s:fi
RESEARCH

Swiss Finance Institute Research Paper Series N°19-27

A Flexible Regime Switching Model for Asset Returns

Marc S. Paolella
University of Zurich and Swiss Finance Institute

Pawel Polak
Columbia University

Patrick S. Walker
University of Zurich

A Flexible Regime Switching Model for Asset Returns

Marc S. Paolella[a,b] Paweł Polak[c*] Patrick S. Walker[a†]

[a] *Department of Banking and Finance, University of Zurich, Switzerland*
[b] *Swiss Finance Institute, Zurich, Switzerland*
[c] *Department of Statistics, Columbia University, New York, United States*

May 16, 2019

Abstract

A non-Gaussian multivariate regime switching dynamic correlation model for financial asset returns is proposed. It incorporates the multivariate generalized hyperbolic law for the conditional distribution of returns. All model parameters are estimated consistently using a new two-stage expectation-maximization algorithm that also allows for incorporation of shrinkage estimation via quasi-Bayesian priors. It is shown that use of Markov switching correlation dynamics not only leads to highly accurate risk forecasts, but also potentially reduces the regulatory capital requirements during periods of distress. In terms of portfolio performance, the new regime switching model delivers consistently higher Sharpe ratios and smaller losses than the equally weighted portfolio and all competing models. Finally, the regime forecasts are employed in a dynamic risk control strategy that avoids most losses during the financial crisis and vastly improves risk-adjusted returns.

Keywords: GARCH; Markov Switching; Multivariate Generalized Hyperbolic Distribution; Portfolio Optimization; Value-at-Risk.

JEL Classification: C32; C51; C53; G11; G17; G32.

*Corresponding author. Email: pawel.polak@columbia.edu

†This research was partially conducted while Walker was a visiting scholar at the Statistics Department at Columbia University. He is grateful for the financial support by the Swiss National Science Foundation (SNSF) through the Doc.Mobility fellowship no. P1ZHP1171760.

1 Introduction

Within the field of financial time series analysis, models for conditional heteroskedasticity of asset returns now have a long history and are widely applied to account for clustering effects in the time-varying volatility of asset returns; see, e.g., Bauwens et al. (2006) and Bollerslev (2010) for surveys. Several GARCH-type models in the multivariate setting, hereafter MGARCH, have been proposed, some of which are, in terms of parameter estimation, suitable for large dimensions. This includes the arguably most popular construction, namely the constant conditional correlation (CCC) model of Bollerslev (1990). For CCC, the filtered innovations from individual Gaussian univariate GARCH models applied to each constituent series are used via the usual plug-in estimator to estimate the correlation matrix. This assumes the correlations to be constant, but via the time-varying volatility from the individual fitted GARCH recursions, the covariance matrix changes through time. Based on the stylized facts of asset returns, of considerable interest are models that address the strong non-Gaussianity of the filtered GARCH innovation sequences. MGARCH models that allow for this include Aas et al. (2005), using the multivariate normal inverse Gaussian (NIG); Jondeau et al. (2007, Sec. 6.2) and Wu et al. (2015), using the multivariate skew-Student density; Santos et al. (2013), using a multivariate Student-t; and Virbickaite et al. (2016), using a Dirichlet location-scale mixture of multivariate normals. Multivariate generalizations of the univariate mixed normal GARCH model have been proposed and investigated by Bauwens et al. (2007) and Haas et al. (2009). Paolella and Polak (2015b) use the flexible class of multivariate generalized hyperbolic (hereafter MGHyp) distributions in a full maximum-likelihood framework; similar recent models using other non-Gaussian distribution families are Bianchi et al. (2016) and Slim et al. (2017).

One of the critiques of the popular CCC model of Bollerslev (1990) is that conditional covariances are time-varying only because of the univariate GARCH dynamics, while the correlations are constant. Addressing the stylized fact of time-varying correlations has become a highly active research area, with the seminal dynamic conditional correlation, or DCC, model of Engle (2002, 2009), and the varying correlation, or VC, model of Tse and Tsui (2002). These authors augment the baseline CCC model with a simple, two-parameter structure that allows for motion in the correlation coefficients. The fact that only two parameters are invoked to model the evolution of a correlation matrix is obviously a limitation, but can also be seen as its strong point: Estimation is straightforward, parameter proliferation in high dimensions is avoided, and the DCC and VC models can still pick up enough signal to improve forecasts over the CCC. Clearly, such a model (like GARCH itself) is a simplistic representation of a more complicated underlying process, and its efficacy is judged not on an economic modeling justification, but rather by an improved forecasting ability. Pelletier (2006) took a different approach, suggesting an MGARCH model with Markov switching between two CCC structures. We refer to this model subsequently as Gaussian-RSDC and build on this approach, addressing the non-Gaussianity of the underlying innovation process, and doing so in a coherent stochastic framework, as opposed to an ad-hoc two step approach commonly used to (possibly incorrectly) cobble together, e.g., a Gaussian GARCH-DCC or -RSDC model with a, say, multivariate Student-t distribution; see Paolella and Polak (2017) for a discussion of this issue.

In this paper, our goal is to introduce a model that lends itself to feasible estimation in high

dimensions, addresses the aforementioned stylized facts, and, crucially, leads to superior out-of-sample forecasting ability. Before discussing this new model, we continue with some relevant literature review upon which our model is inspired. So and Yip (2012) extend the DCC framework by grouping similar correlations into clusters and allowing different dynamics for the different clusters. Similarly, a clustering of variances in the DCC model is investigated by Aielli and Caporin (2014), whereas Kasch and Caporin (2013) extend the DCC framework by allowing the correlation dynamics to depend on the variances through a threshold structure. Billio and Caporin (2005) generalize the DCC framework through Markov switching in both the parameters of the DCC specification and the unconditional correlations. Models that allow for excess kurtosis and asymmetry of the distribution of innovations in the DCC model have been proposed in Paolella and Polak (2017) and Urga et al. (2011). The so-called DECO equicorrelation construction of Engle and Kelly (2012) models the correlation matrices as a parsimonious time-varying convex combination of the unity matrix and a matrix of ones. With our proposed model we build upon the strand of literature that invokes Markov switching. These models continue to gain prominence in their ability to disentangle differing states in the data generating process of financial asset returns; see e.g. Haas et al. (2004), Haas and Paolella (2012), and Henry (2009) in the univariate setting; and Billio and Pelizzon (2000), Chevallier and Goutte (2015), Chollete et al. (2009), Fink et al. (2017), and Pelletier (2006) in the multivariate case. Our model is closest to the Gaussian-RSDC of the latter article, in which a superior in-sample fit, compared to the GARCH-DCC model, is demonstrated. While improved in-sample fit does not necessarily imply improved (let alone economically significant) out-of-sample forecasting ability, it suggests that the changing correlations could be better modeled as being constant over short periods of time, and allowed to change according to a discrete Markov chain. With respect to out-of-sample performance of both density and risk prediction, we find, arguably unsurprisingly, that the Gaussianity assumption plays a large and detrimental role. The goal of this paper is to (i) address this factor in a coherent, non-ad-hoc way, such that parameter estimation in large dimensions is computationally feasible, straightforward, and consistent; and (ii) conduct extensive out-of-sample forecasting exercises to measure and compare its effectiveness to the DCC model, as well as the various models nested within our proposed structure. This includes the Gaussian-RSDC, the univariate switching model of Hamilton (1989, 1993), as well as univariate GARCH models that use non-Gaussian special cases of the generalized hyperbolic distributional assumption; see, among others, Jensen and Lunde (2001) and Aas and Haff (2006). Another special case is the unconditional MGHyp model, as examined in McNeil et al. (2015).

The proposed model also generalizes the class of so-called COMFORT models introduced in Paolella and Polak (2015b). An interesting feature of the MGHyp distribution employed in the COMFORT model is the introduction of a univariate stochastic term that elegantly and coherently gives rise to the non-Gaussianity. It can be interpreted as a common market factor, responsible for modeling news arrivals that jointly affect the distributions of all current period returns. While COMFORT is a very powerful model that outperforms all its special cases in terms of density forecasts, its dynamics cannot handle changes in the correlation structure because it is also endowed with a CCC structure. In this paper, we relax this assumption and allow for regime switches in the dependency matrix, suggesting COMFORT-RSDC as the name for our new model class. The regime switching correlation dynamics are empirically motivated by the well documented leverage

or down market effect, referring to the negative correlation between volatility and stock returns, as investigated, e.g., in the classical work of Black (1976); and the contagion effect, which describes the tendency of correlations between assets to shoot up during market crashes.
Crucially, the newly proposed model class is straightforward to estimate, even with a large number of assets, owing to a new two-step expectation maximization (EM) algorithm for likelihood maximization. This lends itself to straightforward estimation of the multivariate predictive density, as discussed below, and this for potentially large dimensions.
As detailed in the empirical section below, the COMFORT-RSDC model with two regimes significantly outperforms the single-regime case and all Gaussian-based competitors in an extensive in- and out-of-sample analysis. This includes a forecasting exercise for the daily return distribution and financial risk measures, as well as an investigation of out-of-sample portfolio optimization. To further motivate the use of our likelihood-based estimation algorithm, issues related to its convergence properties are studied.
The remainder of the paper is structured as follows. Section 2 presents the new model, while Section 3 discusses the proposed method of estimation and some of its properties. Section 4 provides a detailed empirical study of the performance of the various models. Section 5 concludes and discusses future potential work based on the new RSDC model. Finally, several appendices gather additional aspects of the model, including Appendix C, which provides mathematical details on the new estimation algorithm.

2 Model

Let $\mathbf{Y}_t = (Y_{t,1}, Y_{t,2}, \ldots, Y_{t,K})'$ denote a return vector of K financial assets at time t, for $t = 1, 2, \ldots, T$. The equally spaced realization of the return vector is denoted $\mathbf{Y} = [\mathbf{Y}_1 \mid \mathbf{Y}_2 \mid \cdots \mid \mathbf{Y}_T]$ and the information set at time t, defined as the sigma algebra generated by the history of returns $\{\mathbf{Y}_1, \ldots, \mathbf{Y}_t\}$, is denoted $\mathbf{\Phi}_t$. In what follows, we assume that $\mathbf{Y}_t$ has a time varying conditional distribution with the COMFORT representation, from Paolella and Polak (2015b) given by

$$\begin{aligned} \mathbf{Y}_t \mid \mathbf{\Phi}_{t-1} &\stackrel{d}{=} \boldsymbol{\mu} + \boldsymbol{\gamma} G_t + \boldsymbol{\varepsilon}_t, \quad \text{with} \\ \boldsymbol{\varepsilon}_t &= \mathbf{H}_t^{1/2} \sqrt{G_t} \mathbf{Z}_t, \end{aligned} \tag{1}$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)'$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)'$ are column vectors in $\mathbb{R}^K$; $\mathbf{H}_t$ is a symmetric, positive definite dispersion matrix of order K; $\mathbf{Z}_t \overset{\text{iid}}{\sim} \mathrm{N}(\mathbf{0}, \mathbf{I}_K)$ is a sequence of independent and identically distributed (i.i.d.) normal random variables; and $G_t \sim \mathrm{GIG}(\lambda, \chi, \psi)$ are i.i.d. mixing random variables, independent of $\mathbf{Z}_t$, with the generalized inverse Gaussian (GIG) density given by

$$f_G(x; \lambda, \chi, \psi) = \frac{\chi^{-\lambda} \left(\sqrt{\chi\psi}\right)^{\lambda}}{2\mathcal{K}_\lambda\left(\sqrt{\chi\psi}\right)} x^{\lambda-1} \exp\left(-\frac{1}{2}\left(\chi x^{-1} + \psi x\right)\right), \quad x > 0; \tag{2}$$

$\mathcal{K}_\lambda(x)$ is the modified Bessel function of the third kind, given by

$$\mathcal{K}_\lambda(x) = \frac{1}{2} \int_0^\infty t^{\lambda-1} \exp\left(-\frac{x}{2}\left(t + t^{-1}\right)\right) \mathrm{d}t, \quad x > 0; \tag{3}$$

and $\chi > 0$, $\psi \geq 0$ if $\lambda < 0$; $\chi > 0$, $\psi > 0$ if $\lambda = 0$; and $\chi \geq 0$, $\psi > 0$ if $\lambda > 0$. The same MGHyp distribution arises from the parameter constellation $(\lambda, \chi/c, c\psi, \boldsymbol{\mu}, c\mathbf{H}_t, c\boldsymbol{\gamma})$ for any $c > 0$;

hence for identification purposes, we fix either χ or ψ. The model assumes $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)'$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)'$, as well as the GIG parameters (λ, χ, ψ) to be time invariant. For the GIG distribution to be well defined, $\mathbb{E}[G_t \mid \boldsymbol{\Phi}_{t-1}]$ and $\mathbb{E}\left[G_t^2 \mid \boldsymbol{\Phi}_{t-1}\right]$ have to be positive; sufficient conditions are given in Paolella and Polak (2015b).

The conditional, symmetric, positive definite dispersion matrix $\mathbf{H}_t$ is decomposed as

$$\mathbf{H}_t = \mathbf{S}_t \boldsymbol{\Gamma}_t \mathbf{S}_t, \tag{4}$$

where $\mathbf{S}_t$ is a diagonal matrix holding the strictly positive conditional scale terms $s_{k,t}$, $k = 1, \ldots, K$, and $\boldsymbol{\Gamma}_t$ is a positive definite dependency matrix. The univariate scale terms $s_{k,t}$ are each modeled by a GARCH-type process. To describe the volatility clustering we use the GARCH(1,1) model defined by

$$s_{k,t}^2 = \omega_k + \alpha_k \varepsilon_{k,t-1}^2 + \beta_k s_{k,t-1}^2, \tag{5}$$

where $\varepsilon_{k,t} = y_{k,t} - \mu_k - \gamma_k G_t$, is the kth element of the $\boldsymbol{\varepsilon}_t$ vector in (1), and $\omega_k > 0$, $\alpha_k \geq 0$, $\beta_k \geq 0$, for $k = 1, 2, \ldots, K$.

In our model, $\boldsymbol{\mu}$ and $\mathbf{H}_t$ are the location vector and the dispersion matrix of the conditional distribution of $\mathbf{Y}_t$, respectively, while the mean and the covariance matrix are given by

$$\mathbb{E}[\mathbf{Y}_t \mid \boldsymbol{\Phi}_{t-1}] = \boldsymbol{\mu} + \mathbb{E}[G_t \mid \boldsymbol{\Phi}_{t-1}]\, \boldsymbol{\gamma} \tag{6}$$

and

$$\mathrm{Cov}(\mathbf{Y}_t \mid \boldsymbol{\Phi}_{t-1}) = \mathbb{E}[G_t \mid \boldsymbol{\Phi}_{t-1}]\, \mathbf{H}_t + \mathbb{V}(G_t \mid \boldsymbol{\Phi}_{t-1})\, \boldsymbol{\gamma}\boldsymbol{\gamma}', \tag{7}$$

respectively, where $\mathbb{V}(G_t \mid \boldsymbol{\Phi}_{t-1}) = \mathbb{E}\left[G_t^2 \mid \boldsymbol{\Phi}_{t-1}\right] - (\mathbb{E}[G_t \mid \boldsymbol{\Phi}_{t-1}])^2$. Furthermore, by definition of the mean-variance mixture distribution of $\mathbf{Y}_t \mid \boldsymbol{\Phi}_{t-1}$, the matrix $\boldsymbol{\Gamma}_t$ is a (time conditional) correlation matrix, conditionally on the realization of the mixing process G_t. For convenience and to make the analogy to the Gaussian-based literature, we use the names dependency matrix and correlation matrix interchangeably for $\boldsymbol{\Gamma}_t$. The CCC model entails using the time invariant dependency matrix $\boldsymbol{\Gamma}_t = \boldsymbol{\Gamma}$, as is also the case in the COMFORT model of Paolella and Polak (2015b). Dynamics in the dependency matrix can be incorporated into their model with the DCC approach of Engle (2002), as shown in Paolella and Polak (2015c).

Unfortunately, the latter authors demonstrate that the lion's share of the improved gains in out-of-sample forecasts are generated from relaxing the Gaussian assumption (with or without GARCH), followed by use of a GARCH-type process for the margins, as compared to an i.i.d. assumption. While in a Gaussian framework, use of DCC does offer improvements to forecasting compared to CCC, when using (several special cases of) the MGHyp, the added benefit of DCC over CCC is negligible. We conjecture that, in a relatively heavily misspecified model such as the Gaussian-CCC, any reasonable additional structure to the data generating process will help improve forecasts. However, when the non-Gaussianity aspect is addressed, the sparsely parametrized DCC structure no longer is able to convey a measurable advantage. This is in no way to be interpreted as either that the returns do not have time-varying correlations, nor that the DCC model should be discarded. Quite on the contrary, not only do asset returns display prominent time-varying correlation, but also the DCC structure is admirable for its use of only two additional parameters to account for some of this movement. In light of these observations, we believe that a richer DCC structure, such as the matrix-variate versions proposed in Billio et al.

(2006) and Cappiello et al. (2006), would be able to extract further signal that leads to enhanced forecast ability. The problem with such constructions is the proliferation of parameters amid a finite amount of data, and the more practical aspect of the need to estimate them jointly - this being infeasible for even modest dimensions K, let alone large asset portfolios corresponding to those used by major financial institutions. In other words, we believe that such models are "asking too much" from the data, and instead of pursuing a model for the time-varying correlations that allows continuous changes in correlations at every point in time, we advocate the use of a (small) set of fixed constant correlation matrices incorporated into a Markov switching framework, as in Pelletier (2006) in the Gaussian setting. The stochastic dependency matrix $\mathbf{\Gamma}_t$ at time t under a Markov switching model with N regimes is given by

$$\mathbf{\Gamma}_t = \sum_{n=1}^{N} \mathbf{1}_{\{\Delta_t = n\}} \mathbf{\Gamma}_n, \tag{8}$$

where $\mathbf{1}_{\{\}}$ is an indicator function; Δ_t is a latent random variable governed by a first order, homogeneous Markov chain, independent of $\mathbf{Z}_t$ and G_t, which can take one of N possible values; and $\mathbf{\Gamma}_n$, $n = 1, \ldots, N$, are state-specific $K \times K$ symmetric, positive definite correlation matrices such that $\mathbf{\Gamma}_n \neq \mathbf{\Gamma}_m$ for $n \neq m$. The probability law governing Δ_t is defined by its time-invariant transition probability matrix, denoted by $\mathbf{\Pi} = [\pi_{n,m}]_{n,m=1,\ldots,N}$, where $\pi_{n,m}$ is the probability of going from state n in period t to state m in period $t+1$.
The unconditional covariance of $\mathbf{Y}_t$ can be expressed in terms of the unconditional covariance function of $\mathbf{Y}_t \mid G_t$ as

$$\operatorname{Cov}(\mathbf{Y}_t) = \mathbb{E}\left[\operatorname{Cov}(\mathbf{Y}_t \mid G_t)\right] + \mathbb{V}(G_t)\,\boldsymbol{\gamma}\boldsymbol{\gamma}'. \tag{9}$$

Thus, our model inherits the lack of an explicit form for $\operatorname{Cov}(\mathbf{Y}_t)$ from the CCC model.

The predictive multivariate probability density function of returns, i.e., the density of the distribution of $\mathbf{Y}_{t+1} \mid \mathbf{\Phi}_t$, is a finite mixture of MGHyp densities with different correlation matrices, weighted by the forecasted probabilites of next period's regimes

$$f_{\mathbf{Y}_{t+1}|\mathbf{\Phi}_t}(\mathbf{y}_{t+1}) = \sum_{n=1}^{N} \xi_{n,t+1|t}\; f_{\mathbf{Y}_{t+1}|\mathbf{\Phi}_t}(\mathbf{y}_{t+1} \mid \Delta_{t+1} = n), \tag{10}$$

where $\xi_{n,t+1|t}$, for $n = 1, \ldots, N$, is the forecasted probability that regime n will manifest itself at time $t+1$, given the information up to time t.

In the remainder of this paper, we restrict our attention to three non-Gaussian special cases of the MGHyp family, namely the multivariate Laplace, normal inverse Gaussian, and the Student-t distribution. In agreement with Prause (1999) and Protassov (2004), Paolella and Polak (2015b) find that the MGHyp exhibits a relatively flat likelihood in some of the parameters, and advocate fixing some parameters to overcome numerical issues. These special cases are faster and numerically more reliable to estimate, but still retain the flexibility required for modeling asset returns. Crucially, they allow for higher kurtosis than the normal distribution and also for individual asset asymmetry parameters. The multivariate asymmetric Laplace (MALap) or variance-gamma

distribution is obtained for $G_t \sim \text{GIG}(\lambda, \chi, \psi)$ with $\lambda > 0$, $\chi = 0$ and $\psi = 2$; the multivariate asymmetric normal inverse Gaussian (NIG) distribution for $\lambda = -1/2$, $\chi > 0$ and $\psi = 1$; and the multivariate asymmetric Student-t (MAt) distribution for $\lambda = -\chi/2$ and $\psi = 0$. The corresponding symmetric distributions with $\boldsymbol{\gamma} = \mathbf{0}$ are abbreviated as MLap, SNIG, and Mt respectively. Note that all univariate margins in the MGHyp distribution family possess the same tail thickness, which is a potential concern of this distributional assumption when the constituent series have markedly different tail behaviors. It is not obvious how to generalize the current framework to allow for different tail behaviors and still preserve the analytic tractability of the portfolio distribution (weighted sum of the margins) arising from the multivariate predictive density. As an example, copula structures can easily address the heterogeneous tail behavior of the margins issue, but at the price of losing tractability of the portfolio distribution (and thus require simulation).

3 Two-Stage Estimation for the Regime Switching Correlation Model

Paolella and Polak (2015b) develop an expectation conditional maximization either (ECME) algorithm appropriate for estimating the model under the assumption of a constant conditional correlation matrix, i.e., the single-component case. To generalize this to the N-component case, in principle, one could extend that algorithm into a nested double ECME algorithm, with the inner loop estimating the regime switching correlations. However, the rapid increase of the number of parameters combined with the dynamic structure in the conditional correlation matrix of the RSDC model makes even this, nested ECME, estimation method impractical for large K. To overcome this, a new, two-stage estimation procedure is proposed. It replaces a double iterative procedure by two separate procedures applied iteratively, whereby the second procedure is conducted conditional on the results from the first one.

For notational convenience later, we collect the parameters of the model into four vectors (process, distribution, correlation matrices and Markov transition matrix)

$$\begin{aligned} \boldsymbol{\theta}_P &= \left(\boldsymbol{\mu}', \boldsymbol{\gamma}', \boldsymbol{\omega}', \boldsymbol{\alpha}', \boldsymbol{\beta}'\right)', \quad \boldsymbol{\theta}_D = (\lambda, \chi, \psi)', \\ \boldsymbol{\theta}_C &= \left(\text{vech}\left(\boldsymbol{\Gamma}_1\right)', \ldots, \text{vech}\left(\boldsymbol{\Gamma}_N\right)'\right)', \text{ and } \boldsymbol{\theta}_M = \left(\text{vec}_{N-1}\left(\boldsymbol{\Pi}\right)\right)', \end{aligned} \tag{11}$$

where $\boldsymbol{\omega}, \boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are K-dimensional vectors of GARCH$(1,1)$ parameters from (5); (λ, χ, ψ) denote the constant GIG parameters; $\text{vech}\left(\boldsymbol{\Gamma}_n\right)$, for $n = 1, \ldots, N$, denotes a column vector of the elements above the main diagonal of matrix $\boldsymbol{\Gamma}_n$; and $\text{vec}_{N-1}\left(\boldsymbol{\Pi}\right)$ is a column vector of first $N-1$ columns of the transition probabilities matrix $\boldsymbol{\Pi}$. For just the CCC model, $\boldsymbol{\theta}_C$ reduces to $\text{vech}\left(\boldsymbol{\Gamma}\right)$ and $\boldsymbol{\theta}_M = \emptyset$ because there is no Markov chain to govern regimes; for brevity in this case we omit $\boldsymbol{\theta}_M$ in the formulas.

The estimation algorithm is based on the decomposition of the complete conditional log-likelihood function into a sum of three terms: the normal log-likelihood (i.e. conditionally on the realization of the mixing random variable and the state of Markov chain), the log-likelihood function of the mixing random variable (which is state independent), and the log-likelihood of the Markov chain

variable i.e.,

$$\log L_{\mathbf{Y},\mathbf{G},\boldsymbol{\Delta}}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_D,\boldsymbol{\theta}_C,\boldsymbol{\theta}_M\right)=\log L_{\mathbf{Y}|\mathbf{G},\boldsymbol{\Delta}}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_C,\boldsymbol{\theta}_M\right)+\log L_{\mathbf{G}}\left(\boldsymbol{\theta}_D\right)+\log L_{\boldsymbol{\Delta}}(\boldsymbol{\theta}_M). \tag{12}$$

Next, the normal log-likelihood can be split as in Bollerslev (1990) and Pelletier (2006), i.e.,

$$\log L_{\mathbf{Y}|\mathbf{G},\boldsymbol{\Delta}}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_C,\boldsymbol{\theta}_M\right)=\log L^{\mathrm{MV}}_{\mathbf{Y}|\mathbf{G}}\left(\boldsymbol{\theta}_P\right)+\log L^{\mathrm{Corr}}_{\mathbf{Y}|\mathbf{G},\boldsymbol{\Delta}}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_C,\boldsymbol{\theta}_M\right), \tag{13}$$

where $L^{\mathrm{MV}}_{\mathbf{Y}|\mathbf{G}}\left(\boldsymbol{\theta}_P\right)$ is the mean-volatility term given by

$$\begin{aligned}\log L^{\mathrm{MV}}_{\mathbf{Y}|\mathbf{G}}\left(\boldsymbol{\theta}_P\right)=-\frac{1}{2}\sum_{t=1}^{T}\Big[&K\log\left(2\pi\right)+\log|\mathbf{S}_t|^2\\&+g_t^{-1}\left(\mathbf{y}_t-\boldsymbol{\mu}-\boldsymbol{\gamma}g_t\right)'\mathbf{S}_t^{-1}\mathbf{S}_t^{-1}\left(\mathbf{y}_t-\boldsymbol{\mu}-\boldsymbol{\gamma}g_t\right)+\log g_t\Big],\end{aligned} \tag{14}$$

and $L^{\mathrm{Corr}}_{\mathbf{Y}|\mathbf{G},\boldsymbol{\Delta}}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_C,\boldsymbol{\theta}_M\right)$ is the correlation term given by

$$\log L^{\mathrm{Corr}}_{\mathbf{Y}|\mathbf{G},\boldsymbol{\Delta}}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_C,\boldsymbol{\theta}_M\right)=-\frac{1}{2}\sum_{n=1}^{N}\left(\sum_{t:\Delta_t=n}\left[\log|\boldsymbol{\Gamma}_n|+\boldsymbol{e}_t'\boldsymbol{\Gamma}_n^{-1}\boldsymbol{e}_t-\boldsymbol{e}_t'\boldsymbol{e}_t\right]\right), \tag{15}$$

where $\boldsymbol{e}_t=g_t^{-1/2}\mathbf{S}_t^{-1}\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\varepsilon}_t=\mathbf{y}_t-\boldsymbol{\mu}-\boldsymbol{\gamma}g_t$ are the residuals from (1).

Owing to the mixture structure of the MGHyp, $L_{\mathbf{Y}|\mathbf{G},\boldsymbol{\Delta}}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_C,\boldsymbol{\theta}_M\right)$ is a multivariate Gaussian likelihood with a GARCH structure for the scales and a given conditional correlation model. As such, maximization of $L_{\mathbf{Y}|\mathbf{G},\boldsymbol{\Delta}}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_C,\boldsymbol{\theta}_P\right)$ can be done in two steps. First, with the correlation structure ignored, the GARCH parameters in (5) are estimated for each of the K assets separately (or concurrently with parallel computing) by maximizing $L_{\mathbf{Y}|\mathbf{G}}\left(\boldsymbol{\theta}_P\mid\boldsymbol{\theta}_C=\mathbf{I}_K\right)$, and, in the second step, the correlation parameters $\boldsymbol{\theta}_C$ and $\boldsymbol{\theta}_M$ are estimated from the first-step standardized residuals.

In case of a constant dependency matrix, as in Paolella and Polak (2015b), the $\boldsymbol{\Gamma}$ matrix is estimated by the usual empirical correlation estimator (the MLE under normality) of the standardized residuals $g_t^{-1/2}\mathbf{S}_t^{-1}\widehat{\boldsymbol{\varepsilon}}_t$, with the unobserved realizations of $g_t^{-1/2}$ replaced by their conditional expectations from the E-step below; this being computationally trivial. When $\boldsymbol{\Gamma}_t$ has a regime switching structure, the correlation step would be considerably slower.

Therefore, we propose to keep correlation estimates fixed at some $\boldsymbol{\theta}_C^*$ and $\boldsymbol{\theta}_M^*$ in the ECME algorithm and turn directly to the maximization of the second term in the decomposition (12). Given the $\boldsymbol{\theta}_P$ estimates, denoted as usual by $\widehat{\boldsymbol{\theta}}_P$, estimate mixing process parameters $\boldsymbol{\theta}_D$ by maximizing $L_{\mathbf{Y}}\big(\boldsymbol{\theta}_D\mid\widehat{\boldsymbol{\theta}}_P,\boldsymbol{\theta}_C^*,\boldsymbol{\theta}_M^*\big)$. This implies that, in each iteration of the algorithm, we update $\boldsymbol{\theta}_P$ and $\boldsymbol{\theta}_D$, and keep $\boldsymbol{\theta}_C^*$ and $\boldsymbol{\theta}_M^*$ fixed. Given all these estimates, we proceed with the next E-step update of the unobserved mixing random variable $\mathbf{G}$ (still with fixed $\boldsymbol{\theta}_C^*$ and $\boldsymbol{\theta}_M^*$) and continue to iterate until convergence. This completes the Stage-I ECME algorithm.

The updates of the correlation dynamics, $\boldsymbol{\theta}_C$ and $\boldsymbol{\theta}_M$, are obtained in the second stage by a separate EM algorithm, conditional on the estimates in the first stage. By iterating the two algorithms we obtain the proposed two-stage EM. We now present the details of our proposed two-stage EM algorithm for parameter estimation of the COMFORT-RSDC model.

Stage-I ECME algorithm

Stage-I E-step: Calculate $\mathbb{E}\big[\log L_{\mathbf{Y},\mathbf{G}} \mid \boldsymbol{\Phi}; \boldsymbol{\theta}_P^{[\ell]}, \boldsymbol{\theta}_D^{[\ell]}, \boldsymbol{\theta}_C^*, \boldsymbol{\theta}_M^*\big]$. The log-likelihood (14) is linear with respect to g_t and g_t^{-1}, so that we can replace the unobserved realizations of G_t and G_t^{-1} in (12) by their conditional expectations $\mathbb{E}\big[G_t^{\pm 1} \mid \boldsymbol{\Phi}; \boldsymbol{\theta}_P^{[\ell]}, \boldsymbol{\theta}_D^{[\ell]}, \boldsymbol{\theta}_C^*, \boldsymbol{\theta}_M^*\big]$, where

$$\left(G_t \mid \boldsymbol{\Phi}; \boldsymbol{\theta}_P^{[\ell]}, \boldsymbol{\theta}_D^{[\ell]}, \boldsymbol{\theta}_C^*, \boldsymbol{\theta}_M^*\right) \sim \text{GIG}\left(\lambda^{[\ell]} - K/2, \chi^{[\ell]} + \|\mathbf{Y}_t - \boldsymbol{\mu}^{[\ell]}\|^2_{\mathbf{H}_t^{[\ell]}}, \psi^{[\ell]} + \|\boldsymbol{\gamma}^{[\ell]}\|^2_{\mathbf{H}_t^{[\ell]}}\right) \tag{16}$$

and $\|\mathbf{x}\|^2_{\mathbf{H}} = \mathbf{x}'\mathbf{H}^{-1}\mathbf{x}$, for any vector $\mathbf{x}$ and a square matrix $\mathbf{H}$ of proper dimension.

Stage-I CM1-step: Update $\boldsymbol{\theta}_P$ by computing

$$\boldsymbol{\theta}_P^{[\ell+1]} = \arg\max_{\boldsymbol{\theta}_P} \log L^{MV}_{\mathbf{Y}|\mathbf{G}}\left(\boldsymbol{\theta}_P\right), \tag{17}$$

where $L^{MV}_{\mathbf{Y}|\mathbf{G}}\left(\boldsymbol{\theta}_P\right)$ is a Gaussian likelihood with zero correlations. This allows to estimate the parameters $(\mu_k, \gamma_k, \omega_k, \alpha_k, \beta_k)$ separately for each asset by maximizing the corresponding GARCH likelihood.

Stage-I CM2-step: Given the CM1-step estimates of $\boldsymbol{\theta}_P$, update the estimates of $\boldsymbol{\theta}_D$ by maximizing the incomplete data log-likelihood

$$\boldsymbol{\theta}_D^{[\ell+1]} = \arg\max_{\boldsymbol{\theta}_D} \log L_{\mathbf{Y}}\left(\boldsymbol{\theta}_D \mid \boldsymbol{\theta}_P^{[\ell+1]}, \boldsymbol{\theta}_C^*, \boldsymbol{\theta}_M^*\right). \tag{18}$$

Iterate these E- and CM-steps until convergence, to get $\hat{\boldsymbol{\theta}}_P$ and $\hat{\boldsymbol{\theta}}_D$.

Then proceed with Stage-II to update the regime switching parameters $\boldsymbol{\theta}_C$ and $\boldsymbol{\theta}_M$. This is done conditional on $\hat{\boldsymbol{\theta}}_P$ and $\hat{\boldsymbol{\theta}}_D$ by the maximization

$$\arg\max_{\boldsymbol{\theta}_C, \boldsymbol{\theta}_M} \log L_{\mathbf{Y}}\big(\boldsymbol{\theta}_C, \boldsymbol{\theta}_M \mid \hat{\boldsymbol{\theta}}_P, \hat{\boldsymbol{\theta}}_D\big). \tag{19}$$

The Stage-II algorithm is itself an EM algorithm, conditional on the Stage-I estimates, and generalizes the algorithm of Hamilton (1993, 1994) to the case of MGHyp innovations. The smoothing formula for inference of the regime probabilities of Kim (1994) is adopted. A quasi-Bayesian approach is utilized for shrinkage estimation of the correlation matrices as in Hamilton (1991) and Paolella (2015) to reduce the estimation error in finite samples and to increase the forecasting performance of the model (see below).

Stage-II ECME algorithm

Stage II-1: Set the iteration count to $\ell = 1$ and choose appropriate starting values for $\boldsymbol{\theta}_C^{[1]} = \left[\boldsymbol{\Gamma}_1^{[1]}, \ldots, \boldsymbol{\Gamma}_N^{[1]}\right]$, and $\boldsymbol{\theta}_M^{[1]} = \boldsymbol{\Pi}^{[1]}$.

Stage-II E-step-1: Recursively calculate the forecasted and inferred state probabilities $\boldsymbol{\xi}_{t+1|t}^{[\ell]}$ and $\boldsymbol{\xi}_{t|t}^{[\ell]}$, as well as the smoothed inference $\boldsymbol{\xi}_{t|T}^{[\ell]}$, as derived in Kim (1994), from iterating on

$$\boldsymbol{\xi}_{t|t}^{[\ell]} = \frac{\boldsymbol{\xi}_{t|t-1}^{[\ell]} \odot \boldsymbol{\eta}_t^{[\ell]}}{\mathbf{1}_N'\left(\boldsymbol{\xi}_{t|t-1}^{[\ell]} \odot \boldsymbol{\eta}_t^{[\ell]}\right)},\ \boldsymbol{\xi}_{t+1|t}^{[\ell]} = \boldsymbol{\Pi}^{[\ell]}\boldsymbol{\xi}_{t|t}^{[\ell]},\ \boldsymbol{\xi}_{t|T}^{[\ell]} = \boldsymbol{\xi}_{t|t}^{[\ell]} \odot \left\{\boldsymbol{\Pi}^{[\ell]\prime}\left[\boldsymbol{\xi}_{t+1|T}^{[\ell]} \div \boldsymbol{\xi}_{t+1|t}^{[\ell]}\right]\right\}, \tag{20}$$

where $\odot$ and $\div$ denote the element by element product and division, respectively, and

$$\boldsymbol{\eta}_t^{[\ell]} = \begin{bmatrix} f\left(\mathbf{y}_t \mid \boldsymbol{\Phi}_{t-1}, \Delta_t = 1; \hat{\boldsymbol{\theta}}_P, \hat{\boldsymbol{\theta}}_D, \boldsymbol{\Gamma}_1^{[\ell]}\right) \\ f\left(\mathbf{y}_t \mid \boldsymbol{\Phi}_{t-1}, \Delta_t = 2; \hat{\boldsymbol{\theta}}_P, \hat{\boldsymbol{\theta}}_D, \boldsymbol{\Gamma}_2^{[\ell]}\right) \\ \vdots \\ f\left(\mathbf{y}_t \mid \boldsymbol{\Phi}_{t-1}, \Delta_t = N; \hat{\boldsymbol{\theta}}_P, \hat{\boldsymbol{\theta}}_D, \boldsymbol{\Gamma}_N^{[\ell]}\right) \end{bmatrix}. \tag{21}$$

Here, $\boldsymbol{\xi}_{s|t}^{[\ell]}$ denotes the N-dimensional vector of probabilities of observing each of the N states at time s, based on the data obtained through date t and based on knowledge of the population parameters (11). Moreover, $\mathbf{1}_N$ is an $N \times 1$ vector of ones. The starting values $\hat{\boldsymbol{\xi}}_{1|0}$ are chosen by maximum likelihood estimation by extending the parameter space accordingly.

Stage-II E-step-2: Calculate $\mathbb{E}\big[\log L_{\mathbf{Y},\mathbf{G},\boldsymbol{\Delta}} \mid \boldsymbol{\Phi}_t; \hat{\boldsymbol{\theta}}_P, \hat{\boldsymbol{\theta}}_D, \boldsymbol{\theta}_C^{[\ell]}, \boldsymbol{\theta}_M^{[\ell]}\big]$. The log-likelihood (14) is linear with respect to g_t and g_t^{-1}, so we replace the unobserved realizations of G_t and G_t^{-1} in (12) by their conditional expectations $\mathbb{E}\big[G_t^{\pm 1} \mid \boldsymbol{\Phi}_t, \boldsymbol{\Delta}_t; \hat{\boldsymbol{\theta}}_P, \hat{\boldsymbol{\theta}}_D, \boldsymbol{\theta}_C^{[\ell]}, \boldsymbol{\theta}_M^{[\ell]}\big]$. In particular,

$$\boldsymbol{g}_t^{\pm 1} = \begin{bmatrix} \mathbb{E}\left[G_t^{\pm 1} \mid \boldsymbol{\Phi}_t, \Delta_t = 1; \hat{\boldsymbol{\theta}}_P, \hat{\boldsymbol{\theta}}_D, \boldsymbol{\theta}_C^{[\ell]}, \boldsymbol{\theta}_M^{[\ell]}\right] \\ \mathbb{E}\left[G_t^{\pm 1} \mid \boldsymbol{\Phi}_t, \Delta_t = 2; \hat{\boldsymbol{\theta}}_P, \hat{\boldsymbol{\theta}}_D, \boldsymbol{\theta}_C^{[\ell]}, \boldsymbol{\theta}_M^{[\ell]}\right] \\ \vdots \\ \mathbb{E}\left[G_t^{\pm 1} \mid \boldsymbol{\Phi}_t, \Delta_t = N; \hat{\boldsymbol{\theta}}_P, \hat{\boldsymbol{\theta}}_D, \boldsymbol{\theta}_C^{[\ell]}, \boldsymbol{\theta}_M^{[\ell]}\right] \end{bmatrix}, \tag{22}$$

for

$$\left(G_t \mid \boldsymbol{\Phi}_t; \hat{\boldsymbol{\theta}}_P, \hat{\boldsymbol{\theta}}_D, \boldsymbol{\theta}_C^{[\ell]}, \Delta_t = n\right) \sim \text{GIG}\left(\lambda^{[\ell]}, \chi_{n,t}^{[\ell]}, \psi_{n,t}^{[\ell]}\right), \tag{23}$$

where $\lambda^{[\ell]} = \hat{\lambda} - K/2$, $\chi_{n,t}^{[\ell]} = m_{n,t}^{[\ell]} + \hat{\chi}$, $\psi_{n,t}^{[\ell]} = \hat{\psi} + \hat{\boldsymbol{\gamma}}' \left(\mathbf{S}_{n,t}\boldsymbol{\Gamma}_n^{[\ell]}\mathbf{S}_{n,t}\right)^{-1} \hat{\boldsymbol{\gamma}}$ and

$$m_{n,t}^{[\ell]} = (\mathbf{y}_t - \hat{\boldsymbol{\mu}})' \left(\mathbf{S}_{n,t}\boldsymbol{\Gamma}_n^{[\ell]}\mathbf{S}_{n,t}\right)^{-1} (\mathbf{y}_t - \hat{\boldsymbol{\mu}}). \tag{24}$$

The conditional expectations of G_t and G_t^{-1} in (22) can then be calculated from the explicit formula for the moments of a GIG random variable, see e.g. Paolella (2007, Sec. 9.4). The $\mathbf{S}_{n,t}$ are diagonal matrices collecting the estimated scale terms generated by equation (5) using the first-step estimates of $\boldsymbol{\theta}_P$ and $\boldsymbol{\theta}_D$, and $g_{n,t}^{-1}$ the regime-specific estimates of G_t in (22) corresponding to regime n. Note that these estimates depend on the regimes because we have to condition on a regime to compute them.

Stage-II CM-1-step: Update the estimates of the correlation matrices by a modification of the quasi-Bayesian estimator of Hamilton (1991), analogous to Paolella (2015),

$$\tilde{\boldsymbol{\Gamma}}_n^{[\ell+1]} = \frac{a_n\mathbf{B}_n + \sum_{t=1}^T g_{n,t}^{-1}\mathbf{S}_{n,t}^{-1} (\mathbf{y}_t - \hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\gamma}} g_{n,t}) (\mathbf{y}_t - \hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\gamma}} g_{n,t})' \mathbf{S}_{n,t}^{-1} \xi_{n,t|T}^{[\ell]}}{a_n + \sum_{t=1}^T \xi_{n,t|T}^{[\ell]}}, \tag{25}$$

where the tilde indicates that the matrix must be standardized to be a correlation matrix, and fixed quantities $a_n \geq 0$ and $\mathbf{B}_n$ represent the weight, or strength, and the target of the nth prior

correlation matrix of the shrinkage estimator. Every $\mathbf{B}_n$ must be symmetric, positive definite with 1 on the diagonal and all off-diagonal elements between -1 and 1. By $\xi_{n,t|T}^{[\ell]}$ we denote the nth element of the probability vector $\boldsymbol{\xi}_{t|T}^{[\ell]}$ from (20). Hatted parameters stem from the Stage-I ECME algorithm.

The matrix (25) must be rescaled to be a correlation matrix. This is done as in Pelletier (2006) by setting

$$\boldsymbol{\Gamma}_n^{[\ell+1]} = \mathbf{D}_n^{-1[\ell+1]} \tilde{\boldsymbol{\Gamma}}_n^{[\ell+1]} \mathbf{D}_n^{-1[\ell+1]}, \tag{26}$$

for $n = 1, \ldots, N$ with

$$\mathbf{D}_n^{[\ell+1]} = \operatorname{diag}\left(\sqrt{\tilde{\boldsymbol{\Gamma}}_{1,1,n}^{[\ell+1]}}, \ldots, \sqrt{\tilde{\boldsymbol{\Gamma}}_{K,K,n}^{[\ell+1]}} \right). \tag{27}$$

Stage-II CM-2-step: The Markov transition matrix $\boldsymbol{\Pi} = (\pi_{n,m})_{n,m=1,\ldots,N}$ is then updated using the formula in Hamilton (1994, eq. (22.4.16)) by

$$\pi_{n,m}^{[\ell+1]} = \sum_{t=2}^{T} \xi_{m,t|T}^{[\ell]} \frac{\pi_{n,m}^{[\ell]} \xi_{n,t-1|t-1}^{[\ell]}}{\xi_{m,t|t-1}^{[\ell]}} \Bigg/ \sum_{t=2}^{T} \xi_{n,t-1|T}^{[\ell]}. \tag{28}$$

These steps are iterated until convergence. This closes the Stage-II EM algorithm.

In every iteration of the algorithm in Stage-I, $\boldsymbol{\theta}_C^*$, and $\boldsymbol{\theta}_M^*$ are set to the correlation estimates from Stage-II in the previous iteration. In particular, $\boldsymbol{\theta}_P$ still can be estimated from (17) but the Stage-I E-step and Stage-I CM2-step need to be updated with the new $\boldsymbol{\theta}_C^*$, and $\boldsymbol{\theta}_M^*$. As demonstrated in Appendix C, each stage of the algorithm preserves the monotonic increase of the incomplete data likelihood function in every step and after every iteration between the two algorithm stages. Therefore, when iterating between the two stages until convergence, the self-consistency of the EM algorithm guarantees, under standard regularity conditions, that our two-stage estimates converge to the maximum likelihood estimates of all parameters. In particular, desirable large sample properties such as consistency, efficiency, and asymptotic normality hold for our two-stage EM estimator.

The gains in computational speed compared with the aforementioned nested EM algorithm are due to the fact that, in our empirical analysis, we find that just the first iteration using $\boldsymbol{\theta}_C^* = \mathbf{I}_K$, and ignoring the correlation structure in the Stage-I algorithm, is usually enough to obtain precise parameter estimates in finite samples of common sizes. Intuitively, neither the filtered G_t from Stage-I E-step, nor the estimates of $\boldsymbol{\theta}_D$ in the Stage-I CM2-step are strongly impacted by ignoring the correlations. The former, because it depends on the correlation matrix only through the shape of ellipsoids in (16), and the latter because of the aforementioned flat likelihood function problem with respect to the GIG parameters.

We emphasize that, even under the normality assumption, the efficiency and asymptotic normality of the two-stage procedure in Pelletier (2006) are not guaranteed. Instead, the author uses the two-stage estimates as initial values for the numerical optimizer of the full likelihood function, as suggested earlier in Pagan (1986), and achieves convergence in a few steps of a Newton-Raphson method.

In our setting, there are far too many parameters for sole use of generic optimization routines:

For example, with $K = 29$ assets and $N = 2$ regimes (as used below in our main empirical application), we have 960 free parameters. Without the ability to directly optimize the likelihood by conventional means (let alone ensure obtaining the global maximum of the log-likelihood), along with the fact that this (and all) models are misspecified with probability one, our paradigm needs to be judged on, first and foremost, its efficacy for forecasting, and, of second order concern but still important for practical purposes, the numeric reliability and speed of our estimation paradigm for a realistic number of assets K. With respect to the latter concern, our proposed algorithm converged for every moving window of all data sets considered, with an average speed of 120 seconds (for $K = 29$ assets, $N = 2$ regimes, and a window size of $T = 1000$), using an Intel i7 CPU at 4.00 GHz (and coding in Matlab 2014A). With respect to our primary concern of forecasting performance, we will see in the subsequent Section 4 that it outperforms all special cases, as well as the Gaussian-CCC and -DCC models.

4 Empirical Application

The main data set of our analysis consists of the 3923 daily returns of $K = 29$ components of the Dow Jones Industrial Index (DJ30) from June 1st 1999, to December 31st 2014. The constituents are based on the DJ30 composition in July 2015 with VISA being removed because the stock was not listed before March 18th 2008. To further our investigation, we additionally consider the top $K = 100$ firms by market capitalization of the Standard & Poor's 500 (SP500) index from January 2nd 1997 until December 31st 2014; and the $K = 19$ stocks from the Swiss Market Index (SMI) from January 5th 2004 until December 30th 2015, where Julius Baer Group is removed due to the company's restructuring and new market launch on October 1st 2009. Finally, we consider $K = 6$ foreign exchange (FX) rates against the Swiss Franc from July 1st 2004 until December 31st 2015. All sample periods are chosen as to include the global financial crisis for the out-of-sample analysis, and all data sets are obtained from Bloomberg Professional or from Wharton Research Data Services. Returns for each asset are computed as continuously compounded percentage returns, given by $y_{k,t} = 100 \log (p_{k,t}/p_{k,t-1})$, where $p_{k,t}$ is the split- and dividend-adjusted price of an asset k at time t.

In our analysis, we compare various models from the literature, namely the CCC model of Bollerslev (1990), the DCC model of Engle (2002), and the RSDC model of Pelletier (2006), all denoted with a prefix MN- for Multivariate Normal distribution of the innovations, against numerous MGHyp-based models proposed in this paper (MALap and MLap, both under CCC and RSDC correlation structures, and the same for NIG, SNIG, MAt and Mt). The scale term dynamics from (5) are used in all models except the MGHyp-IID models, which we also include in the forecasting analysis. This yields a total of 21 different models.

We start in Section 4.1 by comparing the in-sample fit of several models based on the full sample period of the DJ30 data set with 3923 observations. In Section 4.2, we consider the density forecasting performance across different models by using a large number of daily rolling windows to create one-day-ahead forecasts. The first practical application is given in Section 4.3, where the rolling window approach is employed to derive one-day-ahead forecasts of Value-at-Risk (VaR). Finally, in Section 4.4, we perform out-of-sample portfolio optimization based on the density forecasts of the competing models.

Summarizing the empirical results, the COMFORT-RSDC models yield a valuable prediction of which state will get realized next period, and deliver the best in-sample fit and one-step-ahead density forecasts among all considered models. In particular, we find (i) a large improvement moving from the normal to the MGHyp both under CCC and RSDC; (ii) the best performing subclass of MGHyp distributions is the Student-t; (iii) shrinkage estimation of the correlation matrices improves forecasting performance and is necessary to empower the full flexibility of the regime switching mechanism; (iv) the optimal level of correlation shrinkage strength is similar for all considered distributions, and forecasting performance is rather insensitive to the choice of this parameter above a certain threshold; (v) while asymmetry in the correlation structure via RSDC helps forecasting, this does not hold true for asymmetry in the MGHyp distribution: In almost all cases, the symmetric distributions outperform the asymmetric ones in forecasting. As expected, the accuracy of the VaR forecasts is largely determined by the heavy-tailed nature of the return distribution and not by the correlation dynamics. However, use of the latter results in lower VaR forecasts during turbulent times as compared to the CCC models while causing the same VaR violations. Such results could have meaningful economic benefits for financial institutions, e.g., potentially lower regulatory capital requirements. In the portfolio context, combining the MGHyp distribution with regime switching correlations improves risk-adjusted returns and reduces extreme losses.

4.1 In-Sample Performance

The in-sample fit of the various models under consideration for the entire DJ30 data set of 3923 daily returns is presented in Table 1. The second column compares the attained maximum of the log-likelihood function, whereas the fourth and fifth columns report the Akaike and the Bayesian information criteria (AIC and BIC, respectively) to account for model complexity. The table is sorted by AIC in descending order. When considering either AIC or BIC, we find that all non-Gaussian models are superior to their Gaussian counterparts. Noteworthy is that the three symmetric special cases of the MGHyp distribution considered here only have one additional parameter compared to the normal distribution, this being enough to add great flexibility by means of the mixture structure. Based on AIC, the Mt-RSDC model is best, and all COMFORT-RSDC models outperform the corresponding CCC models. The ranking changes under the BIC measure, which penalizes large numbers of parameters more heavily. In particular, under BIC, the COMFORT-CCC models are now preferred over their RSDC counterparts, due to having almost half as many parameters.

It is remarkable that all COMFORT-RSDC and -CCC models are superior to the Gaussian-RSDC model of Pelletier (2006) under any measure of in-sample fit. Moving from the normal distribution to any special case of the MGHyp distribution strongly enhances model fit; this finding being in line with those of McNeil et al. (2015) and Paolella and Polak (2015c). While in terms of likelihood and AIC, all RSDC models are consistently better than the corresponding CCC models, this improvement is largest for the Gaussian distribution. Once we use the more flexible MGHyp distribution, it becomes more difficult to improve in-sample fit by better modeling the correlations. In all COMFORT-CCC and -RSDC models, the best MGHyp-subclass is the Student-t family, whereas the NIG family is superior when assuming i.i.d. returns.

Model	*Likelihood*	*Parameters*	*AIC*	*BIC*
Mt-RSDC	-182404.3	931	**366670.6**	372512.1
Mt-RSDC-3	**-182042.5**	1341	366767.0	375181.3
MAt-RSDC	-182495.4	960	366910.9	372934.2
SNIG-RSDC	-182658.0	931	367177.9	373019.3
NIG-RSDC	-182723.4	960	367366.8	373390.2
MLap-RSDC	-182883.3	931	367628.7	373470.1
MALap-RSDC	-182873.7	960	367667.4	373690.8
Mt-CCC	-183382.3	523	367810.6	**371092.1**
MAt-CCC	-183354.6	552	367813.1	371276.6
SNIG-CCC	-183697.9	523	368441.8	371723.2
NIG-CCC	-183677.5	552	368459.0	371922.5
MLap-CCC	-183809.5	523	368665.1	371946.5
MALap-CCC	-183793.8	552	368691.6	372155.0
SNIG-IID	-186337.7	**465**	373605.3	374346.4
NIG-IID	-186329.8	494	373647.7	374434.9
MLap-IID	-186859.6	**465**	374649.3	375390.3
MALap-IID	-186851.7	494	374691.3	375478.6
MN-RSDC-3	-186799.5	1340	376279.1	384687.0
MN-RSDC	-187264.6	930	376389.2	382224.4
Mt-IID	-187944.8	**465**	376819.6	377560.7
MAt-IID	-187938.1	494	376864.1	377651.4
MN-CCC	-190673.4	522	382390.7	385665.9

Table 1: Comparison of the in-sample fit using 3923 daily returns of 29 companies of the DJ30 index, from 01.06.1999 to 31.12.2014

Finally, the in-sample model fit facilitates the choice of the number of regimes to optimally explain and forecast the dynamics of conditional correlations for the given data set. Theoretically most problematic is the case where the true number of regimes is $N = 1$, and one attempts to estimate a model with $N = 2$ or higher. This concern, in turn, is mitigated by the fact that, first, the entire model is–as with all models for financial asset returns–misspecified, and our paradigm should instead be viewed as an approximation, judged on its ability for efficacy in terms of out-of-sample predictive performance and practicality in terms of numerical optimization stability. Second, the improvement in terms of in-sample fit and out-of-sample performance (see below) when moving from $N = 1$ to $N = 2$ is very strong, suggesting that $N = 1$ is more misspecified than the $N = 2$ case. Moreover, the stylized facts of multivariate financial asset returns data over substantial periods of time strongly suggest movements in the conditional correlation, so that any model that allows for time-varying conditional correlations, whether it be DCC or our RSDC model, should be preferred to the simpler CCC setting.

To investigate the question whether $N = 3$ would be an even better choice, we fit the Gaussian-RSDC and (one special case of) the COMFORT-RSDC model also with three correlation regimes, abbreviated MN-RSDC-3 and Mt-RSDC-3, respectively, in Table 1. In both cases, as expected, the attained maximum-likelihood is higher under $N = 3$, but the AIC is improved only for the Gaussian-RSDC model and not for the COMFORT-RSDC model. Referring to the BIC, using $N = 3$ regimes in both RSDC models generates a poorer fit than using $N = 2$ or even $N = 1$. We conclude that there is no clear improvement in using $N = 3$ regimes, and thus focus on $N = 2$ in the following out-of-sample analysis.

4.2 Out-of-Sample Density Forecasting

Our main interest centers on the quality of one-step-ahead predictions of the multivariate probability density function of returns, as this is a crucial input for risk prediction and portfolio optimization. Hence the object of interest in this section is the entire probability distribution,

instead of a particular moment of it, or a tail risk measure. Furthermore, looking at out-of-sample predictions naturally overcomes the problem of data overfitting and makes the use of the AIC and BIC measures superfluous, because the performance of the model is evaluated at a data point that is not part of the training set. The avoidance of AIC and BIC is favorable for another reason; we will improve the out-of-sample performance of our models by means of different ways of shrinkage estimation (see below); however, as shrinkage estimation restricts the degrees of freedom in parameter estimation, it becomes unclear how to count parameters. Note that, in the previous section on in-sample fit measured by AIC and BIC, we did not make use of any shrinkage estimation methods.

Evaluating the quality of forecasts of entire probability distributions instead of estimates of, say, the mean, has gained in popularity both in the statistics literature, as well as in numerous applications, see Elliott and Timmermann (2008), Timmermann (2000), and Tay and Wallis (2000) for an overview of economic and financial applications, as well as Amisano and Giacomini (2007) for an associated statistical test. To assess the out-of-sample predictive ability of a model we use the normalized sum of the realized predictive log-likelihood (short RPLL), also called the predictive log-score, as proposed and discussed in Weigend and Shi (2000), Paolella and Polak (2015a), and Paolella (2015). It is given by

$$S_T\left(\mathcal{M}\right) = \frac{1}{T}\sum_{t=1}^{T} \pi_t\left(\mathcal{M}\right), \tag{29}$$

where

$$\pi_t\left(\mathcal{M}\right) = \log f^{\mathcal{M}}_{t+1|t}\left(\mathbf{Y}_{t+1} \mid \boldsymbol{\theta}\right), \tag{30}$$

$f^{\mathcal{M}}_{t+1|t}\left(\mathbf{Y}_{t+1} \mid \boldsymbol{\theta}\right)$ is the forecasted density of $\mathbf{Y}_{t+1}$ using information up to and including time t, and $\mathcal{M}$ indicates the model. This measure puts equal emphasize on the entire distribution, whereas using weighted log-likelihoods, one can put the focus on certain regions of the density, for instance the tails, as discussed in Amisano and Giacomini (2007) and Gneiting and Ranjan (2011). We abstain from using such weighting schemes, as they require to integrate over the multivariate density function, which, in high dimensions, is computationally rather costly.
For the RSDC models, the correlation matrix of next day's returns is stochastic because the latent regime follows a Markov chain. Hence, we extend the above definition of $\pi_t\left(\mathcal{M}\right)$ by replacing (30) with a conditional expectation analogue, i.e.,

$$\pi_t\left(\mathcal{M}\right) = \log \mathbb{E}\left[f^{\mathcal{M}}_{t+1|t}\left(\mathbf{Y}_{t+1} \mid \widehat{\boldsymbol{\theta}}\right) \mid \boldsymbol{\Phi}_t\right] = \log\left(\sum_{n=1}^{N} \xi_{n,t+1|t}\; f^{\mathcal{M}}_{t+1|t}\left(\mathbf{Y}_{t+1} \mid \Delta_{t+1} = n; \widehat{\boldsymbol{\theta}}\right)\right), \tag{31}$$

where $\xi_{n,t+1|t}$ is the nth element of the vector $\boldsymbol{\xi}_{t+1|t}$, given in (20).

Dow Jones 30 Data Set
Using the predictive log-likelihood (29), we compare the forecasting performance of the aforementioned models. We estimate each model on 2922 rolling windows of 1000 data points, using the same DJ30 data set as above.
In line with numerous forecasting applications that advocate the use of shrinkage to enhance out-of-sample forecasting performance, see, e.g., Diebold and Li (2006), we do so in the Stage-II

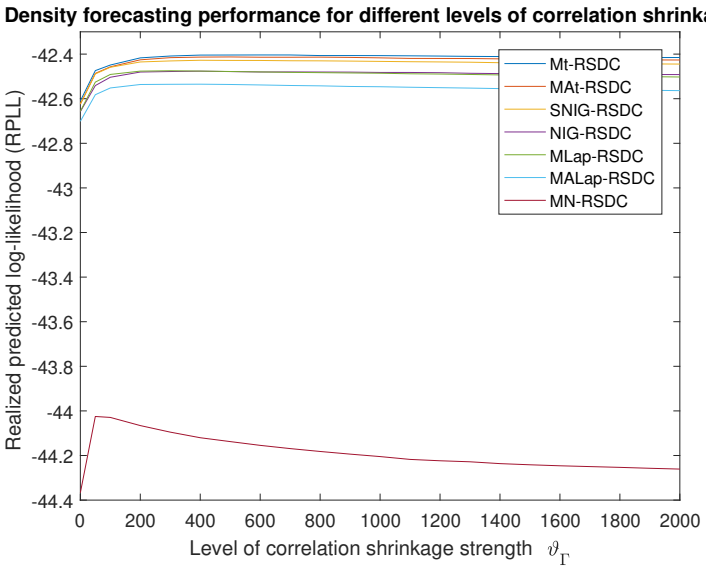

Figure 1: Realized predictive log-likelihood measures $S_T(\mathcal{M})$ for different models $\mathcal{M}$, plotted as a function of the correlation shrinkage strength parameter $0 \leq \vartheta_{\mathbf{\Gamma}} \leq 2000$, data set consisting of 29 stocks of DJ30, out-of-sample period from 23.05.2003 to 31.12.2014.

EM algorithm, where estimation of the regime-specific correlation matrices $\mathbf{\Gamma}_n$, $n = 1, \ldots, N$, is enhanced by the quasi-Bayesian shrinkage estimator (25). For the shrinkage targets B_n, $n = 1, \ldots, N$, in (25), we refer to the one-step estimated COMFORT-CCC model, by using the resulting CCC correlation matrix estimate $\widehat{\mathbf{\Gamma}}_{\text{CCC}}$ as our target. For simplicity, we use the same target $B_n = \widehat{\mathbf{\Gamma}}_{\text{CCC}}$ for all correlation matrices $\mathbf{\Gamma}_n$, but we impose different levels of shrinkage strength $a_n \geq 0$ for the various regimes. In the case $N = 2$ considered in this empirical section, we couple the shrinkage strengths using a hyper-parameter: we set $a_1 = \omega\vartheta_{\mathbf{\Gamma}}$ and $a_2 = \vartheta_{\mathbf{\Gamma}}/\omega$, with $\vartheta_{\mathbf{\Gamma}} \geq 0$ a scalar strength hyper-parameter and $\omega > 0$ a scalar parameter determining the level of similarity of both regimes. Less prior strength is assigned to the second regime correlation matrix, allowing more flexibility in the correlation matrix and potentially resulting in higher correlations during turbulent market periods. In a multivariate finite mixture model, Paolella (2015) uses the value $\omega = 2$. We do not optimize over this parameter, but instead choose a priori $\omega = 3$ to allow for even larger spreads in correlations between regimes and to accommodate the fact the we use the same target for both correlation matrices (which is not the case in Paolella, 2015). Other choices of shrinkage targets are of course feasible and might well be superior to our target, e.g., one could shrink towards the unit matrix or a matrix of ones, akin to the DECO model Engle and Kelly (2012). It is important to remark that choosing a good shrinkage target for the correlation matrices is non-trivial, as it increases the misspecification error, while a reduction of the estimation error is often hard to achieve. Furthermore, while one could optimize the choice of ω by choosing it such that the out-of-sample forecasting performance is maximized as a function of it, this could be deemed a form of backtest overfitting (see, e.g., Bailey et al., 2014, and Zhu et al., 2017).

The only parameter left to choose is the shrinkage strength hyper-parameter $\vartheta_{\mathbf{\Gamma}} \geq 0$. Our objective for the optimal value $\vartheta^*_{\mathbf{\Gamma}}$ is to maximize the out-of-sample density forecasting performance, measured by the RPLL, denoted $\pi_t(\mathcal{M})$ in (29). This is done in a computationally intense grid-

Rank	*Model*	*RPLL*	$\vartheta^*_{\mathbf{\Gamma}}$
1	Mt-RSDC	**-42.4038**	600
2	MAt-RSDC	-42.4127	500
3	SNIG-RSDC	-42.4276	400
4	MLap-RSDC	-42.4746	300
5	NIG-RSDC	-42.4765	400
6	MALap-RSDC	-42.5349	400
7	Mt-CCC	-42.6861	-
8	MAt-CCC	-42.7166	-
9	SNIG-CCC	-42.7251	-
10	MLap-CCC	-42.7601	-
11	NIG-CCC	-42.7617	-
12	MALap-CCC	-42.8010	-
13	SNIG-IID	-43.1385	-
14	NIG-IID	-43.1608	-
15	MLap-IID	-43.3109	-
16	MALap-IID	-43.3337	-
17	Mt-IID	-43.5238	-
18	MAt-IID	-43.5448	-
19	MN-RSDC	-44.0250	50
20	MN-DCC	-44.6677	-
21	MN-CCC	-44.6721	-

Table 2: Comparison of out-of-sample density prediction performance measured by realized predictive log-likelihood, data set consisting of 29 stocks of DJ30, out-of-sample period 23.05.2003 - 31.12.2014; RSDC models with optimal levels of correlation shrinkage $\vartheta^*_{\mathbf{\Gamma}}$

search, spanning a wide range of values $\vartheta_{\mathbf{\Gamma}} \geq 0$. Figure 1 presents the results by plotting the RPLL against the level of correlation shrinkage $\vartheta_{\mathbf{\Gamma}}$ for all RSDC models. We find that the optimal shrinkage strength $\vartheta^*_{\mathbf{\Gamma}}$ depends on the specific distribution but the difference in optimal strength between the MGHyp distributions is small (see also Table 2). Furthermore, the forecasting ability is quite robust to the choice of shrinkage strength for levels of $\vartheta_{\mathbf{\Gamma}} \geq 200$, as shown by the concave shape of the curves flattening out very slowly for higher levels of shrinkage. This allows the statistician to choose a priori a well-functioning, albeit not optimal, level of correlation shrinkage strength in order to avoid data snooping. The CCC structure constitutes the limiting case for $\vartheta_{\mathbf{\Gamma}} \to \infty$ only when the CCC correlation matrix is chosen for both shrinkage targets B_1 and B_2. In addition to improving density forecasts, correlation shrinkage has the pleasant effect of significantly reducing computational time of the Stage-II EM algorithm. For example, estimation time can be reduced by up to a factor of three when using shrinkage levels of $\vartheta_{\mathbf{\Gamma}} \geq 400$.
Using the optimal levels of correlation shrinkage $\vartheta^*_{\mathbf{\Gamma}}$ for all RSDC models according to Figure 1, we summarize the density prediction comparison in Table 2. For comparison we also include the i.i.d. case for all MGHyp distributions considered herein and the Gaussian-DCC model of Engle (2002). We observe that the biggest gain is achieved by replacing the normal distribution with any of the MGHyp distributions. Within the MGHyp distribution family, the Student-t subclass is best for the conditional models, i.e., for CCC and RSDC, and the NIG and Laplace subclasses show very similar performance. Most notably, all RSDC models deliver significantly better density forecasts than the corresponding CCC models. The improvement from the RSDC extension is largest for the Gaussian-based model but the Gaussian-RSDC model is vastly inferior to even the worst performing COMFORT-CCC model. Moreover, the weakest COMFORT-RSDC model still clearly outperforms the best COMFORT-CCC model, demonstrating the importance of dynamic correlation modeling for multivariate density forecasting. Even the MGHyp-based i.i.d. models are more potent for density forecasting than the Gaussian-CCC, -DCC and -RSDC models, underlining the utmost importance of non-Gaussian distributions for return modeling. Within the

i.i.d. models, the SNIG performs best and the Student-t worst, contrary to the conditional setting. This can be explained by the well-known fact that GARCH processes induce excess kurtosis, even when a thin-tailed innovation distribution is used. Hence the optimal tail thickness of the innovation process will differ between unconditional and conditional models.

Noteworthy also is that, even for the simple i.i.d. models, the elliptical MGHyp distributions all outperform the asymmetric variants in this out-of-sample analysis. This finding is investigated closer below but it seems that for out-of-sample forecasting in the MGHyp class, the simpler elliptical distributions are at least as good as those that allow for asymmetric margins.

Comparing the two Gaussian models with dynamic conditional correlations, we see that the Gaussian-RSDC is better suited for density forecasting than the Gaussian-DCC model, which performs virtually as poorly as the Gaussian-CCC due to its restrictive functional form of correlation dynamics.

Overall the best model is the symmetric Student-RSDC, in line with the in-sample results. Comparison of the COMFORT-CCC models in Table 2 with the left-end points in the plots in Figure 1 shows that, even without correlation shrinkage, the COMFORT-RSDC models are better than their CCC counterparts. Summarizing, the regime switching correlation dynamics improve the out-of-sample forecasts even without the quasi-Bayesian prior (25), but the predictive performance can be further improved with it; the latter result also holds for the Gaussian-RSDC model of Pelletier (2006), as seen from Figure 1. Hence as a byproduct of our work, we provide an easy method to improve the Gaussian-RSDC model considerably by using shrinkage estimation of the correlation matrices with the CCC correlation matrix as target.

Further parameter estimates that could potentially benefit from shrinkage estimation are the skewness parameters $\boldsymbol{\gamma} \neq \mathbf{0}$ of the non-elliptical cases of MGHyp distributions (abbreviated MAt, MALap and NIG herein). As observed above, for both in- and out-of-sample fit, the elliptical variants of the COMFORT-RSDC model perform better than the non-elliptical ones. The same result is found in the density forecasting analysis for the COMFORT-CCC and MGHyp-IID models. It seems particularly difficult to accurately estimate the asymmetry parameters, and shrinkage estimation could offer improvement, especially for the out-of-sample density forecasting performance. In our analysis we try three different shrinkage estimators for $\boldsymbol{\gamma}$. First, we use explicit linear shrinkage estimation by forming the convex combination of the parameter estimates from the Stage-I ECME algorithm and a fixed shrinkage target. Let $\boldsymbol{\gamma}_{\text{2S}}$ be the asymmetry estimate from the ECME algorithm. We define $\boldsymbol{\gamma}_{\text{SE1}} := (1 - \vartheta_{\boldsymbol{\gamma}})\boldsymbol{\gamma}_{\text{2S}} + \vartheta_{\boldsymbol{\gamma}}\boldsymbol{\gamma}_{\text{ST}}$, where $\vartheta_{\boldsymbol{\gamma}} \in [0, 1]$ is the shrinkage strength and $\boldsymbol{\gamma}_{\text{ST}}$ is the shrinkage target vector. The second and third approaches use implicit shrinkage via L^p-penalization of the Stage-I CM1-step likelihood function (17) with the penalty term $\vartheta_{\boldsymbol{\gamma}}||\boldsymbol{\gamma} - \boldsymbol{\gamma}_{\text{ST}}||_p$, where $\vartheta_{\boldsymbol{\gamma}} \geq 0$ is the penalty strength, $\boldsymbol{\gamma}_{\text{ST}}$ is the shrinkage target and $||\cdot||_p$ is the L^1-norm in the second and the L^2-norm in the third approach.

In all estimators, we shrink towards a vector of zeros, i.e. $\boldsymbol{\gamma}_{\text{ST}} = \mathbf{0}$, and find the optimal shrinkage strength $\vartheta^*_{\boldsymbol{\gamma}}$ in terms of the best forecasting performance by a computationally costly search on a specified grid of values for $\vartheta_{\boldsymbol{\gamma}}$. We find that the simple linear shrinkage estimator improves the density forecasting ability only slightly and less than the penalized likelihood approaches. Use of L^p-penalization additionally stabilizes and speeds up the Stage-I ECME algorithm by a factor of up to two. These favorable features are known from general regularization methods in numerical optimization, because the objective function (with possibly many local optima) is

smoothed by adding a convex penalty term. The quality of density prediction is similar for L^1- and L^2-penalization, but the L^2-norm produces smoother parameter estimates over the rolling windows and is thus favored.
However, and most importantly, none of the asymmetry parameter shrinkage approaches delivers better density forecasts than the corresponding elliptical models within a reasonable range of shrinkage strength. The elliptical case, corresponding to $\vartheta_{\boldsymbol{\gamma}} \to \infty$, performs best out-of-sample for all models under consideration. At least for the COMFORT-CCC, this contrasts the better in-sample fit of the asymmetric distributions (see Table 1) and hints towards in-sample over-fitting. The fact that the lower forecasting capacity cannot be overcome by various methods of shrinkage estimation supports the adequacy of elliptical MGHyp distributions for multivariate return modeling. To still profit from the acceleration effect of regularized optimization, we have used the L^2-penalized estimator with a low level of shrinkage strength ($\vartheta_{\boldsymbol{\gamma}} = 0.1$) for estimating all non-elliptical MGHyp models in this out-of-sample section.

Another object of interest are the out-of-sample forecasts of regime probabilities, i.e., the one-step-ahead forecasts of which regime will occur next. More precisely, we make a prediction at time t, based on a rolling window of the last 1000 observations, of the probability that the volatile market regime will occur at time $t+1$. Using the optimal level of correlation shrinkage, Figure 2 presents the out-of-sample forecasts $\boldsymbol{\xi}_{t+1|t}$ of the probabilities governing which state is going to realize next in the Gaussian-RSDC and the symmetric Student-RSDC model. The actual estimates $\hat{\boldsymbol{\xi}}_{t+1|t}$ are plotted as dots and a local regression smoother is added to facilitate interpretation. Comparing the two figures, we see that the forecasted state probabilities under the Gaussian-RSDC and the COMFORT-RSDC models are very similar, but the variation in forecasts is lower for the simpler Gaussian model. The major difference in this example is the different behavior of the competing models towards the end of the sample period, where the Gaussian-RSDC model predicts a resurgence of the crisis regime, whereas the COMFORT-RSDC model predicts a calming market.
In general, the timing with which the COMFORT-RSDC model forecasts changes in market regimes is closely linked to actual historical economic events in the US stock markets. In the right plot of Figure 2 we can identify seven time periods that match specific periods of market conditions. The first such period starts in mid-2003, lasts until summer 2005, and is characterized by a low probability of a crisis, due to the ongoing bull market during that time. From the end of 2005 until mid 2006, the probability of a crisis jumps up, reflecting an increasing nervousness in the market and first signs of a looming financial crisis. From fall 2006 until mid 2007, market volatility falls again. Then, from mid 2007 (when Bear Stearns declared bankruptcy) until mid 2009, the global financial crisis manifests itself in a strongly dominant and stable volatile regime. The probability of the crisis ending tomorrow is (except for the two isolated dots in mid 2008), always estimated as less than 20% during these two years.
The next time frame is spanned from fall 2009 until the end of 2011, and is characterized by an increase in the probability of a recovery and a less clear separation of regimes. The high uncertainty of this period stems from the recovery in the financial markets, whereas at the same time the real economy was still struggling heavily, casting doubt on the further performance of equity markets. During the years 2011 and 2012, the regime forecasts predict a resurgence of the financial crisis.

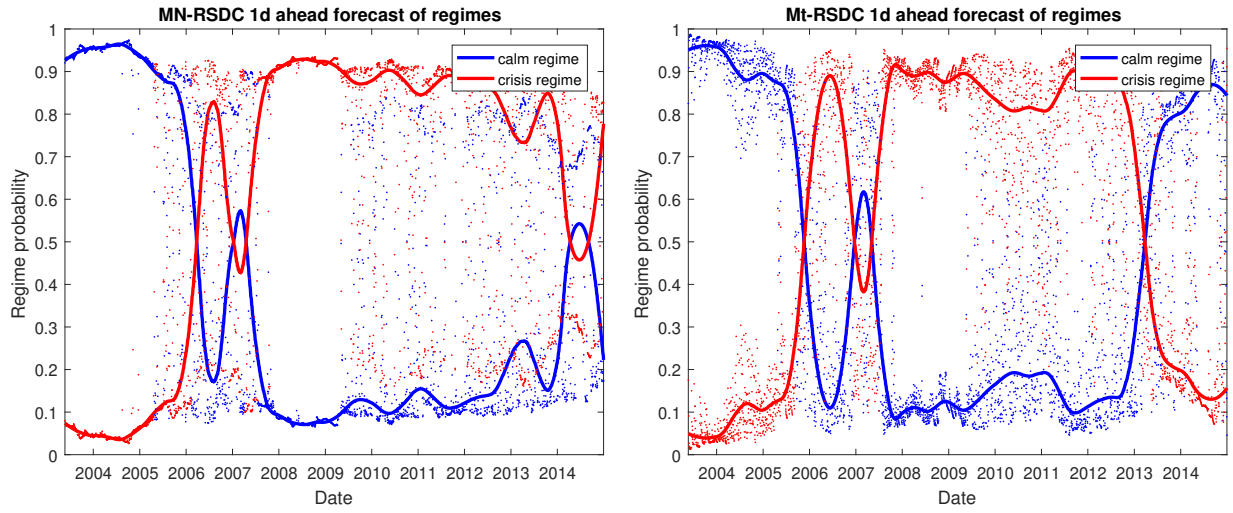

Figure 2: *Left*: One-day-ahead forecasts of regimes of the Gaussian-RSDC model for the data set consisting of 29 stocks of the DJ30, out-of-sample period from 23.05.2003 - 31.12.2014. The plot shows the forecasted probability of the volatile regime under optimal correlation shrinkage $\vartheta^*_{\mathbf{\Gamma}} = 50$. *Right*: Same plot for the symmetric Student-RSDC model with optimal correlation shrinkage $\vartheta^*_{\mathbf{\Gamma}} = 600$

This fear was spurred by the debt crisis in Europe, which reached its peak when Ireland had to be bailed out in November 2010 and Portugal had to be rescued in April 2011. The struggle to keep Greece solvent and remaining within the euro zone went on until fall 2012, when the ECB signaled to undertake large debt purchases, and the euro zone members agreed on a joint liability for Greece's debt.
Finally, starting in spring 2013, the calm regime dominates again, because of the strong bull markets fueled by central bank money worldwide and the fading of the euro crisis. Interestingly, the Gaussian-RSDC model does not capture this bull market and instead predicts a resurgence of the volatile regime that did not take place. The corresponding plots for other MGHyp-distributions used in the COMFORT-RSDC model look very similar and are available upon request.

The time evolution of the Markov transition matrix $\Pi = (\pi_{i,j})_{i,j=1,2}$, governing the regime switching correlations, is plotted in Figure 3 over the 2922 rolling windows of our out-of-sample analysis. The blue line describes the probability to remain in the calm regime when the process is in the calm regime, and the red line shows the probability to remain in the crisis regime once we are in a crisis. Both RSDC models show considerable time variation of the transition probabilities, indicating that use of a non-homogeneous Markov chain with time-dependent transition probabilities might be appropriate (which is at the moment left open for future research). The most striking difference between the Gaussian and the heavy-tailed RSDC models is that the latter returns to probabilities of 80% and higher of remaining in a given regime after the year 2013, making the calm regime more persistent in the years 2013-2014. On the contrary, the Gaussian-RSDC model

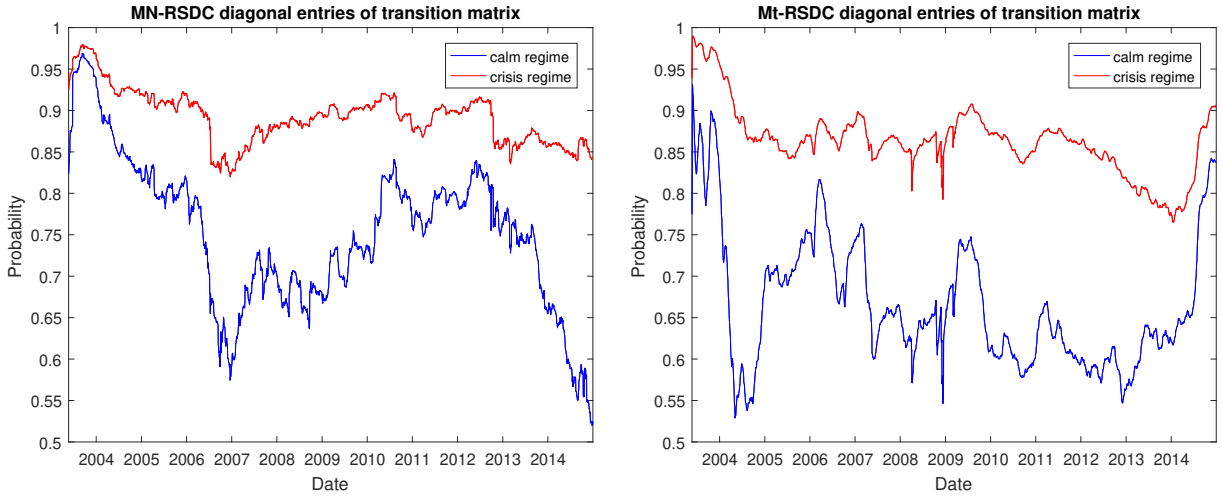

Figure 3: *Left*: Diagonal entries of the transition matrix Π in the Gaussian-RSDC model plotted over the rolling windows. *Right*: Same plot for the symmetric Student-RSDC model

does not revert back to pre-crisis levels of transition probabilities, having the effect that recovery of the calm regime is not persistent, as seen in Figure 2.

Referring to the discussion of the optimal number of regimes in the RSDC models, for completeness, we compare the forecasting performance of the Gaussian-RSDC and COMFORT-RSDC models with $N = 2$ and $N = 3$ under different levels of correlation shrinkage. We find that three regimes are able to forecast equally well, but not better, as two regimes, with the important drawback of $N = 3$ being that one has to strongly increase the correlation shrinkage strength. Hence, while we use more regimes, these must be forced to be more similar to each other. As $N = 3$ does not increase the forecasting performance, but requires heavier shrinkage and considerably more computation time, we again conclude that the $N = 2$ case is sufficient. In fact, the improved density forecasts, when the number of free parameters in the case $N = 2$ is restricted by means of shrinkage estimation, indicate that the model complexity under $N = 2$ is sufficiently high.

Subperiods of the Dow Jones Data Set

We want to evaluate the extent to which the superior forecasting performance of the COMFORT-RSDC model is attributed to the specific sample period. We perform a robustness check by considering two sub-periods of our data: the crisis years 2008 and 2009; and the period spanning from 2008 until 2014. We do not repeat the search for optimal correlation shrinkage levels but instead use the ones found above for the entire data period. In Table 3 we summarize the out-of-sample density forecasting performances on both subsets. During the acute financial crisis, we observe for most distributions that the CCC structure is very slightly advantageous over the RSDC specification; only for the heavily misspecified Gaussian distribution does regime switching improve density forecasts. The underlying reason for the mildly inferior performance of the RSDC models in this period is that during these years both the Gaussian- and the symmetric Student-

Rank	*Model*	*RPLL*	$\vartheta_{\mathbf{\Gamma}}$
1	MAt-CCC	**-54.5732**	-
2	NIG-CCC	-54.5802	-
3	Mt-CCC	-54.5902	-
4	SNIG-CCC	-54.5923	-
5	MLap-RSDC	-54.5937	300
6	SNIG-RSDC	-54.6005	400
7	MALap-CCC	-54.6096	-
8	MALap-RSDC	-54.6122	400
9	NIG-RSDC	-54.6125	400
10	MLap-CCC	-54.6132	-
11	Mt-RSDC	-54.6185	600
12	MAt-RSDC	-54.6273	500
13	MN-RSDC	-55.9805	50
14	MN-CCC	-56.2814	-

Rank	*Model*	*RPLL*	$\vartheta_{\mathbf{\Gamma}}$
1	Mt-RSDC	**-42.7503**	600
2	MAt-RSDC	-42.7610	500
3	SNIG-RSDC	-42.7749	400
4	MLap-RSDC	-42.8179	300
5	NIG-RSDC	-42.8460	400
6	MALap-RSDC	-42.8909	400
7	Mt-CCC	-43.1507	-
8	SNIG-CCC	-43.1688	-
9	MAt-CCC	-43.1703	-
10	NIG-CCC	-43.1943	-
11	MLap-CCC	-43.1968	-
12	MALap-CCC	-43.2399	-
13	MN-RSDC	-44.3240	50
14	MN-CCC	-44.8183	-

Table 3: *Left:* Comparison of out-of-sample density prediction performance in the financial crisis subperiod from 02.01.2008 - 31.12.2009 (RSDC models with same levels of correlation shrinkage $\vartheta_{\mathbf{\Gamma}}$ as in the previous analysis). *Right:* Same exercise for the sample period from 02.01.2008 - 31.12.2014.

RSDC models predict with high probability that the volatile regime will persist. In order words, use of two regimes is superfluous during the acute financial crisis. A second finding is that for the years 2008 and 2009, the non-elliptical COMFORT-CCC models perform better than the elliptical ones. The negative skewness is very pronounced in this period due to many large negative returns, so that in the COMFORT-CCC model, the misspecification under $\boldsymbol{\gamma} \neq \mathbf{0}$ outweighs the estimation error when the asymmetry parameters are estimated (via shrinkage estimation as described above). When considering the sample period from 02.01.2008 until 31.12.2014, as reported in the right side of Table 3, again all RSDC models outperform their CCC counterparts. Moreover, on the longer data set, the elliptical models beat the asymmetric ones, even in the CCC case.
We conclude that the added flexibility of Markov switching correlations enhances out-of-sample predictions only if the data is diverse enough to justify the use of several regimes. This result is anticipated and reasonable, given the additional statistical estimation errors of a more complex model. Moreover, we infer that using an MGHyp distribution with asymmetric marginals is more adequate in times of extraordinary financial distress but over longer time periods the elliptical MGHyp distributions are preferable because the asymmetry is much less pronounced in regular market conditions and the accurate estimation of these additional parameters is problematic.

Further Data Sets
To further investigate the flexibility and superior forecasting performance of the COMFORT-RSDC model, we perform out-of-sample density forecasting on three additional asset return data sets. Due to computational limitations we renounce to search for the optimal correlation shrinkage levels and instead make an a priori choice according to scheme presented in Table 12 in Appendix D. Using fixed tuning parameters has the additional advantage of preventing the statistician from data-snooping the optimal parameters and enables us to demonstrate the model performance in a more conservative and realistic setting.
The first data set consists of the $K = 100$ largest companies of the Standard & Poor's 500 (SP500) index, as measured by market capitalization on January 2nd 1997, and ranges from this

Rank	*Model*	*RPLL*
1	Mt-RSDC	**-149.8447**
2	MAt-RSDC	-149.8469
3	SNIG-RSDC	-149.8865
4	NIG-RSDC	-149.9116
5	MLap-RSDC	-149.9324
6	MALap-RSDC	-149.9849
7	Mt-CCC	-150.4813
8	MAt-CCC	-150.4843
9	MLap-CCC	-150.5368
10	SNIG-CCC	-150.5565
11	MALap-CCC	-150.5822
12	NIG-CCC	-150.6119
13	MN-RSDC	-155.2882
14	MN-CCC	-155.6347

Rank	*Model*	*RPLL*
1	Mt-RSDC	**-29.8699**
2	MAt-RSDC	-29.8737
3	SNIG-RSDC	-29.8901
4	NIG-RSDC	-29.8905
5	MLap-RSDC	-29.9158
6	MALap-RSDC	-29.9194
7	Mt-CCC	-29.9537
8	MAt-CCC	-29.9590
9	SNIG-CCC	-29.9914
10	NIG-CCC	-30.0061
11	MLap-CCC	-30.0484
12	MALap-CCC	-30.0803
13	MN-RSDC	-30.9762
14	MN-CCC	-31.3902

Rank	*Model*	*RPLL*
1	MLap-RSDC	-0.9312
2	MLap-CCC	-0.9591
3	SNIG-RSDC	-1.0845
4	Mt-RSDC	-1.1045
5	Mt-CCC	-1.1598
6	SNIG-CCC	-1.1610
7	MN-RSDC	-1.2833
8	MN-CCC	-1.5556

Table 4: *Left*: Comparison of density prediction performance on the top 100 market capitalization stocks of SP500 from 15.12.2004 - 31.12.2014 (RSDC models with fixed a priori chosen level of correlation shrinkage). *Middle:* Same analysis for 19 stocks of the SMI from 24.01.2008 - 30.12.2015. *Right:* Same analysis for six foreign exchange rates against the Swiss franc, out-of-sample period 26.07.2007 - 31.12.2015.

day until December 31st 2014, equaling 4529 observations. We use a rolling window of 2000 daily observations to accommodate the larger size of the correlation matrix, hence the out-of-sample period starts on December 15th 2004. The levels of correlation shrinkage are fixed to $\vartheta_\Gamma = 1000$ for all RSDC models except for the Gaussian-RSDC which uses $\vartheta_\Gamma = 200$. The realized predictive log-likelihood values are reported on the left of Table 4. In line with the findings on the DJ30 data, the Gaussian-CCC model performs worst, followed by the Gaussian-RSDC. Changing from Gaussian to MGHyp innovations vastly improves density forecasts, while within the MGHyp class, the best distribution is again the symmetric Student-t, for both the CCC and RSDC correlation structures. All RSDC models markedly improve upon the CCC analogues, to the extent that the worst COMFORT-RSDC model is considerably better than the best CCC one.

Our next data set consists of $K = 19$ stocks of the Swiss Market Index (SMI) that were listed from January 5th 2004 until December 30th 2015 giving a sample of 2946 daily observations. Here we use $\vartheta_\Gamma = 250$ for all non-Gaussian COMFORT-RSDC models and $\vartheta_\Gamma = 25$ for the Gaussian-RSDC, along with a rolling window of 1000 days. The out-of-sample period starts on January 24th 2008, amid the early phase of the financial crisis. Completely in accordance with the main findings on the previous data sets, all COMFORT-RSDC models beat their CCC rivals; the Student-t is the best and the Gaussian is the worst distributional assumption; and all symmetric variants of the MGHyp distributions perform better than the asymmetric versions, again for both the RSDC and the CCC specifications.

The third data set contains six series of daily FX returns. The currencies are euro, Australian dollar, Japanese yen, pound sterling, Hong Kong dollar and US dollar against the Swiss franc, from July 1st 2004 until December 31st 2015. We use the same shrinkage strength as for the SMI data set and a window size of 800 observations, effectively starting the out-of-sample forecasts on July 26th 2007. We observe that this data set is characterized by extreme movements in the asymmetry parameter estimates over the rolling windows; they attain large negative values on the first several hundred rolling windows, jump upwards to the positive range for a short period of time, and then stay close to zero for the remaining, largest part of the sample period. This causes problems in

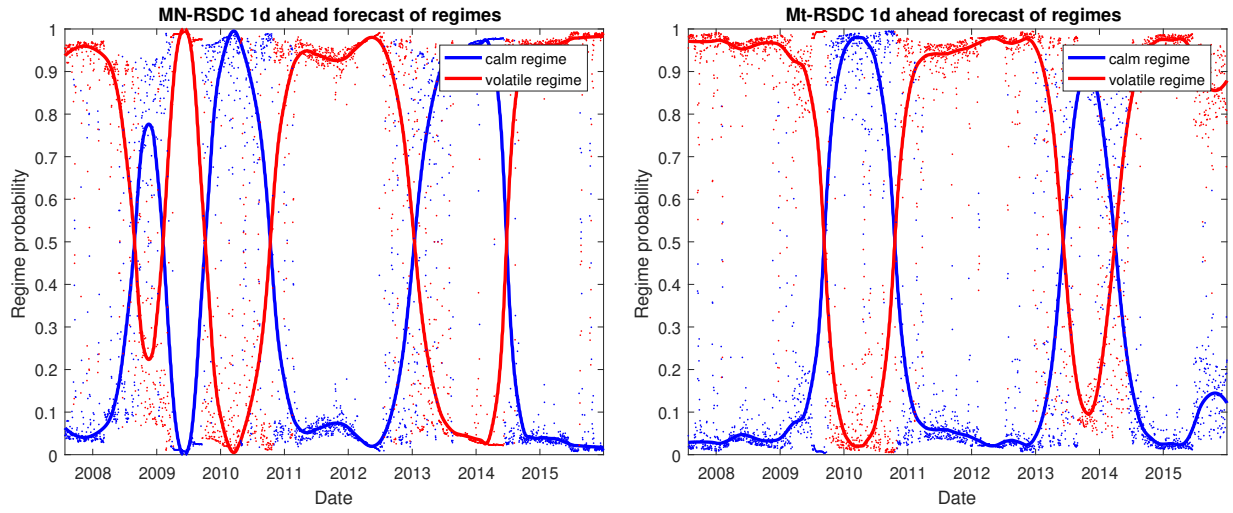

Figure 4: *Left*: Out-of-sample forecasts of regime probabilities of the volatile regime of the Gaussian-RSDC model on the FX data with a priori chosen level of correlation shrinkage, *Right*: same plot for the symmetric Student-RSDC model

the optimization routine used in Stage-I of the ECME algorithm for the non-elliptical MGHyp distributions. For this reason we only consider elliptical distributions in this analysis and point to the additional advantage of elliptical MGHyp distributions to be numerically less challenging to estimate. An interesting result, reported on the right side of Table 4, is that the symmetric Laplace distribution performs so well for the FX data that both the MLap-CCC and MLap-RSDC beat all competitors. Having thinner tails than the Student-t and NIG distribution the Laplace distribution offers a better statistical fit because FX returns have smaller excess kurtosis than stock returns. Nevertheless, the exponentially decreasing tails of the Gaussian distribution are so misspecified that both the CCC and RSDC model are vastly inferior to all MGHyp-based models. Most importantly, each RSDC model outperforms the corresponding CCC model for density forecasting on this data set, also. Again, the gain from using Markov switching correlations is by far largest for the Gaussian distribution, in line with previous results. The one-day-ahead forecasts of regime probabilities of the Gaussian-RSDC and the COMFORT-RSDC model are plotted in Figure 4, with the major difference being a short resurgence of the calm regime in the Gaussian-RSDC that is not present in the COMFORT-RSDC model.

The main findings from this out-of-sample density forecasting exercise support the conclusion that the best forecasts are achieved by combining the distributional flexibility of the MGHyp family with the correlation dynamics of a Markov regime switching model, along with judicious use of shrinkage estimation. The superior performance of the COMFORT model is demonstrated to hold also under an a priori, non-optimal shrinkage strength and on data sets of various dimensionality and asset classes.

4.3 Value-at-Risk Forecasting and Backtesting

We now apply the new COMFORT-RSDC model to forecasting risk measures, in particular Value-at-Risk (VaR). One major use of GARCH models is to produce concise short-term forecasts of tail risk measures; Santos et al. (2013) compares univariate and multivariate models for this task and favors the multivariate approach. Concentrating on the industry-wide standard risk measure of VaR, we analyze the quality of daily VaR forecasts based on the conditional density forecasts derived from the various models. The forecast quality is evaluated by means of the ubiquitous Christoffersen (1998) backtest, which is still the industry standard despite the development of newer backtest procedures, such as Christoffersen et al. (2001), Christoffersen and Pelletier (2004), Haas (2005), and recently Pelletier and Wei (2016).

Much of the earlier literature on VaR forecasting made use of univariate models applied to the returns induced by a given portfolio strategy or, quite often, just to a stock market index, which serves as a typical investment vehicle (by way of ETFs), as well as a typical return series for testing and comparing the adequacy of several VaR prediction methods; see, e.g., Kuester et al. (2006), Santos et al. (2013), Pelletier and Wei (2016), Slim et al. (2017), and the references therein. More recently, non-Gaussian multivariate models have also been used; see, e.g., Santos et al. (2013), who show that, among the various univariate and multivariate models they considered, the latter tended to perform better; while Härdle et al. (2015) propose and study a copula-based Markov switching multivariate model and illustrate its favorable performance with VaR prediction.

Herein, we wish to compare the VaR prediction performance of the newly proposed COMFORT-RSDC model to its various special cases, such as the CCC and Gausian-based models. Although it is well-known that unconditional models are not suitable for risk measure forecasting, for completeness we include the SNIG-IID model, which was the best i.i.d. model for density forecasting. Our goal is not a horse race among all competing VaR prediction methods, and thus is less ambitious than Kuester et al. (2006) and Santos et al. (2013), as our emphasis is on the study and efficacy of a new model. Anticipating the results discussed in more detail below, we find that the largest gains in VaR forecasting performance are due to the heavier tails of the MGHyp distribution, irrespective of the correlation structure used. While COMFORT-RSDC does not generally improve the outcome of the Christoffersen (1998) VaR backtest compared to COMFORT-CCC, it is economically more frugal by producing lower VaR forecasts during periods of large market volatility.

We restrict ourselves to the data set of 29 stocks from the DJ30, ranging from 01.06.1999 to 31.12.2014. Using a rolling window size of 1000 observations, the out-of-sample period starts on 23.03.2003. Analysis of the other data sets considered earlier leads to similar findings; further results are available on request.

For the RSDC models, we do not make use of correlation shrinkage estimation discussed in Section 4.2 because we find this to reduce the quality of VaR forecasts at all levels of shrinkage strength. This does not contradict the improved density forecasts under correlation shrinkage because VaR forecasts concentrate solely on the left tail of the univariate portfolio return distribution, while the predicted log-likelihood measure considers the full probability distribution of the multivariate asset returns.

The Christoffersen (1998) backtest is performed for 95%, 99%, 99.5%, and 99.9% one-step-ahead

daily VaR forecasts of the equally weighted $(1/K)$ portfolio. We test such high quantiles because of our short forecasting horizon of just one day. The 95% VaR corresponds to having on average one VaR exceedance every 20 trading days, which is certainly too frequent, for example, for a bank. The 99.5% and 99.9% VaR are more relevant as they correspond to one exceedance per 200 and 1000 days, respectively, on average. The backtesting results are presented in Tables 5 - 8. Each table is sorted by the p-values of the conditional coverage test, as this is a critical measure, e.g., to fulfill regulatory requirements. The p-values of the independence, unconditional coverage, and conditional coverage tests are calculated from the asymptotic chi-square distribution with one, one, and two degrees of freedom, respectively. This asymptotic distribution is a good approximation here because we use a fairly large sample size of 2922 out-of-sample VaR forecasts.
For the 95% VaR, the MALap- and MLap-CCC models perform best, while all models under consideration, except the SNIG-IID, pass the backtest on any conventional confidence level. The RSDC specification performs slightly worse than CCC but all RSDC models still easily pass the test. Even for the 95% VaR, the Gaussian assumption is not recommended as most MGHyp-based models perform better; nevertheless all three Gaussian models pass the test easily and perform identically well irrespective of the correlation structure.
The superiority of the non-Gaussian models becomes much more striking for 99% VaR forecasts. The Gaussian models, as well as the SNIG-IID, drastically fail the backtest with p-values of the conditional coverage test close to zero due to the vast number of violations. The three Gaussian models deliver identical backtest results because they produce the exact same sequence of VaR violations. In contrast, all COMFORT-CCC and -RSDC models pass the VaR backtest. Overall, the MLap-RSDC, MAt-CCC, MAt-RSDC, and Mt-CCC models perform best.
For 99.5% and 99.9% VaR, the findings are very similar. The Gaussian and the the SNIG-IID models provide extremely poor VaR forecasts and are thus strongly rejected. Again all COMFORT-CCC and -RSDC models easily pass the backtest. Several models perform equally well, among them being different cases of the MGHyp distribution and both the CCC and RSDC correlation structures.
It is not surprising that the i.i.d. assumption is worst in the independence test for all quantiles; even for high quantiles, the Gaussian-CCC, -DCC and -RSDC models perform better in this respect because of the sensitivity of the GARCH based VaR forecasts to increases in market volatility. Despite its excess kurtosis, the SNIG-IID model fails the backtest for all quantiles considered herein. We find the same for all other MGHyp-based i.i.d. models. We conclude, which is nothing new, that any reasonable model for risk forecasts has to account for volatility clustering, and therefore the i.i.d. assumption should be avoided irrespective of the probability distribution used. The predominance of conditional models, especially those that feature the stylized fact of heavy-tailed financial returns, for risk measure forecasting has been noted, amongst others, in Christoffersen et al. (2001), Kuester et al. (2006), and Righi and Ceretta (2015).
Summarizing the VaR backtest, we find that only those models that combine the flexible, semi-heavytailed MGHyp class with an MGARCH structure to account for volatility clustering pass the backtest for all quantiles. Overall, both the COMFORT-CCC and -RSDC class deliver superb forecasts of daily VaR, and the MAt-RSDC, Mt-CCC, and MALap-RSDC models appear to be particularly powerful across all quantiles.
Based on the backtest results in Tables 5 - 8, there is no clear reason to prefer the COMFORT-

95% **VaR**								
Model	*Stat. CC*	*p-value CC*	*Stat. UC*	*p-value UC*	*Stat. IND*	*p-value IND*	*Violations*	*Coverage Rate*
MALap-CCC	**0.2414**	**0.8863**	0.0095	0.9222	0.2318	0.6302	145	0.0496
MLap-CCC	0.2697	0.8738	**0.0002**	**0.9898**	0.2696	0.6036	**146**	**0.0499**
NIG-CCC	0.4384	0.8032	0.3740	0.5408	0.0643	0.7998	139	0.0476
MLap-RSDC	0.4575	0.7955	0.0581	0.8095	0.3994	0.5274	149	0.0510
NIG-RSDC	0.7434	0.6896	0.6844	0.4081	0.0590	0.8080	156	0.0534
SNIG-CCC	0.7781	0.6777	0.7589	0.3837	**0.0192**	**0.8897**	136	0.0465
MAt-CCC	0.7972	0.6713	0.2434	0.6217	0.5537	0.4568	152	0.0520
Mt-CCC	0.9436	0.6239	0.3331	0.5639	0.6106	0.4346	153	0.0523
SNIG-RSDC	1.0385	0.5950	0.6844	0.4081	0.3541	0.5518	156	0.0534
MAt-RSDC	1.3825	0.5010	1.1576	0.2820	0.2248	0.6354	159	0.0544
MN-CCC	1.5109	0.4698	0.1059	0.7449	1.4050	0.2359	150	0.0513
MN-DCC	1.5109	0.4698	0.1059	0.7449	1.4050	0.2359	150	0.0513
MN-RSDC	1.5109	0.4698	0.1059	0.7449	1.4050	0.2359	150	0.0513
MALap-RSDC	1.5620	0.4579	0.6844	0.4081	0.8776	0.3489	156	0.0534
Mt-RSDC	2.3981	0.3015	1.9745	0.1600	0.4236	0.5152	163	0.0558
SNIG-IID	13.7050	0.0011	3.0613	0.0802	10.6437	0.0011	126	0.0431

Table 5: Backtest results of 95% daily VaR forecasts of equally weighted portfolio, DJ30 data set with out-of-sample period spanning 23.05.2003 - 31.12.2014, rolling window size of 1000, RSDC models without correlation shrinkage.

99% **VaR**								
Model	*Stat. CC*	*p-value CC*	*Stat. UC*	*p-value UC*	*Stat. IND*	*p-value IND*	*Violations*	*Coverage Rate*
MLap-RSDC	**0.6427**	**0.7252**	**0.0203**	**0.8867**	0.6224	0.4302	**30**	**0.0103**
MAt-CCC	**0.6427**	**0.7252**	**0.0203**	**0.8867**	0.6224	0.4302	**30**	**0.0103**
MAt-RSDC	**0.6427**	**0.7252**	**0.0203**	**0.8867**	0.6224	0.4302	**30**	**0.0103**
Mt-CCC	**0.6427**	**0.7252**	**0.0203**	**0.8867**	0.6224	0.4302	**30**	**0.0103**
MLap-CCC	0.7710	0.6801	0.1062	0.7446	0.6648	0.4149	31	0.0106
NIG-CCC	0.8413	0.6566	0.3745	0.5406	0.4669	0.4944	26	0.0089
MALap-CCC	0.9659	0.6170	0.2572	0.6120	0.7087	0.3999	32	0.0109
SNIG-CCC	1.0817	0.5823	0.6502	0.4200	**0.4315**	**0.5113**	25	0.0086
SNIG-RSDC	1.9308	0.3808	1.0822	0.2982	0.8487	0.3569	35	0.0120
MALap-RSDC	2.3733	0.3052	1.4751	0.2245	0.8982	0.3433	36	0.0123
Mt-RSDC	2.3733	0.3052	1.4751	0.2245	0.8982	0.3433	36	0.0123
NIG-RSDC	3.4297	0.1800	2.4283	0.1192	1.0014	0.3170	38	0.0130
SNIG-IID	15.2984	0.0005	7.3781	0.0066	7.9203	0.0049	45	0.0154
MN-CCC	22.2573	0.0000	22.2357	0.0000	0.0216	0.8831	58	0.0198
MN-DCC	22.2573	0.0000	22.2357	0.0000	0.0216	0.8831	58	0.0198
MN-RSDC	22.2573	0.0000	22.2357	0.0000	0.0216	0.8831	58	0.0198

Table 6: Same as Table 5 but for 99% daily VaR forecasts

99.5% **VaR**								
Model	*Stat. CC*	*p-value CC*	*Stat. UC*	*p-value UC*	*Stat. IND*	*p-value IND*	*Violations*	*Coverage Rate*
NIG-CCC	**0.1612**	**0.9226**	0.0264	0.8710	**0.1348**	**0.7135**	14	0.0048
MAt-RSDC	**0.1612**	**0.9226**	0.0264	0.8710	**0.1348**	**0.7135**	14	0.0048
MLap-CCC	**0.1612**	**0.9226**	0.0264	0.8710	**0.1348**	**0.7135**	14	0.0048
MLap-RSDC	**0.1612**	**0.9226**	0.0264	0.8710	**0.1348**	**0.7135**	14	0.0048
SNIG-CCC	**0.1612**	**0.9226**	0.0264	0.8710	**0.1348**	**0.7135**	14	0.0048
MALap-CCC	0.1649	0.9209	**0.0101**	**0.9199**	0.1548	0.6940	**15**	**0.0051**
MAt-CCC	0.1649	0.9209	**0.0101**	**0.9199**	0.1548	0.6940	**15**	**0.0051**
Mt-CCC	0.1649	0.9209	**0.0101**	**0.9199**	0.1548	0.6940	**15**	**0.0051**
Mt-RSDC	0.9567	0.6198	0.7336	0.3917	0.2231	0.6367	18	0.0062
SNIG-RSDC	2.0630	0.3565	1.7873	0.1813	0.2757	0.5996	20	0.0068
NIG-RSDC	2.7720	0.2501	2.4679	0.1162	0.3040	0.5814	21	0.0072
MALap-RSDC	4.4777	0.1066	4.1127	0.0426	0.3650	0.5458	23	0.0079
SNIG-IID	16.7572	0.0002	12.4605	0.0004	4.2967	0.0382	30	0.0103
MN-CCC	33.2220	0.0000	32.0550	0.0000	1.1670	0.2800	41	0.0140
MN-DCC	33.2220	0.0000	32.0550	0.0000	1.1670	0.2800	41	0.0140
MN-RSDC	33.2220	0.0000	32.0550	0.0000	1.1670	0.2800	41	0.0140

Table 7: Same table for 99.5% daily VaR forecasts

99.9% **VaR**								
Model	*Stat. CC*	*p-value CC*	*Stat. UC*	*p-value UC*	*Stat. IND*	*p-value IND*	*Violations*	*Coverage Rate*
NIG-CCC	**0.0082**	**0.9959**	**0.0020**	**0.9642**	0.0062	0.9374	**3**	**0.0010**
Mt-CCC	**0.0082**	**0.9959**	**0.0020**	**0.9642**	0.0062	0.9374	**3**	**0.0010**
SNIG-CCC	**0.0082**	**0.9959**	**0.0020**	**0.9642**	0.0062	0.9374	**3**	**0.0010**
MAt-RSDC	0.3312	0.8474	0.3284	0.5666	**0.0027**	**0.9583**	2	0.0007
Mt-RSDC	0.3312	0.8474	0.3284	0.5666	**0.0027**	**0.9583**	2	0.0007
MAt-CCC	0.3668	0.8324	0.3559	0.5508	0.0110	0.9166	4	0.0014
MALap-CCC	1.2329	0.5399	1.2158	0.2702	0.0171	0.8958	5	0.0017
MLap-CCC	1.2329	0.5399	1.2158	0.2702	0.0171	0.2702	5	0.0017
MLap-RSDC	1.2329	0.5399	1.2158	0.2702	0.0171	0.8958	5	0.0017
NIG-RSDC	1.2329	0.5399	1.2158	0.2702	0.0171	0.8958	5	0.0017
SNIG-RSDC	1.2329	0.5399	1.2158	0.2702	0.0171	0.8958	5	0.0017
MALap-RSDC	2.5037	0.2860	2.4790	0.1154	0.0247	0.8751	6	0.0021
SNIG-IID	29.5862	0.0000	18.6816	0.0000	10.9045	0.0000	13	0.0044
MN-CCC	31.9734	0.0000	31.7745	0.0000	0.1990	0.6556	17	0.0058
MN-DCC	31.9734	0.0000	31.7745	0.0000	0.1990	0.6556	17	0.0058
MN-RSDC	31.9734	0.0000	31.7745	0.0000	0.1990	0.6556	17	0.0058

Table 8: Same table for 99.9% daily VaR forecasts

RSDC model over the simpler CCC structure. All COMFORT-RSDC and CCC models pass the backtest easily but the best model can either be of the RSDC or the CCC type, depending on the VaR level and most likely on the data set. However, the Christoffersen (1998) backtest considers merely the number and timing of violations but does not explicitly account for the level of VaR forecasts. Two competing models that produce the same series of VaR violations can have different VaR forecasts and hence could potentially imply different regulatory capital requirements. Naturally, given that a model passes the regulatory backtest, banks prefer a model that is more capital efficient and consequently less costly. We make the simplifying assumption that lower VaR numbers translate directly into lower regulatory capital requirements, i.e., that capital is preserved when VaR forecasts are lower but pass the backtest. This cost advantage is expressed on the same scale as the VaR numbers, so that, for example, a VaR forecast of 7% versus 8% translates into the bank having to hold 1% less in regulatory capital. Furthermore, we assume it to be desirable when VaR forecasts are less volatile over time (while still passing the backtest), because more volatile VaR numbers might translate into more variation in regulatory capital requirements or could be disadvantageous for internal risk budgeting.

A graphical inspection of daily VaR forecasts therefore sheds additional light on the adequacy and capital efficiency of VaR forecasts of competing models. From Figure 5 it is apparent that the Gaussian-CCC model produces huge VaR forecasts during the financial crisis due to its variance shooting up. This is economically very costly and thus the Gaussian distribution for VaR calculation should be avoided even for lower quantiles, such as the 95% VaR, where the failure rate is still acceptable according to the backtest. More crucially, Figure 5 shows that our COMFORT-RSDC models produce lower VaR forecasts throughout the entire financial crisis of 2008-2009, and hence lower capital costs than the CCC models of (roughly) the same backtesting quality. This comes at the cost of the RSDC model having slightly higher VaR forecasts during calm markets, best seen in the lower plot of Figure 5. A second advantage of the COMFORT-RSDC model is that it produces slightly smoother VaR forecasts over time compared the CCC model.

Figure 6 investigates in more detail the timing when use of the COMFORT-RSDC reduces capital requirements over the CCC model. The upper plot shows the relationship between the average GARCH-implied market volatility and the capital savings of the NIG-RSDC over the NIG-CCC

model for the 99% VaR. Positive values of capital saving stand for the RSDC model being more capital efficient than the CCC model. The scatter plot and the linear regression line depict a strong positive relationship between the two variables. The linear regression has an adjusted R^2 of 0.798 and the regression coefficient of the market volatility is 1.176. Hence, in periods of volatile markets, the RSDC model helps to save costs by reducing capital requirements. This is even more important because, during times of distress, liquidity is scarcest and borrowing capital is most costly. A bank thus prefers a risk model that provides accurate VaR forecasts, i.e., passes the backtest, and implies the lowest capital costs during volatile or crisis periods.
The bottom plot of Figure 6 compares the timing of capital savings of the RSDC model with the evolution of the costs of borrowing capital on the interbank market. We approximate these costs by considering the three month USD LIBOR rate, which is one of the most important interbank market rates, and accommodate the varying general level of interest rates by calculating the spread between this LIBOR rate and the effective FED funds rate. All data is gathered from The Federal Reserve Bank of St. Louis. The plot shows a striking congruence of time points when the RSDC model reduces capital requirements and when borrowing capital is costly for a bank, most notably the financial crisis 2008-2009 and the climax of the euro crisis in 2011. When performing an Engle-Granger test, the null hypothesis of no cointegration between the two time series is strongly rejected with a p-value below 0.001. If the St. Louis FED Financial Stress Index (STLFSI) is considered instead of the interbank rates, we find that the sharp increases in capital savings of the COMFORT-RSDC model at the end of 2003 and in the second half of 2011 correspond to jumps in the stress index which are not reflected in the interbank rates.
It should be noted that, for all non-Gaussian distributions, the COMFORT-RSDC implies capital savings with the same time pattern as depicted in Figure 6. Moreover, the average level of VaR forecasts over the whole sample period of all COMFORT-RSDC models (except for the asymmetric Student-t distribution) is smaller than that of the forecasts from the corresponding COMFORT-CCC models. Hence, even irrespective of different levels of capital costs, most COMFORT-RSDC models reduce the overall capital costs and do so most notably in times of financial distress. In stark contrast to this, the Gaussian-RSDC model is more costly than the Gaussian-CCC during the financial crisis 2008-2009, i.e., the spike in the bottom plot of Figure 6 has a negative sign and attains a values of -1.56% in this case. This presents another persuasive argument against the use of Gaussian models for risk management: not only are Gaussian models unable to pass the VaR backtests for quantiles higher than 95%, but even using more sophisticated correlation dynamics can be counterproductive under this highly restrictive distributional assumption.

Summarizing our findings, we suggest the COMFORT-RSDC model as a new methodology for highly accurate VaR forecasts that in addition appears to be particularly capital efficient during times of financial distress and causes the least variation in daily VaR forecasts (except for the SNIG-IID, which we reject due to its largely inferior performance). We also acknowledge that there is no clear edge when asymmetric MGHyp distributions are used instead of elliptical ones. The most distinct finding, and which is not new, is that the level of excess kurtosis and the ability to account for clustering in volatility are pivotal for good risk measure forecasts.

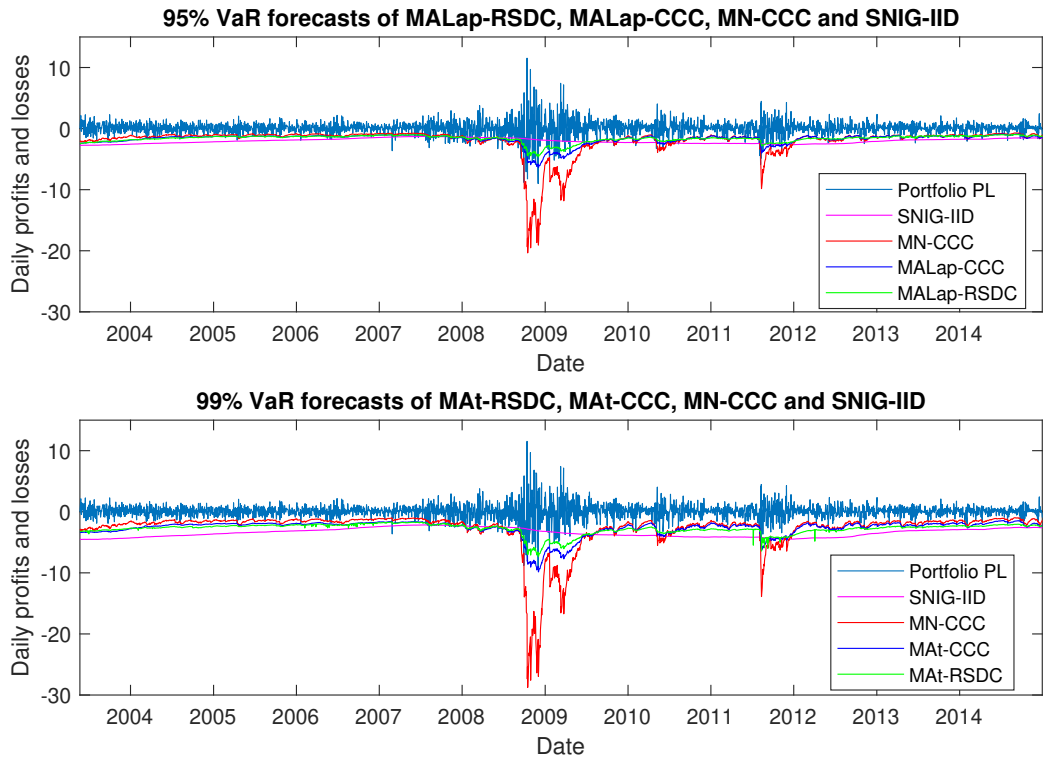

Figure 5: Daily profits and losses of the $1/K$ portfolio and daily VaR forecasts for several models, *Top:* 95% VaR forecasts of MALap-RSDC, MALap-CCC, SNIG-IID and MN-CCC; *Bottom:* 99% VaR forecasts of MAt-RSDC, MAt-CCC, SNIG-IID and MN-CCC

4.4 Out-of-Sample Portfolio Optimization

In Sections 4.1 and 4.2 we consider the in- and out-of-sample fit of the COMFORT-RSDC model class from a statistical perspective, and in Section 4.3 use the superior out-of-sample forecasting ability to forecast VaR. In this section we apply the COMFORT-RSDC model in the context of out-of-sample portfolio optimization. The importance of accounting for market downturns in correlation modeling and portfolio allocation was nicely highlighted by Greenspan (1999): "[...] joint distributions estimated over periods without panics will misestimate the degree of correlation between asset returns during panics. [...] Consequently, the benefits of portfolio diversification will tend to be overestimated when the rare panic periods are not taken into account."
Again, we form 2922 rolling windows, each of length 1000 days, and use the one-step-ahead density predictions to compute the optimal portfolio weights. Our portfolio optimization methodology is the minimum-expected shortfall (short: min-ES) framework, i.e., in every time step we minimize the expected portfolio loss over a one-day holding period subject to the constraint of being fully invested in the stock market. It is well known that portfolio strategies that solely minimize a risk measure, such as variance or expected shortfall, and do not require an estimate of expected returns, have more stable asset weights than, e.g., the Markowitz mean-variance portfolio. With respect to the choice of the risk measure, expected shortfall has many desirable properties and

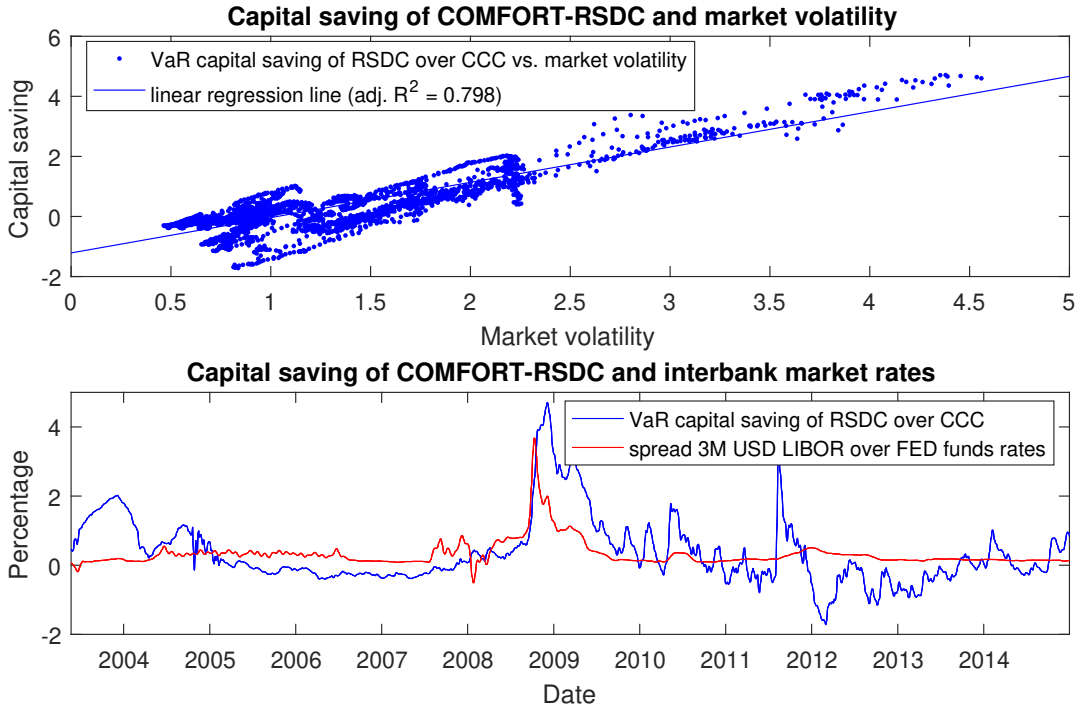

Figure 6: *Top:* Scatter plot and linear regression line of market volatility vs. cost savings of the NIG-RSDC over the NIG-CCC model for 99% VaR forecasting, *Bottom:* Time variation of cost savings of the NIG-RSDC over the NIG-CCC model for 99% VaR forecasting and of the costs of borrowing capital on the interbank market measured by the spread between the three month USD LIBOR rate over the effective FED funds rate

takes the full tail of the distribution into account; the latter argument being of importance when leaving the Gaussian world. In Paolella and Polak (2015c), the min-ES portfolio based on a broad range of multivariate models is shown to achieve consistently higher Sharpe ratios than the corresponding Markowitz mean-variance portfolio. In our implementation we use the approach of Rockafellar and Uryasev (2000, 2002) to translate the task of computing the min-ES portfolio into a convex optimization problem. Let the K-dimensional space of no-shortsale, fully invested portfolios be

$$\mathbf{X} := \{\mathbf{w} : w_i \geq 0 \; \forall i = 1, \ldots, K, |\mathbf{w}| = 1\} \subseteq \mathbb{R}^K, \tag{32}$$

and define an auxiliary function on $\mathbf{X} \times \mathbb{R}$ by

$$H_\alpha(\mathbf{w}, l) := l + \frac{1}{1-\alpha} \int_{\mathbf{y} \in \mathbb{R}^K} (-\mathbf{w}'\mathbf{y} - l)^+ f^{\mathcal{M}}_{t+1|t}(\mathbf{y}) d\mathbf{y}, \tag{33}$$

where $f^{\mathcal{M}}_{t+1|t}$ is the one-step-ahead density forecast based on observations up to time t. This function is convex and continuously differentiable in $(\mathbf{w}, l)$ so that optimization is computationally fast and delivers the global extremum. The min-ES portfolio is then computed as the global minimum of H_α, i.e.,

$$\min_{\mathbf{w} \in \mathbf{X}} \mathrm{ES}_\alpha(\mathbf{w}) = \min_{(\mathbf{w}, l) \in \mathbf{X} \times \mathbb{R}} H_\alpha(\mathbf{w}, l), \tag{34}$$

see Rockafellar and Uryasev (2000, Theorem 2). This approach does not require computing VaR in advance. Minimization of this well behaved convex function is fast even when K is large and the search space $\mathbf{X}$ is high dimensional. This allows us to perform an extensive comparison of out-of-sample portfolio performance of the various models under consideration.

The results of the portfolio analysis are summarized in Table 9 and sorted by Sharpe ratios. Comparing risk adjusted returns, we find that all MGHyp based models (CCC and RSDC) outperform the corresponding Gaussian models but the gain in Sharpe ratio is moderate. In particular, the Gaussian-RSDC beats the MAt-CCC and SNIG-CCC models (despite having inferior density forecasts). Even the worst performing models, the Gaussian-DCC and -CCC, beat the passive, equally weighted $1/K$ strategy in terms of higher Sharpe ratio, and lower portfolio volatility and maximum drawdown. An important finding is that the increase in Sharpe ratios when moving from CCC to RSDC is higher (around 0.1 on average) than the gain from using a MGHyp over a Gaussian distribution. Other measures of risk adjusted returns such as the Sortino and Starr ratios lead to the same conclusion. This indicates that, for portfolio allocation, modeling the dynamic dependency structure between assets is more important than using a non-normal distribution. It should be pointed out that the Markov regime switching model used in this study, albeit delivering better portfolio results, is a very simplistic approach that probably leaves room for further improvements.

While the benchmark $1/K$ strategy has the highest expected daily return and realized total return, it also has the highest volatility and maximum drawdown, as well as the lowest Sharpe-, Sortino- and Starr-ratios. Passive strategies, such as the equally weighted portfolio or value-weighted market ETFs, have large downside risk in market turmoils and the $1/K$ portfolio underperforms all min-ES portfolios in our analysis in terms of risk-adjusted returns on a long time horizon. In contrast, the min-ES strategies derived from the non-Gaussian COMFORT-CCC and -RSDC models can be regarded as particularly safe strategies for risk-averse investors. Compared to COMFORT-CCC, all corresponding COMFORT-RSDC models lead to higher expected returns and Sharpe ratios. Another clear advantage of the COMFORT-RSDC over the -CCC model is the increase in realized total return of around 15%. In addition, the COMFORT-RSDC model leads to a reduction in maximum drawdown of about $3-4\%$.

Overall, the best model in terms of Sharpe ratio is the symmetric Student-RSDC and, more importantly, all COMFORT-RSDC models deliver higher Sharpe ratios and realized total returns than their CCC competitors.

Finally and more generally, it should be noted that better out-of-sample density forecasting performance does not always translate into better portfolio performance. For example, the Gaussian-RSDC model attains a higher Sharpe ratio and total return than the MAt-CCC and the SNIG-CCC models, although it is vastly inferior in forecasting the return density. Same results are found for the impact of correlation shrinkage on the resulting portfolio performance; in contrast to positive impact on density prediction, which is rather insensitive to the level $\vartheta_{\mathbf{\Gamma}}$ of shrinkage strength (see Figure 1), we observe no clear pattern in the impact of correlation shrinkage on the performance as measured by the Sharpe ratio. A plot, analogous to Figure 1 but having the Sharpe ratio on the vertical axis, shows an erratic, non-monotone relationship between shrinkage strength $\vartheta_{\mathbf{\Gamma}}$ and Sharpe ratio. However, using no shrinkage at all seems to be inferior to using any positive shrink-

Model	*Exp. Daily Return*	*Volatility*	*Total Return*	*Max. Drawdown*	*Sharpe*	*Sortino*	*Starr*
Mt-RSDC	0.0425	0.8725	124.2199	22.8742	**0.7732**	**1.1058**	**0.0115**
MAt-RSDC	0.0426	0.8760	124.5814	23.1305	0.7724	1.1052	**0.0115**
MLap-RSDC	0.0423	0.8750	123.6263	22.5707	0.7673	1.0955	0.0113
NIG-RSDC	0.0425	0.8838	124.2862	22.8413	0.7638	1.0945	0.0114
MALap-RSDC	0.0420	0.8831	122.8223	**22.5410**	0.7554	1.0839	0.0112
SNIG-RSDC	0.0409	0.8715	119.6801	22.6054	0.7458	1.0653	0.0110
NIG-CCC	0.0383	0.8660	111.9449	25.9797	0.7020	1.0025	0.0104
Mt-CCC	0.0373	0.8657	108.9755	27.7583	0.6836	0.9669	0.0100
MALap-CCC	0.0370	**0.8624**	108.2771	25.5303	0.6819	0.9684	0.0101
MLap-CCC	0.0370	0.8648	108.1175	27.7149	0.6790	0.9621	0.0100
MN-RSDC	0.0369	0.8668	107.9421	26.0883	0.6763	0.9543	0.0098
MAt-CCC	0.0364	0.8642	106.4670	26.4901	0.6691	0.9490	0.0098
SNIG-CCC	0.0363	0.8661	106.1463	27.2155	0.6656	0.9416	0.0097
MN-CCC	0.0341	0.8736	99.5647	30.5057	0.6190	0.8752	0.0090
MN-DCC	0.0332	0.8698	96.9519	32.2004	0.6053	0.8538	0.0088
1/K	**0.0443**	1.1816	**129.4294**	42.1190	0.5949	0.8410	0.0085

Table 9: Comparison of min-ES portfolios at 99%-ES; $K = 29$ stocks from DJ30; out-of-sample trading period 23.05.2003 - 31.12.2014; rolling window size 1000 data points; RSDC models with same levels of correlation shrinkage $\vartheta_{\mathbf{\Gamma}}$ as in density forecasting analysis; all returns measured in percent

age strength. To avoid data overfitting we do not search for an optimal (highly data dependent) shrinkage strength and instead use the level $\vartheta^*_{\mathbf{\Gamma}}$ that is optimal for density forecasting.

Subperiods of the Dow Jones Data Set

When considering the two year crisis subperiod of 2008-2009, only the COMFORT-RSDC models attain a positive expected daily return, Sharpe ratio, and total return, due to the lower extreme losses compared to the COMFORT-CCC and the Gaussian models. The latter perform particularly poorly during the crisis years and attain the lowest expected daily and total return of all strategies, even lower than the equally weighted portfolio. The inferior performance of the Gaussian models during this turbulent two year horizon highlights that the normality assumption is dangerous in extremely volatile markets. On the other hand, each of the actively traded min-ES strategies (even under Gaussianity) significantly reduces the maximum drawdown compared to the passive $1/K$ strategy; this being a warning towards overconfidence in passive strategies (such as index ETFs) that are purportedly low-risk strategies. The active min-ES strategies, despite the restriction to be fully invested in the stock market, reduce losses during the acute crisis by diversifying risk much more successfully than the apparently well diversified $1/K$ portfolio.

High Dimensional Data Set

As second data set for portfolio optimization we consider the $K = 100$ top market capitalization companies of the SP500. The convex optimization formulation of Rockafellar and Uryasev (2000, 2002) allows to efficiently compute the min-ES portfolio even in this high dimensional setting. The results of the min-ES strategy for the various models under consideration are presented in Table 10. The first main finding is that the COMFORT-RSDC model class performs best again; all COMFORT-RSDC models beat the CCC and Gaussian models by means of higher Sharpe ratio and total returns, as well as lower maximum drawdown. A more surprising finding is the strong portfolio performance of the Gaussian-CCC model, which beats the Gaussian-RSDC and -DCC, as well as the COMFORT-CCC models. A possible explanation is the increased parameter

Model	*Exp. Daily Return*	*Volatility*	*Total Return*	*Max. Drawdown*	*Sharpe*	*Sortino*	*Starr*
MLap-RSDC	**0.0347**	0.7798	**87.6785**	33.4409	**0.7058**	**0.9960**	**0.0103**
NIG-RSDC	0.0346	0.7803	87.4555	34.3773	0.7035	0.9954	0.0102
SNIG-RSDC	0.0345	0.7799	87.3465	34.1237	0.7030	0.9924	0.0102
MAt-RSDC	0.0343	0.7843	86.7394	**32.6645**	0.6942	0.9825	0.0101
MALap-RSDC	0.0337	0.7809	85.2834	34.1478	0.6855	0.9674	0.0099
Mt-RSDC	0.0335	0.7834	84.6130	34.1920	0.6780	0.9575	0.0098
MN-CCC	0.0334	0.7872	84.4467	36.1323	0.6733	0.9443	0.0095
MAt-CCC	0.0327	0.7721	82.5880	37.6406	0.6714	0.9321	0.0093
MLap-CCC	0.0326	0.7730	82.4284	38.3570	0.6694	0.9265	0.0092
SNIG-CCC	0.0323	0.7729	81.7417	39.2862	0.6639	0.9186	0.0091
MALap-CCC	0.0319	**0.7705**	80.6091	38.7214	0.6567	0.9092	0.0091
MN-DCC	0.0325	0.7863	82.0689	36.1193	0.6551	0.9166	0.0092
Mt-CCC	0.0317	0.7729	80.2750	38.8401	0.6520	0.9012	0.0090
NIG-CCC	0.0312	0.7687	78.8837	39.1845	0.6442	0.8915	0.0089
MN-RSDC	0.0310	0.7810	78.2798	39.5186	0.6292	0.8741	0.0088
1/K	0.0238	1.2952	84.0289	66.1503	0.2918	0.4047	0.0043

Table 10: Comparison of the min-ES portfolios at 99%-ES; $K = 100$ top market cap stocks of SP 500; trading period from 15.12.2004 - 31.12.2014; rolling window size 2000 data points; RSDC models with a priori fixed level of correlation shrinkage $\vartheta_{\mathbf{\Gamma}}$; all returns measured in percent

estimation error in the high dimensional data set, which is more pronounced in the EM-algorithm for fitting the MGHyp distributions, and the covariance dynamics of the DCC and RSDC structures. Irrespective of which statistical model is chosen to compute the min-ES portfolio, this active strategy vastly outperforms the passive $1/K$ benchmark. The latter has, by far, the lowest risk-adjusted returns, the highest daily volatility, and the highest maximum drawdown. Equally diversifying the portfolio over all $K = 100$ stocks turns out to be a rather risky investment approach that delivers inferior risk-adjusted returns over this decade-long trading period. The large asset universe in this example allows the min-ES strategy to exploit diversification opportunities much more successfully than in the case of fewer assets. This results in vastly reduced volatility and drawdown, and, ultimately, in Sharpe ratios more than twice as high.

Dynamic Risk Control Based on Regime Forecasts

The probability forecasts of regimes from the COMFORT-RSDC model can be utilized for creating a portfolio strategy with the so-called Dynamic Risk Control (DRC) mechanism, where the stock portfolio is liquidated and cash is held if the volatile regime is likely to arise. For this analysis we focus on the symmetric Mt-RSDC model but the other cases of the MGHyp distribution perform similarly. In order to avoid excessive switching in and out of the stock market we compute the exponentially weighted moving average (EWMA) of the forecasts of regime probabilities $\hat{\boldsymbol{\xi}}_{t+1|t}$ and use this as signal for the market timing strategy. We hold the min-ES portfolio (computed as above) if and only if the smoothed forecast of the volatile regime is below a threshold $\tau \in [0, 1]$, otherwise we hold cash without interests. In practice one would move to high quality bonds or buy put options as protection. Both the number of lags of the EWMA smoother and the threshold τ are chosen by cross-validation, using a window size of 250 daily returns, such that the (in-sample) Sharpe ratio is maximized. We emphasize the important aspect here that no future data points are used to compute the DRC signal.

Figure 7 presents the out-of-sample rolling-window cumulative returns of the COMFORT-RSDC-based min-ES strategy with and without DRC, and the same results for the Gaussian-RSDC model, and Table 11 reports the performance measures. The regime-based DRC mechanism

Model	*Exp. Daily Return*	*Volatility*	*Total Return*	*Max. Drawdown*	*Sharpe*	*Sortino*	*Starr*
Mt-RSDC + DRC	0.0309	**0.4661**	79.5669	**5.8851**	**1.0527**	**1.5597**	**0.0156**
Mt-RSDC	0.0396	0.8995	102.0194	26.5049	0.6997	0.9987	0.0103
MN-RSDC + DRC	0.0196	0.4670	50.5603	10.8500	0.6676	0.9437	0.0094
MN-RSDC	0.0352	0.8913	90.5023	29.4612	0.6265	0.8833	0.0091
1/K	**0.0412**	1.1691	**120.3210**	48.7481	0.5589	0.7885	0.0079

Table 11: Comparison of min-ES portfolios with and without dynamic risk control (DRC); $K = 29$ stocks from DJ30; out-of-sample trading period 13.10.2004 - 31.12.2014.

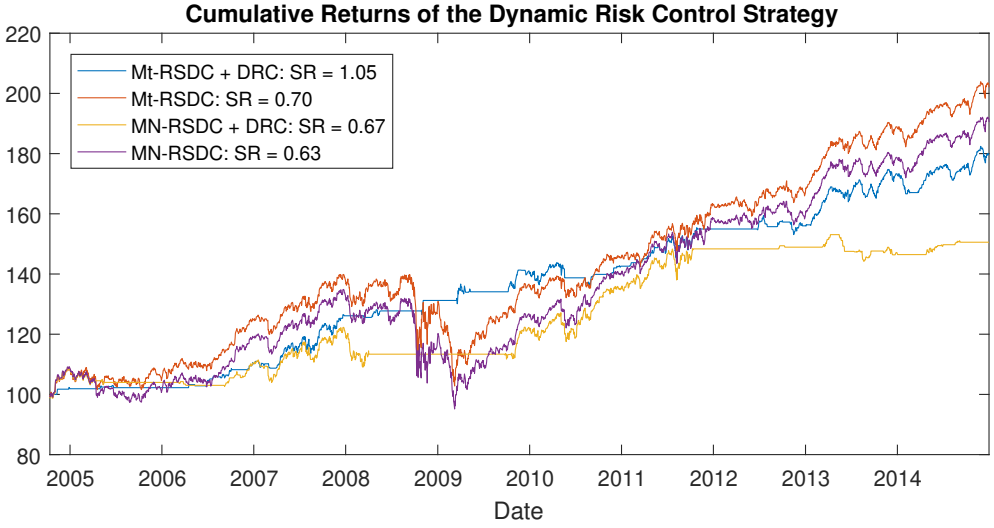

Figure 7: Cumulative returns of the min-ES portfolio based on COMFORT-RSDC with and without dynamic risk control, same for Gaussian-RSDC, out-of-sample trading period from 13.10.2004 to 31.12.2014.

allows to circumvent most losses in the financial crisis, drastically reducing the portfolio volatility and maximum drawdown. The DRC mechanism works best for the non-Gaussian COMFORT-RSDC model because the forecasts of regimes are enhanced compared to the Gaussian setup. In particular, the Gaussian-based models switch slower into the risk-free asset when market troubles arise, thereby not avoiding losses as effectively as the COMFORT-RSDC model. The gain in risk-adjusted returns in the Gaussian case is therefore only around 6%, whereas under the MGHyp-based model the use of the DRC increases the Sharpe ratio by 50%. The improvement of the risk-return profile stems purely from risk reduction, while the cumulative returns under DRC are lower than those of the baseline strategies. It should be noted that the DRC strategy does not use any external information about the stress level of the market, but is based solely on past return data, making the method applicable in markets where a stress index is not available. Finally, the improvements due to DRC are not limited to the acute financial crisis of 2008 but also help to avoid smaller market downturns, as in mid 2010 and in early 2012.

Summarizing the portfolio optimization applications of the COMFORT-RSDC model, we find that using our new model boosts portfolio performance in terms of consistently higher Sharpe ratios and total returns while at the same time reduces the extreme losses during the financial crisis. Overall, the influence of replacing the CCC by an RSDC structure appears to be at least

as important for portfolio performance than replacing the Gaussian distribution by a fat-tailed one. This underscores the finding that modeling the dependencies between assets is of utmost importance for asset allocation, and that the best results are achieved by combining the RSDC dynamics with a non-Gaussian MGHyp distribution to account for the excess kurtosis of returns. This reasoning also applies to our dynamic risk control mechanism. It is able to avoid heavy losses in the financial crisis, based on the regime forecasts of the COMFORT-RSDC model.

5 Conclusion

We have introduced a new, parametric, multivariate model class for financial asset returns that incorporates univariate GARCH-type dynamics, non-Gaussian conditional returns, and hidden Markov regime switching dynamics for the correlations. The model is coherent in the sense of being a well-defined stochastic process, as opposed to an ad-hoc assemblage of models estimated sequentially, such as a Gaussian-DCC or -RSDC overlaid with a multivariate Student-t with the latter estimated based on the residuals of the former.

We proposed and tested a likelihood-based estimation procedure that splits the estimation problem into two stages, both of which are amenable to estimation via an EM algorithm. Necessary and sufficient conditions for the consistency and asymptotic normality of the resulting iterative two-stage estimator are given. Out-of-sample model performance is enhanced by shrinkage estimation using a quasi-Bayesian prior on the regime-specific correlation matrices. A welcome side effect of our proposed methodology is, in addition to improved out-of-sample predictive performance, correlation shrinkage often greatly reduces the computation time.

In the empirical analysis, we investigated two non-Gaussian models of conditional correlations: (i) the Constant Conditional Correlation (COMFORT-CCC), and (ii) the Regime Switching Conditional Correlation model (COMFORT-RSDC), for various non-elliptic and elliptic special cases of the MGHyp distribution, including the Gaussian special case. On the basis of in-sample fit and out-of-sample density forecasting performance, all non-Gaussian models outperformed their Gaussian counterparts, including Gaussian-CCC and Gaussian-RSDC. In turn and importantly, all COMFORT-RSDC models outperformed their COMFORT-CCC counterparts, both in- and out-of-sample, with the best overall model being the symmetric Student-RSDC. There is some robustness to our empirical findings, in that various data sets, of different dimensionality and asset classes, were used, and very similar conclusions were found.

In an economic application, we have demonstrated that the new model leads to improved VaR forecasting, in terms of lower capital requirements during financial distress while easily passing regulatory backtests. When compared to the Gaussian-CCC model, the vastly superior risk measure forecasting performance is due to two factors: First, and obviously, addressing the blatant non-Gaussianity of the GARCH-filtered innovation sequences, and, secondly, via improved correlation prediction. The latter implies that applications explicitly centering on correlation predictions could also benefit from our model. We address this in an out-of-sample analysis of portfolio optimization—which is of obvious interest to major financial institutions and large investors such as pension funds. The poor VaR forecasts of the MGHyp-IID model highlights the utmost importance of addressing time-varying volatilities for risk management purposes, even under the greatly flexible MGHyp distribution.

In a financial application, the COMFORT-RSDC model is shown to deliver superior results in portfolio optimization. Using the min-ES methodology, the new model delivers higher risk-adjusted returns while also minimizing losses in times of financial distress compared to the COMFORT-CCC and Gaussian models. Importantly, the model easily outperforms the equally weighted benchmark portfolio and has preferable risk characteristics for risk-averse investors. Our results highlight that the best results are achieved when a sophisticated model for correlation dynamics is supplemented with a non-Gaussian distribution that accounts for the excess kurtosis of returns. The regime forecasts generated by the COMFORT-RSDC model can be used to detect market crashes and to avoid most losses, even in severe market downturns such as the financial crisis, by shifting the stock portfolio to a risk-free asset. While returns are inevitably lowered, the risk-adjusted performance of this dynamic risk control strategy is vastly enhanced.

New light is also shed on the importance of modeling the skewness of multivariate return series. Comparing the likelihoods of the COMFORT model, the asymmetric cases of the MGHyp distribution perform slightly better in-sample than the elliptical variants, but with respect to AIC and BIC this advantage is lost; similar findings are reported for the i.i.d. case in McNeil et al. (2015). Additionally, in accordance with our findings, the best distribution overall in Hu and Kercheval (2010) and McNeil et al. (2015) is the symmetric Student-t. In our out-of-sample analysis, the elliptical MGHyp models perform consistently better than their asymmetric counterparts. While returns are almost surely asymmetric (or more generally non-elliptical), e.g., exhibit different and non-zero asymmetry coefficients and tail coefficients (Paolella and Polak, 2015a), these properties are difficult to capture and more so to forecast. In a multivariate parametric model, due to the proliferation of the parameters and the bias-variance tradeoff, modeling these characteristics can induce too much estimation variance that harms the forecasting performance. For our MGHyp-based model, several approaches of shrinkage estimation of the asymmetry parameters have been tried but could not overcome this issue. As such, it seems safe to conclude that one can forgo the use of non-elliptical MGHyp models, i.e., set all asymmetry parameters to zero, in this context because meaningful estimation and forecasting of this parameter vector simply ask too much from the data.

Future research could investigate more general scale term dynamics than the GARCH(1,1)-model used herein, e.g., using APARCH or GJR-GARCH models. While this should improve modeling performance, we expect the advantage to be relatively small compared to the gains achieved when moving from the Gaussian to an MGHyp distribution, and from CCC to RSDC. As the regimes tend to be aligned with low and high volatility markets, another plausible future generalization would be to assume other features of the DGP also to be different between regimes, such as the MGHyp shape parameters, or possibly the persistence of volatility as dictated by the GARCH parameters. Allowing the tail shape to be regime switching is shown in Liu (2017) to successfully capture non-linearities in systemic risk contributions of institutions to a financial system, and is likely to be helpful also for modeling asset returns. As such, one could entertain a more general structure allowing for this, though note that doing so is not an immediate, trivial generalization of the framework herein, and the applicability of the two-stage EM algorithm would be jeopardized, thus rendering the model inapplicable for large scale problems.

It should be emphasized that, like Engle (2002) and Pelletier (2006), our model does not explicitly use any exogenous (say, macroeconomic) information. Future work (as kindly and astutely

suggested by an anonymous conference referee on an earlier draft) could explore if the regime shifts are related to the business cycle, policy regimes, or other such structural macroeconomic measures. Having now established the theoretical framework of model estimation, properties, and also demonstration of its efficacy in terms of risk prediction, such considerations will be of interest for future, possibly more applied, work.

Appendices

A In-sample Regime Probabilities and the Volatility Effect

In Pelletier (2006), the in-sample regime probabilities of the Gaussian-RSDC model are plotted, whereas out-of-sample forecasts of regimes are not analyzed. Figure 8 presents the corresponding in-sample regime probabilities for the Gaussian-RSDC and the symmetric Student-RSDC model, calculated with the smoothing algorithm of Kim (1994), as in Pelletier (2006). For many points in time, both models provide a separation of regimes, i.e., the probability to be in a given regime is either close to zero or one. However, a considerable number of regime probabilities are far away from the boundary of the unit interval. Furthermore, the second regime is strikingly prevalent in both models during the vast majority of trading days, making it difficult to assign an economic interpretation to the two regimes. Comparing these plots with those in Pelletier (2006), we observe that, for our model, switching between regimes occurs at a higher frequency and that the regime probabilities appear to be more noisy. However, one needs to be aware that Pelletier (2006) studies a $K = 4$ dimensional data set of daily FX rates, whereas we study a $K = 29$ dimensional stock return data set. The different structure of our in-sample regime probabilities could be due to the different type of assets, or due to the vastly larger dimensionality of the correlation matrices. Clearly, the more return series are analyzed, the more difficult it is to find a common trend in correlations.

A time-homogeneous Markov chain will switch fast between different states when the probability to remain in a given regime (the diagonal entries of the estimated transition matrix) are not close to 1. The probability to stay, e.g., in the first regime for a considerable number of consecutive days, declines rapidly in this case. For the Gaussian-RSDC model, the diagonal entries of the transition matrix are $\hat{\pi}_{11} = 0.78$ and $\hat{\pi}_{22} = 0.89$, which is in the same range as the values reported in Pelletier (2006) for their FX data set. For the symmetric Student-RSDC however, our estimates are $\hat{\pi}_{11} = 0.41$ and $\hat{\pi}_{22} = 0.82$, so that the small probability to remain in the first regime causes the second regime to be dominant.

Naively restricting the diagonal entries of the transition matrix to be closer to 1 deteriorates both in- and out-of-sample fit and indicates it is indeed desirable to switch back to the second regime (at least from a statistical point of view). Moreover, as shown in Table 1, our Markov regime switching model improves the in-sample fit due to its greater flexibility, even when accounting for increased complexity of the model, by way of the AIC. Still, we would like the two regimes to correspond to a calm and a volatile market environment, respectively, with different characteristics of the correlations mapped to those regimes. We address this issue in two ways in this paper. First, as subsequently discussed, we use information about the current market volatility to help infer the latent regime at hand, i.e., we exploit the well known volatility effect to infer the hidden

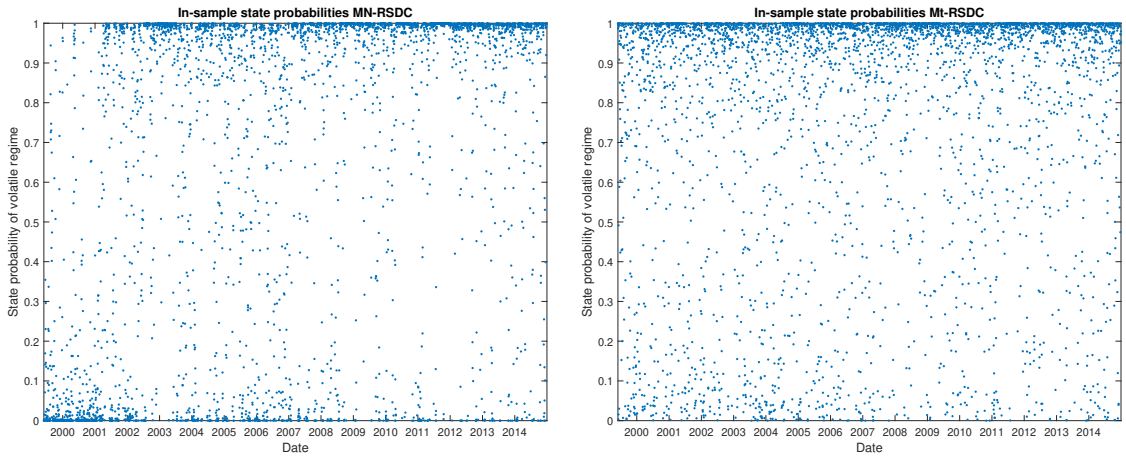

Figure 8: *Left*: In-sample smoothed state probabilities of the second (volatile) regime of the Gaussian-RSDC model on the DJ30 data, *Right*: same plot for the symmetric Student-RSDC model

market regimes. Second, as described in much detail earlier in Section 4.2, we use different levels of shrinkage estimation of the correlation matrices to differentiate stronger between the two regime-specific correlation matrices and to reduce estimation noise, resulting in both better density forecast quality and a clear identification of regimes.

The relationship between the general level of market volatility and the size of conditional correlations , i.e., the so-called volatility effect, is shown in Bauwens and Otranto (2016) to be statistically and economically significant for modeling and forecasting regimes in the Gaussian-RSDC model of Pelletier (2006). The authors use information about the GARCH implied market volatility to enhance the regime dynamics via time-varying transition probabilities that are given as a function of market volatility. Our approach is much simpler and we do not deviate from the setting of a time-homogeneous Markov chain but instead use a simple heuristic rule, based on the GARCH-implied average market volatility, to alter the in-sample regime probabilities (which are themselves the basis for estimating the transition matrix).

In particular, in the Stage-II EM algorithm, we set $\hat{\xi}_{2,t|T} = 1$, i.e., we enforce the second, volatile regime when the GARCH-implied market volatility exceeds its own longterm mean by more than one standard deviation; the first, calm regime is imposed analogously when the market volatility falls below one standard deviation of the mean. We recall that the univariate GARCH processes have been estimated in the Stage-I ECME algorithm and can readily be used in Stage-II. Other thresholds are of course possible, but we do not optimize this tuning parameter. We call our approach *regime forcing* henceforth. The resulting in-sample regime probabilities under regime forcing, presented in Figure 9, are less frantic and identify the dot-com crash, the Iraq war, the financial crisis of 2008-2009, and the peak of the euro crisis.

Regime forcing increases the persistence of regimes by altering the estimates of transition probabilities indirectly through (28): For the Gaussian-RSDC, the new estimates are $\hat{\pi}_{11} = 0.78$ and $\hat{\pi}_{22} = 0.94$, while for the symmetric Student-RSDC they are $\hat{\pi}_{11} = 0.72$ and $\hat{\pi}_{22} = 0.92$. These numbers are now very similar to those reported for daily FX returns in Pelletier (2006), indicating

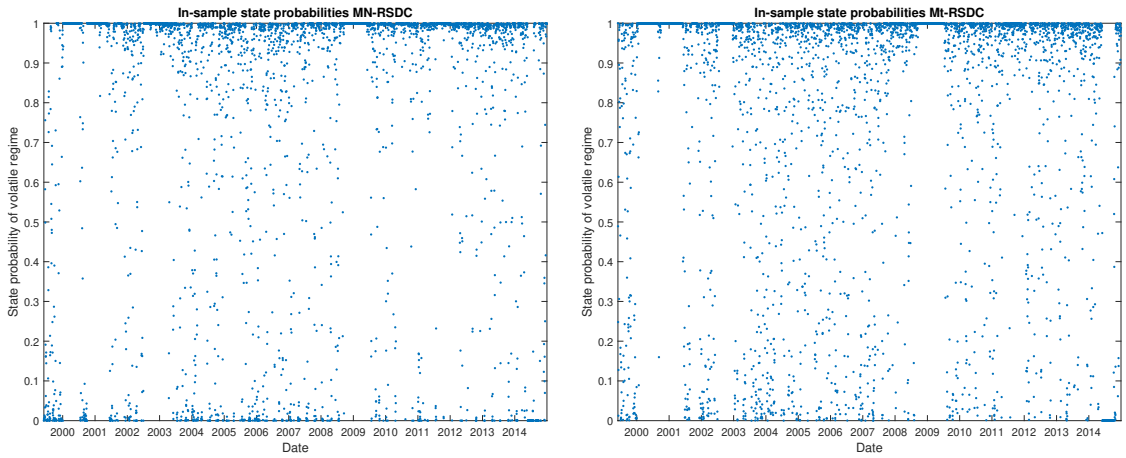

Figure 9: *Left*: In-sample state probabilities of the volatile regime of the Gaussian-RSDC model with regime forcing on the DJ30 data, *Right*: same plot for the symmetric Student-RSDC model

that the frequency of switching of our model is indeed in line with the results reported in the literature. With respect to the in-sample fit, regime forcing has, as expected, a negative impact and slightly lowers the attained maximum likelihoods of all RSDC models (by approximately -100 compared to the values in Table 1, which are computed without regime forcing) but not enough to alter the ranking.

In order to further investigate the regime probabilities, we fit the COMFORT-RSDC model to the six-dimensional data set of daily FX rate returns. These rates are euro, Australian dollar, Japanese yen, pound sterling, Hong Kong dollar and US dollar against the Swiss franc from July 1st 2004 until December 31st 2015. The in-sample smoothed state probabilities, both without and with regime forcing, are shown in Figure 10, and display two findings. First, the RSDC structure in fact makes use of both regimes, even without regime forcing, and very similar to the findings in Pelletier (2006). Second, the regime forcing procedure does not change the timing of regime switching under both the Gaussian- and the Student-t distribution when the regimes are already well inferred from the RSDC model without regime forcing. This supports the claim that the RSDC structure readily detects periods of high market volatility, and also, that our regime forcing approach does not drive the regimes but only supports their timely detection.

The out-of-sample density forecasts in Section 4.2 are performed using regime forcing because it increases the predictive log-likelihood by approximately 0.05 to 0.10. Furthermore, the plots of the one-day-ahead forecasts of regimes are visually improved and provide a clearer separation of regimes due to the changed estimates of the transition matrix.

The influence of regime forcing on the quality of VaR forecasts in Section 4.3 and on the portfolio performance in Section 4.4 is negligible, and for the latter, not always positive. For the sake of simplicity, we produce these results without regime forcing.

Concluding the discussion on in-sample regime probabilities, despite a reasonable use of regime forcing to enhance identification of regimes, on both data sets considered, the in-sample regime probabilities appear to be noisier compared to Pelletier (2006) (despite the transition matrices be-

ing close), and the second regime dominates through most of the sample period. While this might seem to be a shortcoming of our model (or the model implementation), we have seen in Section 4.2 that the out-of-sample forecasting analysis (for which we use the regime forcing approach in density and regime forecasting) shows the superiority of our two-component Markov RSDC model over the CCC, i.e., the one-component case. Our models use both regimes in a way that lends itself to reasonable economic interpretation; see the accordance of crisis forecasts and economic events described earlier. We emphasize that, also without regime forcing, all COMFORT-RSDC models clearly outperform all CCC models in density forecasting, i.e., the strong results of our new model are not driven by the regime forcing approach. We use the method only for the slightly enhanced out-of-sample forecasts of densities and regimes.

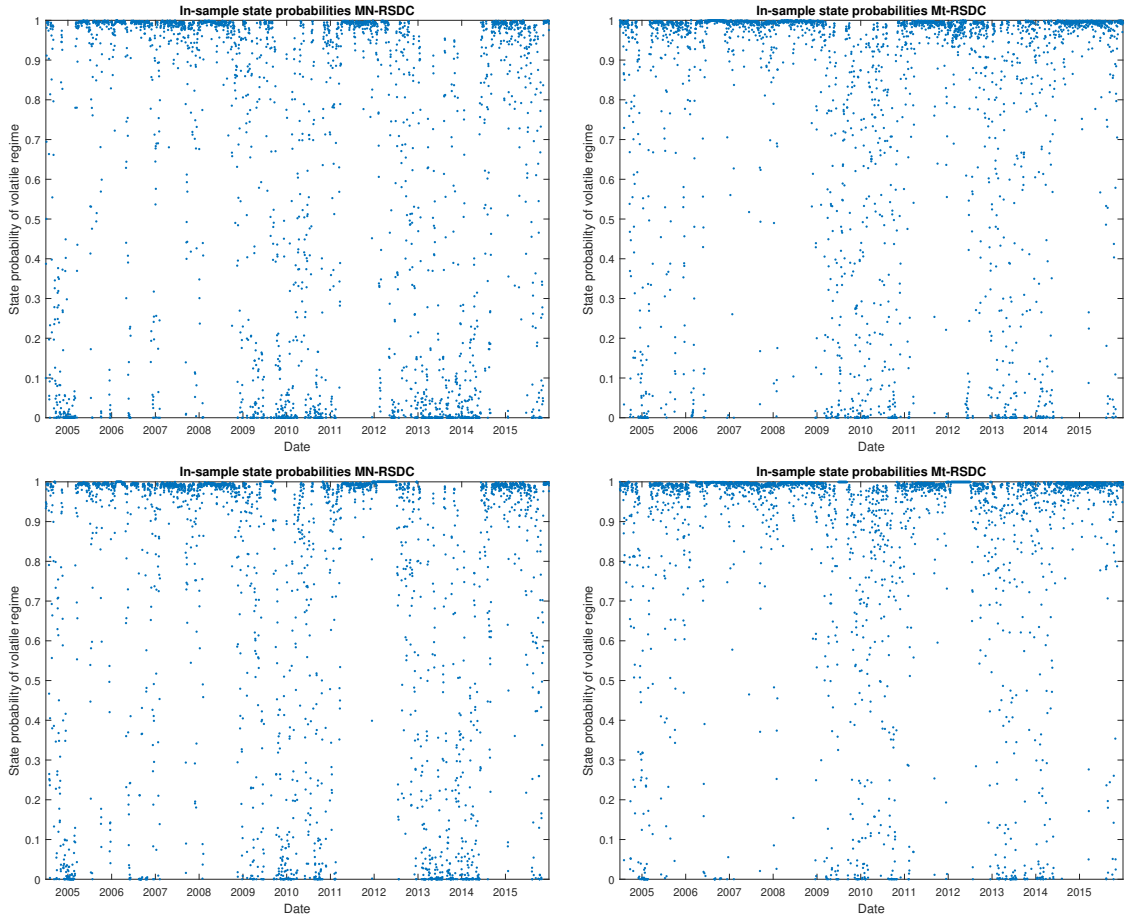

Figure 10: *Top left*: In-sample state probabilities of the volatile regime of the Gaussian-RSDC without regime forcing for the FX data. *Top right*: same for the symmetric Student-RSDC without regime forcing. *Bottom left*: same for the Gaussian-RSDC with regime forcing, *Bottom right*: same for the symmetric Student-RSDC with regime forcing

B Single Component Approximation of the Predictive Probability Density Function

The predictive density function of the COMFORT-RSDC model given in (10) is a finite mixture of MGHyp densities that differ only in their correlation matrices. Numerically, we find this finite mixture distribution to be well approximated by a single component MGHyp density, with the correlation matrix replaced by the probability weighted average of regime-specific correlation matrices,

$$f_{t+1|t}\left(\mathbf{y}_{t+1}\right) = \sum_{n=1}^{N} \xi_{n,t+1|t}\; f_{t+1|t}\left(\mathbf{y}_{t+1} \mid \Delta_{t+1} = n\right) \approx f_{t+1|t}\left(\mathbf{y}_{t+1} \middle| \sum_{n=1}^{N} \xi_{n,t+1|t} \mathbf{\Gamma}_n\right). \tag{35}$$

In terms of density and risk measure forecasting, approximation (35) is very precise and delivers almost identical results, while the computational effort in computing the VaR estimates is reduced almost by a factor of two, due to the smaller number of evaluations of the MGHyp density. This finding does not contradict our modeling approach of using a regime switching two-component mixture of MGHyp distributions. In contrast to the popular use of finite mixtures of Gaussian distributions with different variances in order to generate a higher kurtosis, we account for excess kurtosis of returns by use of semi-heavy or heavy tailed MGHyp distribution components that only differ in their correlation matrices; in particular the variances and the GIG distribution that determines the tail thickness in the mean-variance mixture distribution is the same in both components of our finite MGHyp mixture. Moreover, the differences between two correlation matrices in our RSDC model are small because we use correlation shrinkage with the same target matrix. The fact that the single-component approximation (35) works well and outperforms the COMFORT-CCC model indicates that the regime switching structure of correlation matrices is a better approximation of the true data generating process than CCC, to the extent that it clearly outweighs the approximation error made in (35). Irrespective of this, all reported results are attained from the formally correct two-component mixture distribution.

C Convergence of the Two-Step EM Algorithm

We show here that, under the usual assumption that the model reflects the true DGP except for the unknown parameters, the proposed two-stage EM algorithm monotonically increases the incomplete-data likelihood function. The proof is done in three steps. First, in Lemma C.1, we show that ignoring the correlation structure does not change estimates of $\boldsymbol{\theta}_P$ in the Stage-I ECME algorithm. Second, in Theorem C.1, we show that the monotonic increase of the incomplete-data likelihood is preserved when the Stage-I is performed on the subspace of the parameters space with fixed $\boldsymbol{\theta}_C$ and $\boldsymbol{\theta}_M$. Finally, in Theorem C.2, we show that the Stage-II EM algorithm also monotonically increases the incomplete-data likelihood function.

The first round of the Stage-I ECME algorithm is performed on the zero correlations subspace. Hence, the Markov chain Δ is omitted from the likelihood functions, $\boldsymbol{\theta}_M$ is not in the parameter space, and $\boldsymbol{\theta}_C = \mathbf{I}_K$ is fixed.

Let the incomplete-data likelihood function be denoted by $L_{\mathbf{Y}}\left(\boldsymbol{\theta}_P, \boldsymbol{\theta}_D, \boldsymbol{\theta}_C, \boldsymbol{\theta}_M\right)$. Then, by Bayes'

rule and the complete log-likelihood function decomposition (13),

$$\begin{aligned}\log L_{\mathbf{Y}}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_D,\boldsymbol{\theta}_C,\boldsymbol{\theta}_M\right) &= \log L_{\mathbf{Y},G}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_D,\boldsymbol{\theta}_C,\boldsymbol{\theta}_M\right) - \log k_{G|\mathbf{Y}}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_D,\boldsymbol{\theta}_C,\boldsymbol{\theta}_M\right)\\ &= \log L^{\mathrm{MV}}_{\mathbf{Y}|G}\left(\boldsymbol{\theta}_P\right) + \log L^{\mathrm{Corr}}_{\mathbf{Y}|G}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_C,\boldsymbol{\theta}_M\right)\\ &\quad + \log L_G\left(\boldsymbol{\theta}_D\right) - \log k_{G|\mathbf{Y}}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_D,\boldsymbol{\theta}_C,\boldsymbol{\theta}_M\right), \end{aligned} \tag{36}$$

where $k_{G|\mathbf{Y}}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_D,\boldsymbol{\theta}_C,\boldsymbol{\theta}_M\right)$ denotes the conditional probability density function of $G \mid \mathbf{Y}$. In the Stage-I E-step of the algorithm we take the conditional expectation under the probability measure given by last step estimates of $\boldsymbol{\theta}_P$ and $\boldsymbol{\theta}_D$ and with fixed $\boldsymbol{\theta}^*_C$ and $\boldsymbol{\theta}^*_M$. Thus we start by taking the expectation of both sides of (36) with respect to the conditional distribution of G_t given $\boldsymbol{Y}_t$, using the ℓth fit of parameters $(\boldsymbol{\theta}_P,\boldsymbol{\theta}_D)$, together with $\boldsymbol{\theta}^*_C$ and $\boldsymbol{\theta}^*_M$ fixed. This results in

$$\begin{aligned}\log L_{\mathbf{Y}}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_D,\boldsymbol{\theta}_C,\boldsymbol{\theta}_M\right) &= \mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell]},\boldsymbol{\theta}_D^{[\ell]},\boldsymbol{\theta}^*_C,\boldsymbol{\theta}^*_M\right)}\left[\log L^{\mathrm{MV}}_{\mathbf{Y}|G}\left(\boldsymbol{\theta}_P\right) + \log L^{\mathrm{Corr}}_{\mathbf{Y}|G}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_C,\boldsymbol{\theta}_M\right) \mid \boldsymbol{\Phi}_t\right]\\ &\quad + \mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell]},\boldsymbol{\theta}_D^{[\ell]},\boldsymbol{\theta}^*_C,\boldsymbol{\theta}^*_M\right)}\left[\log L_G\left(\boldsymbol{\theta}_D\right) \mid \boldsymbol{\Phi}_t\right]\\ &\quad - \mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell]},\boldsymbol{\theta}_D^{[\ell]},\boldsymbol{\theta}^*_C,\boldsymbol{\theta}^*_M\right)}\left[\log k_{G|\mathbf{Y}}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_D,\boldsymbol{\theta}_C,\boldsymbol{\theta}_M\right) \mid \boldsymbol{\Phi}_t\right]. \end{aligned} \tag{37}$$

Lemma C.1. *Let ℓ and $\ell+1$ be the consecutive steps in the Stage-I ECME algorithm performed on the fixed correlation subspace $(\boldsymbol{\theta}^*_C,\boldsymbol{\theta}^*_M)$. Then*

$$\boldsymbol{\theta}_P^{[\ell+1]} = \arg\max \mathcal{L}^{\mathrm{MV}}_{\mathbf{Y}|G}\left(\boldsymbol{\theta}_P\right) := \arg\max \mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell]},\boldsymbol{\theta}_D^{[\ell]},\boldsymbol{\theta}^*_C,\boldsymbol{\theta}^*_M\right)}\left[\log L^{\mathrm{MV}}_{\mathbf{Y}|G}\left(\boldsymbol{\theta}_P\right) \mid \boldsymbol{\Phi}_t\right], \tag{38}$$

maximizes also

$$\mathcal{L}_{\mathbf{Y}|G}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}^*_C,\boldsymbol{\theta}^*_M\right) := \mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell]},\boldsymbol{\theta}_D^{[\ell]},\boldsymbol{\theta}^*_C,\boldsymbol{\theta}^*_M\right)}\left[\log L^{\mathrm{MV}}_{\mathbf{Y}|G}\left(\boldsymbol{\theta}_P\right) + \log L^{\mathrm{Corr}}_{\mathbf{Y}|G}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}^*_C,\boldsymbol{\theta}^*_M\right) \mid \boldsymbol{\Phi}_t\right]. \tag{39}$$

Proof. We show that both sets of first order conditions with respect to $\boldsymbol{\theta}_P$ are satisfied for the same vector of parameters. Hence the $\ell+1$ step estimates of $\boldsymbol{\theta}_P$ from the fixed correlation model and the full model coincide. Define $\tilde{\varepsilon}_{k,t} = (Y_{k,t} - \mu_k - \gamma_k g_t)$ and vectors $\tilde{\boldsymbol{\varepsilon}}_t = (\mathbf{Y}_t - \boldsymbol{\mu} - \boldsymbol{\gamma} g_t)$ and $M_{k,t} = \left[0,\ldots,\left(Y_{k,t} - \mu_k - \gamma_k g_t\right) s_{k,t}^{-1},\ldots,0\right]'$. Then $\boldsymbol{\theta}_P^{[\ell+1]}$ are chosen so that:

$$\frac{\partial \mathcal{L}^{\mathrm{MV}}_{\mathbf{Y}|G}\left(\boldsymbol{\theta}_P\right)}{\partial \theta_{1,k,j}} = 0 \quad \Leftrightarrow$$

for $\theta_{1,k,j}$ equal ω_k, α_k or β_k we have

$$\mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell]},\boldsymbol{\theta}_D^{[\ell]},\boldsymbol{\theta}^*_C,\boldsymbol{\theta}^*_M\right)}\left[-\frac{1}{s_{k,t}}\frac{\partial s_{k,t}}{\partial \theta_{1,k,j}} + g_t^{-1}\tilde{\varepsilon}_{k,t}s_{k,t}^{-1}s_{k,t}^{-1}\tilde{\varepsilon}_{k,t}\frac{1}{s_{k,t}}\frac{\partial s_{k,t}}{\partial \theta_{1,k,j}}\right] = 0; \tag{40}$$

for $\theta_{1,k,j}$ equal μ_k we have

$$\mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell]},\boldsymbol{\theta}_D^{[\ell]},\boldsymbol{\theta}^*_C,\boldsymbol{\theta}^*_M\right)}\left[-\frac{1}{s_{k,t}}\frac{\partial s_{k,t}}{\partial \mu_k} + g_t^{-1}\tilde{\varepsilon}_{k,t}s_{k,t}^{-1}s_{k,t}^{-1}\left(\tilde{\varepsilon}_{k,t}\frac{1}{s_{k,t}}\frac{\partial s_{k,t}}{\partial \mu_k} - 1\right)\right] = 0; \tag{41}$$

for $\theta_{1,k,j}$ equal γ_k we have

$$\mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell]},\boldsymbol{\theta}_D^{[\ell]},\boldsymbol{\theta}^*_C,\boldsymbol{\theta}^*_M\right)}\left[-\frac{1}{s_{k,t}}\frac{\partial s_{k,t}}{\partial \gamma_k} + g_t^{-1}\tilde{\varepsilon}_{k,t}s_{k,t}^{-1}s_{k,t}^{-1}\left(\tilde{\varepsilon}_{k,t}\frac{1}{s_{k,t}}\frac{\partial s_{k,t}}{\partial \gamma_k} - g_t\right)\right] = 0. \tag{42}$$

On the other hand, the $\boldsymbol{\theta}_P^{[\ell+1]}$ that maximizes (39) must satisfy

$$\frac{\partial \mathcal{L}_{\mathbf{Y}|G}\left(\boldsymbol{\theta}_P, \boldsymbol{\theta}_C, \boldsymbol{\theta}_M\right)}{\partial \theta_{1,k,j}} = 0 \quad \Leftrightarrow$$

for $\theta_{1,k,j}$ equal ω_k, α_k or β_k we have

$$\mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell]}, \boldsymbol{\theta}_D^{[\ell]}, \boldsymbol{\theta}_C^*, \boldsymbol{\theta}_M^*\right)} \left[-\frac{1}{s_{k,t}} \frac{\partial s_{k,t}}{\partial \theta_{1,k,j}} + g_t^{-1} \tilde{\boldsymbol{\varepsilon}}_t \mathbf{S}_t^{-1} \boldsymbol{\Gamma}_t^{-1} M_{k,t} \frac{1}{s_{k,t}} \frac{\partial s_{k,t}}{\partial \theta_{1,k,j}} \right] = 0; \tag{43}$$

for $\theta_{1,k,j}$ equal μ_k we have

$$\mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell]}, \boldsymbol{\theta}_D^{[\ell]}, \boldsymbol{\theta}_C^*, \boldsymbol{\theta}_M^*\right)} \left[-\frac{1}{s_{k,t}} \frac{\partial s_{k,t}}{\partial \mu_k} + g_t^{-1} \tilde{\boldsymbol{\varepsilon}}_t \mathbf{S}_t^{-1} \boldsymbol{\Gamma}_t^{-1} \left(M_{k,t} \frac{1}{s_{k,t}} \frac{\partial s_{k,t}}{\partial \mu_k} - s_{k,t}^{-1} \right) \right] = 0; \tag{44}$$

for $\theta_{1,k,j}$ equal γ_k we have

$$\mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell]}, \boldsymbol{\theta}_D^{[\ell]}, \boldsymbol{\theta}_C^*, \boldsymbol{\theta}_M^*\right)} \left[-\frac{1}{s_{k,t}} \frac{\partial s_{k,t}}{\partial \gamma_k} + g_t^{-1} \tilde{\boldsymbol{\varepsilon}}_t \mathbf{S}_t^{-1} \boldsymbol{\Gamma}_t^{-1} \left(M_{k,t} \frac{1}{s_{k,t}} \frac{\partial s_{k,t}}{\partial \gamma_k} - s_{k,t}^{-1} g_t \right) \right] = 0. \tag{45}$$

When $\theta_{1,k,j}$ equals ω_k, α_k, or β_k, by applying the trace operator, we get that

$$g_t^{-1} \left(\mathbf{Y}_t - \boldsymbol{\mu} - \boldsymbol{\gamma} g_t \right) \mathbf{S}_t^{-1} \boldsymbol{\Gamma}_t^{-1} \left[0, \ldots, \left(Y_{k,t} - \mu_k - \gamma_k g_t \right) s_{k,t}^{-1}, \ldots, 0 \right]'$$

is a random variable with the same mean as $g_t^{-1} \left(Y_{k,t} - \mu_k - \gamma_k g_t \right) s_{k,t}^{-1} s_{k,t}^{-1} \left(Y_{k,t} - \mu_k - \gamma_k g_t \right)$. The analogous result follows for $\theta_{1,k,j}$ equal to μ_k or γ_k.
From this we conclude that the value of $\boldsymbol{\theta}_P^{[\ell+1]}$ that maximizes (38) also maximizes (39).

□

Next, we show the monotonicity of the Stage-I ECME algorithm.

Theorem C.1 (Monotonicity of Stage-I ECME)**.** *Let ℓ and $\ell+1$ be the consecutive steps in the ECME algorithm performed on the subspace with fixed correlation parameters ($\boldsymbol{\theta}_C^*$, $\boldsymbol{\theta}_M^*$), then the incomplete-data likelihood function increases for each iteration of the algorithm, i.e.,*

$$L_{\mathbf{Y}} \left(\boldsymbol{\theta}_P^{[\ell+1]}, \boldsymbol{\theta}_D^{[\ell+1]}, \boldsymbol{\theta}_C^*, \boldsymbol{\theta}_M^* \right) \geq L_{\mathbf{Y}} \left(\boldsymbol{\theta}_P^{[\ell]}, \boldsymbol{\theta}_D^{[\ell]}, \boldsymbol{\theta}_C^*, \boldsymbol{\theta}_M^* \right). \tag{46}$$

Proof. The CM1-step of the algorithm maximizes the first term in (37) and does not change the second term. The CM2-step finds $\boldsymbol{\theta}_D^{[\ell+1]}$ such that

$$\log L_{\mathbf{Y}} \left(\boldsymbol{\theta}_P^{[\ell+1]}, \boldsymbol{\theta}_D^{[\ell+1]}, \boldsymbol{\theta}_C^*, \boldsymbol{\theta}_M^* \right) \geq \log L_{\mathbf{Y}} \left(\boldsymbol{\theta}_P^{[\ell+1]}, \boldsymbol{\theta}_D, \boldsymbol{\theta}_C^*, \boldsymbol{\theta}_M^* \right), \text{ for all } \boldsymbol{\theta}_D. \tag{47}$$

For the last term in (37), by Jensen's inequality and the concavity of the logarithmic function, for all $\boldsymbol{\theta} = (\boldsymbol{\theta}_P, \boldsymbol{\theta}_D, \boldsymbol{\theta}_C)$,

$$\begin{aligned}
&\mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell]}, \boldsymbol{\theta}_D^{[\ell]}, \boldsymbol{\theta}_C^*, \boldsymbol{\theta}_M^*\right)} \left[\log k_{G_t|\boldsymbol{\Phi}_t} \left(g \mid \mathbf{y}_t; \boldsymbol{\theta} \right) \mid \boldsymbol{\Phi}_t \right] - \mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell]}, \boldsymbol{\theta}_D^{[\ell]}, \boldsymbol{\theta}_C^*, \boldsymbol{\theta}_M^*\right)} \left[\log k_{G_t|\boldsymbol{\Phi}_t} \left(g \mid \mathbf{y}_t; \boldsymbol{\theta}_P^{[\ell]}, \boldsymbol{\theta}_D^{[\ell]}, \boldsymbol{\theta}_C^*, \boldsymbol{\theta}_M^* \right) \mid \boldsymbol{\Phi}_t \right] \\
&= \mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell]}, \boldsymbol{\theta}_D^{[\ell]}, \boldsymbol{\theta}_C^*, \boldsymbol{\theta}_M^*\right)} \left[\log \frac{k_{G_t|\boldsymbol{\Phi}_t} \left(g \mid \mathbf{y}_t; \boldsymbol{\theta} \right)}{k_{G_t|\boldsymbol{\Phi}_t} \left(g \mid \mathbf{y}_t; \boldsymbol{\theta}_P^{[\ell]}, \boldsymbol{\theta}_D^{[\ell]}, \boldsymbol{\theta}_C^*, \boldsymbol{\theta}_M^* \right)} \mid \boldsymbol{\Phi}_t \right] \\
&\leq \log \mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell]}, \boldsymbol{\theta}_D^{[\ell]}, \boldsymbol{\theta}_C^*, \boldsymbol{\theta}_M^*\right)} \left[\frac{k_{G_t|\boldsymbol{\Phi}_t} \left(g \mid \mathbf{y}_t; \boldsymbol{\theta} \right)}{k_{G_t|\boldsymbol{\Phi}_t} \left(g \mid \mathbf{y}_t; \boldsymbol{\theta}_P^{[\ell]}, \boldsymbol{\theta}_D^{[\ell]}, \boldsymbol{\theta}_C^*, \boldsymbol{\theta}_M^* \right)} \mid \boldsymbol{\Phi}_t \right] \\
&= \log \int \frac{k_{G_t|\boldsymbol{\Phi}_t} \left(g \mid \mathbf{y}_t; \boldsymbol{\theta} \right)}{k_{G_t|\boldsymbol{\Phi}_t} \left(g \mid \mathbf{y}_t; \boldsymbol{\theta}_P^{[\ell]}, \boldsymbol{\theta}_D^{[\ell]}, \boldsymbol{\theta}_C^*, \boldsymbol{\theta}_M^* \right)} k_{G_t|\boldsymbol{\Phi}_t} \left(g \mid \mathbf{y}_t; \boldsymbol{\theta}_P^{[\ell]}, \boldsymbol{\theta}_D^{[\ell]}, \boldsymbol{\theta}_C^*, \boldsymbol{\theta}_M^* \right) \mathrm{d}g \\
&= \log \int k_{G_t|\boldsymbol{\Phi}_t} \left(g \mid \mathbf{y}_t; \boldsymbol{\theta} \right) \mathrm{d}g = 0.
\end{aligned}$$

Combining these three arguments proves (46) and establishes monotonicity of the Stage-I ECME algorithm. □

Finally, we show that the Stage-II EM algorithm also monotonically increases the incomplete-data log-likelihood. This EM algorithm estimates $\boldsymbol{\theta}_C$ and $\boldsymbol{\theta}_M$, and is performed conditionally on Stage-I ECME estimates $\widehat{\boldsymbol{\theta}}_P$ and $\widehat{\boldsymbol{\theta}}_D$. In the E-step of the algorithm we take the conditional expectation under the probability measure given by the estimates of $\boldsymbol{\theta}_P$ and $\boldsymbol{\theta}_D$ and with the last update of $\boldsymbol{\theta}_C$ and $\boldsymbol{\theta}_M$. Therefore, we start by taking the expectation of both sides of (36) with respect to the conditional distribution of G_t given $\boldsymbol{Y}_t$, $\widehat{\boldsymbol{\theta}}_P$, and $\widehat{\boldsymbol{\theta}}_D$, and using the ℓth fit of parameters in $\boldsymbol{\theta}_C$ and $\boldsymbol{\theta}_M$. This results in

$$\begin{aligned}\log L_{\mathbf{Y}}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_D,\boldsymbol{\theta}_C,\boldsymbol{\theta}_M\right) &= \mathbb{E}\left[\log L^{\mathrm{MV}}_{\mathbf{Y}|G}\left(\boldsymbol{\theta}_P\right) \mid \boldsymbol{\Phi}_t\right] + \mathbb{E}\left[\log L^{\mathrm{Corr}}_{\mathbf{Y}|G,\Delta}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_C,\boldsymbol{\theta}_M\right) \mid \boldsymbol{\Phi}_t\right] \\ &+ \mathbb{E}\left[\log L_G\left(\boldsymbol{\theta}_D\right) \mid \boldsymbol{\Phi}_t\right] - \mathbb{E}\left[\log k_{G|\mathbf{Y},\Delta}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_D,\boldsymbol{\theta}_C,\boldsymbol{\theta}_M\right) \mid \boldsymbol{\Phi}_t\right]. \\ &+ \mathbb{E}\left[\log L_{\Delta}\left(\boldsymbol{\theta}_M\right) \mid \boldsymbol{\Phi}_t\right] - \mathbb{E}\left[\log k_{\Delta|\mathbf{Y}}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_D,\boldsymbol{\theta}_C,\boldsymbol{\theta}_M\right) \mid \boldsymbol{\Phi}_t\right], \end{aligned} \tag{48}$$

where all the expectations are taken conditionally on $\left(\widehat{\boldsymbol{\theta}}_P,\widehat{\boldsymbol{\theta}}_D,\boldsymbol{\theta}_C^{[\ell]},\boldsymbol{\theta}_M^{[\ell]}\right)$.

Theorem C.2 (Monotonicity of Stage-II EM)**.** *Let ℓ and $\ell+1$ be the consecutive steps in the second-stage EM algorithm performed conditionally on the Stage-I estimates $\widehat{\boldsymbol{\theta}}_P$ and $\widehat{\boldsymbol{\theta}}_D$. Then the incomplete-data likelihood function increases for each iteration of the algorithm, i.e.,*

$$L_{\mathbf{Y}}\left(\widehat{\boldsymbol{\theta}}_P,\widehat{\boldsymbol{\theta}}_D,\boldsymbol{\theta}_C^{[\ell+1]},\boldsymbol{\theta}_M^{[\ell+1]}\right) \geq L_{\mathbf{Y}}\left(\widehat{\boldsymbol{\theta}}_P,\widehat{\boldsymbol{\theta}}_D,\boldsymbol{\theta}_C^{[\ell]},\boldsymbol{\theta}_M^{[\ell]}\right). \tag{49}$$

Proof. In each iteration, conditional on the Stage-I estimates $\widehat{\boldsymbol{\theta}}_P$ and $\widehat{\boldsymbol{\theta}}_D$, we update only the parameters in $\boldsymbol{\theta}_C$ and $\boldsymbol{\theta}_M$. From (48), $\log L_{\mathbf{Y}}\left(\widehat{\boldsymbol{\theta}}_P,\widehat{\boldsymbol{\theta}}_D,\boldsymbol{\theta}_C,\boldsymbol{\theta}_M\right)$ depends on $\boldsymbol{\theta}_C$ and $\boldsymbol{\theta}_M$ only through

$$\begin{aligned}&\mathbb{E}_{\left(\widehat{\boldsymbol{\theta}}_P,\widehat{\boldsymbol{\theta}}_D,\boldsymbol{\theta}_C^{[\ell]},\boldsymbol{\theta}_M^{[\ell]}\right)}\left[\log L^{\mathrm{Corr}}_{\mathbf{Y}|G,\Delta}\left(\widehat{\boldsymbol{\theta}}_P,\boldsymbol{\theta}_C,\boldsymbol{\theta}_M\right) + \log L_{\Delta}(\boldsymbol{\theta}_M) \mid \boldsymbol{\Phi}_t\right], \\ &\mathbb{E}_{\left(\widehat{\boldsymbol{\theta}}_P,\widehat{\boldsymbol{\theta}}_D,\boldsymbol{\theta}_C^{[\ell]},\boldsymbol{\theta}_M^{[\ell]}\right)}\left[\log k_{G|\mathbf{Y},\Delta}\left(\widehat{\boldsymbol{\theta}}_P,\widehat{\boldsymbol{\theta}}_D,\boldsymbol{\theta}_C,\boldsymbol{\theta}_M\right) \mid \boldsymbol{\Phi}_t\right], \text{ and} \\ &\mathbb{E}_{\left(\widehat{\boldsymbol{\theta}}_P,\widehat{\boldsymbol{\theta}}_D,\boldsymbol{\theta}_C^{[\ell]},\boldsymbol{\theta}_M^{[\ell]}\right)}\left[\log k_{\Delta|\mathbf{Y}}\left(\widehat{\boldsymbol{\theta}}_P,\widehat{\boldsymbol{\theta}}_D,\boldsymbol{\theta}_C,\boldsymbol{\theta}_M\right) \mid \boldsymbol{\Phi}_t\right].\end{aligned}$$

First, note that, under no correlation shrinkage,

$$\boldsymbol{\theta}_C^{[\ell+1]} = \arg\max_{\boldsymbol{\theta}_C} \mathbb{E}_{\left(\widehat{\boldsymbol{\theta}}_P,\widehat{\boldsymbol{\theta}}_D,\boldsymbol{\theta}_C^{[\ell]},\boldsymbol{\theta}_M^{[\ell]}\right)}\left[L^{\mathrm{Corr}}_{\mathbf{Y}|G,\Delta}\left(\widehat{\boldsymbol{\theta}}_P,\boldsymbol{\theta}_C,\boldsymbol{\theta}_M\right) \mid \boldsymbol{\Phi}_t\right],$$

where $\hat{\mathbf{e}}_{n,t} = g_{n,t}^{-1/2}\mathbf{S}_{n,t}^{-1}\boldsymbol{\varepsilon}_{n,t}$, and $\boldsymbol{\varepsilon}_{n,t} = \mathbf{y}_t - \hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\gamma}} g_{n,t}$ are the regime-specific residuals with $g_{n,t}$ filtered in the E-step-2 of Stage-II algorithm.
Second, following the derivation in Hamilton (1990),

$$\boldsymbol{\theta}_M^{[\ell+1]} = \arg\max_{\boldsymbol{\theta}_M} \mathbb{E}_{\left(\widehat{\boldsymbol{\theta}}_P,\widehat{\boldsymbol{\theta}}_D,\boldsymbol{\theta}_C^{[\ell]},\boldsymbol{\theta}_M^{[\ell]}\right)}\left[\log L_{\mathbf{Y}|\mathbf{G},\boldsymbol{\Delta}}\left(\widehat{\boldsymbol{\theta}}_P,\widehat{\boldsymbol{\theta}}_D,\boldsymbol{\theta}_C^{[\ell]},\boldsymbol{\theta}_M\right) + \log L_{\boldsymbol{\Delta}}\left(\boldsymbol{\theta}_M\right) \mid \boldsymbol{\Phi}_t\right].$$

Then, using Jensen's inequality and the same argument as in the proof of the Stage-I ECME algorithm,

$$\mathbb{E}_{\left(\widehat{\boldsymbol{\theta}}_P,\widehat{\boldsymbol{\theta}}_D,\boldsymbol{\theta}_C^{[\ell]},\boldsymbol{\theta}_M^{[\ell]}\right)}\left[\log k_{G|\mathbf{Y},\Delta}\left(\widehat{\boldsymbol{\theta}}_P,\widehat{\boldsymbol{\theta}}_D,\boldsymbol{\theta}_C^{[\ell+1]},\boldsymbol{\theta}_M^{[\ell+1]}\right) - \log k_{G|\mathbf{Y},\Delta}\left(\widehat{\boldsymbol{\theta}}_P,\widehat{\boldsymbol{\theta}}_D,\boldsymbol{\theta}_C^{[\ell]},\boldsymbol{\theta}_M^{[\ell]}\right) \mid \boldsymbol{\Phi}_t\right] \leq 0$$

Number of assets	*Corr. shrinkage* ϑ_Γ *Gaussian*	*Corr. shrinkage* ϑ_Γ *MGHyp*	*Window Size*
2-10	25	250	800
11-25	25	250	1000
26-50	100	750	1000
51-100	200	1000	2000

Table 12: A priori rule for choosing the correlation shrinkage strength and rolling window size

for all $\boldsymbol{\theta}_C$ and $\boldsymbol{\theta}_M$, and

$$\mathbb{E}_{\left(\widehat{\boldsymbol{\theta}}_P,\widehat{\boldsymbol{\theta}}_D,\boldsymbol{\theta}_C^{[\ell]},\boldsymbol{\theta}_M^{[\ell]}\right)}\left[\log k_{\Delta|\mathbf{Y}}\left(\widehat{\boldsymbol{\theta}}_P,\widehat{\boldsymbol{\theta}}_D,\boldsymbol{\theta}_C^{[\ell+1]},\boldsymbol{\theta}_M^{[\ell+1]}\right)-\log k_{\Delta|\mathbf{Y}}\left(\widehat{\boldsymbol{\theta}}_P,\widehat{\boldsymbol{\theta}}_D,\boldsymbol{\theta}_C^{[\ell]},\boldsymbol{\theta}_M^{[\ell]}\right)\mid\boldsymbol{\Phi}_t\right]\leq 0$$

for all $\boldsymbol{\theta}_C$ and $\boldsymbol{\theta}_M$. This completes the proof. □

The consequence of Theorems C.1 and C.2 is that, if one performs Stage-I and Stage-II of our two-stage EM algorithm iteratively, if $\vartheta_{\boldsymbol{\Gamma}}=0$, and if the starting values for both algorithms, $\left(\boldsymbol{\theta}_P^{[1]},\boldsymbol{\theta}_D^{[1]}\right)$ and $\left(\boldsymbol{\theta}_C^{[1]},\boldsymbol{\theta}_M^{[1]}\right)$, are sufficiently close to the true parameters, or if the log likelihood is unimodal in the parameter space, then monotonicity in the likelihood values of the estimates $\left(\boldsymbol{\theta}_P^{[\ell]},\boldsymbol{\theta}_D^{[\ell]}\right)$ and $\left(\boldsymbol{\theta}_C^{[\ell]},\boldsymbol{\theta}_M^{[\ell]}\right)$ guarantees their convergence to the corresponding maximum likelihood estimates. Thus, the iterative two-stage EM algorithm, under the aforementioned assumptions, converges to the global maximum of the incomplete data likelihood function with respect to $(\boldsymbol{\theta}_P,\boldsymbol{\theta}_D,\boldsymbol{\theta}_C,\boldsymbol{\theta}_M)$. If correlation shrinkage is employed, i.e., in case $\vartheta_{\boldsymbol{\Gamma}}>0$, then the objective function is altered along the lines of Hamilton (1991), and the proof follows with the likelihood $L_{\mathbf{Y}}$ replaced by the new objective function as in Hamilton (1994, 22.3.11).

D A priori choice of correlation shrinkage strength

Performing a grid choice for the correlation shrinkage strength $\vartheta_{\boldsymbol{\Gamma}}^*$ that maximizes the out-of-sample density forecasting ability is very time consuming and not feasible in practice. We provide a rule of thumb for an a priori choice of $\vartheta_{\boldsymbol{\Gamma}}$ that accounts for different dimensions of the data sets. This rule of thumb, presented in Table 12, chooses $\vartheta_{\boldsymbol{\Gamma}}$ that are almost surely non-optimal but are shown in Section 4.2 to perform well. It is worthy to note again that, on the DJ30 data set investigated, the COMFORT-RSDC outperforms the single component case under any level of correlation shrinkage and that, according to Figure 1, the exact level is almost irrelevant as long as it is not too small. The values given in Table 12 can be regarded as recommended lower bounds for $\vartheta_{\boldsymbol{\Gamma}}$.

References

Aas, K. and Haff, I. H. (2006). The Generalized Hyperbolic Skew Students t-Distribution. *Journal of Financial Econometrics*, 4(2):275–309.

Aas, K., Haff, I. H., and Dimakos, X. K. (2005). Risk Estimation using the Multivariate Normal Inverse Gaussian Distribution. *Journal of Risk*, 8(2):39–60.

Aielli, G. P. and Caporin, M. (2014). Variance Clustering Improved Dynamic Conditional Correlation MGARCH Estimators. *Computational Statistics & Data Analysis*, 76:556 – 576.

Amisano, G. and Giacomini, R. (2007). Comparing Density Forecasts via Weighted Likelihood Ratio Tests. *Journal of Business & Economic Statistics*, 25(2):177–190.

Bailey, D. H., Borwein, J. M., López de Prado, M., and Zhu, Q. J. (2014). Pseudo-Mathematics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-of-Sample Performance. *Notices of the American Mathematical Society*, 61(5):458–471.

Bauwens, L., Hafner, C. M., and Rombouts, J. V. K. (2007). Multivariate Mixed Normal Conditional Heteroskedasticity. *Computational Statistics & Data Analysis*, 51(7):3551–3566.

Bauwens, L., Laurent, S., and Rombouts, J. K. V. (2006). Multivariate GARCH Models: A Survey. *Journal of Applied Econometrics*, 21(1):79–109.

Bauwens, L. and Otranto, E. (2016). Modeling the Dependence of Conditional Correlations on Market Volatility. *Journal of Business & Economic Statistics*, 34(2):254–268.

Bianchi, M. L., Tassinari, G. L., and Fabozzi, F. J. (2016). Riding with the Four Horsemen and the Multivariate Normal Tempered Stable Model. *International Journal of Theoretical and Applied Finance*, 19(04):1650027.

Billio, M. and Caporin, M. (2005). Multivariate Markov Switching Dynamic Conditional Correlation GARCH Representations for Contagion Analysis. *Statistical Methods and Applications*, 14(2):145–161.

Billio, M., Caporin, M., and Gobbo, M. (2006). Flexible Dynamic Conditional Correlation Multivariate GARCH Models for Asset Allocation. *Applied Financial Economics Letters*, 2(2):123–130.

Billio, M. and Pelizzon, L. (2000). Value-at-Risk: A Multivariate Switching Regime Approach. *Journal of Empirical Finance*, 7(5):531–554.

Black, F. (1976). Studies of Stock Price Volatility Changes. *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics Section*, pages 177–181.

Bollerslev, T. (1990). Modeling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Approach. *Review of Economics and Statistics*, 72:498–505.

Bollerslev, T. (2010). Glossary to ARCH (GARCH). In Bollerslev, T., Russell, J., and Watson, M., editors, *Volatility and Time Series Econometrics: Essays in Honor of Robert Engle*, chapter 8, pages 137–163. Oxford University Press, Oxford.

Cappiello, L., Engle, R. F., and Sheppard, K. (2006). Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns. *Journal of Financial Econometrics*, 4(4):537–572.

Chevallier, J. and Goutte, S. (2015). Detecting Jumps and Regime Switches in International Stock Markets Returns. *Applied Economics Letters*, 22(13):1011–1019.

Chollete, L., Heinen, A., and Valdesogo, A. (2009). Modeling International Financial Returns with a Multivariate Regime-Switching Copula. *Journal of Financial Econometrics*, 7(4):437–480.

Christoffersen, P. F. (1998). Evaluating Interval Forecasts. *International Economic Review*, pages 841–862.

Christoffersen, P. F., Hahn, J., and Inoue, A. (2001). Testing and Comparing Value-at-Risk Measures. *Journal of Empirical Finance*, 8(3):325–342.

Christoffersen, P. F. and Pelletier, D. (2004). Backtesting Value-at-Risk: A Duration-Based Approach. *Journal of Financial Econometrics*, 2(1):84–108.

Diebold, F. X. and Li, C. (2006). Forecasting the Term Structure of Government Bond Yields. *Journal of Econometrics*, 130(2):337–364.

Elliott, G. and Timmermann, A. (2008). Economic Forecasting. *Journal of Economic Literature*, 46(1):3–56.

Engle, R. F. (2002). Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business and Economic Statistics*, 20:339–350.

Engle, R. F. (2009). *Anticipating Correlations: A New Paradigm for Risk Management*. Princeton University Press, Princeton.

Engle, R. F. and Kelly, B. (2012). Dynamic Equicorrelation. *Journal of Business & Economic Statistics*, 30(2):212–228.

Fink, H., Klimova, Y., Czado, C., and Stöber, J. (2017). Regime Switching Vine Copula Models for Global Equity and Volatility Indices. *Econometrics*, 5(1):3.

Gneiting, T. and Ranjan, R. (2011). Comparing Density Forecasts Using Threshold-and Quantile-Weighted Scoring Rules. *Journal of Business & Economic Statistics*, 29(3):411–422.

Greenspan, A. (1999). New Challenges for Monetary Policy - A Symposium Sponsored by the Federal Reserve Bank of Kansas City. Opening Remarks of the Chairmain of the Board of Governors of the Federal Reserve System.

Haas, M. (2005). Improved Duration-Based Backtesting of Value-at-Risk. *The Journal of Risk*, 8(2):17–38.

Haas, M., Mittnik, S., and Paolella, M. S. (2004). A New Approach to Markov Switching GARCH Models. *Journal of Financial Econometrics*, 2(4):493–530.

Haas, M., Mittnik, S., and Paolella, M. S. (2009). Asymmetric Multivariate Normal Mixture GARCH. *Computational Statistics & Data Analysis*, 53(6):2129–2154.

Haas, M. and Paolella, M. S. (2012). Mixture and Regime–switching GARCH Models. In Bauwens, L., Hafner, C. M., and Laurent, S., editors, *Handbook of Volatility Models and their Applications*, number 3, Hoboken, New Jersey. John Wiley & Sons, Inc.

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57:257–84.

Hamilton, J. D. (1990). Analysis of Time Series Subject to Changes in Regime. *Journal of Econometrics*, 45(1):39–70.

Hamilton, J. D. (1991). A Quasi-Bayesian Approach to Estimating Parameters for Mixtures of Normal Distributions. *Journal of Business and Economic Statistics*, 9(1):21–39.

Hamilton, J. D. (1993). *Handbook of Statistics Vol. 11*, chapter Estimation, Inference, and Forecasting of Time Seris Subject to Changes in Regime. New York: North-Holland.

Hamilton, J. D. (1994). *Time Series Analysis*. Princton University Press.

Härdle, W. K., Okhrin, O., and Wang, W. (2015). Hidden Markov Structures for Dynamic Copulae. *Econometric Theory*, 31(05):981–1015.

Henry, O. T. (2009). Regime Switching in the Relationship between Equity Returns and Short-Term Interest Rates in the UK. *Journal of Banking & Finance*, 33(2):405–414.

Hu, W. and Kercheval, A. N. (2010). Portfolio Optimization for Student t and Skewed t Returns. *Quantitative Finance*, 10(1):91–105.

Jensen, M. B. and Lunde, A. (2001). The NIG-S&ARCH Model: A Fat-Tailed, Stochastic and Autoregressive Conditional Heteroskedastic Volatility Model. *Econometrics Journal*, 4:319–342.

Jondeau, E., Poon, S.-H., and Rockinger, M. (2007). *Financial Modeling Under Non-Gaussian Distributions*. Springer, London.

Kasch, M. and Caporin, M. (2013). Volatility Threshold Dynamic Conditional Correlations: An International Analysis. *Journal of Financial Econometrics*, 11(4):706–742.

Kim, C.-J. (1994). Dynamic Linear Models with Markov-Switching. *Journal of Econometrics*, 60(1-2):1–22.

Kuester, K., Mittnik, S., and Paolella, M. S. (2006). Value–at–Risk Prediction: A Comparison of Alternative Strategies. *Journal of Financial Econometrics*, 4:53–89.

Liu, X. (2017). Measuring Systemic Risk with Regime Switching in Tails. *Economic Modelling*, 67:55–72.

McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools (Princeton Series in Finance)*. Princeton University Press, revised edition.

Pagan, A. (1986). Two Stage and Related Estimators and their Applications. *The Review of Economic Studies*, 53(4):517–538.

Paolella, M. S. (2007). *Intermediate Probability: A Computational Approach*. Wiley-Interscience.

Paolella, M. S. (2015). Multivariate Asset Return Prediction with Mixture Models. *European Journal of Finance*, 21(13-14):1214–1252.

Paolella, M. S. and Polak, P. (2015a). ALRIGHT: Asymmetric Large-Scale (I) GARCH with Hetero-Tails. *International Review of Economics & Finance*, 40:282–297.

Paolella, M. S. and Polak, P. (2015b). COMFORT: A Common Market Factor Non-Gaussian Returns Model. *Journal of Econometrics*, 187(2):593–605.

Paolella, M. S. and Polak, P. (2015c). Portfolio Selection with Active Risk Monitoring. Technical report, Swiss Finance Institute Research Paper Series No.15-17.

Paolella, M. S. and Polak, P. (2017). Density and Risk Prediction with Non-Gaussian COMFORT Models. Mimeo.

Pelletier, D. (2006). Regime Switching for Dynamic Correlations. *Journal of Econometrics*, 131:445–473.

Pelletier, D. and Wei, W. (2016). The Geometric-VaR Backtesting Method. *Journal of Financial Econometrics*, 14(4):725.

Prause, K. (1999). *The Generalized Hyperbolic Model: Estimation, Financial Derivatives, and Risk Measures*. PhD thesis, University of Freiburg.

Protassov, R. (2004). EM-based Maximum Likelihood Parameter Estimation for Multivariate Generalized Hyperbolic Distributions with Fixed Lambda. *Statistics and Computing*, 14:1:67–77.

Righi, M. B. and Ceretta, P. S. (2015). A Comparison of Expected Shortfall Estimation Models. *Journal of Economics and Business*, 78:14–47.

Rockafellar, R. and Uryasev, S. (2000). Optimization of Conditional Value-at-Risk. *Journal of Risk*, 2:21–42.

Rockafellar, R. and Uryasev, S. (2002). Conditional Value-at-Risk for General Loss Distributions. *Journal of Banking & Finance*, 26(7):1443–1471.

Santos, A. A. P., Nogales, F. J., and Ruiz, E. (2013). Comparing Univariate and Multivariate Models to Forecast Portfolio Value–at–Risk. *Journal of Financial Econometrics*, 11(2):400–441.

Slim, S., Koubaa, Y., and BenSaida, A. (2017). Value-at-Risk under Lévy GARCH models: Evidence from global stock markets. *Journal of International Financial Markets, Institutions and Money*, 46:30–53.

So, M. K. P. and Yip, I. W. H. (2012). Multivariate GARCH Models with Correlation Clustering. *Journal of Forecasting*, 31(5):443–468.

Tay, A. S. and Wallis, K. F. (2000). Density Forecasting: A Survey. *Journal of Forecasting*, 19(4):124–143.

Timmermann, A. (2000). Density Forecasting in Economics and Finance. *Journal of Forecasting*, 19(4):231–234.

Tse, Y. K. and Tsui, A. K. C. (2002). A Multivariate Generalized Autoregressive Conditional Heteroscedasticity Model With Time-Varying Correlations. *Journal of Business and Economic Statistics*, 20(3):351–362.

Urga, G., Cajigas, J.-P., and Ghalanos, A. (2011). Dynamic Conditional Correlation Models with Asymmetric Multivariate Laplace Innovations. Working Paper.

Virbickaite, A., Ausin, M. C., and Galeano, P. (2016). A Bayesian Non-Parametric Approach to Asymmetric Dynamic Conditional Correlation Model with Application to Portfolio Selection. *Computational Statistics & Data Analysis*, 100:814–829.

Weigend, A. S. and Shi, S. (2000). Predicting Daily Probability Distributions of S&P500 returns. *Journal of Forecasting*, 19(4):375–392.

Wu, L., Meng, Q., and Velazquez, J. C. (2015). The Role of Multivariate Skew-Student Density in the Estimation of Stock Market Crashes. *European Journal of Finance*, 21(13–14):1144–1160.

Zhu, Q. J., Bailey, D. H., López de Prado, M., and Borwein, J. M. (2017). The Probability of Backtest Overfitting. *Journal of Computational Finance*, 20(4):39–69.

Swiss Finance Institute

Swiss Finance Institute (SFI) is the national center for fundamental research, doctoral training, knowledge exchange, and continuing education in the fields of banking and finance. SFI's mission is to grow knowledge capital for the Swiss financial marketplace. Created in 2006 as a public–private partnership, SFI is a common initiative of the Swiss finance industry, leading Swiss universities, and the Swiss Confederation.

swiss:finance:institute

c/o University of Geneva, Bd. Du Pont d'Arve 42, CH-1211 Geneva 4
T +41 22 379 84 71, rps@sfi.ch, www.sfi.ch