UNIVERSITY OF MANNHEIM

Mannheim Working Paper Series on
Risk Theory, Portfolio Management and Insurance

No. 186

Tail Risk Hedging and Regime Switching

Markus Huggenberger, Peter Albrecht, Alexandr Pekelis

August 2016

Tail Risk Hedging and Regime Switching

Markus Huggenberger[a,*], Peter Albrecht[a], Alexandr Pekelis[a]

[a] *University of Mannheim, Schloss, 68131 Mannheim, Germany*

This version: August 2016

Abstract: We analyze hedging strategies that minimize tail risk measured by Value-at-Risk (VaR) or Conditional-Value-at-Risk (CVaR). In particular, we derive first-order conditions characterizing VaR- and CVaR-minimal hedging with futures in regime-switching models. Using cross-hedging examples, we theoretically and empirically demonstrate that tail-risk-minimal strategies can noticeably deviate from standard minimum-variance policies in the presence of crash regimes. In such examples, VaR- and CVaR-minimal strategies based on regime-switching models are able to attain additional tail risk reductions, which can be confirmed by nonparametric and extreme-value-theory-based methods. These results imply that the proposed methodology for tail risk management can cut losses during financial crises and reduce capital requirements for institutional investors.

Keywords: Value-at-Risk, Conditional-Value-at-Risk, regime-switching models, elliptical distributions, futures hedging

JEL classifications: G11, G32, C58

* Corresponding author. Tel. +49 621 181 16 79. E-mail: huggenberger@bwl.uni-mannheim.de.

We are grateful for helpful comments from the participants of the Asian Finance Association Annual Meeting 2015, the China International Conference in Finance 2015, the International Finance and Banking Association Conference 2015 and the World Risk and Insurance Economics Congress 2015. The theoretical results presented in this paper constitute a part of the first author's dissertation. This article was previously titled 'VaR- and CVaR-minimal futures hedging: An analytical approach'.

1 Introduction

After maturing into standard tools for risk measurement, especially for setting capital requirements and risk limits, Value-at-Risk (VaR) and Conditional-Value-at-Risk (CVaR) have been increasingly adopted as decision tools for active risk management in financial institutions. Focusing on the latter, the objective of this paper is to develop static futures hedging strategies that minimize tail risk measured by VaR or CVaR.[1] Such strategies might be of special interest for institutional investors, who can lower the utilization of their risk budgets or reduce capital requirements by basing their hedging decisions on the same measure which is used for risk control. In addition, tail-risk-minimal strategies are of general interest if avoiding large losses[2] is given preference over minimizing the overall variance of the position, which is the standard paradigm for futures hedging following Johnson (1960) and Ederington (1979). Hence, tail-risk-minimal hedging is useful for investors who are particularly concerned about performance under extreme market circumstances such as financial crises.

Implementing VaR or CVaR as objectives in portfolio optimization is technically more demanding than solving variance-based problems because these risk measures – in general – depend on the full distribution of the portfolio return and not just on the first two moments. In addition, as compared to pure risk measurement applications, portfolio and hedging decisions require a multivariate model, which narrows down the range of applicable techniques for the calculation of VaR or CVaR. A popular approach is to assume jointly elliptically distributed returns, which implies that the loss distribution – as opposed to the general case – *is* fully characterized by the first two moments and the distribution type. Within this framework, influential portfolio selection studies incorporating (C)VaR objectives or restrictions include Alexander and Baptista (2002, 2004, 2008) as well as Bertsimas et al. (2004).[3] From a pure hedging perspective, this approach is less promising because for elliptical distributions, (C)VaR-minimal hedging strategies deviate from minimum-variance hedges only due to the

[1] We consider VaR due to its importance in regulation and industry applications although it is often criticized for its lack of coherence and for not taking the severity of the highest losses into account.

[2] Thereby, this approach relates to the literature on safety-first investors. See, for example, Arzac and Bawa (1977).

[3] These authors note that a mean-variance-based analysis of VaR and CVaR can be extended beyond the elliptical setup using results like Chebyshev's inequality, which provide upper bounds on these tail risk measures.

impact of expected returns. This is attributable to the following properties of elliptical models: They cannot capture i) univariate asymmetries, ii) differing tail behaviors of their margins and iii) nonlinear dependence, in particular dependence asymmetries. We therefore believe that going beyond the elliptical setup is crucial for hedging tail risk.

Avoiding restrictive modeling assumptions, a number of studies work with nonparametric methods for the derivation of VaR- or CVaR-optimal portfolios or hedging rules (Rockafellar and Uryasev, 2002; Harris and Shen, 2006). In addition, semiparametric (Cao et al., 2010; Hilal et al., 2011; Barbi and Romagnoli, 2014) and very flexible multivariate parametric models based on copulas (Patton, 2004) are applied in the risk and portfolio management literature, focusing on non-normalities. However, such models usually do not allow for a tractable analytic characterization of the resulting aggregated return distribution and therefore rely on a combination of simulation and numerical optimization methods to derive tail-risk-optimal policies. With VaR, this approach can cause numerical problems due to the nonconvexity of this risk measure. Furthermore, purely nonparametric methods might suffer from high estimation risk caused by a small number of tail observations. This problem can be adressed by introducing worst-case modifications of VaR and CVaR and using techniques from robust optimization, which are reviewed in Fabozzi et al. (2010).

In contrast to this literature, we propose a more flexible but still tractable parametric modeling approach for the minimization of the original VaR or CVaR. In particular, we analyze tail risk management with regime-switching (RS) models. This approach naturally incorporates the presence of crash regimes into hedging decisions and, more generally, it allows to account for the relevant nonelliptical features mentioned above.

RS models were first introduced by Hamilton (1989) in a univariate setting and applied to portfolio choice by Ang and Bekaert (2002). Assuming normally or t-distributed components, multivariate RS models allow for the analytic derivation of the aggregate return distribution but can at the same time reproduce flexible univariate distribution shapes and asymmetric dependence structures. Their capability for tail risk *measurement* has been emphasized by Billio and Pelizzon (2000) as well as Guidolin and Timmermann (2006). The flexible shape of RS models has also been utilized to solve portfolio selection problems with skewness and kurtosis preferences (Guidolin and Timmermann, 2008). Moreover, various studies exploit the temporal dependencies implied by the models to construct conditional strategies within a variance-based setup (Alizadeh et al., 2008). Chang (2010) analyzes univariate VaR-minimal hedging using, however, a numerical search algorithm to determine the optimal policy. Related to our work is also Buckley et al. (2008), who demonstrate the usefulness of multivariate normal mixture distributions for lower-partial-moment-based portfolio optimization.

To the best of our knowledge, we are the first to present an analytical characterization of VaR- and CVaR-minimal hedging rules that applies to RS models. Our theoretical contribution is as follows: First, we use results on quantile derivatives from Hong (2009) and Hong and Liu (2009) to derive first-order conditions for tail-risk-minimal hedging strategies that cover general multivariate density models under relatively weak continuity and differentiability assumptions. Second, we provide the specific form of these conditions for finite mixture distributions with elliptical components. Third, we discuss the implementation of our strategies for mixtures[4] and RS processes with normally and t-distributed components. Based on these analytical results, we provide a stylized example showing that crash regimes can produce substantial differences between minimum-variance and tail-risk-optimal hedging strategies. Furthermore, we provide an upper bound for these differences in the special case of standard elliptical distributions.

In the empirical part of this paper, we present cross-hedging examples with well-known market indices demonstrating the in- and out-of-sample benefits of tail-risk-minimal hedging based on RS models. We estimate multivariate RS models with Gaussian conditional distributions, and confirm that they produce reliable tail risk estimates in our examples. Then, we compare hedging strategies derived from these models and some benchmark specifications. First, we find that in- as well as out-of-sample non-negligible additional tail risk reductions can be attained by switching from minimum-variance to VaR- and CVaR-based hedging in our examples. The relative risk reductions over the minimum-variance benchmark reach 18% with conditional as well as unconditional hedging strategies. Second, we compare the out-of-sample performance of tail-risk-optimal hedges derived from different econometric specifications and

[4] A technically similar result has recently been derived by Litzenberger and Modest (2010), who analyze a mixture-based stress testing framework for portfolio selection with hedge funds.

find promising results for the RS approach. In this context, we also provide a small simulation study on the statistical significance and robustness of the reported benefits. Finally, we illustrate our methodology with a composite hedging application and present a selection of robustness checks. The tail risk reductions in our examples can be confirmed – independent from our model – by standard univariate nonparametric or extreme-value-theory-based estimators, which is especially important if such procedures are used to set the capital requirements or risk limits for the optimized positions.

The remainder of our paper is structured as follows: In Section 2, we give a formal problem statement and derive our most general characterization of tail-risk-minimal hedging rules. Section 3 contains the derivation of first-order conditions for hedging with mixtures, the application of these results to RS models and a stylized example for the differences between tail-risk-minimal and minimum-variance hedging. In Section 4, we present our empirical findings and robustness checks. Section 5 concludes. We provide omitted proofs in the Appendix.

2 Tail Risk Hedging with Quantile Derivatives

2.1 Problem Statement

We analyze a multivariate static hedging problem over a fixed investment horizon $[t, t+1]$. The portfolio we want to hedge consists of N positions – typically in the spot market. The discrete returns of these positions over $[t, t+1]$ are denoted by $R_{S,i},\ i = 1, \ldots, N$. The corresponding portfolio weights are given by $w_i = \frac{v_{S,i}}{v_P}$, $i = 1, \ldots, N$, where $v_{S,i}$ is the value of the ith position in t and $v_P = \sum_{i=1}^{N} v_{S,i}$.

Furthermore, we assume that M futures instruments are available to temporarily reduce the risk of the spot positions. The relative price changes of these instruments will also be described by their discrete returns $R_{F,j},\ j = 1, \ldots, M$.[5] Abstracting from initial margins, futures positions will have no effect on the portfolio value in t. We therefore define hedging weights h_j relative to v_P, i.e., $h_j = \frac{v_{F,j}}{v_P}$, $j = 1, \ldots, M$, where $v_{F,j}$ is the nominal value of a short position in the jth futures contract. Collecting the returns and the weights in column

[5] Denoting the price of the jth futures by $F_{t,j}$, we use the usual return definition $R_{F,j} = \frac{F_{t+1,j} - F_{t,j}}{F_{t,j}}$, although futures do not require an initial investment of their nominal value. An alternative return definition for futures uses the spot price in the denominator (Figlewski, 1984).

vectors $\boldsymbol{R}_S = (R_{S,i})$, $\boldsymbol{R}_F = (R_{F,j})$, $\boldsymbol{w} = (w_i)$ and $\boldsymbol{h} = (h_j)$, we obtain for the return of the hedged (net) position $R_H(\boldsymbol{h}) := R_H = \boldsymbol{w}' \cdot \boldsymbol{R}_S - \boldsymbol{h}' \cdot \boldsymbol{R}_F$. Thus, the percentage loss of the hedged position is given by

$$L_H(\boldsymbol{h}) := L_H := -\boldsymbol{w}' \cdot \boldsymbol{R}_S + \boldsymbol{h}' \cdot \boldsymbol{R}_F. \tag{1}$$

The standard approach following Johnson (1960) and Ederington (1979) to determine optimal hedging weights is to minimize the variance of this loss variable or, equivalently, the variance of the return, i.e., to solve $\min_{\boldsymbol{h}\in\mathbb{R}^M} \operatorname{var}[L_H(\boldsymbol{h})] = \min_{\boldsymbol{h}\in\mathbb{R}^M} \operatorname{var}[R_H(\boldsymbol{h})]$, which requires that $R_{S,i} \in \mathcal{L}^2$ and $R_{F,j} \in \mathcal{L}^2$ for $i = 1, \dots, N$ and $j = 1, \dots, M$. It is easy to show that the hedging policy $\boldsymbol{h}^*_{\text{var}}$ solving this problem is given by

$$\boldsymbol{h}^*_{\text{var}} = (\operatorname{cov}[\boldsymbol{R}_F])^{-1} \cdot \operatorname{cov}[\boldsymbol{R}_F, \boldsymbol{R}_S] \cdot \boldsymbol{w}. \tag{2}$$

Much of the literature on futures hedging is centered around implementing dynamic specifications for the covariance terms in (2) that are conditional on the filtration $\mathcal{F}_t$ generated by the return process. In fact, many studies investigate the performance of time-varying conditional hedging strategies based on multivariate GARCH models following Baillie and Myers (1991).

In contrast, our focus lies on hedging strategies that minimize the tail risk or the corresponding capital requirement, which are usually measured in terms of VaR and CVaR. For a simple definition of these risk measures, we assume that $L_H \in \mathcal{L}^1$ and that it has a positive density. Then, VaR_α and CVaR_α at the confidence level $1 - \alpha$ with $\alpha \in (0, 1)$ satisfy

$$\text{P}(L_H \leq \text{VaR}_\alpha[L_H]) = 1 - \alpha \quad \text{and} \quad \text{CVaR}_\alpha[L_H] = \mathbb{E}[L_H \mid L_H \geq \text{VaR}_\alpha[L_H]]\,. \tag{3}$$

Accordingly, VaR_α can be understood as the loss value, which is not exceeded with a probability of $1 - \alpha$. Formally, VaR_α simply corresponds to the $(1 - \alpha)$-quantile $q_{1-\alpha}[L_H]$ of the loss distribution.[6] CVaR_α is the expected loss in the worst $100 \cdot \alpha\%$ of the cases. Comparing both measures, VaR_α is still dominant in industry applications, although CVaR_α is preferable

[6] More generally, VaR_α is usually defined as the *lower* $(1 - \alpha)$-quantile, i.e., $\text{VaR}_\alpha[L_H] = \inf\{l \in \mathbb{R} | \text{P}(L_H \leq l) \geq 1 - \alpha\}$. This definition will be used in Section 4 for the analysis of empirical distributions.

from an axiomatic point of view as a coherent risk measure in the sense of Artzner et al. (1999).[7] Moreover, VaR_α might be questionable if the aim is to avoid large losses because it does not consider the extent of losses in the very tail of the distribution. However, the choice between VaR_α and CVaR_α remains a matter of debate in academia and industry.[8] We therefore consider both measures in our analysis. Writing these risk measures as functions of the hedging weights, i.e., $v_\alpha(\boldsymbol{h}) := \mathrm{VaR}_\alpha[L_H(\boldsymbol{h})]$ and $c_\alpha(\boldsymbol{h}) := \mathrm{CVaR}_\alpha[L_H(\boldsymbol{h})]$, we analyze

$$\min_{\boldsymbol{h}\in\mathbb{R}^M} v_\alpha(\boldsymbol{h}) = \min_{\boldsymbol{h}\in\mathbb{R}^M} \mathrm{VaR}_\alpha[L_H(\boldsymbol{h})], \tag{4}$$

$$\min_{\boldsymbol{h}\in\mathbb{R}^M} c_\alpha(\boldsymbol{h}) = \min_{\boldsymbol{h}\in\mathbb{R}^M} \mathrm{CVaR}_\alpha[L_H(\boldsymbol{h})]. \tag{5}$$

Univariate versions of these problems have recently been analyzed by Harris and Shen (2006) and Cao et al. (2010) in a non- and semiparametric framework. Furthermore, Barbi and Romagnoli (2014) analyzed tail-risk-minimal hedging strategies with copula models. More often, similar problems have been studied in a portfolio selection context. In particular, the sample-based approach of Rockafellar and Uryasev (2002), which allows to solve problems of the second type using LP techniques, has gained a lot of attention. Although these studies focus on the unconditional distribution, we emphasize that (4) and (5) can of course also be applied conditionally on $\mathcal{F}_t$. For a general discussion of conditional quantile risk *measurement*, we refer to McNeil and Frey (2000). Hilal et al. (2011) present an application to CVaR_α hedging using an elaborate combination of time series modeling and multivariate extreme value theory. Although we do not systematically assess conditional versus unconditional risk modeling here, some of the results presented in our empirical section might be of relevance for this problem.

---

[7] Without the assumption of a continuous loss distribution, this requires a CVaR_α definition that takes the possibility of a point mass at the VaR_α into account (Rockafellar and Uryasev, 2002). We will adopt such a definition when working with empirical distributions in Section 4, see (31).

[8] See Kellner and Rösch (2016) for a recent contribution to this debate and for a brief review of the literature. There has been a renewed interest in CVaR_α, also known as Expected Shortfall, following the proposal of the Basel Committee on Banking Supervision to use this measure instead of VaR_α for market risk measurement.

2.2 A General Solution

Complementing the above-mentioned results on non- and semiparametric VaR_α and CVaR_α hedging, we are interested in analytic characterizations of the solutions to (4) and (5). These can be derived under the following regularity conditions on the distribution of $(\boldsymbol{R}_S', \boldsymbol{R}_F')'$, adapted from Hong (2009) and Hong and Liu (2009).[9]

(R1) $R_{S,i} \in \mathcal{L}^1$ and $R_{F,j} \in \mathcal{L}^1$ for $i = 1, \ldots, N$ and $j = 1, \ldots, M$.

(R2) For all $\boldsymbol{h} \in \mathbb{R}^M$, $L_H(\boldsymbol{h})$ has a continuous and strictly positive density. Moreover, for all h_j, $j = 1, \ldots, M$, the partial derivative of $F_{L_H}(l; \boldsymbol{h}) = \mathrm{P}(L_H(\boldsymbol{h}) \leq l)$ with respect to h_j exists and is continuous in l and h_j.

(R3) For all $j = 1, \ldots, M$, the conditional expectations $\mathbb{E}[R_{F,j} \mid L_H = l]$ are continuous as functions of l.

Assumption (R1) is obviously weaker than the corresponding integrability requirements needed for the variance-based approach. However, Assumptions (R2) and (R3) define some additional continuity and differentiability conditions. Note that due to (R1) and (R2) the requirements for using the simple VaR_α and CVaR_α representations from (3) are satisfied.

We are now ready to state a first analytic characterization of tail-risk-based hedging strategies.

Proposition 1 *Under (R1) - (R3),* VaR_α*- and* CVaR_α*-minimal hedging strategies* $\boldsymbol{h}^*_{\mathrm{VaR}}$ *and* $\boldsymbol{h}^*_{\mathrm{CVaR}}$*, i.e., solutions to* (4) *and* (5)*, satisfy*

$$\mathbb{E}[\boldsymbol{R}_F \mid L_H(\boldsymbol{h}^*_{\mathrm{VaR}}) = v_\alpha(\boldsymbol{h}^*_{\mathrm{VaR}})] = \boldsymbol{0}_M, \tag{6}$$

$$\mathbb{E}[\boldsymbol{R}_F \mid L_H(\boldsymbol{h}^*_{\mathrm{CVaR}}) \geq v_\alpha(\boldsymbol{h}^*_{\mathrm{CVaR}})] = \boldsymbol{0}_M. \tag{7}$$

This characterization is an application of results on quantile derivatives to the hedging problem. In particular, (6) and (7) follow as FOCs of (4) and (5) from Theorem 2 in Hong (2009)

[9] See the proof of Proposition 1 for the relation between the assumptions given here and the original statements made in Hong (2009) and Hong and Liu (2009).

and Theorem 3.1 in Hong and Liu (2009). Some technical details of this reasoning can be found in the Appendix.[10]

Note that Proposition 1 makes no statement on the existence of optimal strategies. As already observed in Alexander and Baptista (2004) for the case of portfolio selection strategies, it is possible that VaR_α and CVaR_α minimizations have no solutions even with normally distributed returns. Moreover, there is an important difference between using v_α and c_α as objective functions. Whereas the c_α-FOC (7) is only fulfilled by the global minimizer of (5), the v_α-FOC (6) might also be satisfied by other stationary points. This is due to the fact that CVaR_α is in general a coherent risk measure, which implies that (5) is always a convex optimization problem. VaR_α will, however, only be subadditive and convex under specific combinations of distributional assumptions on $(\boldsymbol{R}_S', \boldsymbol{R}_F')'$ and confidence levels. In such cases, (6) will uniquely characterize the global VaR_α-minimal hedging vector (if such a strategy exists).

We note that it might be interesting to analyze the tail risk of the demeaned loss variables instead of the losses themselves. Therefore, we define $\mathrm{MVaR}_\alpha[L_H] := \mathrm{VaR}_\alpha[L_H - \mathbb{E}[L_H]]$ and $\mathrm{MCVaR}_\alpha[L_H] := \mathrm{CVaR}_\alpha[L_H - \mathbb{E}[L_H]]$. By construction, MVaR_α and MCVaR_α do not allow reducing the risk of the position by increasing its expected return. Under (R1) - (R3), FOCs for the strategies that minimize these risk measures are obtained by subtracting $\mathbb{E}[\boldsymbol{R}_F]$ from the left hand side of (6) and (7).[11]

Proposition 1 applies to a wide range of continuous return distributions because the conditions in (R1) - (R3) are rather weak. However, at this level of generality, we cannot provide explicit representations for the conditional expectations in Equations (6) and (7). We therefore analyze more specific distributional assumptions in the following section.

[10] Earlier results on quantile derivatives, for example Gourieroux et al. (2000) or Scaillet (2004), could also be applied to obtain (6) and (7).

[11] This follows from $\frac{\partial}{\partial \boldsymbol{h}}\mathbb{E}[L_H(\boldsymbol{h})] = \mathbb{E}[\boldsymbol{R}_F]$ and the translation invariance property of VaR_α and CVaR_α.

3 Tail Risk Hedging with Mixture Distributions

3.1 Mixtures of Elliptical Distributions

The main idea in this section is to combine the econometric flexibility of mixture modeling with the analytic tractability of elliptical distributions. We will derive explicit forms of the FOCs in Proposition 1 under the assumption that the joint distribution of $\boldsymbol{R} = (\boldsymbol{R}'_S, \boldsymbol{R}'_F)'$ is a multivariate finite mixture with elliptical components.

First, we briefly recall a density-based definition of elliptical distributions, which largely corresponds to definition c) in Owen and Rabinovitch (1983). Let $\boldsymbol{\mu}$ be a real-valued $P \times 1$ vector and let $\boldsymbol{\Sigma}$ denote a symmetric, positive definite $P \times P$ matrix for $P \in \mathbb{N}$. A $P \times 1$ random vector $\boldsymbol{Y}$ with a density $f_{\boldsymbol{\mu},\boldsymbol{\Sigma},g}$ follows an elliptical distribution if this density is of the form

$$f_{\boldsymbol{\mu},\boldsymbol{\Sigma},g}(\boldsymbol{y}) = \det(\boldsymbol{\Sigma})^{-\frac{1}{2}} \; g_P\big((\boldsymbol{y}-\boldsymbol{\mu})' \cdot \boldsymbol{\Sigma}^{-1} \cdot (\boldsymbol{y}-\boldsymbol{\mu})\big)\,, \tag{8}$$

where g_P is a non-negative scalar function on $\mathbb{R}$. This function is referred to as density generator. To define a distribution over several dimensions, a collection of generators $g = (g_P)_{P\in\mathbb{N}}$ is needed because g_P is parameterized by the dimension of $\boldsymbol{Y}$. We use the notation $\boldsymbol{Y} \sim \mathcal{E}_P(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ if $\boldsymbol{Y}$ has an elliptical distribution with parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and the generator (family) g. The widespread use of this model is partly explained by its favorable distributional properties. An example being particularly relevant in a portfolio context is the behavior under linear transformations (Owen and Rabinovitch, 1983, P.1).[12]

Second, we build on the following definition of finite mixture models. $\boldsymbol{Y}$ has a mixture distribution with component densities f_k, $k = 1, \ldots, K$ and component weights π_k, $k = 1, \ldots, K$, $\sum_{k=1}^{K} \pi_k = 1$ if its density is of the form

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \sum_{k=1}^{K} \pi_k f_k(\boldsymbol{y}). \tag{9}$$

As we will detail later, this structure allows for very flexible univariate and multivariate distribution shapes even if relatively simple components like normal distributions are com-

[12] For a full account of elliptical distributions, we refer to Fang et al. (1990).

bined.[13] Let us for the moment just note that the mixture framework can be motivated by introducing an unobserved state variable S with values in $\{1, \ldots, K\}$, which is often assumed to describe the state of the relevant market. If the distribution of S is given by $\mathrm{P}(S = k) = \pi_k$ and the component densities of the mixture correspond to the conditional distributions of $\boldsymbol{Y}$ given $S = k$, the structure in (9) is obtained from the law of total probability.

Combining (8) with (9) and adding the requirement that the density is strictly positive, we obtain the following assumption:

(M1) The vector $\boldsymbol{R} = (\boldsymbol{R}_S', \boldsymbol{R}_F')'$ follows a multivariate K-state mixture of elliptical distributions with continuous and strictly positive density generators $g_{N+M,k}$, i.e., its density is of the form

$$f_{\boldsymbol{R}}(\boldsymbol{r}) = \sum_{k=1}^{K} \pi_k \ \det(\boldsymbol{\Sigma}_k)^{-\frac{1}{2}} \ g_{N+M,k} \ \big((\boldsymbol{r} - \boldsymbol{\mu}_k)' \cdot \boldsymbol{\Sigma}_k^{-1} \cdot (\boldsymbol{r} - \boldsymbol{\mu}_k)\big) \tag{10}$$

for $\pi_k \in (0, 1)$, $\sum_{k=1}^{K} \pi_k = 1$, $\boldsymbol{\mu}_k \in \mathbb{R}^{N+M}$ and positive definite $(N + M) \times (N + M)$ covariance matrices $\boldsymbol{\Sigma}_k$.

Using the state variable approach described above, we can give the following equivalent formulation of (M1):

(M1') $\boldsymbol{R}|S = k \sim \mathcal{E}_{N+M}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, g_k)$ for $k = 1, \ldots, K$ with continuous, strictly positive density generators $g_{N+M,k}$ and $\mathrm{P}(S = k) = \pi_k$.

This setting obviously includes popular modeling choices like mixtures of multivariate normals or multivariate t-distributions.

We first provide the solution to the minimum-variance hedging problem for (M1) with the additional assumption that all elements of $\boldsymbol{R}$ are in $\mathcal{L}^2$. Therefore, note that for $\boldsymbol{Y} \sim \mathcal{E}_N(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$, it holds that $\mathbb{E}[\boldsymbol{Y}] = \boldsymbol{\mu}$ and $\mathrm{cov}[\boldsymbol{Y}] = c_g \cdot \boldsymbol{\Sigma}$, which under (M1) implies

$$\mathbb{E}[\boldsymbol{R}] = \sum_{k=1}^{K} \pi_k \ \boldsymbol{\mu}_k \quad \text{and} \quad \mathrm{cov}[\boldsymbol{R}] = \sum_{k=1}^{K} \pi_k \left[c_{g_k} \boldsymbol{\Sigma}_k + \boldsymbol{\mu}_k \cdot \boldsymbol{\mu}_k'\right] - \mathbb{E}[\boldsymbol{R}] \cdot \mathbb{E}\big[\boldsymbol{R}'\big]\,. \tag{11}$$

[13] See, for example, McLachlan and Peel (2000) for a comprehensive discussion of the properties of this modeling approach and illustrations of its flexibility.

Using $\boldsymbol{\mu}_k = \begin{pmatrix} \boldsymbol{\mu}_{S,k} \\ \boldsymbol{\mu}_{F,k} \end{pmatrix}$ and $\boldsymbol{\Sigma}_k = \begin{pmatrix} \boldsymbol{\Sigma}_{S,k} & \boldsymbol{\Sigma}_{SF,k} \\ \boldsymbol{\Sigma}'_{SF,k} & \boldsymbol{\Sigma}_{F,k} \end{pmatrix}$, we obtain from (2) and (11) for the traditional minimum-variance hedging weights[14]

$$\boldsymbol{h}^*_{\text{var}} = \left[\sum_{k=1}^{K} \pi_k \left[c_{g_k} \boldsymbol{\Sigma}_{F,k} + \boldsymbol{\mu}_{F,k} \cdot \boldsymbol{\mu}'_{F,k} \right] - \sum_{k=1}^{K} \pi_k \ \boldsymbol{\mu}_{F,k} \cdot \sum_{k=1}^{K} \pi_k \ \boldsymbol{\mu}'_{F,k} \right]^{-1} \cdot \left[\sum_{k=1}^{K} \pi_k \cdot \left[c_{g_k} \boldsymbol{\Sigma}'_{SF,k} + \boldsymbol{\mu}_{F,k} \cdot \boldsymbol{\mu}'_{S,k} \right] - \sum_{k=1}^{K} \pi_k \ \boldsymbol{\mu}_{F,k} \cdot \sum_{k=1}^{K} \pi_k \ \boldsymbol{\mu}'_{S,k} \right] \cdot \boldsymbol{w}. \quad (12)$$

For the analysis of tail risk hedging under (M1), we first observe that the distribution of the portfolio loss is also a mixture with elliptical components, i.e.,

$$L_H(\boldsymbol{h}) \mid S = k \sim \mathcal{E}_1(\mu_{L,k},\, \sigma^2_{L,k},\, g_k), \quad (13)$$

where

$$\mu_{L,k} := \mu_{L,k}(\boldsymbol{h}) = -\boldsymbol{w}' \cdot \boldsymbol{\mu}_{S,k} + \boldsymbol{h}' \cdot \boldsymbol{\mu}_{F,k}, \quad (14)$$

$$\sigma^2_{L,k} := \sigma^2_{L,k}(\boldsymbol{h}) = \boldsymbol{w}' \cdot \boldsymbol{\Sigma}_{S,k} \cdot \boldsymbol{w} - 2\ \boldsymbol{w}' \cdot \boldsymbol{\Sigma}_{SF,k} \cdot \boldsymbol{h} + \boldsymbol{h}' \cdot \boldsymbol{\Sigma}_{F,k} \cdot \boldsymbol{h}, \quad (15)$$

which follows from the behavior of elliptical distributions under linear transformations. We write $f_{L,k} := f_{L_H|S=k}$ and $F_{L,k} := F_{L_H|S=k}$ for the corresponding component pdfs and cdfs. According to (8) and (9), the component densities and the unconditional density $f_L := f_{L_H}$ satisfy

$$f_{L,k}(l) = \sigma^{-1}_{L,k} \cdot g_{1,k} \left(\frac{(l - \mu_{L,k})^2}{\sigma^2_{L,k}} \right) \quad \text{and} \quad f_L(l) = \sum_{k=1}^{K} \pi_k \ f_{L,k}(l). \quad (16)$$

The tail risk measures that we analyze are given by

$$1 - \alpha = \sum_{k=1}^{K} \pi_k F_{L,k}(v_\alpha(\boldsymbol{h})), \quad (17)$$

$$c_\alpha(\boldsymbol{h}) = \frac{1}{\alpha} \sum_{k=1}^{K} \pi_k \ \mathbb{E}[L_H \ \mathbb{1}(L_H \geq v_\alpha(\boldsymbol{h})) \mid S = k]\,. \quad (18)$$

[14] This corresponds to the strategy analyzed by Alizadeh et al. (2008) in a univariate, two-state setting.

The simple VaR_α characterization in (17) is sufficient due to the positivity of the density generators. Note that by introducing $Z_k \sim \mathcal{E}_1(0,1,g_k)$ for $k=1,\dots,K$ and setting

$$z_k(\boldsymbol{h}) := \frac{v_\alpha(\boldsymbol{h}) - \mu_{L,k}(\boldsymbol{h})}{\sigma_{L,k}(\boldsymbol{h})}, \qquad \lambda_k(\boldsymbol{h}) := \mathbb{E}[Z_k \mid Z_k \geq z_k(\boldsymbol{h})]\,, \tag{19}$$

we can rewrite (18) in terms of the location and scale parameters of the mixture as

$$c_\alpha(\boldsymbol{h}) = \frac{1}{\alpha}\sum_{k=1}^{K} \pi_k\ (1 - F_{L,k}(v_\alpha(\boldsymbol{h})))\ \left[\mu_{L,k}(\boldsymbol{h}) + \sigma_{L,k}(\boldsymbol{h})\ \lambda_k(\boldsymbol{h})\right]. \tag{20}$$

Given $v_\alpha(\boldsymbol{h})$, (20) can usually be evaluated explicitly for specific density generators $k = 1,\dots,K$. In contrast, the implicit VaR_α definition in (17) can, even in basic cases like normally distributed components, not be written explicitly. Therefore, the derivation of FOCs that characterize minimum-VaR_α and minimum-CVaR_α hedging vectors is not straightforward.[15] However, applying Proposition 1, we are able to obtain such conditions, which we present in the following Theorem.

Theorem 1 *If (R1) and (M1) hold, the* VaR_α*-minimal hedging strategy* $\boldsymbol{h}^*_{\mathrm{VaR}}$ *solves*

$$\sum_{k=1}^{K} \frac{\pi_k f_{L,k}(v_\alpha(\boldsymbol{h}^*_{\mathrm{VaR}}))}{f_L(v_\alpha(\boldsymbol{h}^*_{\mathrm{VaR}}))}\left[\boldsymbol{\mu}_{F,k} + \frac{\boldsymbol{\Sigma}_{FL,k}(\boldsymbol{h}^*_{\mathrm{VaR}})}{\sigma_{L,k}(\boldsymbol{h}^*_{\mathrm{VaR}})} z_k(\boldsymbol{h}^*_{\mathrm{VaR}})\right] = \mathbf{0}_M, \tag{21}$$

where $\boldsymbol{\Sigma}_{FL,k}(\boldsymbol{h}) = -\boldsymbol{\Sigma}'_{SF,k}\cdot\boldsymbol{w} + \boldsymbol{\Sigma}_{F,k}\cdot\boldsymbol{h}$*. Under the same conditions, the* CVaR_α*-minimal hedging strategy* $\boldsymbol{h}^*_{\mathrm{CVaR}_\alpha}$ *satisfies*

$$\sum_{k=1}^{K} \frac{\pi_k (1 - F_{L,k}(v_\alpha(\boldsymbol{h}^*_{\mathrm{CVaR}})))}{\alpha}\left[\boldsymbol{\mu}_{F,k} + \frac{\boldsymbol{\Sigma}_{FL,k}(\boldsymbol{h}^*_{\mathrm{CVaR}_\alpha})}{\sigma_{L,k}(\boldsymbol{h}^*_{\mathrm{CVaR}_\alpha})}\lambda_k(\boldsymbol{h}^*_{\mathrm{CVaR}_\alpha})\right] = \mathbf{0}_M. \tag{22}$$

See the Appendix for a proof of Theorem 1. Note that the conditions in (21) and (22) could be multiplied by $f_L(v_\alpha(\boldsymbol{h}^*_{\mathrm{VaR}}))$ and α, respectively. We omitted this simplification to emphasize that the weights of the summands correspond to modified state probabilities implied by Bayes' Theorem. For the case of the VaR_α-minimal strategy it, for example, holds

[15] Litzenberger and Modest (2010) present an alternative reasoning for mixtures of normal distributions that relies on differentiating the implicit VaR_α definition in (17).

that

$$\mathrm{P}(S=k|L_H=v_\alpha(\boldsymbol{h}^*_{\mathrm{VaR}}))=\frac{\mathrm{P}(S=k)\ f_{L,k}(v_\alpha(\boldsymbol{h}^*_{\mathrm{VaR}}))}{\sum_{j=1}^K \mathrm{P}(S=j)\ f_{L,j}(v_\alpha(\boldsymbol{h}^*_{\mathrm{VaR}}))}=\frac{\pi_k\ f_{L,k}(v_\alpha(\boldsymbol{h}^*_{\mathrm{VaR}}))}{f_L(v_\alpha(\boldsymbol{h}^*_{\mathrm{VaR}}))}. \tag{23}$$

The corresponding MVaR_α- and MCVaR_α-minimal strategies are obtained by subtracting $\mathbb{E}[\boldsymbol{R}_F]=\sum_{k=1}^K \pi_k\ \boldsymbol{\mu}_{F,k}$.

Of course, Theorem 1 can also be used to derive VaR_α- and CVaR_α-minimal hedging strategies for the special case $K=1$, i.e. for simple multivariate elliptical distributions. We provide a Corollary with the corresponding FOCs in the online appendix.[16] In particular, these results imply that tail-risk-minimal strategies are identical to the minimum-variance approach if either $\mathbb{E}[\boldsymbol{R}_F]=\boldsymbol{0}_M$ or the demeaned risk measures MVaR_α and MCVaR_α are used as objective functions. This parallels a well known result from portfolio selection (Embrechts et al., 2002, Theorem 1) and emphasizes that tail-risk-minimal and minimum-variance strategies only differ due to the impact of expected returns in the elliptical case. We, moreover, provide a formal analysis of $K=N=M=1$, for which tail-risk-minimal hedging strategies and the resulting tail risk values can be characterized fully explicitly. For this case, we show that $\mathrm{VaR}_\alpha(h^*_{\mathrm{var}})-\mathrm{VaR}_\alpha(h^*_{\mathrm{VaR}})\le b$ and $\mathrm{CVaR}_\alpha(h^*_{\mathrm{var}})-\mathrm{CVaR}_\alpha(h^*_{\mathrm{CVaR}})\le b$ with[17]

$$b=|\mathbb{E}[R_F]|\cdot\sqrt{\frac{\mathrm{var}\,[R_S]}{\mathrm{var}\,[R_F]}\cdot(1-\mathrm{corr}[R_F,R_S]^2)}. \tag{24}$$

This confirms the importance of the mean return for tail risk hedging to be beneficial in an elliptical framework and furthermore shows that a non-negligible level of basis risk is required.

These results are not surprising given that elliptical return models cannot capture asymmetries, which might be important sources of differences between tail-risk-minimal and variance-based hedging. Equally important is that – although elliptical models allow for heavy tailed marginals – the heaviness of tails is determined by the density generator, for example by the degree of freedom parameter, and is therefore not influenced by the hedging weights. At this point, there is a crucial difference between this simple, restricted model on the one hand and

[16] In contrast to the mixture case, these results could also be derived from the explicit VaR_α and CVaR_α expressions available in this case, without relying on Proposition 1.

[17] In contrast to this upper bound, the exact differences, which are provided in the online appendix, additionally depend on the significance level and the choice of the tail risk measure.

the full mixture approach on the other hand, which we will illustrate with the examples at the end of this section and in the empirical part of this paper.

3.2 Regime-Switching Models

In this subsection, we discuss the application of Theorem 1 for the regime-switching approach introduced by Hamilton (1989). Therefore, we extend the setting provided at the beginning of Section 2 to a time series context by introducing a discrete-time return process $(\boldsymbol{R}_t)_{t\in\mathbb{N}}$ and a state process $(S_t)_{t\in\mathbb{N}}$. The latter is assumed to be a time-homogeneous Markov chain with state space $\{1,\ldots,K\}$ and transition matrix $\boldsymbol{Q} = (q_{ij})_{i,j=1,\ldots,K}$, i.e. $\mathrm{P}(S_{t+1} = j|S_t = i) = q_{ij}$ for $i,j = 1,\ldots,K$ and $t \in \mathbb{N}$. Under the additional assumptions that the Markov chain is aperiodic and irreducible, it has a unique invariant (ergodic) distribution $\boldsymbol{\pi}^e = (\pi_k^e)_{k=1,\ldots,K}$. Finally, assuming that $(S_t)_{t\in\mathbb{N}}$ starts from this distribution implies that the model is stationary with $\mathrm{P}(S_t = k) = \pi_k^e$ for all $t \in \mathbb{N}$. The (conditional) distribution of the return vector $\boldsymbol{R}_{t+1}$ is assumed to be given by (M1), replacing the state variable S by S_{t+1}, i.e., $\boldsymbol{R}_{t+1}|S_{t+1} = k \sim \mathcal{E}_{N+M}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, g_k)$.

Maintaining the assumption that $(S_t)_{t\in\mathbb{N}}$ is unobservable, our hedging decisions must rely on the (marginal) distribution of $\boldsymbol{R}_{t+1}$, which according to (M1) exhibits a mixture structure. Due to the temporal dependence introduced by $(S_t)_{t\in\mathbb{N}}$, we have to distinguish two important cases for the component weights. An unconditional hedging strategy would rely on the stationary distribution of $(S_t)_{t\in\mathbb{N}}$. It would thus use $\boldsymbol{\pi}^e$ to weight the distribution components. A conditional approach would infer predictive weights $\mathrm{P}(S_{t+1} = k|\boldsymbol{R}_t,\ldots,\boldsymbol{R}_1)$ from the history of the return process, which can be recursively obtained using the Hamilton filter (Hamilton, 1989). In both cases, Theorem 1 can be applied to obtain VaR_α- and CVaR_α-minimal strategies.

A standard approach in mixture and RS modeling is to assume Gaussian component densities. Then, all components have the same density generator $g(s) = (2\pi)^{-P/2}\exp(-1/2\ s)$ and the Z_k in (19) are all standard normally distributed. This comparatively simple setup already allows for very flexible univariate distribution shapes (Timmermann, 2000), and it can reproduce asymmetric exceedance correlations as shown, for example, by Ang and Bekaert

(2002). For this setup, tail risk measures and the corresponding FOCs from Theorem 1 can be implemented with $F_{L,k}(v_\alpha(\boldsymbol{h})) = \Phi(z_k(\boldsymbol{h}))$ and

$$\lambda_k(\boldsymbol{h}) = \mathbb{E}[Z \mid Z \geq z_k] = \frac{\varphi(z_k(\boldsymbol{h}))}{1 - \Phi(z_k(\boldsymbol{h}))}, \tag{25}$$

where φ is the pdf and Φ is the cdf of a standard normally distributed random variable Z.

Although the mixture of normals approach already allows for a high level of econometric flexibility, it might have two weaknesses in the scope of tail risk modeling. First, the marginal distributions show exponentially decaying tails. Second, the dependence structure implied by a finite mixture of multivariate normals is not capable of describing asymptotic tail dependence (Garcia and Tsafack, 2011). To overcome these problems, tail risk hedging can be implemented with mixtures of multivariate t-distributions. The density generator of the standardized t-distribution is

$$g_{P,k}(s;\nu_k) = \frac{\Gamma(\frac{(P+\nu_k)}{2})}{((\nu_k-2)\ \pi)^{\frac{P}{2}}\ \Gamma(\frac{\nu_k}{2})} \left(1 + \frac{s}{\nu_k - 2}\right)^{-\frac{P+\nu_k}{2}} \qquad \text{for } \nu_k > 2. \tag{26}$$

The degrees of freedom parameter ν_k determines the heaviness of the tails of the mixture components. It corresponds to the tail index of the distribution, so that we need $\nu_k > 2$ for the standardized version of the distribution to be well defined. Denoting the resulting pdf and cdf by f_t^* and F_t^*, we obtain $F_{L,k}(v_\alpha(\boldsymbol{h})) = F_t^*(z_k(\boldsymbol{h});\nu_k)$ and

$$\lambda_k(\boldsymbol{h}) = \frac{f_t^*(z_k(\boldsymbol{h});\nu_k)}{1 - F_t^*(z_k(\boldsymbol{h});\nu_k)} \frac{\nu_k - 2 + (z_k(\boldsymbol{h}))^2}{\nu_k - 1} \tag{27}$$

for the implementation of VaR_α and CVaR_α and the corresponding FOCs. This model can be calibrated with equal degrees of freedom parameters for all components or with individual ν_k, $k = 1, \ldots, K$.

Although basic regime-switching models, as defined above, can already capture persistence in (all) conditional moments of $(\boldsymbol{R}_t)_{t\in\mathbb{N}}$, in particular autocorrelation of the returns and volatility clustering, the temporal dependence introduced by the Markov chain is often augmented with traditional time series filters (Alizadeh et al., 2008). As our focus is on the distributional and tail characteristics of the return model, we will not consider such extensions. We, however, note that Theorem 1 also applies to these extensions by replacing $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ with the conditional moments predicted by the time series filters for state k. Furthermore, Theorem 1 applies to conditional mixture distributions derived from RS models with time-varying transition probabilities, which have for example been used in Perez-Quiros and Timmermann (2000). Finally, there are also a number of finance applications that work within the simpler setting of mixture distributions, in which $(S_t)_{t\in\mathbb{N}}$ is an i.i.d. sequence (Buckley et al., 2008).

3.3 A Stylized Example

Concluding the theoretical section, we illustrate differences between (C)VaR$_\alpha$-minimal hedging strategies and minimum-variance hedging in the case of nonelliptical distributions. In particular, we compare the optimal strategies, the remaining tail risk and the moments of the hedged portfolio return for a stylized model with a crash regime.

We use a simple two-state mixture of normals for the joint distribution of spot and futures returns.[18] The first state is assumed to be a low probability crash scenario with negative means, high standard deviations and a high correlation. In particular, we assume $\pi_1 = 10\%$ for the state probability and

$$\mu_{S,1} = \mu_{F,1} = -5\%, \qquad \sigma_{S,1} = \sigma_{F,1} = 10\%, \qquad \rho_{SF,1} = 95\%, \tag{28}$$

where $\sigma_{S,k}$, $\sigma_{F,k}$ and $\rho_{SF,k}$ denote the state specific standard deviations and the correlation of spot and futures returns in state k. The second state describes a normal market environment with positive means, lower standard deviations and a somewhat lower correlation.[19] Specifically, we set $\pi_2 = 90\%$ for its probability and assume

$$\mu_{S,2} = \mu_{F,2} = 0.5\%, \qquad \sigma_{S,2} = 3\%, \qquad \sigma_{F,2} = 6\% \qquad \text{and} \qquad \rho_{SF,2} = 75\%. \tag{29}$$

[18] As discussed before, this can be the unconditional or the relevant conditional distribution of an RS model.

[19] This setup can be motivated by the importance of basis risk found under the assumption of elliptical distributions. More generally, the focus on cross-hedging can be motivated by recent empirical evidence showing that also variance-based strategies only have relevant advantages over the naive benchmark if there is a non-negligible amount of basis risk (Alexander and Barbosa, 2007).

Table 1: Hedging Strategies, Risk Measures and Moments – Mixture Example

strategy	uh	naive	var	$\text{VaR}_{1\%}$	$\text{CVaR}_{1\%}$
h	0.00	100.00	51.07	67.51	70.79
$\text{VaR}_{1\%}(h)$	15.25	9.70	7.98	6.74	6.78
$\text{CVaR}_{1\%}(h)$	19.04	11.16	10.02	7.95	7.87
mean R_H	0.40	0.00	0.20	0.13	0.12
std R_H	4.21	4.10	2.60	2.81	2.90
skew R_H	-1.40	0.00	-0.73	-0.12	-0.06
kurt R_H	8.48	3.13	5.35	3.18	3.06

Optimal hedging strategies and characteristics of the hedged portfolio return. Optimal hedging weights, strategies, moments and risk measures are calculated analytically under the assumption of the two-state mixture presented in this section. The columns describe different strategies: uh is the unhedged spot portfolio, var is the minimum-variance hedge, $\text{VaR}_{1\%}$ and $\text{CVaR}_{1\%}$ are tail-risk-minimal strategies with the corresponding objective functions. h is the hedging weight of each strategy. $\text{VaR}_{1\%}(h)$ and $\text{CVaR}_{1\%}(h)$ are the resulting tail risk measures for the strategies. mean, std, skew and kurt correspond to the expected value, the standard deviation, the skewness and the kurtosis of the hedged return. h, mean, std, $\text{VaR}_{1\%}$ and $\text{CVaR}_{1\%}$ in percent.

Due to the differences in standard deviations and correlations, larger hedge ratios are more effective in the crash state than in the normal state. This intuitively explains why the presence of such a crash regime drives a wedge between minimum-variance and tail-risk-optimal hedging.

We document the characteristics of these strategies in Table 1. In particular, we compare the naive hedge with a minimum-variance strategy and VaR_α- as well as CVaR_α-minimal policies for $\alpha = 1\%$. The hedging weight h^*_{var} that minimizes the variance is obtained from (12) and the tail-risk-optimal policies h^*_{VaR} and h^*_{CVaR} are calculated by applying Theorem 1. First, we observe that the hedging weight is increased by switching from a minimum-variance to a tail-risk-optimal strategy. Second, this increase creates sizable additional tail risk reductions measured by VaR_α or CVaR_α, whose magnitudes are comparable to the reduction obtained by adopting the minimum-variance strategy instead of the naive hedge. Third, the example quantifies the negative effect of moving away from the minimum-variance strategy on the standard deviation and it shows a negative effect on the expected portfolio return under the chosen parameterization. However, the higher moments improve when switching to the tail-risk-optimal policies, i. e., the amount of negative skewness and the level of kurtosis are reduced. This makes the left tail of the return distribution less dangerous and explains the additional tail risk reduction mentioned before.

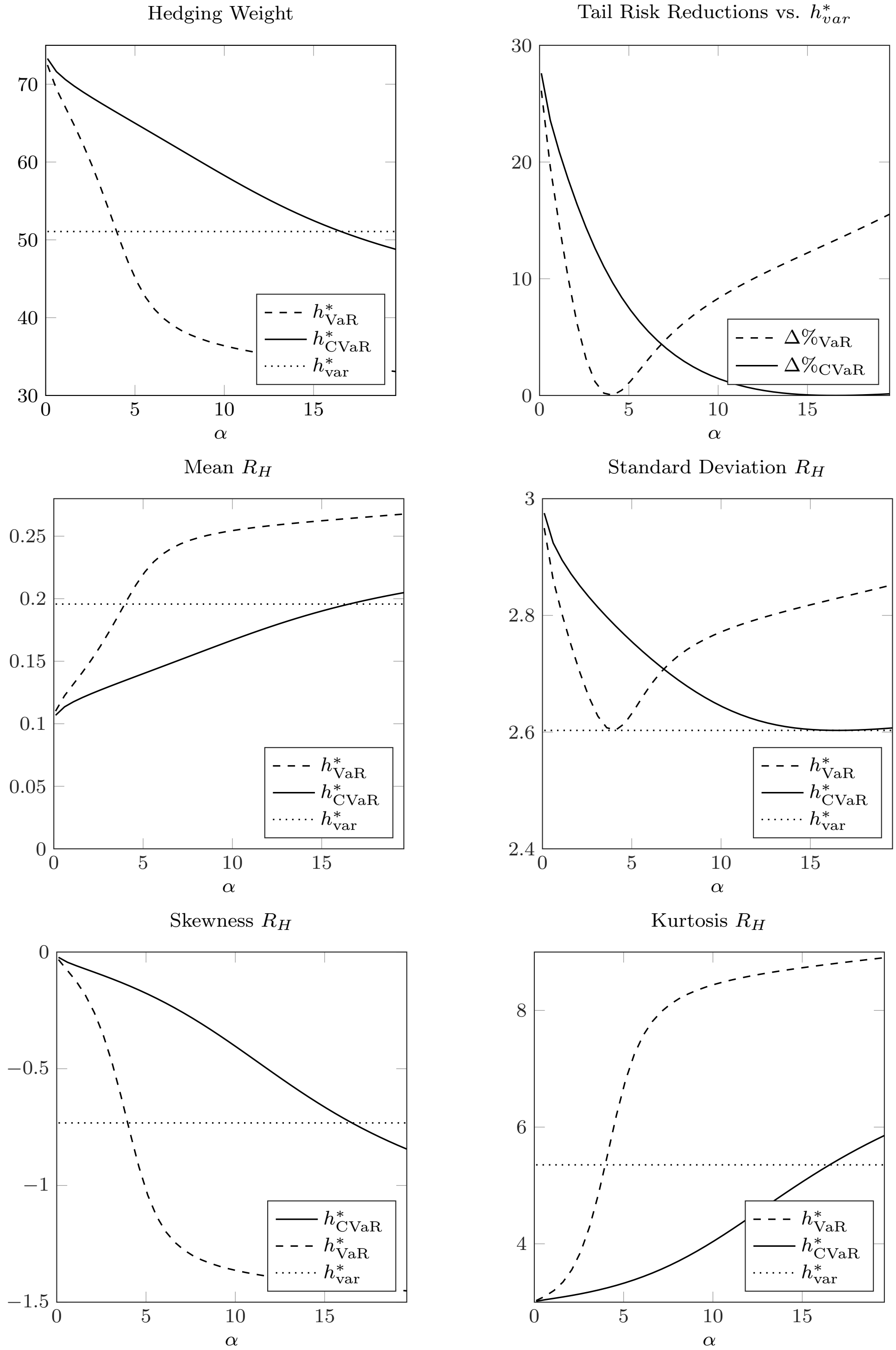

Figure 1: Hedging Strategies – Mixture Example. The two graphs in the first row show the optimal hedging weights under different objective functions and the attainable tail risk reductions by applying (C)VaR$_\alpha$-minimal policies. These reductions are calculated according to (30). The remaining four graphs illustrate the moments of the hedged portfolio return R_H for different strategies. All results are derived from the multivariate mixture model presented in this section. Hedging weights, risk reductions, means, standard deviations and α in percent.

In Figure 1, we illustrate how the hedging weights and the characteristics of the hedged return distribution change with the (C)VaR$_\alpha$-parameter α. First, we observe that for typical parameter values ($\alpha \leq 5\%$) the hedging weights increase with the confidence level $1 - \alpha$. Furthermore, we plot relative tail risk reductions over the minimum-variance strategy, i. e.,

$$\Delta\%_{\mathrm{VaR}} = 1 - \frac{\mathrm{VaR}_\alpha(\boldsymbol{h}^*_{\mathrm{VaR}})}{\mathrm{VaR}_\alpha(\boldsymbol{h}^*_{\mathrm{var}})} \qquad \text{and} \qquad \Delta\%_{\mathrm{CVaR}} = 1 - \frac{\mathrm{CVaR}_\alpha(\boldsymbol{h}^*_{\mathrm{CVaR}})}{\mathrm{CVaR}_\alpha(\boldsymbol{h}^*_{\mathrm{var}})}. \tag{30}$$

These reductions are also increasing in the confidence level, which means that the difference between minimum-variance and tail-risk-optimal hedging becomes more important further out in the tails. This is an interesting finding as compared to the elliptical setup studied before, in which VaR$_\alpha$- and CVaR$_\alpha$-minimal strategies converge to minimum-variance rules for $\alpha \to 0$. Looking at the curves of return skewness and kurtosis, we can see that these differences are related to changes in higher moments.

4 Empirical Examples

In this section, we analyze differences between tail risk hedging derived from RS models and minimum-variance hedging in applications with standard financial time series. In addition, we show that the benefits of the proposed method can be verified by nonparametric estimators both in- and out-of-sample. As laid out before, our focus is again on cross-hedging examples. We compare futures-based hedging strategies that are used to temporarily minimize the tail risk of diversified investment portfolios on an asset allocation level. Such hedging problems may be caused by risk limits, capital requirements or tactical considerations. In line with an investment perspective, we use a monthly hedging horizon, which, in addition, allows us to keep the time series structure of the models relatively simple.

4.1 Data

In our baseline analysis, we consider two portfolios representing the risky part of a broad asset allocation. The first example portfolio (P1) is invested in stocks and bonds represented by the MSCI World Index and the Bank of America Merrill Lynch U.S. High Yield 100 Index. For the second example portfolio (P2), we add the FTSE/NAREIT U.S. All REITs Index. In

Table 2: Descriptive Statistics

	Spot indexes			Futures	Portfolios	
	MSCI	HY	REITs	S&P	(P1)	(P2)
mean [%]	0.83	0.75	0.77	0.50	0.80	0.81
	(0.23)	(0.11)	(0.26)	(0.23)	(0.15)	(0.17)
median [%]	1.33	0.96	1.14	0.88	1.17	1.06
std [%]	4.44	2.23	4.95	4.42	2.99	3.26
min [%]	-20.99	-15.42	-35.99	-22.83	-18.17	-23.76
max [%]	11.13	7.15	24.67	12.41	8.43	14.14
skewness	-0.91	-1.41	-1.71	-0.97	-1.35	-1.72
kurtosis	5.49	11.71	15.00	6.01	8.38	13.31
JB	148.76	1310.77	2433.56	200.15	565.53	1848.11
p_{JB} [%]	0.10	0.10	0.10	0.10	0.10	0.10
corr$(\cdot, F)$	0.88	0.58	0.57	1.00	0.87	0.81
ex-corr$(\cdot, F; q_{0.2})$	0.86	0.57	0.62	1.00	0.83	0.77
ex-corr$(\cdot, F; q_{0.8})$	0.70	0.16	0.26	1.00	0.50	0.47

Descriptive statistics of spot and futures time series. Monthly log-returns from April 1983 to June 2014, $T = 375$ return observations. MSCI: MSCI World Total Return Index, HY: BofA Merrill Lynch US High Yield 100 Total Return Index, REIT: FTSE/NAREIT All REITs Total Return Index, S&P: Chicago Mercantile Exchange S&P 500 Index futures. Equally weighted multi-asset spot portfolios: (P1): MSCI/HY, (P2): MSCI/HY/REITs. JB refers to the Jarque-Bera test statistic for normality and p_{JB} denotes the corresponding p-value. ex-corr$(\cdot, F; q_\alpha)$ measures the correlation of spot and S&P futures returns, given that both returns fall below $(\alpha = 0.2)$ or exceed $(\alpha = 0.8)$ their α-quantile.

both portfolios, the investments are equally weighted. As a liquid hedging instrument, we use S&P 500 Index futures traded on the Chicago Mercantile Exchange. This choice is motivated by the importance of the S&P 500 as an indicator for the overall U.S. market and a relatively high correlation with the spot indexes used.[20] Total return indexes for the spot investments and a perpetual price index[21] for the futures contract were obtained from Datastream.

Our sample spans from March 1983 to June 2014, which corresponds to 376 monthly price observations. Following common practice in the literature on RS models, we use continuously compounded returns.[22] Descriptive statistics of the return series are presented in Table 2.

The returns on all individual assets as well as on our portfolios exhibit pronounced skewness and excess kurtosis so that the normality assumption is formally rejected by Jarque-Bera tests for all series. Comparing the spot portfolios, the returns of (P2) exhibit stronger asymmetries and fatter tails than those of (P1). The kurtosis of the former is twice as high as that of

[20] We also considered using U.S. T-Bond futures to improve the hedging quality for the bond component, but we found that these have a very low or even negative correlation with our high yield bond index.

[21] This index is computed from returns of the nearest futures with switch over following the last trading day. For days when contracts are rolled forward, calculating spurious returns with prices of different futures is avoided by considering the prices of two successive securities.

[22] The usage of log-returns is a standard approximation for the exact approach based on discrete returns discussed in Section 2. In Section 4.4, we present an example for hedging with discrete returns, obtaining very similar results.

the futures returns. According to nonparametric estimates of exceedance correlations, we find evidence for dependence asymmetries in the bivariate distributions of spot and futures returns.

4.2 Parameter Estimates and Model Fit

To hedge long positions in (P1) and (P2), we fit RS models with two and three[23] normal components to the bivariate distributions of portfolio and futures returns.[24] The parameters that attained the highest likelihood in repeated maximum-likelihood estimations from randomly chosen initial values are displayed in Table 3.[25]

To ensure the irreducibility and aperiodicity of the state process, we restrict the elements of the transition matrix to be positive. Label switching is applied to obtain a state ordering according to $q_{11} < q_{22} < q_{33}$. The structure of the two-state models is very similar: There is a joint bearish state with a low probability of occurrence, negative means, high standard deviations and high correlations. Allowing for a third component, the first state becomes a severe crash scenario in both cases.

In Panel A of Table 4, we provide some evidence on the fit of these models and simple elliptical distributions for the bivariate return samples.[26] According to AIC and BIC, at least one of the RS models is favored over nonswitching specifications. While AIC prefers three-state models, BIC favors two-state models. We also perform Jarque-Bera tests on the distribution fit after transforming the sample data to normality with the Berkowitz (2001) approach. In contrast to the more restrictive specifications, the fit of the predictive distributions derived from the three-state RS models cannot be rejected at conventional significance levels.

---

[23] The restriction on two or three regimes will be shown to be adequate given the results of risk measurement backtests in this section. Furthermore, it enables us to run out-of-sample analyses with relatively small estimation samples.

[24] Although our approach allows for a full asset-level description of the joint distribution of spot and futures returns, we prefer aggregating the spot returns into portfolio returns first to keep the dimension of the model as low as possible.

[25] As described in Section 3.2, we assume that the state process starts from its stationary distribution, which excludes the use of the standard analytic EM algorithm (Hamilton, 1990). Results obtained with this algorithm are, however, similar, as shown in the online appendix to this paper. The online appendix also contains further omitted estimation results.

[26] The degrees of freedom parameters estimated for the multivariate distributions correspond to 4.5 and 4.1 for (P1) and (P2), respectively. The other model parameters can be found in the online appendix.

Table 3: In-Sample Parameter Estimates

	(P1)				(P2)			
	RS2		RS3		RS2		RS3	
	par	s.e.	par	s.e.	par	s.e.	par	s.e.
State 1								
$\mu_{S,1}$	-0.57	(1.01)	-3.79	(4.20)	-2.61	(1.58)	-4.00	(2.33)
$\mu_{F,1}$	-1.61	(1.40)	-6.33	(6.72)	-4.19	(1.76)	-6.13	(2.51)
$\sigma_{S,1}$	4.90	(0.59)	4.85	(1.00)	6.61	(1.63)	7.86	(1.86)
$\sigma_{F,1}$	6.96	(0.68)	6.24	(2.77)	7.44	(1.19)	8.12	(1.49)
$\rho_{SF,1}$	87.31	(2.64)	82.45	(16.02)	83.13	(4.25)	82.27	(5.74)
State 2								
$\mu_{S,2}$	1.18	(0.14)	1.46	(0.32)	1.24	(0.17)	1.05	(0.38)
$\mu_{F,2}$	1.08	(0.20)	1.37	(0.44)	1.10	(0.26)	0.69	(0.46)
$\sigma_{S,2}$	2.03	(0.12)	2.49	(0.35)	2.18	(0.16)	3.13	(0.30)
$\sigma_{F,2}$	3.17	(0.17)	4.02	(0.39)	3.43	(0.25)	3.57	(0.37)
$\rho_{SF,2}$	83.50	(2.18)	81.04	(3.27)	76.33	(2.60)	92.11	(2.09)
State 3								
$\mu_{S,3}$			0.97	(0.19)			1.14	(0.16)
$\mu_{F,3}$			0.92	(0.29)			1.01	(0.29)
$\sigma_{S,3}$			1.80	(0.15)			2.08	(0.14)
$\sigma_{F,3}$			2.54	(0.26)			3.73	(0.28)
$\rho_{SF,3}$			88.36	(2.24)			73.99	(3.45)
Transition matrix								
q_{11}	83.0	(7.7)	61.2	(30.0)	63.0	(11.7)	61.1	(14.2)
q_{12}			38.7	(46.6)			10.2	(9.5)
q_{21}	4.5	(1.7)	4.6	(3.8)	4.7	(2.4)	1.5	(1.3)
q_{22}			92.8	(3.0)			97.2	(2.1)
q_{31}			2.8	(2.8)			2.8	(1.6)
q_{32}			0.8	(1.2)			0.1	(0.1)
Stationary distribution								
π_1	21.1		9.1		11.2		5.9	
π_2	78.9		53.2		88.8		22.8	
π_3			37.7				71.3	

Parameter estimates for bivariate two-state and three-state RS models with normal components. The models describe the bivariate distributions of the portfolio returns and the returns of the S&P futures. (P1) and (P2) are defined in Table 2. The parameters are obtained by MLE using the Hamilton filter, assuming that the state process started from its stationary distribution. For each model the estimation was repeated several times from randomly chosen initial values to avoid local maxima. We report robust standard errors derived from the Hessian of the log-likelihood and the outer product of the scores. For (P2) and $K = 3$, a boundary solution was found due to the low value of q_{32}. All parameter values in percent.

Before analyzing the hedging performance of the RS models, we assess their risk *measurement* quality and compare it to multivariate elliptical distributions. We focus on the 99% confidence level, which will also be considered in the hedging analysis. In particular, we analyze risk forecasts for an unhedged long position in both portfolios and a short position in the S&P futures derived from each of the bivariate return models. For the RS models, we distinguish between unconditional forecasts $\widehat{\text{VaR}}_\alpha^{RS,u}$ and $\widehat{\text{CVaR}}_\alpha^{RS,u}$ derived from the stationary

Table 4: Model Fit and Risk Backtesting

	Panel A				*Panel B*					
	Statistical fit				Risk spot long			Risk futures short		
	LL	AIC	BIC	p_{berk}	p_{uc}	p_{cc}	p_{CVaR}	p_{uc}	p_{cc}	p_{CVaR}
(P1)										
np	-	-	-	-	69.0	90.2	33.1	69.0	90.2	22.6
evt	-	-	-	-	89.4	8.6	53.3	89.4	94.9	49.0
mv-n	1681.9	-3353.7	-3334.1	0.1	0.2	0.0	1.1	9.1	23.8	8.2
mv-t	1726.1	-3440.2	-3416.7	5.2	5.5	0.0	28.9	-	-	-
RS2 stat	1749.2	-3474.4	-3427.3	9.7	89.4	8.6	16.1	9.1	23.8	91.3
RS2 pred	-	-	-	-	28.0	11.5	43.5	69.0	90.2	99.1
RS3 stat	1769.4	-3496.8	-3414.3	50.0	69.0	90.2	27.2	69.0	90.2	66.3
RS3 pred	-	-	-	-	69.0	90.2	35.1	32.1	60.5	65.7
(P2)										
np	-	-	-	-	69.0	4.1	27.9	69.0	90.2	22.7
evt	-	-	-	-	53.3	11.6	77.9	89.4	94.9	48.7
mv-n	1593.5	-3177.0	-3157.4	0.1	0.7	0.2	1.1	9.1	23.8	8.1
mv-t	1660.3	-3308.6	-3285.1	3.6	0.7	0.2	31.3	-	-	-
RS2 stat	1679.5	-3335.0	-3287.9	50.0	69.0	4.1	24.0	69.0	90.2	65.9
RS2 pred	-	-	-	-	69.0	90.2	70.0	69.0	90.2	53.1
RS3 stat	1701.3	-3360.5	-3278.1	50.0	69.0	4.1	37.8	69.0	90.2	54.8
RS3 pred	-	-	-	-	89.4	94.9	81.9	69.0	90.2	51.6

Panel A refers to the statistical fit of the models for the bivariate processes of portfolio and futures returns. LL is the log-likelihood of the models. AIC and BIC refer to the Akaike information criterion and the Bayesian information criterion. p_{berk} is the p-value of a Jarque-Bera test applied to the sample data transformed with its predictive cdf and the inverse cdf of the normal distribution. (P1) and (P2) are defined in Table 2. The tests in Panel B are applied to model-based risk estimates for a long position in the spot portfolio and a short position in the S&P futures. p_{uc} and p_{cc} are p-values of Christoffersen (1998) tests on correct unconditional and conditional coverage. p_{CVaR} refers to p-values of one-sided CVaR_α tests according to McNeil et al. (2005, p. 163). np and evt are nonparametric and EVT-based risk estimates for the corresponding loss series. mv-n and mv-t refer to multivariate normal and multivariate standardized t-distributions. RS2/3 denote RS models with $K = 2$ and $K = 3$ normal components. stat refers to backtesting results for the unconditional risk estimates, and pred contains the corresponding results for conditional risk forecasts.

distribution and series of conditional forecasts $\widehat{\text{VaR}}_\alpha^{RS,c}$ and $\widehat{\text{CVaR}}_\alpha^{RS,c}$ based on the predictive distribution. Both are calculated using (17) and (18) with (25).

Furthermore, we also consider nonparametric risk estimators derived from the empirical distribution and semiparametric estimators relying on extreme value theory (EVT). These are calculated from the relevant univariate loss sample $(l_t)_{t=1,\dots,T}$. As nonparametric estimators we apply the lower quantile of the empirical distribution, i.e., $\widehat{\text{VaR}}_\alpha^{np} = l_{(\lceil T(1-\alpha)\rceil)}$, and the CVaR_α estimator from Rockafellar and Uryasev (2002, Proposition 8), which can be written as

$$\widehat{\text{CVaR}}_\alpha^{np} = \frac{1}{\alpha}\left[\frac{1}{T}\sum_{i=\lceil(1-\alpha)T\rceil+1}^{T} l_{(i)} + \left(\frac{\lceil T(1-\alpha)\rceil}{T} - (1-\alpha)\right) l_{(\lceil T(1-\alpha)\rceil)}\right], \tag{31}$$

where $l_{(i)}$ is the ith rank statistic of the loss sample. The EVT-based estimates are calculated from the subsample of losses exceeding a threshold u.[27] We fit a generalized Pareto distribution (GPD) to the corresponding loss exceedances $l_t - u$. From the estimated shape and scale parameters $\hat{\xi}$ and $\hat{\beta}$ and the number of exceedances n_u, the risk estimators (McNeil and Frey, 2000)

$$\widehat{\text{VaR}}_\alpha^{evt} = u + \frac{\hat{\beta}}{\hat{\xi}}\left[\left(\alpha\frac{T}{n_u}\right)^{-\hat{\xi}} - 1\right], \tag{32}$$

$$\widehat{\text{CVaR}}_\alpha^{evt} = \widehat{\text{VaR}}_\alpha^{evt} + \frac{\hat{\beta} + \hat{\xi}(\widehat{\text{VaR}}_\alpha^{evt} - u)}{1 - \hat{\xi}} \tag{33}$$

are determined.

We apply the conditional and unconditional coverage tests proposed by Christoffersen (1998) and the CVaR_α test introduced in McNeil et al. (2005, p. 163) for the formal evaluation of our tail risk estimators. Corresponding test results can be found in Panel B of Table 4. The VaR_α estimates derived from the predictive distributions of the RS models, the nonparametric and the EVT-based techniques are never rejected according to unconditional coverage tests at standard significance levels, whereas the risk forecasts derived from the elliptical specifications can be rejected at the 10% level. Furthermore, the predictive VaR_α-series of the RS models pass all conditional coverage tests in contrast to the unconditional specifications considered here. However, at the 1% significance level, there is no rejection of the correct conditional coverage hypothesis for $\widehat{\text{VaR}}_\alpha^{RS,u}$. Hence, the evidence in favor of dynamic risk forecasting is not very strong for our monthly data. The CVaR_α tests do not seem to have much discriminatory power between the models. They only reject the multivariate normal specifications in our examples.

4.3 Tail Risk Hedging Results

Turning to the core of our empirical analysis, we investigate tail-risk-minimal hedging strategies derived from RS models using Theorem 1 and compare their performance to minimum-variance hedges according to (12). We provide in- and out-of-sample results. The

[27] We set the threshold to the 0.9-quantile of the empirical distribution.

former are based on the models presented above and for the latter we use a growing estimation window. In particular, we reserve 175 observations for the first estimation and re-estimate the parameters monthly.[28]

We analyze VaR_α- and CVaR_α-minimal strategies for $\alpha = 1\%$. Motivated by the recent proposals of the Basel Committee[29], we also consider the CVaR_α with $\alpha = 2.5\%$, which, in addition, allows for a more accurate evaluation of the hedging performance. This performance is measured by the standard nonparametric risk estimators $\widehat{\text{VaR}}_\alpha^{np}$ and $\widehat{\text{CVaR}}_\alpha^{np}$ applied to the hedged returns of the strategies under consideration. Just as in (30), we calculate realized relative tail risk reductions $\Delta\%_{\text{VaR}}$ and $\Delta\%_{\text{CVaR}}$ over the minimum-variance strategy based on $\widehat{\text{VaR}}_\alpha^{np}$ and $\widehat{\text{CVaR}}_\alpha^{np}$. The results for *conditional* hedging strategies based on the predictive distributions of three-state RS models are presented in Table 5.[30]

First, note that the average hedging weights of tail-risk-minimal strategies are higher than the corresponding values of the minimum-variance approach. The strategies minimizing the CVaR_α with $\alpha = 1\%$ have the highest average hedging weights. The hedging weights of the strategies minimizing the VaR_α with $\alpha = 1\%$ and the CVaR_α with $\alpha = 2.5\%$ are between the minimum-variance hedge and the CVaR_α hedge with $\alpha = 1\%$. The former two strategies are almost identical for (P1) but show some differences for (P2), where the CVaR_α hedge with $\alpha = 2.5\%$ is closer to the CVaR_α-based strategy with $\alpha = 1\%$.

Second, we compare the risk of the hedged portfolios. We find that minimum-variance cross-hedges already successfully remove a large fraction of tail risk. However, the increase in the hedging amount implied by the tail-risk-optimal strategies leads to additional tail risk reductions in- and out-of-sample in all but one of the cases[31]. These additional reductions range between 2% and 18% for our data sets. Similar to the theoretical example discussed earlier, the differences are higher for the CVaR_α than for the VaR_α when looking at $\alpha = 1\%$ and they are higher for $\alpha = 1\%$ than for $\alpha = 2.5\%$ when comparing the CVaR_α objectives.

---

[28] To get stable results in the first estimations with a limited number of observations, we include a parameter constraint on the transition matrix ensuring that the stationary regime probabilities are at least equal to five percent. This avoids problems with regimes describing a very small number of observations.

[29] For normally distributed risks, the CVaR_α with $\alpha = 2.5\%$ and the VaR_α with $\alpha = 1\%$ are almost equal.

[30] A selection of results for two-state models is provided in the context of the model comparison in Table 7.

[31] In-sample, there is a slightly negative effect for (P2) when using the CVaR_α as objective function with $\alpha = 2.5\%$.

Table 5: Conditional Tail Risk Hedging

	(P1) MSCI+HY					(P2) MSCI+HY+REIT				
strategy	uh	var	VaR_α	CVaR_α	CVaR_α	uh	var	VaR_α	CVaR_α	CVaR_α
α			1%	2.5%	1%			1%	2.5%	1%
Panel A: In-Sample Results										
mean h	0.00	58.17	63.30	64.91	68.76	0.00	58.78	63.88	68.32	75.84
std h	0.00	3.36	5.64	5.19	4.56	0.00	12.02	12.29	10.10	7.14
$\widehat{\text{VaR}}^{np}_{1\%}$	9.44	3.62	3.39			10.98	5.20	4.69		
$\Delta\%$			*6.47*					*9.84*		
$\widehat{\text{CVaR}}^{np}_{2.5\%}$	9.39	3.60		3.37		10.90	4.56		4.63	
$\Delta\%$				*6.40*					*-1.53*	
$\widehat{\text{CVaR}}^{np}_{1\%}$	12.38	4.37			3.99	15.13	6.60			5.64
$\Delta\%$					*8.80*					*14.58*
mean R_H	0.80	0.52	0.50	0.49	0.47	0.81	0.53	0.51	0.48	0.44
std R_H	2.99	1.49	1.54	1.54	1.58	3.26	1.77	1.83	1.87	1.99
skew R_H	-1.35	-0.06	0.45	0.46	0.48	-1.72	-0.16	0.23	0.23	0.21
kurt R_H	8.38	5.34	6.46	6.35	6.15	13.31	8.03	7.75	7.28	6.04
Panel B: Out-of-Sample										
mean h	0.00	54.31	56.60	57.07	58.67	0.00	56.49	60.57	61.18	62.54
std h	0.00	5.31	8.80	9.11	9.92	0.00	12.88	13.21	14.26	16.12
$\widehat{\text{VaR}}^{np}_{1\%}$	9.44	4.28	3.80			10.98	5.39	5.27		
$\Delta\%$			*11.20*					*2.21*		
$\widehat{\text{CVaR}}^{np}_{2.5\%}$	10.66	4.29		3.99		13.03	6.73		6.06	
$\Delta\%$				*7.00*					*9.95*	
$\widehat{\text{CVaR}}^{np}_{1\%}$	14.18	6.03			5.01	18.03	10.53			8.59
$\Delta\%$					*17.00*					*18.40*
mean R_H	0.54	0.38	0.38	0.37	0.36	0.62	0.41	0.42	0.41	0.38
std R_H	3.31	1.41	1.41	1.40	1.40	3.84	2.12	1.99	1.99	1.98
skew R_H	-1.42	-1.02	-0.34	-0.33	-0.28	-1.61	-1.34	-0.95	-1.01	-1.12
kurt R_H	8.08	8.68	6.68	6.68	7.00	11.57	14.67	10.35	10.49	10.74

Hedging weights, tail risk measures and moments of selected conditional hedging strategies for (P1) and (P2) as defined in Table 2. The strategies in this table are derived from the predictive distribution of three-state RS models. The columns describe different strategies: uh is the unhedged spot portfolio, var is the minimum-variance hedge, $\text{VaR}_{1\%}$, $\text{CVaR}_{2.5\%}$ and $\text{CVaR}_{1\%}$ are tail-risk-minimal strategies with the corresponding objective functions. h is the hedging weight of the S&P futures. $\widehat{\text{VaR}}^{np}_\alpha$ and $\widehat{\text{CVaR}}^{np}_\alpha$ refer to nonparametric tail risk estimates. The emphasized values denoted by $\Delta\%$ are the relative tail risk reductions of the optimal strategy over the minimum-variance strategy defined in (30). mean, std, skewness and kurtosis are the moments of the hedged portfolio return R_H. Panel A contains in-sample results derived from the full sample and the three-state RS models presented in the Table 3. Panel B contains out-of-sample results based on a growing estimation window, using 175 observations for the first estimation and updating the strategies monthly. Hedge ratios, risk estimates, means and standard deviations in percent.

Third, we analyze the moments of the return distributions of the net positions to gain insight into the sources of the risk reduction. In line with our theoretical example, we find that the returns of strategies with tail risk objectives typically have a less negative skewness and a lower kurtosis. An exception is the in-sample hedging of (P1), for which only the skewness is increased but the kurtosis of the tail-risk-based strategies is somewhat higher than for the minimum-variance hedge.

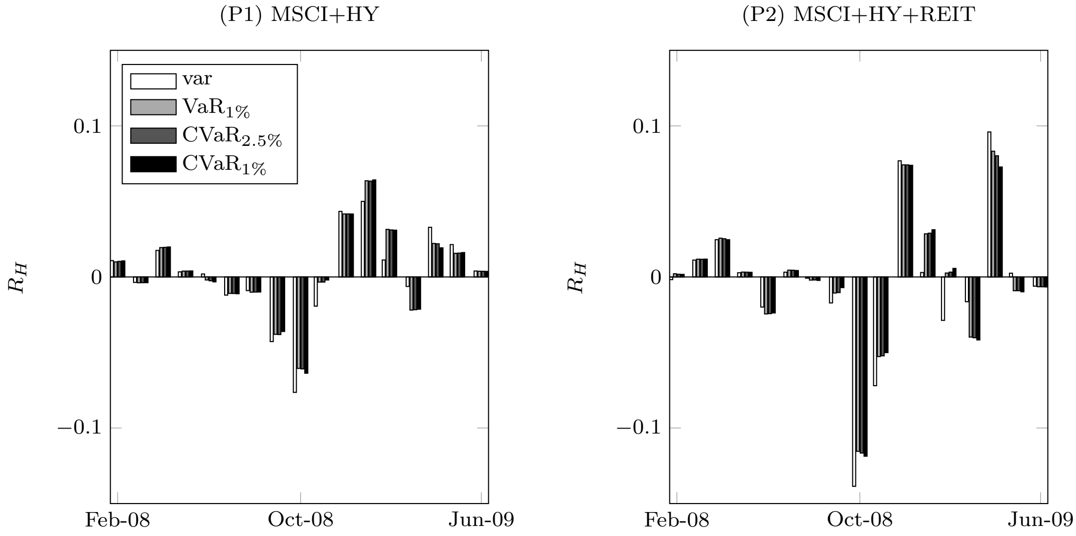

Figure 2: Hedged Portfolio Returns During the Financial Crisis: These graphs show the hedged returns of the two example portfolios under different hedging strategies during the financial crises in 2008/2009. var denotes minimum-variance strategies. $\mathrm{VaR}_{1\%}$, $\mathrm{CVaR}_{2.5\%}$ and $\mathrm{CVaR}_{1\%}$ are tail-risk-based strategies which minimize the corresponding objective function. All strategies are derived from the predictive distributions of three-state RS models which were estimated out-of-sample. See Table 2 for the definitions of (P1) and (P2) and see Table 5 for further details on the strategies and their estimation.

In addition to the formal VaR_α and CVaR_α analyses, Figure 2 shows the returns of the proposed strategies during the subprime crisis. These returns are obtained with out-of-sample hedging weights. It can be seen that the tail-risk-based hedges limit the largest losses more effectively than the minimum-variance strategy. This comes at the cost of slightly increased losses under less extreme circumstances and sometimes reduced gains. For (P2) the largest losses are more extreme and the differences between the strategies are more pronounced.

We next investigate simpler *unconditional* hedging strategies derived from the stationary distribution of the RS models. For this case, an unconditional minimum-variance hedge is used as the reference strategy.[32] We focus on CVaR_α-optimal strategies with $\alpha = 1\%$ and extend the range of tail risk estimation methods. In particular, we include the EVT-based risk estimator $\widehat{\mathrm{CVaR}}_\alpha^{evt}$ and model-based risk estimates[33] to evaluate the performance of the strategies. The corresponding results are presented in Table 6.

[32] We implement this strategy with a sample estimator of the covariance matrix. Minimum-variance strategies derived from the stationary distribution of the RS models are very similar.

[33] This is possible for in-sample analyses based on the stationary distribution, for which a single model and a single set of state probabilities are used.

Table 6: Unconditional Tail Risk Hedging

	(P1) MSCI+HY			(P2) MSCI+HY+REIT		
strategy	uh	var	$\text{CVaR}_{1\%}$	uh	var	$\text{CVaR}_{1\%}$
Panel A: In-Sample Results						
h	0.00	58.43	70.20	0.00	60.05	80.46
$\widehat{\text{CVaR}}_{1\%}^{RS,u}$	12.08	4.88	4.48	15.74	7.41	6.31
$\Delta\%$			*8.07*			*14.81*
$\widehat{\text{CVaR}}_{1\%}^{np}$	12.38	4.59	4.16	15.13	7.55	6.39
$\Delta\%$			*9.48*			*15.37*
$\widehat{\text{CVaR}}_{1\%}^{evt}$	12.58	4.84	4.17	16.02	7.28	6.41
$\Delta\%$			*13.92*			*11.98*
mean R_H	0.80	0.51	0.45	0.81	0.51	0.40
std R_H	2.99	1.50	1.59	3.26	1.89	2.10
skew R_H	-1.35	-0.23	0.29	-1.72	-0.79	-0.01
kurt R_H	8.38	5.55	5.09	13.31	10.29	5.08
$Q_{0.9}[-R_H]$	2.45	1.37	1.52	2.65	1.78	2.12
ξ	0.05	0.19	-0.06	0.26	0.33	-0.07
β	2.78	0.73	0.90	2.40	0.84	1.48
Panel B: Out-of-Sample Results						
mean h	0.00	53.33	61.63	0.00	50.62	61.79
std h	0.00	3.68	6.97	0.00	6.22	14.10
$\widehat{\text{CVaR}}_{1\%}^{np}$	14.18	6.12	5.17	18.03	11.27	9.43
$\Delta\%$			*15.54*			*16.32*
$\widehat{\text{CVaR}}_{1\%}^{evt}$	13.95	6.47	5.26	17.50	10.31	9.45
$\Delta\%$			*18.59*			*8.32*
mean R_H	0.54	0.36	0.35	0.62	0.43	0.36
std R_H	3.31	1.42	1.34	3.84	2.20	1.99
skew R_H	-1.42	-1.13	-0.46	-1.61	-1.60	-1.35
kurt R_H	8.08	8.94	7.30	11.57	14.61	12.23
$Q_{0.9}[-R_H]$	3.19	1.28	1.22	3.25	1.89	1.93
ξ	-0.04	0.34	0.30	0.08	0.30	0.44
β	3.51	0.75	0.65	3.71	1.35	0.85

Hedging weights, tail risk measures and moments of selected unconditional hedging strategies for (P1) and (P2) as defined in Table 2. The strategies are derived from the stationary distribution of three-state RS models. The columns describe different strategies. See Table 5 for definitions of these strategies and for the return characteristics that we provide. In addition, we report analytic risk estimates $\text{CVaR}_\alpha^{RS,u}$ and EVT-based estimates $\widehat{\text{CVaR}}_\alpha^{evt}$. $Q_{0.9}[-R_H]$, ξ and β describe the GPD models fitted to the upper tail of the corresponding loss distributions. Panel A contains in-sample results derived from the full sample and the three-state RS models presented in Table 3. Panel B contains out-of-sample results based on a growing estimation window, using 175 observations for the first estimation and updating the strategies monthly. Hedging weights, risk estimates, risk reductions, means, standard deviations, $Q_{0.9}[-R_H]$ and β in percent.

The constant in-sample hedge ratios are again higher for tail-risk-based strategies than for the minimum-variance approach. This also applies to the out-of-sample analyses, which show some variation due to the re-estimation of the models. Moreover, we again find non-negligible reductions in tail risk as compared to the benchmark strategy. The magnitude of the realized tail risk reductions according to $\widehat{\text{CVaR}}_\alpha^{np}$ is between 9% and 16%. Similar reductions are confirmed by the EVT-estimates and, for the in-sample analysis, also by model-based tail risk figures. Furthermore, we again observe that these reductions are related to changes in the higher moments of the return distributions, which decrease the probability mass in the left tail. This observation is confirmed by the GPD models calibrated to the upper tails of the loss distributions. In particular, these results show that tail risk hedging lowers the shape parameter and thus reduces the heaviness of the relevant tail in three out of four cases. This reduction can come at the cost of increasing the 90%-quantile or the scale parameter of the GPD, but it overcompensates for these effects according to $\widehat{\mathrm{CVaR}}_\alpha^{evt}$ estimates.

Concluding the presentation of our baseline results, we provide a complementary view on the observed differences between hedging strategies based on the unconditional dependence structure of the spot and futures returns. For the returns of (P2) and the S&P futures, we present sample estimates of exceedance correlations and model-implied values in Figure 3. We observe higher correlations in joint crash states than in states with large positive returns in both assets. This explains the reduction in CVaR_α by *increasing* the hedging weight. A three-state RS model can capture this dependence structure closely matching the nonparametric correlation estimates. Similar evidence for an increased dependence between spot and futures returns in bear markets is obtained by comparing the nonparametric estimate of the lower tail dependence function (Garcia and Tsafack, 2011) with the corresponding values implied by a normal distribution, which are also provided in Figure 3. Again, the values of the RS model[34] are close to the nonparametric estimates, which fluctuate quite strongly due to the small sample size.

The results presented so far show that tail risk hedging based on RS models can be beneficial in- and out-of-sample with conditional as well as unconditional hedging strategies. Moreover, we observed that the differences can be attributed to nonelliptical features of the data.

4.4 Model Comparisons

In this section, we compare the out-of-sample performance of the proposed tail-risk-minimal hedging approach based on RS models with alternative implementations. We focus on (P2)

[34] Interestingly, we find that over the plotted range, these values are even higher than those of the t-model (copula).

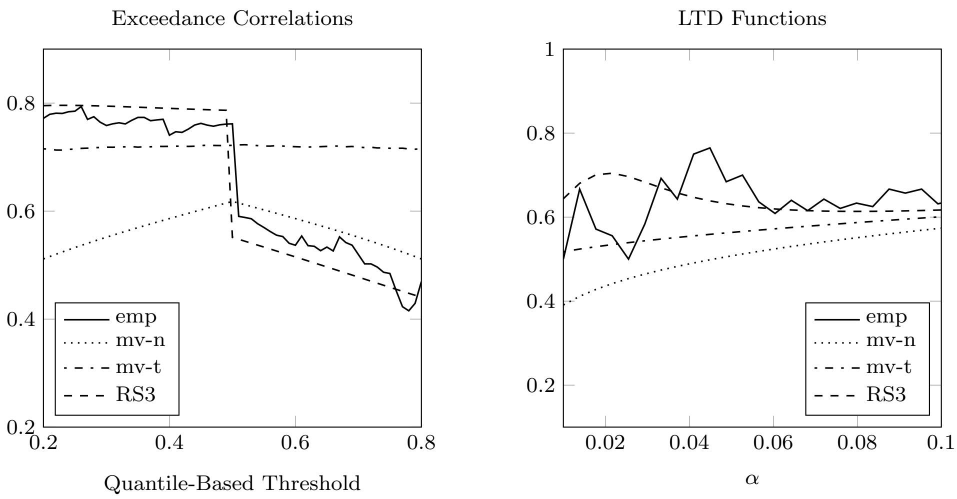

Figure 3: Dependence Structure of (P2) and the S&P Futures. The left graph shows exceedance correlations between the portfolio and futures returns for a quantile-based threshold (Patton, 2004). The right graph depicts lower tail dependence functions (Garcia and Tsafack, 2011) for these returns. emp, mv-n and mv-t refer to values calculated based on the empirical, the multivariate normal and the multivariate t-distribution. RS3 refers to values implied by a three-state RS model with normal components.

and again consider unconditional CVaR_α hedging. This setting allows for a fair comparison with solely unconditional techniques such as strategies derived from nonswitching specifications and nonparametric CVaR_α minimizations (Rockafellar and Uryasev, 2002).[35]

In our first comparison, we consider four parametric specifications: a Gaussian return model, the multivariate t-distribution and RS models with two as well as three normally distributed components.[36] Furthermore, we include the numerical minimization of the nonparametric CVaR_α estimate corresponding to the approach by Rockafellar and Uryasev (2002). Tail risk reductions are measured over a standard minimum-variance hedge. We analyze CVaR_α-minimal strategies with $\alpha = 2.5\%$ as well as $\alpha = 1\%$ and use the same risk measures for performance evaluation. The corresponding results are presented in Table 7.

We find that the CVaR_α strategies based on elliptical models perform worse than the minimum-variance benchmark. This is not surprising given that these models cannot capture relevant features of the data, which was confirmed by the relatively weak risk backtesting results for these models. In contrast, the two-state RS model clearly attains additional tail

35 Note that the linear programming techniques presented in Rockafellar and Uryasev (2002) only apply to CVaR_α in contrast to the methods proposed here.

36 RS models with t-distributed components will be briefly analyzed at the end of this section.

Table 7: Unconditional Tail Risk Hedging – Model Comparison

strategy/model	uh	var	mv-n	mv-t	RS2	RS3	np
Panel A: CVaR$_\alpha$*-hedging with* $\alpha = 2.5\%$							
mean h	0.00	50.62	48.65	43.34	62.46	59.83	57.04
std h	0.00	6.22	6.45	4.84	10.03	9.88	8.24
$\widehat{\text{CVaR}}^{np}_{2.5\%}$	13.03	7.25	7.36	7.91	6.54	6.53	6.69
$\Delta\%$			*-1.46*	*-9.01*	*9.84*	*10.05*	*7.83*
$\widehat{\text{CVaR}}^{evt}_{2.5\%}$	13.20	7.19	7.38	8.08	6.72	6.38	6.54
$\Delta\%$			*-2.75*	*-12.45*	*6.50*	*11.17*	*8.95*
Panel B: CVaR$_\alpha$*-hedging with* $\alpha = 1\%$							
mean h	0.00	50.62	48.89	43.73	66.77	61.79	69.98
std h	0.00	6.22	6.42	4.80	11.47	14.10	15.60
$\widehat{\text{CVaR}}^{np}_{1\%}$	18.03	11.27	11.44	11.98	9.56	9.43	9.78
$\Delta\%$			*-1.56*	*-6.34*	*15.10*	*16.32*	*13.19*
$\widehat{\text{CVaR}}^{evt}_{1\%}$	17.50	10.31	10.53	11.76	9.33	9.45	9.80
$\Delta\%$			*-2.10*	*-14.09*	*9.54*	*8.32*	*4.96*

Out-of-sample hedging weights, tail risk measures and sample moments of CVaR$_\alpha$-minimal strategies for (P2) as defined in Table 2. Panel A contains results for $\alpha = 2.5\%$ and Panel B shows results for $\alpha = 1\%$. The first two columns describe the reference strategies: the unhedged portfolio (uh) and a minimum-variance hedge (var). mv-n and mv-t are CVaR$_\alpha$-minimal hedges derived from a multivariate normal and a multivariate t-model. RS2 and RS3 correspond to hedging strategies based on RS models with two and three states. np corresponds to the nonparametric benchmark technique proposed by Rockafellar and Uryasev (2002). mean h and std h are the average hedging weight and its standard deviation. $\widehat{\text{CVaR}}^{np}_{\alpha}$ and $\widehat{\text{CVaR}}^{evt}_{\alpha}$ are nonparameteric and EVT-based risk estimates. The emphasized values denoted by $\Delta\%$ are the corresponding tail risk reductions over the minimum-variance strategy. All models and strategies are calculated out-of-sample based on a growing estimation window, using 175 observations for the first estimation and updating the strategies monthly. All values in percent.

risk reductions and performs nearly as good as the three-state model discussed in the previous section. Moreover, we find that the RS model with three states outperforms the nonparametric CVaR$_\alpha$-minimal benchmark in this example. This might indicate a lack of robustness of the latter technique. In line with our theoretical example, risk reductions are higher for $\alpha = 1\%$ than for $\alpha = 2.5\%$.

We next provide the results of simulation experiments to confirm the out-of-sample performance of our hedging policies with larger sample sizes. We again focus on (P2) and adopt the unconditional hedging strategy derived from the corresponding three-state RS model. Furthermore, we include the nonparametric CVaR$_\alpha$-minimal strategy to gain a better understanding of the differences between our approach and this benchmark. We consider three different simulations: First, we assume that the fitted RS model is the true data-generating process and simulate random paths starting from its stationary distribution. Second, we sample from the empirical distribution (with replacement). Third, we simulate from a meta model, consisting of a t-copula and skewed-t margins. This choice combines an elliptical dependence structure allowing for (symmetric) tail dependence and nonelliptical marginal distributions. We generate $10,000$ return samples with $T = 1,000$ observations. This sample size allows us to focus on $\alpha = 1\%$. We do not re-estimate the models but apply the hedging weights estimated from the original data for all strategies.

The results of this simulation study are reported in Table 8. Simulating from the estimated model, the average CVaR_α reduction confirms our analytic results from Panel A of Table 6. Looking at the quantiles of the reduction series obtained from the simulations, we find that the tail risk reduction of RS CVaR_α hedging as compared to the minimum-variance strategy is positive in 90% of the cases under sampling from the model and the empirical distribution. This implies a statistical significance of this reduction at the 10% level. The same quantiles are negative for the nonparametric CVaR_α minimization, even under sampling from the empirical distribution, which reveals a strong reliance of this technique on specific realizations in the given sample and shows an interesting advantage of the model-based approach for this example. Remarkably, the parametric RS CVaR_α hedging strategy also attains a reduction in 75% of the samples simulated from the copula model, which indicates a certain robustness against model misspecification. At the same time, we observe that the extent of the reduction decreases, confirming a positive contribution of dependence asymmetries to the reported effects.

4.5 Model Extensions and Robustness Checks

In this section, we focus again on tail risk hedging with RS models and investigate whether the documented benefits of this approach can be confirmed for *composite* hedging with two futures contracts. Furthermore, we analyze the performance of tail risk hedging based on RS models with t-distributed components and provide some robustness checks.

In the remaining examples, we analyze CVaR_α hedging with $\alpha = 1\%$ and provide in-sample results for unconditional strategies, which allows us to calculate analytic risk reductions according to (30) in addition to nonparametric and EVT-based tail risk measurements.

Table 8: Unconditional Tail Risk Hedging – Simulation

	RS3				Bootstrap				t-copula, skewed-t margins			
strategy	uh	var	RS3-CVaR	np-CVaR	uh	var	RS3-CVaR	np-CVaR	uh	var	RS3-CVaR	np-CVaR
h	0.00	60.05	80.46	91.54	0.00	60.05	80.46	91.54	0.00	60.05	80.46	91.54
mean $\widehat{\text{CVaR}}^{np}_{1\%}$	15.38	7.24	6.25	6.53	14.92	7.46	6.30	6.03	14.00	7.99	7.43	7.62
mean $\Delta\%$			12.89	8.37			14.08	16.62			6.32	3.23
$q_{0.5}[\Delta\%]$			13.86	9.92			15.26	18.70			6.67	3.89
$q_{0.25}[\Delta\%]$			8.47	1.77			8.98	7.94			2.02	-3.55
$q_{0.1}[\Delta\%]$			2.53	-7.57			2.78	-3.04			-2.3	-10.61
$q_{0.05}[\Delta\%]$			-1.81	-13.33			-1.50	-10.47			-5.28	-15.03
$q_{0.01}[\Delta\%]$			-10.54	-26.50			-9.78	-24.66			-10.99	-23.82
mean R_H	0.81	0.50	0.40	0.34	0.81	0.51	0.40	0.35	0.81	0.51	0.40	0.35
std R_H	3.22	1.89	2.11	2.37	3.25	1.89	2.09	2.35	3.24	1.98	2.21	2.47
skewness R_H	-1.58	-0.43	0.13	0.24	-1.66	-0.75	-0.01	0.20	-1.50	-1.18	-0.22	0.13
kurtosis R_H	11.77	8.31	5.44	4.88	12.72	9.92	5.04	4.50	21.06	24.16	14.43	10.97

Out-of-sample simulations of hedging portfolio (P2) with the S&P futures. RS3 denotes the simulation from the estimated three-state RS model with normal components, bootstrap refers to sampling from the empirical distribution. The results in the last panel are based on simulations from a model that combines a t-copula with skewed-t margins. The columns uh and var describe the unhedged spot portfolio and the minimum-variance hedge. RS3-CVaR and np-CVaR correspond to CVaR_α-minimal strategies with $\alpha = 1\%$ derived from the RS model and the nonparametric benchmark approach. h is the hedging weight of the S&P futures. mean $\widehat{\text{CVaR}}^{np}_{1\%}$ is the average CVaR_α across simulations, mean $\Delta\%$ is the average risk reduction over the minimum-variance hedge across simulations. $q_u[\Delta\%]$ refers to the u-quantile of the additional risk reductions obtained across simulations. The last rows correspond to the moments of the hedged portfolio return across all simulated samples. Hedging weights, risk estimates, risk reductions, means and standard deviations in percent.

To assess the performance of composite CVaR_α hedging, we consider a third portfolio (P3), which consists of equally weighted investments into stocks, bonds and the S&P GSCI Total Return Commodity Index. In addition to the S&P futures, we include an oil futures contract, i.e., the NYMEX Light Crude Oil futures, as a second hedging instrument. We fit a three-state RS model to the joint return distribution of the spot portfolio and the two futures due to the promising results of these models in Section 4.3. The corresponding hedging weights and resulting risk estimates can be found in Table 9. Although the hedging amount in the oil futures does not differ much between the hedging strategies, we can again observe a reduction in tail risk by switching from the minimum-variance hedge to a CVaR_α-based approach, which ranges between 10% and 13% depending on the measurement technique. As in the univariate examples discussed before, this improvement is attained by increasing the hedging weight of the S&P futures. For comparison, we also include results of univariate hedging strategies using the S&P futures only. In this case, we again observe a large additional tail risk reduction by switching to the CVaR_α-optimal strategy. As expected, the remaining tail risk is higher than for the composite hedge.

Table 9: Composite Hedging

	(P3) univariate			(P3) composite		
strategy	uh	var	$\text{CVaR}_{1\%}$	uh	var	$\text{CVaR}_{1\%}$
h_1	0.00	45.96	73.12	0.00	43.82	58.69
h_2				0.00	15.56	14.85
$\widehat{\text{CVaR}}_{1\%}^{RS,u}$	14.65	9.01	7.01	14.02	5.97	5.35
$\Delta\%$			*22.24*			*10.31*
$\widehat{\text{CVaR}}_{1\%}^{np}$	13.51	8.26	7.07	13.51	5.73	4.97
$\Delta\%$			*14.35*			*13.30*
$\widehat{\text{CVaR}}_{1\%}^{evt}$	20.20	8.38	6.95	20.20	5.50	4.83
$\Delta\%$			*17.13*			*12.09*

In-sample results for composite hedging using two futures contracts. (P3): equally weighted investments into stocks, bonds and the S&P GSCI Total Return Commodity Index. See Table 6 for definitions of the strategies and the risk measures. h_1 denotes the hedging weight of the S&P futures. h_2 is the hedging weight of the oil futures.

In Table 10, we provide results for univariate hedging strategies derived from RS models with t-distributed components. For (P1) the findings are similar to the specification with normal components. The hedging weight for (P2), however, is close to that of the minimum-variance strategy. Looking at the parameters presented in the online appendix, we see that the model does not identify a crash state in this case, emphasizing the importance of this feature for our results.

Table 10: RS Models with t-Distributed Components

	(P1)			(P2)		
strategy	uh	var	$\text{CVaR}_{1\%}$	uh	var	$\text{CVaR}_{1\%}$
h	0.00	58.43	70.04	0.00	60.05	58.56
$\widehat{\text{CVaR}}_{1\%}^{RS,u}$	12.05	4.88	4.49	10.37	5.55	5.55
$\Delta\%$			*7.83*			*0.11*
$\widehat{\text{CVaR}}_{1\%}^{np}$	12.38	4.59	4.16	15.13	7.55	7.70
$\Delta\%$			*9.42*			*-1.95*
$\widehat{\text{CVaR}}_{1\%}^{evt}$	12.58	4.84	4.18	16.02	7.28	7.27
$\Delta\%$			*13.66*			*0.13*

Hedging results for three-state RS models with standardized t-distributed components and equal degrees of freedom across the components. (P1) and (P2) are defined in Table 2. h is the hedging weight of the S&P futures. See Table 6 for the row and column definitions.

Table 11 provides four robustness checks. In Panel A, we show that the results remain almost unchanged if discrete returns are used or if the MCVaR_α is optimized instead of the CVaR_α. In Panel B, we consider modifications of our data set. The results obtained using different indexes for the assets in the spot portfolio are similar to those of the original specification and we confirm that similar reductions can be attained without data from the subprime crisis, using the first half of the sample.[37]

Table 11: Robustness Checks

Panel A	(P2) MCVaR_α			(P2) Discrete returns		
strategy	uh	var	$\text{MCVaR}_{1\%}$	uh	var	$\text{CVaR}_{1\%}$
h	0.00	60.05	81.49	0.00	59.58	78.81
$\widehat{\text{(M)CVaR}}_{1\%}^{RS,u}$	16.55	7.92	6.71	15.00	7.15	6.18
$\Delta\%$			*15.23*			*13.67*
$\widehat{\text{(M)CVaR}}_{1\%}^{np}$	15.94	8.06	6.73	13.93	6.98	6.04
$\Delta\%$			*16.41*			*13.44*
$\widehat{\text{(M)CVaR}}_{1\%}^{evt}$	16.83	7.79	6.78	14.60	6.68	6.08
$\Delta\%$			*12.97*			*8.96*
Panel B	(P2) First sample half			(P2) Different spot indexes		
strategy	uh	var	$\text{CVaR}_{1\%}$	uh	var	$\text{CVaR}_{1\%}$
h	0.00	46.66	60.75	0.00	70.67	90.79
$\widehat{\text{CVaR}}_{1\%}^{RS,u}$	11.61	4.88	4.27	17.36	8.19	7.05
$\Delta\%$			*12.43*			*13.99*
$\widehat{\text{CVaR}}_{1\%}^{np}$	11.51	3.92	3.53	17.33	8.72	7.39
$\Delta\%$			*10.12*			*15.17*
$\widehat{\text{CVaR}}_{1\%}^{evt}$	13.39	3.85	3.39	17.32	8.41	7.15
$\Delta\%$			*11.94*			*14.97*

In Panel A, we provide results for using the MCVaR_α instead of the CVaR_α and for an analysis with discrete returns. (P2) is defined in Table 2. Panel B shows results for (P2) with modified time series. We replace our spot indexes with the MSCI All Country World Total Return Index, the BofA Merrill Lynch High Yield Master II Total Return Index and the FTSE/EPRA NAREIT North America Total Return Index using 290 return observations from May 1990 to June 2014. Finally, we report results for the first half of our original sample, for which we estimated RS models with two states. See Table 6 for descriptions of the rows and the columns.

5 Conclusion

In this paper, we study the use of finite mixtures and, in particular, regime-switching models for tail risk *management*. We provide a general characterization of VaR_α- and CVaR_α-minimal futures hedging strategies relying on results on quantile derivatives and show how to implement these characterizations for mixtures of elliptical distributions. Based on these results, we theoretically and empirically demonstrate that VaR_α and CVaR_α minimizations can change hedging strategies and tail risk characteristics as compared to variance minimizations

[37] For this case, we fitted a two-state RS model due to the small sample size.

in the presence of crash regimes. This observation might be especially relevant for institutional investors who can benefit from a better utilization of risk budgets and reduced capital requirements when hedging with such policies.

An interesting direction for future studies is the implementation of tail-risk-minimal hedging based on RS models with more elaborate time series structures, which could be particularly relevant for the application of such strategies with daily and weekly data. Last but not least, applications of the RS approach to other portfolio selection problems involving tail risk constraints or objectives seem to be an interesting area for further research.

Appendix

Proof of Proposition 1: First, we define the loss function $l_H : \mathbb{R}^N \times \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}$ for a given vector of portfolio weights $\boldsymbol{w}$

$$l_H(\boldsymbol{r}_S, \boldsymbol{r}_F, \boldsymbol{h}) = -\boldsymbol{w}' \cdot \boldsymbol{r}_S + \boldsymbol{h}' \cdot \boldsymbol{r}_F, \tag{34}$$

such that $L_H(\boldsymbol{h}) = l_H(\boldsymbol{R}_S, \boldsymbol{R}_F, \boldsymbol{h})$. With this definition, the Assumptions (R1) - (R3) imply that the conditions for Theorem 2 in Hong (2009) are satisfied. The conditions from Assumption 1 in Hong (2009), i.e., the partial differentiability of the loss function and its Lipschitz continuity, are implied by the linear structure of the function in (34) and the integrability constraints in (R1). (R2) is a global version of Assumption 2 in Hong (2009) with the additional requirement that the density is positive, which ensures the uniqueness of the VaR_α. Since eventually, $\frac{\partial l_H}{\partial h_j} = r_{F,j}$, (R3) corresponds to Assumption 3, such that we can invoke Theorem 2 from Hong (2009) for the $(1-\alpha)$-quantile $q_{1-\alpha}[l_H(\boldsymbol{R}_S, \boldsymbol{R}_F, \boldsymbol{h})] = v_\alpha(\boldsymbol{h})$ to obtain

$$\frac{\partial v_\alpha(\boldsymbol{h})}{\partial h_j} = \mathbb{E}\left[\frac{\partial l_H}{\partial h_j}(\boldsymbol{R}_S, \boldsymbol{R}_F, \boldsymbol{h}) \mid l_H(\boldsymbol{R}_S, \boldsymbol{R}_F, \boldsymbol{h}) = v_\alpha(\boldsymbol{h})\right]. \tag{35}$$

Again, with $\frac{\partial l_H}{\partial h_j} = r_{F,j}$ the componentwise application of this result for $\boldsymbol{h}$ implies that (6) contains the FOCs for (4). These FOCs must be satisfied by the global minimizer of v_α since the optimization problem is unconstrained. However, due to $\boldsymbol{h} \in \mathbb{R}^M$, the objective function may be unbounded, in which case (4) has no solution. This is also true for (5). (7) follows as FOC for this problem from Theorem 3.1 in Hong and Liu (2009), which may be applied since (R1) - (R3) imply that also the necessary conditions therein are satisfied. In particular, the differentiability of v_α follows from the first part of this proof. We thus obtain

$$\frac{\partial c_\alpha(\boldsymbol{h})}{\partial h_j} = \mathbb{E}\left[\frac{\partial l_H}{\partial h_j}(\boldsymbol{R}_S, \boldsymbol{R}_F, \boldsymbol{h}) \mid l_H(\boldsymbol{R}_S, \boldsymbol{R}_F, \boldsymbol{h}) \geq v_\alpha(\boldsymbol{h})\right], \tag{36}$$

which proves (7).

Proof of Theorem 1: First, we note that it is not difficult to show that the joint distribution of $\boldsymbol{R}_F$ and L_H is given by

$$\begin{pmatrix} \boldsymbol{R}_F \\ L_H \end{pmatrix} \mid S = k \sim \mathcal{E}_{M+1}(\begin{pmatrix} \boldsymbol{\mu}_{F,k} \\ \mu_{L,k} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{F,k} & \boldsymbol{\Sigma}_{FL,k} \\ \boldsymbol{\Sigma}'_{FL,k} & \sigma^2_{L,k} \end{pmatrix}, g_k), \tag{37}$$

where the parameters are calculated according to (14), (15) and $\boldsymbol{\Sigma}_{FL,k} = -\boldsymbol{\Sigma}'_{SF,k} \cdot \boldsymbol{w} + \boldsymbol{\Sigma}_{F,k} \cdot \boldsymbol{h}$. To derive the FOCs for VaR_α-minimal hedging we first rewrite the general expressions presented in Proposition 1 in terms of conditional expectations for the component distributions and then use the properties of elliptical distributions to give explicit representations of these expectations. Due to the positivity of the density generators in (M1), we can write the expectation from (6) as $\mathbb{E}[\boldsymbol{R}_F \mid L_H = l] = f_L(l)^{-1} \cdot \mathbb{E}[\boldsymbol{R}_F \ \mathbb{1}(L_H = l)]$, with f_L given by (16). Using (37), this expectation can be decomposed into

$$\mathbb{E}[\boldsymbol{R}_F \mid L_H = l] = \sum_{k=1}^{K} \frac{\pi_k}{f_L(l)} \ \mathbb{E}[\boldsymbol{R}_F \ \mathbb{1}(L_H = l) \mid S = k] \tag{38}$$

$$= \sum_{k=1}^{K} \frac{\pi_k \ f_{L,k}(l)}{f_L(l)} \ \mathbb{E}[\boldsymbol{R}_F \mid L_H = l, \, S = k]\,. \tag{39}$$

We now exploit the fact that the component distributions are elliptical. In particular, we use the regression property of elliptical distributions (Owen and Rabinovitch, 1983, P.2) and obtain

$$\mathbb{E}[\boldsymbol{R}_F \mid L_H = l] = \sum_{k=1}^{K} \frac{\pi_k \ f_{L,k}(l)}{f_L(l)} \ \left[\boldsymbol{\mu}_{F,k} + \frac{\boldsymbol{\Sigma}_{FL,k}}{\sigma^2_{L,k}} \ (l - \mu_{L,k}) \right]. \tag{40}$$

This proves (21) for $l = v_\alpha(\boldsymbol{h})$ and $z_k(\boldsymbol{h}) = \frac{v_\alpha(\boldsymbol{h}) - \mu_{L,k}}{\sigma_{L,k}}$. Since we assumed the density generators to be continuous, this also holds for the involved densities in (40) so that $\mathbb{E}[R_{F,j} \mid L_H = l]$ as a function of l is continuous for all $j = 1, \ldots, M$, which implies that (R3) is valid under (M1).

For the derivation of the CVaR_α-minimal hedging strategy, we conclude by the same reasoning that

$$\mathbb{E}[\boldsymbol{R}_F \mid L_H \geq l] = \sum_{k=1}^{K} \frac{\pi_k \ \mathrm{P}(L_H \geq l | S = k)}{\mathrm{P}(L_H \geq l)} \ \mathbb{E}[\boldsymbol{R}_F \mid L_H \geq l, S = k]\,. \tag{41}$$

Denoting the density of L_H conditional on $L_H \geq l$ and $S = k$ by $f_{L_H | L_H \geq l, \, S = k}$, we can rewrite the involved conditional expectations as

$$\mathbb{E}[\boldsymbol{R}_F \mid L_H \geq l, S = k] = \int_l^\infty \mathbb{E}[\boldsymbol{R}_F \mid L_H = x, \, S = k] \cdot f_{L_H | L_H \geq l, S = k}(x) \ \lambda(dx). \tag{42}$$

Again using the regression property of elliptical distributions and the linearity of the integration operator, it follows that

$$\mathbb{E}[\boldsymbol{R}_F \mid L_H \geq l, \, S = k] = \boldsymbol{\mu}_{F,k} + \frac{\boldsymbol{\Sigma}_{FL,k}}{\sigma^2_{L,k}} \ \left[\mathbb{E}[L_H \mid L_H \geq l, S = k] - \mu_{L,k} \right]. \tag{43}$$

We conclude that

$$\mathbb{E}[\boldsymbol{R}_F \mid L_H \geq l]$$
$$= \sum_{k=1}^{K} \frac{\pi_k \ (1 - F_{L,k}(l))}{\mathrm{P}(L_H \geq l)} \left[\boldsymbol{\mu}_{F,k} + \frac{\boldsymbol{\Sigma}_{FL,k}}{\sigma^2_{L,k}} \ (\mathbb{E}[L_H \mid L_H \geq l, S = k] - \mu_{L,k}) \right]. \quad (44)$$

With $Z_k \sim \mathcal{E}_1(0, 1, g_k)$, it holds that

$$\mathbb{E}[L_H \mid L_H \geq l, S = k] = \mu_{L,k} + \sigma_{L,k} \ \mathbb{E}\left[Z_k \mid Z_k \geq \frac{l - \mu_{L,k}}{\sigma_{L,k}} \right]. \quad (45)$$

Again, for $l = v_\alpha(\boldsymbol{h})$ and with the definitions of $z_k(\boldsymbol{h})$ and $\lambda_k(\boldsymbol{h})$, we obtain (22) because $\mathrm{P}(L_H \geq v_\alpha(\boldsymbol{h})) = \alpha$. It remains to verify that Assumption (R2) is satisfied in our setting. This follows from the assumed continuity of the density generators and the observation that the cdf of L_H can be written as $F_{L_H}(l, \boldsymbol{h}) = \sum_{k=1}^{K} \pi_k \ F_{Z_k} \left(\frac{l - \mu_{L,k}(\boldsymbol{h})}{\sigma_{L,k}(\boldsymbol{h})} \right)$ with $Z_k \sim \mathcal{E}_1(0, 1, g_k)$.

References

Alexander, C. and Barbosa, A. (2007). Effectiveness of minimum-variance hedging. *Journal of Portfolio Management*, 33(2):46–59.

Alexander, G. J. and Baptista, A. M. (2002). Economic implications of using a mean-VaR model for portfolio selection: A comparison with mean-variance analysis. *Journal of Economic Dynamics and Control*, 26(7–8):1159–1193.

Alexander, G. J. and Baptista, A. M. (2004). A comparison of VaR and CVaR constraints on portfolio selection with the mean-variance model. *Management Science*, 50(9):1261–1273.

Alexander, G. J. and Baptista, A. M. (2008). Active portfolio management with benchmarking: Adding a value-at-risk constraint. *Journal of Economic Dynamics and Control*, 32(3):779–820.

Alizadeh, A. H., Nomikos, N., and Pouliasis, P. K. (2008). A Markov regime switching approach for hedging energy commodities. *Journal of Banking & Finance*, 32(9):1970–1983.

Ang, A. and Bekaert, G. (2002). International asset allocation with regime shifts. *Review of Financial Studies*, 15(4):1137–1187.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent Measures of Risk. *Mathematical Finance*, 9(3):203–228.

Arzac, E. R. and Bawa, V. S. (1977). Portfolio choice and equilibrium in capital markets with safety-first investors. *Journal of Financial Economics*, 4(3):277–288.

Baillie, R. T. and Myers, R. J. (1991). Bivariate garch estimation of the optimal commodity futures Hedge. *Journal of Applied Econometrics*, 6(2):109–124.

Barbi, M. and Romagnoli, S. (2014). A copula-based quantile risk measure approach to estimate the optimal hedge ratio. *Journal of Futures Markets*, 34(7):658–675.

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, 19(4):465–474.

Bertsimas, D., Lauprete, G., and Samarov, A. (2004). Shortfall as a risk measure: properties, optimization and applications. *Journal of Economic Dynamics and Control*, 28(7):1353–1381.

Billio, M. and Pelizzon, L. (2000). Value-at-risk: a multivariate switching regime approach. *Journal of Empirical Finance*, 7(5):531–554.

Buckley, I., Saunders, D., and Seco, L. (2008). Portfolio optimization when asset returns have the gaussian mixture distribution. *European Journal of Operational Research*, 185(3):1434–1461.

Cao, Z., Harris, R. D. F., and Shen, J. (2010). Hedging and value at risk: A semi-parametric approach. *Journal of Futures Markets*, 30(8):780–794.

Chang, K.-L. (2010). The optimal value-at-risk hedging strategy under bivariate regime switching ARCH framework. *Applied Economics*, 43(21):2627–2640.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4):841–862.

Ederington, L. H. (1979). The Hedging Performance of the New Futures Markets. *Journal of Finance*, 34(1):157–170.

Embrechts, P., McNeil, A. J., and Straumann, D. (2002). Correlation and dependence in risk management: Properties and pitfalls. In Dempster, M., editor, *Risk Management: Value at Risk and Beyond*, pages 176–223. Cambridge University Press.

Fabozzi, F. J., Huang, D., and Zhou, G. (2010). Robust portfolios: contributions from operations research and finance. *Annals of Operations Research*, 176(1):191–220.

Fang, K.-T., Kotz, S., and Ng, K. W. (1990). *Symmetric multivariate and related distributions*. Chapman & Hall, London.

Figlewski, S. (1984). Hedging Performance and Basis Risk in Stock Index Futures. *The Journal of Finance*, 39(3):657–669.

Garcia, R. and Tsafack, G. (2011). Dependence structure and extreme comovements in international equity and bond markets. *Journal of Banking & Finance*, 35(8):1954–1970.

Gourieroux, C., Laurent, J., and Scaillet, O. (2000). Sensitivity analysis of Values at Risk. *Journal of Empirical Finance*, 7(3-4):225–245.

Guidolin, M. and Timmermann, A. (2006). Term structure of risk under alternative econometric specifications. *Journal of Econometrics*, 131(1–2):285–308.

Guidolin, M. and Timmermann, A. (2008). International asset allocation under regime switching, skew, and kurtosis preferences. *Review of Financial Studies*, 21(2):889–935.

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and The Business Cycle. *Econometrica*, 57(2):357–384.

Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45(1–2):39–70.

Harris, R. D. F. and Shen, J. (2006). Hedging and value at risk. *Journal of Futures Markets*, 26(4):369–390.

Hilal, S., Poon, S.-H., and Tawn, J. (2011). Hedging the black swan: Conditional heteroskedasticity and tail dependence in S&P500 and VIX. *Journal of Banking & Finance*, 35(9):2374–2387.

Hong, L. J. (2009). Estimating Quantile Sensitivities. *Operations Research*, 57(1):118–130.

Hong, L. J. and Liu, G. (2009). Simulating Sensitivities of Conditional Value at Risk. *Management Science*, 55(2):281–293.

Johnson, L. L. (1960). The Theory of Hedging and Speculation in Commodity Futures. *The Review of Economic Studies*, 27(3):139–151.

Kellner, R. and Rösch, D. (2016). Quantifying market risk with Value-at-Risk or Expected Shortfall? – Consequences for capital requirements and model risk. *Journal of Economic Dynamics and Control*, 68(C):45–63.

Litzenberger, R. H. and Modest, D. M. (2010). Crisis and noncrisis risk in financial markets: A unified approach to risk management. In Diebold, F. X., Doherty, N. A., and Herring, R. J., editors, *The Known, the Unknown, and the Unknowable in Financial Risk Management*, pages 74–102. Princeton University Press.

McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. Wiley-Interscience, New York.

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7(3–4):271–300.

McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton University Press, Princeton, Oxford.

Owen, J. and Rabinovitch, R. (1983). On the Class of Elliptical Distributions and their Applications to the Theory of Portfolio Choice. *Journal of Finance*, 38(3):745–752.

Patton, A. J. (2004). On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *Journal of Financial Econometrics*, 2(1):130–168.

Perez-Quiros, G. and Timmermann, A. (2000). Firm Size and Cyclical Variations in Stock Returns. *The Journal of Finance*, 55(3):1229–1262.

Rockafellar, R. T. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26(7):1443–1471.

Scaillet, O. (2004). Nonparametric estimation and sensitivity analysis of expected shortfall. *Mathematical Finance*, 14(1):115–129.

Timmermann, A. (2000). Moments of Markov switching models. *Journal of Econometrics*, 96(1):75–111.