Tail Risk Hedging and Regime Switching

Markus Huggenberger*, Peter Albrecht, Alexandr Pekelis

University of Mannheim, Germany

This version: June 21, 2015

First version: August 15, 2011

Abstract: In this paper, we analyze futures-based hedging strategies which minimize tail risk measured by Value-at-Risk (VaR) and Conditional-Value-at-Risk (CVaR). In particular, we first deduce general characterizations of VaR- and CVaR-minimal hedging policies from results on quantile derivatives. We then derive first-order conditions for tail-risk-minimal hedging in mixture and regime-switching (RS) models. Using cross hedging examples, we show that CVaR-minimal hedging can noticeably deviate from standard minimum-variance hedging if the return data exhibit nonelliptical features. In our examples, we find an increase in hedging amounts if RS models identify a joint crash scenario and we confirm a reduction in tail risk using empirical and EVT-based risk estimators. These results imply that switching from minimum-variance to CVaR-minimal hedging 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.

1 Introduction

After maturing into standard tools for risk measurement, especially for setting capital requirements, Value-at-Risk (VaR) and Conditional-Value-at-Risk (CVaR) are increasingly adopted as decision tools for active risk management in financial institutions. Focusing on the latter, this paper aims to develop static futures hedging policies that minimize tail risk measured by VaR or CVaR.[1] This approach is of special interest for agents facing risk limits or capital requirements set with these measures. In addition, it is 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 the 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.[3] Within this framework, influential portfolio selection studies incorporating (C)VaR objectives or restrictions are Alexander and Baptista (2002, 2004) as well as Bertsimas et al. (2004).[4] 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 impact

---

[1]   Both VaR and CVaR quantify the extent of losses in the upper tail of the loss distribution. VaR has often been criticized for not considering the severity of the highest losses. Therefore, CVaR, which is also coherent, might be the better measure of tail risk. However, due to the importance of VaR, we include both measures in our analysis.

[2]   Thereby, this approach relates to the traditional literature on safety-first and lower partial moment hedging and portfolio optimization (Telser, 1955; Fishburn, 1977; Arzac and Bawa, 1977).

[3]   Technically this means the density generator, which may contain additional parameters like the degrees of freedom in case of the t-distribution. However, these additional parameters are a property of the multivariate model which is invariant under different portfolio compositions.

[4]   The authors argue that a mean-variance-based approach to VaR risk management can be justified as approximation by Tschebycheff's inequality. Bertsimas et al. (2004) provides a similar variance-based bound for CVaR.

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, 2000, 2002; Campbell et al., 2001; Agarwal and Naik, 2004; Gaivoronski and Pflug, 2005; 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 are applied in the risk and portfolio management literature, focusing on non-normalities (Patton, 2004).[5] However, such models do usually 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.

Against this background, we propose to use regime-switching (RS) models based on elliptical distributions for tail risk management decisions. Regime switching models were first introduced by Hamilton (1989) in a univariate setting and then 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 (Timmermann, 2000) and asymmetric dependence structures (Ang and Chen, 2002). 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 dynamic strategies within a variance-based setup (Tu, 2010; Alizadeh et al., 2008). Chang (2010) analyzes univariate VaR-minimal hedging, however using a numerical search algorithm to determine the optimal policy. Related to our work is in particular Buckley et al. (2008), who

---

[5] A further alternative, recently proposed in a number of studies, is the use of robust optimization techniques with VaR and CVaR. See Fabozzi et al. (2010) for a comprehensive overview.

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 rules which 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[6] and RS processes with normally and t-distributed components.

In the empirical part of our paper, we present cross-hedging examples demonstrating the advantage of tail-risk-minimal hedging over minimum-variance hedging when the mixture approach is used. In particular, futures hedging for multi-asset investment portfolios with returns exhibiting nonelliptical features is investigated. We consider a monthly hedging horizon, which allows us to focus on distributional aspects and keep the time series structure of our models relatively simple. We estimate multivariate RS models with Gaussian conditional distributions, and find that they produce reliable tail risk estimates. The stationary distribution of these models is then used to derive CVaR-minimal hedging rules for the selected portfolios. In all cases, we find an increase in the hedging demand compared to the traditional minimum-variance approach, which can be attributed to a *joint* (low-probability) crash state identified by the RS models. We show that the reduction in tail risk obtained by switching from minimum-variance to tail-risk-minimal hedging can reach 20%. This result is confirmed – independent from our model – by univariate empirical and EVT-based estimators, which is especially important if such standard procedures are used to set the capital requirements or risk limits for the optimized positions. We confirm our findings in out-of-sample backtests and perform a simulation experiment that allows for a more reliable risk estimation than within the relatively small samples available for the backtests. We finally give evidence for a superior performance of our approach in dynamic and composite hedging setups.

---

[6]   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.

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 and the application of these results to RS models. In Section 4, we document 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 $R_{F,j}$, $j = 1, \ldots, M$.[7] 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 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

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

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 re-

---

[7] 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.

quires that $R_{S,i} \in \mathcal{L}^2$ and $R_{F,j} \in \mathcal{L}^2$ for $i = 1, \ldots, N$ and $j = 1, \ldots, M$. It is easy to show that the hedging policy $\boldsymbol{h}^*_{\text{var}}$ solving this problem is given by

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

Large parts of the literature on futures hedging are centered around implementing dynamic specifications for the covariance terms in (2), conditioning these 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), Kroner and Sultan (1993) and Brooks et al. (2002).

In contrast, our focus lies on hedging strategies that minimize the tail risk or the corresponding capital requirement. This is usually measured in terms of VaR_α or CVaR_α, which for $\alpha \in (0, 1)$ and the confidence level $1 - \alpha$ are defined as[8]

$$(3) \qquad \text{VaR}_\alpha[L_H] = \inf \{l \in \mathbb{R} | \text{P}(L_H \leq l) \geq 1 - \alpha\}$$

and

$$(4) \qquad \text{CVaR}_\alpha[L_H] = \frac{\text{P}(L_H > \text{VaR}_\alpha[L_H])}{\alpha} \cdot \mathbb{E}[L_H \mid L_H > \text{VaR}_\alpha[L_H]] \\ + \left(\frac{\alpha - \text{P}(L_H > \text{VaR}_\alpha[L_H])}{\alpha}\right) \text{VaR}_\alpha[L_H].$$

Accordingly, VaR_α can be understood as the smallest loss value, which is not exceeded with a probability of at least $1 - \alpha$. Formally, VaR_α simply corresponds to the lower $(1 - \alpha)$-quantile $q_{1-\alpha}[L_H]$ of the loss distribution. CVaR_α is the expected loss in the worst $100 \cdot \alpha\%$ of the cases. In general, it is thus defined as a convex combination of VaR_α, which has a positive weight for $\text{P}(L_H > \text{VaR}_\alpha[L_H]) < \alpha$, and the conditional expectation of losses exceeding VaR_α. Comparing both measures, VaR_α is still dominant in industry applications, although CVaR_α is preferable from an axiomatic point of view as a coherent risk measure in the sense of Artzner et al. (1999). Moreover, VaR_α might be questionable if the aim is to avoid large losses since it does not consider the extent of losses in the very tail of the distribution.

---

[8] See Rockafellar and Uryasev (2002) for definitions of this type. If $\text{P}(L_H > \text{VaR}_\alpha[L_H]) = 0$, which is possible for discrete loss distributions, we set $\text{CVaR}_\alpha[L_H] = \text{VaR}_\alpha[L_H]$.

The choice between VaR_α and CVaR_α, however, remains a matter of practical and academic debate (Embrechts and Hofert, 2014). We, therefore, consider both measures in our analysis. Writing these risk measures as functions of the hedging weights, i.e., $v_\alpha(\boldsymbol{h}) := \text{VaR}_\alpha[L_H(\boldsymbol{h})]$ and $c_\alpha(\boldsymbol{h}) := \text{CVaR}_\alpha[L_H(\boldsymbol{h})]$, we analyze

$$(5) \qquad \min_{\boldsymbol{h} \in \mathbb{R}^M} v_\alpha(\boldsymbol{h}) = \min_{\boldsymbol{h} \in \mathbb{R}^M} \text{VaR}_\alpha[L_H(\boldsymbol{h})],$$

$$(6) \qquad \min_{\boldsymbol{h} \in \mathbb{R}^M} c_\alpha(\boldsymbol{h}) = \min_{\boldsymbol{h} \in \mathbb{R}^M} \text{CVaR}_\alpha[L_H(\boldsymbol{h})].$$

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 (2000, 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 (5) and (6) 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 issue.

2.2 A General Solution

Complementing the mentioned results on non- and semiparametric VaR_α and CVaR_α hedging, we are interested in analytic characterizations of the solutions to (5) and (6). 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]

(A1) $R_{S,i} \in \mathcal{L}^1$ and $R_{F,j} \in \mathcal{L}^1$ for $i = 1, \ldots, N$ and $j = 1, \ldots, M$.

---

[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).

(A2) 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.

(A3) For all $j = 1, \ldots, M$, the conditional expectations $\mathbb{E}[R_{F,j} \mid L_H = l]$ are continuous as functions of l.

(A1) is obviously weaker than the corresponding integrability requirements needed for the variance-based approach. However, (A2) and (A3) define some additional continuity and differentiability conditions. Note that (A2) implies the following simplified VaR_α and CVaR_α representations[10]

$$(7) \qquad \mathrm{P}(L_H \leq \mathrm{VaR}_\alpha[L_H]) = 1 - \alpha \quad \text{and} \quad \mathrm{CVaR}_\alpha[L_H] = \mathbb{E}[L_H \mid L_H \geq \mathrm{VaR}_\alpha[L_H]].$$

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

Proposition 1 *Under (A1) - (A3) VaR_α- and CVaR_α-minimal hedging policies $\boldsymbol{h}_{\mathrm{VaR}}^*$ and $\boldsymbol{h}_{\mathrm{CVaR}}^*$, i.e., solutions to (5) and (6), satisfy*

$$(8) \qquad \qquad \mathbb{E}[\boldsymbol{R}_F \mid L_H(\boldsymbol{h}_{\mathrm{VaR}}^*) = v_\alpha(\boldsymbol{h}_{\mathrm{VaR}}^*)] = \boldsymbol{0}_M,$$

$$(9) \qquad \qquad \mathbb{E}[\boldsymbol{R}_F \mid L_H(\boldsymbol{h}_{\mathrm{CVaR}}^*) \geq v_\alpha(\boldsymbol{h}_{\mathrm{CVaR}}^*)] = \boldsymbol{0}_M.$$

This characterization is an application of results on quantile derivatives to the hedging problem. In particular, (8) and (9) follow as FOCs of (5) and (6) from Theorem 2 in Hong (2009) and Theorem 3.1 in Hong and Liu (2009).[11] Some technical details of this reasoning can be found in the Appendix.[12]

Note that Proposition 1 makes no statement on the existence of optimal strategies. As already pointed out in Alexander and Baptista (2004) portfolio selection strategies, it is possible that

---

[10] While we work with these simplified versions throughout the theoretical sections, (3) and (4) will be used in the empirical analysis in Section 4.

[11] Gourieroux et al. (2000) and Hong (2009) demonstrate that quantile derivatives could also be used to implement gradient-based search algorithms for the solution of portfolio optimization problems involving VaR_α and CVaR_α.

[12] Earlier results on quantile derivatives, e.g., Gourieroux et al. (2000), Tasche (2002) or Scaillet (2004) could also be applied to obtain (8) and (9).

VaR_α and CVaR_α minimizations have no solutions even with normally distributed returns.[13] Moreover, there is an important difference between using v_α and c_α as objective functions. Whereas the c_α-FOC (9) is only fulfilled by the global minimizer of (6), the v_α-FOC (8) 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 (6) always is 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.[14] In such cases, (8) will uniquely characterize the global VaR_α-minimal hedging vector (if such a strategy exists).

We note that it might be interesting to apply tail risk measures to the demeaned loss variables instead of the losses themselves. In contrast to the variance, both VaR_α and CVaR_α depend on the expected value of the underlying loss random variable. They contain an implicit trade-off between the location and the dispersion of the loss distribution. In order to improve the comparability between minimum (C)VaR_α and minimum-variance strategies, we consider the following demeaned modifications[15] of these tail risk measures

$$(10) \qquad \mathrm{MVaR}_\alpha[L_H] := \mathrm{VaR}_\alpha[L_H - \mathbb{E}[L_H]] = \mathrm{VaR}_\alpha[L_H] - \mathbb{E}[L_H],$$

$$(11) \qquad \mathrm{MCVaR}_\alpha[L_H] := \mathrm{CVaR}_\alpha[L_H - \mathbb{E}[L_H]] = \mathrm{CVaR}_\alpha[L_H] - \mathbb{E}[L_H].$$

By construction, MVaR_α and MCVaR_α do not allow to reduce the risk of the position by increasing its expected return. The corresponding optimization problems are

$$(12) \qquad \min_{\boldsymbol{h} \in \mathbb{R}^M} \mathrm{MVaR}_\alpha[L_H(\boldsymbol{h})] \qquad \text{and} \qquad \min_{\boldsymbol{h} \in \mathbb{R}^M} \mathrm{MCVaR}_\alpha[L_H(\boldsymbol{h})].$$

Under (A1) - (A3), FOCs for the solutions to (12) follow from Proposition 1 by noting that $\frac{\partial}{\partial \boldsymbol{h}} \mathbb{E}[L_H(\boldsymbol{h})] = \mathbb{E}[\boldsymbol{R}_F]$. Therefore, such strategies must satisfy

$$(13) \qquad \mathbb{E}[\boldsymbol{R}_F \mid L_H(\boldsymbol{h}^*_{\mathrm{MVaR}}) = v_\alpha(\boldsymbol{h}^*_{\mathrm{MVaR}})] - \mathbb{E}[\boldsymbol{R}_F] = \boldsymbol{0}_M,$$

$$(14) \qquad \mathbb{E}[\boldsymbol{R}_F \mid L_H(\boldsymbol{h}^*_{\mathrm{MCVaR}}) \geq v_\alpha(\boldsymbol{h}^*_{\mathrm{MCVaR}})] - \mathbb{E}[\boldsymbol{R}_F] = \boldsymbol{0}_M.$$

---

[13] In our online appendix, we provide conditions that guarantee the existence of solutions to (5) and (6) under stronger distributional assumptions.

[14] We refer to Daníelsson et al. (2013) for an overview on recent findings concerning this issue.

[15] See Rockafellar et al. (2006) for a general treatment of the so-called deviation measures.

Since the conditions in (A1) - (A3) are rather weak, Proposition 1 applies to a wide range of continuous return distributions. However, at this level of generality, we cannot provide explicit representations for the conditional expectations in equations (8), (9) and (13), (14). We therefore analyze more specific distributional assumptions in the following section.

3 Tail 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

$$(15) \qquad f_{\boldsymbol{\mu}, \boldsymbol{\Sigma}, g}(\boldsymbol{y}) = \det(\boldsymbol{\Sigma})^{-\frac{1}{2}} \, g_P\big((\boldsymbol{y} - \boldsymbol{\mu})' \cdot \boldsymbol{\Sigma}^{-1} \cdot (\boldsymbol{y} - \boldsymbol{\mu})\big),$$

where g_P is a non negative scalar function on $\mathbb{R}$. This function is referred to as density generator. Since g_P is parameterized by the dimension of $\boldsymbol{Y}$, we need a collection of generators $g = (g_P)_{P \in \mathbb{N}}$ to define a distribution over several dimensions. 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, in a portfolio context especially the behavior under linear transformations (Owen and Rabinovitch, 1983, P.1).[16]

---

[16]   For a full account of elliptical distributions, we refer to Kelker (1970), Fang et al. (1990) or McNeil et al. (2005).

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

$$(16) \qquad f_{\boldsymbol{Y}}(\boldsymbol{y}) = \sum_{k=1}^{K} \pi_k f_k(\boldsymbol{y}).$$

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 combined.[17] 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 (16) is obtained from the law of total probability.

Combining (15) with (16), 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

$$(17) \qquad f_{\boldsymbol{R}}(\boldsymbol{r}) = \sum_{k=1}^{K} \pi_k \det(\boldsymbol{\Sigma}_k)^{-\frac{1}{2}} g_{N+M,k} \left((\boldsymbol{r} - \boldsymbol{\mu}_k)' \cdot \boldsymbol{\Sigma}_k^{-1} \cdot (\boldsymbol{r} - \boldsymbol{\mu}_k) \right)$$

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 $\mathrm{P}(S = k) = \pi_k$.

---

[17] For extensive discussions of the properties of this modeling approach and for illustrations of its flexibility, we refer to McLachlan and Peel (2000) or Frühwirth-Schnatter (2006).

This setting obviously includes popular modeling choices like mixtures of multivariate normals or multivariate t-distributions.[18]

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

$$(18) \qquad \mathbb{E}[\boldsymbol{R}] = \sum_{k=1}^{K} \pi_k \, \boldsymbol{\mu}_k \quad \text{and} \quad \mathrm{cov}[\boldsymbol{R}] = \sum_{k=1}^{K} \pi_k \left[c_{g_k} \boldsymbol{\Sigma}_k + \boldsymbol{\mu}_k \cdot \boldsymbol{\mu}_k' \right] - \mathbb{E}[\boldsymbol{R}] \cdot \mathbb{E}[\boldsymbol{R}'] .$$

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 (18) for the traditional minimum-variance hedging weights[19]

$$(19) \quad \boldsymbol{h}_{\mathrm{var}}^* = \left[\sum_{k=1}^{K} \pi_k \left[c_{g_k} \boldsymbol{\Sigma}_{F,k} + \boldsymbol{\mu}_{F,k} \cdot \boldsymbol{\mu}_{F,k}' \right] - \sum_{k=1}^{K} \pi_k \, \boldsymbol{\mu}_{F,k} \cdot \sum_{k=1}^{K} \pi_k \, \boldsymbol{\mu}_{F,k}' \right]^{-1}$$
$$\cdot \left[\sum_{k=1}^{K} \pi_k \cdot \left[c_{g_k} \boldsymbol{\Sigma}_{SF,k}' + \boldsymbol{\mu}_{F,k} \cdot \boldsymbol{\mu}_{S,k}' \right] - \sum_{k=1}^{K} \pi_k \, \boldsymbol{\mu}_{F,k} \cdot \sum_{k=1}^{K} \pi_k \, \boldsymbol{\mu}_{S,k}' \right] \cdot \boldsymbol{w}.$$

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.,

$$(20) \qquad\qquad L_H(\boldsymbol{h}) \mid S = k \sim \mathcal{E}_1(\mu_{L,k}, \, \sigma_{L,k}^2, \, g_k),$$

where

$$(21) \qquad\qquad \mu_{L,k} := \mu_{L,k}(\boldsymbol{h}) = -\boldsymbol{w}' \cdot \boldsymbol{\mu}_{S,k} + \boldsymbol{h}' \cdot \boldsymbol{\mu}_{F,k},$$

$$(22) \qquad\qquad \sigma_{L,k}^2 := \sigma_{L,k}^2(\boldsymbol{h}) = \boldsymbol{w}' \cdot \boldsymbol{\Sigma}_{S,k} \cdot \boldsymbol{w} - 2\,\boldsymbol{w}' \cdot \boldsymbol{\Sigma}_{SF,k} \cdot \boldsymbol{h} + \boldsymbol{h}' \cdot \boldsymbol{\Sigma}_{F,k} \cdot \boldsymbol{h},$$

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 (15) and (16), the component densities and the unconditional density $f_L := f_{L_H}$

---

[18] See Kamdem (2009) for a general discussion of mixtures of elliptical distributions in a risk measurement context.

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

satisfy

$$(23) \qquad f_{L,k}(l) = \sigma_{L,k}^{-1} \cdot g_{1,k}\left(\frac{(l-\mu_{L,k})^2}{\sigma_{L,k}^2}\right) \quad \text{and} \quad f_L(l) = \sum_{k=1}^{K} \pi_k \, f_{L,k}(l).$$

The tail risk measures that we analyze are given by

$$(24) \qquad 1-\alpha = \sum_{k=1}^{K} \pi_k F_{L,k}(v_\alpha(\boldsymbol{h})),$$

$$(25) \qquad c_\alpha(\boldsymbol{h}) = \frac{1}{\alpha}\sum_{k=1}^{K} \pi_k \, \mathbb{E}[L_H \, \mathbb{1}(L_H \geq v_\alpha(\boldsymbol{h})) \mid S = k].$$

The simple VaR_α characterization in (24) 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 $k = 1, \ldots, K$, and setting

$$(26) \qquad z_k(\boldsymbol{h}) := \frac{v_\alpha(\boldsymbol{h}) - \mu_{L,k}(\boldsymbol{h})}{\sigma_{L,k}(\boldsymbol{h})}, \qquad \lambda_k(\boldsymbol{h}) := \mathbb{E}[Z_k \mid Z_k \geq z_k(\boldsymbol{h})],$$

we can rewrite (25) in terms of the location and scale parameters of the mixture as

$$(27) \qquad c_\alpha(\boldsymbol{h}) = \frac{1}{\alpha}\sum_{k=1}^{K} \pi_k \, (1 - F_{L,k}(v_\alpha(\boldsymbol{h}))) \, [\mu_{L,k}(\boldsymbol{h}) + \sigma_{L,k}(\boldsymbol{h}) \, \lambda_k(\boldsymbol{h})].$$

Given $v_\alpha(\boldsymbol{h})$, (27) can usually be evaluated explicitly for specific density generators $k = 1, \ldots, K$. In contrast, the implicit VaR_α definition in (24) 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.[20] However, applying Proposition 1, we are able to obtain such conditions, which we present in the following Theorem.

Theorem 1 *If (A1) and (M1) hold, the* VaR_α*-minimal hedging strategy* $\boldsymbol{h}^*_{\text{VaR}}$ *solves*

$$(28) \qquad \sum_{k=1}^{K} \frac{\pi_k f_{L,k}(v_\alpha(\boldsymbol{h}^*_{\text{VaR}}))}{f_L(v_\alpha(\boldsymbol{h}^*_{\text{VaR}}))} \left[\boldsymbol{\mu}_{F,k} + \frac{\boldsymbol{\Sigma}_{FL,k}(\boldsymbol{h}^*_{\text{VaR}})}{\sigma_{L,k}(\boldsymbol{h}^*_{\text{VaR}})} z_k(\boldsymbol{h}^*_{\text{VaR}})\right] = \boldsymbol{0}_M,$$

---

[20] Litzenberger and Modest (2010) present an alternative reasoning for mixtures of normal distributions that relies on differentiating the implicit VaR_α definition in (24).

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

$$(29) \qquad \sum_{k=1}^{K} \frac{\pi_k (1 - F_{L,k}(v_\alpha(\boldsymbol{h}^*_{\mathrm{CVaR}})))}{\alpha} \left[\boldsymbol{\mu}_{F,k} + \frac{\Sigma_{FL,k}(\boldsymbol{h}^*_{\mathrm{CVaR}_\alpha})}{\sigma_{L,k}(\boldsymbol{h}^*_{\mathrm{CVaR}_\alpha})} \lambda_k(\boldsymbol{h}^*_{\mathrm{CVaR}_\alpha}) \right] = \mathbf{0}_M.$$

See the Appendix for a proof of Theorem 1. Note that the conditions in (28) and (29) could be multiplied by $f_L(v_\alpha(\boldsymbol{h}^*_{\mathrm{VaR}}))$ and α, respectively. We omitted this simplification to underline 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, e.g., holds that

$$(30) \qquad \mathrm{P}(S = k | L_H = v_\alpha(\boldsymbol{h}^*_{\mathrm{VaR}})) = \frac{\mathrm{P}(S = k) \, f_{L,k}(v_\alpha(\boldsymbol{h}^*_{\mathrm{VaR}}))}{\sum_{j=1}^{K} \mathrm{P}(S = j) \, f_{L,j}(v_\alpha(\boldsymbol{h}^*_{\mathrm{VaR}}))} = \frac{\pi_k \, f_{L,k}(v_\alpha(\boldsymbol{h}^*_{\mathrm{VaR}}))}{f_L(v_\alpha(\boldsymbol{h}^*_{\mathrm{VaR}}))}.$$

Using (13) and (14), 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.[21] In particular, these results imply that tail-risk-minimal strategies are identical to the minimum-variance approach if either $\mathbb{E}[\boldsymbol{R}_F] = \mathbf{0}_M$ 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 $\mathrm{VaR}_\alpha(h^*_{\mathrm{var}}) - \mathrm{VaR}_\alpha(h^*_{\mathrm{VaR}}) \leq b$ and $\mathrm{CVaR}_\alpha(h^*_{\mathrm{var}}) - \mathrm{CVaR}_\alpha(h^*_{\mathrm{CVaR}}) \leq b$ with[22]

$$(31) \qquad b = |\mathbb{E}[R_F]| \cdot \sqrt{\frac{\mathrm{var}\,[R_S]}{\mathrm{var}\,[R_F]}} \cdot (1 - \mathrm{corr}[R_F, R_S]^2).$$

---

[21]   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.

[22]   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.

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.[23]

These results are not surprising because 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, e.g., the degree of freedom parameter, and is therefore not influenced by the hedging weights. At this point, there is a very crucial difference between this simple, restricted model on the one hand and the full mixture approach on the other hand.

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 $\{1,\dots,K\}$ and transition matrix $\boldsymbol{Q} = (q_{ij})_{i,j=1,\dots,K}$, i.e. $\mathrm{P}(S_{t+1}=j|S_t=i) = q_{ij}$ for $i,j = 1,\dots,K$ and $t \in \mathbb{N}$. Under the additional assumptions that the Markov chain is aperiodic and irreducible, it has a unique invariant (ergodic) distribution $\boldsymbol{\pi}^e = (\pi_k^e)_{k=1,\dots,K}$. 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 com-

---

[23]   We eventually provide a numerical example in the online appendix, which shows that the differences in the hedging amount and in the corresponding tail risk values are small for typical parameter constellations.

ponents. 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, 1994). 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 (26) are all standard normally distributed. This comparatively simple setup already allows for very flexible univariate distribution shapes (Timmermann, 2000) and as shown by Ang and Chen (2002) it can reproduce asymmetric Longin and Solnik (2001) exceedance correlations.[24] 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

$$(32) \qquad \lambda_k(\boldsymbol{h}) = \mathbb{E}[Z \mid Z \geq z_k] = \frac{\varphi(z_k(\boldsymbol{h}))}{1 - \Phi(z_k(\boldsymbol{h}))},$$

where φ and Φ are the pdf and cdf, respectively, 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, we now provide additional results for tail risk hedging with mixtures of multivariate t-distributions.[25] We use a standardized version of the t-distribution, which is defined by the density generators

$$(33) \qquad g_{P,k}(s; \nu_k) = \frac{\Gamma(\frac{(P+\nu_k)}{2})}{((\nu_k - 2)\,\pi)^{\frac{P}{2}} \Gamma(\frac{\nu_k}{2})} \left(1 + \frac{s}{\nu_k - 2}\right)^{-\frac{P+\nu_k}{2}} \quad \text{for } \nu_k > 2.$$

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[26], so that we need $\nu_k > 2$ for

---

[24] We refer to Ang and Timmermann (2012) for a comprehensive review of its properties and a wide selection of applications.

[25] See, e.g., McLachlan and Peel (2000); Haas (2009) who use this model specification.

[26] See McNeil et al. (2005, Example 7.29).

the standardized version of the distribution to be well defined. Denoting the resulting pdf and cdf by f_t^* and F_t^* respectively, we obtain $F_{L,k}(v_\alpha(\boldsymbol{h})) = F_t^*(z_k(\boldsymbol{h}); \nu_k)$ and

$$(34) \qquad \lambda_k(\boldsymbol{h}) = \frac{f_t^*(z_k(\boldsymbol{h}); \nu_k)}{1 - F_t^*(z_k(\boldsymbol{h}); \nu_k)} \frac{\nu_k - 2 + (z_k(\boldsymbol{h}))^2}{\nu_k - 1}$$

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 $\nu_k, k = 1, \ldots, K$.[27]

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 (Rydén et al., 1998), the temporal dependence introduced by the Markov chain is often augmented with traditional time series filters (Alizadeh et al., 2008). Since 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 such models by replacing $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ with the conditional moments predicted by the time series filters for state k. Moreover, there are also a number of finance applications which work within the simpler setting of mixture distributions, where $(S_t)_{t \in \mathbb{N}}$ is an i. i. d. sequence (Kon, 1984; Buckley et al., 2008).

4 Empirical Results

We demonstrate our approach using three cross-hedging examples.[28] In particular, we compare futures-based hedging strategies that are used to temporarily minimize the tail risk of 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 and to focus on unconditional hedging.[29]

---

[27] Since Haas (2009) provides evidence for a limited advantage of the more flexible approach, we will consider the equal degrees of freedom setting in the empirical section.

[28] This setup can be motivated by the importance of basis risk found under the assumption of elliptical distributions. Moreover, a non-negligible amount of basis risk was shown to be important for the advantage of model-based hedging over naive strategies in general (Alexander and Barbosa, 2007).

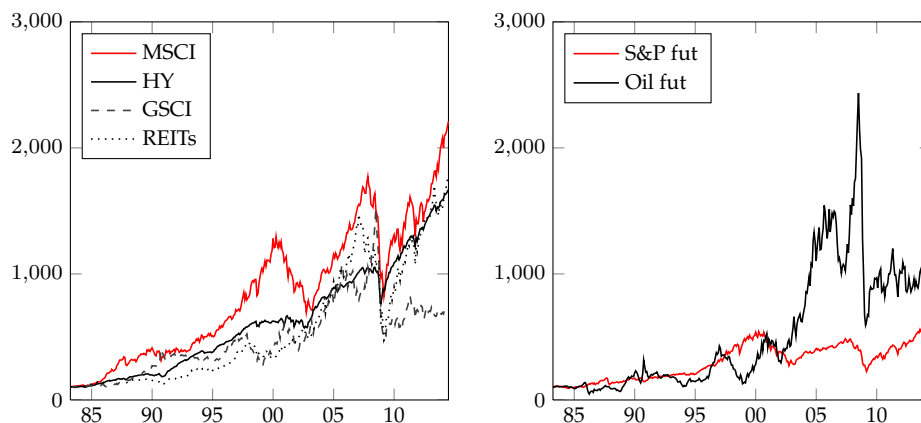[29] See Section 4.5 for a conditional version of our approach.

Figure 1: Spot and Futures Prices

4.1 Data

We consider portfolios representing the risky part of a broad asset allocation using the total return indices of the MSCI World, the Bank of America Merrill Lynch U.S. High Yield 100, the S&P GSCI and the FTSE/NAREIT U.S. All REITs. We form three equally-weighted multi-asset portfolios from these indices. Portfolio (P1) is invested into the MSCI and the HY index. For portfolios (P2) and (P3), we include the GSCI and the REIT index respectively. The S&P 500 Index futures traded on Chicago Mercantile Exchange and the NYMEX Light Crude Oil futures are considered as hedging instruments. The choice of these futures is motivated by liquidity, data availability and of course a relatively high correlation with the spot indices used.[30] Price data were obtained from Datastream.[31]

Our sample spans from March 1983 to June 2014, which corresponds to 376 monthly price observations. We plot the spot indices and futures series in Figure 1. Following common practice in the literature on RS models, we use continuously compounded returns.[32] Descriptive statistics of the return series are presented in Table 1.

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

---

[30]   We also considered using US Treasury 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.

[31]   We use a perpetual price index for the futures, which 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.

[32]   The usage of log-returns is a standard approximation for the exact approach based on discrete returns discussed in Section 2. In Section 4.5, we present an example for hedging with discrete returns, obtaining very similar results.

Table 1: Descriptive Statistics

	Spot Indices				Futures		Portfolios		
	MSCI	HY	GSCI	REITs	S&P Fut	Oil Fut	(P1)	(P2)	(P3)
mean [%]	0.83	0.75	0.53	0.77	0.50	0.62	0.80	0.75	0.81
	(0.23)	(0.11)	(0.29)	(0.26)	(0.23)	(0.53)	(0.15)	(0.16)	(0.17)
median [%]	1.33	0.96	0.73	1.14	0.88	1.45	1.17	0.92	1.06
std [%]	4.44	2.23	5.67	4.95	4.42	10.19	2.99	3.05	3.26
min [%]	-20.99	-15.42	-33.13	-35.99	-22.83	-42.29	-18.17	-22.91	-23.76
max [%]	11.13	7.15	20.65	24.67	12.41	40.68	8.43	9.81	14.14
skewness	-0.91	-1.41	-0.62	-1.71	-0.97	-0.36	-1.35	-1.76	-1.72
kurtosis	5.49	11.71	6.67	15.00	6.01	5.54	8.38	13.39	13.31
JB	148.76	1310.77	235.17	2433.56	200.15	109.33	565.53	1880.77	1848.11
p_{JB} [%]	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
corr$(\cdot, F_1)$	0.88	0.58	0.17	0.57	1.00	0.06	0.87	0.67	0.81
corr$(\cdot, F_2)$	0.09	0.04	0.82	0.05	0.06	1.00	0.08	0.56	0.07
ex-corr$(\cdot, F_1; q_{0.2})$	0.86	0.57	0.66	0.62	1.00	0.58	0.83	0.63	0.77
ex-corr$(\cdot, F_1; q_{0.8})$	0.70	0.16	-0.06	0.26	1.00	0.08	0.50	0.41	0.47

Note: Descriptive statistics of spot and futures instruments. Monthly log-returns from April 1983 to June 2014, $T = 375$ return observations. JB refers to the Jarque-Bera test statistic for normality and p_{JB} denotes the corresponding p-value. ex-corr$(\cdot, F_1; q_\alpha)$ 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. MSCI: MSCI World Total Return Index, HY: BofA Merrill Lynch US High Yield 100 Total Return Index, GSCI: S&P GSCI Commodity Total Return, REIT: FTSE/NAREIT All REITs Total Return Index, S&P Fut: Chicago Mercantile Exchange S&P 500 Index futures, Oil Fut: NYMEX Light Crude Oil futures. Equally-weighted multi-asset spot portfolios: (P1): MSCI/HY, (P2): MSCI/HY/GSCI, (P3): MSCI/HY/REITs.

tests for all series. Comparing the spot portfolios, the returns of (P2) and (P3) exhibit stronger asymmetries and fatter tails than those of (P1). The kurtosis of the former is twice as high as that of the futures returns. According to empirical exceedance correlations, we find evidence for asymmetric dependencies as shown for the bivariate distributions of spot and S&P futures returns.

4.2 Parameter Estimates and Model Fit

For our baseline analysis, we hedge long positions in (P1) - (P3) with the futures on the S&P 500 Index. We first fit RS models with two and three normal components to the bivariate distributions of portfolio and futures returns.[33] The parameters that attained the highest likelihood in repeated maximum-likelihood estimations from randomly chosen initial values are displayed in Table 2.[34] In order 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 three

---

[33]   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, in order to keep the dimension of the model as low as possible.

[34]   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 documented in the online appendix.

Table 2: In-Sample Parameter Estimates

	(P1)				(P2)				(P3)			
	$K=2$		$K=3$		$K=2$		$K=3$		$K=2$		$K=3$	
	par	s.e.	par	s.e.	par	s.e.	par	s.e.	par	s.e.	par	s.e.
State 1												
$\mu_{S,1}$	-0.57	(0.85)	-3.79	(4.34)	-0.74	(1.64)	-10.34	(2.37)	-2.61	(1.57)	-4.00	(2.46)
$\mu_{F,1}$	-1.61	(1.18)	-6.33	(6.95)	-2.27	(2.29)	-8.88	(1.99)	-4.19	(1.75)	-6.13	(2.50)
$\sigma_{S,1}$	4.9	(0.58)	4.85	(1.02)	5.14	(1.37)	5.59	(1.87)	6.61	(1.63)	7.86	(2.02)
$\sigma_{F,1}$	6.96	(0.67)	6.24	(2.89)	6.90	(1.07)	4.56	(1.48)	7.44	(1.19)	8.12	(1.60)
$\rho_{SF,1}$	87.31	(2.64)	82.44	(16.62)	73.47	(4.73)	99.35	(0.55)	83.12	(4.24)	82.27	(5.74)
State 2												
$\mu_{S,2}$	1.18	(0.14)	1.46	(0.32)	1.12	(0.13)	0.78	(0.40)	1.24	(0.17)	1.05	(0.36)
$\mu_{F,2}$	1.08	(0.20)	1.37	(0.44)	1.19	(0.22)	-0.20	(0.90)	1.10	(0.26)	0.69	(0.45)
$\sigma_{S,2}$	2.03	(0.12)	2.49	(0.36)	2.09	(0.19)	3.33	(0.38)	2.18	(0.16)	3.13	(0.31)
$\sigma_{F,2}$	3.17	(0.17)	4.02	(0.39)	3.20	(0.33)	5.92	(0.75)	3.43	(0.25)	3.57	(0.38)
$\rho_{SF,2}$	83.50	(2.18)	81.04	(3.25)	52.38	(5.45)	67.69	(5.96)	76.33	(2.60)	92.11	(2.10)
State 3												
$\mu_{S,3}$			0.97	(0.20)			1.07	(0.16)			1.14	(0.17)
$\mu_{F,3}$			0.92	(0.29)			1.13	(0.23)			1.01	(0.29)
$\sigma_{S,3}$			1.80	(0.15)			1.98	(0.13)			2.08	(0.21)
$\sigma_{F,3}$			2.54	(0.26)			2.92	(0.20)			3.73	(0.30)
$\rho_{SF,3}$			88.36	(2.23)			52.20	(5.99)			73.99	(4.74)
Transition matrix												
q_{11}	83.0	(7.6)	61.2	(30.4)	80.8	(13.6)	29.2	(18.0)	63.0	(11.7)	61.1	(14.2)
q_{21}	4.5	(1.7)	4.6	(4.0)	4.6	(1.7)	2.6	(1.9)	4.7	(2.3)	1.5	(2.1)
q_{31}			2.8	(2.6)			0.8	(0.9)			2.8	(1.6)
q_{12}			38.7	(48.4)			54.3	(25.7)			10.2	(25.1)
q_{22}			92.8	(3.0)			88.7	(7.3)			97.2	(8.5)
q_{32}			0.8	(1.2)			3.8	(2.1)			0.0	(15.6)
Stationary distribution												
π_1	21.1		9.1		19.5		1.9		11.2		5.9	
π_2	78.9		53.2		80.5		31.6		88.8		22.8	
π_3			37.7				66.5				71.2	

Note: Parameter estimates for bivariate two-state and three-state RS models with normal components. 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 in order 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 (P3) and $K=3$ a boundary solution was found due to the low value of q_{32}.

two-state models is very similar: For all bivariate distributions, there is a joint bearish state with a low probability of occurrence, negative means[35], high standard deviations and high correlations. Allowing for a third component, the first state becomes a severe crash scenario in all cases. In particular, (P2) shows a very high correlation in this state, which is almost twice the correlation in state three.

In Panel A of Table 3, we provide some evidence on the fit of these models and simple elliptical distributions for the bivariate return samples.[36] According to information criteria, at least one of the RS models is always favored over nonswitching specifications. While AIC prefers

---

[35] The mean estimates exhibit substantial standard errors because of the low unconditional state probabilities.

[36] The degrees of freedom parameters estimated for the three multivariate distributions correspond to 4.5, 5.3 and 4.1 for (P1) - (P3), respectively. The other model parameters can be found in the online appendix.

Table 3: Model Fit and 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)										
emp	-	-	-	-	69.0	90.2	33.5	69.0	90.2	22.0
pot	-	-	-	-	89.4	8.6	53.5	89.4	94.9	49.1
mv-n	1681.9	-3353.7	-3334.1	0.1	0.2	0.0	1.1	9.1	23.8	8.4
mv-tstd	1726.1	-3440.2	-3416.7	5.2	5.5	0.0	29.0	-	-	0.0
RS $K = 2$ stat	1749.2	-3474.4	-3427.3	9.7	89.4	8.6	16.4	9.1	23.8	92.5
RS $K = 2$ pred	-	-	-	-	28.0	11.5	44.0	69.0	90.2	99.1
RS $K = 3$ stat	1769.4	-3496.8	-3414.3	50.0	69.0	90.2	27.5	69.0	90.2	66.4
RS $K = 3$ pred	-	-	-	-	69.0	90.2	34.9	32.1	60.5	64.8
(P2)										
emp	-	-	-	-	69.0	0.0	39.0	69.0	90.2	22.2
pot	-	-	-	-	89.4	0.1	98.1	89.4	94.9	49.8
mv-n	1526.1	-3042.1	-3022.5	0.1	2.1	0.3	0.5	9.1	23.8	7.9
mv-tstd	1570.2	-3128.4	-3104.9	4.6	2.1	0.3	6.7	9.1	23.8	92.2
RS $K = 2$ stat	1589.9	-3155.8	-3108.7	7.2	89.4	0.1	19.2	69.0	90.2	96.9
RS $K = 2$ pred	-	-	-	-	53.3	77.0	51.9	69.0	90.2	83.9
RS $K = 3$ stat	1609.9	-3177.9	-3095.4	50.0	32.1	1.0	36.0	9.1	23.8	92.2
RS $K = 3$ pred	-	-	-	-	53.3	77.0	99.7	32.1	60.5	97.5
(P3)										
emp	-	-	-	-	69.0	4.1	29.9	69.0	90.2	21.7
pot	-	-	-	-	53.3	11.6	78.6	89.4	94.9	49.6
mv-n	1593.5	-3177.0	-3157.4	0.1	0.7	0.2	1.0	9.1	23.8	7.9
mv-tstd	1660.3	-3308.6	-3285.1	3.6	0.7	0.2	31.6	-	-	0.0
RS $K = 2$ stat	1679.5	-3335.0	-3287.9	50.0	69.0	4.1	23.3	69.0	90.2	65.5
RS $K = 2$ pred	-	-	-	-	69.0	90.2	69.7	69.0	90.2	53.8
RS $K = 3$ stat	1701.3	-3360.5	-3278.1	50.0	69.0	4.1	39.2	69.0	90.2	55.5
RS $K = 3$ pred	-	-	-	-	89.4	94.9	82.3	69.0	90.2	52.1

Note: Panel A refers to the statistical fit of the multivariate models. 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. 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). emp and pot are empirical and Peaks-over-Threshold risk estimates for the corresponding loss series. mv-n and mv-tstd refer to multivariate normal and multivariate standardized t-distribution models. RS denotes regime-switching models with $K = 2$ and $K = 3$ states. stat refers to backtest results for the unconditional risk estimates and pred contains the corresponding results for conditional risk forecasts.

three-state models, BIC favors two-state models. We also perform JB-Tests on the distribution fit after transforming the sample data to normality with the Berkowitz (2001) approach. Whereas the predictive distributions of all RS models pass this test at the 5% significance level, the simpler multivariate normal and t-models are mostly rejected, which hints at a misspecification of the tails for these models.

Before we compare the hedging performance derived from these models, we assess their risk *measurement* quality. We focus on the 99% confidence level, which we will also use for the hedging analysis. In particular, we analyze risk forecasts for an unhedged long position in (P1) - (P3) and a short position in the S&P futures derived from each of the bi-

variate 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 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 (24) and (25) with (32). We use empirical and extreme-value-theory-based risk estimates as benchmarks for our analysis. Both are calculated from a univariate loss sample $(l_t)_{t=1,\ldots,T}$. As nonparametric VaR_α and CVaR_α estimators, the sample counterparts of (3) and (4) are used, i. e. $\widehat{\text{VaR}}_\alpha^e = l_{(\lceil T(1-\alpha)\rceil)}$ and[37]

$$(35) \qquad \widehat{\text{CVaR}}_\alpha^e = \frac{1}{\alpha}\left[\frac{1}{T}\sum_{i=\lceil(1-\alpha)T\rceil+1}^{T} l_{(i)} + \left(\frac{\lceil T(1-\alpha)\rceil}{T} - (1-\alpha)\right)l_{(\lceil T(1-\alpha)\rceil)}\right],$$

where $l_{(i)}$ is the ith rank statistic of the loss sample. For the calculation of Peaks-over-Threshold (POT) risk estimates, we consider the subsample of losses exceeding a threshold[38] u and fit a generalized Pareto distribution to the loss exceedances $l_t - u$. From the estimated shape and scale parameters $\hat{\xi}$ and $\hat{\beta}$ and the number of exceedances n_u we obtain the following risk estimators[39]

$$(36) \qquad \widehat{\text{VaR}}_\alpha^{pot} = u + \frac{\hat{\beta}}{\hat{\xi}}\left[\left(\alpha\frac{T}{n_u}\right)^{-\hat{\xi}} - 1\right],$$

$$(37) \qquad \widehat{\text{CVaR}}_\alpha^{pot} = \widehat{\text{VaR}}_\alpha^{pot} + \frac{\hat{\beta} + \hat{\xi}(\widehat{\text{VaR}}_\alpha^{pot} - u)}{1 - \hat{\xi}}.$$

We use 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 the tail risk estimates obtained from these models. Corresponding test results can be found in Panel B of Table 3. The VaR_α estimates derived from the RS models and the benchmark techniques are never rejected according to unconditional coverage tests at conventional significance levels, whereas the risk forecasts derived from the elliptical specifications are all rejected at the 10% significance level. According to the p-values of conditional coverage tests, we observe a uniform improvement by using the predictive risk series. However, at the 1% significance level, there is only a single rejection of the correct conditional coverage

---

[37] See Rockafellar and Uryasev (2002, P.8) for this estimator.
[38] We use the 0.9-quantile as threshold for our estimations.
[39] See e. g. McNeil and Frey (2000) for these estimators.

hypothesis for $\widehat{\text{VaR}}_\alpha^{RS,u}$ (long position in (P2)). 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. These tests reject the normal and t-models only in some cases.

4.3 In-Sample Hedging Results

Turning to the core of our analysis, we now investigate unconditional hedging strategies derived from the stationary distribution of the fitted RS models using Theorem 1. These are compared to minimum-variance hedges[40] and CVaR_α-minimal hedging strategies obtained from the nonparametric approach by Rockafellar and Uryasev (2000). The latter serve as a benchmark for the maximum reduction of the empirical CVaR_α estimates in a static in-sample analysis. In addition to hedging weights, we of course provide CVaR_α values for the hedged positions, which we estimated with our models and also according to the non- and semiparametric estimators from (35) and (37). We measure the reduction of tail risk attained by switching from a simple minimum-variance strategy $\boldsymbol{h}^*_{\text{var}}$ to the CVaR_α-minimal policy $\boldsymbol{h}^*_{\text{CVaR}}$ by

$$(38) \qquad \Delta\% = 1 - \frac{\text{CVaR}_\alpha(\boldsymbol{h}^*_{\text{CVaR}})}{\text{CVaR}_\alpha(\boldsymbol{h}^*_{\text{var}})}.$$

We focus on the results for three-state models, which we provide in Table 4.[41]

First, note that the hedging weights of the CVaR_α-minimal strategies are always higher than the corresponding minimum-variance weights. We find a 10% increase in the amount of hedging for (P1)[42], whereas the hedging positions for (P2) and (P3) are about 20% greater than those of minimum-variance strategies. Moreover, the RS CVaR_α hedging strategies are always close to the in-sample optimum as measured by the empirical approach.

---

[40]　The differences between using a model-free OLS estimate and the model-based minimum-variance hedge according to (19) is negligible. Thus, (19) is only relevant for conditional hedging strategies, which we consider in Section 4.5.

[41]　Corresponding results for the two-state models are provided in the online appendix. Although differences to the minimum-variance strategy are less pronounced, all effects remain similar.

[42]　Nevertheless, even the effect in this case, esp. in terms of risk reduction, is still in line with improvements typically reported in the futures hedging literature.

Table 4: Hedging Results: In-Sample

	(P1)				(P2)				(P3)			
	uh	var	RS	emp	uh	var	RS	emp	uh	var	RS	emp
Hedging weights [%]												
h_1		58.43	70.20	72.46		45.96	73.12	66.82		60.05	80.46	91.54
Risk measures [%]												
$\widehat{\text{CVaR}}_\alpha^{RS,u}$	12.08	4.88	4.48	4.50	14.65	9.01	7.01	7.15	15.74	7.41	6.31	6.59
$\Delta\%$			8.07	7.79			22.24	20.65			14.81	11.11
$\widehat{\text{CVaR}}_\alpha^{e}$	12.38	4.59	4.16	4.11	13.51	8.26	7.07	7.02	15.13	7.55	6.39	6.06
$\Delta\%$			9.48	10.40			14.35	14.96			15.37	19.81
$\widehat{\text{CVaR}}_\alpha^{pot}$	12.58	4.84	4.17	4.11	20.20	8.38	6.95	7.10	16.02	7.28	6.41	6.10
$\Delta\%$			13.92	15.06			17.13	15.33			11.98	16.21
Moments (model/empirical)												
mean R_H [%]	0.80	0.51	0.45	0.44	0.76	0.52	0.38	0.41	0.81	0.50	0.40	0.34
	0.80	0.51	0.45	0.44	0.75	0.52	0.38	0.41	0.81	0.51	0.40	0.35
std R_H [%]	2.97	1.49	1.57	1.61	3.02	2.26	2.55	2.43	3.23	1.89	2.11	2.37
	2.99	1.50	1.59	1.62	3.05	2.27	2.57	2.45	3.26	1.89	2.10	2.35
skewness R_H	-1.18	-0.24	0.14	0.19	-1.58	-0.75	0.09	-0.03	-1.63	-0.44	0.14	0.25
	-1.35	-0.23	0.29	0.36	-1.76	-0.88	0.16	-0.03	-1.72	-0.79	-0.01	0.21
kurtosis R_H	6.99	5.39	4.80	4.75	11.28	6.60	3.74	3.91	12.18	8.65	5.55	4.94
	8.38	5.55	5.09	5.14	13.39	7.51	3.80	3.99	13.31	10.29	5.08	4.53
Tail characteristics												
$\hat{q}_{0.9}$	2.45	1.37	1.52	1.59	2.61	2.12	2.72	2.53	2.65	1.78	2.12	2.41
$\hat{\xi}$	0.05	0.19	-0.06	-0.10	0.57	0.16	-0.04	-0.03	0.26	0.33	-0.07	-0.52
$\hat{\beta}$	2.78	0.73	0.90	0.92	1.34	1.39	1.38	1.47	2.40	0.84	1.48	2.41

Note: Hedging weights and risk of minimum-variance and CVaR_α hedging strategies; in-sample results for $\alpha = 0.01$. uh: unhedged spot portfolios, var: minimum-variance hedging strategy, RS: stationary CVaR_α hedging strategy, emp: CVaR_α hedging strategy based on Rockafellar and Uryasev (2002). h_1 is the hedging weight of the S&P futures. $\widehat{\text{CVaR}}_\alpha^{RS,u}$ refers to parametric risk estimates based on the stationary distribution of the fitted RS models. $\widehat{\text{CVaR}}_\alpha^{e}$ refers to empirical risk estimates and $\widehat{\text{CVaR}}_\alpha^{pot}$ are POT-based risk estimates. $\Delta\%$ is the relative tail risk reduction as compared to the minimum-variance hedging strategy.

Second, we find that minimum-variance cross-hedges already successfully remove a large fraction of tail risk, in particular for (P1) and (P3).[43] However, we see that the increase in the hedging amount implied by the CVaR_α policies further reduces the tail risk of the net positions. The risk reductions obtained by switching from a variance-based to a CVaR_α-minimal policy range between 8% and 22%. The evidence for advantages of CVaR_α hedging in our examples is conclusive because all measurement methods confirm a tail risk reduction as compared to the minimum-variance strategy.

Third, we analyze the moments of the return distributions of the net positions to gain insights into the sources of the risk reduction. We find that in all cases CVaR_α hedging attains a tail risk reduction by increasing skewness and lowering kurtosis of the hedged returns.[44] The

---

[43] For (P2), we attain similar reductions by adding the oil futures in Section 4.5.
[44] Results show that the empirical and the model-implied moments match at least approximately.
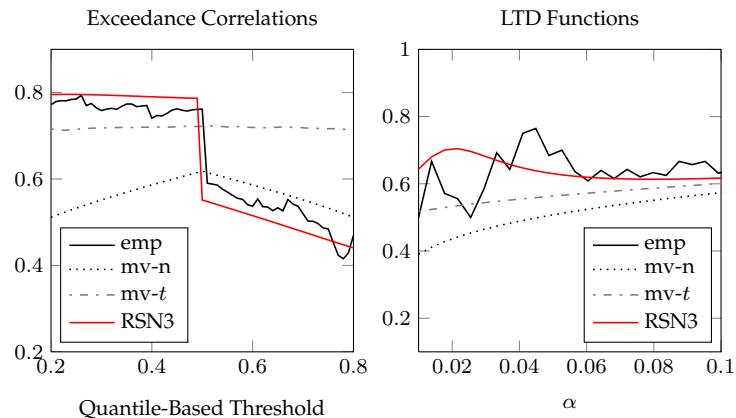
Figure 2: Exceedance Correlations and Lower Tail Dependence Functions of (P3) and the S&P Futures.

reduction is higher for (P2) and (P3), for which the returns of the minimum-variance strategy still exhibit a sizable amount of skewness and excess kurtosis.

Fourth, we analyze the upper tails of the loss distributions as described by the GPD. CVaR_α hedges lower the shape parameter, and thus reduce the heaviness of the relevant tail in all examples. This reduction comes at cost of increasing the 90%-quantile or the scale parameter of the GDP, but it overcompensates for these effects according to the POT-CVaR_α-estimates.

Concluding the presentation of our in-sample results, we provide a complementary view on the differences between minimum-variance and CVaR_α-based hedging in our examples. For (P3) and the S&P futures, we plotted empirical exceedance correlations (Longin and Solnik, 2001) and the corresponding model implied values in Figure 2.[45] For all three portfolios we observe correlations to be higher in joint crash states than in joint good states, which explains the reduction in CVaR_α by *increasing* the hedging weight. RS models (as depicted for $K = 3$) can capture this dependence structure closely matching the empirical correlation estimates. Similar evidence for an increased dependence between spot and futures returns in bear markets is obtained by comparing the empirical lower tail dependence functions (Garcia and Tsafack, 2011) with the corresponding values implied by a normal distribution, which are also provided in Figure 2. Interestingly, we find that over the plotted range, the values derived from the stationary distribution of the RS model are even higher than those corresponding to a t-model (copula) and they are close to the empirical values, which, however, fluctuate quite strongly due to the small sample size.

---

[45]   We use a quantile-based threshold for the implementation of the correlations (Patton, 2004, p. 138).

Table 5: Hedging Results: Out-of-Sample

	(P1)				(P2)				(P3)			
	uh	var	RS	emp	uh	var	RS	emp	uh	var	RS	emp
Hedging weights [%]												
min h		47.30	50.78	58.29		27.70	25.76	16.19		44.36	50.56	52.84
mean h		53.33	60.12	65.53		36.2	43.35	52.08		50.62	65.72	69.98
max h		58.54	65.70	75.99		46.12	72.09	87.30		60.30	84.28	91.54
std h		3.68	3.83	7.51		6.07	13.41	19.96		6.22	12.11	15.60
Risk measures [%]												
$\widetilde{\mathrm{CVaR}}_\alpha^e$	14.18	6.12	5.62	5.1	16.81	11.92	11.57	9.85	18.03	11.27	9.56	9.78
$\Delta\%$			8.19	16.68			2.91	17.34			15.10	13.18
$\widetilde{\mathrm{CVaR}}_\alpha^{pot}$	13.95	6.47	5.58	4.93	18.27	11.67	10.85	10.00	17.50	10.31	9.64	9.80
$\Delta\%$			13.78	23.72			7.05	14.28			6.48	4.94
Moments (empirical)												
mean R_H [%]	0.54	0.36	0.34	0.30	0.46	0.32	0.26	0.22	0.62	0.43	0.34	0.32
std R_H [%]	3.31	1.42	1.36	1.36	3.67	2.67	2.56	2.50	3.84	2.2	2.03	2.11
skewness R_H	-1.42	-1.13	-0.81	-0.32	-1.71	-1.42	-1.34	-0.73	-1.61	-1.60	-1.42	-1.51
kurtosis R_H	8.08	8.94	8.19	7.55	11.05	10.05	9.93	6.05	11.57	14.61	12.25	12.58

Note: Hedging weights and risk of minimum-variance and CVaR_α hedging strategies based on RS models with two normal components. For all strategies, we calculate hedging weights with a growing estimation window, using 175 observations for the first estimation and updating the hedging weights monthly. Risk estimates and sample moments are based on the resulting 200 hedged return observations.

4.4 Out-of-Sample Hedging Results

In this section, we complement the in-sample performance evaluation with the results of two out-of-sample experiments.

We begin with out-of-sample backtests in the presented datasets. We reserve the first 175 observations for the first estimation and work with a growing estimation window. In total, 200 two-state RS models per portfolio are estimated and used to derive the equal number of RS CVaR_α hedging weights.[46] We use the same estimation windows to determine hedging weights for the minimum-variance and the empirical minimum-CVaR_α hedges. See Figure 3 for a plot of the resulting strategies. Note that there is a significant increase in the hedging weights for all three portfolios as soon as data from the subprime crisis enter the estimation window. However, it is important to remark that the hedging amount implied by CVaR_α strategies was already higher than with minimum-variance hedging *before* the financial crisis occurred – at least for (P1) and (P3).

The results of these backtests are summarized in Table 5. Although we still use an unconditional hedging approach, the hedging policies are now time-varying due to re-estimation.

---

[46] We estimated two-state models because in comparison with three-state models their calibration is more stable with the limited amount of data available for the first estimations.
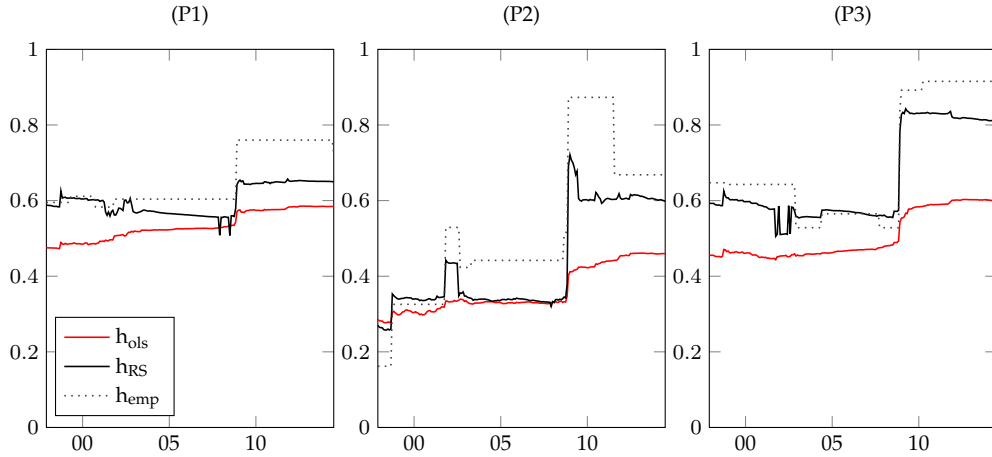
Figure 3: Out-of-Sample Hedging Weights

Looking at their descriptive statistics, we find that the average hedging amount of an RS CVaR_α-minimal strategy is greater than for the minimum-variance approach with the difference ranging from 6% to 15%. Moreover, the standard deviation is higher for CVaR_α hedging – with much of the variation in hedging weights being caused by the financial crisis. To evaluate the risk reductions attained by the strategies, we again use $\widehat{\text{CVaR}}_\alpha^e$ and $\widehat{\text{CVaR}}_\alpha^{pot}$.[47] According to both measures, the tail risk reduction from switching to CVaR_α hedging is always positive in our examples. The reductions over all portfolios and the two estimation methods range between 3% and 15%. Again, these reductions come with an increase in return skewness and a decrease in kurtosis as compared to the minimum-variance approach. Finally, these results emphasize that differences between tail risk and minimum-variance hedging can already be attained using the most basic two-state RS models.

We next provide the results of simulation experiments to confirm the out-of-sample performance of our hedging policies with larger sample sizes. We focus on (P3) and adopt the hedging strategy derived from the corresponding three-state RS model. 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.[48] This choice combines an elliptical dependence structure allowing for (symmetric) tail dependence and nonelliptical marginal

---

[47]   Note that the reliability of the empirical CVaR_α is seriously affected by the small sample size.
[48]   The parameter estimates are also provided in the online appendix.

Table 6: Out-of-Sample Simulation Results

	RS $K = 3$				bootstrap				t-Copula + skewed-t margins			
	uh	var	RS	emp	uh	var	RS	emp	uh	var	RS	emp
Hedging weights [%]												
h_1		60.05	80.45	91.54		60.05	80.45	91.54		60.05	80.45	91.54
Risk measures [%]												
mean $\widehat{\mathrm{CVaR}}_\alpha^e$	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
Moments (model/empirical)												
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

Note: Out-of-sample simulations for portfolio (P3). h_1 denotes the hedging weight in the S&P futures. $q_\alpha[\Delta\%]$ refers to the α-quantile of the risk reductions obtained from the simulations.

distributions. We simulate $10,000$ return samples each of length $T = 1,000$ observations. 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 6. Simulating from the estimated model, the average CVaR_α reduction confirms our analytic results from Table 4. Looking at the quantiles of the reduction series obtained from our 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 simulations under sampling from the model and the empirical distribution. This implies a (weak) statistical significance of this reduction at the 10% level. The same quantiles are negative for the nonparametric reference strategy, even under sampling from the empirical distribution, which reveals a strong reliance of this technique on the specific characteristics of the given sample. Remarkably, RS CVaR_α hedging also attained 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, indicating a positive contribution of dependence asymmetries to the reported effects.

Table 7: In-Sample Results, Composite Hedging

	(P2)			
	uh	var	RS	emp
Hedging weights [%]				
h_1		43.82	58.69	62.27
h_2		15.56	14.85	17.18
Risk measures [%]				
$\widehat{\mathrm{CVaR}}_\alpha^{RS,u}$	14.02	5.97	5.35	5.50
$\Delta\%$			10.31	7.77
$\widehat{\mathrm{CVaR}}_\alpha^{e}$	13.51	5.73	4.97	4.55
$\Delta\%$			13.30	20.59
$\widehat{\mathrm{CVaR}}_\alpha^{pot}$	20.20	5.50	4.83	4.79
$\Delta\%$			12.09	12.83

Note: In-sample results for the composite hedging of portfolio (P2) using two futures. h_1 denotes the hedging weight in the S&P futures. h_2 is the hedging weight in the oil futures.

4.5 Model Extensions and Robustness Checks

In this section, we investigate whether the documented advantage of tail-risk-minimal hedging can be confirmed for more complex setups. In particular, we first analyze an example for *multivariate* CVaR_α hedging and then report the performance of *conditional* CVaR_α hedging strategies derived from the predictive distribution of the RS models. Eventually, we provide a battery of robustness checks on our modeling assumptions and datasets.

To assess the performance within a multivariate setting, we again consider (P2), which contains the GSCI, and use the oil futures as a second hedging instrument in addition to the S&P futures. Due to the promising results of three-state models in Section 4.3, we also fit a three-state model for the joint return distribution of the spot portfolio and the two futures.[49] The corresponding hedging weights and resulting risk estimates can be found in Table 7. Although the hedging amount in the oil futures does not differ much between the three 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 case, this improvement is attained by increasing the hedging weight of the S&P futures.

Next, we analyze the effect of using the predictive distribution of the RS models for CVaR_α-based hedging and now use a dynamic minimum-variance strategy according to (19) as a benchmark. We report the corresponding in-sample results for our three univariate exam-

---

[49] The estimation results are provided in the online appendix.

Table 8: Dynamic Hedging

	(P1)				(P2)				(P3)			
	uh	RS-var	RS-CVaR	emp	uh	RS-var	RS-CVaR	emp	uh	RS-var	RS-CVaR	emp
Hedging weights [%]												
min h_1		54.22	63.98	72.46		43.12	69.94	66.82		48.30	69.90	91.54
mean h_1		58.17	68.76	72.46		44.48	71.01	66.82		58.77	75.84	91.54
max h_1		63.40	80.29	72.46		82.81	82.81	66.82		79.54	97.24	91.54
std h_1		3.36	4.56	0.00		5.25	1.84	0.00		12.02	7.14	0.00
Risk measures [%]												
$\widehat{\mathrm{CVaR}}_\alpha^e$	12.38	4.37	3.99	4.11	13.51	6.25	6.43	7.02	15.13	6.60	5.64	6.06
$\Delta\%$			8.80	5.93			-2.91	-12.34			14.58	8.24
$\widehat{\mathrm{CVaR}}_\alpha^{pot}$	12.58	4.63	3.93	4.11	20.20	6.35	6.26	7.10	16.02	6.36	5.76	6.10
$\Delta\%$			15.07	11.14			1.38	-11.75			9.53	4.17

Note: In-sample results for conditional hedging based on the predictive distribution of RS models with $K = 3$ normal components. h_1 denotes the hedging weight in the S&P futures.

ples in Table 8. Here the evidence is mixed: Although we again find CVaR_α reductions from 8% to 15% for (P1) and (P3), the effect for (P2) is quite weak and changes its sign with the evaluation method. Note, however, that for all portfolios, our strategy outperforms unconditional CVaR_α-optimal hedging based on the empirical distribution.

Table 9 eventually provides the results of our robustness checks.[50] First, we test whether improvements can be obtained by fitting RS models with t-distributed components. We report our results in Panel A. For (P1) to (P2) the findings are similar to the specification with normal components. The hedging weight for (P3), however, is close to that of the minimum-variance hedge. 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. In Panel B we document different implementations of our CVaR_α-minimal approach, focusing on (P3). Here we find that the results remain almost unchanged if discrete returns are used or if the MCVaR_α is optimized instead of the CVaR_α. However, we find that differences between minimum-CVaR_α and minimum-variance hedging decrease in the confidence level. In Panel C, we validate our results based on another dataset. The results obtained using different indices for the assets in the spot portfolio are similar to those of the original specification.[51] We also show that our findings cannot be attributed to the particular choice of the rollover strategy by reproducing the hedge with the S&P spot series. We even-

---

[50]   All estimation results can again be found in the online appendix.
[51]   See the notes below Table 9 for the description of the data.

Table 9: Robustness Checks

Panel A	(P1) RS-t				(P2) RS-t				(P3) RS-t			
	uh	var	RS	emp	uh	var	RS	emp	uh	var	RS	emp
Hedging weights [%]												
h_1		58.43	70.04	72.46		45.96	70.1	66.82		60.05	58.56	91.54
Risk measures [%]												
$\widehat{\text{CVaR}}_\alpha^{RS,u}$	12.05	4.88	4.49	4.51	14.23	8.8	7.23	7.27	10.37	5.55	5.55	7.97
$\Delta\%$			7.83	7.50			17.87	17.44			0.11	-43.62
$\widehat{\text{CVaR}}_\alpha^{e}$	12.38	4.59	4.16	4.11	13.51	8.26	7.05	7.02	15.13	7.55	7.70	6.06
$\Delta\%$			9.42	10.40			14.64	14.96			-1.95	19.81
$\widehat{\text{CVaR}}_\alpha^{pot}$	12.58	4.84	4.18	4.11	20.20	8.38	6.99	7.10	16.02	7.28	7.27	6.10
$\Delta\%$			13.66	15.06			16.58	15.33			0.13	16.21

Panel B	(P3) $\alpha = 0.025$				(P3) MCVaR$_\alpha$				(P3) Discrete returns			
	uh	var	RS	emp	uh	var	RS	emp	uh	var	RS	emp
Hedging weights [%]												
h_1		60.05	72.46	67.13		60.05	81.49	91.54		59.58	78.81	86.88
Risk measures [%]												
$(\text{M})\widehat{\text{CVaR}}_\alpha^{RS,u}$	11.43	5.22	4.92	4.97	16.55	7.92	6.71	6.93	15.00	7.15	6.18	6.33
$\Delta\%$			5.70	4.68			15.23	12.45			13.67	11.52
$(\text{M})\widehat{\text{CVaR}}_\alpha^{e}$	10.90	5.19	5.02	4.99	15.94	8.06	6.73	6.40	13.93	6.98	6.04	5.93
$\Delta\%$			3.34	3.79			16.41	20.53			13.44	15.05
$(\text{M})\widehat{\text{CVaR}}_\alpha^{pot}$	11.23	5.19	5.14	5.17	16.83	7.79	6.78	6.45	14.60	6.68	6.08	5.97
$\Delta\%$			0.91	0.39			12.97	17.19			8.96	10.69

Panel C	(P3) 1st sample half				(P3) Different spot indices				(P3) S&P spot			
	uh	var	RS	emp	uh	var	RS	emp	uh	var	RS	emp
Hedging weights [%]												
h_1		46.66	60.75	64.72		70.67	90.79	105.85		60.34	79.93	89.58
Risk measures [%]												
$\widehat{\text{CVaR}}_\alpha^{RS,u}$	11.61	4.88	4.27	4.31	17.36	8.19	7.05	7.63	15.75	7.43	6.40	6.62
$\Delta\%$			12.43	11.59			13.99	6.89			13.92	10.98
$\widehat{\text{CVaR}}_\alpha^{e}$	11.51	3.92	3.53	3.43	17.33	8.72	7.39	6.76	15.13	7.57	6.49	6.15
$\Delta\%$			10.12	12.54			15.17	22.45			14.25	18.72
$\widehat{\text{CVaR}}_\alpha^{pot}$	13.39	3.85	3.39	3.42	17.32	8.41	7.15	7.13	16.02	7.13	6.50	6.30
$\Delta\%$			11.94	11.16			14.97	15.23			8.91	11.71

Note: Panel A contains hedging results for three-state RS models with standardized t-distributed components and equal degrees of freedom across the components. In Panel B we provide robustness results for using a different confidence level, the demeaned CVaR$_\alpha$ and discrete returns. Panel C shows results for (P3) with different time series. We replace our spot indices 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 and substitute the perpetual S&P futures returns with the spot returns. Finally, we report results for the first half of our original sample, for which we estimated RS models with two states.

tually confirm that similar reductions can be attained without data from the subprime crisis, using the first half of the sample.[52]

---

[52] As in the out-of-sample setup, we fitted a two-state model here due to the small sample size.

5 Conclusion

In this paper, we studied the use of finite mixtures and in particular regime-switching models for tail risk *management*. We provided a general characterization of VaR_α- and CVaR_α-minimal futures hedging strategies relying on results on quantile derivatives and showed how to implement these characterizations for mixtures of elliptical distributions. Using multivariate regime-switching models, we empirically demonstrate that CVaR_α minimizations can change hedging strategies and tail risk characteristics as compared to variance minimizations if the investments under consideration exhibit nonelliptical return distributions. This observation might be especially useful for institutional investors who can benefit from reduced capital requirements when implementing our policies.

An interesting direction for future studies is the implementation of RS tail-risk-minimal hedging with more elaborate time series structures, in particular for the usage and evaluation of these strategies with daily and weekly data. This would include a systematic analysis of dynamic (conditional) tail risk hedging against the unconditional approach, which we favored throughout most of our work. Last but not least, the application of the ideas presented here to derive portfolio selections under tail risk constraints or objectives seems to be an interesting object of investigation.

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 $\boldsymbol{w}$

$$l_H(\boldsymbol{r}_S, \boldsymbol{r}_F, \boldsymbol{h}) = -\boldsymbol{w}' \cdot \boldsymbol{r}_S + \boldsymbol{h}' \cdot \boldsymbol{r}_F, \tag{39}$$

such that $L_H(\boldsymbol{h}) = l_H(\boldsymbol{R}_S, \boldsymbol{R}_F, \boldsymbol{h})$. With this definition, (A1) - (A3) imply that the conditions for Theorem 2 in Hong (2009) are satisfied. Assumption 1 in Hong (2009), i.e., partial differentiability of the loss function and its Lipschitz continuity, are implied by the linear structure of the function in (39) and the integrability constraints in (A1). (A2) 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, $\frac{\partial l_H}{\partial h_j} = r_{F,j}$, (A3) 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

$$(40) \qquad \frac{\partial v_\alpha(\boldsymbol{h})}{\partial h_j} = \frac{\partial q_{1-\alpha}}{\partial h_j}[l_H(\boldsymbol{R}_S, \boldsymbol{R}_F, \boldsymbol{h})] = \mathbb{E}\left[\frac{\partial l_H}{\partial h_j}(\boldsymbol{R}_S, \boldsymbol{R}_F, \boldsymbol{h}) \mid l_H(\boldsymbol{R}_S, \boldsymbol{R}_F, \boldsymbol{h}) = v_\alpha(\boldsymbol{h})\right].$$

Again, with $\frac{\partial l_H}{\partial h_j} = r_{F,j}$ the componentwise application of this result for $\boldsymbol{h}$ implies that (8) contains the FOCs for (5). 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 (5) has no solution. This is also true for (6). (9) follows as FOC for this problem from Theorem 3.1 in Hong and Liu (2009), which may be applied since (A1) - (A3) 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

$$(41) \qquad \frac{\partial c_\alpha(\boldsymbol{h})}{\partial h_j} = \mathbb{E}\left[\frac{\partial l_H}{\partial h_j}(\boldsymbol{R}_S, \boldsymbol{R}_F, \boldsymbol{h}) \mid l_H(\boldsymbol{R}_S, \boldsymbol{R}_F, \boldsymbol{h}) \geq v_\alpha(\boldsymbol{h})\right],$$

which proves (9).

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

$$(42) \qquad \begin{pmatrix} \boldsymbol{R}_F \\ L_H \end{pmatrix} \mid S = k \sim \mathcal{E}_{M+1}\left(\begin{pmatrix} \boldsymbol{\mu}_{F,k} \\ \mu_{L,k} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{F,k} & \boldsymbol{\Sigma}_{FL,k} \\ \boldsymbol{\Sigma}'_{FL,k} & \sigma^2_{L,k} \end{pmatrix}, g_k\right),$$

where the parameters are calculated according to (21), (22) 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 derived 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 (8) 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 (23). Using (42), this expectation can be decomposed into

$$(43) \qquad \mathbb{E}[\boldsymbol{R}_F \mid L_H = l] = \sum_{k=1}^{K} \frac{\pi_k}{f_L(l)} \mathbb{E}[\boldsymbol{R}_F \, \mathbb{1}(L_H = l) \mid S = k]$$

$$(44) \qquad = \sum_{k=1}^{K} \frac{\pi_k \, f_{L,k}(l)}{f_L(l)} \mathbb{E}[\boldsymbol{R}_F \mid L_H = l, S = k].$$

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

$$(45) \qquad \mathbb{E}[\boldsymbol{R}_F \mid L_H = l] = \sum_{k=1}^{K} \frac{\pi_k \, f_{L,k}(l)}{f_L(l)} \left[\boldsymbol{\mu}_{F,k} + \frac{\boldsymbol{\Sigma}_{FL,k}}{\sigma^2_{L,k}}(l - \mu_{L,k})\right].$$

This proves (28) 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 (45) 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 (A3) is valid under (M1).

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

$$(46) \qquad \mathbb{E}[\boldsymbol{R}_F \mid L_H \geq l] = \sum_{k=1}^{K} \frac{\pi_k\, \mathrm{P}(L_H \geq l | S = k)}{\mathrm{P}(L_H \geq l)}\, \mathbb{E}[\boldsymbol{R}_F \mid L_H \geq l, S = k].$$

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

$$(47) \qquad \mathbb{E}[\boldsymbol{R}_F \mid L_H \geq l, S = k] = \int_l^\infty \mathbb{E}[\boldsymbol{R}_F \mid L_H = x,\, S = k] \cdot f_{L_H | L_H \geq l, S=k}(x)\, \lambda(dx).$$

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

$$(48) \qquad \mathbb{E}[\boldsymbol{R}_F \mid L_H \geq l,\, S = k] = \boldsymbol{\mu}_{F,k} + \frac{\boldsymbol{\Sigma}_{FL,k}}{\sigma_{L,k}^2}\, [\mathbb{E}[L_H \mid L_H \geq l, S = k] - \mu_{L,k}].$$

We conclude that

$$(49) \qquad \mathbb{E}[\boldsymbol{R}_F \mid L_H \geq l] = \sum_{k=1}^{K} \frac{\pi_k\, (1 - F_{L,k}(l))}{\mathrm{P}(L_H \geq l)} \left[\boldsymbol{\mu}_{F,k} + \frac{\boldsymbol{\Sigma}_{FL,k}}{\sigma_{L,k}^2}\, (\mathbb{E}[L_H \mid L_H \geq l, S = k] - \mu_{L,k}) \right].$$

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

$$(50) \qquad \mathbb{E}[L_H \mid L_H \geq l, S = k] = \mu_{L,k} + \sigma_{L,k}\, \mathbb{E}\!\left[Z_k \mid Z_k \geq \frac{l - \mu_{L,k}}{\sigma_{L,k}} \right].$$

Again, for $l = v_\alpha(\boldsymbol{h})$ and with the definitions of $z_k(\boldsymbol{h})$ and $\lambda_k(\boldsymbol{h})$, we obtain (29) because $\mathrm{P}(L_H \geq v_\alpha(\boldsymbol{h})) = \alpha$. It remains to verify that (A2) is statisfied in our setting. This follows again 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

Agarwal, V. and Naik, N. Y. (2004). Risks and portfolio decisions involving hedge funds. *Review of Financial Studies*, 17(1):63–98.

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.

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.

Ang, A. and Chen, J. (2002). Asymmetric correlations of equity portfolios. *Journal of Financial Economics*, 63(3):443–494.

Ang, A. and Timmermann, A. (2012). Regime changes and financial markets. *Annual Review of Financial Economics*, 4(1):313–337.

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.

Brooks, C., Henry, Ó. T., and Persand, G. (2002). The Effect of Asymmetries on Optimal Hedge Ratios. *Journal of Business*, 75(2):333–352.

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.

Campbell, R., Huisman, R., and Koedijk, K. (2001). Optimal portfolio selection in a value-at-risk framework. *Journal of Banking & Finance*, 25(9):1789–1804.

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.

Daníelsson, J., Jorgensen, B. N., Samorodnitsky, G., Sarma, M., and de Vries, C. G. (2013). Fat tails, VaR and subadditivity. *Journal of Econometrics*, 172(2):283–291.

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

Embrechts, P. and Hofert, M. (2014). Statistics and quantitative risk management for banking and insurance. *Annual Review of Statistics and Its Application*, 1(1):493–514.

Embrechts, P., McNeil, A., and Straumann, D. (2002). Correlation and dependence in risk management: Properties and pitfalls. In Dempster, M., editor, *Risk Management: Value at Risk and Beyond*. 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.

Fishburn, P. C. (1977). Mean-risk analysis with risk associated with below-target returns. *The American Economic Review*, 67(2):116–126.

Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*. Springer Series in Statistics. Springer, New York.

Gaivoronski, A. A. and Pflug, G. (2005). Value-at-risk in portfolio optimization: properties and computational approach. *Journal Of Risk*, 7(2):1–31.

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.

Haas, M. (2009). Value-at-risk via mixture distributions reconsidered. *Applied Mathematics and Computation*, 215(6):2103–2119.

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.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton Univers. Press, Princeton, N.J.

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.

Kamdem, J. S. (2009). Delta-var and delta-tvar for portfolios with mixture of elliptic distributions risk factors and dcc. *Insurance: Mathematics and Economics*, 44(3):325–336.

Kelker, D. (1970). Distribution Theory of Spherical Distributions and a Location-Scale Parameter Generalization. *Sankhyā: The Indian Journal of Statistics, Series A*, 32(4):419–430.

Kon, S. J. (1984). Models of stock returns–a comparison. *Journal of Finance*, 39(1):147.

Kroner, K. F. and Sultan, J. (1993). Time-Varying Distributions and Dynamic Hedging with Foreign Currency Futures. *Journal of Financial and Quantitative Analysis*, 28(4):535–551.

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.

Longin, F. and Solnik, B. (2001). Extreme correlation of international equity markets. *Journal of Finance*, 56(2):649–676.

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.

Rockafellar, R. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, 2(3):21–42.

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

Rockafellar, R. T., Uryasev, S., and Zabarankin, M. (2006). Generalized deviations in risk analysis. *Finance and Stochastics*, 10(1):51–74.

Rydén, T., Teräsvirta, T., and Åsbrink, S. (1998). Stylized facts of daily return series and the hidden markov model. *Journal of Applied Econometrics*, 13(3):217–244.

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

Tasche, D. (2002). Conditional Expectation as Quantile Derivative. *unpublished*.

Telser, L. G. (1955). Safety first and hedging. *The Review of Economic Studies*, 23(1):1–16.

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

Tu, J. (2010). Is regime switching in stock returns important in portfolio decisions? *Management Science*, 56(7):1198–1215.