \phi_1)f_{i-1}$ are increasing and that for both expressions there exists a k such that $F_{k-1} < 0 \leq F_k$ and $F_{k-1}^{(1)} < 0 \leq F_k^{(1)}$. For example, consider Case 1. If $\psi_1 \geq 0$ and $\phi_1 \leq f_2$ this implies $\phi_1 < f_i$ for all $i > 2$ and hence the nonnegativity of all ψ_i by equation (2.8). If $\phi_1 > f_2$, then there exists a k such that $\phi_1 \leq f_k$ and so ψ_{k-1} implies $\psi_i \geq 0$ for all $i \geq k$ because f_i is increasing. Also, $\psi_{k-1} \geq 0$ and $f_{k-1} < 0$ imply $\psi_i \geq 0$ for all $i \leq k-2$. Case 2 can be treated analogously using equation (2.9).

Next, we compare Corollary 2.1 with the already existing sufficient conditions for the FIGARCH($1, d, 1$) suggested in Baillie et al. (1996a), Bollerslev and Mikkelsen (1996) and Chung (1999). Baillie et al. (1996a), p. 22, provide the following sufficient constraints

$$0 \leq \beta_1 \leq \phi_1 + d \qquad \text{and} \qquad 0 \leq d \leq 1 - 2\phi_1$$

which are equivalent to $\psi_1 \geq 0$ and $F_2 \geq 0$. Alternatively, Bollerslev and Mikkelsen (1996), p. 159, state the inequality constraints

$$\beta_1 - d \leq \phi_1 \leq \frac{2-d}{3} \qquad \text{and} \qquad d\left[\phi_1 - \frac{1-d}{2}\right] \leq \beta_1(\phi_1 - \beta_1 + d).$$

which are equivalent to $\psi_1, \psi_2 \geq 0$ and $F_3 \geq 0$. Hence, these inequality constraints reflect the first condition in Case 1 of Corollary 2.1 or the arbitrary choice of $k = 3$

(again Case 1). The Bollerslev and Mikkelsen (1996) conditions are weaker than the Baillie et al. (1996a) conditions, but restrict $\phi_1 \leq f_3$. Note, that both stets of sufficient conditions do not cover Case 2 where $-1 < \beta_1 < 0$, since the corollary requires $F_3^{(1)} \geq 0$. Finally, Chung (1999) suggests a third set of sufficient constraints which is given by

$$0 \leq \phi_1 \leq \beta_1 \leq d < 1 \tag{2.10}$$

and provides two examples:

(i) $\phi_1 = 0.6$, $\beta_1 = 0.7$ and $d = 0.8$

(ii) $\phi_1 = 0.5$, $\beta_1 = 0.2$ and $d = 0.25$.

The first set of parameters satisfies equation (2.10) but not the Bollerslev and Mikkelsen (1996) conditions, while the second satisfies the Bollerslev and Mikkelsen (1996) conditions but not equation (2.10). Chung (1999), p. 18, concludes: "The examples show that there may be parameter values that cannot satisfy either set of sufficient conditions while still allow all ψ_i coefficients to be positive."

The corollary above provides necessary and sufficient conditions and thereby solves this problem. One can easily check that the parameters in both examples satisfy the conditions of Corollary 2.1. Moreover, in comparison to the Bollerslev and Mikkelsen (1996) sufficient conditions it widens the range of admissible parameters: (i) if $f_{k-1} < \phi_1 \leq f_k$ with $k > 3$ parameters can still be admissible and (ii) we allow for $\beta_1 < 0$.

Figure 2.1 illustrates how the inequality constraints from Corollary 2.1, Case 1, extends the sufficient set from Bollerslev and Mikkelsen (1996) to the necessary and sufficient set for two fixed values of d, i.e. for $d \in \{0.1, 0.9\}$ and $\phi_1 > 0$.[5] The set denoted B+M is given by the Bollerslev and Mikkelsen (1996) conditions, while the set denoted C+H is the area which is allowed for by Corollary 2.1, Case 1, but not by the Bollerslev and Mikkelsen (1996) conditions. The dashed line separating the two sets corresponds to $\phi_1 = f_3$. For a given value of d, f_3 is the upper bound for ϕ_1 in the Bollerslev and Mikkelsen (1996) conditions. The joint set, i.e. B+M $\cup$ C+H, is the necessary and

[5]Note that we exclude $\phi_1 = \beta_1$ by the assumption that $\Phi(L)$ and $B(L)$ have no common roots.

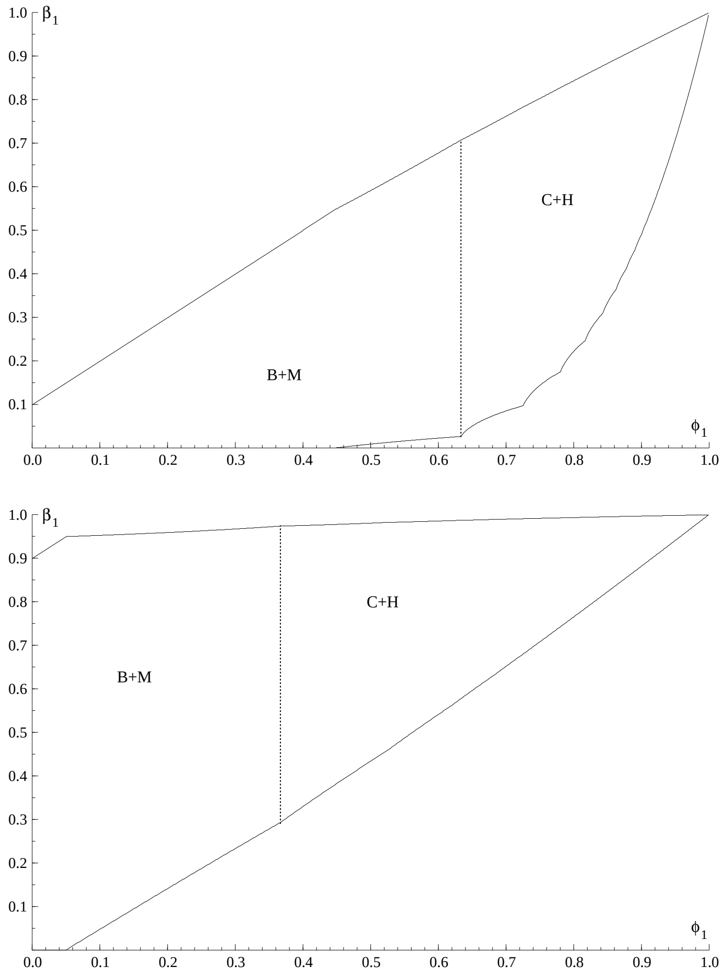

Figure 2.1: Necessary and sufficient parameter set for FIGARCH$(1, d, 1)$ (Case 1 and $\phi_1 > 0$) with $d = 0.1$ (upper) and $d = 0.9$ (lower).

sufficient one. The necessary and sufficient set given by Figure 2.2 covers Case 1 and 2 for $d = 0.3$ and $-1 < \phi_1 < 1$. As can be easily seen, Corollary 2.1 dramatically enlarges the set of parameter values which is allowed for and thereby allows for a greater flexibility in model specification. In particular, note that in contrast to the GARCH model the conditional variance of the FIGARCH can be nonnegative although $\phi_1 < 0$ and $\beta_1 < 0$ and on the other hand it can become negative although all parameters are positive. When d is approaching zero the parameter set described by Corollary 2.1 converges to the well known necessary and sufficient set for a GARCH$(1, 1)$ with parameters $\alpha_1 = \phi_1 - \beta_1$

and β_1. When d approaches one the FIGARCH$(1, d, 1)$ collapses to a GARCH$(1, 2)$ with parameters $\alpha_1 = 1 + \phi_1 - \beta_1$, $\alpha_2 = -\phi_1$ and β_1, which add to one. The admissible parameter set again coincides with the parameter set given by Nelson and Cao (1992) for this model.

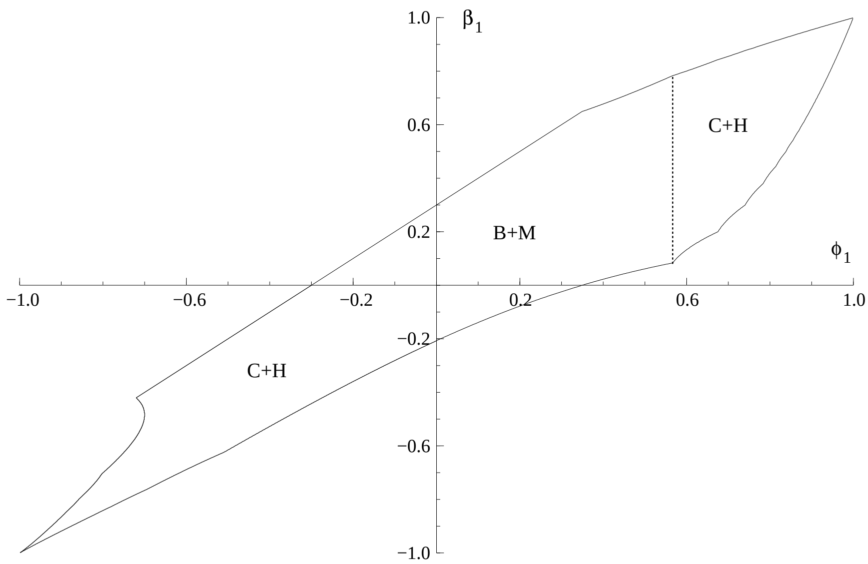

Figure 2.2: Necessary and sufficient parameter set for FIGARCH$(1, d, 1)$ with $d = 0.3$.

For the GARCH$(2, 2)$ model He and Teräsvirta (1999c) show that the Nelson and Cao (1992) necessary and sufficient conditions imply richer shapes of the ACF of the squared residuals compared to shapes implied by the Bollerslev (1986) sufficient conditions. They discover four possible types of ACFs which can be generated. Type 1 is characterized by a smooth monotonic decay from ρ_1 onwards, while type 2 reaches its peak at $\rho_2 > \rho_1$ with monotonic decay from ρ_2 onwards. The autocorrelations may be oscillating either with peak at ρ_1 or at ρ_2, which are the types 3 and 4. Expressions for the ACF of the squared residuals in the LMGARCH(p, d, q) were derived in Karanasos et al. (2004b).[6] While the long-run behavior of the ACF is governed by the fractional differencing parameter d, the short run behavior is determined by ϕ_1 and β_1. We plot

[6]Recall that in contrast to the FIGARCH the LMGARCH is covariance stationary. However, as pointed out by Karanasos et al. (2004b) both models have the same 'second-order structure'.

in Figure 2.3 the ACFs of a LMGARCH$(1, d, 1)$ with $d = 0.3$ and certain combinations of parameters ϕ_1 and β_1 lying in the necessary and sufficient set (see Figure 2.2). It is evident that even the LMGARCH$(1, d, 1)$ can generate all four types of ACFs described by He and Teräsvirta (1999c) for the GARCH$(2, 2)$. Interestingly, type 4 is generated by parameter values which do not lie in the Bollerslev and Mikkelsen (1996) sufficient set and we could not to find any combination of parameters with $\beta_1 > 0$ leading to this type. This finding suggests that the enlarged parameter set directly translates into an increased flexibility in characterizing the autocorrelation structure of the squared residuals.

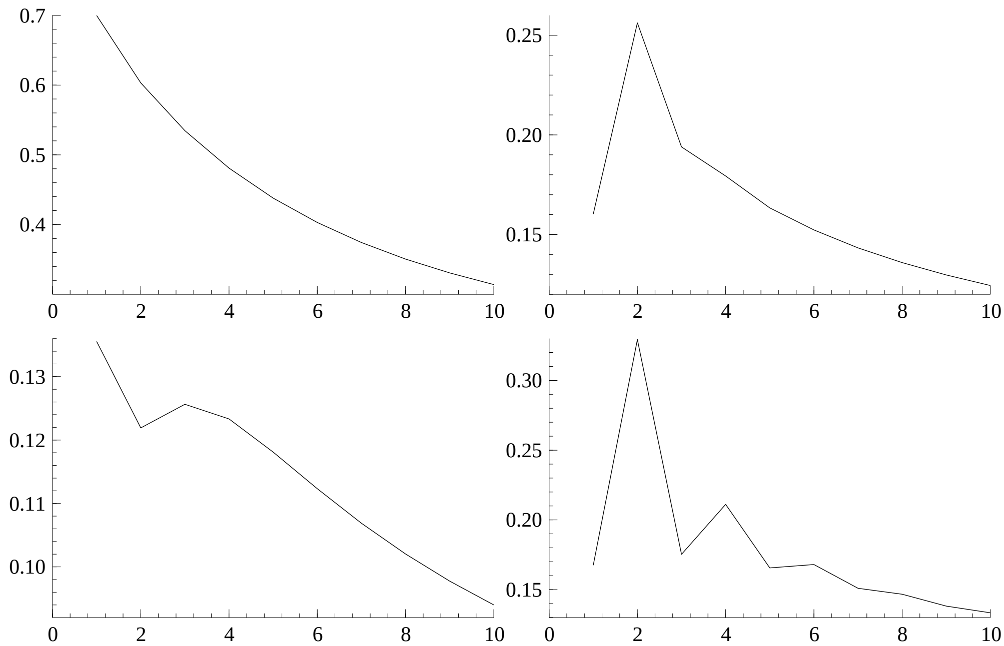

Figure 2.3: Four types of ACFs for LMGARCH$(1, d, 1)$ with $d = 0.3$. Type 1: $\phi_1 = 0.7$, $\beta_1 = 0.5$ (upper left), type 2: $\phi_1 = -0.2$, $\beta_1 = 0.05$ (upper right), type 3: $\phi_1 = 0.3$, $\beta_1 = 0.53$ (lower left) and type 4: $\phi_1 = -0.5$, $\beta_1 = -0.25$ (lower right).

Example 2.1. *Baillie et al. (2002) can serve as an example which illustrates the importance of our result. ARFIMA-FIGARCH*$(1, d, 1)$ *models are estimated to several inflation series. In Table 3, p. 507, the estimated parameters for the French inflation* *data are* $\widehat{\beta}_1 = 0.899, \widehat{d} = 0.331, \widehat{\phi}_1 = 0.859$. *Even though these parameters do not satisfy the Bollerslev and Mikkelsen (1996) conditions (clearly* $\widehat{\phi}_1 > \widehat{f}_3$*), the parameters are in accordance with the conditions given by Corollary 2.1, Case 1.*

Remark 2.1. *Corollary 2.1 can be directly applied to the bivariate constant correlation FIGARCH*$(1, d, 1)$ *model given by the equations*

$$\begin{aligned} h_{11,t} &= \frac{\omega_{11}}{1-\beta_{11}} + \left[1 - \frac{(1-\phi_{11}L)(1-L)^{d_1}}{(1-\beta_{11}L)}\right]\varepsilon_{1,t}^2 \\ h_{22,t} &= \frac{\omega_{22}}{1-\beta_{22}} + \left[1 - \frac{(1-\phi_{22}L)(1-L)^{d_2}}{(1-\beta_{22}L)}\right]\varepsilon_{2,t}^2 \\ \text{and} \quad h_{12,t} &= \rho\sqrt{h_{11,t}h_{22,t}}. \end{aligned}$$

Positive definiteness of the variance-covariance matrix is guaranteed if and only if $|\rho| < 1$ *and the parameters* $(\phi_{jj}, d_j, \beta_{jj})$ *satisfy the condition given in Corollary 2.1 for* $j = 1, 2$. *This model is used e.g. by Brunetti and Gilbert (2000) to investigate long memory in oil price data. Brunetti and Gilbert (2000) check the sufficient Bollerslev and Mikkelsen (1996) constraints for each equation.*

Remark 2.2. *Jasiak (1998) extends the Engle and Russel (1998) ACD*(p, q) *model to the FIACD*(p, d, q) *model, i.e. he assumes that the duration time* x_i *between the* i*-th and* $(i-1)$*-th event can be modelled as*

$$x_i = \delta_i Z_i \qquad \text{and} \qquad \delta_i = \frac{\omega}{1-\beta_1} + \left[1 - \frac{(1-\phi_1 L)(1-L)^d}{(1-\beta_1 L)}\right] x_i^2$$

with Z_i *being a sequence of i.i.d. random variables with expectation one and* δ_i *the conditional expectation of the* i*-th duration. Applied in this context, Corollary 2.1 ensures the nonnegativity of the conditional duration time* δ_i.

The FIGARCH$(1, d, 1)$ nests two interesting submodels: the FIGARCH$(1, d, 0)$ and the FIGARCH$(0, d, 1)$. Although Corollary 2.1 does not explicitly cover the cases with $\phi_1 = 0$ and $\beta_1 = 0$ they can be treated along the same lines of argumentation.

Corollary 2.2. *The conditional variance of the FIGARCH*$(0, d, 1)$ *is nonnegative a.s. iff*

1. $\psi_1 \geq 0 \Leftrightarrow d + \phi_1 \geq 0$

2. $F_2 \geq 0 \Leftrightarrow (1-d)/2 - \phi_1 \geq 0$

If $\beta_1 = 0$ the recursion given by equation (2.8) reduces to $\psi_i = (f_i - \phi_1)(-g_{i-1})$ for $i \geq 2$. We impose $\psi_1 \geq 0$. Recall that $-g_i > 0$ for $i \geq 1$. The nonnegativity of ψ_2 requires $f_2 - \phi_1 \geq 0$. Since f_i is increasing, $\psi_2 \geq 0$ implies $\psi_i \geq 0$ for all $i > 2$.[7]

Example 2.2. *Conrad and Karanasos (2005,b) [see Chapter 5] estimate ARFIMA-FIGARCH models for the inflation rates of ten European countries. For Belgium the preferred specification for the conditional variance is a FIGARCH(*$0, d, 1$*) with estimated parameters* $\widehat{d} = 0.330$ *and* $\widehat{\phi}_1 = -0.280$. *Since* $\widehat{d} > -\widehat{\phi}_1$ *and* $\widehat{\phi}_1 < 0$ *it follows from Corollary 2.2 that these parameters guarantee the nonnegativity of the conditional variance for all* t. *The autocorrelation structure implied by these parameters is of type 2.*

Corollary 2.3. *The conditional variance of the FIGARCH(*$1, d, 0$*) is nonnegative a.s. iff* [8]

Case 1: $0 < \beta_1 < 1$

$$\psi_1 \geq 0 \Leftrightarrow d - \beta_1 \geq 0$$

Case 2: $-1 < \beta_1 < 0$

$$\psi_2 \geq 0 \Leftrightarrow \left(d - \sqrt{2(2-d)}\right)/2 \leq \beta_1$$

If $\phi_1 = 0$ the recursions given by equations (2.8) and (2.9) reduce to $\psi_i = \beta_1\psi_{i-1} - g_i$ for $i \geq 2$ and $\psi_i = \beta_1^2\psi_{i-2} + (\beta_1 + f_i)(-g_{i-1})$ for $i \geq 3$. For $0 < \beta_1 < 1$ assuming $\psi_1 \geq 0$ together with $-g_i \geq 0$ for all $i \geq 1$ imply $\psi_i \geq 0$ for all i. For $-1 < \beta_1 < 0$ it is easy

[7] Similarly, from the proof of Theorem 2.1 it follows that the conditional variance of the FIGARCH($0, d, q$) is nonnegative a.s. iff *1.* $\psi_1, \ldots, \psi_q \geq 0$ and *2.* $F_{q+1} \geq 0$. Note, that $F_{q+1} \geq 0$ is trivially fulfilled if $\phi_i \leq 0$ for $i = 1, \ldots, q$.

[8] Note that in Baillie et al. (1996a), p. 11, only Case 1 was considered and $0 \leq \beta_1 < d \leq 1$ is stated as a necessary and sufficient condition for the conditional variance of the FIGARCH($1, d, 0$) model to be positive almost surely for all t.

to see that $\psi_2 \geq 0$ implies $F_2^{(1)} = \beta_1 + f_2 \geq 0$. Since f_i is increasing it follows that $F_i^{(1)} = \beta_1 + f_i > 0$ for all $i \geq 3$. Hence, $\psi_1 \geq 0$ (ensured by $\beta_1 < 0$) and $\psi_2 \geq 0$ imply $\psi_i \geq 0$ for all i.

Above we illustrated the consequences of our less severe parameter constraints on the shapes of the ACF for the LMGARCH$(1, d, 1)$. The implications of allowing for $\beta_1 < 0$ on the degree of persistence that can be modelled is now illustrated by considering the ACF and the impulse response function (IRF) of the LMGARCH$(1, d, 0)$. As an example we assume $d = 0.45$. According to Corollary 2.3, the range of values for β_1 which guarantee

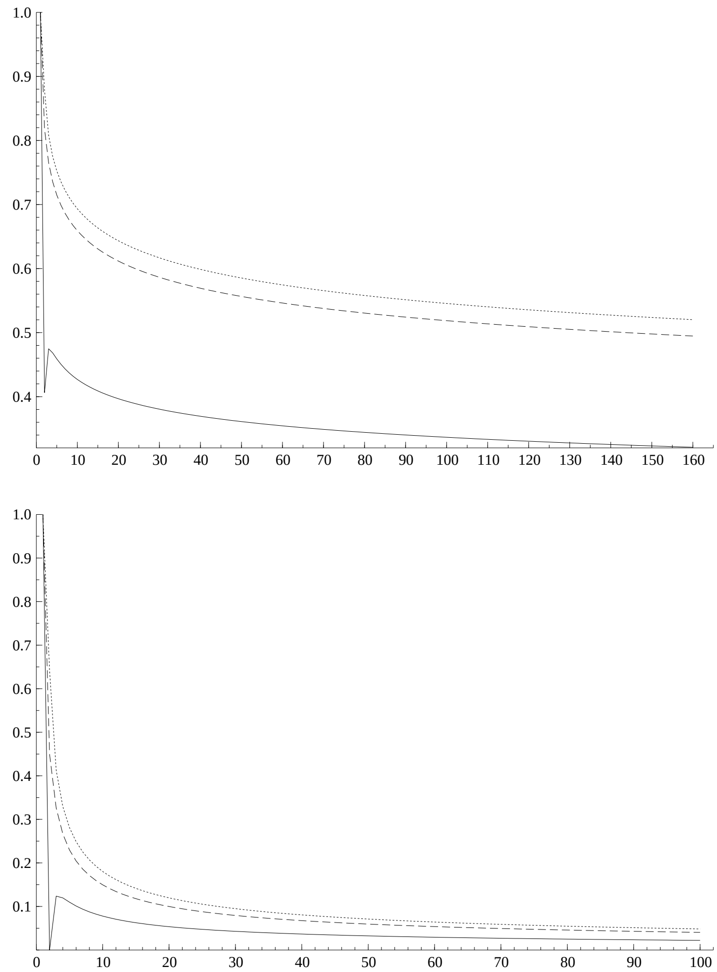

Figure 2.4: ACFs of ε_t^2 (upper) and IRFs (lower) for LMGARCH$(1, d, 0)$ with $d = 0.45$ and $\beta_1 = 0.45$ (solid), $\beta_1 = 0$ (dashed) and $\beta_1 = -0.1925$ (dotted), respectively.

the nonnegativity of the conditional variance is given by $-0.1925 \leq \beta_1 \leq 0.45$. The LMGARCH$(1, d, 0)$ model with the restriction $\beta_1 \geq 0$ can not produce ACFs in the area above the ACF with $\beta_1 = 0$ as can be seen from Figure 2.4 (upper), i.e. the degree of persistence is unnecessarily limited. Expressions for the IRF of the LMGARCH(p, d, q) have been obtained by Conrad and Karansos (2006) [see Chapter 3] . Figure 2.4 (lower) plots the IRF of the LMGARCH$(1, d, 0)$ with $d = 0.45$ and $\beta \in \{-0.1925, 0, 0.45\}$. Clearly, allowing for $\beta_1 < 0$ increases the flexibility of the IRF. In accordance with the result for the ACF, a negative β_1 increases the persistence of the process.

2.3.2 $\mathbf{FIGARCH(2, d, q)}$

Before we consider cases with $p \geq 2$ we derive a recursive representation of the $\{\psi_i\}$ sequence. Again, let $(\lambda_{(1)}, \ldots, \lambda_{(p)})$ be some ordering of the λ_i. We will make use of the representation in the proofs of the subsequent theorems.

Lemma 2.1. *The sequence $\{\psi_i, i = 1, 2, \ldots\}$ can be written as*

$$\psi_i = \psi_i^{(p)} \qquad \text{where}$$
$$\psi_i^{(r)} = \lambda_{(r)}\psi_{i-1}^{(r)} + \psi_i^{(r-1)} \qquad 1 < r \leq p,\ i \geq 1,$$

and the sequence of $\{\psi_i^{(1)}\}$ is given by

$$\psi_i^{(1)} = -c_i + \sum_{j=1}^{\min\{i,q\}} \phi_j c_{i-j} \qquad \text{for } i = 1, \ldots, q \quad \text{and with} \quad c_i = \sum_{j=0}^{i} \lambda_{(1)}^{i-j} g_j$$
$$\psi_i^{(1)} = \lambda_{(1)}\psi_{i-1}^{(1)} + F_i(-g_{i-q}) \qquad \text{for } i > q$$

with starting values $\psi_0^{(r)} = -1$, $r = 1, \ldots, p$.

Now, we turn to the case $p = 2$. Without loss of generality we assume that $\lambda_1 \geq \lambda_2$. No inequality constraints – not even sufficient – have been established for $p \geq 2$ in the literature on long memory GARCH models so far. We will firstly consider the FIGARCH$(2, d, 0)$ and then combine the results with those from the FIGARCH$(1, d, q)$.

Proposition 2.1. *The conditional variance of the FIGARCH($2, d, 0$) is nonnegative a.s. iff (recall that in this case: $\lambda_1 + \lambda_2 = \beta_1$ and $\lambda_1 \cdot \lambda_2 = -\beta_2$)*

Case 1: $1 > \lambda_1 \geq \lambda_2 > 0$, *i.e.* $\beta_1 > 0$, $\beta_2 < 0$

$$\psi_1 \geq 0 \Leftrightarrow d \geq \lambda_1 + \lambda_2 = \beta_1$$

Case 2: $1 > \lambda_1 > 0 > \lambda_2 > -1$ *and* $\lambda_1 \geq |\lambda_2|$, *i.e.* $\beta_1 \geq 0$, $\beta_2 > 0$

$$\psi_1 \geq 0 \text{ and } \psi_2 \geq 0$$

Case 3: $1 > \lambda_1 > 0 > \lambda_2 > -1$ *and* $\lambda_1 < |\lambda_2|$, *i.e.* $\beta_1 < 0$, $\beta_2 > 0$

Either if $\psi_2^{(1)} \geq 0 \Leftrightarrow \lambda_2(d - \lambda_2) + f_2 d \geq 0$ *or* $\psi_2, \psi_4, \ldots, \psi_{\overline{k}-2} \geq 0$, *where* $\overline{k} = \min_{\tilde{k} \text{ even}} \{\psi_{\tilde{k}}^{(1)} > 0\}$ *with* $\lambda_{(1)} = \lambda_2$ *and* $\lambda_{(2)} = \lambda_1$.

Case 4: $0 > \lambda_1 \geq \lambda_2 > -1$, *i.e.* $\beta_1 < 0$, $\beta_2 < 0$

$$\psi_2^{(1)} \geq 0 \text{ and } \psi_2 \geq 0 \text{ where } \lambda_{(1)} = \lambda_1 \text{ and } \lambda_{(2)} = \lambda_2.$$

Proposition 2.1, Case 1, states that the conditional variance can be nonnegative although $\beta_2 < 0$. Case 3 shows that with $p = 2$ we can allow for $\beta_1 < 0$ (in the GARCH($2, 2$) one needs at least $\beta_1 > 0$). Finally, Case 4 illustrates that in contrast to the GARCH model with $p = 2$, where at least one root must be nonnegative, in the FIGARCH with $p = 2$ we can allow for both roots being negative. Note however that this case will rarely appear for financial data in practise since estimating $\beta_1 < 0$ and $\beta_2 < 0$ is very unlikely. We are not aware of any application where such a parameter combination has been estimated.

Example 2.3. *Beine and Laurent (2003) estimate AR(1)-FIGARCH($2, d, 0$) models for the exchange rate of the Japanese yen, French franc and British pound against the US dollar. The parameter estimates for all three currencies presented in Table 2, p. 651, are such that Case 2 applies. Checking $\widehat{\psi}_1$ and $\widehat{\psi}_2$ immediately proves that all $\widehat{\psi}_i$ coefficients are nonnegative and thereby confirms the validity of the chosen model specifications.*

Theorem 2.2. *The conditional variance of the FIGARCH*$(2, d, q)$ *is nonnegative a.s. iff*

Case 1: $1 > \lambda_1 \geq \lambda_2 > 0$, *i.e.* $\beta_1 > 0$, $\beta_2 < 0$

1. $\psi_1 \geq 0$, *and for* $q \geq 2$: $\psi_1, \ldots, \psi_{q-1} \geq 0$
2. *either* $\psi_q^{(1)} \geq 0$ *and* $F_{q+1} \geq 0$ *or for* k *with* $F_{k-1} < 0 \leq F_k$ *either* $\psi_{k-1}^{(1)} \geq 0$ *or* $\psi_{\overline{k}-1} \geq 0$, *where*

$$\overline{k} = \underset{\widetilde{k}>k}{min}\{-\lambda_{(1)}\psi_{k-1}^{(1)} < \sum_{j=0}^{\widetilde{k}} F_{k+j}(-g_{k-q+j})\lambda_{(1)}^{-j}\}$$

and $\lambda_{(1)} = \lambda_1$, $\lambda_{(2)} = \lambda_2$.

Case 2: $1 > \lambda_1 > 0 > \lambda_2 > -1$, $\lambda_1 \geq |\lambda_2|$, *i.e.* $\beta_1 \geq 0$, $\beta_2 > 0$

1. $\psi_1 \geq 0$, *and for* $q \geq 2$: $\psi_1, \ldots, \psi_{q-1} \geq 0$
2. *either* $\psi_q \geq 0$ *and* $F_{q+1} \geq 0$ *or for* $k \geq q+2$ *with* $F_{k-1} < 0 \leq F_k$ *we have* $\psi_1, \ldots, \psi_{k-1} \geq 0$.

Case 3: $1 > \lambda_1 > 0 > \lambda_2 > -1$, $\lambda_1 < |\lambda_2|$, *i.e.* $\beta_1 < 0$, $\beta_2 > 0$

1. $\psi_1 \geq 0$, *and for* $q \geq 2$: $\psi_2, \ldots, \psi_{q-1} \geq 0$
2. *either* $\psi_q^{(1)} \geq 0$, $\psi_{q+1}^{(1)} \geq 0$ *and* $F_{q+2}^{(1)} \geq 0$ *or for* k *with* $F_{k-1}^{(1)} < 0 \leq F_k^{(1)}$ *either* $\psi_{k-1}^{(1)} \geq 0$ *and* $\psi_{k-2}^{(1)} \geq 0$ *or* $\psi_{q+1}, \ldots, \psi_{\overline{k}-1} \geq 0$, *where*

$$\begin{aligned}\overline{k} \;\; = \;\; & \underset{\widetilde{k}>k}{min}\{-\lambda_{(1)}^2\psi_{k-2}^{(1)} < \sum_{j=0}^{\widetilde{k}} F_{k+2j}^{(1)}(-g_{k-q+2j-1})\lambda_{(1)}^{-2j}, \\ & -\lambda_{(1)}^2\psi_{k-1}^{(1)} < \sum_{j=0}^{\widetilde{k}} F_{k+2j}^{(1)}(-g_{k-q+2j-1})\lambda_{(1)}^{-2j}\}\end{aligned}$$

and $\lambda_{(1)} = \lambda_2$, $\lambda_{(2)} = \lambda_1$.

Case 4: $0 > \lambda_1 \geq \lambda_2 > -1$, $\lambda_1 + \lambda_2 > -1$, *i.e.* $-1 < \beta_1 < 0$, $\beta_2 < 0$

1. $(\lambda_1, d, \phi_1, \ldots, \phi_q)$ *satisfy the conditions for the FIGARCH*$(1, d, q)$

2. $\psi_1, \ldots, \psi_{k-1} \geq 0$, *where* $k > q + 2$ *is s.t.* $F_{k-1}^{(2)} < 0 \leq F_k^{(2)}$

Case 5: $0 > \lambda_1 \geq \lambda_2 > -1$, $\lambda_1 + \lambda_2 \leq -1$ *i.e.* $\beta_1 \leq -1$, $\beta_2 < 0$

1. $(\lambda_{(1)}, d, \phi_1, \ldots, \phi_q)$ *satisfy the conditions for the FIGARCH(*$1, d, q$*)*

2. There exists a $\bar{k}$ *such that* $S_{\bar{k},k} > 0$, *where* k *such that* $F_{k-1}^{(1)} < 0 \leq F_k^{(1)}$ *and*

$$S_{j,i} = \lambda_{(1)} \Lambda_2 \sum_{l=1}^{j} \lambda_{(1)}^{2(i-l)} F_{i+2l}^{(1)}(-g_{i-q+2l-1}) + F_{i+2j}^{(2)}(-g_{i-q+2j-1})$$

3. $\psi_1, \ldots, \psi_{\bar{k}} \geq 0$

Again, note that Case 4 and 5 are not of empirical interest, and in particular Case 5 which implies $\beta_1 < -1$. Moreover, Theorem 2.2 illustrates that with increasing p the number of cases which have to be analyzed grows exponentially. As the conditions are already quite complex when $p = 2$ the analysis becomes even worse when $p \geq 3$.

Remark 2.3 (Relation to GARCH(p, q)). *Since the simple GARCH model is nested within the FIGARCH we treat it as a special case. If we set* $d = 0$, *we obtain*

$$\Psi(L) \quad = \quad \frac{B(L) - \Phi(L)}{B(L)} = \frac{\alpha(L)}{B(L)} = \sum_{i=1}^{\infty} \psi_i L^i.$$

In analogy to Lemma 2.1 the $\{\psi_i\}$ *sequence can be obtained in the GARCH(p,* q*) as*

$$\begin{aligned} \psi_i &= \psi_i^{(p)} \quad \text{where} \\ \psi_i^{(r)} &= \lambda_{(r)} \psi_{i-1}^{(r)} + \psi_i^{(r-1)} \text{ for } 1 < r \leq p,\ i \geq 1 \text{ with } \psi_0^{(r)} = 0, \end{aligned}$$

where $\psi_i^{(1)}$ *in the GARCH(*$1, q$*) is given by* $\psi_i^{(1)} = \sum_{j=1}^{i} \lambda_{(1)}^{i-j} \alpha_j$ *for* $i = 1, 2, \ldots, q$, *and* $\psi_i^{(1)} = \lambda_{(1)}^{i-q} \psi_q$ *for all* $i > q$, *and for some ordering* $(\lambda_{(1)}, \ldots, \lambda_{(p)})$.

For $p = 1$ *or* $p = 2$ *it can be easily shown that our methodology leads to the same necessary and sufficient nonnegativity constraints as were derived by Nelson and Cao (1992).*

2.3.3 $\mathbf{FIGARCH(p, d, q)}$

As in the GARCH(p, q) necessary and sufficient conditions are more difficult to derive for the general FIGARCH(p, d, q) process. Instead, we state two sets of sufficient conditions. A first set which is more restrictive on the parameters, but – when satisfied – immediately implies $\psi_i \geq 0$ for all i. A second set which is less restrictive on the parameters, but requires to check the nonnegativity of the first k ψ_i.

Theorem 2.3. *The conditional variance of the FIGARCH(p, d, q) is nonnegative a.s. if*

1. *Let s be the number of inverse roots of $B(L)$ which are positive. If p is even we require $s \geq p/2 - 1$ and if p is odd we require $s \geq (p-1)/2$.*
2. $\psi_1 = d + \phi_1 - (\lambda_1 + \lambda_2 + \ldots + \lambda_p) \geq 0$
3. (a) *If $p = s$, then there must be a λ_i s.t. $(\lambda_i, d, \phi_1, \ldots, \phi_q)$ satisfy the conditions for the FIGARCH($1, d, q$)*

   (b) *If $p > s$, then there must exist an ordering $(\lambda_{(1)}, \lambda_{(2)}, \ldots, \lambda_{(p)})$ of the roots λ_i s.t.*

      i. $(\lambda_{(1)}, d, \phi_1, \ldots, \phi_q)$ *satisfy the conditions for the FIGARCH($1, d, q$)*

      ii. $0 > \lambda_{(2(p-s-\mathbf{1})+1)} \geq \lambda_{(2(p-s-1-\mathbf{1})+1)} \geq \ldots \geq \lambda_{(5)} \geq \lambda_{(3)}$ *and*
      $\lambda_{(2)} \geq |\lambda_{(3)}|, \lambda_{(4)} \geq |\lambda_{(5)}|, \ldots, \lambda_{(2(p-s-\mathbf{1}))} \geq |\lambda_{(2(p-s-\mathbf{1})+1)}|$

      iii. $\psi_2^{(3)}, \psi_2^{(5)}, \ldots, \psi_2^{(2(p-s-\mathbf{1})+1)} \geq 0$

      where $\mathbf{1} = 1$ *if* $\lambda_{(1)} < 0$ *and* 0 *otherwise.*

In a slightly modified version the same arguments can be applied to the GARCH(p, q) model using the representation given in Remark 2.3. Such a sufficient condition is more restrictive than the sufficient condition stated in Nelson and Cao (1992), which is given by $\lambda_{(1)} > \max_{i=2,\ldots,p}\{|\lambda_{(i)}|\}$, but – in contrast to the Nelson and Cao (1992) condition – directly implies $\psi_i \geq 0$ for all i.

Now, we come to the second and less restrictive sufficient condition. For this condition we have to find $0 \leq p_1 \leq p_2 \leq p$ with $p_2 - p_1$ even, such that the ordering of the p

inverse roots of $B(L)$ is in the following way

$$
\begin{aligned}
&\lambda_{(1)} \leq \cdots \leq \lambda_{(p_1)} < 0 \\
&\lambda_{(p_1+1)} > 0, \lambda_{(p_1+2)} < 0, \ldots, \lambda_{(p_2-1)} > 0, \lambda_{(p_2)} < 0 \\
&\qquad\qquad \text{with } \lambda_{(p_1+2i-1)} + \lambda_{(p_1+2i)} \geq 0, \quad i = 1, \ldots, (p_2 - p_1)/2 \\
&\lambda_{(p_2+1)} \geq \cdots \geq \lambda_{(p)} > 0
\end{aligned}
$$

This ordering is of course not unique as there can always be taken positive and negative roots to build a new pair as well as pairs can be separated, such that p_1 and p_2 differ. But it is always possible to find such an ordering.

Theorem 2.4. *If in the FIGARCH(p, d, q) there exists an ordering of the roots such that $\Lambda_{p_1} > -1$ then there exists a k such that $\psi_i \geq 0$ for all $i > k$.*

From this theorem it is clear that if $\Lambda_{p_1} > -1$ it is sufficient to check a finite number of $\psi_i, i \leq k$ to find out if for specific parameter values the conditional variance is nonnegative almost surely for all t. The existence of such a k under a weak condition is a strong result, since ψ_i is a i-th order polynomial in all parameters. The unknown k can be found going along the proof of this theorem. This procedure can easily be implemented. However, the condition is not necessary, i.e. it is possible to find parameter values such that $\psi_i \geq 0$ for all i and $\Lambda_{p_1} < -1$ for every ordering of the roots. The set which is not covered by this theorem is expected to be small, e.g. in the FIGARCH($2, d, q$) the theorem would cover four out of the five cases considered in Theorem 2.2. Since for most economic data we would expect that $\Lambda_p = \beta_1 > 0$ it suffices to check a finite number of coefficients. In this case the theorem provides a necessary and sufficient condition.

For higher order models – in which the estimated parameters do not satisfy the sufficient condition given by Theorem 2.4 – the sequence ψ_i can be calculated using Lemma 2.1. By plotting the sequence for sufficiently high lags one can obtain an indication whether the conditional variance will stay positive or not. However, this does of course not guarantee the nonnegativity of the conditional variance for all t.

Theorem 2.4 can be seen as being analogous to the sufficient condition stated in Nelson and Cao (1992) for the GARCH(p, q).

2.4 Empirical Example

In order to illustrate the importance of our results we investigate an empirical times series. We employ daily exchange rate data for the Japanese Yen vs. US Dollar sourced from the Datastream database for the period 1^{st} November 1993 to 18^{th} November 2003, giving a total of 2,621 observations. The continuously compounded returns are computed as $r_t = 100 \cdot [\log(p_t) - \log(p_{t-1})]$ where p_t is the price on day t. Table 2.1 presents the quasi-maximum likelihood parameter estimates for a FIGARCH$(1, d, 1)$ model ($r_t = \mu + \varepsilon_t$) estimated with the G@RCH package. Additionally to these parameter estimates

Table 2.1: FIGARCH$(1, d, 1)$ estimates.

$\widehat{\mu}$	$\widehat{\omega}$	$\widehat{d}$	$\widehat{\phi}_1$	$\widehat{\beta}_1$
0.012	0.027	0.264	0.592	0.727
(0.928)	(0.896)	(2.899)	(2.131)	(2.821)

Notes: Numbers in parentheses are t-statistics.

G@RCH provides the following output:

> *"The positivity constraint for the FIGARCH* $(1, d, 1)$ *is not observed.* $\Rightarrow$ *See Bollerslev and Mikkelsen (1996) for more details."*

Obviously, $\widehat{\phi}_1 > \widehat{f}_3$ in this case, and thus the Bollerslev and Mikkelsen (1996) conditions are violated. We check the conditions from Corollary 2.1, since it allows for $\widehat{\phi}_1 > \widehat{f}_3$. Step 1 is to determine k which is given by $(1 + \widehat{d})/(1 - \widehat{\phi}_1) \leq k$ and so step 2 is to verify that $\widehat{\psi}_{k-1} \geq 0$ which suffices for $\widehat{\psi}_i \geq 0$ for all $i = 1, 2, \ldots$. For our empirical example we find $k = 4$ and so $\widehat{\psi}_3$ has to be calculated by the recursions for the $(1, d, 1)$. It can be easily seen that $\widehat{\psi}_3 > 0$ for our parameter estimates. Hence, the set of parameters guarantees that the conditional variance is nonnegative almost surely for all t. Figure 2.5 illustrates the set of necessary and sufficient ($\phi_1 > 0$, $\beta_1 > 0$) parameter values for $\widehat{d} = 0.264$. The dashed line bounds the Bollerslev and Mikkelsen (1996) set and is given by $\phi_1 = \widehat{f}_3$. The cross represents the estimated parameter combination which lies in the necessary and sufficient set.[9] The practitioner solely relying on the G@RCH output – and hence on the Bollerslev and Mikkelsen (1996) conditions – would have falsely rejected the model.

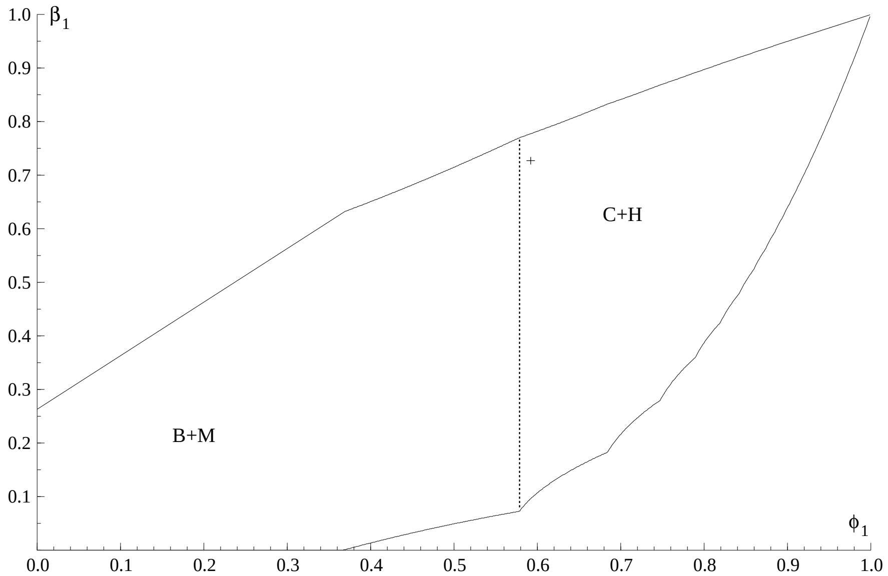

Figure 2.5: Necessary and sufficient parameter set for FIGARCH$(1, d, 1)$ (Case 1 and $\phi_1 > 0$) with $\widehat{d} = 0.264$.

2.5 Conclusions

In this chapter we derive necessary and sufficient conditions which ensure the nonnegativity of the conditional variance in the FIGARCH model of the order $p \leq 2$ and sufficient conditions for the general model. These conditions are important since any

[9]Note that the estimated parameters also violate the sufficient constraints suggested by Chung (1999).

practitioner estimating FIGARCH models – in particular when using parameter estimates for forecasting volatility – has to make sure that the model is well defined in the sense that it can not lead to negative conditional variances. This issue is even more important when considering long memory GARCH models since – unlike with the short memory GARCH models – one can not easily deduce the nonnegativity of the conditional variance from the sign of the estimated parameters. So far only a sufficient condition for the FIGARCH$(1, d, 1)$ was available and no conditions existed for higher order models. We demonstrate graphically how the necessary und sufficient conditions for the $(1, d, 1)$ enlarge the feasible parameter set which has important implications for the permitted shapes of the autocorrelation and impulse response functions of the LM-GARCH and thereby widens the range of data features that can be handled. The lack of knowledge concerning conditions which ensure the nonnegativity of the conditional variance in higher order models is presumably one reason why these models have been applied rarely in the literature. Studies as Caporin (2003) which are concerned with identification and order selection in long memory GARCH models restrict their analysis to the set of parameters defined by Bollerslev and Mikkelsen (1996) and hence to models of order $(1, d, 1)$, $(1, d, 0)$ and $(0, d, 0)$ only. Our work is intended to close this gap. As with the Nelson and Cao (1992) conditions we suggest that econometric packages should state not only the estimated parameters but also whether those satisfy the necessary and sufficient conditions derived in this chapter. An interesting avenue for future research would be to analyze the implications of imposing the necessary and sufficient restrictions directly on the maximum likelihood estimation.

Our results extend to more sophisticated long memory specifications such as the asymmetric power FIGARCH, the multivariate constant correlation FIGARCH and long memory ACD models in which it must be ensured that the conditional duration time does not take negative values.

2.6 Appendix

We first derive a recursive representation of the ψ_i sequence for the FIGARCH$(1, d, q)$ model. This representation will be used to prove Theorem 2.1. In the FIGARCH$(1, d, q)$ we can write

$$\begin{aligned}
\Psi(L) &= 1 - \Phi(L)(1-L)^d(1-\beta_1 L)^{-1} \\
&= 1 - (1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_q L^q) \cdot \sum_{i=0}^{\infty} c_i L^i \\
&= 1 - \sum_{i=0}^{\infty} (c_i - \phi_1 c_{i-1} - \ldots - \phi_q c_{i-q}) L^i = \sum_{i=1}^{\infty} \psi_i L^i
\end{aligned}$$

where

$$c_i = \sum_{j=0}^{i} \beta_1^{i-j} g_j \quad \text{for} \quad i \geq 0 \quad \text{and} \quad c_i = 0 \quad \text{for} \quad i < 0.$$

Hence, the sequence $\{\psi_i, i = 1, 2, \ldots\}$ can be written as

$$\psi_i = -c_i + \sum_{j=1}^{\min\{i,q\}} \phi_j c_{i-j} \qquad \text{for } i > 0.$$

Note, that the following recursion applies: $c_i = \beta_1 c_{i-1} + g_i$

Proof of Theorem 2.1.

Case 1: $0 < \beta_1 < 1$

"$\Leftarrow$"

1. $\psi_1, \ldots, \psi_{q-1} \geq 0$ by assumption.

2. (i) If $\psi_q \geq 0$ and $F_{q+1} \geq 0$ this ensures $\psi_i \geq 0$ for all $i > q$, since F_i is increasing and

$$\begin{aligned}
\psi_i &= -c_i + \phi_1 c_{i-1} + \ldots + \phi_q c_{i-q} \qquad \text{for } i \geq q+1 \\
&= -(\beta_1 c_{i-1} + g_i) + \phi_1(\beta_1 c_{i-2} + g_{i-1}) + \ldots + \phi_q(\beta_1 c_{i-q-1} + g_{i-q}) \\
&= \beta_1(-c_{i-1} + \phi_1 c_{i-2} + \ldots + \phi_q c_{i-q-1}) + \\
&\qquad (f_i f_{i-1} \ldots f_{i-q+1} - \phi_1 f_{i-1} \ldots f_{i-q+1} - \ldots - \phi_q)(-g_{i-q}) \\
&= \beta_1 \psi_{i-1} + F_i(-g_{i-q}) \geq 0. \qquad (2.11)
\end{aligned}$$

(ii) If $F_{q+1} < 0$, then for ψ_i with $q < i < k$ it holds that

$$\begin{aligned} \psi_i &= \beta_1 \psi_{i-1} + F_i(-g_{i-q}) \geq 0 \\ \Leftrightarrow \beta_1 \psi_{i-1} &\geq F_i g_{i-q} > 0 \\ \Rightarrow \psi_{i-1} &\geq 0 \end{aligned}$$

Thus, $\psi_i \geq 0$ implies $\psi_{i-1} \geq 0$. As $\psi_{k-1} \geq 0$ it follows recursively that $\psi_i \geq 0$ for all $q \leq i < k$. For $i \geq k$ we have

$$\psi_i = \beta_1 \psi_{i-1} + F_i(-g_{i-q})$$

and hence, from $\psi_{k-1} \geq 0$ follows $\psi_k \geq 0$ since $F_k \geq 0$. $\Rightarrow \psi_i \geq 0$ for all $i > k$ by induction.

"$\Rightarrow$"

1. The first condition and $\psi_q, \psi_{k-1} \geq 0$ are trivially fulfilled.

2. Either $F_{q+1} \geq 0$ or $F_{q+1} \leq 0$, but since $F_{i-1} \leq F_i$ and $F_i \to 1 - \sum_{j=1}^{q} \phi_i > 0$ there exists a k s.t. $F_i \geq 0$ for all $i \geq k$.

Case 2: $-1 < \beta_1 < 0$

"$\Leftarrow$"

1. $\psi_1, \ldots, \psi_{q-1} \geq 0$ by assumption.

2. We make use of the following recursion

$$\psi_i = \beta_1^2 \psi_{i-2} + F_i^{(1)}(-g_{i-q-1}) \text{ for } i \geq q+2 \tag{2.12}$$

3. (i) If $F_{q+2}^{(1)} \geq 0$ then $\psi_q \geq 0$ and $\psi_{q+1} \geq 0$ ensure that $\psi_i \geq 0$ for all $i \geq q+2$.

   (ii) If $F_{q+2}^{(1)} < 0$ then for ψ_i with $q+2 < i < k$ it holds that

$$\begin{aligned} \psi_i &= \beta_1^2 \psi_{i-2} + F_i^{(1)}(-g_{i-q-1}) \geq 0 \\ \Leftrightarrow \beta_1^2 \psi_{i-2} &\geq -F_i^{(1)}(-g_{i-q-1}) \geq 0 \\ \Rightarrow \psi_{i-2} &\geq 0 \end{aligned}$$

Thus, $\psi_i \geq 0$ implies $\psi_{i-2} \geq 0$. As $\psi_{k-1} \geq 0$ and $\psi_{k-2} \geq 0$ it follows recursively that $\psi_i \geq 0$ for all $q \leq i < k$. For $i \geq k$ we use equation (2.12) and hence, from $\psi_{k-1} \geq 0$ and $\psi_{k-2} \geq 0$ it follows that $\psi_i \geq 0$ for all $i > k$ by induction.

"$\Rightarrow$"

1. The first condition and $\psi_q, \psi_{q+1}, \psi_{k-2}, \psi_{k-1} \geq 0$ are trivially fulfilled.
2. Either $F_{q+1}^{(1)} \geq 0$ or $F_{q+1}^{(1)} \leq 0$, but since $F_{i-1}^{(1)} \leq F_i^{(1)}$ and $F_i^{(1)} \to \beta_1(1-\phi_1-\ldots-\phi_q)+(1-\phi_1-\ldots-\phi_q) > 0$ there exists a k s.t. $F_i^{(1)} \geq 0$ for all $i \geq k$.

□

Proof of Lemma 2.1.

$$\begin{aligned}
\Psi(L) &= 1-\frac{\Phi(L)(1-L)^d}{B(L)} = 1-\Phi(L)(1-L)^d(1-\lambda_{(1)}L)^{-1}\cdot\ldots\cdot(1-\lambda_{(p)}L)^{-1} \\
&= 1+\sum_{i=0}^{\infty}\psi_i^{(1)}L^i\cdot\sum_{i=0}^{\infty}\lambda_{(2)}^i L^i\cdot\ldots\cdot\sum_{i=0}^{\infty}\lambda_{(p)}^i L^i \\
&= 1+\sum_{i=0}^{\infty}\psi_i^{(2)}L^i\cdot\sum_{i=0}^{\infty}\lambda_{(3)}^i L^i\cdot\ldots\cdot\sum_{i=0}^{\infty}\lambda_{(p)}^i L^i \\
&\vdots \\
&= 1+\sum_{i=0}^{\infty}\psi_i^{(p)}L^i = \sum_{i=1}^{\infty}\psi_i^{(p)}L^i
\end{aligned}$$

since

$$\sum_{i=0}^{\infty}\psi_i^{(r-1)}L^i\cdot\sum_{i=0}^{\infty}\lambda_{(r)}^i L^i = \sum_{i=0}^{\infty}\psi_i^{(r)}L^i$$

with

$$\begin{aligned}
\psi_i^{(r)} &= \sum_{j=0}^{i}\lambda_{(r)}^j\psi_{i-j}^{(r-1)} = \psi_i^{(r-1)}+\sum_{j=1}^{i}\lambda_{(r)}^j\psi_{i-j}^{(r-1)} \\
&= \psi_i^{(r-1)}+\lambda_{(r)}\sum_{j=0}^{i-1}\lambda_{(r)}^j\psi_{i-j-1}^{(r-1)} = \psi_i^{(r-1)}+\lambda_{(r)}\psi_{i-1}^{(r)}
\end{aligned}$$

□

Using the representation from Lemma 2.1 and equation (2.11) for $\psi_i^{(1)}$ we deduce for $i > q+1$

$$\psi_i^{(r)} = \sum_{k=0}^{r-1} \psi_{i-2}^{(r-k)} \lambda_{(r-k)} (\Lambda_r - \Lambda_{r-k-1}) + F_i^{(r)}(-g_{i-q-1}). \tag{2.13}$$

Repeated application of equation (2.13) leads to

$$\begin{aligned}\psi_{i+2m}^{(r)} &= \lambda_{(r)}^{2m} \psi_i^{(r)} + \sum_{k=1}^{r-1} \sum_{j=1}^{m} \lambda_{(r)}^{2(m-j)} \lambda_{(r-k)} (\Lambda_r - \Lambda_{r-k-1}) \psi_{i+2j-2}^{(r-k)} \\ &\quad + \sum_{j=1}^{m} \lambda_{(r)}^{2(m-j)} F_{i+2j}^{(r)} (-g_{i-q+2j-1})\end{aligned} \tag{2.14}$$

for $m = 1, 2, \ldots$ and $i > q+1$.

We will make us of equation (2.13) and (2.14) in the subsequent proofs.

Proof of Proposition 2.1.

"$\Leftarrow$"

Case 1: $1 > \lambda_1 \geq \lambda_2 > 0$.

Set $\lambda_{(1)} = \lambda_1$, $\lambda_{(2)} = \lambda_2$ and note that $\psi_i^{(1)}$ is identical with ψ_i from Case 1 of the FIGARCH$(1, d, 0)$, since $\lambda_{(1)} = \lambda_1 > 0$.

Observe that $\psi_1 = \psi_1^{(2)} = d - (\lambda_{(1)} + \lambda_{(2)}) \geq 0$ implies $\psi_1^{(1)} = d - \lambda_{(1)} \geq 0$. Furthermore, from Proposition 2.3, Case 1, we know that $\psi_1^{(1)} \geq 0$ implies $\psi_i^{(1)} \geq 0$ for all i. Hence, it follows that $\psi_i = \lambda_{(2)} \psi_{i-1} + \psi_i^{(1)} \geq 0$ for all $i \geq 2$.

Case 2: $1 > \lambda_1 > 0 > \lambda_2 > -1$ and $\lambda_1 \geq |\lambda_2|$.

Set $\lambda_{(1)} = \lambda_1$, $\lambda_{(2)} = \lambda_2$. Then for $i > 2$ we can write

$$\begin{aligned}\psi_i &= \lambda_{(2)} \psi_{i-1} + \psi_i^{(1)} \\ &= \lambda_{(2)} \psi_{i-1} + \lambda_{(1)} \psi_{i-1}^{(1)} - g_i \\ &= \lambda_{(2)} \psi_{i-1} + \lambda_{(1)} (\psi_{i-1} - \lambda_{(2)} \psi_{i-2}) - g_i \\ &= (\lambda_{(1)} + \lambda_{(2)}) \psi_{i-1} - \lambda_{(1)} \cdot \lambda_{(2)} \psi_{i-2} - g_i\end{aligned}$$

Since $\lambda_{(1)} + \lambda_{(2)} \geq 0$, $\lambda_{(1)} \cdot \lambda_{(2)} < 0$ and $g_i < 0$, it suffices to assume that $\psi_1 \geq 0$ and $\psi_2 \geq 0$ to ensure that $\psi_i \geq 0$ for all $i > 2$.

Case 3: $1 > \lambda_1 > 0 > \lambda_2 > -1$ and $\lambda_1 < |\lambda_2|$.

Note that $\psi_i^{(1)}$ is identical with ψ_i from the FIGARCH$(1, d, 0)$, Case 2, since $\lambda_{(1)} = \lambda_2 < 0$.

1. $\psi_1^{(1)} = d - \lambda_{(1)} \geq 0$ is obviously satisfied.

2. (i) If $\lambda_{(1)}$ is such that $\psi_2^{(1)} \geq 0$ which implies $\psi_i^{(1)} \geq 0$ for all i (Proposition 2.3, Case 2) we immediately obtain $\psi_i \geq 0$ for all i.

   (ii) If $\lambda_{(1)}$ is such that $\psi_2^{(1)} \leq 0$ we make use of the recursion

$$\psi_i^{(1)} = \lambda_{(1)}^2 \psi_{i-2}^{(1)} + (\lambda_{(1)} + f_i)(-g_{i-1}) \text{ for } i \geq 3 \tag{2.15}$$

   Since we know that there exists a k s.t. $\lambda_{(1)} + f_{k-1} < 0 \leq \lambda_{(1)} + f_k$, we can conclude that there exists an even $\overline{k} = k + 2i + \mathbf{1}_k$ with

$$\begin{aligned}\psi_{k+2i+\mathbf{1}_k}^{(1)} &= \lambda_{(1)}^{2i+2} \psi_{k-2+\mathbf{1}_k}^{(1)} + \sum_{j=0}^{i} (\lambda_{(1)} + \\ &\qquad f_{k+2j+\mathbf{1}_k})(-g_{k-1+2j+\mathbf{1}_k}) \lambda_{(1)}^{2(i-j)} \geq 0\end{aligned}$$

   since

$$0 \leq -\lambda_{(1)}^2 \psi_{k-2+\mathbf{1}_k}^{(1)} < \sum_{j=0}^{i} (\lambda_{(1)} + f_{k+2j+\mathbf{1}_k})(-g_{k-1+2j+\mathbf{1}_k}) \lambda_{(1)}^{-2j}$$

   where the rhs is diverging and $\mathbf{1}_k$ is defined as

$$\mathbf{1}_k = \begin{cases} 1 & \text{if } k \text{ odd,} \\ 0 & \text{otherwise.} \end{cases}$$

   By definition $\psi_4^{(1)}, \psi_6^{(1)}, \ldots, \psi_{\overline{k}-2}^{(1)} \leq 0$ and from $\psi_i^{(1)} = \lambda_{(1)} \psi_{i-1}^{(1)} - g_i$ it follows that
   $\psi_3^{(1)}, \psi_5^{(1)}, \ldots, \psi_{\overline{k}-1}^{(1)} \geq 0$.
   Again from equation (2.15) we deduce that $\psi_i^{(1)} \geq 0$ for all $i \geq \overline{k} + 1$.

Observe that $\psi_1 = d - (\lambda_{(1)} + \lambda_{(2)}) \geq 0$ without further assumptions. For $i \geq 3$ we can apply the following recursion

$$\begin{aligned} \psi_i &= \lambda_{(2)}\psi_{i-1} + \psi_i^{(1)} \\ &= \lambda_{(2)}^2\psi_{i-2} + \lambda_{(2)}\psi_{i-1}^{(1)} + \psi_i^{(1)} = \lambda_{(2)}^2\psi_{i-2} + (\lambda_{(1)} + \lambda_{(2)})\psi_{i-1}^{(1)} - g_i \end{aligned} \tag{2.16}$$

Given that $\psi_1 \geq 0$ and knowing that all $\psi_i^{(1)} \leq 0$ with i even we obtain $\psi_i \geq 0$ for all $i < \overline{k}$ and i odd. By observing that $\psi_i^{(1)} \geq 0$ for all $i \geq \overline{k}$ we conclude from equation (2.16) that $\psi_i \geq 0$ for all $i \geq \overline{k}$. It remains to assume that $\psi_2, \psi_4, \ldots, \psi_{\overline{k}-2} \geq 0$.

Case 4: $0 > \lambda_1 \geq \lambda_2 > -1$.

Note that $\psi_i^{(1)}$ is identical with ψ_i from the FIGARCH$(1, d, 0)$, Case 2 with $\lambda_{(1)} = \lambda_1$. Again, observe that $\psi_1 = d-(\lambda_{(1)}+\lambda_{(2)}) \geq 0$ without further assumptions. Now consider ψ_2:

$$\begin{aligned} \psi_2 &= \lambda_{(2)}\psi_1 + \psi_2^{(1)} \\ &= \lambda_{(2)}(d - (\lambda_{(1)} + \lambda_{(2)})) + \lambda_{(1)}(d - \lambda_{(1)}) + f_2 d \\ &= (\lambda_{(1)} + \lambda_{(2)} + f_2)d - \lambda_{(1)}^2 - \lambda_{(2)}(\lambda_{(1)} + \lambda_{(2)}) \\ \psi_2 &\geq 0 \\ \Leftrightarrow &\quad (\lambda_{(1)} + \lambda_{(2)} + f_2)d \geq \lambda_{(1)}^2 + \lambda_{(2)}(\lambda_{(1)} + \lambda_{(2)}) \geq 0 \\ \Rightarrow &\quad (\lambda_{(1)} + \lambda_{(2)}) + f_2 \geq 0 \end{aligned}$$

Notice that $F_2^{(2)} = \Lambda_2 + f_2 \geq 0$ implies $F_i^{(2)} = \Lambda_2 + f_i \geq 0$ for all $i \geq 0$. Finally, for $i \geq 3$ we can apply equation (2.13)

$$\psi_i = \lambda_{(2)}^2\psi_{i-2} + \lambda_{(1)}\Lambda_2\psi_{i-2}^{(1)} + F_i^{(2)}(-g_{i-1}) \tag{2.17}$$

For ψ_i being nonnegative by equation (2.17) we must require that $\psi_i^{(1)} \geq 0$ for all i, which is the case iff $\psi_2^{(1)} \geq 0$. Given that $\psi_1 \geq 0$ and by assuming that $\psi_2 \geq 0$ it follows from the recursion that $\psi_i \geq 0$ for all i.

"$\Rightarrow$"

If $\psi_i \geq 0$ for all i, then all assumptions are trivially satisfied.

□

Before we prove the next theorem we need to establish the rate of convergence of the g_j coefficients. Since for the hypergeometric function it holds that

$$H(-d, 1; 1; L) = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(1+j)} L^j,$$

where $\Gamma(\cdot)$ is the gamma function which is defined by $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$ for $x > 0$, $\Gamma(x) = \infty$ for $x = 0$ and $\Gamma(x) = x^{-1}\Gamma(1+x)$ for $x < 0$, we have the representation

$$g_j = \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(1+j)} = O(j^{-d-1})$$

by applying Sterling's formula. Hence,

$$-g_j\lambda_i^{-j} \longrightarrow +\infty \quad \text{as } j \longrightarrow \infty,$$

which will be made use of in the subsequent proofs.

Proof of Theorem 2.2.

Case 1: $1 > \lambda_1 \geq \lambda_2 > 0$.

"$\Leftarrow$"

Note, that $\psi_i^{(1)}$ is identical with ψ_i from the FIGARCH$(1, d, q)$, Case 1, with $\lambda_{(1)} = \lambda_1$.

1. We assume $\psi_1 \geq 0$, $\psi_2, \ldots, \psi_{q-1} \geq 0$. Note, that $\psi_1 \geq 0$ implies $\psi_1^{(1)} \geq 0$.

2. (i) If either $\psi_q^{(1)} \geq 0$ and $F_{q+1} \geq 0$ or for k with $F_{k-1} < 0 \leq F_k$ we have that $\psi_{k-1}^{(1)} \geq 0$, by the same arguments as in the FIGARCH$(1, d, q)$, Case 1, we can conclude that $\psi_i^{(1)} \geq 0$ for $i = q, \ldots$ and hence $\psi_{q-1} \geq 0$ (note, for $q = 1$ we require $\psi_1 \geq 0$ instead) implies $\psi_i \geq 0$ for all $i \geq q$.

   (ii) If $\psi_{k-1}^{(1)} \leq 0$, then there exist $(\underline{k}, \overline{k})$ with $\underline{k} = \min\{j \mid \psi_j^{(1)} \leq 0, j = q, \ldots\}$, $\overline{k} = \min\{j \mid \psi_j^{(1)} \geq 0, j = k, \ldots\}$ (only if $q = 1$ do we have $q < \underline{k}$) s.t.

$\psi_i^{(1)} \geq 0 \; \forall \; i \in \{q, \ldots, \underline{k}-1\} \cup \{\overline{k}, \ldots\}$ and $\psi_i^{(1)} \leq 0 \; \forall \; i \in \{\underline{k}, \ldots, \overline{k}-1\}$. $\overline{k}$ exists because we can write $\psi_{k+i}^{(1)}$ as

$$\psi_{k+i}^{(1)} = \lambda_{(1)}^{i+1}\psi_{k-1}^{(1)} + \sum_{j=0}^{i} F_{k+j}(-g_{k-q+j})\lambda_{(1)}^{i-j}$$

and hence we must have that

$$0 \leq -\lambda_{(1)}\psi_{k-1}^{(1)} < \sum_{j=0}^{i} F_{k+j}(-g_{k-q+j})\lambda_{(1)}^{-j}$$

for some i (the rhs is diverging) which gives $\overline{k} = k + i$. The existence of $\underline{k}$ is obvious, since $\psi_{k-1}^{(1)} \leq 0$. This implies $\psi_i \geq 0$ for $i = 1, \ldots, \underline{k}-1$. Assuming that $\psi_{\overline{k}-1} \geq 0$ and starting with $i = \overline{k}-1$ we derive recursively

$$\begin{aligned} \psi_i &= \lambda_2\psi_{i-1} + \psi_i^{(1)} \geq 0 \\ \Leftrightarrow \lambda_2\psi_{i-1} &\geq -\psi_i^{(1)} \geq 0 \\ \Leftrightarrow \psi_{i-1} &\geq 0 \end{aligned}$$

which implies that $\psi_i \geq 0$ with $i \in \{\underline{k}, \ldots, \overline{k}-2\}$. Finally, $\psi_i \geq 0$ for $i \geq \overline{k}$, since $\psi_{\overline{k}-1} \geq 0$ and $\psi_i^{(1)} \geq 0 \; \forall \; i \geq \overline{k}$.

" $\Rightarrow$ "

$\psi_1 \geq 0$ and $\psi_{\overline{k}-1} \geq 0$ is trivially satisfied. For k with $F_{k-1} < 0 \leq F_k$ we either have that $\psi_{k-1}^{(1)} \geq 0$ or $\psi_{k-1}^{(1)} \leq 0$, but in the latter case there exists a $\overline{k}$ s.t. $\psi_{\overline{k}}^{(1)} \geq 0$ as shown above.

Case 2: $1 > \lambda_1 > 0 > \lambda_2 > -1$ and $\lambda_1 \geq |\lambda_2|$.

"$\Leftarrow$"

1. We assume $\psi_1 \geq 0$, $\psi_2, \ldots, \psi_{q-1} \geq 0$.

2. Set $\lambda_{(2)} = \lambda_2$ and $\lambda_{(1)} = \lambda_1$. Similar as in the FIGARCH$(2, d, 0)$ we obtain for $i \geq q+1$

$$\psi_i = \lambda_{(2)}\psi_{i-1} + \psi_i^{(1)} = (\lambda_{(1)} + \lambda_{(2)})\psi_{i-1} - \lambda_{(1)} \cdot \lambda_{(2)}\psi_{i-2} + F_i(-g_{i-q})$$

(i) Either $\psi_q \geq 0$ and $F_{q+1} \geq 0$ which implies $\psi_i \geq 0$ for all $i \geq q$ or

(ii) since $\lambda_{(1)} + \lambda_{(2)} \geq 0$, $\lambda_{(1)} \cdot \lambda_{(2)} < 0$, $g_i < 0$ and there exists a $k \geq q+2$ such that $F_{k-1} < 0 \leq F_k$ it suffices to assume that $\psi_1, \ldots, \psi_{k-1} \geq 0$ to ensure that $\psi_i \geq 0$ for all i.

"$\Rightarrow$"

$\psi_1, \ldots, \psi_{k-1} \geq 0$ is trivially satisfied and the existence of k follows from $\lambda_{(1)} < 1$.

Case 3: $1 > \lambda_1 > 0 > \lambda_2 > -1$ and $\lambda_1 < |\lambda_2|$.

"$\Leftarrow$"

Note that $\psi_i^{(1)}$ is identical with ψ_i from the FIGARCH$(1, d, q)$, Case 2, with $\lambda_{(1)} = \lambda_2$.

1. We assume $\psi_1 \geq 0$, $\psi_2, \ldots, \psi_{q-1} \geq 0$.

2. (i) If either $\psi_q^{(1)} \geq 0$, $\psi_{q+1}^{(1)} \geq 0$ and $F_{q+2}^{(1)} \geq 0$ or for k with $F_{k-1} < 0 \leq F_k^{(1)}$ we have that $\psi_{k-1}^{(1)} \geq 0$ and $\psi_{k-2}^{(1)} \geq 0$, by the same arguments as in the FIGARCH$(1, d, q)$, Case 2, we can conclude that $\psi_i^{(1)} \geq 0$ for $i = q, \ldots$ and hence $\psi_{q-1} \geq 0$ (note, for $q = 1$ we require $\psi_1 \geq 0$ instead) implies $\psi_i \geq 0$ for all $i \geq q$.

   (ii) If $\psi_{k-1}^{(1)} \leq 0$ and/or $\psi_{k-2}^{(1)} \leq 0$, then there exists $\overline{k}$ with $q < k < \overline{k}$ s.t. $\psi_i^{(1)} \geq 0 \; \forall \; i \geq \overline{k}$. $\overline{k}$ exists because we can write $\psi_{k+2i}^{(1)}$ and $\psi_{k+1+2s}^{(1)}$ as

$$\begin{aligned} \psi_{k+2i}^{(1)} &= \lambda_{(1)}^{2i+2} \psi_{k-2}^{(1)} + \sum_{j=0}^{i} F_{k+2j}^{(1)} (-g_{k-q+2j-1}) \lambda_{(1)}^{2(i-j)} \\ \psi_{k+1+2s}^{(1)} &= \lambda_{(1)}^{2s+2} \psi_{k-1}^{(1)} + \sum_{j=0}^{s} F_{k+1+2j}^{(1)} (-g_{k-q+2j}) \lambda_{(1)}^{2(s-j)} \end{aligned}$$

   and hence we must have that

$$\begin{aligned} 0 \leq & - \lambda_{(1)}^2 \psi_{k-2}^{(1)} < \sum_{j=0}^{i} F_{k+2j}^{(1)} (-g_{k-q+2j-1}) \lambda_{(1)}^{-2j} \\ 0 \leq & - \lambda_{(1)}^2 \psi_{k-1}^{(1)} < \sum_{j=0}^{s} F_{k+1+2j}^{(1)} (-g_{k-q+2j}) \lambda_{(1)}^{-2j} \end{aligned}$$

for some i, s (the rhs is diverging) which gives $\overline{k} = k + 2 \cdot \max\{i, s\}$. By assuming that $\psi_{q+1}, \ldots, \psi_{\overline{k}-1} \geq 0$ the nonnegativity of ψ_i for all i follows directly.

" $\Rightarrow$ "

$\psi_1 \geq 0$ and $\psi_{\overline{k}-1} \geq 0$ is trivially satisfied. For k with $F^{(1)}_{k-1} < 0 \leq F^{(1)}_k$ we either have that $\psi^{(1)}_{k-1} \geq 0$ and $\psi^{(1)}_{k-2} \geq 0$ or $\psi^{(1)}_{k-1} \leq 0$ and/or $\psi^{(1)}_{k-2} \leq 0$, but in the latter case there exists a $\overline{k}$ s.t. $\psi^{(1)}_{\overline{k}} \geq 0$ as shown above.

Case 4: $0 > \lambda_1 \geq \lambda_2 > -1$, $\lambda_1 + \lambda_2 > -1$

" $\Leftarrow$ "

Set $\lambda_{(1)} = \lambda_1$ and $\lambda_{(2)} = \lambda_2$. As in Case 2 we can write ψ_i as

$$\psi_i = \lambda_{(2)}\psi_{i-1} + \psi^{(1)}_i$$

with $\psi^{(1)}_i$ coming from the FIGARCH$(1, d, q)$ model. But now, obviously we need to require $\psi^{(1)}_i \geq 0$ for all i. Using equation (2.13) we obtain for $i \geq q + 2$

$$\psi_i = \lambda^2_{(2)}\psi_{i-2} + \lambda_{(1)}\Lambda_2\psi^{(1)}_{i-2} + F^{(2)}_i(-g_{i-q-1})$$

We know that there exists a k with $F^{(2)}_{k-1} < 0 \leq F^{(2)}_k$. Therefore, assuming that $\psi_1, \ldots, \psi_{k-1} \geq 0$ implies $\psi_i \geq 0$ for all $i \geq k$.

" $\Rightarrow$ "

$\psi_1, \ldots, \psi_{k-1} \geq 0$ and the existence of k is trivially satisfied. Moreover, $\psi^{(2)}_i > 0$ implies $\psi^{(1)}_i > 0$ for all i.

Case 5: $0 > \lambda_1 \geq \lambda_2 > -1$, $\lambda_1 + \lambda_2 < -1$

" $\Leftarrow$ "

Choose $\lambda_{(1)}$ and $\lambda_{(2)}$ such that $\lambda_{(1)}/\lambda_{(2)} < 1$. Then by equation (2.14)

$$\begin{aligned}
\psi_{i+2m}^{(2)} &= \lambda_{(2)}^{2m}\psi_i^{(2)} + \sum_{j=1}^{m}\lambda_{(2)}^{2(m-j)}\lambda_{(1)}\Lambda_2\psi_{i+2j-2}^{(1)} + \sum_{j=1}^{m}\lambda_{(2)}^{2(m-j)}F_{i+2j}^{(2)}(-g_{i-q+2j-1}) \\
&= \lambda_{(2)}^{2m}\psi_i^{(2)} + \sum_{j=1}^{m}\lambda_{(2)}^{2(m-j)}\lambda_{(1)}\Lambda_2\left[\lambda_{(1)}^{2j-2}\psi_i^{(1)} + \sum_{l=1}^{j}\lambda_{(1)}^{2(i-l)}F_{i+2l}^{(1)}(-g_{i-q+2l-1})\right] \\
&\quad + \sum_{j=1}^{m}\lambda_{(2)}^{2(m-j)}F_{i+2j}^{(2)}(-g_{i-q+2j-1}) \\
&= \lambda_{(2)}^{2m}\psi_i^{(2)} + \lambda_{(2)}^{2m}\lambda_{(1)}^{-1}\Lambda_2\psi_i^{(1)}\sum_{j=1}^{m}\left(\frac{\lambda_{(1)}}{\lambda_{(2)}}\right)^{2j} + \sum_{j=1}^{m}\lambda_{(2)}^{2(m-j)}S_{j,i} \qquad (2.18)
\end{aligned}$$

Now, choose k such that $F_{k-1}^{(1)} < 0 \leq F_k^{(1)}$ and observe that

$$0 \leq \sum_{l=1}^{\infty}\lambda_{(1)}^{2(k-l)}F_{k+2l}^{(1)}(-g_{k-q+2l-1}) < \infty$$

Hence, if there exists $\bar{k}$ with $S_{\bar{k},k} > 0$, then $S_{j,k} > 0$ for all $j > \bar{k}$ by monotonicity. Therefore, if $\psi_1, \ldots, \psi_{\bar{k}} \geq 0$ it follows by equation (2.18) that $\psi_i \geq 0$ for all i.

" $\Rightarrow$ "

$\psi_1, \ldots, \psi_{\bar{k}} \geq 0$ is trivially satisfied. If $\bar{k}$ does not exist it follows from equation (2.18) that ψ_{i+2m} will become negative for some m. As in Case 4, $\psi_i^{(2)} > 0$ implies $\psi_i^{(1)} > 0$ for all i. □

Proof of Theorem 2.3.

1. If $p = s$ and there is a λ_i such that $(\lambda_i, d, \phi_1, \ldots, \phi_q)$ satisfy the conditions for the FIGARCH$(1, d, q)$, it is immediately clear from the recursion derived in Lemma 2.1 that $\psi_1 \geq 0$ is sufficient for $\psi_i \geq 0$ for all i.

2. Let $p > s$. First, observe that $\psi_1 \geq 0$ implies $\psi_1^{(1)}, \ldots, \psi_1^{(p-1)} \geq 0$.

   If $(\lambda_{(1)}, d, \phi_1, \ldots, \phi_q)$ satisfy the conditions for the FIGARCH$(1, d, q)$ we can conclude that $\psi_i^{(1)} \geq 0$ for all i.

   Since $\psi_i^{(1)} \geq 0$ for all i and $\lambda_{(2)} > 0$ we immediately obtain $\psi_i^{(2)} \geq 0$ for all i.

$\psi_i^{(3)}$ can be written as

$$\psi_i^{(3)} = \lambda_{(3)}^2 \psi_{i-2}^{(3)} + \underbrace{(\lambda_{(2)} + \lambda_{(3)})}_{\geq 0} \psi_{i-1}^{(2)} + \psi_i^{(1)} \text{ for } i \geq 3.$$

Since $\psi_1^{(3)} = d + \phi_1 - (\lambda_{(1)} + \lambda_{(2)} + \lambda_{(3)}) \geq 0$, assuming $\psi_2^{(3)} \geq 0$ ensures that $\psi_i^{(3)} \geq 0$ for all i.

By the same arguments we can show that $\psi_i^{(r)} \geq 0$ for all i follows from $\psi_i^{(r)} = \lambda_{(r)}\psi_{i-1}^{(r)} + \psi_i^{(r-1)}$ if $\psi_i^{(r-1)} \geq 0$ for all i and $r \leq 2(p-s)$ even and from $\psi_i^{(r)} = \lambda_{(r)}^2\psi_{i-2}^{(r)} + (\lambda_{(r-1)} + \lambda_{(r)})\psi_{i-1}^{(r-1)} + \psi_i^{(r-2)}$ by assuming $\psi_2^{(r)} \geq 0$ if $r \leq 1 + 2(p-s)$ odd.

For $r > 1 + 2(p-s)$ the recursion $\psi_i^{(r)} = \lambda_{(r)}\psi_{i-1}^{(r)} + \psi_i^{(r-1)}$ applies again.

□

Proof of Theorem 2.4.

Assume there exists an ordering with $\Lambda_{p_1} > -1$. Recall form equation (2.6) that $F_i^{(r)} \to c > 0$ for all $r = 1, \ldots, p_1$.

We use the recursions (see Lemma 2.1)

$$\begin{aligned} \psi_i^{(1)} &= \lambda_{(1)}\psi_{i-1}^{(1)} + F_i(-g_{i-q}) & & i > q \\ \psi_i^{(r)} &= \lambda_{(r)}\psi_{i-1}^{(r)} + \psi_i^{(r-1)} & & 1 < r \leq p,\ i \geq 1, \text{ with } \quad \psi_0^{(r)} = -1 \end{aligned}$$

to deduce for $i > q+1$

$$\psi_i^{(r)} = \sum_{k=0}^{r-1} \psi_{i-2}^{(r-k)} \lambda_{(r-k)} (\Lambda_r - \Lambda_{r-k-1}) + F_i^{(r)}(-g_{i-q-1}) \tag{2.19}$$

where $\Lambda_r - \Lambda_{r-k-1} < 0$ for all k as long as $r = 2, \ldots, p_1$. Repeated application of equation (2.19) leads to

$$\begin{aligned} \psi_{i+2m}^{(r)} &= \lambda_{(r)}^{2m}\psi_i^{(r)} + \sum_{k=1}^{r-1}\sum_{j=1}^{m} \lambda_{(r)}^{2(m-j)} \lambda_{(r-k)} (\Lambda_r - \Lambda_{r-k-1}) \psi_{i+2j-2}^{(r-k)} \\ &\qquad + \sum_{j=1}^{m} \lambda_{(r)}^{2(m-j)} F_{i+2j}^{(r)}(-g_{i-q+2j-1}) \end{aligned} \tag{2.20}$$

for $m = 1, 2, \dots$ and $i > q + 1$.

In the following we show that for each r there exists a k_r such that $\psi_j^{(r)} > 0$ for all $j \geq k_r$. From the FIGARCH$(1, d, q)$ we know that there exists a $k_1 > q$ such that $\psi_j^{(1)} > 0$ for all $j > k_1$. This holds also in the case where $\lambda_1 > 0$, i. e. $p_1 = 0$.

(I) $1 < r \leq p_1$

Assume that we have shown the existence of k_{r-1}. We know that $F_i^{(r)} = \Lambda_r F_{i-1} + F_i f_{i-q} \geq \widetilde{c} > 0$ for all $i > \widetilde{k}$. Then, for $i = \max\{k_{r-1}, \widetilde{k}\}$ or $i = \max\{k_{r-1}, \widetilde{k}\} + 1$ it is clear that every term on the right side of equation (2.20) is positive except for $\psi_i^{(r)}$. If $\psi_i^{(r)} > 0$ it follows directly that $\psi_{i+2m}^{(r)} \geq 0$ for all m and so we set $k_r = i$.

If $\psi_i^{(r)} < 0$ we see that $\psi_{i+2m}^{(r)} > 0$ is equivalent to

$$-\psi_i^{(r)} < \sum_{k=1}^{r-1} \sum_{j=1}^{m} \lambda_{(r)}^{-2j} \lambda_{(r-k)} (\Lambda_r - \Lambda_{r-k-1}) \psi_{i+2j-2}^{(r-k)} + \sum_{j=1}^{m} \lambda_{(r)}^{-2j} F_{i+2j}^{(r)} (-g_{i+2j-q-1}) \qquad (2.21)$$

Now the first sum on the right side is positive for all j and the second sum tends to infinity as $F_{i+2j} \to c > 0$. Then it is obvious that there exists a $\bar{m}$ from which on this inequality is fulfilled. Then we set $k_r = \bar{m}$.

(II) $p_1 < r \leq p_2$

Consider first the cases where $\lambda_{(r)} > 0$, i. e. where $r - p_1$ is odd. Here we have that

$$\psi_i^{(r)} = \lambda_{(r)} \psi_{i-1}^{(r)} + \sum_{k=1}^{(r-p_1-1)/2} \left(\lambda_{(r-2k+1)}^2 \psi_{i-2}^{(r-2k+1)} + (\lambda_{(r-2k+1)} + \lambda_{(r-2k)}) \psi_{i-1}^{(r-2k)} \right) + \psi_i^{(p_1)}$$

and the iterated version

$$\begin{aligned} \psi_{i+m}^{(r)} \quad = \quad & \lambda_{(r)}^m \psi_i^{(r)} + \sum_{j=1}^{m} \lambda_{(r)}^{m-j} \psi_{i+j}^{(p_1)} \\ & + \sum_{j=1}^{m} \lambda_{(r)}^{m-j} \sum_{k=1}^{(r-p_1-1)/2} \left(\lambda_{(r-2k+1)}^2 \psi_{i+j-2}^{(r-2k+1)} + (\lambda_{(r-2k+1)} + \lambda_{(r-2k)}) \psi_{i+j-1}^{(r-2k)} \right) \end{aligned}$$

For $i = k_{r-1}$ every term on the right hand side is positive except for $\psi_i^{(r)}$ which might be negative. If it is positive we set $k_r = k_{r-1}$, if it is negative, we plug in equation (2.20) for $\psi_{i+j}^{(p_1)}$ and obtain

$$\begin{aligned}
\psi_{i+m}^{(r)} &= \lambda_{(r)}^m \psi_i^{(r)} + \sum_{\substack{j=1 \\ j \text{ even}}}^{m} \lambda_{(r)}^{m-j} \sum_{l=1}^{j/2} \lambda_{(p_1)}^{j-2l} F_{i+2l}^{(p_1)}(-g_{i-q+2l-1}) \\
&\quad + \sum_{\substack{j=1 \\ j \text{ odd}}}^{m} \lambda_{(r)}^{m-j} \psi_{i+j}^{(p_1)} + \sum_{\substack{j=1 \\ j \text{ even}}}^{m} \lambda_{(r)}^{m-j} \lambda_{(p_1)}^{j} \psi_i^{(p_1)} \\
&\quad + \sum_{\substack{j=1 \\ j \text{ even}}}^{m} \lambda_{(r)}^{m-j} \sum_{k=1}^{p_1-1} \sum_{l=1}^{j/2} \lambda_{(p_1)}^{j-2l} \lambda_{(p_1-k)} (\Lambda_{p_1} - \Lambda_{p_1-k-1}) \psi_{i+2l-2}^{(p_1-k)} \\
&\quad + \sum_{j=1}^{m} \lambda_{(r)}^{m-j} \sum_{k=1}^{(r-p_1-1)/2} \left(\lambda_{(r-2k+1)}^2 \psi_{i+j-2}^{(r-2k+1)} + (\lambda_{(r-2k+1)} + \lambda_{(r-2k)}) \psi_{i+j-1}^{(r-2k)} \right)
\end{aligned}$$

Next we use the same argument as in equation (2.21): Dividing the whole equation by $\lambda_{(r)}^m$ we argue that the last sum on the right side diverges, as

$$\sum_{\substack{j=1 \\ j \text{ even}}}^{m} \lambda_{(r)}^{-j} \sum_{l=1}^{j/2} \lambda_{(p_1)}^{j-2l} F_{i+2l}^{(p_1)}(-g_{i-q+2l-1}) \geq \sum_{\substack{j=1 \\ j \text{ even}}}^{m} \lambda_{(r)}^{-j} F_{i+j}^{(p_1)}(-g_{i-q+j-1})$$

and the right sum tends to infinity by the usual argument. From this the existence of k_r follows.

Next consider $\lambda_{(r)} < 0$, i. e. $r - p_1$ is even. We then have for $i - 2 > k_{r-1}$ that

$$\begin{aligned}
\psi_i^{(r)} &= \sum_{k=1}^{(r-p_1)/2+1} \left(\lambda_{(r-2k)}^2 \psi_{i-2}^{(r-2k)} + (\lambda_{(r-2k)} + \lambda_{(r-2k-1)}) \psi_{i-1}^{(r-2k-1)} \right) \\
&\quad + \lambda_{(r)}^2 \psi_{i-2}^{(r)} + (\lambda_{(r)} + \lambda_{(r-1)}) \psi_{i-1}^{(r-1)} + \psi_i^{(p_1)}
\end{aligned} \tag{2.22}$$

Every term on the right hand side is positive except for $\psi_{i-2}^{(r)}$ which is possibly negative. If it is negative, iterating and inserting equation (2.20) show the existence of k_r by the same arguments as in the case where $r - p_1$ is odd.

(III) $p_2 < r \leq p$

Here we use the representation

$$\psi_i^{(r)} = \lambda_{(r)} \psi_{i-1}^{(r)} + \sum_{k=1}^{r-p_2+1} \lambda_{(r-k)} \psi_{i-1}^{(r-k)} + \psi_i^{(p_2)}$$

for $i > k_{r-1}$. If $\psi_{i-1}^{(r)}$ is positive, then every term on the right side is positive and we set $k_r = k_{r-1}$. If $\psi_{i-1}^{(r)}$ is negative then we plug in equation (2.22) for $\psi_i^{(p_2)}$ and then equation (2.19) for $\psi_i^{(p_1)}$. Iteration and the same argument as in equation (2.20) shows the existence of k_r.

Finally, set $k = k_p + 1$.

□

Chapter 3

The Impulse Response Function of the Long Memory GARCH Process

3.1 Introduction

The topic of long memory and persistence has recently attracted considerable attention in terms of the second moment of a process. An excellent survey of major econometric work on long memory processes and their applications in economics and finance is given by Baillie (1996). The issue of temporal dependence on financial time series has been the focus of attention since information on persistence can also guide the search for an economic explanation of movements in asset returns. For example, as Baillie et al. (1996a) point out, there is a direct relation between the long-term dependence in the conditional variances of daily spot exchange rates and the long memory in the forward premium. This relation could explain the systematic rejection of the unbiasedness hypothesis as

---

This chapter was published as: Conrad, C., and M. Karanasos (2006). "The impulse response function of the long memory GARCH process." *Economics Letters* 90, 34–41.

an artefact due to the unbalanced regression of the return on the premium (Baillie et al., 1996a).

Robinson (1991) was the first to consider the long memory potential of a model which he called linear ARCH (LARCH). Subsequently, many researchers have proposed extensions of GARCH-type models which can produce long memory behavior (see, for example, Teyssière, 1998, Davidson, 2004 and Giraitis et al. 2004, and the references therein). The empirical relevance of long memory conditional heteroscedasticity has emerged in a variety of studies of time series of economic and financial variables (see, for example, Conrad and Karanasos, 2005a,b [see Chapter 4 and 5]). Kirman and Teyssière (2005) assemble models from economic theory providing plausible micro foundations for the occurrence of long memory in economics. Recent research has been aimed at both extending our understanding of these well established models, and widening the range of data features that can be handled. For example, Giraitis et al. (2005) provide an overview of recent theoretical findings on the long memory GARCH (LMGARCH) processes.

Moreover, Baillie et al. (1996a) measure the persistence of shocks to the conditional variance using impulse response functions (IRFs). To appreciate how such measures work in practice they consider the fractionally integrated GARCH (FIGARCH) process of order $(1, d, 0)$. To that end, in this chapter the IRF is analyzed in the framework of an LMGARCH(p, d, q) process. We also look at the persistence of shocks in the conditional variance process for the LMGARCH model as compared with the persistence of shocks for the stable and integrated GARCH models. An important related study by Karanasos et al. (2004b) derives convenient representations for the autocorrelation function (ACF) of the squared values of LMGARCH processes, and some of our results can been seen as complementary to theirs.

This chapter is organized as follows. Section 3.2 lays out the model of interest, assumptions and notation. Section 3.3 presents expressions for the cumulative IRF of the LMGARCH(p, d, q) process and discusses an empirical example. In the conclusions we suggest future developments.

3.2 The Long Memory GARCH Model

To establish terminology and notation, we define an LMGARCH(p, d, q) process $\{\varepsilon_t\}$ by the equations (see Karanasos et al., 2004b)

$$\varepsilon_t = Z_t \sqrt{h_t}, \quad t \in \mathbb{Z}, \tag{3.1}$$

and

$$h_t = \omega + [\Omega(L) - 1]\, v_t, \tag{3.2}$$

with

$$\Omega(L) = \sum_{j=0}^{\infty} \omega_j L^j \triangleq \frac{B(L)}{\Phi(L)(1-L)^d},$$

where $v_t \triangleq \varepsilon_t^2 - h_t$. Here and in the remainder of this chapter, L stands for the lag operator and the symbol '$\triangleq$' is used to indicate equality by definition. We assume hereafter that $\omega \in (0, \infty)$, $0 < d < 0.5$ and that the finite order polynomials $\Phi(L) \triangleq 1 - \sum_{i=1}^{q} \phi_i L^i = \prod_{i=1}^{q}(1 - \zeta_i L)$, $B(L) \triangleq -\sum_{i=1}^{p} \beta_i L^i$ ($\beta_0 \triangleq -1$) have zeros outside the unit circle in the complex plane.

The rescaled innovations Z_t are assumed to be *i.i.d.* with $\mathbf{E}(Z_t) = \mathbf{E}(Z_t^2 - 1) = 0$. By (3.1) and the *i.i.d.*-ness of the Z_t, $\mathbf{E}(v_t|\mathcal{F}_{t-1}) = 0$ where $\mathcal{F}_t$ is the σ-field of events induced by $\{\varepsilon_s, s \leqslant t\}$. For notational convenience, in what follows we denote $\widetilde{\omega}_m \triangleq \sum_{j=0}^{\infty} \omega_j \omega_{j+m}$ ($m = 0, 1, 2, \ldots$). Note that $d < 0.5$ implies $\widetilde{\omega}_0 < \infty$. The ε_t^2 has finite first moment equal to ω. In addition, simple manipulations suggest that $\mathbf{E}(e_t^4) < \infty$ and $[\mathbf{E}(e_t^4) - 1]\widetilde{\omega}_0 < \mathbf{E}(e_t^4)$ imply covariance stationarity of the ε_t^2. Under these conditions the autocorrelations $\{\rho_m(\varepsilon_t^2) \triangleq \mathbf{Corr}(\varepsilon_{t+m}^2, \varepsilon_t^2)\}$ are $\rho_m(\varepsilon_t^2) = \widetilde{\omega}_m / \widetilde{\omega}_0$ (see Karanasos et al., 2004b).

Moreover, h_t has an ARCH(∞) representation, i.e. it can be expressed as an infinite distributed lag of ε_{t-j}^2 terms as $(h_t - \omega) = \Psi(L)(\varepsilon_t^2 - \omega)$, where $\Psi(L) = \sum_{j=1}^{\infty} \psi_j L^j \triangleq [1 - \Phi(L)(1-L)^d / B(L)]$. In this specification the conditional variance and the squared errors are expressed in deviations from the unconditional variance ω. To guarantee the non-negativity of the conditional variance one has to impose inequality constraints which ensure that $\psi_j \geq 0$ for $j = 1, 2, \ldots$. Necessary and sufficient constraints for $p \leq 2$ and sufficient constraints for $p > 2$ can be found in Conrad and Haag (2006) [see Chapter 2].

Furthermore, under (3.1) – (3.2), the coefficients ω_j decay at a slow hyperbolic rate so that $\omega_j = O(j^{d-1})$. This in turn implies that the autocorrelations satisfy $\rho_m(\varepsilon_t^2) = O(m^{2d-1})$, provided $\mathbf{E}(\varepsilon_t^4) < \infty$. Hence, when the fourth moment of the ε_t exists, ε_t^2 is a weakly stationary process which exhibits long memory for all $d \in (0, 0.5)$, in the sense that the series $\sum_{m=0}^{\infty} |\rho_m(\varepsilon_t^2)|$ is properly divergent. For this reason, we refer to a process ε_t satisfying (3.1) and (3.2) as an LMGARCH(p, d, q) process.

Finally, it is interesting to note the difference between the LMGARCH process and the FIGARCH model. The ARCH(∞) formulation of the latter is $h_t = \omega + \Psi(L)\varepsilon_t^2$. It is noteworthy that this model, unless $\mathbf{E}(Z_t^2) < 1$, is not compatible with covariance stationary ε_t. However, Zaffaroni (2004) points out that even this weak covariance stationarity condition for the levels ε_t rules out long memory in the squares ε_t^2.

3.3 The IRF of the LMGARCH$(\mathbf{p}, \mathbf{d}, \mathbf{q})$ Model

In the LMGARCH class of models, the short-run behavior of the time series is captured by the conventional ARCH and GARCH parameters, while the long-run dependence is conveniently modelled through the fractional differencing parameter.

Since in the LMGARCH the conditional variance is parameterized as a linear function of the past squared innovations, the persistence of the conditional variance is most simply characterized in terms of the impulse response coefficients for the optimal forecast of the future conditional variance as a function of the current innovation v_t. Following Baillie et al. (1996a) we define the IRF and cumulative IRF as follows:

Definition 3.1. *The IRF of the LMGARCH(p, d, q) model is given by the sequence δ_k, $k = 0, 1, \ldots,$ where*

$$\delta_k \triangleq \frac{\partial \mathbf{E}(\varepsilon_{t+k}^2 | \mathcal{F}_t)}{\partial v_t} - \frac{\partial \mathbf{E}(\varepsilon_{t+k-1}^2 | \mathcal{F}_t)}{\partial v_t},$$

with $\delta_0 \triangleq 1$. Accordingly, the cumulative IRF is given by the sequence $\lambda_k \triangleq \sum_{l=0}^{k} \delta_l$.

The impulse response coefficients δ_k can be simply obtained by considering the first difference of the squared errors

$$(1 - L)\varepsilon_t^2 = \Delta(L)v_t, \tag{3.3}$$

where $\Delta(L) = \sum_{j=0}^{\infty} \delta_j L^j = (1 - L)\Omega(L)$. Since by definition the impulse response coefficients δ_k are related to the cumulative impulse response weights λ_k by $\Delta(L) = (1 - L)\Lambda(L)$, with $\Lambda(L) = \sum_{k=0}^{\infty} \lambda_k L^k$, it follows that the cumulative impulse response weights λ_k coincide with the ω_k coefficients defined by equation (3.2).[1]

Further, let F be the Gaussian hypergeometric function defined by

$$F(a, b; c; z) \triangleq \sum_{j=0}^{\infty} \frac{(a)_j (b)_j z^j}{(c)_j j!},$$

where $(a)_j \triangleq \prod_{i=0}^{j-1}(a + i)$ with $(a)_0 = 1$ is Pochhammer's shifted factorial. Then, the fractional differencing operator $(1 - L)^d$ in (3.2) is most conveniently expressed in terms of the hypergeometric function

$$(1 - L)^d = F(-d, 1; 1; L) = \sum_{j=0}^{\infty} \frac{\Gamma(j - d)}{\Gamma(j + 1)\Gamma(-d)} L^j \triangleq \sum_{j=0}^{\infty} g_{-d}(j) L^j,$$

where $\Gamma(\cdot)$ is the gamma function. Note that using this notation we can write $\Delta(L) = F(d - 1, 1; 1; L) \cdot B(L)/\Phi(L)$.

Next, we establish a representation for the cumulative impulse response function of the LMGARCH(p, d, q) process.

Theorem 3.1. *The cumulative IRF λ_k, $k = 0, 1, \ldots,$ of the LMGARCH(p, d, q) model is given by*

$$\begin{aligned} \lambda_k \;=\; & \left(1 - \sum_{i=1}^{\min\{k,p\}} \beta_i\right) \sum_{i=0}^{\max\{k-p,0\}} \xi_i g_d(\max\{k - p, 0\} - i) \\ & + \sum_{i=\max\{k-p,0\}+1}^{k} \left(1 - \sum_{l=1}^{k-i} \beta_l\right) c_i, \end{aligned} \tag{3.4}$$

[1] Note that since equation (3) is satisfied by both the LMGARCH and the FIGARCH process the results that follow can be applied to the latter as well.

where

$$c_i \triangleq \sum_{l=0}^{i} \xi_{i-l} g_{d-1}(l), \quad \textit{with} \quad \xi_i \triangleq \sum_{l=1}^{q} \theta_l \zeta_l^i, \quad \textit{and} \quad \theta_i \triangleq \frac{\zeta_i^{q-1}}{\prod_{l=1,l\neq i}^{q}(\zeta_i - \zeta_l)}.$$

(Recall that the ζ_i are defined by (3.2)).

Proof. In view of (3.3), we have

$$\Delta(L) = \sum_{j=0}^{\infty}\sum_{i=0}^{p} -c_{j-i}\beta_i L^j, \quad \text{or} \quad \delta_j = \sum_{i=0}^{p} -c_{j-i}\beta_i,$$

where c_i is defined in Theorem 3.1 and $c_{-i} = 0$ $(i = 1, \ldots, p)$.

Therefore, the cumulative IRF is given by

$$\lambda_k = \sum_{i=0}^{k} \delta_i = \left(- \sum_{i=0}^{\min\{k,p\}} \beta_i \right) \sum_{i=0}^{\max\{k-p,0\}} c_i + \sum_{i=\max\{k-p,0\}+1}^{k} \left(1 - \sum_{l=1}^{k-i} \beta_l \right) c_i,$$

where we use the convention that $\sum_{j=1}^{0} \beta_j = 0$.

Hence, in view of the fact that

$$\sum_{i=0}^{\max\{k-p,0\}} c_i = \sum_{i=0}^{\max\{k-p,0\}} \xi_i g_d(\max\{k-p,0\} - i),$$

equation (3.4) follows. □

The long-run impact of past shocks for the volatility process may now be assessed in terms of the limit of the cumulative impulse response weights, i.e.,

$$\Delta(1) = \lim_{k\to\infty} \sum_{l=0}^{k} \delta_l = \lim_{k\to\infty} \lambda_k.$$

Note that the results in Theorem 3.1 hold for $0 \leq d \leq 1$.

As noted by Baillie et al. (1996a), for $0 \leq d < 1$, $F(d-1,1;1;1) = 0$, so that for the LMGARCH(p,d,q) model with $0 < d < 0.5$ and the stable GARCH model with $d = 0$, shocks to the conditional variance will ultimately die out in a forecasting sense. In contrast, for $d = 1$, $F(d-1,1;1;1) = 1$, and the cumulative impulse response weights will converge to the nonzero constant $\Delta(1) = B(1)/\Phi(1)$. Thus, from a forecasting perspective shocks to the conditional variance of the integrated GARCH (IGARCH) model persist indefinitely.

To illustrate the general result we consider the LMGARCH$(1, d, 1)$ process. In this case $B(L) = 1 - \beta_1 L$ and $\Phi(L) = 1 - \phi_1 L$.

Lemma 3.1. *The cumulative IRF λ_k, $k = 0, 1, \ldots$, of the LMGARCH$(1, d, 1)$ are given by*

$$\lambda_k = g_d(k) + (\phi_1 - \beta_1) \sum_{i=1}^{k} \phi_1^{i-1} g_d(k - i). \tag{3.5}$$

Moreover, when $\phi_1 = 0$, equation (3.5) gives the cumulative impulse response weights of the LMGARCH $(1, d, 0)$ model: $\lambda_k = [1 - \beta_1 - (1 - d)/k] \cdot g_d(k - 1)$.[2] By restricting d to being zero in (3.5) and observing that $g_0(0) = 1$ and $g_0(j) = 0$ for $j > 0$, we obtain the cumulative IRF of the GARCH$(1, 1)$ model: $\lambda_k = (\phi_1 - \beta_1)\phi_1^{k-1}$. Finally, when $d = 0$ and $\phi_1 = 1$, we obtain the λ_k of the IGARCH$(1, 1)$ model: $\lambda_k = (1 - \beta_1)$.

The impulse response functions can be used:

(a) to distinguish between short and long memory specifications. Conrad and Karanasos (2005a) [see Chapter 4] model the conditional variance of the monthly US inflation rate for the period 1962 - 2000 as GARCH$(1, 1)$, IGARCH$(1, 1)$ and FIGARCH$(1, d, 1)$, respectively. Figure 3.1 (upper) plots the cumulative IRFs of their parameter estimates for the GARCH$(1, 1)$ model with $\widehat{\phi}_1 = 0.976$ and $\widehat{\beta}_1 = 0.822$, the IGARCH$(1, 1)$ process with $\widehat{\beta}_1 = 0.819$ and the FIGARCH$(1, d, 1)$ specification with $\widehat{\phi}_1 = 0.325$, $\widehat{\beta}_1 = 0.768$ and $\widehat{d} = 0.692$. While a shock to the optimal forecast of the future conditional variance decays at an exponential rate in the GARCH model, and remains important for forecasts of all horizons in the IGARCH model, it vanishes out at a slow hyperbolic rate in the FIGARCH model.

(b) to investigate the persistence properties of a particular LMGARCH specification for different parameter values. For example, Conrad and Haag (2006) [see Chapter 2, Corollary 2.3] show that in the LMGARCH$(1, d, 0)$ model one can allow for $\beta_1 < 0$ (there is a lower bound for β_1 depending on the value of d). Figure 3.1 (lower) plots the IRF for the LMGARCH$(1, d, 0)$ for d fixed at 0.45 and with $\beta \in \{-0.1925, 0, 0.45\}$,

[2] The cumulative impulse response function of the LMGARCH$(1, d, 0)$ model was first derived by Baillie et al. (1996a) (see also, equation (87) in Baillie, 1996).

which is the range of β_1 values allowed for by Corollary 2.1 in Chapter 2. Clearly, the IRFs for the three different values of β_1 help to show that persistence is decreasing in β_1.

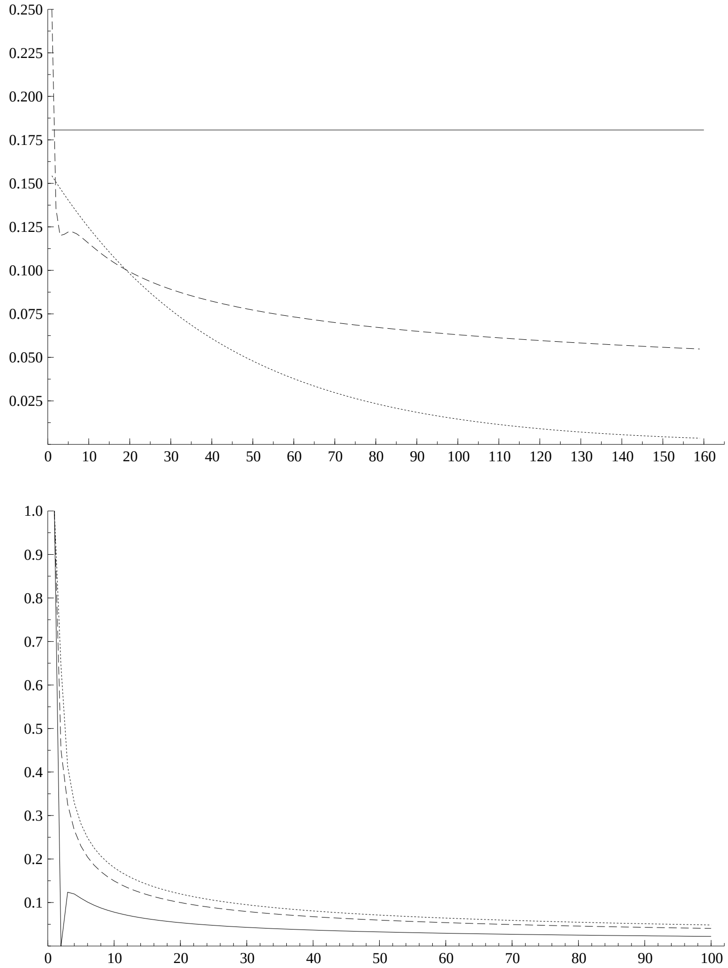

Figure 3.1: Cumulative IRFs for GARCH (dotted), IGARCH (solid) and FIGARCH (dashed) models (upper) and cumulative IRFs for LMGARCH$(1, d, 0)$ (lower) with $d = 0.45$ and $\beta_1 = 0.45$ (solid), $\beta_1 = 0$ (dashed) and $\beta_1 = -0.1925$ (dotted), respectively.

Illustrative Example

As an empirical illustration, we examine the properties of the continuously compounded daily rate of returns for the Deutschmark exchange rate vis-a-vis the US dollar over the period from 31^{st} October 1983 to 31^{st} December 1992 (2,394 observations in total). This data set was also used by Karanasos et al. (2004b). We compare the cumulative IRF of LMGARCH$(1, d, 0)$, LMGARCH$(0, d, 1)$ and LMGARCH$(1, d, 1)$ models. The cumulative impulse response weights were evaluated using the formula in equation (3.5) and the quasi-maximum likelihood (QML) parameter estimates, reported in Table 3.1, obtained under the assumption of conditional Gaussianity.[3] Note, that for all three models the

Table 3.1: QML estimates for LMGARCH models.

	LMGARCH$(1, d, 0)$	LMGARCH$(0, d, 1)$	LMGARCH$(1, d, 1)$
$\widehat{d}$	0.2326 (0.0365)	0.1847 (0.0237)	0.3805 (0.0680)
$\widehat{\phi}_1$	-	-0.1260 (0.0306)	0.2742 (0.0471)
$\widehat{\beta}_1$	0.1973 (0.0460)	-	0.6114 (0.0620)

Notes: Figures in parentheses are asymptotic standard errors.

estimated parameters satisfy the condition in Conrad and Haag (2006) [see Chapter 2, Corollary 2.1] ensuring the non-negativity of the conditional variance. Figure 3.2 plots the cumulative IRF for lags up to 160. The cumulative impulse response weights of the LMGARCH$(1, d, 0)$ and LMGARCH$(0, d, 1)$ decrease much faster than that of the LMGARCH$(1, d, 1)$. The plots of the ACFs in Karanasos et al. (2004b) show a very similar pattern.

3.4 Conclusions

In this chapter we have obtained convenient representations for the cumulative impulse response weights of a process with long memory conditional heteroscedasticity. Since the long memory GARCH model includes the stable and integrated GARCH models as special cases our theoretical results provide a useful tool which facilitates comparison between these major classes of GARCH models. It is worth noting that our results

[3]The parameter estimates from Table 3.1 correspond to those in Karanasos et al. (2004b), p 278.

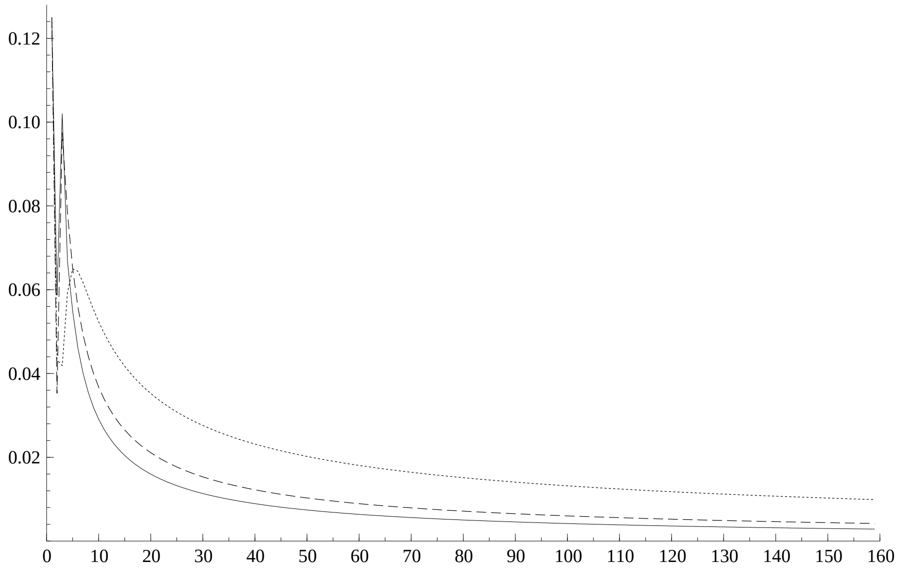

Figure 3.2: Cumulative IRFs for LMGARCH$(1, d, 1)$ (dotted), LMGARCH$(1, d, 0)$ (dashed) and LMGARCH$(0, d, 1)$ (solid) models from Table 3.1.

on the IRF of the general LMGARCH(p, d, q) model extend the results in Baillie et al. (1996a) on the first order LMGARCH model. We should also mention that the methodology used in this chapter can be applied to obtain the impulse response weights of more sophisticated long memory GARCH models, e.g. ARFIMA, asymmetric power, and multivariate LMGARCH models.

Part II

Long Memory GARCH Models: Applications

Chapter 4

On the Inflation-Uncertainty Hypothesis in the USA, Japan and the UK

4.1 Introduction

The issue of the welfare costs of inflation has been one of the most researched topics in macroeconomics both on the theoretical and empirical front. Friedman (1977) argues that a rise in the average rate of inflation leads to more uncertainty about the future rate of inflation. The opposite type of causation between inflation and its uncertainty has also been analyzed in the theoretical macroeconomics literature. Cukierman and Meltzer (1986) argue that central banks tend to create inflation surprises in the presence of more inflation uncertainty. Clarida et al. (1999) emphasize the fact that since the late 1980s a stream of empirical work has presented evidence that monetary policy may have important effects on real activity. Consequently, there has been a great resurgence

---

This chapter was published as: Conrad, C., and M. Karanasos (2005a). "On the inflation-uncertainty hypothesis in the USA, Japan and the UK: a dual long memory approach." *Japan and the World Economy* 17, 327–343.

of interest in the issue of how to conduct monetary policy. If an increase in the rate of inflation causes an increase in inflation uncertainty, one can conclude that greater uncertainty – which many have found to be negatively correlated to economic activity – is part of the costs of inflation. Thus, if we hope ever to give a really satisfactory answer to the questions:

- What actions should the central bankers take?
- What is the optimal strategy for the monetary authorities to follow?

we must first develop some clear view about the temporal ordering of inflation and nominal uncertainty.

In this chapter, the above issues are analyzed empirically for the USA, Japan and the UK with the use of a parametric model of long memory in both the conditional mean and the conditional variance. Our emphasis on these three countries is justified by a number of considerations. First, the USA is the best-documented case and the American experience has played an important role in setting the agenda for previous analysis of monetary policy.[1] Second, the USA and Japan are the two largest economies in the world and changes in their inflation rates (variabilities) and average growth rates have repercussions in the world economy. Third, all three countries experienced wide variations in their conduct of monetary policy in the last forty years. For example, the increase in oil prices in late 1973 was a major shock for Japan, with substantial adverse effects on inflation, economic growth, and the government's budget. In response to an increase in the inflation rate to a level above 20% in 1974 the bank of Japan,

[1] As emphasized by Bernanke and Mishkin (1992), the conduct of monetary policy in the USA is conventionally divided into three regimes. During the 1970s the Fed did not consider meeting money growth targets to be of high priority, placing greater weight on reducing unemployment while maintaining a relatively smooth path for interest rates. However, the change in Fed operating procedures in 1979 was accompanied by a decision by the Fed to place greater weight on monetary targets and to tolerate high and volatile interest rates in order to bring down inflation. The main objectives during the latter part of the 1980s were exchange rate stabilization, financial market stability (particularly after the October 1987 stock market crash) and the maintenance of low and stable inflation.

like other central banks, began to pay more attention to money growth rates.[2] Prior to 1978 the Bank of Japan was committed only to monitoring rather than controlling money growth. However, after 1978 there did appear to be a substantive change in policy strategy, in the direction of being more 'money-focused'. Particularly striking was the different response of monetary policy to the second oil price shock in 1979. The difference in the inflation outcome in this episode was also striking, as inflation increased only moderately with no adverse effects on the unemployment situation. Beginning 1989 asset prices came down as money growth slowed, economic activity weakened and there was a slowdown in lending by Japanese banks. In responding to these developments the Bank of Japan permitted a considerable increase in the variability of broad money growth after the late 1980s (Bernanke and Mishkin, 1992). Finally, these three countries represent 'independent observations' in the sense that, no two of them belonged to a common system of fixed exchange rates.[3] Other countries with independent monetary policies, such as Australia, would be interesting to study but are excluded because of space.

The development of GARCH techniques allows the measurement of inflation uncertainty by the conditional variance of the inflation series and the more accurate testing of the Friedman and the Cukierman and Meltzer hypotheses. Several researchers have examined the inflation-uncertainty relationship using GARCH measures of inflation uncertainty. Many studies on the relationship between inflation and its uncertainty used

[2]See Bernanke and Mishkin (1992) for an excellent discussion of the monetary policy in Japan.

[3]However, as Bernanke and Mishkin (1992) point out, there are some parallels between the recent histories of British and American monetary policies. As in the USA, the British introduced money targeting in the mid-1970s in response to mounting inflation concerns. During the pre-1979 period, the British monetary authorities, like their American counterparts, were not taking their money growth targets very seriously. As in the United States, the perception of an inflationary crisis led to a change in strategy in 1979. Overall, a comparison with the US and the other countries does not put British monetary policy in a favorable light. However, in the 1980s British inflation performance did improve considerably, remaining well below the 1970s level and becoming significantly less variable (Bernanke and Mishkin, 1992).

GARCH type models with a joint feedback between the conditional mean and variance of inflation (i.e, Brunner and Hess, 1993, and Fountas et al., 2003). In contrast, Grier and Perry (1998) and Fountas and Karanasos (2005) use the estimated conditional variance from GARCH type models and employ Granger methods to test for the direction of causality between average inflation and inflation uncertainty for the G7.[4] All the preceding works use traditional ARMA processes to model the conditional mean of inflation. On the other hand, Brunner and Hess (1993) argue that the US inflation rate was stationary before the 1960s, but that it has possessed a unit root since this time. Subsequently, Hassler and Wolters (1995) have found that the inflation rates of five industrial countries were well explained by different orders of integration, which varied around the stationarity border of 0.5. Baum et al. (1999) also found evidence that both CPI- and WPI- based inflation rates for many industrial as well as developing countries are fractionally integrated with a differencing parameter that is significantly different from zero or unity. Along these lines, Baillie et al. (1996b) and Hwang (2001) estimate various ARFIMA-GARCH-in mean models where lagged inflation is included in the variance equation.

In a recent paper, Baillie et al. (2002) found that inflation has the rather curious property of persistence in both its first and its second conditional moments. They introduce the ARFIMA-fractionally integrated GARCH (ARFIMA-FIGARCH) model, which is sufficiently flexible to handle the dual long memory behavior encountered in inflation. To this end, this study uses an ARFIMA-FIGARCH type model to generate a time-varying conditional variance of surprise inflation. This model has a distinct advantage for this application: it nests several alternative GARCH models of conditional heteroscedasticity. With this conditional variance as a measure of inflation uncertainty,

[4]We should also mention that several empirical studies (i.e., Grier and Perry, 2000, and Fountas et al. 2002) use bivariate GARCH type models to estimate simultaneously the conditional means, variances and covariances of inflation and output growth. These models make it possible to test for evidence on all the bidirectional causality relationships between inflation, output growth, and uncertainty about inflation and output growth.

it then employs Granger methods to test the direction of causality between average inflation and uncertainty. The Granger-causality approach is adopted in this chapter instead of the simultaneous-estimation approach because it allows us to capture the lagged effects between the variables of interest. In addition, the former approach minimizes the number of estimated parameters whereas the latter approach is subject to criticism on the grounds of the potential negativity of the variance.

This chapter is organized as follows: Section 4.2 considers the hypotheses about the causality between inflation and inflation uncertainty in more detail and provides the model. Section 4.3 discusses the data and the results and Section 4.4 summarizes the main conclusions.

4.2 Theory and Model

4.2.1 The Relation between Inflation and Inflation-Uncertainty

Friedman (1977) outlined an informal argument regarding the real effects of inflation. Friedman's point comes in two parts. In the first part, an increase in inflation may induce an erratic policy response by the monetary authority and therefore lead to more uncertainty about the future rate of inflation. The second part of Friedman's hypothesis predicts that inflation uncertainty causes an adverse output effect.[5] Ball (1992), using an asymmetric information game, offers a formal derivation of Friedman's hypothesis that higher inflation causes more inflation uncertainty. It is also possible that more inflation will lead to a lower level of inflation uncertainty. The argument advanced by Pourgerami and Maskus (1987) is that in the presence of rising inflation agents may invest more resources in forecasting inflation, thus reducing uncertainty about inflation. A formal analysis of this effect is presented in Ungar and Zilberfarb (1993).

The causal effect of inflation uncertainty on inflation has been analyzed in the theo-

[5] Investigating the causal impact of nominal uncertainty on output growth is an interesting avenue for future research, but is beyond the scope of the chapter.

retical macro literature by Cukierman and Meltzer (1986). Using the well-known Barro-Gordon model, Cukierman and Meltzer (1986) show that an increase in uncertainty about money growth and inflation will raise the optimal average inflation rate because it provides an incentive to the policymaker to create an inflation surprise in order to stimulate output growth. Therefore, the prediction of the Cukierman-Meltzer analysis is that higher inflation uncertainty leads to more inflation. Holland (1995) has supplied a different argument based on the stabilization motive of the monetary authority, the so-called "stabilizing Fed hypothesis". He claims that as inflation uncertainty rises due to increasing inflation, the monetary authority responds by contracting money supply growth, in order to eliminate inflation uncertainty and the associated negative welfare effects. Hence, Holland's argument supports a negative causal effect of inflation uncertainty on inflation.

4.2.2 The Econometric Specification

In this section we present the ARFIMA-FIGARCH model, which generates the long memory property in both the first and second conditional moments, and is thus sufficiently flexible to handle the dual long memory behavior encountered in inflation.

In the ARFIMA($p, d_m, 0$)-FIGARCH($1, d_v, 1$) model the mean equation is defined as

$$(1 - \varphi_1 L - \varphi_{12} L^{12} - \varphi_{24} L^{24})(1 - L)^{d_m}(\pi_t - \mu) = \varepsilon_t, \quad (4.1)$$

where π_t denotes the inflation rate and $0 \leq d_m \leq 1$ is the fractional differencing parameter of the mean.[6] The innovation ε_t is conditionally normal with mean zero and variance h_t. That is $\varepsilon_t | \mathcal{F}_{t-1} \sim \mathcal{N}(0, h_t)$, where $\mathcal{F}_{t-1}$ is the information set up to time $t-1$. The structure of the conditional variance is

$$(1 - \beta L) h_t = \omega + \left[(1 - \beta L) - (1 - \phi L)(1 - L)^{d_v}\right] \varepsilon_t^2, \quad (4.2)$$

[6] The fractional differencing operator $(1 - L)^{d_m}$ is most conveniently expressed in terms of the hypergeometric function. For details see Chapter 3.3.

where $0 \leq d_v \leq 1$, $\omega > 0$, and $\phi, \beta < 1$. Moreover, the parameters (β, d_v, ϕ) have to satisfy the conditions in Conrad and Haag (2006) [see Chapter 2, Corollary 2.1] to guarantee the nonnegativity of the conditional variance.

If $h_t = \omega$, a constant, the process reduces to the ARFIMA$(p, d_m, 0)$ model and π_t will be covariance stationary and invertible for $-0.5 < d_m < 0.5$ and will be mean reverting for $d_m < 1$. Although the ARFIMA-FIGARCH process is strictly stationary and ergodic for $0 \leq d_v \leq 1$, it will have an infinite unconditional variance for all d_m given a $d_v \neq 0$.

The ARFIMA-FIGARCH model in (4.1) and (4.2) has a distinctive feature. It allows us simultaneously to estimate the degree of persistence in both inflation and uncertainty about inflation. It also has the advantage of keeping the analytical elegance of the ARMA-GARCH model while enhancing its dynamics. Put differently, the ARFIMA-FIGARCH model has at least two important implications for our understanding of inflation and inflation uncertainty. First, it recognizes the long memory aspect of the inflation rate and provides an empirical measure of inflation uncertainty that accounts for long memory in the second conditional moment of the inflation process. Second, it allows for a more systematic comparison of many possible models that can capture the features of the inflation series.

4.3 Empirical Analysis

4.3.1 Data

We use monthly data on the CPI (Consumer Price Index) obtained from the *OECD Statistical Compendium* as proxies for the price level.[7] The data range from 1962:01 to

[7]Since most of the studies use CPI based inflation measures (i.e., Caporale and McKierman, 1997, Grier and Perry, 1998, and Baillie et al., 2002) we construct our inflation and inflation uncertainty measures from the Consumer Price Index. Alternatively, one can use either the Producer Price Index (PPI) or the GNP deflator. Brunner and Hess (1993) use all three measures of inflation but they discuss only the results using CPI inflation. Grier and Perry (2000) use both (CPI and PPI) indices and find that the CPI and PPI results are virtually identical.

2000:12 and cover three industrial countries, namely, the USA, the UK and Japan. Inflation is measured by the monthly difference of the log CPI, i.e. $\pi_t = 100 \cdot \log(\text{CPI}_t/\text{CPI}_{t-1})$. Allowing for differencing leaves 469 usable observations. The inflation rates of the USA, UK and Japan are plotted in Figure 4.1. Table 4.1 presents summary statistics for the

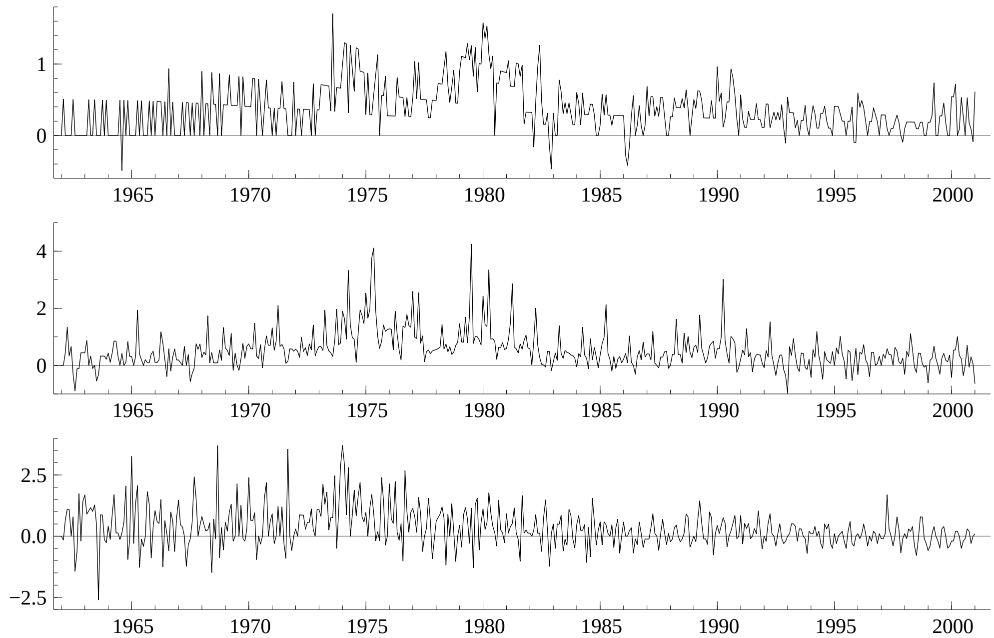

Figure 4.1: Monthly inflation rates for the US (upper), the UK (middle) and Japan (bottom) for the period 1962:01 to 2000:12.

three inflation rates. The results indicate that the distributions of all three inflation series are skewed to the right and leptokurtic. The large values of the Jarque-Bera (JB) statistic imply a deviation from normality, and the significant $Q-$statistics of the squared deviations of the inflation rate from its sample mean indicate the existence of ARCH effects. This evidence is also supported by the LM statistics, which are highly significant. The autocorrelations of CPI inflation for the three countries (not reported) exhibit the clear pattern of slow decay and persistence. The autocorrelations of the first differenced inflation series (not reported) appear to be overdifferenced with large negative autocorrelations at lag one.

Table 4.1: Summary statistics for inflation rates.

	μ	σ	K	S	JB	Q^2_{12}	LM
USA	0.376	0.341	3.77 [0.00]	0.81 [0.00]	63.47 [0.00]	1123.81 [0.00]	130.88 [0.00]
UK	0.550	0.656	9.46 [0.00]	1.90 [0.00]	1098.20 [0.00]	250.09 [0.00]	30.53 [0.00]
Japan	0.348	0.794	5.47 [0.00]	0.93 [0.00]	186.94 [0.00]	129.27 [0.00]	9.02 [0.00]

Notes: μ denotes the average inflation rate over the period 1962:01 to 2000:12, and σ its standard deviation. K and S are the estimated kurtosis and skewness respectively. JB is the Jarque-Bera statistic for normality. Q^2_{12} is the 12-th order Ljung-Box test for serial correlation in the squared deviations of the inflation rate from its sample mean. The Engle test for ARCH effects is denoted by LM. Numbers in brackets are p-values.

Phillips and Perron (1988) (henceforth PP) and Kwiatkowski, Phillips, Schmidt and Shin (1992) (henceforth KPSS) develop two alternative approaches to testing for unit roots. Table 4.2 presents the results of applying the PP and KPSS tests to the three inflation series. In all cases we reject both the KPSS and PP statistics. Hence, for all three countries there is evidence that inflation may not be generated by either an I(0) or I(1) process and this is at least indicative of fractional integration (see also Baillie et al., 1996b).

Table 4.2: Tests for the order of integration of different countries' inflation series.

	H_0: I(1)		H_0: I(0)	
	$Z(t_\mu)$	$Z(t_\tau)$	$\widehat{\eta}_\mu$	$\widehat{\eta}_\tau$
USA	7.11***	8.49***	0.84***	0.68***
UK	5.80***	8.33***	1.03***	0.64***
Japan	3.96***	10.47***	2.31***	0.29***

Notes: $Z(t_\mu)$ and $Z(t_\tau)$ are the Phillips-Perron adjusted t-statistics of the lagged dependent variable in a regression with intercept only, and intercept and time trend included respectively. The 0.01 critical values for $Z(t_\mu)$ and $Z(t_\tau)$ are 3.43 and 3.96. $\widehat{\eta}_\mu$ and $\widehat{\eta}_\tau$ are the KPSS test statistics based on residuals from regressions with an intercept and intercept and time trend, respectively. The 0.01 critical values for $\widehat{\eta}_\mu$ and $\widehat{\eta}_\tau$ are 0.739 and 0.216. *** denotes significance at the 0.01 level.

4.3.2 Estimated Models of Inflation

We proceed with the estimation of the ARFIMA$(p, d_m, 0)$-FIGARCH$(1, d_v, 1)$ model described by equations (4.1) and (4.2) in order to take into account the serial correlation observed in the levels and squares of our time series data, and to capture the possible long memory in the conditional mean and the conditional variance. We estimate the ARFIMA-FIGARCH models using the quasi-maximum likelihood estimation (QMLE) method as implemented by Laurent and Peters (2002) in Ox.[8] Table 4.3 reports the results for the period 1962-2001.

The seasonal autoregressive parameters ($\varphi_1, \varphi_{12}, \varphi_{24}$) were necessary to account for the significant seasonality, which is evident for all three countries.[9] The φ_1 parameter was only significant for the USA. The estimated ARCH parameters ($\widehat{\beta}, \widehat{d}_v, \widehat{\phi}$) for the US inflation are significant and satisfy the set of necessary and sufficient conditions derived in Conrad and Haag (2006) [see Chapter 2, Corollary 2.1] guaranteeing the nonnegativity of the conditional variance. For the UK and Japan, the Akaike and Schwarz information criteria (hereafter, AIC and SIC respectively) come out in favor of the FIGARCH$(0, d_m, 0)$ model. The estimated long memory conditional mean parameter is in the range $0 < \widehat{d}_m < 0.37$. The estimated value of d_m for Japan in Table 4.3 is 0.043, which is significantly different from zero at the 0.26% level and implies some very mild long-memory features.[10] In all countries the estimates for the fractional differencing parameter ($\widehat{d}_v$) are relatively large and are statistically significant.[11] Finally, with all three countries, the hypothesis of uncorrelated standardized and squared standardized residuals is well supported by the Ljung-Box test statistics, indicating that there is no statistically significant evidence of misspecification.

[8]The consistency and asymptotic normality of the QMLE has been established only for specific special cases of the ARFIMA and/or FIGARCH model. However, a detailed Monte-Carlo study, where ARFIMA-FIGARCH type models were simulated, was performed by Baillie et al. (2002) and it was found that the quality of the application of the QMLE is generally very satisfactory.

[9]Alternatively, we also estimated a moving average specification with parameters ($\theta_1, \theta_{12}, \theta_{24}$), but the AIC and SIC information criteria came out in favor of the autoregressive specification.

[10]Although d_m is insignificant for Japan it seems to improve the performance of the model since restricting d_m to being zero results in Ljung-Box test statistics which indicate serial correlation in the standardized residuals. Moreover, note that in Baillie et al. (1996b) the estimate of the d_m parameter for Japan is also insignificant.

[11]The estimates for d_m and d_v in Baillie et al. (2002) are quite close to the ones we obtain. In particular, for the USA and the UK they estimated a $d_m(d_v)$ of 0.414 (0.644) and 0.364 (0.633) respectively, while for Japan they estimate a d_v of 0.317.

Table 4.3: ARFIMA-FIGARCH models 1962:01–2000:12.

	$\widehat{\mu}$	$\widehat{d}_m$	$\widehat{\varphi}_{12}$	$\widehat{\varphi}_{24}$	$\widehat{\omega}$	$\widehat{d}_v$	$\widehat{\phi}$	$\widehat{\beta}$	Q_{12}	Q^2_{12}
USA	0.296	0.342	0.161	0.139	0.002	0.692	0.325	0.768	22.21	3.40
	(2.018)	(4.787)	(3.491)	(2.719)	(1.538)	(1.892)	(1.747)	(6.863)	[0.04]	[0.99]
UK	0.100	0.367	0.407	0.331	0.022	0.474	-	-	14.01	12.27
	(0.208)	(4.479)	(6.911)	(5.828)	(1.689)	(4.025)			[0.30]	[0.42]
Japan	0.098	0.043	0.361	0.263	0.026	0.288	-	-	20.05	12.82
	(1.084)	(1.133)	(6.313)	(5.033)	(1.089)	(5.194)			[0.07]	[0.38]

Notes: For each of the three inflation series, the table reports QMLE parameter estimates for the ARFIMA-FIGARCH model. The numbers in parentheses are t-statistics. Q_{12} and Q^2_{12} are the 12-th order Ljung-Box tests for serial correlation in the standardized and squared standardized residuals respectively. The numbers in [·] are *p*-values. For the USA we estimate a $\widehat{\varphi}_1$ of -0.148 (-2.024).

Table 4.4: Tests of USA fractional differencing parameters

	H_0:ARMA (d_m=0)		H_0: GARCH (d_v=0)	H_0:ARMA-GARCH (d_m=0, d_v=0)	
$\widehat{d}_m$	LR	$\widehat{d}_v$	LR	LR	W
0.342	35.45	0.692	4.14	36.56	15.97
{0.071}	[0.00]	{0.365}	[0.04]	[0.00]	[0.00]

Notes: Columns 2, 4 and 5 report the value of the following likelihood ratio test: $LR = 2[ML_u - ML_r]$, where ML_u and ML_r denote the maximum log-likelihood values of the unrestricted and restricted models respectively. The last column reports the Wald statistic. The numbers in $[\cdot]$ are p-values. The numbers in $\{\cdot\}$ are standard errors.

To test for the persistence in the first two conditional moments of the three inflation series, we examine the likelihood ratio (LR) tests for the linear constraints $d_m = 0$ ('ARMA' model) and $d_v = 0$ ('GARCH' model). We also test the joint hypothesis that $d_m = d_v = 0$ using both an LR test and a Wald (W) statistic. As seen in Table 4.4 for the USA the LR and W statistics clearly reject the 'ARMA' and/or the 'GARCH' null hypotheses against the ARFIMA-FIGARCH model. Similar results are obtained for the UK and Japan but are omitted for space considerations. The evidence obtained from the Wald and LR tests is reinforced by the model ranking provided by the AIC and SIC model selection criteria. In almost all cases the criteria (not reported) favor the ARFIMA-FIGARCH model over both the ARMA-FIGARCH and ARFIMA-GARCH models.[12] Hence, from the various diagnostic statistics it appears that monthly CPI inflation has long memory behavior in both its first and its second conditional moments.

[12]For Japan the LR test, the W statistic and the selection criteria favor the ARMA-FIGARCH model over the ARFIMA-FIGARCH model.

Predictability of Higher Levels of Inflation

In the USA inflation accelerated two years after the augmentation of defense spending in connection with the Vietnam war which took place in the mid-1965. Moreover, the increase in the oil prices by OPEC in the fourth quarter of 1973 and the progressive elimination of control on prices and wages amplify the acceleration of American inflation in 1974. Finally, the considerable fluctuation of oil prices during the period 1979-1980 led the federal reserve to implement a new restrictive monetary policy. Since the early 1980s the American economy embarked on a productivity growth phase supported by a decrease of oil prices and the reduction of the inflation rate. The USA inflation and inflation uncertainty series are shown in Figure 4.2, which plots the inflation rate and its corresponding conditional standard deviation from the ARFIMA-FIGARCH model.

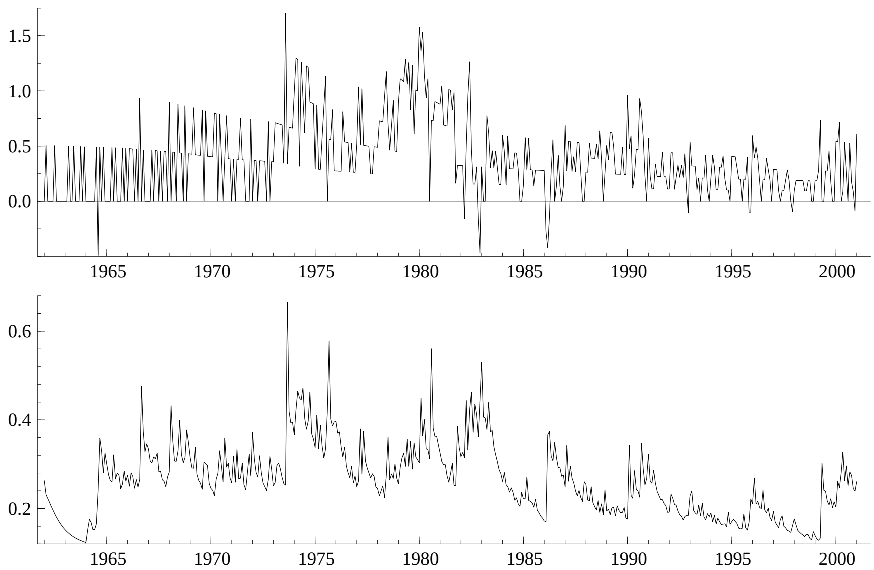

Figure 4.2: US inflation rate (upper) and conditional standard deviation (lower) for the period 1962:01 to 2000:12.

Some researchers, such as Cosimano and Jansen (1988), have failed to find strong evidence that higher rates of inflation are less predictable. Using the dual long memory specification, we compare our results with theirs. In contrast to the conclusion of these studies, Figure 4.2 provides evidence that higher levels of inflation are less predictable. According to our estimates, the conditional standard deviation average (annual rate) in the low-inflation 1960s is about 3.1%. In the high-inflation 1970s, the conditional standard deviation average (annual rate) is about 3.9%. Finally, in the low-inflation environment of the 1990s, the average of the conditional standard deviation is only 2.4%. Similar figures for the UK and Japan are omitted for reasons of brevity but are available from the authors on request.

Persistence in Volatility

In order to illustrate how a shock to the conditional variance decays over time in the FIGARCH$(1, d, 1)$ model we plot in Figure 4.3 the cumulative impulse response function for the USA. The cumulative impulse response weights λ_k for the optimal forecast of the future conditional variance in the FIGARCH(1,d,1) are derived in Conrad and Karanasos (2006) [see Chapter 3, Lemma 3.1] as

$$\lambda_k = \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)} + (\phi_1 - \beta_1) \sum_{i=1}^{k} \phi_1^{i-1} \frac{\Gamma(k-i+d)}{\Gamma(k-i+1)\Gamma(d)}. \tag{4.3}$$

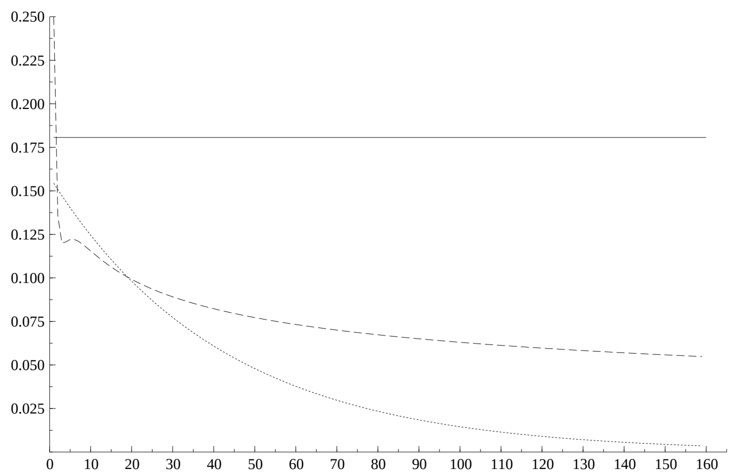

Figure 4.3: Cumulative IRFs for the conditional variance of the US inflation rate: GARCH (dotted), IGARCH (solid) and FIGARCH (dashed).

The cumulative impulse response function of the FIGARCH model is compared with the one of the stable GARCH and integrated GARCH (IGARCH).[13] For the stable GARCH we have $\lambda_k = (\phi - \beta)\phi^{k-1}$ for all $k \geq 1$ and for the IGARCH we have $\lambda_k = 1 - \beta$ for all $k \geq 1$. In the stable GARCH case a shock decays at a fast exponential rate whereas in the IGARCH case it persists forever. In sharp contrast, the shock decays at a slow hyperbolic rate in the FIGARCH case.[14]

4.3.3 Granger-Causality Tests

In this section we report results of Granger-causality tests to provide some statistical evidence on the nature of the relationship between average inflation and nominal uncertainty. Following Granger (1969) the following bivariate autoregression is used to test for causality between the inflation rate and its uncertainty

$$\begin{bmatrix} \pi_t \\ h_t \end{bmatrix} = \begin{bmatrix} \alpha_\pi \\ \alpha_h \end{bmatrix} + \sum_{i=1}^{k} \begin{bmatrix} c_{\pi\pi,i} & c_{\pi h,i} \\ c_{h\pi,i} & c_{hh,i} \end{bmatrix} \begin{bmatrix} \pi_{t-i} \\ h_{t-i} \end{bmatrix} + \begin{bmatrix} e_{\pi t} \\ e_{ht} \end{bmatrix}, \tag{4.4}$$

where $e_t = [e_{\pi t}, e_{ht}]'$ is a bivariate white noise with mean zero and nonsingular covariance matrix Σ_e. The test of whether π_t (h_t) strictly Granger-causes h_t (π_t) is simply a test of the joint restriction that all the $c_{h\pi,i}$ ($c_{\pi h,i}$), $i = 1, \ldots, k$, are zero. In each case, the null hypothesis of no Granger-causality is rejected if the exclusion restriction is rejected. Bidirectional feedback exists if all the elements $c_{\pi h,i}$, $c_{h\pi,i}$, $i = 1, \ldots, k$, are jointly significantly different from zero. However, if the variables are non-stationary, Park and Phillips (1989) and Sims et al. (1990) have shown that the conventional asymptotic theory is not applicable to hypothesis testing in levels VAR's. In addition, Tsay and Chung (2000) in their analysis of spurious regression with fractionally integrated processes find that no matter whether the dependent variable and the regressor are sta-

[13] For the stable GARCH we estimated $\widehat{\beta} = 0.822$ and $\widehat{\phi} = 0.976$. The $\widehat{\beta}$ coefficient in the IGARCH was 0.819.

[14] Similar plots are available for the other two countries but are omitted for reasons of brevity.

tionary or not, as long as their fractional orders of integration sum up to a value greater than 0.5, the t ratios become divergent.

Therefore, we utilize the methodology developed by Toda and Yamamoto (1995) to test for causality between the inflation rate and its uncertainty, which leads to a χ^2 distributed test statistic despite any possible non-stationarity or cointegration between the two series.[15] In other words, the advantage of this procedure is that it does not require a knowledge of cointegrated properties of the system (see Zapata and Rambaldi, 1997). The test is performed in two steps. In the first step the optimal lag length (k) of the system is determined by utilizing the AIC and SIC information criteria. In the second step a VAR of order $k^* = k + d_{\max}$ is estimated (where $d_{\max}$ is the maximal (integer) order of integration suspected to occur in the system) and a modified Wald (MW) test is applied to the first k VAR coefficient matrices to make Granger-causal inference. This MW test statistic has an asymptotic χ^2 distribution with k degrees of freedom. For the USA, Japan and the UK both the AIC and SIC information criteria came out in favor of a VAR with 8, 4 and 12 lags, respectively. Since all variables are fractionally integrated with $d_m, d_v < 1$ we set $d_{\max} = 1$ and estimate VAR models with $k^* = k + 1$ lags. To ensure that our results are not sensitive to the choice of the lag length we report in Table 4.5 for all three countries the MW tests using 4, 8 and 12 lags, as well as the sums of lagged coefficients.

Panel A reports evidence on the Friedman hypothesis. Statistically significant effects are present for all countries. Panel B reports the results of the causality tests where causality runs from the nominal uncertainty to the rate of inflation. This panel provides strong evidence in favor of the Cukierman and Meltzer hypothesis for Japan. For the USA the effect of inflation uncertainty on average inflation is positive but insignificant at any lag length. In contrast, we obtain mixed evidence for the UK. In particular, at 8 lags uncertainty about inflation has a positive impact on inflation as predicted by Cukierman and Meltzer, whereas the value of the MW test statistic and the sign of the sum of lagged coefficients at 12 lags (optimal lag length) provide support for the

[15] We are grateful to an anonymous referee for calling this paper to our attention.

Holland hypothesis.

Table 4.5: Granger-causality tests 1962:01–2000:12.

	USA	UK	JAPAN
Panel A. H_0: Inflation does not Granger-cause inflation uncertainty			
4 (5)	15.39***(+)	14.29***(+)	**13.47***(+)**
8 (9)	**19.39**(+)**	35.35***(+)	16.08**(+)
12 (13)	27.36***(+)	**36.82***(+)**	28.02***(+)
Panel B. H_0: Inflation uncertainty does not Granger-cause inflation			
4 (5)	2.58(+)	3.04(–)	**19.58***(+)**
8 (9)	**7.16(+)**	31.41***(+)	22.25***(+)
12 (13)	11.68(+)	**63.21***(-)**	40.79***(+)

Notes: The figures are *MW* statistics. The numbers in the first column give the lag structure and in parentheses the order of the VAR. The bold numbers indicate the optimal lag length chosen by AIC and SIC. A +(-) indicates that the sum of the lagged coefficients is positive (negative). ***, ** and * denote significance at the 0.01, 0.05 and 0.10 levels, respectively.

When Grier and Perry (1998) look for institutional reasons why the inflation response to increased uncertainty varies across countries, they note that countries associated with an opportunistic response have much lower central bank independence than the countries associated with a stabilizing response. We use measures of central bank independence provided by Alesina and Summers (1993), which constructed a 1-4 (maximum independence) scale of central bank independence. The USA and Japan have relatively independent central banks with a score of 3. However, in Japan increased inflation uncertainty raises inflation, while in the USA uncertainty does not Granger-cause inflation. Thus, one cannot argue that the most independent central banks are in countries where inflation falls in response to increased uncertainty. The UK has a relatively dependent central bank, with a score of 2. However, in the UK the sign (and significance) of the effect varies with the lag length. Thus, a lack of independence does not correspond to 'opportunistic behavior'.

4.3.4 Comparison with other Work

The GARCH time series studies that examine the link between inflation rates and inflation uncertainty use various sample periods, frequency data sets and empirical methodologies.[16] Some GARCH studies of this issue utilize the simultaneous estimation approach. In particular, Caporale and McKierman (1997), and Fountas (2001) estimate GARCH-type processes where lagged inflation is included in the conditional variance equation to test Friedman's hypothesis. Brunner and Hess (1993) model the conditional variance as a nonlinear function of lagged values of inflation. Baillie et al. (1996b), Hwang (2001) and Fountas et al. (2003) employ univariate GARCH models that allow for simultaneous feedback between the conditional mean and variance of inflation while Grier and Perry (2000) use a bivariate GARCH-in-mean specification.[17] Some researchers employ the Granger-causality approach. For example, Grier and Perry (1998) and Fountas and Karanasos (2005) estimate univariate component GARCH models, Fountas et al. (2004) employ an EGARCH specification, while Fountas et al. (2002) use a bivariate constant correlation GARCH formulation.

More specifically, Baillie et al. (1996b) show that for the low-inflation countries (except the UK) there is no link between the inflation rate and its uncertainty whereas for the high-inflation countries strong evidence is provided of a bidirectional feedback between nominal uncertainty and inflation. Grier and Perry (1998) find that in all G7 countries inflation significantly raises inflation uncertainty. Fountas et al. (2004), in five out of six European countries, and Fountas and Karanasos (2005) in six of the G7 countries also find support for Friedman's hypothesis. In sharp contrast, for Germany

[16] Fountas (2001) employs annual data on UK CPI for the period 1885-1998 while Fountas et al. (2004) use quarterly data from six European countries for the period 1960-1999.

[17] These studies either use the conditional variance or the conditional standard deviation as a regressor in the conditional mean.

the study by Fountas and Karanasos (2005) finds that the effect of inflation on nominal uncertainty is negative as predicted by Pourgerami and Maskus (1987). These three studies find less robust evidence regarding the direction of the impact of a change in inflation uncertainty on inflation. That is, they find evidence in favor of the Cukierman-Meltzer hypothesis for some countries and in favor of the Holland hypothesis for other countries.

In all three countries our results on the Friedman hypothesis are identical to those of the above-mentioned studies. That is, in all studies which use the two-step approach and in most of the studies which use the simultaneous approach increased inflation affects nominal uncertainty positively. In sharp contrast, Baillie et al. (1996b) find, for the USA and Japan, that inflation uncertainty is independent of changes in inflation. Hwang (2001) finds that the US inflation affected its uncertainty weakly and negatively. For the USA, our result that there is no causal effect of nominal uncertainty on inflation squares with the findings of most of the recent studies (e.g., Baillie et al., 1996b, Grier and Perry, 2000, Hwang, 2001, and Fountas and Karanasos, 2005). However, Grier and Perry (1998) find that uncertainty about future inflation has a negative impact on inflation whereas Fountas et al. (2003) find evidence for a positive effect of nominal uncertainty on inflation. Note that both studies estimate short-memory GARCH models. For Japan we find that uncertainty about inflation has a positive effect on inflation, as predicted by Cukierman and Meltzer. This result is in agreement with the conclusion of Grier and Perry (1998) and Fountas and Karanasos (2005). In sharp contrast, Fountas et al. (2002) provide strong empirical support for Holland's hypothesis. Our work differs from Fountas et al. (2002) in the chosen econometric methodology (univariate dual long-memory GARCH-type models) and the use of CPI in measuring inflation. The authors estimate short-memory bivariate GARCH models and use PPI data. Moreover, Baillie et al. (1996b) fail to find any effect of nominal uncertainty on inflation for Japan. Our work differs from theirs in that we use more than one lag of monthly inflation and uncertainty to look for a link between the two. Finally, for the UK our result that uncertainty about future inflation appears to have a mixed impact on inflation is consistent with the findings of previous studies by Fountas et al. (2004) and Fountas and Karanasos (2005). Note also that our result on the mild evidence (at lag 12) that increased nominal uncertainty lowers inflation is identical to that of Grier and Perry (1998).

4.4 Conclusions

We have used monthly data on inflation in the USA, Japan, and the UK to examine the possible relationship between inflation and nominal uncertainty, and hence test a number of economic theories. The results in this chapter highlight the importance of modeling long memory not only in the conditional mean of inflation but in the conditional variance as well. The application of the ARFIMA-FIGARCH approach allows us to derive two important conclusions: First, the Friedman hypothesis that inflation leads to more inflation uncertainty applies in all countries. Since an increase in the rate of inflation causes an increase in inflation uncertainty, we conclude that greater uncertainty – which many have found to be negatively correlated to economic activity – is part of the costs of inflation. This result may have important implications for the inflation-output relationship. Gylfason and Herbertsson (2001) argue that inflation can be detrimental to economic growth through four different channels. It would be interesting to find whether this negative effect may work also indirectly via the inflation uncertainty channel. For example, since the Japanese economy during the 1990s was plagued by a deflationary episode associated with low or zero rates of inflation and low output growth rates it would be interesting to find out whether the low output growth rates can be associated with the rate of inflation and the corresponding inflation uncertainty as predicted by Friedman (1977). However, as emphasized by Gylfason and Herbertsson (2001), one can not preclude the possibility that low inflation may be harmless to growth, perhaps even beneficial. Krugman (1998) has recommend more rapid monetary expansion and inflation in Japan in order to reduce real interest rates below zero and thereby stimulate investment and growth.

Second, less robust evidence is found regarding the direction of the impact of a change in nominal uncertainty on inflation. No effect was present for the USA whereas we obtained mixed evidence for the UK. At twelve lags we find evidence against the Cukierman-Meltzer hypothesis. This evidence partially favors the 'stabilization hypothesis' put forward by Holland (1995). He claims that for countries where inflation leads to nominal uncertainty and real costs, we would expect the policy maker to stabilize inflation, hence a negative effect of inflation uncertainty on inflation. This result squares with the findings of recent studies by Fountas and Karanasos (2005) and Fountas et al. (2004). Both studies found that uncertainty about inflation causes negative real effects in the UK. In Japan we found strong evidence in favor of the Cukierman-Meltzer hypothesis. According to Devereux (1989) inflation uncertainty can have a positive impact on inflation via the real uncertainty channel. If the variability of real shocks is the predominant cause of nominal uncertainty, then inflation uncertainty and inflation are positively correlated. As real shocks become more variable the optimal degree of indexation declines. The inflation rate rises only after the degree of indexation falls. Assuming that changes in the degree of indexation take time to occur, greater inflation uncertainty precedes higher inflation.

Chapter 5

Dual Long Memory in Inflation Dynamics across Countries of the Euro Area

5.1 Introduction

An extensive body of theoretical literature examines the relationship between the rate of inflation and the nominal uncertainty. It is important to discover whether an increase in inflation precedes an increase in uncertainty, if we are to add to our understanding about the welfare costs of inflation. Different theories emphasize different channels, some pointing to a positive relationship and some to a negative one. Friedman (1977) argues that higher inflation may induce erratic policy responses to counter it, with consequent unanticipated inflation movements. In contrast, Pourgerami and Maskus (1990) point out that a negative effect may exist. The opposite direction of causality than that

---

This chapter was published as: Conrad, C., and M. Karanasos (2005b). "Dual long memory in inflation dynamics across countries of the Euro area and the link between inflation uncertainty and macroeconomic performance." *Studies in Nonlinear Dynamics & Econometrics* 4, Article 5.

examined by Friedman has also been addressed in the theoretical literature. In particular, Cukierman and Meltzer (1986) contend that inflation uncertainty produces greater average inflation due to opportunistic central bank behavior, whereas according to Holland (1995) higher nominal uncertainty leads to lower average rates of inflation. Much empirical work has been done aimed at signing the effects of inflation on its uncertainty and vice versa. Contradictory empirical results are reported by various researchers. Given the theoretical ambiguity, it is not surprising that the statistical evidence is also ambiguous. Moreover, economic theory and empirical work reach a striking variety of conclusions about the responsiveness of output growth to changes in nominal uncertainty. The importance of uncertainty as a distinct channel in explaining the real effects of inflation has recently been given considerable empirical support (Grier et al., 2004, Elder, 2004, and Fountas and Karanasos, 2005). This channel was first highlighted by Friedman (1977). He argues that uncertainty about inflation causes an adverse growth effect. Dotsey and Sarte (2000) using a cash-in-advance framework obtain the opposite result: more nominal uncertainty can increase real growth.

This study has three primary objectives. First, it analyzes the inflation dynamics of several countries belonging to the European Monetary Union and of the UK. One group of countries is formed by Germany, France, Italy, the Netherlands and Spain. These five major countries represent 88 percent of the GDP of the Euro area. Given that the explicit mission of the European Central Bank (ECB) is the preservation of price stability, the analysis of the nature of inflation in the Euro area is of distinct interest. We estimate the two main parameters driving the degree of persistence in inflation and nominal uncertainty using an ARFIMA-FIGARCH process.[1] This model, developed in Baillie et al. (2002), provides a general and flexible framework with which to study a complicated process like inflation. Put differently, it is sufficiently flexible to handle the dual long memory behavior encountered in inflation.

Second, it investigates the possible existence of heterogeneity in inflation dynamics

[1]We refer to a model that is fractional integrated in both the ARMA and GARCH specifications as the ARFIMA-FIGARCH model.

across Euro area countries. Inflation differentials have important implications for the design of the optimal monetary policy. For example, as Benigno and Lòpez-Salido (2002) point out, an inflation targeting policy that assigns higher weight to countries with higher degrees of persistence benefits those countries since once the policy of the central bank is credible, it produces lower inflation rates for them simply because it cares more about those inflation rates.

The third objective of this study is to shed more light on the issue of temporal ordering of inflation and nominal uncertainty. To do this we proceed in two steps. First, we use the estimated conditional variance from the ARFIMA-FIGARCH model as our statistical measure of inflation uncertainty. Having constructed a time series of nominal uncertainty in the second part we employ Granger methods to test for evidence on the bidirectional causality relationship between inflation and uncertainty about inflation. The two-step approach has been employed among others by Grier and Perry (1998). In addition, we test for the causal effect of nominal uncertainty on output growth. The empirical evidence on this link remains scant or nonexistent, as pertains, in particular, to international data in European economies.

Our first finding is that all ten European inflation rates have the rather curious property of persistence in both their first and their second conditional moments. This empirical evidence is consistent with the evidence provided by Baillie et al. (2002) for eight industrial countries. The second result that emerges from this study is the existence of heterogeneity in inflation dynamics across Euro area countries. These countries fall into three groups in terms of the difference in the dynamics of the second moment of their inflation rates. The first group of countries includes France and Sweden and is characterized by a mild long memory GARCH behavior of the inflation rate. The second includes Belgium, Finland, Italy, the Netherlands and the UK, which are characterized by the presence of quite strong long memory in the inflation uncertainty. The third group of countries includes Portugal and Spain and is characterized by a near integrated behavior in the second conditional moment of the inflation rate. This finding is of some significance since inflation differentials are not irrelevant for monetary policy.

Third, we provide overwhelming evidence that increased inflation raises its uncertainty, confirming the theoretical predictions made by Friedman. However, we find that nominal uncertainty has a mixed impact on output growth. This result brings out an important asymmetry in the transmission mechanism of monetary policy in Europe in addition to the difference in the economic sizes of the countries. In particular, since the effects of uncertainty on output growth differ across the Euro zone, a common monetary policy that results in similar inflation rates across countries will have asymmetric real effects, provided these effects work via a change in nominal uncertainty. We also find that increased nominal uncertainty significantly affects average inflation in eight countries but not all in the same manner. These differential responses to nominal uncertainty are correlated with measures of central bank independence.

The remainder of the chapter is organized as follows. Section 5.2 summarizes several empirical studies that investigate the short-term inflation dynamics. Section 5.3 discusses the economic theory and the empirical testing concerning the link between inflation uncertainty and macroeconomic performance. In Section 5.4, we describe the time series model for inflation and nominal uncertainty and discuss its merits. The empirical results are reported in Section 5.5, and Section 5.6 draws some policy implications and proposes possible extensions of the time series model for inflation. Section 5.7 contains summary remarks and conclusions.

5.2 Inflation Dynamics

This section summarizes several empirical studies that investigate the short-term inflation dynamics. The nature of the short-run inflation dynamics is a central issue in macroeconomics and one of the most fiercely debated. There is an extensive theoretical literature that attempts to develop structural models of inflation that provide a good approximation to its dynamics (see, for example, Karanassou and Snower, 2003), and an equally extensive empirical literature that attempts to document the properties of inflationary shocks. Many contradictory results can be found in the empirical literature on the persistence of inflation rates. Several studies (see, for instance, Grier and Perry, 1998) argue that inflation is I(0), whereas a large number of researchers, such as Banerjee and Russell (2001), find evidence for a unit root in inflation. Similarly, Ball and Cecchetti (1990) decompose inflation into a permanent component and a transitory component. As noted by Baillie et al. (1996b) and Caporale and Gil-Alana (2003), the stationarity of real rates of interest and the Fisher relation is consistent with neoclassical models of dynamic growth, superneutrality, and capital asset pricing models. But if both inflation and nominal interest rates have a unit root then they must be cointegrated in order for the ex-post real rates to be stationary. Moreover, a nonstationary inflation process also complicates the derivation of optimal monetary policy rules (see McCallum, 1988).

Some researchers argue that inflation has become more persistent over time. In particular, Brunner and Hess (1993) model US inflation as an I(0) process before 1960 and as an I(1) process after this time. Along these lines, Evans and Watchel (1993) develop a time series model of inflation that switches from purely transitory shocks in the 1960s to purely permanent shocks in the 1970s, and back to transitory shocks in the late 1980s. They use this model to derive measures of nominal uncertainty that account for the prospects of changing inflation regimes. Generally speaking, as is often the case with post war data, one cannot say with confidence whether the two series, that is inflation and its uncertainty, are stationary or nonstationary or cointegrated if nonstationary. Accordingly, Holland (1995) performs three different tests for Granger-causality between the two variables, each corresponding to one of the three different assumptions.

In sharp contrast, Backus and Zin (1993) find that a fractional root shows up very clearly in monthly US inflation. They conjecture that the long memory in inflation is the result of aggregation across agents with heterogeneous beliefs. They also demonstrate that the fractional difference process is a good descriptor of short-term interest rates and suggest that the fractional unit root in the short rates is inherited from money growth and inflation. Hassler and Wolters (1995) find that the inflation rates of five industrial countries are well explained by different orders of integration, which vary around the stationarity border of 0.5. Ooms and Hassler (1997) using data from Hassler and Wolters (1995) and a modified periodogram regression, confirm their findings. Subsequently, Baum et al. (1999) presented statistical evidence in favor of I(d) ($0 < d < 1$) behavior for both CPI-and WPI- based inflation rates for many industrial as well as developing countries.

The preceding works provide quite consistent evidence across time periods and countries that inflation exhibits long memory with an order of integration which differs significantly from zero and one. Overall these findings suggest that the traditional ARMA and ARIMA specifications are incapable of imparting the persistence to inflation that we find in the data. Put differently, by viewing inflation as an I(0) or I(1) process instead of an I(d) process, we bias downward or upward our estimate of inflation persistence. However, the previously mentioned articles have not explored the time-dependent heteroscedasticity in the second conditional moment of the inflation process. Along these lines, Baillie et al. (1996b) utilize the ARFIMA-GARCH model to describe the inflation dynamics for ten countries. They provide strong evidence of long memory with mean reverting behavior for all countries except Japan. Hwang (2001) also estimates various ARFIMA-GARCH-type models for monthly US inflation. He finds strong evidence that inflation dynamics are well described by a fractional process with an order of integration of about 0.33.

In many applications the sum of the estimated GARCH(1,1) parameters is often close to one, which implies integrated GARCH (IGARCH) behavior. For example, Baillie et al. (1996b) emphasize that for all ten countries the inflation series possesses substantial persistence in its conditional variance. In particular, the sum of the GARCH parameters was at least 0.965. Most importantly, Baillie et al. (1996a), using Monte-Carlo simulations, show that data generated from a process exhibiting long memory FIGARCH volatility may be easily mistaken for IGARCH behavior. Therefore recently Baillie et al. (2002) have focused their attention on the topic of long memory and persistence in terms of the second moment of the inflation process. They employ the

FIGARCH specification of Baillie et al. (1996b) to model the apparent long memory in the conditional variance of the inflation series. They find that the inflation rates for many industrial countries display significant fractional integration in both their first and second moments. Similarly, Conrad and Karanasos (2005a) [see Chapter 4] find that the ARFIMA-FIGARCH model was the preferred specification for the monthly CPI-based inflation rates for the UK and the US.

5.3 The Link between Inflation-Uncertainty and Macroeconomic Performance

5.3.1 Theory

In this section, we discuss the economic theory concerning the link between nominal uncertainty and macroeconomic performance. Since Friedman (1977) stressed the harmful effects of nominal uncertainty on employment and production much research has been carried out investigating the relationship between inflation and uncertainty about inflation. The effect of inflation on its unpredictability is theoretically ambiguous. Several researchers contend that since a reduction in inflation causes an increase in the rate of unemployment, a high rate of inflation produces greater uncertainty about the future direction of government policy and the future rates of inflation. Ball's (1992) model formalizes this idea in the context of a repeated game between the monetary authority and the public.

Holland (1993) points out that in the Evans and Wachtel (1993) framework, if regime changes cause unpredictable changes in the persistence of inflation, then lagged inflation squared is positively related to nominal uncertainty. If, on the other hand, regime changes do not affect the persistence of inflation, then no relationship between the rate of inflation and its uncertainty is implied. In contrast, Pourgerami and Maskus (1990) suggest that higher inflation may induce the relevant economic agents to invest more in generating accurate predictions and hence may lead to lower nominal uncertainty. Ungar and Zilberfarb (1993) propose a mechanism that may weaken, offset, or even reverse the direction of the traditional view concerning the inflation-uncertainty relationship.

The models of Ball, Evans and Wachtel, and Holland imply that higher nominal uncertainty is part of the welfare costs of inflation because inflation causes its uncertainty. On the other hand, Cukierman and Meltzer (1986) and Cukierman (1992) using the Barro-Gordon model of Fed behavior show that greater uncertainty about money growth and inflation causes a higher mean rate of inflation by increasing the incentive for the policy-maker to create inflation surprises. In addition, Devereux (1989) emphasizes the fact that higher variability of real shocks lowers the optimal degree of indexation and increases the incentives of the policy maker to create surprise inflation. Therefore, if changes in the degree of indexation take time to occur then higher nominal uncertainty precedes greater inflation. In sharp contrast, Holland (1995) argues that due to the 'stabilization motive' higher nominal uncertainty has a negative effect on inflation.[2]

The impact of nominal uncertainty on output growth has received considerable attention in the theoretical macroeconomic literature. However, there is no consensus among macroeconomists on the direction of this effect. Theoretically speaking, the effect of uncertainty on growth is ambiguous. The second part of Friedman's hypothesis postulates that greater inflation variability, by reducing economic efficiency, has a negative impact on real growth. In particular, increased volatility in inflation rates reduces the ability of markets to convey information to market participants about relative price movements and makes long-term contracts more costly. Dotsey and Sarte (2000) analyze the effects of nominal uncertainty on economic growth in a model where money is introduced via a cash-in-advance constraint. In this setting, they find that variability increases average growth through a precautionary savings motive. Within the confines of their neoclassical growth model higher rates of inflation have negative consequences for growth, while increased inflation variability has a small positive effect on growth. In essence the offsetting growth effects of mean inflation and its uncertainty, along with the

[2] Under this scenario, if higher inflation raises its uncertainty, the policy maker responds by disinflating the economy in order to reduce nominal uncertainty and the associated costs.

fact that these are highly correlated, provide a partial rationale for the weak and somewhat fragile relationship between growth and inflation. Finally, an alternative channel through which uncertainty about inflation might affect output growth is via the real uncertainty.[3] For example, a rising nominal uncertainty would be expected to have a positive impact on output growth via a combination of the Logue-Sweeney and Black effects.

5.3.2 Empirical Evidence

In this section, we discuss previous empirical testing of the link between nominal uncertainty and macroeconomic performance. The relationship between inflation and its uncertainty has been analyzed extensively in the empirical literature. Davis and Kanago (2000) survey this literature. Recent time series studies of nominal uncertainty have focused on the GARCH conditional variance of inflation as a statistical measure of its uncertainty. Some studies use GARCH models that include a function of the lagged inflation rate in the conditional variance equation. In particular, Brunner and Hess (1993) allow for asymmetric effects of inflation shocks on nominal uncertainty and find a weak link between US inflation and its uncertainty. Caporale and McKierman (1997) find a positive relationship between the level and variability of US inflation. Three studies use GARCH type models with a joint feedback between the conditional mean and variance of inflation. Baillie et al. (1996b), for three high inflation countries and the UK, and Karanasos et al. (2004a) for the US, find strong evidence in favor of a positive bidirectional relationship in accordance with the predictions of economic theory. In contrast, Hwang (2001) finds that US inflation affects its uncertainty weakly and negatively. Finally, the recent empirical literature tends to confirm the negative association between nominal uncertainty and real growth in the US. Studies by Grier and Perry (2000),

[3]The positive association between inflation and output variability is known in the literature as the Logue-Sweeney hypothesis (see Karanasos and Kim, 2005, for details). The positive impact of output uncertainty on growth is known in the literature as the Black hypothesis (see Fountas and Karanasos, 2005, for details).

Grier et al. (2004) and Elder (2004) employ bivariate GARCH-in-mean models and find a negative effect. However, all these studies use only US data.

Some studies examine the link between nominal uncertainty and the level of inflation using the two-step approach where an estimate of the conditional variance is first obtained from a GARCH-type model and Granger methods are then employed to test for bidirectional effects. In particular, Grier and Perry (1998) find that in all G7 countries inflation significantly raises its uncertainty. They also find evidence in favor of the Cukierman-Meltzer hypothesis for some countries and in favor of the Holland hypothesis for other countries. Fountas et al. (2004), using quarterly data and employing the EGARCH model, find that in five European countries inflation significantly raises its uncertainty. Their results regarding the direction of the impact of a change in nominal uncertainty on inflation were generally consistent with the existing rankings of central bank independence. Conrad and Karanasos (2005a) [see Chapter 4] for three industrial countries find strong evidence in support of both the Friedman and the Cukierman and Meltzer hypotheses.

Holland (1993) tabulates a number of empirical studies concerning the relationship between nominal uncertainty and real activity (employment or output). Studies based on surveys tend to find a negative relationship between nominal uncertainty and real activity, whereas studies based on ARCH volatility find insignificant or positive relationships. In particular, Coulson and Robins (1985) find that nominal uncertainty has a positive impact on real growth, while Jansen (1989) uses a bivariate ARCH in mean model and reports an insignificant relationship. Dotsey and Sarte (2000) also report empirical work documenting that the effect of uncertainty on growth appears to be non-negative. Grier and Tullock (1989) have been unable to verify the more conventional view that greater volatility in the inflation rate lowers growth, while McTaggart (1992) uses annual data for Australia and finds that inflation variability has a positive influence on the log of output. Levine and Renelt (1992) use cross-country regressions to search for empirical linkages between growth rates and a variety of economic policy indicators. They find that all the results are fragile to small changes in the conditioning information set. The empirical findings in Barro (1996), for a panel of around 100 countries, support the notion that the variability of inflation has no significant relation with growth.

The empirical evidence on the relationship between nominal uncertainty and output growth remains scant or nonexistent, as pertains, in particular, to international data in industrialized economies. An exception is Fountas et al., (2004) and Fountas and Karanasos (2005). They employ the two-step approach in a univariate GARCH framework using data for six European and the G7 countries respectively and find significant evidence in favor of the Friedman hypothesis for some countries and in favor of the Dotsey-Sarte hypothesis for other countries. That is, the evidence regarding the direction of the impact of a change in nominal uncertainty on real growth found in these two studies is not robust across countries.

There are a limited number of studies using international data that are based on GARCH measures of inflation uncertainty. These are Baillie et al. (1996b), Grier and Perry (1998), Fountas et al. (2004), Fountas and Karanasos (2005), and Conrad and Karanasos (2005a) [see Chapter 4]. Only Fountas et al. (2004) investigate the relationship between inflation and its uncertainty for six European countries and only Conrad and Karanasos (2005a) [see Chapter 4] focus on a statistical measure of nominal uncertainty that captures the dual long memory aspect of inflation, namely the ARFIMA-FIGARCH (conditional) variance of inflation. This study aims to fill the gaps arising from the lack of interest in the European case, where the results would have interesting implications for the successful implementation of a common European monetary policy, and from the methodological shortcomings of the previous studies.

5.4 Methodology

5.4.1 The ARFIMA-FIGARCH Process

It appears from the study of Baillie et al. (2002) that the apparent long memory in the inflation rate is also present in nominal uncertainty. Hence, there seems to be a need to have a joint model which incorporates long memory in both the conditional mean and the conditional variance of inflation. In other words, the time series features of inflation seem to require the use of fractional integrated models from two different classes, namely the ARMA and the GARCH.

Along these lines, in this section we describe the time series model for inflation and nominal uncertainty and discuss its merits. First, we denote the inflation rate by π_t and we define its mean equation as

$$(1-L)^{d_m}(1-\varphi_6 L^6-\varphi_9 L^9-\varphi_{12} L^{12}-\varphi_{24} L^{24})(\pi_t-\mu)=\varepsilon_t, \tag{5.1}$$

where $(1-L)^{d_m}$ is the fractional differencing operator with $d_m \leq 1$. That is the inflation rate follows an ARFIMA(24,d_m,0) specification. Second, let us suppose that ε_t is conditionally normal with mean zero and variance h_t. That is $\varepsilon_t|\mathcal{F}_{t-1} \sim \mathcal{N}(0,h_t)$, where $\mathcal{F}_{t-1}$ is the information set up to time $t-1$. Finally, we assume that the structure of the conditional variance is given by a FIGARCH$(1,d,1)$, i.e.

$$(1-\beta L)h_t=\omega+\left[(1-\beta L)-(1-\phi L)(1-L)^{d_v}\right]\varepsilon_t^2, \tag{5.2}$$

where $0 \leq d_v \leq 1$, $\omega > 0$, and ϕ, $\beta < 1$. For necessary and sufficient conditions on the parameters (β, d_v, ϕ) guaranteeing the nonnegativity of the conditional variance in the FIGARCH$(1,d,1)$ model see Conrad and Haag (2006) [see Chapter 2, Corollary 2.1]. The FIGARCH specification reduces to a GARCH model for $d_v=0$ and to an IGARCH model for $d_v=1$. If $h_t=\omega$, a constant, the process reduces to the ARFIMA $(24,d_m,0)$ model. Then the inflation rate will be covariance stationary and invertible for $-0.5<d_m<0.5$ and will be mean reverting for $d_m<1$. Although the ARFIMA-FIGARCH process is strictly stationary and ergodic for $0 \leq d_v \leq 1$, it will have an infinite unconditional variance for all d_m given a $d_v \neq 0$. Clearly, the unit root corresponds to the null hypothesis $H_0: d_m=1$.

5.4.2 Two-Step Strategy

To test for the relationship between inflation and nominal uncertainty, one can use either the two-step or the simultaneous approach. Under the latter one, an ARFIMA-FIGARCH-in-mean model is estimated with the conditional variance equation incorporating lags of the inflation series, thus allowing simultaneous estimation and testing of the bidirectional causality between the two variables. The two-step approach is performed by first obtaining an estimate of the conditional variance from the ARFIMA-FIGARCH model and then Granger methods are employed to test for bidirectional effects. We prefer the two-step strategy for the following reasons (see Grier and Perry, 1998). First, it allows us to capture the lagged effects between the variables of interest. In particular, the in-mean model suffers from the disadvantage that it does not allow the testing of a lagged effect of nominal uncertainty on inflation, which would be expected in a study that employs monthly data. As Grier and Perry (1998) mention, the impact of a change in inflation uncertainty on average inflation, via a change in the stabilization policy of the monetary authority, takes time to materialize and cannot be fairly tested in a model that restricts the effect to being contemporaneous. Second, the simultaneous approach is subject to the criticism of the potential negativity of the conditional variance. This is because there is no way of guaranteeing the nonnegativity of the conditional variance by imposing constraints on the parameters of the conditional variance specification since the sign of the inflation series is time-varying. Third, the two-step approach minimizes the number of estimated parameters.

It is also interesting to note the similarities between the Lagrange Multiplier (LM) test for ARCH effects and the Granger-causality methodology. In the LM statistic the first step is to estimate the conventional regression model for inflation by OLS (i.e., assuming independent errors) and obtain the fitted residuals ($\widehat{\varepsilon}_t$). The second step is to regress $\widehat{\varepsilon}_t^2$ on a constant and lags of $\widehat{\varepsilon}_t^2$. If ARCH effects are present (i.e., if the squared errors are linearly related), the estimated parameters should be statistically significant. In our two-step strategy an estimate of the conditional variance is first obtained from an ARFIMA-FIGARCH model (i.e., assuming that inflation and its uncertainty are uncorrelated) and then causality tests are run to test for bidirectional effects between the two variables.

5.5 Empirical Analysis

5.5.1 Data

In this section we look at some of the time series characteristics of inflation. Monthly data, obtained from the *OECD Statistical Compendium*, are used to provide a reasonable number of observations. The inflation and output growth series are calculated as the monthly difference in the natural log of the Consumer Price Index and Industrial Production Index respectively. The data range from 1962:01 to 2004:01 and cover ten European countries, namely, Belgium, Finland, France, Germany, the Netherlands, Italy, Portugal, Spain, Sweden and the UK. Allowing for differencing this implies 504 usable observations.[4]

The summary statics (not reported) for the ten inflation rates show that the German (Portuguese) inflation rate has the lowest (highest) mean and standard deviation. Furthermore, the summary statistics indicate that the distributions of all ten inflation series are skewed to the right. The large values of the Jarque-Bera (JB) statistic imply a deviation from normality, and the significant Q-statistics of the squared deviations of the inflation rate from its sample mean indicate the existence of ARCH effects. This evidence is also supported by the Lagrange Multiplier (LM) test statistics, which are highly significant.

Next, we employ the PP and KPSS unit root tests, suggested by Phillips and Perron (1988) and Kwiatkowski et al. (1992) respectively. The results are presented in Table 5.1 and can be summarized as follows. For all the inflation series shown, based on the

[4] The only exceptions are Belgium and Spain for which output data was available only from January 1965 onwards. For all countries the industrial production series are seasonally adjusted.

Table 5.1: Tests for order of integration of different countries' inflation series.

	H_0: I(1)		H_0: I(0)	
	$Z(t_\mu)$	$Z(t_\tau)$	$\widehat{\eta}_\mu$	$\widehat{\eta}_\tau$
Belgium	-14.24***	-14.93***	2.45***	0.83***
Finland	-17.00***	-18.31***	3.20***	0.81***
France	-10.29***	-11.48***	3.31***	1.46***
Germany	-16.24***	-16.55***	1.26***	0.34***
Italy	-8.30***	-8.71***	2.19***	1.41***
Netherlands	-21.23***	-22.07***	2.45***	0.34***
Portugal	-18.21***	-18.32***	1.63***	1.27***
Spain	-15.30***	-16.15***	2.80***	1.09***
Sweden	-17.66***	-18.12***	2.20***	1.06***
UK	-13.95***	-14.37***	1.88***	0.90***

Notes: $Z(t_\mu)$ and $Z(t_\tau)$ are the PP adjusted t-statistics of the lagged dependent variable in a regression with intercept only, and intercept and time trend included, respectively. The 0.01 critical values for $Z(t_\mu)$ and $Z(t_\tau)$ are -3.43 and -3.96. $\widehat{\eta}_\mu$ and $\widehat{\eta}_\tau$ are the KPSS test statistics based on residuals from regressions with an intercept and intercept and time trend, respectively. The 0.01 critical values for $\widehat{\eta}_\mu$ and $\widehat{\eta}_\tau$ are 0.739 and 0.216. *** denotes significance at the 0.01 level.

PP test we are able to reject the unit root hypothesis, whereas based on the KPSS test the null hypothesis of stationarity is rejected.[5] In other words, the application of these tests yields contradictory results. With all inflation series we find evidence against

[5]We used a Bartlett kernel for the PP test and chose five as truncation lag, the number of lags included in the KPSS test was set to four. Alternatively, the Augmented Dickey-Fuller (ADF) statistic can be applied to test the unit root hypothesis. The evidence obtained from the PP test statistic is reinforced by the results (not reported) provided by the ADF tests.

the unit root as well as against the stationarity hypothesis. Thus fractional integration allowing for long memory is a plausible model.[6] Finally, the results of the unit root tests applied to the output growth series (not reported) imply that we can treat these series as stationary processes.

5.5.2 Model Estimates

The analysis in Baillie et al. (2002) suggests that the ARFIMA-FIGARCH specification describes the inflation series of several industrial countries well. Within this framework we will analyze the dynamic adjustments of both the conditional mean and the conditional variance of inflation for several European countries, as well as the implications of these dynamics for the direction of causality between nominal uncertainty and macroeconomic performance. Estimates of the ARFIMA-FIGARCH model are shown in Table 5.2. These were obtained by quasi-maximum likelihood estimation (QMLE) as implemented by Laurent and Peters (2002) in Ox. The truncation lag for the fractional differencing operator was chosen to be 500. For a detailed description of the estimation procedure see Baillie et al. (1996a). The consistency and asymptotic normality of the QMLE has been established only for specific special cases of the ARFIMA and/or FIGARCH model. However, a detailed Monte-Carlo study, where ARFIMA-FIGARCH type models were simulated, was performed by Baillie et al. (2002) and it was found that the quality of the application of the QMLE is generally very satisfactory. To check for the robustness of our estimates we used a range of starting values and hence ensured

[6] Of course, these unit root tests are merely suggestive. For example, Lee and Amsler (1997) show that the KPSS statistic cannot distinguish consistently between nonstationary long memory and unit root. We also examine the characteristics of inflation graphically by presenting the autocorrelation function of inflation and changes in inflation. Among other things, the figures (not reported) make clear the long memory property of inflation, that is the inflation series itself show significant positive and slowly decaying autocorrelations while the differenced series appear to follow an MA(1) process. Finally, we plot the autocorrelations of the squared and absolute values of the residuals from an estimated ARFIMA$(24, d_m, 0)$ model. Interestingly, these autocorrelations are extremely persistent, which is suggestive of long memory behavior in the conditional variance.

that the estimation procedure converged to a global maximum.

Several findings emerge from Table 5.2. The estimated long memory conditional mean parameter is in the range $0.141 \leq \widehat{d}_m \leq 0.353$. The value of the coefficient for Portugal (0.141) is markedly lower than the corresponding value for Italy (0.353). However, although the estimated value of d_m for Portugal is relatively small it is significantly different from zero. Furthermore, once long memory in the conditional mean has been accounted for, an AR(24) specification appears to capture the serial correlation in all ten inflation series. That is, all the $\widehat{\varphi}_{12}$ and $\widehat{\varphi}_{24}$ parameters are much larger than their standard errors.

The estimation of a FIGARCH model for Portugal and Spain realized an estimated value of d_v close to 0.9 (0.874 and 0.866 respectively), whereas in sharp contrast, for France and Sweden it realized a value close to 0.1 (0.130 and 0.133 respectively). In other words, the estimates of d_v that govern the dynamics of the conditional heteroscedasticity indicate that the conditional variances of the Portuguese and Spanish inflation are characterized by a near integrated GARCH behavior, whereas the conditional variances of the French and Swedish inflation are characterized by a very mild long memory GARCH behavior. For the other six countries, the values of d_v vary from 0.195 (Netherlands), 0.209 (Finland), and 0.269 (Germany) to 0.330 (Belgium), 0.457 (UK), and 0.529 (Italy). For Finland, France, Germany, the Netherlands, Sweden and the UK the Akaike and Schwarz information criteria (AIC and SIC respectively) come out in favor of the FIGARCH$(0, d_v, 0)$ model, while for Italy, Portugal and Spain (the three countries with the highest $\widehat{d_v}$) the FIGARCH$(1, d_v, 0)$ is the preferred specification. In addition, note that the estimated GARCH parameters for these three countries and for Belgium satisfy the set of conditions which are necessary and sufficient to guarantee the nonnegativity of the conditional variance derived in Conrad and Haag (2006) [see Chapter 2, Corollary 2.1].

The ten European countries fall into three groups in terms of the differences in the sum of the two fractional differencing parameters $(d_m + d_v)$. The first group of countries includes Finland, France, Germany, the Netherlands and Sweden: $0.310 < \widehat{d}_m + \widehat{d}_v <$

0.480. In all these countries, except France, the estimated value of d_m is very close to the estimate of d_v. The second includes Belgium, Italy and the UK: $0.540 < \widehat{d}_m + \widehat{d}_v < 0.890$. The third group of countries includes Portugal and Spain: $\widehat{d}_m + \widehat{d}_v \simeq 1$. Portugal and Spain have very similar estimated mean (0.141, 0.181) and variance (0.874, 0.866) fractional differencing parameters.[7] Interestingly, these are the two countries with the lowest (highest) long memory mean (variance) parameters. Whether the sum of the two estimated fractional differencing parameters is below or above 0.5 will become of importance when analyzing causal relationships between inflation and its uncertainty in the next section. In seven out of the ten countries the estimates of d_m are smaller than the estimates of d_v.

Generally speaking, the parameter estimates support the idea that dual long memory effects are present in the inflation process for all ten European countries, suggesting that the dual long memory is an important characteristic of the inflation data. Finally, with all countries, the hypothesis of uncorrelated standardized and squared standardized residuals is well supported, indicating that there is no statistically significant evidence of misspecification.

[7] The Portugese and Spanish inflation series are also the two series with the highest sample means and standard deviations.

Table 5.2: ARFIMA-FIGARCH models 1962:01–2004:01.

	Belgium	Finland	France	Germany	Italy
$\widehat{\mu}$	0.282	0.125	0.314	0.254	0.186
	(1.984)	(1.258)	(1.440)	(2.510)	(1.186)
$\widehat{d}_m$	0.214	0.189	0.305	0.203	0.353
	(3.481)	(6.169)	(6.712)	(5.049)	(4.124)
$\widehat{\varphi}_{12}$	0.249	0.283	0.212	0.376	0.300
	(4.988)	(4.071)	(3.415)	(7.914)	(6.769)
$\widehat{\varphi}_{24}$	0.092	0.156	0.185	0.253	0.191
	(1.882)	(2.700)	(4.064)	(4.925)	(4.646)
$\widehat{\omega}$	0.013	0.038	0.021	0.014	0.001
	(2.013)	(1.893)	(1.873)	(2.454)	(0.517)
$\widehat{d}_v$	0.330	0.209	0.130	0.269	0.529
	(2.311)	(4.810)	(1.678)	(2.077)	(3.371)
Q_{12}	15.65	8.74	13.94	13.78	21.26
	[0.21]	[0.72]	[0.30]	[0.31]	[0.05]
Q_{12}^2	10.54	5.28	12.03	5.07	14.27
	[0.57]	[0.95]	[0.44]	[0.96]	[0.28]

Notes: For each of the ten inflation series, Table 5.2 reports QML parameter estimates for the ARFIMA-FIGARCH model. The numbers in parentheses are t-statistics. Q_{12} and Q_{12}^2 are the 12-th order Ljung-Box tests for serial correlation in the standardized and squared standardized residuals respectively. The numbers in brackets are p values. The φ_6 and φ_9 coefficients are significant only in Belgium and France. The ϕ coefficient is significant only in Belgium: $\widehat{\phi} = -0.280$ (-1.933).

ARFIMA-FIGARCH models (continued).

	Netherlands	Portugal	Spain	Sweden	UK
$\widehat{\mu}$	0.293	0.163	0.361	0.413	0.339
	(1.932)	(1.050)	(1.836)	(2.885)	(0.831)
$\widehat{d}_m$	0.203	0.141	0.181	0.185	0.340
	(4.093)	(3.413)	(3.185)	(4.202)	(4.414)
$\widehat{\varphi}_{12}$	0.507	0.389	0.334	0.261	0.401
	(6.584)	(7.578)	(5.027)	(4.547)	(7.471)
$\widehat{\varphi}_{24}$	0.173	0.180	0.263	0.216	0.330
	(3.545)	(3.681)	(4.485)	(3.713)	(6.310)
$\widehat{\omega}$	0.021	0.002	0.003	0.106	0.019
	(1.788)	(0.420)	(1.702)	(1.939)	(1.666)
$\widehat{d}_v$	0.195	0.874	0.866	0.133	0.457
	(3.685)	(10.82)	(5.854)	(1.760)	(4.056)
Q_{12}	16.24	17.93	20.01	15.92	14.36
	[0.18]	[0.12]	[0.07]	[0.19]	[0.28]
Q_{12}^2	14.55	6.04	9.66	18.74	13.29
	[0.27]	[0.91]	[0.65]	[0.09]	[0.35]

Notes: As in Table 5.2. For Italy, Portugal and Spain we estimate a β of 0.266 (1.010), 0.772 (9.878) and 0.724 (6.129) respectively.

To test for the persistence of the conditional heteroscedasticity models, we examine the likelihood ratio (LR) tests and the Wald (W) statistics for the linear constraints $d_m = d_v = 0$ (ARMA-GARCH model). As seen in Table 5.3 the LR tests and W statistics clearly reject the ARMA-GARCH null hypotheses against the ARFIMA-FIGARCH model for all ten inflation series. Thus, purely from the perspective of searching for a model that best describes the degree of persistence in both the mean and the variance of the inflation series, the ARFIMA-FIGARCH model appears to be the most satisfactory

Table 5.3: Likelihood Ratio and Wald test statistics.

	Be	Fi	Fr	Ge	It
LR	16.95	47.73	39.64	42.88	113.88
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
W	6.72	29.92	29.12	12.93	9.10
	[0.03]	[0.00]	[0.00]	[0.00]	[0.01]
	Ne	Po	Sp	Sw	UK
LR	58.22	18.62	17.69	8.03	33.98
	[0.00]	[0.00]	[0.00]	[0.02]	[0.00]
W	12.31	71.22	19.35	8.89	15.52
	[0.00]	[0.00]	[0.00]	[0.01]	[0.00]

Notes: The rows denoted by LR report the value of the following likelihood ratio test: $LR = 2 \cdot [ML_u - ML_r]$, where ML_u and ML_r denote the maximum log-likelihood values of the unrestricted (ARFIMA-FIGARCH) and restricted (ARMA-GARCH) models respectively. The rows denoted by W report the corresponding Wald statistics. The numbers in brackets are p-values. Be: Belgium, Fi: Finland Fr: France, Ge: Germany, It: Italy, Ne: Netherlands, Po: Portugal, Sp: Spain, Sw: Sweden.

representation.

Following the work of Grier and Perry (1998) among others, the LR test can be used for model selection. Alternatively, the AIC, SIC and Hannan-Quinn or Shibata information criteria (HQIC, SHIC respectively) can be applied to rank the various ARMA-GARCH type models. These model selection criteria check the robustness of the LR and W testing results discussed above.[8] According to the four information criteria, in all ten

[8]The analysis in Caporin (2003) focuses on the identification problem of FIGARCH models. Caporin performs a detailed Monte-Carlo simulation study and shows that the four information criteria can clearly distinguish between long and short memory data generating processes. Finally, Caporin's results show that when LR tests are applied to time series for which the true data generating process is

cases the optimal GARCH type model is the ARFIMA-FIGARCH.[9] Hence, the model selection criteria are in accordance with the LR and W testing results. Furthermore, we should also mention that although the estimated d_v parameter for Portugal and Spain is not significantly different from unity, it appears that the volatility dynamics in these two countries are better modelled by the fractional differencing parameter since both the LR and W statistics (not reported) clearly reject the 'IGARCH' hypothesis against the FIGARCH model. In addition, the information criteria favor the FIGARCH model over the IGARCH model.

Finally, we test for the similarity of the optimal mean fractional differencing parameters estimated for each of the ten inflation series using a pairwise Wald test:

$$\mathrm{W}_m = \frac{(\widehat{d}_{m,1} - \widehat{d}_{m,2})^2}{(se_{m,1})^2 + (se_{m,2})^2},$$

where $\widehat{d}_{m,i}$, $(i = 1, 2)$ is the mean fractional differencing parameter from the ARFIMA-FIGARCH model estimated for the inflation series for country i and $se_{m,i}$ is the standard error associated with the estimated model for country i. The above W statistic tests whether the mean fractional differencing parameters of the two countries are equal $(\widehat{d}_{m,1} = \widehat{d}_{m,2})$, and is distributed as $\chi^2_{(1)}$. In the majority of the cases the results (not reported) of this pairwise testing procedure provide support for the null hypothesis that the estimated fractional parameters are not significantly different from one another.[10]

FIGARCH with a d_v parameter being close to one, the LR test has no power to distinguish between the fractionally integrated and the IGARCH model.

[9]We do not report the AIC, SIC, HQIC or SHIC values for space considerations.

[10]It should be noted that the mean fractional differencing parameters are related to the other parameters in the ARFIMA-FIGARCH model. In particular, the information matrix between the AR parameters and the fractional parameter is not block diagonal. Hence, comparison of estimated d_m parameters, specially between countries with different model specifications, should be taken with a pinch of salt.

5.5.3 Granger-Causality Tests

In this section we report results of Granger-causality tests to provide some statistical evidence on the nature of the relationship between nominal uncertainty and macroeconomic performance. Tsay and Chung (2000) in their analysis of spurious regression with independent, fractionally integrated processes find that in bivariate regressions no matter whether the dependent variable and the regressor are stationary or not, as long as their fractional orders of integration sum up to a value greater than 0.5, the t ratios become divergent. Recall that in five countries (Belgium, Italy, Portugal, Spain and the UK) the estimated sum of the two long memory parameters ($d_m + d_v$) is greater than 0.5.

Consequently, we utilize the methodology developed by Toda and Yamamoto (1995) to test for causality between nominal uncertainty and either inflation or output growth, which leads to a χ^2 distributed test statistic despite any possible nonstationarity or cointegration between the series.[11] The test is performed in two steps. In the first step, the optimal lag length (k) of the system is determined by utilizing the AIC and SIC. In the second step a VAR of order $k^* = k + d_{\max}$ is estimated (where $d_{\max}$ is the maximal (integer) order of integration suspected to occur in the system) and a modified Wald (MW) test is applied to the first k VAR coefficient matrices to make Granger-causal inference. This MW test statistic has an asymptotic χ^2 distribution with k degrees of freedom. Since inflation and its uncertainty are fractionally integrated with $d_m, d_v < 1$ we set $d_{\max} = 1$ and estimate VAR models with $k^* = k + 1$ lags.[12] The optimal lag length turned out to be either 4, 8 or 12 for all countries. To ensure that our results are

[11]Note, that this procedure also avoids the problem of unbalanced regression, which could occur in regressions involving the I(0) output series and the near-integrated conditional variance series of Portugal and Spain.

[12]Of course, the Toda and Yamamoto (1995) procedure is inefficient and suffers some loss of power since one intentionally over-fits the model. However, if – as in our case – the VAR system has only two variables and long lag length, the inefficiency caused by adding only one more lag is expected to be relatively small.

not sensitive to the choice of the lag length we report in Table 5.4 for all ten countries the MW tests using 4, 8 and 12 lags, as well as the sign of the sums of lagged coefficients in case of significance.

Panel A reports results on the impact of changes in inflation on its uncertainty. We apply the MW tests and use the Newey-West heteroscedasticity and autocorrelation consistent standard errors. Statistically significant effects are present for all countries. There is strong evidence that inflation affects its uncertainty positively, as predicted by Friedman (1977) and Ball (1992).

We then perform Granger-causality tests in order to examine the causal effect of nominal uncertainty on macroeconomic performance. The tests are performed under the assumption that the conditional variances follow GARCH-type processes.[13] Panel B reports the results of the causality tests where causality runs from nominal uncertainty to the rate of inflation. This panel shows a significant positive effect of uncertainty on inflation for three out of the ten countries. The evidence is strong for France and Spain and weaker for Portugal, where it applies for only one of the chosen lags. The results from these three countries support the Cukierman-Meltzer hypothesis that Grier and Perry (1998) label as the 'opportunistic Fed'. Increases in nominal uncertainty raise the optimal average inflation by increasing the incentive for the policy-maker to create inflation surprises. For Sweden, we find strong evidence for a negative effect of nominal uncertainty on inflation, which along with the growth effect of nominal uncertainty squares with Holland's stabilization hypothesis. In other words, this result suggests

[13]In the presence of conditional heteroskedasticity Vilasuso (2001) investigates the reliability of causality tests based on least squares. He demonstrates that when conditional heteroskedasticity is ignored, least squares causality tests exhibit considerable size distortion if the conditional variances are correlated. In addition, inference based on a heteroskedasticity and autocorrelation consistent covariance matrix constructed under the least squares framework offers only slight improvement. Therefore, he suggests that causality tests be carried out in the context of an empirical specification that models both the conditional means and conditional variances. However, if the conditional variances are unrelated, then there is only slight size distortion associated with least-squares tests, and the inconsistency of the least squares standard errors is unlikely to be problematic.

that the ‘stabilizing Fed’ notion is plausible. Increased inflation raises uncertainty, which creates real welfare losses and then leads to monetary tightening to lower inflation and thus also uncertainty (see Grier and Perry, 1998). For Finland there is evidence for Holland’s argument at lag 8 only. We also obtain mixed evidence for Germany, the Netherlands and the UK. In particular, at eight lags uncertainty has a positive impact on inflation, whereas the value of the MW test statistic and the sign of the sum of lagged coefficients at 12 lags (optimal lag length) implies a negative relationship. We view this as support for Holland’s stabilization hypothesis. Since monetary policy takes time to materialize, it is not surprising that a negative effect is found at 12 lags, but not at 4 or 8 lags. A time horizon of 3 to 4 quarters is what one would usually expect for monetary policy to effect the economy. However, neither of the two theories is supported in Belgium and Italy, where inflation is independent from changes in its uncertainty. Thus, increased uncertainty significantly affects future inflation in most of the countries in the sample, but not all in the same manner.

The Granger-causality test results of uncertainty on real growth are given in Panel C. As we show above high-inflation countries are also likely to experience highly volatile inflation rates. If only uncertainty is included in the estimated regression equations, it is impossible to determine whether it is the inflation rate or its uncertainty that is affecting output growth. Hence, in order to control for possible effects of uncertainty on growth that take place via changes in inflation Panel C reports the MW statistics when the regressions include in addition lagged inflation rates. Nominal uncertainty has a mixed impact on output growth. Friedman’s hypothesis regarding the negative real effects of uncertainty receives support in five out of the ten countries. The evidence is strong in Belgium, Sweden and the UK, mild in Italy, and weaker in Germany where it applies for only one of the chosen lags. In contrast, in the other five countries we find that uncertainty has a positive impact on real growth, supporting the Dotsey-Sarte hypothesis. The evidence is strong in Finland, France and the Netherlands, and mild in Portugal and Spain. The fact that many other factors are likely to be related to output growth-either causally or because both are influenced by a third factor makes it more difficult to gauge the significance and magnitude of the impact of uncertainty on growth. Therefore, one should be wary of putting too much faith in the uncertainty-growth relationship. But at the broadest level, the available evidence supports the Friedman hypothesis in some countries and is in favor of the Dotsey-Sarte hypothesis for other countries.

Moreover, note that for France, Portugal and Spain we find evidence for a positive effect of nominal uncertainty on inflation, which along with the output effect of inflation uncertainty squares with the Cukierman-Meltzer ('opportunistic') hypothesis. In other words, the central banks dislike inflation but value the higher employment that results from surprise inflation. Therefore, increases in nominal uncertainty raise the average inflation rate by increasing the incentive for the policy-makers to create inflation surprises (Grier and Perry, 1998).

The three figures in Appendix 5.8 plot for Germany, the Netherlands and the UK (i) the time profiles of inflation and its uncertainty due to shocks in nominal uncertainty and inflation respectively and (ii) the time profile of output growth due to shocks in nominal uncertainty.[14] The maximum effect of inflation on its uncertainty takes place after three (two) months for the Netherlands (Germany and the UK). The negative impact of nominal uncertainty on output growth reaches its peak after nine and twelve months in Germany and the UK respectively. In contrast, in the Netherlands the maximum (positive) effect takes place after five months. Finally, the sign of the effect of nominal uncertainty on inflation varies considerably over time. In all three countries the negative impact reaches its peak after twelve months. In Germany and the Netherlands the effect also seems much smaller in size than the effect of inflation uncertainty on real growth.

To summarize, the results in this section confirm that inflation affects its uncertainty positively. Uncertainty surrounding future inflation appears to have a mixed impact on both inflation and output growth.

[14]Generalized impulse response functions are calculated as suggested in Pesaran and Shin (1998). We do not report Figures for the other countries for space considerations.

Table 5.4: Granger-causality tests 1962:01–2004:01.

	Belgium	Finland	France	Germany	Italy
	Panel A. H_0: Inflation does not Granger-cause inflation uncertainty.				
4	**11.51** [0.02](+)	**17.37** [0.00](+)	17.04 [0.00](+)	**15.65** [0.00](+)	**21.62** [0.00](+)
8	17.14 [0.03](+)	25.81 [0.00](+)	28.79 [0.00](+)	17.57 [0.02](+)	22.57 [0.00](+)
12	20.61 [0.06](+)	27.89 [0.00](+)	**35.32** [0.00](+)	21.28 [0.05](+)	24.90 [0.01](+)
	Panel B. H_0: Inflation uncertainty does not Granger-cause inflation.				
4	3.12 [0.54]	**4.00** [0.40]	**7.42** [0.11](+)	5.89 [0.21]	**4.93** [0.29]
8	**4.06** [0.85]	17.07 [0.03](–)	26.46 [0.00](+)	22.22 [0.00](+)	10.90 [0.21]
12	6.72 [0.87]	15.96 [0.19]	28.50 [0.00](+)	**25.68** [0.01](–)	12.96 [0.37]
	Panel C. H_0: Inflation uncertainty does not Granger-cause output growth.				
4	**7.22** [0.12](–)	**7.27** [0.12](+)	**15.43** [0.00](+)	**0.86** [0.93]	**8.10** [0.09](–)
8	37.47 [0.00](–)	20.44 [0.01](+)	20.47 [0.00](+)	5.70 [0.68]	10.53 [0.23]
12	25.22 [0.01](–)	36.56 [0.00](+)	70.42 [0.00](+)	59.81 [0.00](–)	22.34 [0.03](–)

Notes: The figures are MW statistics. The numbers in the first column give the lag structure for the MW tests, i.e. the orders of the VAR's are 5, 9 and 13. The bold (underlined) numbers indicate the optimal lag length chosen by SIC(AIC). The numbers in $[\cdot]$ are p values. A +(–) indicates that the sum of the lagged coefficients is positive (negative).

Granger-causality tests (continued).

	Netherlands	Portugal	Spain	Sweden	UK
Panel A. H_0: Inflation does not Granger-cause inflation uncertainty.					
4	8.75 [0.07](+)	**16.09** [0.00](+)	**13.02** [0.01](+)	**43.83** [0.00](+)	**54.45** [0.00](+)
8	19.01 [0.01](+)	27.28 [0.00](+)	18.51 [0.02](+)	57.96 [0.00](+)	17.87 [0.02](+)
12	**13.38** [0.34]	35.02 [0.00](+)	24.71 [0.02](+)	52.58 [0.00](+)	36.51 [0.00](+)
Panel B. H_0: Inflation uncertainty does not Granger-cause inflation.					
4	21.84 [0.00](+)	5.63 [0.23]	6.73 [0.15](+)	**17.11** [0.00](–)	4.82 [0.31]
8	20.65 [0.01](+)	**8.96** [0.34]	17.83 [0.02](+)	21.74 [0.00](–)	54.37 [0.00](+)
12	**22.05** [0.04](–)	17.14 [0.14](+)	**39.64** [0.00](+)	26.08 [0.01](–)	**67.12** [0.00](–)
Panel C. H_0: Inflation uncertainty does not Granger-cause output growth.					
4	**8.86** [0.06](+)	**1.37** [0.84]	**8.06** [0.09](+)	**11.53** [0.02](–)	**38.77** [0.00](–)
8	16.71 [0.03](+)	13.67 [0.09](+)	8.87 [0.35]	28.30 [0.00](–)	48.69 [0.00](–)
12	34.96 [0.00](+)	40.96 [0.00](+)	19.00 [0.09](+)	32.49 [0.00](–)	68.56 [0.00](–)

Notes: As in Table 5.4.

5.5.4 Robustness

Monte-Carlo Study

To check the sensitivity of our results to the orders of integration of inflation (d_m) and its uncertainty (d_v), we are also using the inflation series filtered by $(1-L)^{\widehat{d}_m}$ and the series of the estimated conditional variances filtered by $(1-L)^{\widehat{d}_v}$. We carry out the conventional Granger-causality tests using both sets of data, i.e., our original set of data and the one with the two filtered series. If significant effects are obtained for the original series, but not when applying the Toda and Yamamoto (1995) procedure (or using the appropriately differenced series), this could be viewed as evidence for spurious regression in the simple Granger-causality tests. The results (not reported) are very similar to those obtained using the methodology developed in Toda and Yamamoto (1995).[15] In particular, when the original data are used the primary difference lies in the stronger evidence on the Cukierman-Meltzer hypothesis for Portugal at 4 lags. The main difference when the filtered series are used is that now no evidence appears for the Dotsey-Sarte hypothesis in Spain.

Since the results from the simple Granger-causality tests and those obtained by the Toda and Yamamoto (1995) procedure are basically identical, it seems that hardly any spurious effect due to the fractionally integrated variables occurs in our setting. At first sight this result seems to be at odds with the findings of Tsay and Chung (2000) who have shown that regressions involving fractionally integrated regressors can lead to spurious results. In particular, analyzing the bivariate regression of y_t on a constant and x_t where $y_t \sim \mathrm{I}(d_y)$ and $x_t \sim \mathrm{I}(d_x)$ they show that the corresponding t-statistic will be divergent provided $d_y + d_x > 0.5$.

We illustrate that their result does not apply to our setting in a small Monte-Carlo study by simulating the critical values of causality tests which are performed for two

[15]Since we also apply the Toda and Yamamoto (1995) procedure to inflation series for which $\widehat{d}_m + \widehat{d}_v < 0.5$, we should mention that in all these cases the results from the two methodologies were qualitatively identical.

independent series having the same orders of fractional integration as the estimated ones for the UK and Portugal.[16] Recall that both countries satisfy $\widehat{d}_m + \widehat{d}_v > 0.5$. The simulation is performed in the following way:

Step 1. We generate two independent series

$$\begin{aligned}\widetilde{\pi}_t &= (1-L)^{-\widehat{d}_m}\widetilde{\varepsilon}_t^{\pi} = \sum_{j=0}^{500}\psi_j^{\pi}\widetilde{\varepsilon}_{t-j}^{\pi} \qquad t=1,\ldots,504,\\ \widetilde{h}_t &= (1-L)^{-\widehat{d}_v}\widetilde{\varepsilon}_t^{h} = \sum_{j=0}^{500}\psi_j^{h}\widetilde{\varepsilon}_{t-j}^{h} \qquad t=1,\ldots,504,\end{aligned}$$

where $\widehat{d}_m$ and $\widehat{d}_v$ are the estimated orders of fractional integration and $\widetilde{\varepsilon}_t^{\pi}$ and $\widetilde{\varepsilon}_t^{h}$ are $\overset{iid}{\sim}\mathcal{N}(0,1)$. Hence, $\widetilde{\pi}_t$ and $\widetilde{h}_t$ are integrated of order $\widehat{d}_m$ and $\widehat{d}_v$, respectively, and satisfy the assumptions made in Tsay and Chung (2000).

Step 2. For the generated sample $\{\widetilde{\pi}_t, \widetilde{h}_t\}$ we run the following regressions:

$$\widetilde{\pi}_t = \beta_0 + \beta_1^h\widetilde{h}_t + \eta_t^{\pi}, \tag{5.3}$$

$$\widetilde{\pi}_t = \beta_0 + \sum_{j=1}^{k}\beta_j^h\widetilde{h}_{t-j} + \sum_{j=1}^{k}\beta_j^{\pi}\widetilde{\pi}_{t-j} + \eta_t^h \qquad \text{for} \quad k=4,8,12, \tag{5.4}$$

and calculate the corresponding value of the test statistic ($H_0: \beta_j^h = 0,\ j=1,\ldots,k$). Equation (5.3) corresponds to the setting described in Tsay and Chung (2000), while equation (5.4) is our setting from Table 5.4, panel B. Repeating step 1 and 2 for $M = 10000$ times we approximate the distribution of the test statistic. From the simulated distribution we calculate the 5% and 1% critical values.

The theoretical results derived in Tsay and Chung (2000) suggest that spurious regression occurs in equation (5.3), but what about equation (5.4) ? The simulation results presented in Table 5.5 show that spurious regression is a much more severe problem in the bivariate case considered by Tsay and Chung (2000) than in regressions

[16]Note, that the intention of the simulation is to show that the results of Tsay and Chung (2000) do not carry over to our setting and not to obtain the correct critical values for the Granger-causality tests using the original series. For this one would have to generate two series according to the equations (5.1) and (5.2) by using the parameter estimates from Table 5.2 and drawing innovations from the estimated standardized residuals.

including lagged dependent and independent variables with 4 or more lags. In the bivariate case the 5% critical value according to the F-distribution would be 3.86, while the simulated critical values are 17.91 ($\widehat{d}_m = 0.34, \widehat{d}_v = 0.46$) and 10.91 ($\widehat{d}_m = 0.14, \widehat{d}_v = 0.87$). Hence, relying on the F critical value in the bivariate regression would lead to rejection of the null hypothesis of no Granger-causality in by far too many cases. However, the more lags are included, the less different are the critical values of the F-distribution and the simulated distribution. In particular, the 5% critical value for 12 lags is 1.77 for the F-distribution, while the simulated ones are 1.82 ($\widehat{d}_m = 0.34, \widehat{d}_v = 0.46$) and 1.83 ($\widehat{d}_m = 0.14, \widehat{d}_v = 0.87$). Since the difference between the critical values for the F and the simulated distributions are very small in our setting the influence of spurious regression does not seem to play an important role. This explains why the simple Granger-causality results are very similar to those obtained by using the Toda-Yamamoto methodology.

Table 5.5: Simulated critical values.

	$d_m = 0, d_v = 0$		$\widehat{d}_m = 0.34, \widehat{d}_v = 0.46$		$\widehat{d}_m = 0.14, \widehat{d}_v = 0.87$	
	$F(k, 504 - k)$		$F^\star$		$F^\star$	
	5%	1%	5%	1%	5%	1%
Bivariate	3.86	6.69	17.91	34.42	10.91	18.46
$k = 4$	2.39	3.35	2.73	3.75	2.81	3.85
$k = 8$	1.96	2.54	2.05	2.65	2.10	2.71
$k = 12$	1.77	2.22	1.82	2.26	1.83	2.30

Notes: $F^\star$ is the simulated distribution of the test-statistic using $M = 10000$ generated samples. The figures are 5% and 1% critical values. k denotes the lag length in the Granger-causality test.

Simultaneous Approach

This section reports the estimation results of an ARFIMA-FIGARCH-in-mean model with lagged inflation included in the variance specification (the so called level effect). We estimate a system of equations that allows only the current value of the conditional variance to affect average inflation and up to the fifth lag of average inflation to influence the conditional variance. In other words, the model includes the mean equation which adds the variance of inflation (δh_t) to the expressions reported in Table 5.2, and the variance equation augmented by the term $\gamma_i \pi_{t-i}$. In the expressions for the conditional variances reported in Table 5.2, various lags of inflation (from 1 to 5) were considered with the best model chosen on the basis of the minimum value of the AIC. Table 5.6 reports only the two estimated parameters of interest. In four out of the ten countries

Table 5.6: ARFIMA-FIGARCH-in-mean-level models.

	Belgium	Finland	France	Germany	Italy
$\widehat{\gamma}_i$	–	0.097 [4]	0.028 [5]	–	0.003 [2]
		(3.066)	(1.811)		(1.830)
$\widehat{\delta}$	–	0.143	-1.203	–	0.299
		(0.689)	(1.405)		(1.017)
	Netherlands	**Portugal**	**Spain**	**Sweden**	**UK**
$\widehat{\gamma}_i$	0.058 [4]	–	–	0.153 [4]	0.048 [5]
	(3.825)			(6.160)	(2.868)
$\widehat{\delta}$	0.015	–	–	-0.901	-0.097
	(0.048)			(3.982)	(0.514)

Notes: For each of the ten inflation series Table 5.6 reports QML estimates of the in-mean and level parameters for the ARFIMA-FIGARCH-in-mean-level model. The numbers in parentheses are absolute t-statistics. A – indicates that there was no convergence. The numbers in [·] indicate the lag of inflation in the variance equation.

(Belgium, Germany, Portugal and Spain) there is no convergence when we employ the model with the simultaneous feedback. In all other six countries we find a positive association between lagged inflation and nominal uncertainty similar to that found with the two-step approach. However, another disadvantage of the simultaneous methodology is that in some cases the estimates of the conditional variances are negative.

In Finland, France, Italy and the UK the in-mean coefficient is insignificant, a result which is identical to that of the causality tests at lag 4. Similarly, in Sweden, as with the two-step approach, we find evidence for the Holland hypothesis. Moreover, in the four countries where there is no convergence, we estimate the model without the level effect ($\gamma_i = 0$) and the results (not reported) square with the findings of the two-step strategy at lag 4. That is, we do not find a significant effect of uncertainty on inflation. Hence, we generally find the two approaches to be in agreement. The only exception is the case of the Netherlands, where we estimate an insignificant δ, but find significant evidence for the Cukierman-Meltzer hypothesis at lag 4 in the two-step approach. However, it should be re-emphasized that such a result is plausible, since any relationship where uncertainty influences average inflation takes time to materialize and cannot be fairly tested in a model that restricts the effect to being contemporaneous.

European Monetary System

Hyung and Franses (2004) point out that inflation rates may perhaps show long memory because of the presence of neglected occasional breaks in the series rather than being really I(d). Our sample period includes various exchange rate and monetary policy regimes. For example, the Bundesbank set a monetary target in 1975, after the break up of Bretton Woods. Originally, a fixed money target was announced but after two years this was changed to a fixed range. Like many other central banks, the Bundesbank translated its main policy goals (e.g., controlling inflation) into near term interest rate objectives. It in turn supplied bank reserves to meet these objectives. After 1985 the Bundesbank supplied banks with reserves mainly via repurchase agreements. Reunification of course introduced new complexities for monetary management. The British also introduced money targeting in the mid-1970s in response to mounting inflation concerns. Although inflation fell subsequent to the 1973 oil price shock, beginning in 1978 prices in the United Kingdom began to accelerate again, with inflation ultimately reaching nearly 20% by 1980. The perception of an inflationary crisis led to a change in strategy in 1979. A comparison with Germany does not portray British monetary policy in a favorable light. Not only has British inflation had higher mean and greater volatility, but the unemployment rate has also been high and variable. However, in the 1980s British inflation performance did improve considerably, remaining well below the 1970s level and becoming less variable.

Overall, the four decades under investigation are characterized by persistently high inflation, as was the case from the late 1960s through the early 1980s, followed by the relatively shock-free 1990s. Since the early 1980s, there has been a tremendous improvement in macroeconomic performance in European countries. This was the case for two reasons. First, the global reduction in inflationary pressures. Second, some countries joined the European Monetary System (EMS) in 1979 in order to borrow Germany's anti-inflation reputation. This is less so, for the Netherlands, which has traditionally aligned its monetary policy stance to Germany's. Furthermore, both inflation and output growth have become more stable. In what follows we examine whether the transition from the high inflation of the sixties and seventies to an era of low inflation during the 1980s and 1990s affects the dynamic interaction between nominal uncertainty and either inflation or output growth by examining the period that starts in 1980 and continues till to the end of the sample. The choice of this period is also based on the widely accepted notion that with the introduction of the exchange rate mechanism (ERM) in March 1979 monetary stability was achieved in Europe.

Table 5.7 presents QML estimates of d_m and d_v. For all countries the estimated long memory conditional mean parameter is in the range $0.134 \leq \widehat{d}_m \leq 0.379$. The value of the coefficient for the Netherlands (0.134) is markedly lower than the corresponding value for Spain (0.379). However, although the estimated value of d_m for the Netherlands is relatively small it is significantly different from zero. The estimation of a FIGARCH model (not reported) for the Netherlands realized an estimated value of d_v close to and not significantly different from zero. In other words, the conditional variance of this inflation series is characterized by a stable GARCH behavior. For the other nine countries, the values of d_v vary from 0.203 (Sweden) to 0.339 (Germany).

It is noteworthy that for the majority of the countries the estimates of d_m and d_v are similar to the ones for the entire period. Moreover, in Portugal and Spain (the two countries which were characterized by a near integrated GARCH behavior) the estimated values of d_v (d_m) are lower (higher) than the corresponding values for the whole sample. Thus, for these two inflation series there appears to be a trade off between the degree of persistence in the first two conditional moments. In sharp contrast, for Belgium, Italy and the UK which were characterized by the presence of quite strong long memory in the inflation uncertainty, the estimates of both d_m and d_v are lower than the ones for the 1962:01–2004:01 period.

Table 5.7: ARFIMA-FIGARCH models 1980:01–2004:01.

	Belgium	Finland	France	Germany	Italy
$\widehat{d}_m$	0.146	0.136	0.190	0.209	0.289
	(2.102)	(2.973)	(2.333)	(2.571)	(4.718)
$\widehat{d}_v$	0.216	0.233	0.208	0.339	0.277
	(2.309)	(2.166)	(2.119)	(1.927)	(3.179)
	Netherlands	Portugal	Spain	Sweden	UK
$\widehat{d}_m$	0.134	0.218	0.379	0.160	0.275
	(2.407)	(1.983)	(5.257)	(2.122)	(2.339)
$\widehat{d}_v$	-	0.221	0.313	0.203	0.273
		(4.256)	(4.951)	(1.836)	(2.132)

Notes: For each of the ten inflation series Table 5.7 reports QML estimates of the two long memory parameters for the ARFIMA-(FI)GARCH model. The numbers in parentheses are t-statistics.

Generally speaking, in the majority of the cases the estimated values of d_m and d_v are lower than the corresponding values for the entire sample. This result is in agreement with the conclusion of Caporale and Gil-Alana (2003). They investigate the stochastic behavior of the inflation series in three hyperinflation countries. They test for fractional integration and find that when allowing for structural breaks the order of integration of the series decreases considerably. However, the parameter estimates still support the idea that dual long memory effects are present in the inflation process for nine out of the ten European countries. This result is consistent with the findings of a previous study by Bos et al. (1999). They find that the apparent long memory in monthly G7 inflation rates is quite resistant to mean shifts.

Table 5.8 reports the results of causality between inflation, nominal uncertainty and real growth for the various ARFIMA-FIGARCH models for the post-1979 period.[17]

Panel A considers Granger-causality from inflation to uncertainty about inflation. For this subperiod we find evidence that increased inflation raises its uncertainty in nine countries. For the Netherlands inflation has no impact on its uncertainty. Moreover, in this low-inflation period the evidence is mild(weak) for France(Germany) where it applies for two(one) of the chosen lags. Hence, the picture for the post-1979 period is similar to that of the entire period.

Panel B reports the results of the causality tests where causality runs from the nominal uncertainty to inflation. The findings for this subperiod provide support for the Cukierman-Meltzer hypothesis in some countries and for the Holland hypothesis in other countries. For three countries, uncertainty about inflation has a positive impact on inflation. Strong evidence in favour of the Cukierman-Meltzer hypothesis applies for Portugal and Spain. Relatively weak evidence applies for France (12 lags). Holland's hypothesis receives support in five countries, namely, in Germany, Italy, the Netherlands, the UK (optimal lag length) and Sweden (for all lags). None of the two theories is

[17]Table 5.8 reports (in case of significance) only the sign of the sum of lagged coefficients for the optimal lag length. The figures of the MW statistics for the three different lags (4, 8 and 12) and the corresponding p-values are omitted for reasons of brevity.

supported in Belgium and Finland where inflation is independent from changes in nominal uncertainty. The results are qualitatively similar to the analogous results from the entire sample. However, in Italy a negative effect begins to exist after 1979, whereas in France the evidence for the Cukierman-Meltzer hypothesis is weaker for the low-inflation period.

For completeness Panel C reports the results of causality from nominal uncertainty to output growth. For the post-1979 period, as for the entire sample period, we find evidence supporting the negative welfare effects of nominal uncertainty in Germany, Italy, Sweden and the UK. For both periods Dotsey-Sarte's hypothesis regarding the positive growth effects of uncertainty receives support in Finland, the Netherlands, Portugal and Spain. In the post-1979 period the evidence is weaker for the Netherlands and Portugal where it applies for only one of the chosen lags. In France the effect is positive during the entire period but turns to negative in the post-1979 period. In Belgium the effect is negative in the period 1965-2004 but it disappears in the low-inflation period.

Table 5.8: Granger-causality tests 1980:01–2004:01.

	Be	Fi	Fr	Ge	It	Ne	Po	Sp	Sw	UK
Panel A:	+	+	+	+	+	x	+	+	+	+
Panel B:	x	x	+	−	−	−	+	+	−	−
Panel C:	x	+	−	−	−	+	+	+	−	−

Notes: The countries are as in Table 5.3. Panels A, B and C are as in Table 5.8. A +(−) indicates that the sign of the effect is positive (negative). An x indicates that the effect is insignificant.

Comparing the results of the post-1979 period with those of the entire period, we note that for the majority of the countries the three effects for the low-inflation period are very similar to those for the entire period. For those countries where we found changes in the effects either the impact of nominal uncertainty on inflation or output growth became less significant which is not surprising since the inflation series are less volatile in the low-inflation period or are more in line with Holland's stabilization hypothesis.

5.6 Discussion

5.6.1 European Monetary Policy

The link between the inflation rate and its uncertainty acquires significant importance for the member countries of the Euro zone. The evidence that in all ten countries higher inflation causes greater uncertainty which then has negative output effects in five out of the ten countries strengthens the case for the choice of price stability as one of the objectives of monetary policy. Moreover, since the effects of nominal uncertainty on economic growth differ across the Euro zone, a common monetary policy that results in similar inflation rates across countries will have asymmetric real effects, provided these effects work via a change in nominal uncertainty. In other words, a reduction in inflation arising from a contractionary monetary policy applied by the ECB will increase growth in Belgium, Germany, Italy, Sweden, and the UK (where the Friedman hypothesis holds) but reduce it in Finland, France, the Netherlands, Portugal and Spain, where there is a positive effect of uncertainty. Therefore, the lack of uniform evidence supporting the second part of the Friedman hypothesis across the Euro zone countries has important policy implications as it makes a common monetary policy a less effective stabilization policy tool. It is noteworthy that evidence for the Dotsey-Sarte hypothesis obtains for the majority of the countries in the group which is characterized by a mild long memory in the conditional variance, and also for the two countries which exhibit near integrated GARCH behavior. In sharp contrast, evidence for the Friedman hypothesis applies in the three countries which are characterized by the presence of quite strong long memory GARCH behavior.

Moreover, less robust evidence is found regarding the direction of the impact of a change in nominal uncertainty on inflation. Countries like France, Portugal and Spain, for which we find evidence in favor of the Cukierman-Meltzer hypothesis, would be expected to gain significantly from EMU as the surrender of their monetary policy to the ECB would eliminate the policymakers' incentive to create inflation surprises. When Grier and Perry (1998) looked for institutional reasons why the inflation response to increased uncertainty varies across countries, they noted that the countries associated with an opportunistic response have much lower central bank independence ratings than the countries associated with a stabilizing response. We have used measures of central bank independence provided by Alesina and Summers (1993), which constructed a 1-4 (maximum independence) scale of central bank independence. Germany is rated as highly independent, with a score of 4. Netherlands also has a relatively independent central bank with a score of 2.5. In both countries increased inflation uncertainty lowers inflation as the sign at lag 12 (optimal lag length) is negative. Thus, one can argue that the most independent central banks are in countries where inflation falls in response to increased uncertainty. France has a relatively dependent central bank, with a score of 2. On the low side of the independence spectrum, Spain's rating is only 1. A lack of independence does seem to correspond to 'opportunistic behavior' because both countries show a highly significant positive effect of uncertainty on inflation. It is worth noting that evidence for the Cukierman-Meltzer hypothesis obtains for the two countries which exhibit near integrated GARCH behavior. In sharp contrast, evidence for the Holland hypothesis obtains for the majority of the countries which are characterized by the presence of mild long memory in nominal uncertainty. Finally, inflation is independent from changes in its uncertainty for two countries which are characterized by the presence of quite strong long memory in the conditional variance of the inflation rate.

5.6.2 Possible Extensions

The main goal of this chapter has been to investigate the link between nominal uncertainty and macroeconomic performance, and to estimate the two main parameters driving the degree of their persistence, for ten European countries. In that respect we achieved our goal. As Hassler and Wolters (1995) point out, a likely explanation of the significant persistence in the inflation rate series is the aggregation argument, which states that persistence can arise from aggregation of constituent processes, each of which has short memory. Alternatively, Baum et al. (1999) conjecture that the long memory property of monetary aggregates will be transmitted to inflation, given the dependence of long-run inflation on the growth rate of money. However, one might also ask why it is necessary to allow for long memory in the conditional variance of inflation. To answer this we must enquire into the possible theoretical sources of heteroscedasticity in the inflation shocks. It will be very useful to provide a theoretical rationale for the dynamics of inflation. Here the choice of the FIGARCH model is justified solely on empirical grounds.

There is substantial evidence that European inflation rates have long memory, a feature which can be captured by a fractional integrated I(d) model. Hyung and Franses (2002) put forward a joint model which incorporates both long memory and occasional level shifts. Overall, however, they find that the dominant feature in 23 US inflation rates is long memory and that the level shifts are less important. This result suggests several avenues for further research. One promising avenue would be to adapt the ARFIMA-FIGARCH model in a way that incorporates occasional level shifts in both the conditional mean and the conditional variance.

Bos et al. (2002) have emphasized that the introduction of two macroeconomic leading indicators namely, the unemployment rate and the short term interest rate, in the ARFIMA model lower the estimate of the fractional parameter and thus account partly for the persistence in inflation. More importantly, they argue that the multi-step forecast intervals of the ARFIMAX model are more realistic than of the ARIMAX model.[18] In the context of our analysis, incorporating macroeconomic variables either in the ARFIMA or in the FIGARCH specification or in both could be at work. We look forward to sorting this out in future work.

Finally, Morana (2002) suggests that long memory in inflation is due to the output growth process. His model implies that inflation and output growth must share a

[18] ARFIMAX denotes an ARFIMA model with explanatory variables in the mean equation.

common long memory component. Using a bivariate ARFIMA-FIGARCH model, which allows the measurement of uncertainty about inflation and output growth by the respective conditional variances, one can test for the empirical relevance of several theories that have been advanced on the relationship between inflation, output growth, real and nominal uncertainty.

5.7 Conclusions

In this chapter we have used ARFIMA-FIGARCH models to generate the conditional variance of inflation as a proxy of its uncertainty. We then performed Granger-causality tests to examine the bidirectional relationship between the two variables. We provided overwhelming evidence that increased inflation raises nominal uncertainty, confirming the theoretical predictions made by Friedman. Uncertainty surrounding future inflation appeared to have a mixed impact on inflation. The division of countries by how their inflation rates respond to inflation uncertainty appears to be closely related to existing rankings of central bank independence. We also found that increased nominal uncertainty significantly affects output growth in the ten European countries but not all in the same manner. The lack of uniform evidence supporting the second leg of the Friedman hypothesis across the Euro zone countries has important implications as it makes a common monetary policy a less effective stabilization policy tool.

The results in this chapter highlight the importance of modeling long memory not only in the conditional mean of inflation but in its conditional variance as well. We find that in all the cases there is a need to consider the joint ARFIMA-FIGARCH model, as in no case does one of its nested versions yield a better fit. Overall, these findings suggest that much more attention needs to be paid to the consequences of dual long memory when estimates of nominal uncertainty are used in applied research. In other words, as our results indicate, estimates of uncertainty that ignore the effects of dual long memory may seriously underestimate both the degree of persistence of uncertainty and its consequences for the inflation-uncertainty hypothesis.

Possible extensions could go in different directions. One could provide an enrichment of the dual long memory model by allowing lagged values of the conditional variance to affect the inflation. Finally, it is worth pointing to an important issue which we have not addressed. The dual long memory model used in this chapter ignores the possibility of structural instability caused by changing regimes. One could develop a dual long memory Markov switching model that explains the changing time series behavior of inflation in the post war era. This is undoubtedly a challenging yet worthwhile task.

5.8 Appendix

Impulse Response Functions

Figures 5.1 - 5.3 plot the effects of a one-time one-standard-deviation increase in inflation on nominal uncertainty (top, left), in nominal uncertainty on inflation (top, right) and in nominal uncertainty on output (bottom, middle) for Germany, the Netherlands and the UK. The dotted lines indicate $\pm$ two standard deviation bands computed by the asymptotic standard errors.

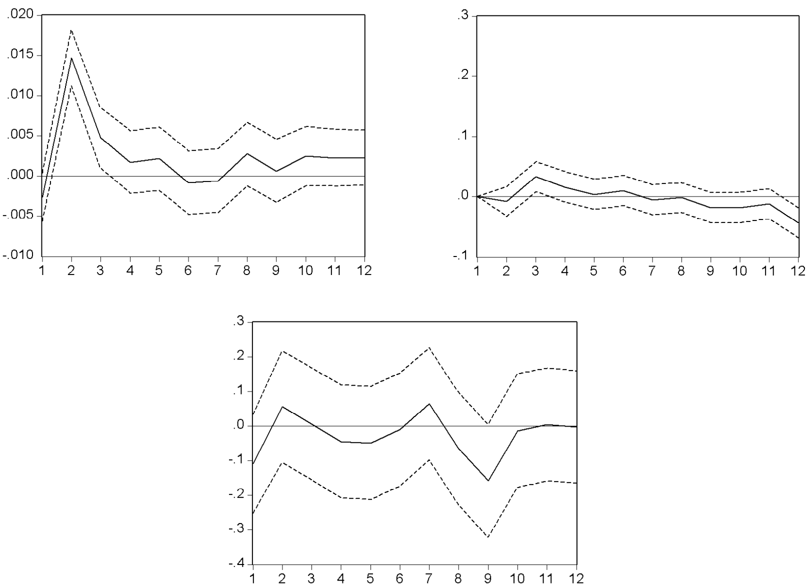

Figure 5.1: IRFs for Germany.

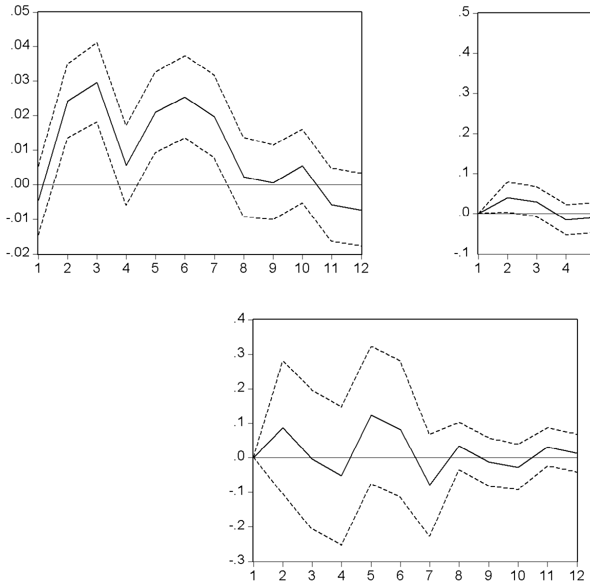

Figure 5.2: IRFs for the Netherlands.

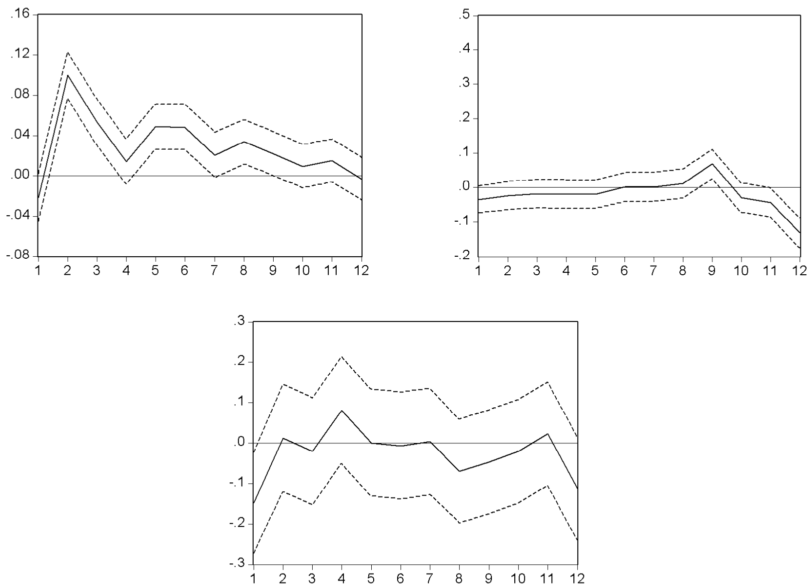

Figure 5.3: IRFs for the UK.

Part III

GARCH-in-Mean Models with Nonparametric Specifications

Chapter 6

A Specification Test for a Class of GARCH-in-Mean Models

6.1 Introduction

Economic theory often predicts a relationship between the level of a macroeconomic or financial variable and its second conditional moment. Typical examples are the relation between risk and expected return, inflation and nominal uncertainty or output growth and output uncertainty. In this chapter we consider an econometric specification for a variable Y_t of the form

$$\mathbf{E}[Y_t|\mathcal{F}_{t-1}] = m(h_t), \tag{6.1}$$

where h_t is the conditional variance of Y_t, typically given by some GARCH-type equation, and $\mathcal{F}_{t-1}$ represents the information available at $t-1$. The function $m(\cdot)$ can be considered as a risk premium. In certain cases economic theory directly implies a particular parametric specification $m = m_\gamma$ with γ being a parameter vector. One of the workhorses in financial econometrics, the GARCH-in-Mean (GARCH-M) model introduced by Engle et al. (1987), is a primary example of such a specification, where m_γ is typically linear or logarithmic in the conditional variance. In this chapter we suggest a test statistic for comparing the fit from such a parametric specification of the risk premium with some nonparametric fit of $m(\cdot)$ based on the integrated squared difference between the two curves, i.e. we provide a specification test for the appropriateness of the particular functional form of the risk premium imposed by a certain parametric GARCH-M specification. Since the conditional variance is unobservable such a nonparametric fit is not readily available. We estimate the conditional variance by an iterative procedure similar to that proposed by Linton and Perron (2003). The procedure starts by estimating the parametric GARCH-M by maximum likelihood. From the parameter estimates and the observed series we then create fitted conditional variances and regress Y_t on those estimates to obtain a nonparametric estimate of the risk premium. This estimate is then used to update the parameters of the variance equation from which new estimates of the conditional variance are obtained. The updated estimate of the conditional variance is a function of a parametric and a nonparametric component. Again, we update the estimate of the risk premium and so on until convergence of the mean function is achieved. Our main result states that the test statistic has a normal limit distribution under the null hypothesis. In particular, the limit distribution is independent of the number of iterations used for estimating the conditional variance. Under the alternative, the iterated estimate of the conditional variance approaches the true unobserved conditional variance although the initial parametric model for the mean was misspecified. Therefore, the test statistic based on the iterated estimate of the conditional variance reveals superior power properties in comparison with the test based on the initial estimate. Since the asymptotic distribution of the test statistic is approached quite slowly as the sample size goes to infinity, we suggest a bootstrap algorithm from which the critical values of the test statistic can be computed. Monte-Carlo simulations show that the bootstrap distribution approximates the distribution of the test statistic under the null hypothesis reasonably well in finite samples. Under the alternative, the test statistic reveals good power properties.

The basic idea of comparing parametric and nonparametric regression fits for testing the appropriateness of a particular parametric specification goes back to e.g. Härdle and Mammen (1993) who concentrated on regressions involving independently and identically distributed observations. The problem of testing for linearity in autoregressive time series models has been considered by Hjellvik and Tjøstheim (1995), Hjellvik et al. (1998) and Poggi and Portier (1997), while Kreiss et al. (2002) test for linearity in a more general times series setting which is not necessarily autoregressive. In all these studies the test statistic is based on the difference between a nonparametric and a parametric regression fit, but in contrast to our study the dependent and independent variable are observed directly. The contribution of this article is to deal with a situation in which the regressor is unobservable and replaced by an appropriate estimate. We show that under certain conditions the asymptotic results for the test statistic based on the iteratively fitted conditional variance are the same as if the conditional variance were observable. To achieve such results we have to base the test statistic on an over-smoothed nonparametric estimator. Our results are of more general interest than only in the context of GARCH-M models because they provide insight into the asymptotic behavior of nonparametric estimators relying on generated regressors. The results indicate in which situations the asymptotic distribution of the nonparametric estimator based on the unobservable regressor and the generated regressor is the same. This was discussed in a parametric context by Pagan (1984) and Pagan and Ullah (1988).

The GARCH-M was primarily motivated by Merton's (1973a) Intertemporal Capital Asset Pricing Model (ICAPM) which suggests that the conditional expected excess return on the stock market should vary positively with the conditional market variance. To capture this "fundamental law of finance" Engle et al. (1987) proposed a specification which assumes that $m_\gamma(h_t) = \mu + \lambda g(h_t)$, where h_t is modelled as a GARCH process.[1] When $\mu = 0$ and $g(h_t) = h_t$ equation (6.1) reflects the exact prediction by Merton (1973a): the conditional expected excess return on the market is proportional to the conditional market variance.

Many attempts have been undertaken to test Merton's (1973a) prediction by using various formulations of the GARCH-M model. The somewhat disappointing result, however, is that most empirical studies on the risk-return relation led to controversial

[1] More generally, the conditional mean can also be a function of lagged Y_t's and other covariates.

findings, some of which indicate a positive relationship such as French et al. (1987), Campbell and Hentschel (1992), Li (2003) or Guo and Neely (2006), some indicate a negative relationship such as Glosten et al. (1993), Pagan and Hong (1990), Li et al. (2005) or Guedhami and Sy (2005) while others do not find a significant relationship at all such as Bodurtha and Mark (1991), Baillie and DeGenarro (1990) or Shin (2005).

The three specifications for $g(h_t)$ employed in the above mentioned studies were either the conditional variance itself, the conditional standard deviation or the logarithm of the conditional variance. All three specifications restrict the shape of $m(\cdot)$ severely. In sharp contrast, Backus and Gregory (1993) using Mehra and Prescott's (1985) dynamic exchange economy model show that the relation between the excess return and its conditional variance can have virtually any shape: increasing, decreasing, flat, U-shaped, inverse U-shaped or non-monotonic depending on both the preferences of the representative agent and the probability structure across states. Similarly, Genotte and Marsh (1993) constructed a general equilibrium model in which the relationship $m(h_t) = \lambda h_t + k(h_t)$ holds, with $k(\cdot)$ depending on preferences and on the parameters of the distribution of asset returns. The Merton (1973a) relationship with $k(\cdot) = 0$ is obtained only as a very special case, namely if the representative agent has logarithmic utility.

We formulate our model such that under the null hypothesis the linear risk-return relation holds while under the alternative a semiparametric model is specified which only assumes the risk premium to be some smooth function. The alternative model thereby allows for shapes of the risk premium motivated by the results of Backus and Gregory (1993) and Genotte and Marsh (1993). In an empirical application we employ monthly and daily excess return data on the CRSP value-weighted index and estimate GARCH$(1, 1)$-M models for several periods. The results from these parametric models are in line with previous studies, i.e. using monthly data we find a positive but insignificant relation between the market excess return and its conditional variance, while we find a highly significant and positive relation using daily data. We then estimate the shape of the risk premium nonparametrically and apply our test procedure. The hypothesis of linearity cannot be rejected for almost all periods and data frequencies. Hence, we find no empirical evidence against the parametric specification suggested by Merton's (1973a) ICAPM. This finding suggests that the previous controversial results concerning the risk-return relation cannot be explained by misspecification of the risk premium.

This chapter is organized as follows. Section 6.2 reviews the empirical literature on testing the risk-return relationship by GARCH-M models. In Section 6.3 we first introduce the semiparametric GARCH-M model and the iterative estimation procedure and then motivate and state the test statistic and derive its asymptotic distribution. Moreover, we explain how our model relates to the parametric GARCH-M and explain the bootstrap procedure. Then we evaluate the empirical properties of the test in a Monte-Carlo simulation study in Section 6.4. Section 6.5 illustrates the method by an application to CRSP excess return data. Finally, we discuss several directions in which our approach could be naturally extended in Section 6.6. Section 6.7 summarizes the main conclusions. All proofs are deferred to the appendix.

6.2 Modelling the Risk-Return Relation

The static CAPM of Sharpe (1964) and Lintner (1965a,b) provides a formal framework for answering a fundamental question in finance: how should the risk of an investment affect its expected return? Merton's (1973a) ICAPM extends the static CAPM to an intertemporal setting with changing investment opportunities. While in the CAPM investors exclusively care about the wealth their portfolio produces at the end of the current period, in the ICAPM they are also concerned with the opportunities they will have to consume or invest the payoff. Therefore, investors choosing a portfolio at time t are concerned with how their wealth at time $t+1$ varies with future state variables such as labor income, prices of consumption goods, inflation and so on. In this model the equilibrium expected return on an asset depends not only on the conditional market risk (systematic risk), but also on conditional intertemporal risks which are measured by the conditional market covariances with the state variables. For simplicity we assume that there is only one state variable S_t. Further we assume that there exists a risk-averse representative agent with indirect utility function $U(W_t, S_t, t)$, where W_t denotes period t wealth and S_t can be viewed as describing the state of the investment opportunity set. Then the equilibrium relation for the market is given by

$$\begin{aligned}\mathbf{E}(r_{M,t} - r_{f,t}|\mathcal{F}_{t-1}) &= \left[-\frac{U_{WW}W_t}{U_W}\right]\mathbf{Var}(r_{M,t} - r_{f,t}|\mathcal{F}_{t-1}) \\ &\quad + \left[-\frac{U_{WS}}{U_W}\right]\mathbf{Cov}(r_{M,t} - r_{f,t}, S_t|\mathcal{F}_{t-1}) \qquad (6.2)\end{aligned}$$

where $r_{M,t}$ denotes the return on the market portfolio, $r_{f,t}$ the return on the risk-free asset and subscripts of U denote partial derivatives with respect to W_t and S_t. In this setting the conditional expected excess return on the market is linear in two components: first in a risk component namely the conditional market variance and second in a hedge component namely the conditional market covariance with the investment opportunities. If we additionally assume that the representative agent obeys a constant relative risk aversion utility function, it follows that $\lambda \triangleq \left[-\frac{U_{WW}W_t}{U_W}\right]$ is a positive constant equal to the Arrow-Pratt measure of relative risk aversion. The coefficient $\lambda_S \triangleq \left[-\frac{U_{WS}}{U_W}\right]$ can be interpreted as the price of intertemporal risk of the state variable. In this framework the equilibrium expected excess return on the market can be approximated as

$$\mathbf{E}(r_{M,t} - r_{f,t}|\mathcal{F}_{t-1}) \approx \lambda \cdot \mathbf{Var}(r_{M,t} - r_{f,t}|\mathcal{F}_{t-1}) \qquad (6.3)$$

either if the partial derivative of the representative agent's utility with respect to wealth is much larger than the partial derivative with respect to the state variable or if the variance of the change in wealth is much larger than the variance of the change in the state variable (see Merton, 1980, p. 329). Finally, the ICAPM reduces to the Sharpe-Lintner CAPM if the investment opportunity set is static or if investors exhibit logarithmic utility. In both cases equation (6.3) holds exactly. Equation (6.3) is often referred to as a conditional single-factor model, while equation (6.2) is labelled a conditional two-factor model. Empirical researchers testing equation (6.3) have to make an assumption concerning the intertemporal nature of the conditional variance of the market. The class of GARCH-M models provides a natural workhorse in which $h_t \triangleq \mathbf{Var}(r_{M,t} - r_{f,t}|\mathcal{F}_{t-1})$ is modelled as some type of GARCH equation. In the following we review some of the results obtained by previews studies investigating the risk-return relation.

First, we discuss some of the studies relying purely on parametric GARCH-M type models. French et al. (1987) estimate GARCH-M models on the daily excess returns of the S&P composite portfolio for the period 1928 to 1984. Using both the conditional variance and the conditional standard deviation specification they provide evidence for a significant positive relationship between excess returns and risk. Employing daily CRSP data and a GARCH-M model with either normal or t-distributed innovations Baillie and DeGenarro (1990) obtain positive but insignificant estimates for λ. Nelson (1991) again investigates CRSP data but uses his exponential GARCH-M (EGARCH-M) specification which allows positive and negative innovations to have an asymmetric effect on the conditional variance. For the data and period he considers, there is evidence for a negative but insignificant relation between market risk and expected return. Glosten et al. (1993) again employ the EGARCH-M model and confirm the findings of Nelson (1991). They include the nominal interest rate as well as October and January seasonal dummies as explanatory variables in the variance equation and report a significant and negative relation between the conditional monthly excess return and its conditional variance.

Although the finding of a negative relation between risk and excess return is at odds with the prediction of the ICAPM, it can be rationalized by general equilibrium models. Whitelaw (2000) investigates the relation between risk and excess return in a general equilibrium exchange economy characterized by a regime-switching consumption process. While a single-regime model generates a positive and essentially linear relation between expected returns and volatility, a two-regime model leads to a complex, nonlinear relation. At the market level this relation will be negative in the long-run. Intuitively this can be explained as follows. Regime shifts introduce large movements in the investors opportunity set, and therefore induce a desire among investors to hedge adverse changes. In some states of the world, the market claim provides such a hedge.

Specifically, when a regime shift is likely, its value is high and its expected return is low as a consequence. These are also the states of the world with high volatility, generating a negative relation between volatility and expected returns.

In contrast to the single-factor models employed in the studies mentioned so far, Scruggs (1998) makes use of a two-factor model. Including long-term government bond returns as a second factor, Scruggs (1998) finds evidence for a positive and significant relation between the excess market return and the conditional market variance. He argues that if the true relationship is a two-factor model then single-factor models are misspecified and their estimates of λ are subject to an omitted variable bias.[2] His empirical example shows that the omitted variable bias in $\widehat{\lambda}$ is sufficiently large to explain the negative and insignificant relation between the excess market return and the conditional market variance found in most previous studies. In contrast, Guedhami and Sy (2005) claim that the often reported negative relationship is not due to the omission of the hedge term associated with the ICAPM. Using an instrumental variables method they estimate a two-factor model including the long-term government bond, but still find evidence for a negative risk-return relation. Guo and Whitelaw (2006) argue that one can neglect the hedge term when using daily data, because investment opportunities change slowly at the business cycle frequency and can be treated as being constant at the daily frequency. However, they also find that expected returns are driven primarily by the hedge component at a monthly or quarterly frequency. Guo and Neely (2006) employing daily international stock market data – neglect the hedge term – and show that the risk-return relation is positive and significant in almost all the markets.

Campbell and Hentschel (1992) and Harvey and Siddique (1999) model the co-movement between the conditional skewness and the conditional variance, the so-called volatility feedback effect. Both studies provide empirical evidence that the conditional

[2] If the true model is two-factor, but a single-factor model is estimated, the bias is given by: $\widehat{\lambda} - \lambda = \lambda_S \mathbf{Cov}(\sigma^2_{M,t} - r_{f,t}, \sigma_{MS,t})/\mathbf{Var}(\sigma^2_{M,t} - r_{f,t})$ where $\sigma_{MS,t}$ denotes the covariance between the market risk premium and the state variable. Note, that Scruggs (1998) assumes that $\lambda_S = -U_{WS}/U_W$ is constant over time.

skewness appears to possess a systematic relation to expected returns and their conditional variance. Therefore, they argue that the omission of the effect of the conditional skewness could explain the puzzling finding of a negative risk-return relationship. Along these lines Li (2003) models daily S&P500, FTSE100 und DAX30 returns as GARCH-M using the skewed t-distribution and allowing the conditional skewness to influence the mean equation. For all three indices, the parameter estimates suggest a positive and significant relation between the conditional variance and the expected excess return.

Overall the evidence provided by the above mentioned studies is mixed. Some pointing to a positive others to a negative or an insignificant relation between excess return and risk. All these studies rely on parametric specifications for the risk premium and the conditional variance. In the following we briefly review some recent studies using nonparametric estimation techniques.

We begin with studies employing nonparametric techniques to estimate the conditional variance. Pagan and Ullah (1988) and Pagan and Hong (1990) argue that the conditional variance is a highly nonlinear function of the past whose form is not adequately captured by parametric GARCH-M models. Therefore, they firstly estimate the conditional variance nonparametrically and then regress the excess return on the estimated conditional variance by least squares methods. Using this procedure they find a negative but insignificant in-mean coefficient. Pagan and Hong (1990) restrict h_t to be a function of the last p observations $\{Y_{t-1}, \ldots, Y_{t-p}\}$ for some fixed p in order to avoid the well known "curse of dimensionality": the optimal rate of convergence decreases with dimensionality p. This restriction however is very problematic since – as has been shown in many other studies – the conditional variance is a highly persistent process and so it is unlikely that its dynamics can be adequately captured by such an estimator. Linton and Mammen (2005) recently suggested an alternative approach based on kernel smoothing and profiled likelihood circumventing the curse of dimensionality and nevertheless allowing the conditional variance to depend on the whole past of the process Y_t. They specify the conditional variance as additive in Y_{t-j} with the restriction that the different additive functions are proportional to each other. This implies that only one univariate function needs to be estimated. Hence their semiparametric ARCH(∞) model is capable of taking into account both nonlinearity and high persistence in the conditional variance.[3] A similar approach is used by Li et al. (2005) who propose a test for the existence of an in-mean effect. The test for the in-mean effect is a simple regression of the excess return on the generated conditional variance series. Investigating twelve international stock markets Li et al. (2005) find a negative and (partly) significant relation between risk and excess returns. Shin (2005) employs the same method to 14 emerging international stock markets and reports a positive but insignificant relationship between stock returns and volatility.

Next, we discuss the two studies which allow for more flexible specifications of the conditional mean. Das and Sarkar (2000) suggest the ARCH-in-Nonlinear-Mean (ARCH-NM) model which defines $g(h_t)$ as a Box-Cox power transformation of the conditional variance. Obviously, this model nests the simpler parametric specifications mentioned above under certain constraints on the power transformation parameter. Although the ARCH-NM specification is favored compared to the standard specification when applied to stock return data, Das and Sarkar (2000) conclude that the model fit is not entirely satisfactory. They conjecture that the ARCH-NM is still not nonlinear enough. Finally, Linton and Perron (2003) suggest an algorithm for estimating a semiparametric (E)GARCH-M model which does not assume a functional form for the shape of the risk premium a-priori. The model is semiparametric in the sense that the conditional variance equation is modelled parametrically as GARCH or EGARCH, while the shape of the conditional mean is estimated nonparametrically.[4] Although no

[3]Ghysels, Santa-Clara and Valkanov (2005) also argue that the GARCH parameterization is not flexible enough to model the conditional variance appropriately. Instead they make use of the so-called mixed data sampling (MIDAS) approach which estimates the conditional variance of monthly returns as a weighted average of lagged squared daily returns where the weights itself are estimated from the data. In a second step the monthly excess returns are regressed on the MIDAS estimated conditional variances. Using this procedure significant evidence for a positive risk premium is obtained.

[4]Masry and Tjøstheim (1995) investigate the problem of nonparametrically estimating both the mean and the conditional variance function. However, their procedure does not allow for a risk premium.

asymptotic theory is provided for their estimator, Monte-Carlo simulations show that the procedure works reasonably well. An application of the semiparametric EGARCH-M to excess returns on the CRSP value-weighted index reveals a hump-shaped pattern of the risk premium which could not be detected by the parametric EGARCH-M model.

6.3 The Semiparametric Model and the Test Statistic

The last section discussed the controversial empirical findings on the risk-return relation. Several possible explanations (misspecification of the conditional variance, omitted variables bias, ect.) were addressed in the literature without convincing success. In this section we focus on the obvious possibility of misspecification of the mean function. The parametric specification of the risk premium implied by the Merton (1973a) ICAPM results from very specific assumptions, and as shown by Genotte and Marsh (1993) and Backus and Gregory (1993), if these assumptions do not hold, the shape of the risk premium can have virtually any form. Therefore, it seems natural to ask for the appropriateness of the commonly applied specifications of the mean function. We consider a general class of in-mean models which nest the standard GARCH-M as a special case characterized by a particular choice of the mean and variance function. For such a model we address the problem of testing for the correct choice of a particular parametric specification of the mean function.

Under the null hypothesis we consider an in-mean model with a parametric mean function depending on a finite-dimensional parameter γ_0:

$$Y_t = m_{\gamma_0}(h_t(\psi_0, \gamma_0)) + \varepsilon_t, \tag{6.4}$$

where $\varepsilon_t = \sqrt{h_t(\psi_0, \gamma_0)} Z_t$ with Z_t being a sequence of independent and identically distributed random variables with expectation zero and variance one. Here h_t is a function of the parameters ψ_0, γ_0 and of $Y_{t-1}, Y_{t-2}, \dots, Y_1, Y_0, Y_{-1}, \dots$. A typical example could be that h_t follows a GARCH$(1,1)$ process, but any specification from the GARCH family is possible. For simplicity, dependence on $Y_{t-1}, Y_{t-2}, \ldots$ is suppressed in the notation. We also write h_t for $h_t(\psi_0, \gamma_0)$ and m_0 for m_{γ_0}. By construction we have $\mathbf{E}[\varepsilon_t|\mathcal{F}_{t-1}] = 0$ and $\mathbf{E}[\varepsilon_t^2|\mathcal{F}_{t-1}] = h_t$, where $\mathcal{F}_t = \sigma(Z_t, Z_{t-1}, \ldots)$ is the σ-field of events generated by $\{Z_s, s \leqslant t\}$. We assume that the true parameter vector $\theta_0 = (\psi_0, \gamma_0)$ is in the interior of Θ, a compact, convex finite dimensional parameter space.

The alternative model is given by a semiparametric version of equation (6.4) with a smooth mean function $m(\cdot)$, but ε_t and h_t as before. The semiparametric alternative has two distinct advantages over previous specifications: (i) it does not rely on any parametric specification of $m(\cdot)$, and (ii) it allows for persistence in the conditional variance process since it does not restrict $\mathcal{F}_{t-1}$ as in Pagan and Hong (1990). The specification under the alternative is closely related to the model considered by Linton and Perron (2003).

6.3.1 Iterative Estimation of Conditional Mean and Variance

For some initial parametric estimators $\widehat{\gamma}$ and $\widehat{\psi}^{(0)}$ we consider an estimate $\widehat{h}_t^{(0)}$ of $h_t(\psi_0, \gamma_0)$ which can be written as a function of $\widehat{\theta}^{(0)} = (\widehat{\psi}^{(0)}, \widehat{\gamma})$ and the past observations $Y_1, \ldots, Y_{t-1}$. We suppress dependence on $Y_1, \ldots, Y_{t-1}$ in the notation and we write $\widehat{h}_t^{(0)} = \widehat{h}_t(\widehat{\theta}^{(0)})$ where $\widehat{h}_t$ is a random function that depends on $Y_1, \ldots, Y_{t-1}$. Note that typically $\widehat{h}_t^{(0)}$ depends also on $\widehat{\gamma}$ because (simultaneously) fitting the residuals ε_t and/or Z_t requires an estimate of m_{γ_0}. This is for instance the case when a parametric GARCH-M is estimated by (quasi-)maximum likelihood in the initial step.

We will use iterative updates of the estimate $\widehat{\psi}^{(0)}$. These updates are denoted by $\widehat{\psi}^{(k)}$ with $k \geq 1$. The estimator of γ_0 will not be updated. This is done for the following reason. Because our semiparametric alternative model contains nonparametric components, updates of the parametric estimators will slow down the rate of convergence to nonparametric rates. Our test for the parametric hypothesis is based on the comparison of estimators of m_{γ_0} on the hypothesis and on the alternative. If the estimate of γ_0 is updated this will introduce an additional bias term that does not cancel out when comparing the estimators on the hypothesis and on the alternative.

The iterative update of the estimators of m_{γ_0}, ψ_0 and h_t works as follows. Given the fit $\widehat{h}_t^{(k-1)}$ of h_t calculated in the $(k-1)$-th cycle, the estimate of m_{γ_0} is updated by smoothing Y_t versus $\widehat{h}_t^{(k-1)}$. The resulting smoother is denoted by $\widehat{m}_b^{(k)}$. Then using the observations and $\widehat{m}_b^{(k)}$, the estimator of ψ_0 and h_t is updated. The resulting estimators are denoted by $\widehat{\psi}^{(k)}$ and $\widehat{h}_t^{(k)}$. We now describe the iteration steps in more detail.

For x in a bounded closed interval I and $k \geq 1$ the updated estimator of m_{γ_0} is defined as

$$\widehat{m}_b^{(k)}(x) = \frac{\widehat{r}_b^{(k)}(x)}{\widehat{f}_b^{(k)}(x)} + m_{\widehat{\gamma}}(x), \tag{6.5}$$

with $\widehat{r}_b^{(k)}(x) = \frac{1}{T}\sum_{t=1}^T K_b(\widehat{h}_t^{(k-1)} - x)[Y_t - m_{\widehat{\gamma}}(\widehat{h}_t^{(0)})]$ and $\widehat{f}_b^{(k)}(x) = \frac{1}{T}\sum_{t=1}^T K_b(\widehat{h}_t^{(k-1)} - x)$ and where $K_b(\cdot) = b^{-1}K(\cdot/b)$ with K being a kernel function and bandwidth parameter b. In the simulations we also use the update

$$\widetilde{m}_b^{(k)}(x) = \widehat{f}_b^{(k)}(x)^{-1}\frac{1}{T}\sum_{t=1}^T K_b(\widehat{h}_t^{(k-1)} - x)Y_t. \tag{6.6}$$

However, the theoretical treatment of $\widehat{m}_b^{(k)}(x)$ is easier because some bias terms cancel in the asymptotic analysis that otherwise could only be analyzed under rather strong additional assumptions. For $x \notin I$ the estimate $\widehat{m}_b^{(k)}(x)$ is put equal to the old estimate $m_{\widehat{\gamma}}(x)$. Thus for $x \notin I$ the estimate of $m_{\widehat{\gamma}}(x)$ is not updated. Alternatively, an updated parametric fit for $x \notin I$ could also be considered. For simplicity, this not pursued here. Furthermore, it could be considered that the choice of the interval I depends on the sample size T and grows to the positive real line for $T \to \infty$. We also do not discuss this here. In the simulations we have taken $I = (0, \infty)$ to avoid the discussion of the choice of I. We conjecture that under our mixing conditions (see Assumption 3 below) differences between different choices of I will be minor for our test.

In a next step the fit of $h_t = h_t(\theta_0)$ is updated. We suppose that the update $\widehat{h}_t^{(k)}$ can be written as a function of $\widehat{m}_b^{(k)}$ and $\widehat{\psi}^{(k)}$ and the observations $Y_1, ..., Y_{t-1}$. Again, we suppress dependence on $Y_1, ..., Y_{t-1}$ in the notation and we write $\widehat{h}_t^{(k)} = \widehat{h}_t(\widehat{\psi}^{(k)}, \widehat{m}_b^{(k)})$

where in abuse of notation we denote the function by $\widehat{h}_t$, as the related function $\widehat{h}_t$ of step 0. We suppose that the function does not depend on k and that $\widehat{h}_t(\widehat{\psi}^{(0)}, m_{\widehat{\gamma}}) = \widehat{h}_t(\widehat{\psi}^{(0)}, \widehat{\gamma})$.

The above procedure can be performed for a finite fixed number of iterations or until a convergence criterium is fulfilled. The asymptotic theory is developed for a fixed number of iterations. In the simulations we use the criterium

$$\delta(k) = \frac{\sum_{j=1}^{J} \left(\widehat{m}_b^{(k)}(x_j) - \widehat{m}_b^{(k-1)}(x_j)\right)^2}{\sum_{j=1}^{J} \left(\widehat{m}_b^{(k-1)}(x_j)\right)^2 + \bar{c}} < \bar{c} \tag{6.7}$$

for some small prespecified $\bar{c}$, where x_j, $j = 1, \ldots, J$, are equally spaced grid points on I. We choose $\bar{c} = 0.001$.

6.3.2 The Test Statistic

We now come to the test statistic which will be based on the difference between a smoothed version of the initial parametric estimator and a Naradaya-Watson kernel estimator of the regression function. The null and alternative hypothesis can be written as

$$\begin{aligned} H_0: &\quad \mathbf{P}(m(\cdot) = m_{\gamma_0}(\cdot)) = 1 \text{ for some } \gamma_0 \in \Theta_\gamma = \{\gamma | (\psi, \gamma) \in \Theta\} \\ \text{and} \quad H_1: &\quad \mathbf{P}(m(\cdot) = m_{\gamma}(\cdot)) < 1 \text{ for any } \gamma \in \Theta_\gamma = \{\gamma | (\psi, \gamma) \in \Theta\}. \end{aligned}$$

The test statistic utilizes the fact that the null hypothesis is equivalent to the condition that the L_2-distance between the two functions is zero.

We consider the following test statistic

$$\widehat{\Gamma}_T^{(k)} = \int \left\{ \frac{\frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t^{(k)} - x)\left[Y_t - m_{\widehat{\gamma}}(\widehat{h}_t^{(0)})\right]}{\frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t^{(k)} - x)} \right\}^2 w(x)dx, \tag{6.8}$$

where $w(x)$ is some nonnegative and bounded weighting function.

Note, that in the test statistic we subtract $m_{\widehat{\gamma}}(\widehat{h}_t^{(0)})$ from Y_t and not $m_{\widehat{\gamma}}(\widehat{h}_t^{(k)})$. This is done to have a parametric rate for $m_{\gamma_0}(h_t) - m_{\widehat{\gamma}}(\widehat{h}_t^{(0)})$ on the hypothesis. In the simulations we also used $m_{\widehat{\gamma}}(\widehat{h}_t^{(k)})$.

Equation (6.8) can be interpreted as the integrated squared difference between a smoothed version of the initial parametric estimate $m_{\widehat{\gamma}}$ and the Naradaya-Watson kernel estimate $\widetilde{m}_b^{(k+1)}$ of the regression function $m(x)$ defined in equation (6.6). The reason for smoothing the parametric estimate is that whereas $m_{\widehat{\gamma}}$ is asymptotically unbiased and converging at rate $\sqrt{T}$, the nonparametric estimate $\widetilde{m}_b^{(k+1)}$ has a kernel smoothing bias and convergence rate $\sqrt{Tb}$. Replacing $m_{\widehat{\gamma}}$ by its smoothed version introduces an artificial bias. As a result, under the null hypothesis the bias of $\widetilde{m}_b^{(k+1)}$ cancels with the one of the smoothed version of the parametric estimate $m_{\widehat{\gamma}}$.

Under the assumption of independent and identically distributed observations, Härdle and Mammen (1993) have shown that under the null hypothesis the above test statistic with h_t observable (and $k = 0$) has an asymptotic normal distribution. Kreiss et al. (2002) extend the results of Härdle and Mammen (1993) to settings with dependent data. Their version of the test statistic can be interpreted as multiplying the weight function $w(x)$ with the squared stationary density of the conditional variance. This particular weighting scheme implies that one down-weights observations in areas where the data are sparse. The results of Kreiss et al. (2002) do not apply directly to our setting since $h_t = h_t(\theta_0)$ is unobservable.

We start with a discussion of the asymptotic behavior of $\widehat{\Gamma}_T^{(k)}$ for $k = 0$. The following assumptions are made:

Assumption 1. *The kernel K has bounded support ($[-1, 1]$, say) and a continuous derivative. The bandwidth b is of order $T^{-\eta}$, i.e.*

$$0 < \liminf_{T\to\infty} T^\eta b \leq \limsup_{T\to\infty} T^\eta b < \infty$$

for a constant η with $0 < \eta < \frac{1}{3}$.

Assumption 2. *It holds that $\mathbf{E}[\exp(\rho|Z_t|)] < \infty$ for $\rho > 0$ small enough.*

Assumption 3. *The process h_t is stationary and β-mixing with mixing coefficients $\beta(j) \leq c\rho^j$ for constants $c > 0$ and $0 < \rho < 1$. The density f_h of h_t is Lipschitz continuous and bounded away from 0 on I. The joint density of h_t and h_{t+s} is bounded on $I \times I$, uniformly in s.*

Assumption 4. *The function $m_\gamma(x)$ is differentiable with respect to γ at the point $\gamma = \gamma_0$ for all $x \in I$ and for the derivative $\dot m_{\gamma_0}$ it holds that*

$$\sup_{x\in I, \|\gamma-\gamma_0\|\le\delta} \left| m_\gamma(x) - m_{\gamma_0}(x) - (\gamma-\gamma_0)^T \dot m_{\gamma_0}(x)\right| = O(\delta^2)$$

for $\delta \to 0$. The derivative $\dot m_{\gamma_0}$ fulfills the following Lipschitz condition

$$\sup_{u,v\in I, \|u-v\|\le\delta} |\dot m_{\gamma_0}(u) - \dot m_{\gamma_0}(v)| = O(\delta^\kappa)$$

for $\delta \to 0$ with a constant $\kappa > 0$. Furthermore, $m_\gamma(x)$ is continuously differentiable with respect to x for $x \in I$.

Assumption 5. *It holds that $\|\widehat\theta^{(0)} - \theta_0\| = O_P(T^{-1/2})$.*

Assumption 6. *There exists a stationary sequence $\dot h_t$ such that*

$$\sup\left|\widehat h_t(\theta) - \widehat h_t(\theta_0) - (\theta-\theta_0)\dot h_t\right| = o_P(T^{-1/2}\log(T)^{-1/2}),$$

where the supremum runs over all t and θ with $T^{1/2-\delta}b \le t \le T$, $\|\theta-\theta_0\| \le CT^{-1/2}$, and with $\widehat h_t(\theta)$ or $\widehat h_t(\theta_0)$ or h_t in I. The process $(\dot h_t, h_t)$ is stationary and β-mixing with $\beta(j) \le c\rho^j$ for constants c and ρ as in Assumption 3. Furthermore $\mathbf{E}|\dot h_t|^r$ is finite for an $r > 2$.

Assumption 7. *For $C > 0$, $T^{1/2-\delta}b \le t \le T, \|\theta-\theta_0\| \le CT^{-1/2}, \|\theta'-\theta_0\| \le CT^{-1/2}$ it holds that*

$$|\widehat h_t(\theta) - \widehat h_t(\theta')| \le R_T\|\theta'-\theta\|^\tau + S_T$$

for random sequences R_T and S_T with $R_T = O_P(T^\varsigma)$ and $S_T = O_P(T^{-1/2-\nu}b)$ for constants ς and $\nu, \tau > 0$.

Assumption 8. *The weight function w is continuous and the closure of its support lies in the interior of I.*

We now shortly discuss the conditions. Assumption 1 is a standard smoothing condition. We do not assume that the bandwidth is of an order that is optimal for estimation under certain smoothness conditions on m_{γ_0}, e.g. that the bandwidth is of order $T^{-1/5}$. Such an assumption would be too restrictive because tests that look for more global deviations from the hypothesis make also sense. Assumption 2 is needed because the techniques from empirical process theory that will be used below require subexponential tails. The assumption could be replaced by higher order moments conditions if more involved mathematical arguments would be used. In particular, Assumption 2 is fulfilled for the standard model of Gaussian Z_t. The β-mixing condition in Assumption 3 could be replaced by the assumption that $\beta(j) \leq aj^{-c}$ for a constant $a > 0$ and for a constant c that is large enough. We avoided an exact check of the necessary size of the constant c because we have no examples of ARCH models where Assumption 3 does not hold but where this weaker assumption applies. Assumption 4 is a condition on the smoothness of the mean function. Assumptions 5 – 7 state conditions on the accuracy of the estimates of θ_0 and h_t and on the smoothness of $\widehat{h}_t(\theta)$ as a function of θ. Assumptions 5 and 6 are needed because we make no assumptions on the specific form of the estimators of the parameters. We remark that Assumption 7 is very weak because it is allowed that the random variable R_T may grow with rate T^ς for an arbitrary positive constant ς. In Assumptions 6 and 7 we allow that h_t has not the required properties for an initial period $1 \leq t < T^{1/2-\delta}b$.

The following theorem states that under the null hypothesis $T\sqrt{b}\ \widehat{\Gamma}_T^{(0)}$ is asymptotically normal. In the proof we show that $T\sqrt{b}\ \widehat{\Gamma}_T^{(0)}$ can be written as a sum of three components whereby the first term is dominating the other summands and determines the asymptotic distribution of the test statistic.

Theorem 6.1. *Assume that Assumptions 1 – 8 apply. Then under H_0 it holds that*

$$T\sqrt{b}\ \frac{\widehat{\Gamma}_T^{(0)} - b^{-1/2}M}{\sqrt{V}} \tag{6.9}$$

converges in distribution to a standard normal distribution. Here

$$\begin{aligned} M &= K^{(2)}(0)\int xw(x)f_h^{-1}(x)dx, \\ V &= 2K^{(4)}(0)\int x^2w^2(x)f_h^{-1}(x)dx, \end{aligned}$$

where $K^{(k)}$ denotes the k-fold convolution of K with itself.

We now discuss the test statistic $\widehat{\Gamma}_T^{(k)}$ for $k \geq 1$. In particular, we will show that replacing $\widehat{h}_t^{(0)}$ by the iterative estimator $\widehat{h}_t^{(k)}$ described above does not effect the asymptotic distribution of the test statistic. The following additional assumptions are needed to obtain our next result on the asymptotic distribution of $\widehat{\Gamma}_T^{(k)}$ for $k \geq 1$.

Assumption 9. *Assumption 1 holds for a constant η with $0 < \eta < \frac{1}{5}$.*

Assumption 10. *For all $C_* > 0$, for a constant $C > \|D_2 m_0\|_\infty$ and for $\iota > 0$ small enough it holds for functions m_1, m_2 with $\|m_j - m_0\|_\infty \leq C_*[(bT)^{-1/2}\log(T) + b^2]$, $\|D_2 m_j\|_\infty \leq C$ and parameters ψ_1, ψ_2 with $\|\psi_j - \psi_0\|_2 \leq C_* b^{3/2}(T)^{-\iota}$ for $j = 1, 2$ that*

$$|\widehat{h}_t(\psi_1, m_1) - \widehat{h}_t(\psi_2, m_2)| \leq V_T\|\psi_1 - \psi_2\|_2 + W_T\|m_1 - m_2\|_\infty.$$

Here V_T and W_T are random variables with $V_T = O_P(T^\upsilon)$ and $W_T = O_P(T^\xi)$ with constants υ and ξ that fulfil $\upsilon < \iota$ and $15\eta + 8\xi < 3$.

Assumption 11. *For $l \leq k$ it holds that*

$$\|\widehat{\psi}^{(l)} - \psi_0\|_2 = O_P(b^{3/2}T^{-\iota}).$$

Note that we now exclude the case that the bandwidth b is of order $T^{-1/5}$. The reason is that we apply uniform convergence results over sets of functions with bounded second derivatives. We need that $\widehat{m}_b^{(k)}$ is an element of this set (with probability tending to one). This requires oversmoothing, i.e. $T^{1/5}b \to \infty$. The rate $T^{-1/5}$ appears as the boundary case that is just excluded. We also conjecture that our results do not hold if the bandwidth is too small. Using the most powerful methods from empirical process theory one cannot achieve a uniform rate of convergence over classes of higher entropies. For the result of the theorem one needs an expansion of $\widehat{m}_b^{(k)}$ that is of order $o_P(T^{-1/2})$. If the interest lies in estimating $m(\cdot)$ then an expansion of order $o_P((bT)^{-1/2})$ is needed. We conjecture that this expansion could be derived by the methods of this chapter also for bandwidths of order $T^{-1/5}$.

Our next theorem states that $\widehat{\Gamma}_T^{(k)}$ has the same asymptotic distribution as $\widehat{\Gamma}_T^{(0)}$.

Theorem 6.2. *Assume that Assumptions 2 – 11 apply. Then under H_0 it holds that*

$$T\sqrt{b}\,\frac{\widehat{\Gamma}_T^{(k)} - b^{-1/2}M}{\sqrt{V}}$$

converges in distribution to a standard normal distribution. Here M and V are defined as in Theorem 6.1.

The advantage of using $\widehat{\Gamma}_T^{(k)}$ with $k \geq 1$ in comparison to $\widehat{\Gamma}_T^{(0)}$ may be explained as follows. The power of the test statistic depends on the accuracy with which the nonparametric estimate of the mean function can approximate the true mean function. Under the alternative, the parametric model for the mean which is initially estimated is misspecified. As a consequence, the nonparametric estimate of the mean function based on the inconsistent estimate $\widehat{h}_t^{(0)}$ will poorly approximate the true mean function. This leads to a low power of the test statistic $\widehat{\Gamma}_T^{(0)}$. The simulations in the next section will show that the iterative estimation procedure overcomes this problem and results in a precise estimate of $m(\cdot)$. The test statistic $\widehat{\Gamma}_T^{(k)}$ which is based on this iterated estimate will dispose of considerably better power properties than $\widehat{\Gamma}_T^{(0)}$.

Note, that we did not distinguish between the bandwidth parameter used for the estimation of the mean function and the one used in the test statistic. In the derivation of the theorems we treat them as identical. In the simulations and the application we choose the bandwidth parameter in the iterative estimation procedure by cross-validation as was suggested in Linton and Perron (2003) and is discussed in the next subsection. To reduce notation we do not equip the bandwidth parameter with an index k. We will report the test statistic for several choices of the bandwidth to document the robustness of the outcome of the test with respect to variations in the bandwidth parameter.

6.3.3 Parametric and Semiparametric GARCH$(1,1)$-M

The model we considered in equation (6.4) neither specifies a particular choice of m_{γ_0} nor of $h_t(\theta_0)$. The parametric GARCH$(1,1)$-M is the most popular version of such a model. In this subsection we describe this model in detail. We review the three most commonly used specifications for the risk premium, the conditions which imply covariance and strict stationarity and we discuss the issue of estimation. Then we will briefly explain the semiparametric GARCH$(1,1)$-M version of Linton and Perron (2003) and relate their approach to ours.

The GARCH$(1,1)$-M model is given by

$$Y_t = m_{\gamma_0}(h_t(\theta_0)) + \varepsilon_t \quad (6.10)$$

$$\varepsilon_t = \sqrt{h_t(\theta_0)} Z_t \quad (6.11)$$

$$h_t(\theta_0) = \omega_0 + \alpha_0 \varepsilon_{t-1}^2 + \beta_0 h_{t-1}(\theta_0). \quad (6.12)$$

The conditional expectation of Y_t is parameterized as $m_{\gamma_0}(h_t(\theta_0)) = \mu_0 + \lambda_0 g(h_t(\theta_0))$. The vector θ contains the parameters of the mean and variance functions, i.e. $\theta_0 = (\psi_0, \gamma_0)$, with $\psi_0 = (\omega_0, \alpha_0, \beta_0)$ and $\gamma_0 = (\mu_0, \lambda_0)$. Three parametric specifications for the function g are commonly applied. The original Engle et al. (1987) specifications assume either $g(h_t(\theta_0)) = h_t(\theta_0)$ or $g(h_t(\theta_0)) = \sqrt{h_t(\theta_0)}$, while Caporale and McKiernan (1996) use $g(h_t(\theta_0)) = \ln(h_t(\theta_0))$. As noted by Pagan and Hong (1990) this latter specification is possibly unsatisfactory, since as $h_t(\theta_0) \to 0$ the conditional variance in logs takes very large negative values and the relationship between the conditional variance and Y_t may be overstated. Of course, when λ_0 is restricted to being zero the GARCH-M reduces to the Bollerslev (1986) GARCH model.

The GARCH$(1,1)$-M process will be strictly stationary and covariance stationary if (i) $Z_t \overset{iid}{\sim} \mathcal{N}(0,1)$ and (ii) $\alpha_0 + \beta_0 < 1$. Note, that strict stationarity and ergodicity of the process only require $\mathbf{E}[\ln(\alpha_0 Z_t^2 + \beta_0)] < 1$ which is weaker then the condition implying covariance stationarity (see Arvanitis and Demos, 2004). Specifically, for the parameters of the conditional variance equation we assume that $\omega_0 > 0$, $0 < \alpha_0 < 1$, $0 < \beta_0 < 1$. These restrictions also imply the non-negativity of the conditional variance. General results on the moments and autocorrelation structure of the GARCH(p,q)-M can be found in Karanasos (2001).

Lee and Hansen (1994) and Lumsdaine (1996) derived the distribution theory for the quasi-maximum likelihood estimator in the GARCH$(1,1)$ model. Lumsdaine (1996) established the consistency and asymptotic normality of the quasi-maximum likelihood estimator under the assumption that the re-scaled innovations $\varepsilon_t/\sqrt{h_t}$ are independent and identically distributed (strong GARCH), while Lee and Hansen (1994) derive the same results under the weaker assumption that the re-scaled innovations are strictly stationary and ergodic but not necessarily independent (semi-strong GARCH). To our knowledge sufficient regularity conditions which ensure consistency and asymptotic normality of the quasi-maximum likelihood estimator for the GARCH-M model have not yet been established. As standard in the literature on GARCH-M we will treat our estimates as if the distribution theory for the GARCH estimator could be directly extended. Note, that in contrast to ARMA-GARCH models which do not allow for an in-mean effect, in the GARCH-M model the information matrix is not block diagonal, and thus consistent estimation of the parameters requires that both the conditional mean and variance functions be correctly specified and estimated simultaneously.

Linton and Perron (2003) propose a semiparametric version of the GARCH$(1,1)$-M model described by equations (6.10) – (6.12) in which the functional dependence of Y_t on its conditional variance, $m(h_t)$, is estimated by nonparametric kernel smoothing methods. The estimation procedure is very similar to the one described above, i.e. based on an iterative updating of both the parameters of the conditional variance equation and the function $m(\cdot)$.

For our simulations we adopt two steps from the Linton and Perron (2003) algorithm. First, the initial parameter estimates $(\widehat{\psi}^{(0)}, \widehat{\gamma})$ will be obtained by estimating the parametric specification described in equations (6.10) – (6.12) by quasi-maximum likelihood. Second, in each iteration step the bandwidth for the nonparametric estimate $\widehat{m}_b^{(k)}$ is chosen as $b = b_0\sigma(\widehat{h}_t^{(k-1)})T^{-1/5}$, where $\sigma(\widehat{h}_t^{(k-1)})$ is the standard deviation of the fitted conditional variance from the $(k-1)$-th iteration step and the value of b is determined as the one which produces the lowest value of the cross-validation function

$$CV(b) = \frac{1}{T}\sum_{t=1}^{T}\left(Y_t - \widehat{m}_{b,-t}^{(k)}(\widehat{h}_t^{(k-1)})\right)^2,$$

where $\widehat{m}_{b,-t}^{(k)}$ is the leave-one-out estimator and b_0 is allowed to vary between 0.5 and 2.5 in increments of 0.1. Recall, that in the simulations we choose $I = (0, \infty)$. For different intervals I the cross-validation will be performed using the observations in the interval only.

In the simulations and in our application we will focus on testing for linearity in the GARCH$(1,1)$-M model. Since many properties of the model such as the behavior of the maximum likelihood estimator are largely unexplored we do not verify our assumptions for this specification. However, it is widely believed that the well known properties of the GARCH$(1,1)$ should also hold for the GARCH$(1,1)$-M. Most of the above assumptions can be easily verified for the GARCH$(1,1)$. Assumption 2 is satisfied by e.g. Gaussian Z_t. Carrasco and Chen (2002) show that h_t in the GARCH$(1,1)$ is β-mixing with exponentially decaying mixing coefficients as required in Assumption 3. Assumption 4 is naturally satisfied when m_γ does not depend on h_t and Assumption 5 holds by the results of Lee and Hansen (1994) and Lumsdaine (1996). Finally, Assumption 7 follows directly from the ARCH(∞) representation of h_t. Note, that the proposed test can also be used to test for the existence of an in-mean effect. In this situation the null hypothesis is given by $m_{\gamma_0}(h_t) = \mu_0$. Such a test can be applied in a first step before one tests for particular parametric specifications of the risk premium.

6.3.4 Parametric Bootstrap

We expect that the theorems can only give a rough idea of the stochastic behavior of our test statistic for small sample sizes. Indeed we will see in the simulations that the normal approximation does not work very well in our setting. Therefore, it seems appropriate not to use the asymptotic critical values but to compute the critical values based on resampling (see Härdle and Mammen, 1993).

Suppose one has obtained initial parameter estimates $(\widehat{\psi}^{(0)}, \widehat{\gamma})$ and final estimates of the conditional variance $\widehat{h}_t^{(k)} = \widehat{h}_t(\widehat{\psi}^{(k)}, \widehat{m}_b^{(k)})$ according to the algorithm described in Section 6.3.1. Then one can approximate $\widehat{\Gamma}_T^{(k)}$ by numerical integration. The bootstrap procedure makes use of the fact that under the null hypothesis we have a parametric specification of the conditional mean and variance and can be described as follows:

Step 1: Generate a bootstrap series $\{Y_t^\star\}_{t=1}^T$ according to equations (6.10)–(6.12) with $m_{\widehat{\gamma}}$ given by the null hypothesis. As a starting value h_0 we use the estimated unconditional variance $\widehat{\omega}^{(0)}/(1-\widehat{\alpha}^{(0)}-\widehat{\beta}^{(0)})$. Innovations $Z_t^\star$ are drawn from the standard normal distribution.

Step 2: Apply the algorithm described in Section 6.3.1 to the bootstrap series $\{Y_t^\star\}_{t=1}^T$ and obtain $m_{\widehat{\gamma}^\star}$ and $\widehat{h}_t^{(k)\star}$. Calculate the value of the bootstrap test statistic $\widehat{\Gamma}_T^{(k)\star}$ by numerical integration.

Step 3: Repeat step 1 and 2 for B times. The bootstrap p-value of $\widehat{\Gamma}_T^{(k)}$ is the relative frequency of the event $\{\widehat{\Gamma}_T^{(k)\star} \geq \widehat{\Gamma}_T^{(k)}\}$ in the B bootstrap resamples.

6.4 Monte-Carlo Simulation

In this section we examine the finite sample properties of the semiparametric estimation procedure and the empirical level and power of the proposed test statistic. We first compare the performance of the parametric GARCH$(1,1)$-M with the semiparametric procedure under the null hypothesis and then under the alternative. Thereafter, we estimate the empirical level and power and demonstrate the robustness of our results with respect to the choice of the bandwidth. We always use an Epanechnikov kernel and weight function $w(\cdot) = \mathbf{1}_{[\underline{h},\overline{h}]}$, where $\underline{h}$ and $\overline{h}$ are chosen such that approximately 90% of the data are covered.[5] For simplicity we will denote the fitted conditional variance and the corresponding test statistic from the last iteration step by $\widehat{h}_t$ and $\widehat{\Gamma}_T$ suppressing the index k. The integral of the test statistic $\widehat{\Gamma}_T$ is numerically approximated on 50 equally spaced grid points on the interval $[\underline{h},\overline{h}]$. The parameters of the conditional variance equation are chosen to be $\omega_0 = 0.01$, $\alpha_0 = 0.1$ and $\beta_0 = 0.85$ which represent typical parameter values in empirical applications (see Section 6.5). The innovations are drawn from the standard normal distribution. All the simulations are carried out for a sample

[5] Alternatively, we used a standard normal kernel and obtained virtually identical results.

size of $T = 1000$ which is realistic when we consider that most applications to financial data such as stock or exchange rate returns are on a daily basis or even higher frequency. The Monte-Carlo experiments are repeated $M = 200$ times and the bootstrap resampling is performed $B = 200$ times for each sample. Initial parameter estimates for the mean and variance equation are obtained by quasi-maximum likelihood. The variance parameters are updated by estimating a parametric GARCH$(1,1)$ on the residuals $Y_t - \widehat{m}_b^{(k)}(\widehat{h}_t^{(k-1)})$. In each iteration step we impose the parameter restrictions described in Section 6.3.3 implying covariance stationarity and nonnegativity of the conditional variance. The bandwidth parameter b is chosen in each iteration step according to the cross-validation criterion discussed in Section 6.3.3. Throughout the simulations we set $I = (0, \infty)$.

6.4.1 Performance of the Estimation Procedure

We first evaluate the performance of the estimation procedure for three linear specifications which reflect the null hypothesis:

$$\begin{aligned} &\text{(N1)} \qquad m(h_t) &=& \; 0.05 \cdot h_t \\ &\text{(N2)} \qquad m(h_t) &=& \; 0.5 \cdot h_t \\ &\text{(N3)} \qquad m(h_t) &=& \; h_t. \end{aligned}$$

Table 6.1 presents in Panel A the median estimates for the mean and variance equation parameters of the parametric GARCH$(1,1)$-M and in Panel B the median estimates of the parameters from the conditional variance equation obtained by the semiparametric procedure.[6] In both panels we also provide the 25% and 75% quantiles for the estimated parameters over the 200 replications. The median parametric parameter estimates presented in Panel A of Table 6.1 are – as expected under the null – very close to the true parameter values of the model for the different values of λ_0. In particular, the

[6]Similar results were obtained for the square root and log specification and are available upon request.

in-mean parameter λ_0 is very well estimated as shown by the 25% and 75% quantiles. However, from the estimates of the quantiles it is evident that the true value λ_0 can be recovered much better for higher values of λ_0 than for smaller ones. From Panel B it becomes clear that the semipametric estimator leads to very precise estimates of the conditional variance equation parameters, although it unnecessarily applies the iterating procedure. We find that the semiparametric estimate of α_0 is in all cases slightly lower than its parametric estimate while the converse holds for the estimates of β_0. The ranges between the 25% and 75% quantiles are approximately the same for the semiparametric procedure and the parametric estimator. Figure 6.1 shows the true mean function, the pointwise median of the parametric and the nonparametric estimate along with the pointwise 25% and 75% quantiles of the nonparametric estimate for model N3. Under the null hypothesis both estimation procedures seem to do equally well in recovering the true structure of the model. Similar figures are available for models N1 and N2, but are omitted for space considerations.

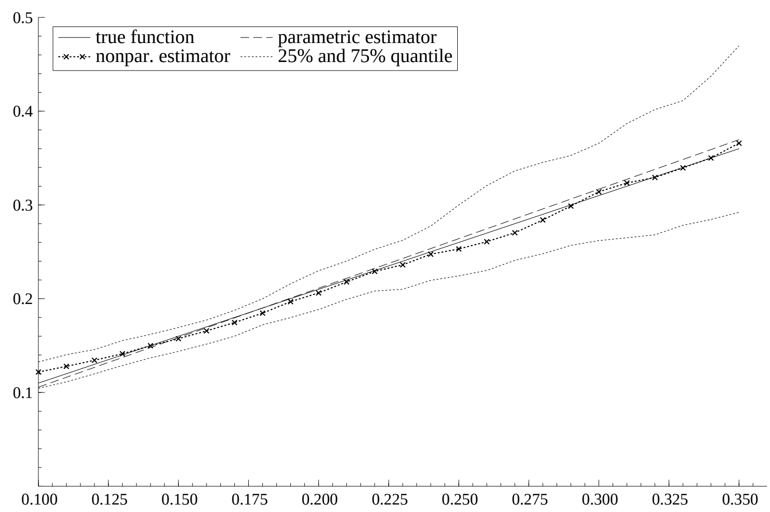

Figure 6.1: Parametric and nonparametric estimate for model N3.

Table 6.1: Monte-Carlo estimates of the parametric and semiparametric regression model.

Panel A: Median parametric estimates					
	$\widehat{\mu}$	$\widehat{\lambda}$	$\widehat{\omega}$	$\widehat{\alpha}$	$\widehat{\beta}$
N1	-0.003	0.050	0.011	0.097	0.841
($\lambda_0 = 0.05$)	(-0.024, 0.019)	(-0.078, 0.184)	(0.009, 0.015)	(0.080, 0.117)	(0.811, 0.868)
N2	-0.005	0.508	0.011	0.098	0.841
($\lambda_0 = 0.5$)	(-0.033,0.017)	(0.386,0.689)	(0.009, 0.015)	(0.082, 0.115)	(0.813, 0.866)
N3	-0.007	1.032	0.010	0.097	0.849
($\lambda_0 = 1$)	(-0.037, 0.024)	(0.879, 1.197)	(0.008, 0.013)	(0.084, 0.109)	(0.824, 0.868)
A1	0.746	-0.809	0.013	0.109	0.825
($\zeta_0 = 0.5$)	(0.682, 0.794)	(-1.094, -0.525)	(0.010, 0.017)	(0.090, 0.122)	(0.798, 0.855)
A2	0.012	0.275	0.011	0.099	0.842
($\zeta_0 = 0.1$)	(-0.022, 0.047)	(0.103, 0.435)	(0.008, 0.014)	(0.086, 0.115)	(0.818, 0.869)
A3	0.077	0.760	0.010	0.096	0.852
($\zeta_0 = 0.12$)	(0.054, 0.102)	(0.627, 0.895)	(0.008, 0.014)	(0.081, 0.109)	(0.823, 0.871)

Notes: Entries in Panel A and B are the median of the estimated parameters over the 200 replications. The entries in the parenthesis are the 25% and 75% quantiles over the 200 replications.

Monte-Carlo estimates of the parametric and semiparametric regression model (continued).

	Panel B: Median semiparametric estimates		
	$\widehat{\omega}$	$\widehat{\alpha}$	$\widehat{\beta}$
N1	0.0102	0.0916	0.8505
($\lambda_0 = 0.05$)	(0.0082, 0.0138)	(0.0768, 0.1100)	(0.8242, 0.8724)
N2	0.0101	0.0911	0.8507
($\lambda_0 = 0.5$)	(0.0082,0.0141)	(0.0765,0.1103)	(0.8251, 0.8735)
N3	0.0101	0.0923	0.8554
($\lambda_0 = 1$)	(0.0076, 0.0131)	(0.0782, 0.1074)	(0.8318, 0.8777)
A1	0.0102	0.0913	0.8541
($\zeta_0 = 0.5$)	(0.0077, 0.0131)	(0.0793, 0.1066)	(0.8323, 0.8762)
A2	0.0101	0.0925	0.8551
($\zeta_0 = 0.1$)	(0.0077, 0.0128)	(0.0784, 0.1061)	(0.8330, 0.8777)
A3	0.0101	0.0910	0.858
($\zeta_0 = 0.12$)	(0.0078, 0.0134)	(0.0774, 0.1024)	(0.8320, 0.8778)

Notes: As in Table 6.1.

Next, we investigate the accuracy of the iterative estimation algorithm under the alternative. We use the following mean functions:

$$
\begin{aligned}
&(\text{A1}) \qquad m(h_t) &=& \; h_t + \zeta_0 \cdot \sin(10 \cdot h_t) \\
&(\text{A2}) \qquad m(h_t) &=& \; 0.5 \cdot h_t + \zeta_0 \cdot \sin(0.5 + 20 \cdot h_t) \\
&(\text{A3}) \qquad m(h_t) &=& \; h_t + \zeta_0 \cdot \sin(3 + 30 \cdot h_t).
\end{aligned}
$$

These alternatives represent shapes of the risk premium which are not covered by the standard specification but can be viewed as motivated by the results on Backus and Gregory (1993), Genotte and Marsh (1993) and the empirical findings of Linton and Perron (2003). Alternative A1 and A2 are inverse U-shaped and U-shaped. A3 stands for a hump-shaped alternative. The parameter ζ_0 can be regarded as a measure for the distance between the linear null hypothesis and the alternative.

Table 6.1 also presents the results for the Monte-Carlo simulations performed for models A1 – A3 with specific values for ζ_0. Again, Panel A reports the mean and variance parameter estimates from the parametric GARCH(1, 1)-M with $m(h_t) = \mu + \lambda h_t$ while Panel B reports the estimates for the conditional variance equation obtained by the semiparametric procedure.[7] Figures 6.2 and 6.3 show the pointwise median parametric and nonparametric estimate along with the 25% and 75% pointwise quantiles of the latter and the true mean function for alternatives A1 and A3. Additionally, we plot the pointwise median estimate of the semiparametric procedure obtained if one does not iterate until convergence but stops after the first iteration step. The figures reveal that the nonparametric estimate of the mean function does again perform very well in uncovering the true mean function. The parametric estimate – which is restricted to being linear – fails to do so. In particular, in model A1 the mean function is inverse U-shaped. For values of the conditional variance up to 0.175 the mean function is increasing while it is decreasing from 0.175 onwards. The parametric estimate of the mean function either over or underestimates the true risk premium. This example shows that one can easily find a negative relationship by applying the parametric model to a non-linear risk premium. A curve similar to A1 is presented by Whitelaw (2000, Figure 3) as a reasonable relationship between the expected return and its volatility in his two regime model when the economy is in a contractionary regime. Merely, the application of the semiparametric procedure makes it possible to obtain the true relationship, i.e. the risk premium is increasing until volatility exceeds a critical value, and then it becomes decreasing. A similar interpretation holds for A2.[8] Finally, A3 is a hump-shaped alternative as suggested by the findings of Linton and Perron (2003). Although, the parametric model captures the overall increasing tendency, it

[7]Again, we do not report the results for the models with $g(h_t) = \sqrt{h_t}$ or $g(h_t) = \ln(g_t)$. These are very similar and available upon request.

[8]The corresponding figure is omitted for reasons of brevity.

would predict very misleading values for the risk premium. The nonparametric fit on the other hand follows closely the true risk premium. These examples clearly illustrate the superiority of the semiparametric approach. Moreover, it is possible to construct non-monotonic shapes of the risk premium which lead to insignificant estimates of the parameter λ_0 and hence would suggest that there is no relationship between h_t and Y_t, while the semiparametric procedure recovers the true relationship. This failure of the parametric estimator may explain the finding of an insignificant $\widehat{\lambda}$ in many studies using the parametric GARCH$(1,1)$-M specification. These graphical intuitions are supported by the estimation results reported in Table 6.1. It is clear that now – as the parametric model is misspecified – the estimates of λ_0 are completely misleading. Nevertheless, the parameters in the conditional variance equation are still surprisingly well estimated using the parametric model. From the simulation it is clear that a misspecified mean function does not necessarily distort the estimates of the parameters of the conditional variance equation. Finally, the semiparametric estimation procedure results in very accurate estimates of the conditional variance parameters ω_0, α_0 and β_0.

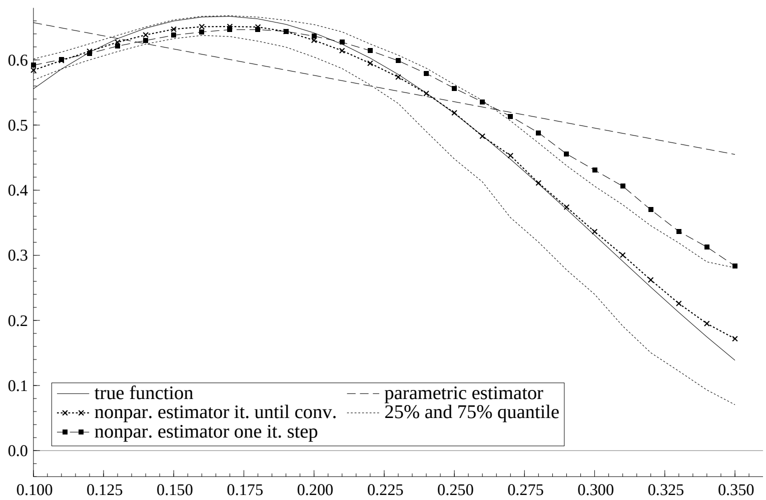

Figure 6.2: Parametric and nonparametric estimate for model A1 ($\zeta_0 = 0.5$).

Figures 6.2 and 6.3 also help to illustrate the gains that are obtained by iterating in the semiparametric estimation procedure. It is evident that the one step iteration

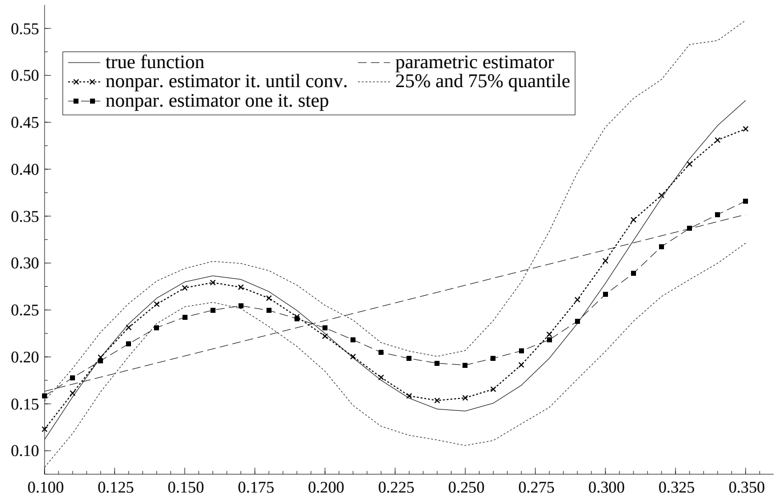

Figure 6.3: Parametric and nonparametric estimate for model A3 ($\zeta_0 = 0.12$).

estimator cannot capture the nonlinearities by the same degree of accuracy as the iteration until convergence estimator. While this seems to be the case for A1 only for large values of h_t, it is generally true for A3 where the one step iteration estimator simply leads to a regression function which is too smooth. It seems that by doing only one iteration step it is not possible to move far enough away from the parametric estimate to be close to the true mean function. This requires further iterations. Such differences can be evaluated by comparing the values of the estimated mean and median integrated squared errors ($MeanISE$ and $MedISE$) over the M Monte-Carlo simulations for the different estimation procedures.[9] We compare the parametric, the one step iteration and the full iteration semiparametric estimate. Since the main focus of the analysis is not on estimation but on the performance of the test statistic we just report exemplary

[9]The ISE for each simulation can be calculated as

$$ISE = \int (\widehat{m}_b(x) - m(x))^2 w(x) dx.$$

Estimates of the $MeanISE$ and $MedISE$ are then the mean and median of the ISE over the 200 Monte-Carlo replications.

the numbers for A1. As expected, we find (numbers are multiplied by 100)

$$\begin{aligned} MeanISE(p.) = 1.233 \quad > \quad & MeanISE(sp.\ 1\ it.) = 0.777 > \\ & \qquad MeanISE(sp.\ it.\ until.\ conv.) = 0.431 \\ MedISE(p.) = 1.148 \quad > \quad & MedISE(sp.\ 1\ it.) = 0.586 > \\ & \qquad MedISE(sp.\ it.\ until.\ conv.) = 0.248 \end{aligned}$$

(with $p.$ = parametric, $sp.\ 1\ it.$ = semiparametric with one iteration step and $sp.\ it.\ until.\ conv.$ = semiparametric with full iteration). Clearly, the one step iteration semiparametric estimate is superior to the parametric estimate, but still the full iteration semiparametric estimate is more much precise in terms of both $MeanISE$ and $MedISE$. We will see in the next subsection that this directly effects the power properties of our test statistic.

6.4.2 Monte-Carlo Estimates of Level and Power

This subsection evaluates the performance of the test statistic. In Table 6.2 we check for models N1, N2 and N3 and for different choices of the bandwidth parameter b whether the estimated level of the test reflects the nominal level. We report the estimated levels in comparison to the nominal 5% and 10% levels. In general, the estimated levels are very stable around the nominal levels of 5% and 10% for a wide range of bandwidth. The lowest bandwidth $b = 0.015$ produces too conservative results, i.e. we observe underrejection. A bandwidth of $b = 0.02$ produces estimates of the level which are in most cases slightly below 5% and 10% respectively, while a bandwidth of $b = 0.045$ leads to estimates slightly above 5% and 10%. Overall, the the bootstrap procedure seems to do a very good job in estimating the 5% and 10% levels close to the nominal ones. The optimal bandwidth as chosen by cross-validation in the last iteration step of the semiparametric procedure is in the neighborhood of $b = 0.02$. Figure 6.4 provides a quantile plot of the test statistic for model N3. It is evident that the test statistic is not normally distributed and therefore one should not rely on

Table 6.2: Monte-Carlo estimates of the level.

	b	0.015	0.020	0.025	0.030	0.035	0.040	0.045
N1	5%	0.030	0.050	0.055	0.060	0.055	0.050	0.055
	10%	0.075	0.070	0.095	0.100	0.105	0.110	0.105
N2	5%	0.025	0.045	0.045	0.050	0.050	0.060	0.070
	10%	0.080	0.090	0.090	0.105	0.110	0.115	0.105
N3	5%	0.025	0.045	0.040	0.040	0.060	0.060	0.070
	10%	0.065	0.080	0.075	0.085	0.085	0.095	0.100

Notes: Entries are rejection rates over the 200 replications at the 5% and 10% nominal level.

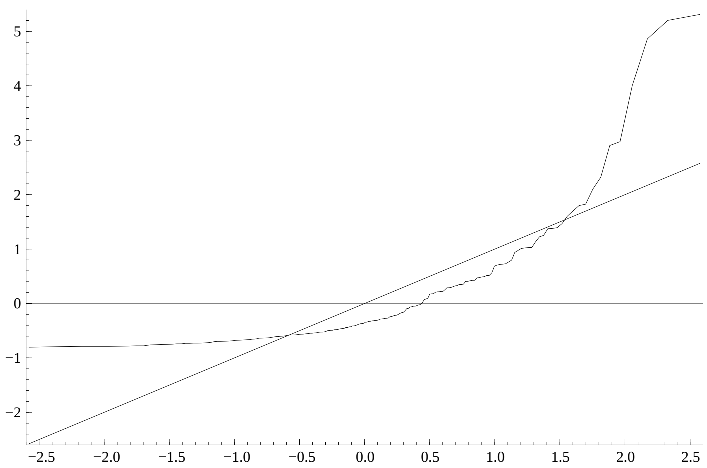

Figure 6.4: Quantile plot of the distribution of the test statistic for model N3 against normal distribution.

the asymptotic critical values. For model N3 we plot the density of $T\sqrt{b}\,\widehat{\Gamma}_T$ and six bootstrap approximations in Figure 6.5 (upper). The figure shows that the bootstrap approximations estimate the distribution of $T\sqrt{b}\,\widehat{\Gamma}_T$ very well when the underlying model reflects the null hypothesis. Figure 6.5 (lower) shows the simulated density of $T\sqrt{b}\,\widehat{\Gamma}_T$ and six bootstrap replications for model A1. Under A1 the simulated density of $T\sqrt{b}\,\widehat{\Gamma}_T$ and the six bootstrap densities are very different, suggesting the test statistic

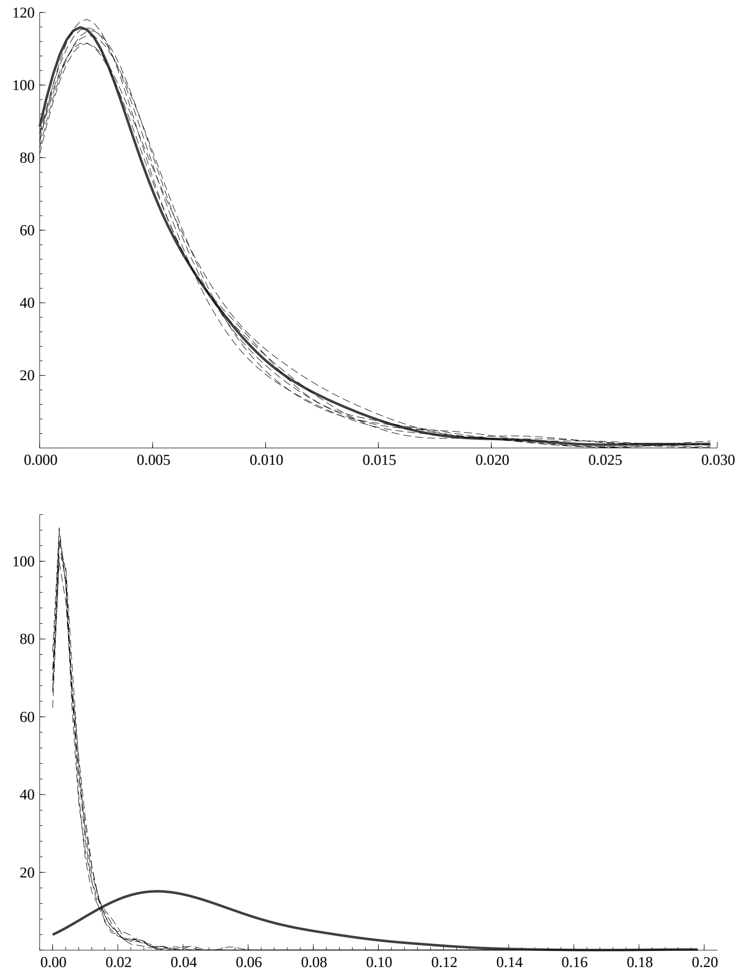

Figure 6.5: Simulated density of test statistic (solid) and six bootstrap approximations (dashed) for model N3 (upper) and A1 (lower).

may have good power properties. Figures 6.6 – 6.8 display the empirical power of the test for alternatives A1, A2 and A3 and three choices of bandwidths. The mean functions under the alternative are constructed such that the models move further away from the null hypothesis as ζ_0 increases. For all three alternatives we find the desired property that the power is monotonically increasing in the value of ζ_0. Moreover, the power is very similar across the three choices for the bandwidth parameter. The overall performance of the test applied under the alternative is very satisfactory. We conclude that the bootstrap procedure works well in our setting.

We also examined the power properties of the one step iteration estimator in comparison to the fully iterated estimator. For all three alternatives the tests based on the full iteration estimator lead to higher power than the corresponding test statistics based on the one step estimator. For instance, for A1 the fully iterated estimator produces empirical powers at the 5% and 10% nominal level of $(0.615, 0.750)$, $(0.875, 0.950)$ and $(0.945, 0.975)$ for $\zeta_0 \in \{0.3, 0.5, 0.7\}$, respectively. The corresponding figures for the one step estimator are $(0.400, 0.595)$, $(0.750, 0.870)$ and $(0.890, 0.945)$. In the light of Figures 6.2 and 6.3 this is not surprising, since the one step estimator is almost everywhere closer to the parametric estimator than the full iteration estimator. Interestingly, the difference in the power decreases with the alternative moving away from the null hypothesis. For alternatives lying sufficiently far away from the null both test statistics reject in approximately the same number of times. Nevertheless, the full iteration estimator approximates the true model much closer as we have seen in the last subsection.

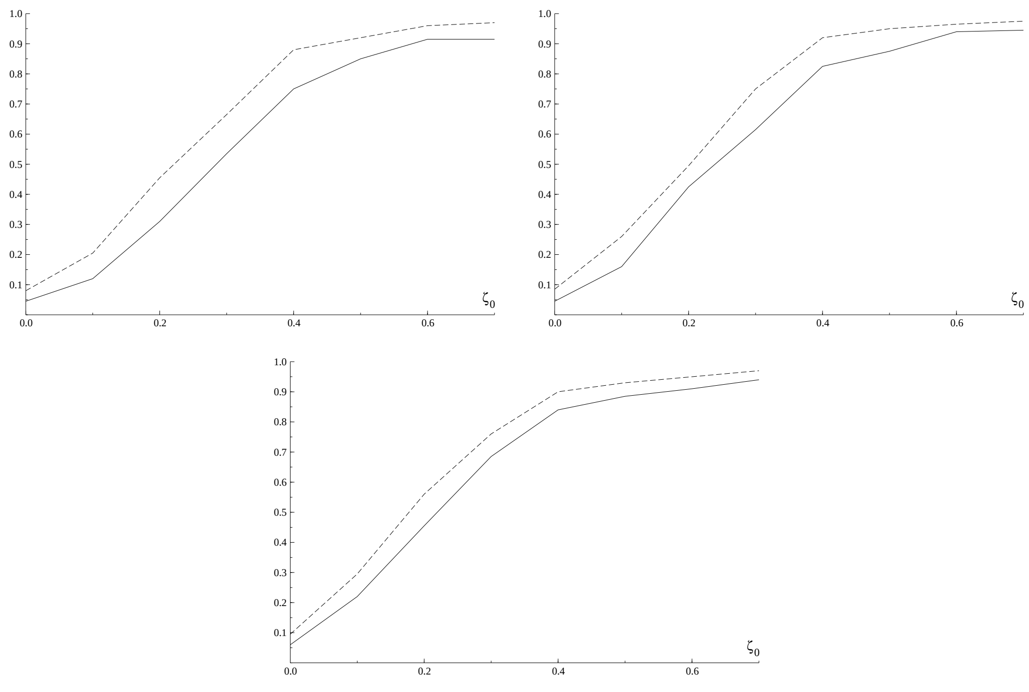

Figure 6.6: Simulated power for model A1 and $b = 0.02$ (upper left), $b = 0.03$ (upper right) and $b = 0.04$ (lower middle). Levels are given by 5% (solid) and 10% (dashed).

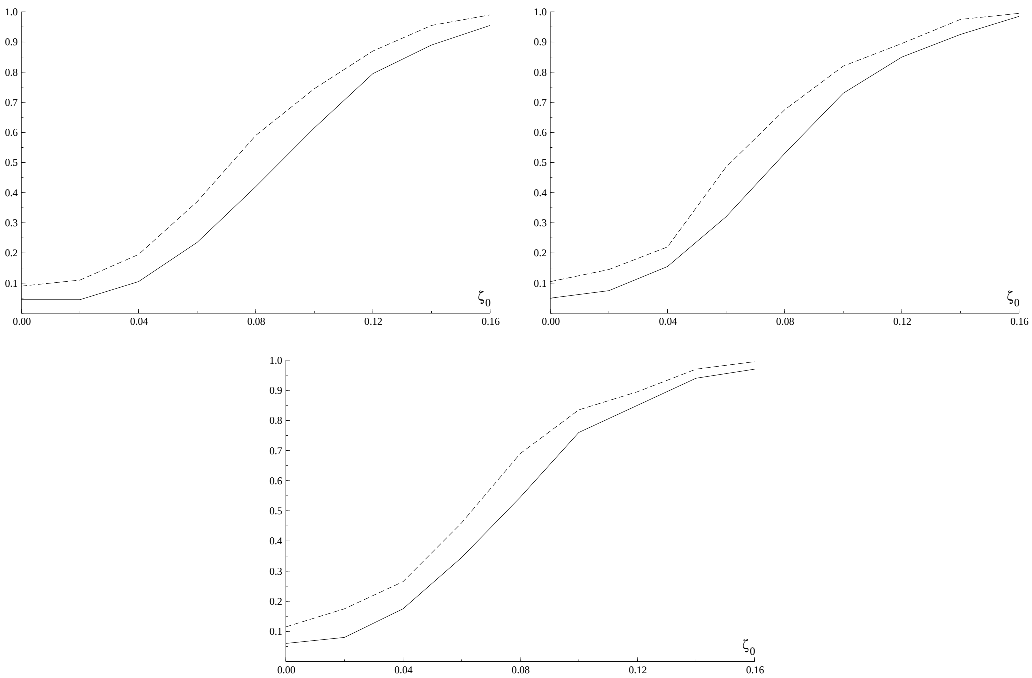

Figure 6.7: Simulated power for model A2 and $b = 0.02$ (upper left), $b = 0.03$ (upper right) and $b = 0.04$ (lower middle). Levels are given by 5% (solid) and 10% (dashed).

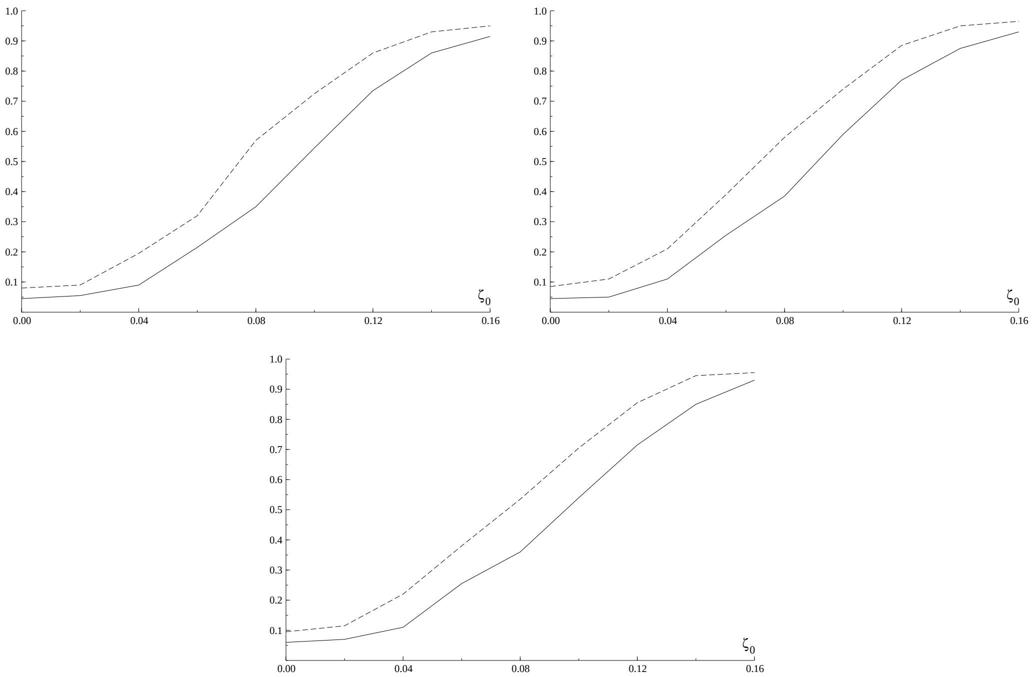

Figure 6.8: Simulated power for model A3 and $b = 0.02$ (upper left), $b = 0.03$ (upper right) and $b = 0.04$ (lower middle). Levels are given by 5% (solid) and 10% (dashed).

6.5 Application: The Shape of the Risk Premium

6.5.1 Data

The usefulness of the specification test will now be assessed in an application to test for linearity in the risk-return relation. For this we employ monthly and daily excess return data on the CRSP value-weighted index, which includes the NYSE, the AMEX and the NASDAQ and can be considered as the best available proxy for the market. Monthly excess returns (including dividends) are calculated as the continuously compounded return on the CRSP minus the yield on a one month Treasury bill (from Ibbotson Associates), $Y_t = r_{M,t} - r_{f,t}$. Daily excess returns are calculated analogously, whereby daily yields are calculated by dividing the monthly yield by the number of trading days in the month and, hence, assuming constant yields for each calender day. The monthly data ranges from January 1926 to December 2001 and was provided by Linton and Perron (2003) who used the same data set for their analysis. Daily return data was obtained from the Kenneth R. French data library for the period July 1963 to July 2005.[10]

The first part of Table 6.3 contains the descriptive statistics for the monthly excess returns on the CRSP. Apart from the statistics for the full sample we also present descriptive statistics for a subsample ranging from 1963:07 to 2001:12. The two samples are labelled as I and II. Sample I corresponds to the period analyzed by Linton and Perron (2003). The average monthly excess return for the two samples is about 0.5% and 0.37% respectively. The distribution of excess returns is negatively skewed and there is evidence for excess kurtosis for both periods reflecting the well known fact that extreme returns occur more often in the market than predicted by the normal distribution. The largest negative return in sample I was realized on September 1931 with -34.26% while the largest positive return was realized in April 1933 with +32.31%. The most extreme excess returns in sample II were realized in October 1987 with -

[10] The data can be downloaded from (last access 11.06.2006):
http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_factors.html

26.08% and in October 1974 with +14.80%. The higher kurtosis in sample I compared to the kurtosis in sample II can be explained by the extreme returns realized during the Great Depression of the early 1930s. The 12-th and 24-th order Ljung-Box statistics in combination with the results of the Engle LM-test for ARCH effects (both not reported) indicate serial correlation in the squared return series and highlight the importance of an appropriate modelling of the conditional variance of the excess returns.

We also investigate two daily excess return series of the CRSP in order to see whether there is any difference between the analysis of daily and monthly data. First, as argued by Andersen and Bollerslev (1998) more precise estimates of conditional volatility may be obtained by employing daily data in comparison with monthly data, and thus a better estimate of the true risk-return relation. Second, as shown by Scruggs (1998) and Guo and Whitelaw (2006) a hedge demand which is not included as an explanatory variable can lead to an omitted variable bias in estimating the risk-return relation. However, since Guo and Whitelaw (2006) find that the investment opportunities change slowly at the business cycle frequency, these changes can be regarded as approximately constant at a daily frequency. Thus, the risk-return relation can be precisely estimated at a daily frequency without explicitly incorporating the hedge demand in the regression equation. The second part of Table 6.3 presents the descriptive statistics for the daily excess return data. Again, we consider two samples. The first one corresponds to the complete sample of daily observations. Sample IV ranges from January 1990 to July 2005. The average daily excess return lies between 0.02% and 0.03%. As for the monthly data, we observe negative skewness and excess kurtosis for both samples. The extreme returns in period III are again realized during October 1987. Moreover, Ljung-Box statistics and Engle LM-tests for ARCH effects (both not reported) point to strong autocorrelation in the squared excess returns.

Table 6.3: Descriptive statistics for monthly and daily CRSP excess returns ($\times 100$).

		Mean	St. dev.	S	K	max	min	T
	monthly data							
I	1926:01 - 2001:12	0.499	5.506	-0.504	9.812	32.307	-34.262	912
II	1963:07 - 2001:12	0.368	4.495	-0.777	6.075	14.797	-26.077	462
	daily data							
III	01.07.1963 - 01.07.2005	0.022	0.890	-0.753	21.160	8.630	-17.160	10593
IV	01.01.1990 - 01.07.2005	0.029	0.986	-0.120	7.045	5.310	-6.650	3929

Notes: The table reports the average excess return over the four periods, and its standard deviation. S and K are the estimated skewness and kurtosis respectively. Max and min refer to the most extreme returns realized in the respective period. T is the sample size.

6.5.2 Parametric GARCH(1, 1)-M Estimates

Next, we estimate parametric GARCH(1, 1)-M models with $m(h_t) = \mu + \lambda h_t$ for the four periods. In all the regressions we include the constant μ to account for market imperfections such as taxes or transaction costs (see Scruggs, 1998). Lanne and Saikkonen (2006) claim that the unnecessary inclusion of the constant term lowers the power properties of tests of the risk-return relation and hence can be responsible for the widely documented controversial results. On the other hand, we find that the excess market return is positive on average and of course all the conditional variances are positive by construction. Therefore, when the constant is omitted it is not surprising if one finds a positive and significant slope parameter.

Parameter estimates are provided in Table 6.4. For periods I and II the GARCH parameter estimates $\widehat{\alpha}$ and $\widehat{\beta}$ are highly significant. The estimated values for α and β are around 0.1 and 0.85 respectively, satisfying the condition for covariance stationarity and implying a high degree of persistence in the conditional variance ($\widehat{\alpha}+\widehat{\beta} = 0.981$ for period I and $\widehat{\alpha}+\widehat{\beta} = 0.949$ for period II). Poterba and Summers (1986) show that only persistent increases in volatility will effect the discount factors applied to future cash flows and thereby current prices. Therefore, they argue that persistence in the volatility is a necessary condition for fluctuations in volatility to have a significant impact on explaining risk premia. Similarly, Baekert and Wu (2000, p. 2) reason that the predicted positive effect of volatility on excess returns relies "first of all on the fact that volatility is persistent". The estimates for λ are positive but insignificant which is in line with the previous literature when the conditional variance was modelled as a GARCH process and monthly data was used. The value estimated for period I is considerably lower than the value estimated for period II. This can be explained by the fact that period I includes the Great Depression which was characterized by extremely high conditional variances associated with large negative returns indicating a temporary distortion of the "normal" risk-return relation. We also estimated a parametric GARCH-M separately for a period including the Great Depression (1926:01 - 1949:12) and obtained an insignificant in-mean parameter $\widehat{\lambda} = -0.193$. This finding is line with Whitelaw's (2000) model which predicts a negative relationship between expected excess returns and the volatility of returns when the economy is in a contractionary regime. A visual investigation of the fitted conditional variance series for period I reveals that the estimated conditional variances from 1950 onwards are dwarfed compared to the estimated conditional variances of the period of the Great Depression. This can be interpreted as a change in the volatility regime and so it is questionable whether one should estimate one single GARCH equation to the full sample of monthly observations. Moreover, as argued by Poterba and Summers (1986) one should be concerned with the fact that the risk-return relation during the Great Depression with its exceptionally high volatility does not provide a useful guide to the current beliefs of market participants. Therefore, we focus our attention on period II and the daily data.

For periods III and IV of daily data, the estimates of α and β are again highly significant and imply an even higher degree of persistence ($\widehat{\alpha} + \widehat{\beta} = 0.995$ for both periods). In sharp contrast to the monthly data, we estimate positive and highly significant in-mean effects. The estimate of λ is significant at the 1% level in sample III and at the 5% level in sample IV. Note, that for the two periods of daily data we find estimates for λ being similar to the estimate of period II of monthly data. This is reasonable since both the risk premium and the conditional variance should be approximately proportional to the length of the measurement interval. For instance, period II and III cover approximately the same period of monthly and daily data with estimates $\widehat{\lambda} = 3.870$ and $\widehat{\lambda} = 3.844$ respectively. If as argued in Guo and Whitelaw (2006) the omitted hedge term does not effect the estimation of the risk-return relation for daily data, the finding of similar $\widehat{\lambda}$'s for monthly and daily data suggests that the omitted variable bias argument of Scruggs (1998) does also not hold at a monthly frequency, because in the presence of such an effect the estimate of λ based on monthly data should be considerably different from the one on daily data. Therefore, the argument by Andersen and Bollerslev (1998) seems to apply, the estimates based on daily data provide a more accurate measure of the conditional volatility and hence allow for a more precise estimation of the risk-return

Table 6.4: GARCH-M estimates for CRSP data.

	daily data		monthly data	
	I	*II*	*III*	*IV*
$\widehat{\mu}$	0.005	-0.003	0.00026	0.00023
	(2.461)	(-0.579)	(2.976)	(1.256)
$\widehat{\lambda}$	0.576	3.870	3.844	4.827
	(0.638)	(1.130)	(2.714)	(2.005)
$\widehat{\omega}$	0.0001	0.0001	$6.66 \cdot 10^{-7}$	$7.94 \cdot 10^{-7}$
	(1.99)	(1.798)	(3.938)	(1.912)
$\widehat{\alpha}$	0.115	0.074	0.089	0.076
	(4.149)	(3.105)	(8.187)	(6.633)
$\widehat{\beta}$	0.866	0.875	0.906	0.917
	(26.295)	(16.561)	(103.919)	(93.605)
Q^2_{12}	4.99	4.61	12.76	10.05
	[0.96]	[0.97]	[0.39]	[0.61]

Notes: Bollerslev and Wooldridge (1992) robust t-statistics are reported in parenthesis $(\cdot)$. Q^2_{12} are the Ljung-Box statistics at the 12-th lag for the squared standardized residuals. Numbers in brackets $[\cdot]$ are p-values.

relation. As a result of this we find a significant in-mean effect using the daily data. Following French et al. (1987) $\widehat{\lambda}$ can be interpreted as an estimate for the parameter of relative risk aversion. The values of $\widehat{\lambda}$ across the four periods are plausible for the coefficient of relative risk aversion. We conclude that the parametric GARCH$(1,1)$-M models deliver convincing evidence for a positive and (partly) significant relation between risk and excess returns.

According to the Ljung-Box statistics the null hypothesis of uncorrelated squared standardized residuals is accepted for all four models. Finally, the GARCH$(1,1)$-M models were preferred by the AIC and BIC information criteria to models of higher order. Since we were concerned that the extreme movements during October 1987 may distort our inferences we reestimated the models for period II and III and omitted this month. We found no qualitative changes in our results.

6.5.3 Testing the Linear Hypothesis

Next, we will apply our specification test to the CRSP excess return data to check whether the functional relationship between excess returns and risk can be confirmed to be linear as assumed by the parametric GARCH$(1,1)$-M. Recall from Section 6.2 that Linton and Perron (2003) using a semiparametric EGARCH-M model found support for a hump-shaped pattern of the risk premium.

The application of the test procedure requires the choice of an appropriate bandwidth b and of an interval $[\underline{h}, \overline{h}]$ on which the test statistic is evaluated.[11] For the four periods we evaluate the test statistic on two different intervals. The larger one is chosen such that it covers 90% of the data, the smaller one covers only 70%. In both situations $\underline{h}$ corresponds to the 5% quantile ($q_{0.05}(\widehat{h}_t)$) of the distribution of the estimated conditional variances from the last iteration step. Accordingly, we choose $\overline{h}$ approximately as the 75% or 95% quantile ($q_{0.75}(\widehat{h}_t)$ and $q_{0.95}(\widehat{h}_t)$). As a guide for choosing the bandwidth we use $b = \sigma(\widehat{h}_t) \cdot T^{-1/5}$, where $\sigma(\widehat{h}_t)$ and T refer only to the observations in $[\underline{h}, \overline{h}]$. This choice of the bandwidth usually results in values slightly above the cross-validated bandwidth from the last iteration step. Since Theorem 6.2 requires oversmoothing in comparison to the optimal bandwidth for estimation, we additionally report the test statistic and the corresponding p-values for two larger choices of b, whereby the largest bandwidth is always based on the full distribution of $\widehat{h}_t$.

The test results for periods I to IV are given in Table 6.5. We begin by discussing the results from periods I and II. Several interesting findings emerge. As can be seen

[11] As in the simulation section, we will denote the fitted conditional variance and the corresponding test statistic from the last iteration step by $\widehat{h}_t$ and $\widehat{\Gamma}_T$ suppressing the index k.

from the table we cannot reject the null hypothesis that the risk premium is linear in the conditional variance for all the periods and intervals at the 5% level. Besides the estimated 95% quantile of the fitted conditional variances $q_{0.95}(\widehat{h}_t)$, we report the median of the 95% quantiles of the fitted conditional variances over the 200 bootstrap replications denoted by $q_{0.95}(h_t^\star)$. For period II we observe that $q_{0.95}(\widehat{h}_t)$ and $q_{0.95}(h_t^\star)$ are very close to each other reflecting the fact that the fitted conditional variances from the bootstrap procedure mimic very well the distribution of the fitted conditional variances from the observed data. For period I this is not the case. The 95% quantile from the regression fit is much larger than the median 95% quantile generated by the bootstrap. Since period I includes the Great Depression, the estimates of the conditional variance are severely higher than the corresponding estimates of period II. The parametric GARCH-M simulated under the null in the bootstrap replications cannot generate the high volatilities fitted for the observed data. Again, this questions the appropriateness of fitting one GARCH-M to the whole sample of monthly observations. Accordingly, we evaluated the test statistic only on the interval which is covered by the bootstrap procedure. Applying the test to a wider interval will always lead to acceptance of the null, since we would evaluate the test in an area where we have only a few bootstrap observations leading to high values of the bootstrap test statistic.

Figure 6.9 shows the parametric and nonparametric estimate of the risk premium for period II. The shape of the nonparametric estimate shows some non-linearity which could be called hump-shaped as in Linton and Perron (2003). Nevertheless, the non-parametric estimate trends very closely with the linear parametric estimate making the test result plausible. Pointwise 95% asymptotic standard errors for the nonparametric estimate are given by

$$\widehat{m}_b^{(k)}(x) \pm 1.96 \cdot \sqrt{\frac{1}{Tb} \frac{x \int K(u)^2 du}{\widehat{f}_h(x)}},$$

see Linton and Perron (2003). In the above expression we use the fact that for the GARCH-M model it holds that $\mathbf{Var}(Y_t | h_t = x) = x$.

The test results for the daily data are provided in the lower part of Table 6.5. For

Table 6.5: Testing for linearity in the risk-return relation.

	I				*II*			
	$q_{0.95}(\widehat{h}_t) = 78.67,\ q_{0.95}(h^\star_t) = 56.77$				$q_{0.95}(\widehat{h}_t) = 29.41,\ q_{0.95}(h^\star_t) = 27.76$			
b	1.74	2.6	5	7.4	0.99	1.25	1.50	1.71
	$[\underline{h}, \overline{h}] = [10, 40]$				$[\underline{h}, \overline{h}] = [12, 25]$			
$T\sqrt{b}\,\widehat{\Gamma}_T$	4.766	4.192	4.144	3.400	0.984	0.787	0.662	0.596
p-value	0.670	0.520	0.250	0.145	0.610	0.545	0.481	0.422
	$[\underline{h}, \overline{h}] = [10, 60]$				$[\underline{h}, \overline{h}] = [12, 30]$			
$T\sqrt{b}\,\widehat{\Gamma}_T$	75.919	53.743	37.576	46.603	1.922	1.561	1.278	1.096
p-value	0.360	0.370	0.260	0.095	0.797	0.754	0.711	0.690
	III				*IV*			
	$q_{0.95}(\widehat{h}_t) = 2.34,\ q_{0.95}(h^\star_t) = 2.38$				$q_{0.95}(\widehat{h}_t) = 2.07,\ q_{0.95}(h^\star_t) = 1.78$			
b	0.05	0.07	0.10	0.14	0.04	0.08	0.13	0.18
	$[\underline{h}, \overline{h}] = [0.2, 1.5]$				$[\underline{h}, \overline{h}] = [0.2, 1.0]$			
$T\sqrt{b}\,\widehat{\Gamma}_T$	8.062	5.913	4.393	4.403	1.467	9.45	9.009	9.608
p-value	0.086	0.136	0.161	0.100	0.118	0.112	0.107	0.020
	$[\underline{h}, \overline{h}] = [0.2, 2.34]$				$[\underline{h}, \overline{h}] = [0.2, 2.07]$			
$T\sqrt{b}\,\widehat{\Gamma}_T$	62.303	56.079	48.484	39.459	33.650	18.860	15.906	8.580
p-value	0.015	0.025	0.075	0.075	0.132	0.162	0.173	0.208

Notes: The smallest bandwidth always corresponds to the smaller interval, while the second smallest bandwidth is chosen according to the larger interval. The two largest bandwidths can be regarded as oversmoothing.

both periods we find that the 95% quantiles of the fitted and bootstrap conditional variances are close to each other. The p-values are now considerably smaller than for the monthly data but still we cannot reject the null of linearity in most of the cases at the 5% level. For period III we reject linearity for some of the bandwidths in the larger interval. The linear specification is accepted for period IV for both intervals with only

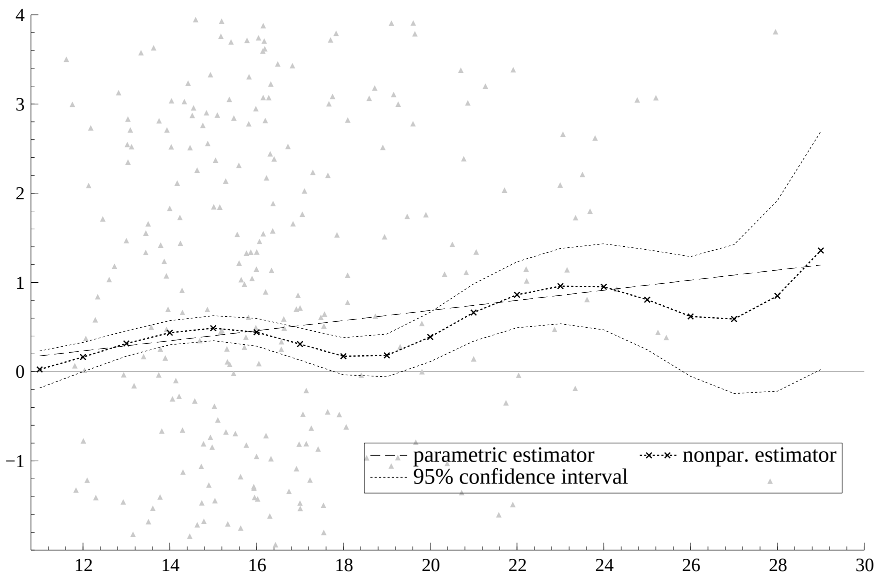

Figure 6.9: Parametric and nonparametric fit for period II.

one exception.

To check for the robustness of our results we also tested the hypothesis of no in-mean effect, i.e. $H_0 : m_\gamma(h_t) = \mu$. This hypothesis was rejected in the overwhelming majority of cases. In summary, we find that there is evidence for an in-mean effect and we cannot reject the hypothesis of the effect being linear.

6.6 Extensions

We apply our test procedure to the standard version of the GARCH-M. More flexible formulations may be required. In the following we provide some motivating examples.

A natural extension of equation (6.10) is to allow for additional explanatory variables. For instance, in the general version of the ICAPM the excess market return is not only explained by its conditional variance but also by state variables. One could assume that these state variables enter linearly in the mean function and as before only $m(h_t)$ is estimated nonparametrically. Another possibility would be to assume that equation (6.10) is an additive function of the conditional variance and the state variables, and each component function could be estimated nonparametrically by backfitting methods as suggested by Mammen, Linton and Nielson (1999).

The GARCH-M model is often applied to variables which require the usage of autoregressive structures in the mean equation. Two popular examples are the relationship between (i) inflation and nominal uncertainty and (ii) output growth and output uncertainty (see for instance Grier and Perry, 2000, Kontonikas, 2004, and Caporale and McKiernan, 1996). Usually, AR-GARCH-M models are employed to test for the effects of inflation uncertainty and output growth uncertainty on average inflation and output growth. Therefore, an appropriate modelling of the mean equation should incorporate lagged values of inflation and output growth. Moreover not only the contemporaneous conditional variance should be allowed to have an effect on inflation and output growth, but also lagged values of the conditional variance. This requires the estimation of a mean equation which includes autoregressive terms as well as several lagged conditional variances.

Of course, more flexible specifications of the conditional variance should be allowed for, such as, the EGARCH or the FI(E)GARCH which capture leverage effects and/or long memory in the conditional variance. Even more generally one could combine the approach presented in this chapter with a nonparametric modelling of the conditional variance as suggested by Linton and Mammen (2005). Such a model would be a natural extension of Masry and Tjøstheim (1995) allowing for an in-mean effect.

Finally, the assumption of normally distributed innovations Z_t should be relaxed, e.g. a more flexible distribution such as the generalized error distribution could be used (see Linton and Perron, 2003).

6.7 Conclusions

We suggest a specification test for a class of parametric GARCH-M models. This class of models is heavily used in the analysis of the risk-return relationship as well as to investigate the causal relationship between the level and the uncertainty of macroeconomic variables such as inflation and output growth. The parametric functional form of the risk premium imposed by the GARCH-M is mainly motivated by the ICAPM or imposed simply for convenience. We suggest a test statistic which compares the fit from a parametric specification of the risk premium with a nonparametric fit obtained by an iterative estimation algorithm. The asymptotic distribution of the test statistic is shown to be normal and independent of the number of iterations used. The critical values can be obtained via a bootstrap algorithm and Monte-Carlo simulations provide convincing evidence that the test works reasonably well in finite samples. Finally, we apply the test procedure to daily as well as monthly return data on the CRSP. Our results suggest that the linear specification for the risk premium is in line with the data and thus we find support in favor of the prediction made by the ICAPM.

6.8 Appendix

Proof of Theorem 6.1.

The test statistic has the following representation: $\widehat{\Gamma}_T^{(0)} = \widehat{\Gamma}_{T,1}^{(0)} + \widehat{\Gamma}_{T,2}^{(0)} + \widehat{\Gamma}_{T,3}^{(0)}$, where

$$\begin{aligned}
\widehat{\Gamma}_{T,1}^{(0)} &= \int \left\{ \frac{\frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t^{(0)} - x)\varepsilon_t}{\frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t^{(0)} - x)} \right\}^2 w(x)dx, \\
\widehat{\Gamma}_{T,2}^{(0)} &= -2\int \left\{ \frac{\frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t^{(0)} - x)\varepsilon_t}{\frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t^{(0)} - x)} \right\} \\
&\quad \times \left\{ \frac{\frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t^{(0)} - x)\left[m_{\widehat{\gamma}}(\widehat{h}_t^{(0)}) - m_{\gamma_0}(h_t)\right]}{\frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t^{(0)} - x)} \right\} w(x)dx, \\
\widehat{\Gamma}_{T,3}^{(0)} &= \int \left\{ \frac{\frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t^{(0)} - x)\left[m_{\widehat{\gamma}}(\widehat{h}_t^{(0)}) - m_{\gamma_0}(h_t)\right]}{\frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t^{(0)} - x)} \right\}^2 w(x)dx.
\end{aligned}$$

We show that

$$\widehat{\Gamma}_{T,1}^{(0)} = \widetilde{\Gamma}_T + o_P(T^{-1}b^{-1/2}), \tag{6.13}$$

$$\widehat{\Gamma}_{T,2}^{(0)} = o_P(T^{-1}b^{-1/2}), \tag{6.14}$$

$$\widehat{\Gamma}_{T,3}^{(0)} = o_P(T^{-1}b^{-1/2}), \tag{6.15}$$

where

$$\widetilde{\Gamma}_T = \frac{1}{T^2} \sum_{s,t=1}^{T} \frac{K^{(2)}(h_t - h_s)}{f_h(h_t) f_h(h_s)} w(h_s) \varepsilon_s \varepsilon_t.$$

For the proof of claim (6.13) one applies first Lemma 6.7, that is stated below. This shows that

$$\widehat{\Gamma}_{T,1}^{(0)} = \int \left\{ \frac{\frac{1}{T}\sum_{t=1}^{T} K_b(h_t - x)\varepsilon_t}{\frac{1}{T}\sum_{t=1}^{T} K_b(h_t - x)} \right\}^2 w(x)dx + o_P(T^{-1}b^{-1/2}).$$

Claim (6.13) now follows from continuity of w and f_h and Lemmas 6.1 and 6.5, see below.

For a proof of claim (6.14) one first applies Assumption 4, Assumption 6, $\widehat{\theta} - \theta_0 = O_P(T^{-1/2})$ and Lemmas 6.1, 6.4, and 6.5 to show that

$$\widehat{\Gamma}_{T,2}^{(0)} = (\widehat{\gamma} - \gamma_0) \frac{1}{T^2} \sum_{1 \le s,t \le T} w_{s,t} \varepsilon_t + (\widehat{\theta} - \theta_0) \frac{1}{T^2} \sum_{1 \le s,t \le T} w_{s,t}^* \varepsilon_t + o_P(T^{-1}b^{-1/2}), \qquad (6.16)$$

with $w_{s,t} = \int_I K_b(h_t - x) K_b(x - h_s) \frac{\dot{m}_{\gamma_0}(x)}{f^2(x)} dx$ and $w_{s,t}^* = \int_I K_b(h_t - x) K_b(x - h_s) \frac{m'_{\gamma_0}(x)}{f^2(x)} dx \dot{h}_s$. We now use $\widehat{\theta} - \theta_0 = O_P(T^{-1/2})$, $b|w_{s,t}| \le C$, $b|w_{s,t}^*| \le C$ for a constant C and Davydov‘s inequality (see Corollary 1.1 in Bosq, 1998). This implies that the right hand side of (6.16) is of order $o_P(T^{-1}b^{-1/2})$ which shows claim (6.14).

Claim (6.15) follows directly from Assumption 4.

For the proof of the theorem it remains to show that $T\sqrt{b}(\widetilde{\Gamma}_T - b^{-1/2}M)/V$ converges in distribution to a standard normal distribution. This can be done by the same arguments as in Fan and Li (1999). □

Lemma 6.1. *Under the assumptions of Theorem 6.1 it holds that*

$$\sup_{x \in I} \left| \frac{1}{T} \sum_{t=1}^{T} K_b(h_t - x) - f_h(x) \right| = O_P(b^2 + \sqrt{\log(T)}(Tb)^{-1/2}).$$

For a proof of this statement see Masry (1996).

In the proof of Theorem 6.1 and in the proofs of the following lemmas we make use of the following exponential inequality for martingales. This inequality is a modification of e.g. Lemma 8.9 in van de Geer (2000).

Lemma 6.2. *For independent mean zero random variables* $..., e_{-1}, e_0, e_1, ..., e_T$ *suppose that* $\sup_t \mathbf{E}[\exp(c|e_t|)] < \infty$ *for a constant* $c > 0$ *small enough. Consider a sequence of random variables* $r_1, r_2, ...$ *where* r_t *is measurable with respect to the* σ*-field generated by* $\{e_s : s < t\}$. *Assume that* $\max_{1\le t\le T} |r_t| \le c/2$. *Then it holds that*

$$\mathbf{E}\left[\exp\left(\sum_{t=1}^{T} r_t e_t\right)\right] \le \left\{\mathbf{E}\left[\exp(C\sum_{t=1}^{T} r_t^2)\right]\right\}^{1/2},$$

where

$$C = \mathbf{E}\left[e_t^2 \exp\left(\frac{c}{2}|e_t|\right)\right].$$

Lemma 6.3. *Under the assumptions of Theorem 6.1 it holds that*

$$\sup_{x\in I}\left|\frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t(\theta_0) - x)\varepsilon_t - \frac{1}{T}\sum_{t=1}^{T} K_b(h_t - x)\varepsilon_t\right| = O_P(T^{-1/2-\kappa})$$

for a $\kappa > 0$.

Proof of Lemma 6.3.

Because of Assumption 1 it suffices to show

$$\sup_{x\in I_T}\left|\frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t(\theta_0) - x)\varepsilon_t - \frac{1}{T}\sum_{t=1}^{T} K_b(h_t - x)\varepsilon_t\right| = O_P(T^{-1/2-\kappa}) \tag{6.17}$$

for a $\kappa > 0$. Here I_T is a grid of points of I with cardinality growing polynomially in T. Equality (6.17) can be proved by application of the exponential bound in Lemma 6.2 and by use of the Markov inequality $\mathbf{P}[\sum_{t=1}^{T} r_t e_t \ge c] \le \exp(-sc)\mathbf{E}[\exp(s\sum_{t=1}^{T} r_t e_t)]$. We apply these bounds with $e_t = Z_t$ and with $r_t = \pm\left\{K_b(\widehat{h}_t(\theta_0) - x) - K_b(h_t - x)\right\}\sqrt{h_t}$ if $T^{-1}\sum_{s=1}^{t} \mathbf{1}(|h_t - x| \le 2b) \le Cb$ and $r_t = 0$ else. Here C is a constant that is chosen large enough. Note that for such a choice

$$T^{-1}\sum_{s=1}^{t} \mathbf{1}(|h_t - x| \le 2b) \le Cb \tag{6.18}$$

for all $x \in I$ with probability tending to one. □

Lemma 6.4. *Under the assumptions of Theorem 6.1 it holds that*

$$\sup_{x\in I}\left|\frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t^{(0)} - x)\varepsilon_t - \frac{1}{T}\sum_{t=1}^{T} K_b(h_t - x)\varepsilon_t\right| = O_P(T^{-1/2-\kappa})$$

for a $\kappa > 0$.

Proof of Lemma 6.4.

Recall that $\widehat{h}_t(\widehat{\theta}^{(0)}) = \widehat{h}_t^{(0)}$. Because of Lemma 6.3 it remains to show that for $C > 0$

$$\sup_{x\in I, \|\theta-\theta_0\|\leq CT^{-\xi}} \left| \frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t(\theta) - x)\varepsilon_t - \frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t(\theta_0) - x)\varepsilon_t \right| = O_P(T^{-1/2-\kappa})$$

for a $\kappa > 0$. This claim can be shown with similar arguments as the statement of Lemma 6.3. In a first step the supremum is replaced by a supremum that runs over a grid of values of x and of θ. Again, the grid has cardinality that polynomially grows with T. □

Lemma 6.5. *Under the assumptions of Theorem 6.1 it holds that*

$$\sup_{x\in I} \left| \frac{1}{T}\sum_{t=1}^{T} K_b(h_t - x)\varepsilon_t \right| = O_P(\sqrt{\log(T)}(Tb)^{-1/2}).$$

Proof of Lemma 6.5.

This lemma can be shown with similar arguments as in the proofs of the last two lemmas. □

Lemma 6.6. *Under the assumptions of Theorem 6.1 it holds that*

$$\sup_{x\in I} \left| \frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t^{(0)} - x) - \frac{1}{T}\sum_{t=1}^{T} K_b(h_t - x) \right| = O_P(T^{-\kappa}\sqrt{b})$$

for a $\kappa > 0$.

Proof of Lemma 6.6.

The lemma directly follows from (6.18) and the bound

$$\sup_{x\in I} \left| K_b(\widehat{h}_t^{(0)} - x) - K_b(h_t - x) \right| = O_P(T^{-\xi}b^{-1}).$$

□

Lemma 6.7. *Under the assumptions of Theorem 6.1 it holds that*

$$\sup_{x\in I} \left| \frac{\frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t^{(0)} - x)\varepsilon_t}{\frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t^{(0)} - x)} - \frac{\frac{1}{T}\sum_{t=1}^{T} K_b(h_t - x)\varepsilon_t}{\frac{1}{T}\sum_{t=1}^{T} K_b(h_t - x)} \right| = O_P(T^{-1/2-\kappa})$$

for a $\kappa > 0$.

Proof of Lemma 6.7.

The statement of Lemma 6.7 follows directly from Lemmas 6.1 – 6.6. □

Proof of Theorem 6.2.

For functions m we define

$$\widehat{\Gamma}_T(\psi, m) = \int \left\{ \frac{\frac{1}{T}\sum_{t=1}^T K_b(\widehat{h}_t(\psi, m) - x)\left[Y_t - m_{\widehat{\gamma}}(\widehat{h}_t^{(0)})\right]}{\frac{1}{T}\sum_{t=1}^T K_b(\widehat{h}_t(\psi, m) - x)} \right\}^2 w(x)dx.$$

Note that $\widehat{\Gamma}_T^{(k)} = \widehat{\Gamma}_T(\widehat{\psi}^{(k)}, \widehat{m}_b^{(k)})$ for $k \geq 1$ and $\widehat{\Gamma}_T^{(0)} = \widehat{\Gamma}_T(\widehat{\psi}^{(0)}, m_{\widehat{\gamma}})$. The statement of Theorem 6.2 follows from the following two claims. For $C > 0$ it holds that

$$\sup_{(\psi_1,m_1),(\psi_2,m_2)\in\mathcal{M}^{C,*}} \left|\widehat{\Gamma}_T(\psi_1, m_1) - \widehat{\Gamma}_T(\psi_2, m_2)\right| = o_P(T^{-1}b^{-1/2}), \quad (6.19)$$

$$\left(\widehat{\psi}^{(k)}, \widehat{m}_b^{(k)}\right) \in \mathcal{M}^{C,*}. \quad (6.20)$$

Here $\mathcal{M}^{C,*}$ denotes the set of all tuples (ψ, m) with $m \in \mathcal{M}^C$ and where ψ fulfills $\|\psi - \psi_0\| \leq b^{3/2}T^{-\iota}$. The set $\mathcal{M}^C$ is the class of all functions m whose second derivative is absolutely bounded by C, which coincide outside of I with $m_{\widehat{\gamma}}$ and which fulfill:

$$\sup_{x\in I} |m(x) - m_{b,0}(x)| \leq C(Tb)^{-1/2}\sqrt{\log(T)},$$

where

$$m_{b,0}(x) = \frac{\mathbf{E}[K_b(h_t - x)m_0(h_t)]}{\mathbf{E}[K_b(h_t - x)]}.$$

For a proof of (6.19) we will show that for all $C > 0$ for $\kappa > 0$ small enough

$$\sup_{x\in I,(\psi_1,m_1),(\psi_2,m_2)\in\mathcal{M}^{C,*}} \left|\frac{1}{T}\sum_{t=1}^T K_b(\widehat{h}_t(\psi_2, m_2) - x)\varepsilon_t - \frac{1}{T}\sum_{t=1}^T K_b(\widehat{h}_t(\psi_1, m_1) - x)\varepsilon_t\right| = O_P(T^{-1/2-\kappa}), \quad (6.21)$$

$$\sup_{x\in I,(\psi_1,m_1),(\psi_2,m_2)\in\mathcal{M}^{C,*}} \left|\frac{1}{T}\sum_{t=1}^T K_b(\widehat{h}_t(\psi_2, m_2) - x) - \frac{1}{T}\sum_{t=1}^T K_b(\widehat{h}_t(\psi_1, m_1) - x)\right| = O_P(\sqrt{b}\, T^{-\kappa}). \quad (6.22)$$

Using these two bounds claim (6.19) follows by similar arguments as in the proof of Theorem 6.1. We now show (6.21) and (6.22). Claim (6.22) follows by a direct bound. We now show claim (6.21). For simplicity we neglect the discussion of the parametric part and we show

$$\sup_{x\in I, m_1, m_2 \in \mathcal{M}^C} \left| \frac{1}{T}\sum_{t=1}^T K_b(\widehat{h}_t(m_2) - x)\varepsilon_t - \frac{1}{T}\sum_{t=1}^T K_b(\widehat{h}_t(m_1) - x)\varepsilon_t \right| = O_P(T^{-1/2-\kappa}), \qquad (6.23)$$

where $\widehat{h}_t(m) = \widehat{h}_t(\psi_0, m)$. For a proof of (6.23) we use a chaining argument and we proceed similarly as e.g. in the proof of Lemma 3.2 in van de Geer (2000). Put $\delta = C(Tb)^{-1/2}\sqrt{\log(T)}$ and for $s \geq 1$ consider $2^{-s}\delta$ covering sets $\mathcal{M}_s^C$ of $\mathcal{M}^C$, i.e. for each $m \in \mathcal{M}^C$ there exists $m^* \in \mathcal{M}_s^C$ with $\|m^* - m\|_\infty \leq 2^{-s}\delta$. The covering sets can be chosen such that their cardinality $\sharp\mathcal{M}_s^C$ does not exceed $C^* \exp[(2^{-s}\delta)^{-1/2}]$ for a constant $C^* > 0$. This is a standard bound for coverings of Sobolev balls, see van de Geer (2000). We now write $\Delta_t(m, m^*) = T^{-1}\{K_b(\widehat{h}_t(m) - x) - K_b(\widehat{h}_t(m^*) - x)\}\varepsilon_t^*$ with $\varepsilon_t^* = \varepsilon_t \mathbf{1}[|\varepsilon_t| \leq C^{**}\log T] - \mathbf{E}\{\varepsilon_t \mathbf{1}[|\varepsilon_t| \leq C^{**}\log T]\}$ for a constant $C^{**} > 0$ that is large enough. Now for $C^{**} > 0$ large enough

$$\sup_{x\in I, m, m^* \in \mathcal{M}^C} \left| \frac{1}{T}\sum_{t=1}^T K_b(\widehat{h}_t(m^*) - x)\varepsilon_t - \frac{1}{T}\sum_{t=1}^T K_b(\widehat{h}_t(m) - x)\varepsilon_t - \sum_{t=1}^T \Delta_t(m, m^*) \right| = O_P(T^{-1/2-\kappa}).$$

For $m_1, m_2 \in \mathcal{M}^C$ we choose now $m_1^s, m_2^s \in \mathcal{M}_s^C$ with $\|m_1^s - m_1\|_\infty \leq 2^{-s}\delta$, $\|m_2^s - m_2\|_\infty \leq 2^{-s}\delta$ and we consider the chain

$$\begin{aligned} \sum_{t=1}^T \Delta_t(m_1, m_2) &= \sum_{t=1}^T \Delta_t(m_1^0, m_2^0) - \sum_{s=1}^{G_T}\sum_{t=1}^T \Delta_t(m_1^{s-1}, m_1^s) + \sum_{s=1}^{G_T}\sum_{t=1}^T \Delta_t(m_2^{s-1}, m_2^s) \\ &\quad - \sum_{t=1}^T \Delta_t(m_1^{G_T}, m_1) + \sum_{t=1}^T \Delta_t(m_2^{G_T}, m_2), \end{aligned}$$

where G_T is the largest integer with $2^{G_T/4}T^{-3/2+\xi}b^{-5/2}\log(T) < \rho$. The constants ρ and ξ were introduced in Assumption 2 and Assumption 10, respectively. We now give a bound on $\mathbf{P}[\sup_{m_1 \in \mathcal{M}^C} \sum_{s=1}^{G_T} \sum_{t=1}^{T} \Delta_t(m_1^{s-1}, m_1^s) > T^{-1/2-\kappa}]$. Similar bounds can be proved for the other terms and for $\mathbf{P}[\inf_{m_1 \in \mathcal{M}^C} \sum_{s=1}^{G_T} \sum_{t=1}^{T} \Delta_t(m_1^{s-1}, m_1^s) < -T^{-1/2-\kappa}]$. We get the following inequality with $\eta_s = c2^{-3/4s}$ where c is chosen such that $\sum_{s=1}^{\infty} \eta_s \leq 1$.

$$\begin{aligned}
&\mathbf{P}\left[\sup_{m_1 \in \mathcal{M}^C} \sum_{s=1}^{G_T} \sum_{t=1}^{T} \Delta_t(m_1^{s-1}, m_1^s) > T^{-1/2-\kappa}\right] \\
&\quad \leq \sum_{s=1}^{G_T} \mathbf{P}\left[\sup_{m_1 \in \mathcal{M}^C} \sum_{t=1}^{T} \Delta_t(m_1^{s-1}, m_1^s) > \eta_s T^{-1/2-\kappa}\right] \\
&\quad \leq \sum_{s=1}^{G_T} \sharp\mathcal{M}_{s-1}^C \sharp\mathcal{M}_s^C \mathbf{P}\left[\sum_{t=1}^{T} \Delta_t(m_1^{s-1}, m_1^s) > \eta_s T^{-1/2-\kappa}\right] \\
&\quad \leq \sum_{s=1}^{G_T} (C^*)^2 \exp[2(2^{-s}\delta)^{-1/2}] \mathbf{P}\left[\sum_{t=1}^{T} \Delta_t(m_1^{s-1}, m_1^s) > \eta_s T^{-1/2-\kappa}\right] \\
&\quad \leq \sum_{s=1}^{G_T} (C^*)^2 \exp[2(2^{-s}\delta)^{-1/2}] \exp[c_* 2^{s/2} T^{1-2\kappa-2\xi} b^4 (\log T)^{-1}]
\end{aligned}$$

with a constant $c_* > 0$. The last inequality follows by application of the exponential inequality of Lemma 6.2. At this point it is also used that $2^{s/4} T^{-3/2+\xi} b^{-5/2} \log(T) < \rho$ for $s \leq G_T$. It can be easily checked that the right hand side of the last inequality converges to 0. This holds for $\kappa > 0$ small enough because of $15\eta + 8\xi < 3$, see Assumption 10. This concludes the proof of (6.19).

For the proof of (6.20) we will argue that for $l \leq k$

$$\sup_{x \in I} \left| \widehat{m}_b^{(l)}(x) - m_{b,0}(x) \right| \leq C(Tb)^{-1/2} \sqrt{\log(T)}, \tag{6.24}$$

$$\sup_{x \in I} \left| D_2 \widehat{m}_b^{(l)}(x) \right| = O_P(1). \tag{6.25}$$

Here, $D_k m$ denotes the k-th derivative of m.

For a proof of (6.24) note that from (6.21) and (6.22) it follows that for $\kappa > 0$ small enough

$$\begin{aligned}
\sup_{x \in I} \left| \frac{1}{T} \sum_{t=1}^{T} K_b(\widehat{h}_t^{(k)} - x)\varepsilon_t - \frac{1}{T} \sum_{t=1}^{T} K_b(\widehat{h}_t^{(0)} - x)\varepsilon_t \right| &= O_P(T^{-1/2-\kappa}), \\
\sup_{x \in I} \left| \frac{1}{T} \sum_{t=1}^{T} K_b(\widehat{h}_t^{(k)} - x) - \frac{1}{T} \sum_{t=1}^{T} K_b(\widehat{h}_t^{(0)} - x) \right| &= O_P(\sqrt{b}\, T^{-\kappa}).
\end{aligned}$$

Thus (6.24) follows from our results on $\frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t^{(0)} - x)\varepsilon_t$ and $\frac{1}{T}\sum_{t=1}^{T} K_b(\widehat{h}_t^{(0)} - x)$ in the proof of Theorem 6.1.

For a proof of (6.25) we write

$$\widehat{m}_b^{(k)}(x) = \frac{\widehat{r}_b^A(x) + \widehat{r}_b^B(x)}{\widehat{f}_b^{(k)}(x)} + m_{\widehat{\gamma}}(x),$$

where $\widehat{r}_b^A(x) = T^{-1}\sum_{t=1}^{T} K_b(\widehat{h}_t^{(k-1)} - x)\varepsilon_t$, $\widehat{r}_b^B(x) = T^{-1}\sum_{t=1}^{T} K_b(\widehat{h}_t^{(k-1)} - x)[m_{\gamma_0}(h_t) - m_{\widehat{\gamma}}(\widehat{h}_t^{(0)})]$, and $\widehat{f}_b^{(k)}(x) = T^{-1}\sum_{t=1}^{T} K_b(\widehat{h}_t^{(k-1)} - x)$. For the proof of (6.20) it suffices to show for $0 \leq j \leq 2$ that $\sup_{x\in I}\left|D_j\widehat{r}_b^A(x)\right| = O_P(1)$, $\sup_{x\in I}\left|D_j\widehat{r}_b^B(x)\right| = O_P(1)$, $\sup_{x\in I}\left|D_j\widehat{f}_b^{(k)}(x)\right| = O_P(1)$ and $\sup_{x\in I}\left|\widehat{f}_b^{(k)}(x)^{-1}\right| = O_P(1)$. This can be done by similar arguments as in the proof of (6.19). □

References

Alesina, A., and L. H. Summers (1993). “Central bank independence and macroeconomic performance: some comparative evidence.” *Journal of Money, Credit and Banking* 25, 151–162.

Andersen, T., and T. Bollerslev (1998). “Answering the skeptics: yes, standard volatility models do provide accurate forecasts.” *International Economic Review* 39, 885–905.

Arvanitis, S., and A. Demos (2004). “Time dependence and moments of a family of time-varying parameter GARCH-in-mean models.” *Journal of Time Series Analysis* 25, 1–25.

Backus, D. K., and A. W. Gregory (1993). “Theoretical relations between risk premiums and conditional variances.” *Journal of Business & Economic Statistics* 11, 177–185.

Backus, D. K., and S. E. Zin (1993). “Long-memory inflation uncertainty: evidence from the term structure of interest rates.” *Journal of Money, Credit and Banking* 25, 681–700.

Baeckert, G. and G. Wu (2000). “Asymmetric volatility and risk in equity markets.” *The Review of Financial Studies* 13, 1–42.

Baillie, R. T. (1996). “Long memory processes and fractional integration in econometrics.” *Journal of Econometrics* 73, 5–59.

Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen (1996a). ”Fractionally integrated generalized autoregressive conditional heteroskedasticity." *Journal of Econometrics* 74, 3–30.

Baillie, R. T., C. Chung, and A. M. Tieslau (1996b). "Analyzing inflation by the fractionally integrated ARFIMA-GARCH model." *Journal of Applied Econometrics* 11, 23–40.

Baillie, R. T. and R. P. DeGennaro (1990). "Stock returns and volatility." *Journal of Financial and Quantitative Analysis* 25, 203–214.

Baillie, R. T., Y. W. Han, and T. Kwon. (2002) "Further long memory properties of inflationary shocks." *Southern Economic Journal* 68, 496–510.

Ball, L. (1992). "Why does high inflation raise inflation uncertainty?" *Journal of Monetary Economics* 29, 371–388.

Ball, L., and S. Cecchetti (1990). "Inflation and uncertainty at short and long horizons." *Brookings Papers on Economic Activity*, 215–254.

Banerjee, A., and B. Russell (2001). "The relationship between the markup and inflation in the G7 economies and Australia." *Review of Economics and Statistics* 83, 377–387.

Barro, R. (1996). "Determinants of economic growth: a cross-country empirical study." Working Paper 5698, National Bureau of Economic Research.

Baum, C. F., J. T. Barkoulas, and M. Caglayan (1999). "Persistence in international inflation rates." *Southern Economic Journal* 65, 900–913.

Bauwens, L., S. Laurent, and J. V. K. Rombouts (2006). "Multivariate GARCH models: a survey." *Journal of Applied Econometrics* 21, 79–109.

Beine, M., and S. Laurent (2003). "Central bank interventions and jumps in double long memory models of daily exchange rates." *Journal of Empirical Finance* 10, 641–660.

Benigno P., and J. D. Lòpez-Salido (2002). "Inflation persistence and optimal monetary policy in the Euro area." Working Paper 178, European Central Bank.

Bera, A. K., and M. L. Higgins (1993). "ARCH models: properties, estimation and testing." *Journal of Economic Surveys* 7, 305–362.

Bernanke, B. S., and F. S. Mishkin (1992). "Central bank behavior and the strategy of monetary policy: observations from six industrialized countries." Working Paper 4082, National Bureau of Economic Research.

Black, F., and M. Scholes (1973). "The pricing of options and corporate liabilities." *Journal of Political Economy* 81, 637–654.

Bodurtha, J. N., and N. C. Mark (1991). "Testing the CAPM with time-varying risks and returns." *The Journal of Finance* XLVI, 1485–1505.

Bollerslev, T. (1986). "Generalized autoregressive conditional heteroskedasticity." *Journal of Econometrics* 31, 307–327.

Bollerslev, T., R. Y. Chou, and K. F. Kroner (1992). "ARCH modeling in finance: a review of the theory and empirical evidence." *Journal of Econometrics* 52, 5–59.

Bollerslev, T., and H. O. Mikkelsen (1996). "Modelling and pricing long memory in stock market volatility." *Journal of Econometrics* 73, 151–184.

Bos, C. S., P. H. Franses, and M. Ooms (1999). "Long memory and level shifts: reanalyzing inflation rates." *Empirical Economics* 24, 427–449.

Bos, C. S., P. H. Franses, and M. Ooms (2002). "Inflation, forecast intervals and long memory regression models." *International Journal of Forecasting* 18, 243–264.

Bosq, D. (1998). *Nonparametric statistics for stochastic processes: estimation and prediction.* Lecture Notes in Statistics, Vol. 110, 2nd Edition, Springer-Verlag, New York.

Bougerol, P., and N. Picard (1992). "Stationarity of GARCH processes and of some non-negative time series." *Journal of Econometrics* 52, 115–128.

Brunetti, C., and C. L. Gilbert (2000). "Bivariate FIGARCH and fractional cointegration." *Journal of Empirical Finance* 7, 509–530.

Brunner, A. D., and G. D. Hess (1993). "Are higher levels of inflation less predictable? A state-dependent conditional heteroscedasticity approach." *Journal of Business and Economic Statistics* 11, 187–197.

Campbell, J. and L. Hentschel (1992). "No news is good news: an asymmetric model of changing volatility in stock returns." *Journal of Financial Economics* 18, 373–399.

Caporale, G. M., and L. A. Gil-Alana (2003). "Long memory and structural breaks in hyperinflation countries." *Journal of Economics and Finance* 2, 136–152.

Caporale, T. and B. McKiernan (1996). "The relationship between output variability and growth: evidence from post war UK data." *Scottish Journal of Political Economy* 43, 229–236.

Caporale, T., and B. McKierman (1997). "High and variable inflation: further evidence on the Friedman hypothesis." *Economics Letters* 54, 65–68.

Caporin, M. (2003). "Identification of long memory in GARCH models." *Statistical Methods & Applications* 12, 133–151.

Carrasco, M., and X. Chen (2002). "Mixing and moment properties of various GARCH and stochastic volatility models." *Econometric Theory* 18, 17–39.

Chung, C.-F. (1999). "Estimating the fractionally integrated GARCH model." Working Paper, National Taiwan University.

Clarida, R., J. Galí, and M. Gertler (1999). "The science of monetary policy: a new Keynesian Perspective." *Journal of Economic Literature* XXXVII, 1661–1707.

Conrad, C., and B. R. Haag (2006). "Inequality constraints in the fractionally integrated GARCH model." *Journal of Financial Econometrics* 4, 413–449.

Conrad, C., and M. Karanasos (2005a). "On the inflation-uncertainty hypothesis in the USA, Japan and the UK: a dual long memory approach." *Japan and the World Economy* 17, 327–343.

Conrad, C., and M. Karanasos (2005b). "Dual long memory in inflation dynamics across countries of the Euro area and the link between inflation uncertainty and macroeconomic performance." *Studies in Nonlinear Dynamics & Econometrics* 4, Article 5.

Conrad, C., and M. Karanasos (2006). "The impulse response function of the long memory GARCH process." *Economics Letters* 90, 34–41.

Cosimano, T., and D. Jansen (1988). "Estimates of the variance of US inflation based upon the ARCH model." *Journal of Money, Credit and Banking*, 20, 409–421.

Coulson, N. E., and R. P. Robins (1985). "Aggregate economic activity and the variance of inflation: another look." *Economics Letters* 17, 71–75.

Cukierman, A. (1992). *Central Bank strategy, credibility and independence.* MIT Press, Cambridge (MA).

Cukierman, A., and A. Meltzer (1986). "A theory of ambiguity, credibility, and inflation under discretion and asymmetric information." *Econometrica* 54, 1099–1128.

Das, S., and N. Sarkar (2000). "An ARCH in the nonlinear mean model." *The Indian Journal of Statistics* 62, 327–344.

Davidson, J. (2004). "Moment and memory properties of linear conditional heteroscedasticity models, and a new model." *Journal of Business & Economic Statistics* 22, 16–19.

Davis, G., and B. Kanago (2000). "The level and uncertainty of inflation: results from OECD forecasts." *Economic Inquiry* 38, 58–72.

Degiannakis, S., and E. Xekalaki (2004). "ARCH models: a review." *Quality Technology & Quantitative Management* 1, 271–324.

Devereux, M. (1989). "A positive theory of inflation and inflation variance." *Economic Inquiry* 27, 105–116.

Ding, Z., and C. W. J. Granger (1996). "Modeling volatility persistence of speculative returns: a new approach." *Journal of Econometrics* 73, 185–215.

Ding, Z., C. W. J. Granger, and R. F. Engle (1993). "A long memory property of stock market returns and a new model." *Journal of Empirical Finance* 1, 83–106.

Dotsey, M., and P. Sarte (2000). "Inflation uncertainty and growth in a cash-in-advance economy." *Journal of Monetary Economics* 45, 631–655.

Elder, J. (2004). "Another perspective on the effects of inflation uncertainty." *Journal of Money, Credit and Banking* 5, 911–928.

Engle, R. F. (1982). "Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation." *Econometrica* 73, 185–215.

Engle, R. F. (2002). "New frontiers in ARCH models." *Journal of Applied Econometrics* 17, 425–446.

Engle, R. F. (2004). "Risk and Volatility: Econometric Models and Financial Practice." *American Economic Review* 94, 326–348.

Engle, R. F., and T. Bollerslev (1986). "Modelling the persistence of conditional variances." *Econometric Reviews* 5, 1–50.

Engle, R. F., and G. G. J. Lee (1993). "A permanent and a transitory component model of stock return volatility." Discussion Paper 9244, University of California, San Diego.

Engle, R. F., D. M. Lilien, and R. P. Robins (1987). "Estimating time varying risk premia in the term structure." *Journal of Business and Economic Statistics* 9, 345–359.

Engle, R. F., and J. R. Russell (1998). "Autoregressive conditional duration: a new model for irregularly spaced transaction data." *Econometrica* 66, 1127–1162.

Evans, M., and P. Wachtel (1993). "Inflation regimes and the sources of inflation uncertainty." *Journal of Money, Credit and Banking* 25, 475–511.

Fan, Y. and Q. Li (1999). "Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing." *Nonparametric Statistics* 10, 245–271.

Fountas, S. (2001). "The relationship between inflation and inflation-uncertainty in the UK: 1885-1998." *Economics Letters* 74, 77–83.

Fountas, S., A. Ioannidi, and M. Karanasos (2004). "Inflation, inflation-uncertainty, and a common European monetary policy." *Manchester School* 2, 221–242.

Fountas, S., M. Karanasos, and M. Karanassou (2003). "A GARCH model of inflation and inflation-uncertainty with simultaneous feedback." Unpublished manuscript, University of York.

Fountas, S., M. Karanasos, and J. Kim (2002). "Inflation and output growth uncertainty and their relationship with inflation and output growth." *Economics Letters* 75, 293–301.

Fountas, S., and M. Karanasos (2005). "Inflation, output growth, and nominal and real uncertainty: empirical evidence for the G7." *Journal of International Money and Finance*, forthcoming.

French, K. R., G. W. Schwert, and R. F. Stambaugh (1987). "Expected stock returns and volatility." *Journal of Financial Economics* 19, 3–29.

Friedman, M. (1977). "Nobel lecture: inflation and unemployment." *Journal of Political Economy* 85, 451–472.

Genotte, G., and T. Marsh (1993). "Valuations in economic uncertainty and risk premiums on capital assets." *European Economic Review* 37, 1021–1041.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2005). "There is a risk-return trade-off after all." *Journal of Financial Economics* 76, 509–548.

Giraitis, L., R. Leipus, and D. Surgailis (2005). "Recent advances in ARCH modelling." In G. Teyssière and A. Kirman (eds.), *Long-Memory in Economics*, Springer-Verlag, Heidelberg.

Giraitis, L., P. M. Robinson, and D. Surgailis (2004). "LARCH, leverage and long memory." *Journal of Financial Econometrics* 2, 177–210.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993). "On the relationship between the expected value and the volatility of the nominal excess return on stocks." *The Journal of Finance* XLVIII, 1779–1801.

Granger, C. W. J. (1969). "Investigating causal relations by econometric models and cross-spectral methods." *Econometrica* 37, 424–438.

Granger, C. W. J. (1980). "Long memory relationships and the aggregation of dynamic models." *Journal of Econometrics* 25, 227–238.

Granger, C. W. J., and R. Joyeux (1980). "An introduction to long memory time series models and fractional differencing." *Journal of Time Series Analysis* 1, 15–39.

Grier, K. B., Ó. T. Henry, N. Olekalns, and K. Shields (2004). "The asymmetric effects of uncertainty on inflation and output growth." *Journal of Applied Econometrics* 5, 551–565.

Grier, K. B., and M. J. Perry (1998). "On inflation and inflation uncertainty in the G7 countries." *Journal of International Money and Finance* 17, 671–689.

Grier, K. B., and M. J. Perry (2000). "The effects of real and nominal uncertainty on inflation and output growth: some GARCH-M evidence." *Journal of Applied Econometrics* 15, 45–58.

Grier, K. B., and G. Tullock (1989). "An empirical analysis of cross-national economic growth: 1951-1980." *Journal of Monetary Economics* 24, 259–276.

Guedhami, O., and O. Sy (2005). "Does the conditional market skewness resolve the puzzeling market risk-return relationship?" *The Quarterly Review of Economics and Statistics* 45, 582–598.

Guo, H., and C. J. Neely (2006). "Investigating the intertemporal risk-return relation in international stock markets with the component GARCH model." Working Paper 006A, Federal Reserve Bank of St. Louis.

Guo, H., and R. Whitelaw (2006). "Uncovering the risk-return relation in the stock market." *Journal of Finance*, forthcoming.

Gylfason, T., and T. T. Herbertsson (2001). "Does inflation matter for growth?" *Japan and the World Economy* 13, 405–428.

Härdle W., and E. Mammen (1993). "Comparing nonparametric versus parametric regression fits." *The Annals of Statistics* 21, 1926–1947.

Harvey, C. R., and A. Siddique (1999). "Autoregressive conditional skewness." *Journal of Financial and Quantitative Analysis* 34, 465–487.

Hassler, U., and J. Wolters (1995). "Long memory in inflation rates: international evidence." *Journal of Business and Economic Statistics* 13, 37–45.

He, C., and T. Teräsvirta (1999a). "Fourth moment structure of the GARCH(p, q) process." *Econometric Theory* 15, 824–846.

He, C., and T. Teräsvirta (1999b). "Properties of moments of a family of GARCH processes." *Journal of Econometrics* 92, 173–192.

He, C., and T. Teräsvirta (1999c). "Properties of the autocorrelation function of squared observations for second-order GARCH processes under two sets of parameter constraints." *Journal of Time Series Analysis* 20, 23–30.

Henry, M., and P. Zaffaroni (2003). “The long range dependence paradigm for macroeconomics and finance.” In P. Doukhan, G. Oppenheim and M. Taqqu (eds.), *The Theory and Applications of Long-Range Dependence*, Birkhauser, Boston.

Hjellvik, V., and D. Tjøstheim (1995). “Nonparametric tests of linearity for time series.” *Biometrika* 82, 351–368.

Hjellvik, V., Q. Yao and D. Tjøstheim (1998). “Linearity testing using local polynomial approximation.” *Journal of Statistical Planning and Inference* 68, 295–321.

Holland, S. (1993). “Comment on inflation regimes and the sources of inflation uncertainty.” *Journal of Money, Credit and Banking* 25, 514–520.

Holland, S. (1995). “Inflation and uncertainty: tests for temporal ordering.” *Journal of Money, Credit and Banking* 27, 827–837.

Hosking, J. R. M. (1981). “Fractional differencing.” *Biometrika* 68, 165–176.

Hwang, Y. (2001). “Relationship between inflation rate and inflation uncertainty.” *Economics Letters* 73, 179–186.

Hyung, N., and P. H. Franses (2002). “Inflation rates: long-memory, level shifts, or both?” Working Paper, Econometric Institute Report 08, Erasmus University Rotterdam.

Hyung, N., and P. H. Franses (2004). “Structural breaks and long memory in US inflation rates: do they matter for forecasting?” Working Paper, Social Science Research Network.

Jansen, D. W. (1989). “Does inflation uncertainty affect output growth? Further evidence.” *Federal Reserve Bank of St. Louis Review*, July/August, 43–54.

Jasiak, J. (1998). “Persistence in intertrade durations.” *Finance* 19, 165–195.

Karanasos, M. (1999). “The second moment and the autocovariance function of the squared errors of the GARCH model.” *Journal of Econometrics* 90, 63–76.

Karanasos, M. (2001). "Prediction in ARMA models with GARCH-in-mean effects." *Journal of Time Series Analysis* 22, 555–576.

Karanasos, M., M. Karanassou, and S. Fountas (2004a). "Analyzing US inflation by a GARCH model with simultaneous feedback." *WSEAS Transactions on Information Science and Applications* 2, 767–772.

Karanasos, M., and J. Kim (2005). "The inflation-output variability relationship in the G3: a bivariate GARCH (BEKK) approach." *Risk Letters*, forthcoming.

Karanasos, M., Z. Psaradakis, and M. Sola (2004b). "On the autocorrelation properties of long-memory GARCH processes." *Journal of Time Series Analysis* 25, 265–281.

Karanassou, M., and D. J. Snower (2003). "An anatomy of the Phillips curve." Unpublished manuscript.

Kirman, A., and G. Teyssière (ed.) (2005). *Long-Memory in Economics.* Springer, Berlin.

Kontonikas, A. (2004). "Inflation and inflation uncertainty in the United Kingdom, evidence from GARCH modelling." *Economic Modelling* 21, 525–543.

Kreiss, J.-P., M. H. Neumann, and Q. Yao (2002). "Bootstrap tests for simple structures in nonparametric time series regression." Preprint, Technische Universität Braunschweig.

Krugman, P. R. (1998). "It's baaack! Japan's slump and the return to the liquidity trap." *Brookings Papers on Economic Activity* 2, 137–187.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin (1992). "Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic series are non-stationary?" *Journal of Econometrics* 54, 159–178.

Lanne, M., and P. Saikkonen (2006). "Why is it so difficult to uncover the risk-return tradeoff in stock returns?" *Economics Letters* 92, 118–125.

Laurent, S., and J. P. Peters (2002). "G@RCH 2.2: an Ox package for estimating and forecasting various ARCH models." *Journal of Economic Surveys* 3, 447–485.

Lee, H. S., and C. Amsler (1997). "Consistency of the KPSS unit root test against fractionally integrated alternative." *Economics Letters* 55, 151–160.

Lee, S.-W., and B. E. Hansen (1994). "Aymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator." *Econometric Theory* 10, 29–52.

Levine, R., and D. Renelt (1992). "A sensitivity analysis of cross-country growth regressions." *American Economic Review* 82, 942–963.

Li, H. (2003). "Mean/Variance relation and the conditional distribution." Working Paper, Swedish School of Economics and Business Administration, Helsinki.

Li, Q., J. Yang, C. Hsiao, and Y.-J. Chang (2005). "The relationship between stock returns and volatility in international stock markets." *Journal of Empirical Finance* 12, 650–665.

Li, W. K., S. Ling, and M. McAleer (2002). "Recent theoretical results for time series models with GARCH errors." *Journal of Economic Surveys* 16, 245–269.

Ling, S., and M. McAleer (2002a). "Necessary and sufficient moment conditions for the GARCH(p, q) and asymmetric power GARCH(p, q) models." *Econometric Theory* 18, 722–729.

Ling, S., and M. McAleer (2002b). "Stationarity and the existence of moments of a family of GARCH processes." *Journal of Econometrics* 106, 109–117.

Lintner, J. (1965a). "The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets." *Review of Economics and Statistics* 47, 12–37.

Lintner, J. (1965b). "Security prices, risk and maximal gains from diversification." *Journal of Finance* 20, 587–615.

Linton, O., and E. Mammen (2005). "Estimating semiparametric ARCH(∞) models by kernel smoothing methods." *Econometrica* 73, 771–836.

Linton, O., and B. Perron (2003). "The shape of the risk premium: evidence from a semiparametric generalized autoregressive conditional heteroscedasticity model." *Journal of Business & Economic Statistics* 21, 354–367.

Lumsdaine, R. L. (1996). "Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models." *Econometrica* 64, 575–596.

Mammen, E., O. Linton, and J. P. Nielson (1999). "The existence and asymptotic properties of a backfitting algorithm under weak conditions." *The Annals of Statistics* 27, 1443–1490.

Markowitz, H. M. (1952). "Portfolio selection." *Journal of Finance* 7, 77–91.

Masry, E. (1996). "Multivariate local polynomial regression for time series: uniform strong consistency and rates." *Journal of Time Series Analysis* 17, 571–599.

Masry, E., and D. Tjøstheim (1995). "Nonparametric estimation and identification of nonlinear ARCH times series." *Econometric Theory* 11, 258–289.

McCallum, B. T. (1988). "Robustness properties of a rule for monetary policy." *Carnegie Rochester Conference Series on Public Policy* 29, 173–203.

McTaggart, D. (1992). "The cost of inflation in Australia." In Blundell-Wignall, A. (ed.), *Inflation, Disinflation and Monetary Policy*, Reserve Bank of Australia, Sydney.

Mehra, R., and E. Prescott (1985). "The equity premium: a puzzle." *Journal of Monetary Economics* 15, 145–161.

Merton, R. C. (1973a). "An intertemporal capital asset pricing model." *Econometrica* 41, 867–886.

Merton, R. C. (1973b). "Theory of rational option pricing." *Bell Journal of Economics and Management Science* 4, 141–183.

Merton, R. C. (1980). "On estimating the expected return on the market - an exploratory investigation." *Journal of Financial Economics* 8, 323–361.

Morana, C. (2002). "Common persistent factors in inflation and excess nominal money growth." *Studies in Nonlinear Dynamics & Econometrics* 3, Article 3.

Nelson, D. B. (1990). "Stationarity and persistence in the GARCH(1,1) model." *Econometric Theory* 6, 318–334.

Nelson, D. B. (1991). "Conditional heteroskedasticity in asset returns: a new approach." *Econometrica* 59, 347–370.

Nelson, D. B., and C. Q. Cao (1992). "Inequality constraints in the univariate GARCH model." *Journal of Business & Economic Statistics* 10, 229–235.

Ooms, M., and U. Hassler (1997). "On the effect of seasonal adjustment on the log-periodogram regression." *Economics Letters* 56, 135–141.

Pagan, A. R (1984). "Econometric issues in the analysis of regressions with generated regressors." *International Economic Review* 25, 221–247.

Pagan, A. R., and Y. S. Hong (1990). "Nonparametric estimation and the risk premium." In Barnett, W. A., J. Powell, and G. E. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics*, 51–75, Cambridge University Press, Cambridge.

Pagan, A. R., and A. Ullah (1988). "The econometric analysis of models with risk terms." *Journal of Applied Econometrics* 3, 87–105.

Park, J. Y., and P. C. B. Phillips (1989). "Statistical inference in regressions with integrated processes: part 2." *Econometric Theory* 5, 95–132.

Pesaran, M. H., and Y. Shin (1998). "Impulse response analysis in linear multivariate models." *Economics Letters* 58, 165–193.

Phillips, P. C. B., and P. Perron (1988). "Testing for a unit root in time series regression." *Biometrika*, 335–346.

Poggi, J.-M., and B. Portier (1997). "A test of linearity for functional autoregressive models." *Journal of Time Series Analysis* 18, 615–639.

Poterba, J. M., and L. H. Summers (1986). "The persistence of volatility and stock market fluctuations." *The American Economic Review* 76, 1142–1151.

Pourgerami, A., and K. E. Maskus (1987). "The effects of inflation on the predictability of price changes in Latin America: some estimates and policy implications." *World Development* 15, 287–290.

Pourgerami, A., and K. E. Maskus (1990). "Inflation and its predictability in high-inflation Latin-American countries: some evidence of two competing hypotheses – a research note." *Journal of International Development* 2, 373–379.

Robinson, P. M. (1991). "Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression." *Journal of Econometrics* 47, 67–84.

Scruggs, J. T. (1998). "Resolving the puzzling intertemporal relation between the market risk premium and the conditional market variance: a two-factor approach." *The Journal of Finance* LIII, 575–603.

Sharpe, W. F. (1964). "Capital asset prices: a theory of market equilibrium under conditions of risk." *Journal of Finance* 19, 425–424.

Shin, J. (2005). "Stock returns and volatility in emerging stock markets." *International Journal of Business and Economics* 4, 31–43.

Sims, C. A., J. H. Stock, and M. W. Watson (1990). "Inference in linear time series models with some unit roots." *Econometrica* 58, 113–144.

Taylor, S. (1986). *Modelling financial time series.* Wiley, New York.

Teyssière, G. (1998). "Multivariate long-memory ARCH modelling for high frequency foreign exchange rates." In *Proceedings of the Second High Frequency Data in Finance (HFDF-II) Conference*, Olsen & Associates, Zürich.

Tobin, J. (1958). "Liquidity preference as behavior towards risk." *Review of Economic Studies* 25, 65–86.

Toda, H. Y., and T. Yamamoto (1995). "Statistical inference in vector autoregressions with possibly integrated processes." *Journal of Econometrics* 66, 225–250.

Tsay, W.J., and C. F. Chung (2000). "The spurious regression of fractionally integrated processes." *Journal of Econometrics* 96, 155–182.

Tse, Y. K. (1998). "The conditional heteroskedasticity of the Yen-Dollar exchange rate." *Journal of Applied Econometrics* 13, 49–55.

Ungar, M., and B. Zilberfarb (1993). "Inflation and its unpredictability – theory and empirical evidence." *Journal of Money, Credit and Banking* 25, 709–720.

Van de Geer, S. (2000). *Empirical processes in M-estimation.* Cambridge University Press, Cambridge.

Vilasuso, J. (2001). "Causality tests and conditional heteroscedasticity: Monte Carlo evidence." *Journal of Econometrics* 1, 25–35.

Whitelaw, R. F. (2000). "Stock market risk and return: an equilibrium approach." *The Review of Financial Studies* 13, 521–547.

Zaffaroni, P. (2004). "Stationarity and memory of ARCH(∞) models." *Econometric Theory* 20, 147–160.

Zakoian, J. M. (1994). "Threshold heteroskedastic models." *Journal of Economic Dynamics and Control* 18, 931–995.

Zapata, H. O., and A. N. Rambaldi (1997). "Monte carlo evidence on cointegration and causation." *Oxford Bulletin of Economics and Statistics* 59, 285–293.

Ehrenwörtliche Erklärung

Hiermit erkläre ich, die vorliegende Dissertation selbständig angefertigt und mich keiner anderen als der in ihr angegebenen Hilfsmittel bedient zu haben. Insbesondere sind sämtliche Entlehnungen aus anderen Schriften als solche gekennzeichnet und mit Quellenangaben versehen.

Mannheim, 16. Juni 2006 *Christian Conrad*

Curriculum Vitae
Christian Conrad

Persönliche Daten

Geburtsdatum:	30. Juni 1977
Geburtsort:	Limburg/Lahn
Anschrift:	Salvatorstrasse 4
	8050 Zürich

Ausbildung

10/2002 – 07/2006	Promotion am *Center for Doctoral Studies in Economics & Management* der Universität Mannheim
10/2000 – 09/2001	Studium der Volkswirtschaftslehre, University of York (UK) Abschluß: MSc in Econometrics and Economics
10/1997 – 09/2002	Studium der Volkswirtschaftslehre, Universität Heidelberg Abschluß: Diplom-Volkswirt
07/1988 – 06/1997	Abitur, Tilemannschule, Limburg/Lahn