

# Swiss Finance Institute Research Paper Series N°22-88

Density and Risk Prediction with Non-Gaussian  
COMFORT Models



**Marc S. Paoletta**  
University of Zurich and Swiss Finance Institute

**Pawel Polak**  
Stony Brook University

# Density and Risk Prediction with Non-Gaussian COMFORT Models\*

Marc S. Paoletta<sup>a,d</sup> Paweł Polak<sup>b,c</sup>

<sup>a</sup>*Department of Banking and Finance, University of Zurich, Switzerland*

<sup>b</sup>*Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, United States*

<sup>c</sup>*Institute for Advanced Computational Science, Stony Brook University, Stony Brook, United States*

<sup>d</sup>*Swiss Finance Institute*

November 11, 2022

## Abstract

The CCC-GARCH model, and its dynamic correlation extensions, form the most important model class for multivariate asset returns. For multivariate density and portfolio risk forecasting, a drawback of these models is the underlying assumption of Gaussianity. This paper considers the so-called COMFORT model class, which is the CCC-GARCH model but endowed with multivariate generalized hyperbolic innovations. The novelty of the model is that parameter estimation is conducted by joint maximum likelihood, of all model parameters, using an EM algorithm, and so is feasible for hundreds of assets. This paper demonstrates that (i) the new model is blatantly superior to its Gaussian counterpart in terms of forecasting ability, and (ii) also outperforms ad-hoc three-step procedures common in the literature to augment the CCC and DCC models with a fat-tailed distribution. An extensive empirical study confirms the COMFORT model's superiority in terms of multivariate density and Value-at-Risk forecasting.

**Keywords:** GJR-GARCH; Multivariate Generalized Hyperbolic Distribution; Non-Ellipticity; Value-at-Risk.

**JEL Classification:** C51; C53; G11; G17.

## 1 Dedication

Both authors, hereafter MP and PP, respectively, have known Michael for many years and were terribly saddened with his passing. We respectfully dedicate this paper to Michael, and wish

---

\*The authors wish to thank two anonymous referees, and the handling editor Alan Wong, for excellent feedback and suggestions that led to a clear improvement in our paper.

Financial support by the Swiss National Science Foundation (SNSF) - through project #150277 for Paoletta and Early Postdoc.Mobility grant #151631 for Polak - is gratefully acknowledged.

to explicitly note that the idea of addressing the limitations of (what we call below the) three-step ad hoc method commonly used in the parameter estimation of non-Gaussian DCC-GARCH constructs came from discussions with Michael, adding to his critiques of the DCC approach in Caporin and McAleer (2013). Michael’s scientific achievements do not require elaboration: A simple google scholar inspection reveals over 24,000 citations from 1,276 research papers (no less with many from 2021, despite his health issues); a feat few academics can claim.

MP had visited Michael (and Chia-Lin Chang, and their associated faculties) numerous times, in Taiwan, Hong Kong, Bangkok, and Chiang Mai. Michael visited Zurich several times, having given seminars in the economics department and also a doctoral level course in the finance department as per MP’s invite. Michael was instrumental in the our first COMFORT paper, Paoletta and Polak (2015): He was pleased with our approach and encouraged us to (first explore its applications to option pricing and) submit it to the special issue at the Journal of Econometrics that he guest edited, where it was eventually published. Michael, having understood our paper in depth, had numerous talks with us about some limitations of the model structure, such as the presence of only one latent “ $G_t$  sequence” (as opposed to setting it to a constant, which results in the Gaussian distribution), and the resulting homogeneity of the tails. This impetus from Michael led us to work on an extension addressing Michael’s correct concerns, resulting, after some years of work and some false starts, in Naf et al. (2019). Michael and Chia-Lin were also intrigued with some related methodology MP was working on a few years back, and asked MP to give the keynote speech at a conference they organized at the National Tsing Hua University in 2014. They also supported the eventual publication of the resulting work in Paoletta (2014).

We cannot overstate our gratitude to Michael for his academic support, his financial support, and, most importantly, his friendship.

## 2 Introduction

The non-normality of asset returns is now fully established. Numerous financial crashes of different scales have been observed in recent years, and part of the reason has been attributed to poor risk management practices, such as neglecting the non-normality of asset returns. In particular, after the recent financial crisis, expressions such as “the formula which killed Wall Street” and “Black-Swan” made their way into the general public. Indeed, the concept of non-Gaussian returns is no longer limited to academic circles. As Greenspan (1997, p. 54) states,

... the biggest problems we now have with the whole evaluation of risk is the fat-tail problem, which is really creating very large conceptual difficulties. Because as we all know, the assumption of normality enables us to drop off a huge amount of complexity of our equations very much to the right of the equal sign. Because once you start putting in non-normality assumptions, which unfortunately is what characterizes the real world, then the issues become extremely difficult.

The use of non-Gaussian distributions for the innovation sequence driving conditional multivariate models, such as constant conditional correlation generalized autoregressive conditional heteroskedasticity (CCC-GARCH), its dynamic CC (DCC-GARCH) extensions, and other multivariate constructions (see e.g., Bauwens et al., 2006 and the references therein), are not yet the

norm. While in the univariate case, models such Student's  $t$  GARCH and others have been available for years in common statistical and econometric software packages, this is not the case in the multivariate setting. Part of the reason is that it is considerably more difficult to estimate all the parameters jointly. Instead, ad hoc three-step procedures are used: The first two steps consist of using the Gaussian-based quasi-MLE to get the set of univariate GARCH parameters and, then, the DCC parameters; and, subsequently, the filtered residuals are used in the third step to estimate the shape parameters of the non-normal multivariate innovations distribution, typically the multivariate Student  $t$ , which is limited by being elliptic, and having the single degree of freedom parameter and thus ruling out heterogeneous tail behavior. One goal of this paper is, as Michael McAleer also noted, to demonstrate flaws with this approach, and propose an alternative, more coherent methodology that also results in superior out of sample performance.

The so-called COMFORT model (Common Market Factor Non-Gaussian Returns Model) of Paolella and Polak (2015) (hereafter PP14), develops an EM-type algorithm to jointly estimate all the model parameters by maximum likelihood. The baseline COMFORT model of PP14 can use the (and, in particular, special cases of the) multivariate generalized hyperbolic, and thus allow for non-ellipticity, but still imposes homogeneous tail behavior. This latter restriction was relaxed in Paolella et al. (2019) by using a regime switching structure on the correlation matrix; and in Näf et al. (2019) by allowing for heterogeneous tails. Both subsequent generalizations stay within the COMFORT framework, though the latter requires more elaborate distribution theory and inversion of characteristic functions (or, as we did, moment generating functions via saddlepoint approximations). Heterogeneous tails can also be successfully and straightforwardly achieved via copula-based constructs, though do not admit analytic representations of the resulting predictive portfolio distribution and instead require simulation; see Paolella and Polak (2018) for an expedient implementation.

The primary goals of this paper are (i) to extend the PP14 construct (which concentrated on the model formulation, parameter estimation method, and its use in option pricing) to a dynamic conditional correlation (DCC) context; and (ii) illustrate the advantage of joint parameter estimation as compared to the "cobbled together" ad hoc method commonly deployed. We make several contributions. The first is that we explicitly show how to correctly use the three-step procedure when estimating a non-Gaussian model. Next, an extensive empirical exercise involving the 30 assets comprising the DJIA demonstrates the general superior predictive performance of non-Gaussian models, compared to their Gaussian counterparts. Moreover, we also show that, in terms of out-of-sample density and Value at Risk (VaR) forecasting, the benefits of the non-Gaussian model class are achieved only when using the COMFORT model, i.e., the full MLE approach. This result can have important implications for market risk management and asset allocation, particularly in light of highly volatile, non-Gaussian market behavior. A final contribution involves shedding light on the value added in terms of forecasting ability of the GJR-GARCH model compared to standard GARCH.

The remainder of this paper is as follows. Section 3 is a brief note about partially adaptive estimation and motivation for considering non-Gaussian distributions. Section 4 discusses some particular features of the generalized hyperbolic that makes it very attractive for financial asset returns data. Section 5 reviews the model under consideration. Section 6 discusses the usual ad hoc procedure used for estimating a CCC or DCC model with non-Gaussian innovations, and

shows the required adjustment in this case to ensure consistent estimation. Section 7 provides an extensive empirical illustration of the model’s performance, both in- and out-of-sample, compared with the usual Gaussian alternatives and non-Gaussian models based on the three-step estimation technique. Section 8 concludes.

### 3 Partially Adaptive Estimation

Part of the appeal of using the normality assumption is model simplicity and ease of parameter estimation. Moreover, it is known that, under certain assumptions on the true data generating process, the method of quasi maximum likelihood (use of normality even though it is incorrect) still leads to a consistent estimate of the covariance matrix or, more generally, the parameters that govern its evolution through time (see, e.g., Francq and Zakoïan, 2004, 2010). This approach, however, can suffer in small samples because of the divergence of the Gaussian and true innovations distribution. In particular, a large deviation can cause an increase in the variance of the estimates, and result in efficiency loss. This is particularly acute in empirical finance applications, for which asset returns data are well-known to be generally heavy-tailed and mildly asymmetric, and even more so in the common case with a large number of assets compared to observations in time; see, e.g., Bickel and Levina (2008); Fan et al. (2008); and the references therein.

Furthermore, if interest centers not just on asymptotically consistent point estimates of some of the model parameters, but rather on the predictive distribution of the returns based on a potentially small finite sample, then the incorrect use of normality may have far-reaching consequences. In particular, modern quantities for portfolio risk, such as value-at-risk and expected shortfall, do not just depend on the mean vector and covariance matrix of returns, but on all the distributional parameters. Such risk measures are features of the predictive distribution, which is the most general object from which all measurable quantities of interest can be derived. Indeed, in both the univariate case (see, e.g., Kuester et al., 2006; Broda et al., 2012; and the references therein) and multivariate case (see, e.g., Paoletta, 2015 and the references therein), risk and density forecasts are greatly enhanced when moving from the normal to flexible, heavier-tailed, asymmetric distributions. Observe also that portfolio allocation that respects the use of downside-risk measures in the asymmetric case, and the leptokurtic nature of asset returns requires the use of non-Gaussian distributions; see, e.g., Doganoglu et al. (2007); Giacometti et al. (2007); and Broda and Paoletta (2009).

It is important to emphasize that the true underlying class of distributions need not be known, but only that the assumed distributional class is flexible enough to capture the salient features of the data (in finance, these being heavy tails and asymmetry, but otherwise a uni-modal, bell-like distribution which nests, or features as a limiting case, the normal distribution). The application of a parametric model that most likely differs from the true underlying one, but is flexible enough to adequately fit the data, is generally referred to as *partially adaptive estimation*; see, e.g., McDonald and Newey (1988); Phillips (1994); and Hansen et al. (2006). It is useful because, besides accounting for the often substantial departures from normality, it avoids some of the disadvantages of nonparametric inference (see, for example, McDonald, 1991, 1997), even if, as already mentioned, the underlying model is not precisely from the specified distributional class. The multivariate generalized hyperbolic (MGHyp) is essentially perfect for this task, given its

great flexibility, but also its other desirable features, as discussed next.

## 4 The Multivariate Generalized Hyperbolic

### 4.1 Construction and Properties

The MGHyp was introduced in Barndorff-Nielsen (1978); popularized as a candidate model for financial returns in Eberlein and Keller (1995), Eberlein et al. (1998), Wang (2009), and McNeil et al. (2015); and extended in Schmidt et al. (2006). Recent work on its development and applications include Fotopoulos et al. (2020), Bianchi et al. (2020), and Saliha and Aboudi (2021). It is a very general and extremely flexible (albeit uni-modal) class of distributions that, as special or limiting cases, includes the Gaussian, and asymmetric versions of the (multivariate) Student's  $t$ , Laplace, normal inverse Gaussian, variance-gamma, and others. An interesting feature of this distribution particularly relevant for financial returns data is that it has what are called semi-heavy tails; these being a compromise between power (fat) tails and exponential (thin) tails, and admitting a moment generating function. See Paoletta (2007, Ch. 9) for a detailed presentation in the univariate case. Moreover, the special and limiting cases mentioned above indicate that it also supports (in the limit) both genuine fat (no moment generating function) and thin tails. Theoretical support for using an MGHyp model comes from the fact that it is consistent with continuous-time models where logarithmic asset prices follow multivariate Lévy processes; see, e.g., Eberlein and Keller (1995).

From an empirical perspective, it has been shown (for both univariate and multivariate data) that the MGHyp provides an excellent fit to the unconditional distribution of financial returns; see, e.g., McNeil et al. (2015) and the reference therein. However, the extension from normality to MGHyp entails estimation of yet more model parameters. Far worse, however, is that, in high dimensions (irrespective of the distribution), estimation via direct optimization of the likelihood becomes problematic because of the proliferation of parameters in the dispersion matrix and, for the COMFORT model, the GARCH parameters governing the time-varying evolution of each of the univariate processes. To address this, a multi-stage Expectation Maximization (EM) algorithm is used for maximum likelihood estimation. This is possible because the MGHyp is expressible as a continuous normal mixture, as reviewed below. Crucially, the new estimation method is applicable for a large number of assets.<sup>1</sup>

Let the  $K$ -variate vector  $\mathbf{Y} = (Y_1, \dots, Y_K)'$  have representation

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\gamma}G + \mathbf{H}^{1/2}\sqrt{G}\mathbf{Z}, \quad (1)$$

where  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_K)'$  and  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_K)'$  are column vectors in  $\mathbb{R}^K$ ;  $\mathbf{H}$  is a positive definite, symmetric, dispersion matrix of order  $K$ ;  $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I}_K)$ ; and  $G \sim GIG(\lambda, \chi, \psi)$  is a generalized inverse Gaussian random variable, independent of  $\mathbf{Z}$ . The GIG pdf is given by

$$f_{GIG}(x; \lambda, \chi, \psi) = \frac{x^{\lambda-1}}{k_\lambda(\chi, \psi)} \exp\left[-\frac{1}{2}(\chi x^{-1} + \psi x)\right] \mathbb{I}_{(0, \infty)}(x), \quad (2)$$

---

<sup>1</sup>Blæsild and Sørensen (1992) developed a program for estimating MGHyp distributions, for which Prause (1999) reports that it can estimate the general case up to three dimensions in a reasonable time. An EM algorithm was proposed by Protassov (2004). All this work was for iid data, and does not include any type of multivariate conditional volatility modeling.

with parameter space by  $\Theta_{\text{GIG}}$ , which consists of three cases given by

$$\begin{aligned}\Theta_{\text{GIG}} := \{(\lambda, \chi, \psi) \in \mathbb{R}^3 : & \quad \lambda \in \mathbb{R}, \chi > 0, \psi > 0 \quad (\text{normal case}), \\ & \text{or} \quad \lambda > 0, \chi = 0, \psi > 0 \quad (\text{boundary case I}), \\ & \text{or} \quad \lambda < 0, \chi > 0, \psi = 0\} \quad (\text{boundary case II}),\end{aligned}$$

and where

$$k_\lambda(\chi, \psi) = 2(\chi/\psi)^{\lambda/2} K_\lambda(\sqrt{\chi\psi}), \quad (3)$$

with  $K_\nu(\cdot)$  being the modified Bessel function of the third kind with index  $\nu$ , given by

$$K_\nu(x) = \frac{1}{2} \int_0^\infty t^{\nu-1} e^{-\frac{1}{2}x(t+t^{-1})} dt, \quad x > 0. \quad (4)$$

Via the location-scale continuous Gaussian mixture structure of the MGhyp,

$$f_{\mathbf{Y}}(\mathbf{y}; \lambda, \chi, \psi, \boldsymbol{\mu}, \mathbf{H}, \boldsymbol{\gamma}) = \int_0^\infty f_{\mathbf{Y}|G}(\mathbf{y} | g; \boldsymbol{\mu}, \mathbf{H}, \boldsymbol{\gamma}) f_G(g; \lambda, \chi, \psi) dg; \quad (5)$$

where  $(\mathbf{Y} | G) \sim N(\boldsymbol{\mu} + \boldsymbol{\gamma}G, \sqrt{G}\mathbf{H}^{1/2})$  and  $G \sim \text{GIG}(\lambda, \chi, \psi)$ .

Observe that component  $G_t \sim \text{GIG}(\lambda, \chi, \psi)$  is common to all margins, implying that the components of  $\mathbf{Y}$  can only be independent in the limiting, Gaussian case. The latent mixing component  $G_t$  can be treated as a so-called *common market factor*. This factor, and how it differs from the volatility persistence driven by the GARCH equations, will be elaborated upon in Section 7.1. Importantly, PP14 shows that it is amenable to prediction, and is the key to allowing the model to embody a stochastic volatility (SV) extension.

Another theoretical property of great importance for portfolio optimization is that, if the vector of asset returns is MGhyp distributed, then the distribution of the portfolio return (a weighted sum of the margins) is univariate GHyp with location and asymmetry parameters being weighted (by portfolio weights) sums of the corresponding vector parameters and the scale parameter being a quadratic form of the corresponding dispersion matrix; see McNeil et al. (2015). That is, now endowing the relevant parameters with time subscript  $t$  in anticipation of our chosen model formulation in Section 5, and letting  $\Phi_{t-1}$  denote the available information set up to time  $t-1$ , if  $(\mathbf{Y}_t | \Phi_{t-1}) \sim \text{MGhyp}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \mathbf{H}_t, \lambda_t, \chi_t, \psi_t)$ , then the distribution of the portfolio return  $P_t = \mathbf{w}'\mathbf{Y}_t$ , with a vector of portfolio weights  $\mathbf{w} \in \mathbb{R}^K \setminus \mathbf{0}$ , is given by

$$(P_t | \Phi_{t-1}) \sim \text{GHyp}(\mathbf{w}'\boldsymbol{\mu}, \mathbf{w}'\boldsymbol{\gamma}, \mathbf{w}'\mathbf{H}_t\mathbf{w}, \lambda_t, \chi_t, \psi_t). \quad (6)$$

Of interest is the limiting behavior of the GHyp density when  $x \rightarrow \pm\infty$ . Define for convenience

$$\mu_{\mathbf{w}} = \mathbf{w}'\boldsymbol{\mu}, \quad \gamma_{\mathbf{w}} = \mathbf{w}'\boldsymbol{\gamma}, \quad \sigma_{t,\mathbf{w}}^2 = \mathbf{w}'\mathbf{H}_t\mathbf{w}. \quad (7)$$

From Barndorff-Nielsen (1997) and Prause (1999, eq. 1.19)), the limiting distribution is characterized by

$$f_{P_t}(x; \mu_{\mathbf{w}}, \gamma_{\mathbf{w}}, \sigma_{t,\mathbf{w}}^2, \lambda_t, \chi_t, \psi_t) \propto |x|^{\lambda_t-1} \exp\left(-\sqrt{\frac{\psi_t + \gamma_{\mathbf{w}}/\sigma_{t,\mathbf{w}}^2}{\sigma_{t,\mathbf{w}}^2}} |x| + \frac{\gamma_{\mathbf{w}}}{\sigma_{t,\mathbf{w}}^2} x\right). \quad (8)$$

Observe how it is a product of the power-tail component  $|x|^{\lambda_t-1}$  and the exponential. The latter is described by the following three parameters: (i) the dispersion matrix,  $\mathbf{H}_t$  (with GARCH effects

and the correlation matrix  $\boldsymbol{\Gamma}$  introduced below); (ii) the asymmetry vector  $\boldsymbol{\gamma}$ ; and (iii) the GIG distribution parameter,  $\psi_t$ . The power-tail factor is described by only one parameter,  $\lambda_t$ . As this parameter is estimated, an estimate of the tail thickness of the data can be captured by the model, *under the assumption of the assumed data generating process*. The folly of attempting to estimate the maximally existing moment of data from parametric assumptions, and the inaccuracies of attempting to do so via semi-nonparametric tail estimation procedures, are discussed at length in Paolella (2018, Ch. 9).

## 4.2 Special Cases of the MGHyp

As reported by Protassov (2004), and also confirmed by our studies, one or more of the MGHyp shape parameters can have a relatively flat likelihood (already after fixing  $\chi$  or  $\psi$  for identification purposes), implying possible numeric problems when maximizing the likelihood. While the class of MGHyp models is identified, it is too general for use with typical sample sizes, and so we advocate use of special cases. In particular, we use the multivariate asymmetric Laplace (MALap), this distribution being also referred to as the variance gamma (VG; see below), the multivariate normal inverse Gaussian (MNIG), and the multivariate asymmetric  $t$ -distribution (MAt). Each allows for individual asset asymmetry parameters, as well as higher kurtosis than the normal distribution (in the case of existing fourth moments). Note how most of these special cases have a moment generating function, obviating explicit concern about existence of moments.

Fixing some of the MGHyp distribution parameters in a judiciously chosen way can result in a distribution that is not only virtually as capable of modeling the features of financial data as the full general MGHyp case, but can even result in superior results in the same way that parameter shrinkage results in lower mean squared error. To this end, we use three special cases of the MGHyp distribution in which: (i)  $G_t$  is gamma distributed with shape parameter  $\lambda > 0$  and unit scale parameter (the multivariate asymmetric Laplace, or MALap, model); (ii) the  $\lambda$  parameter is fixed at  $-1/2$  (the multivariate normal inverse Gaussian, MNIG, model); and (iii)  $G_t$  is inverse gamma distributed with the scale and the shape parameters both equal to  $v/2$  (equivalently  $G_t$  is GIG with  $\lambda = -0.5v$ ,  $\chi = v$  and  $\psi = 0$ ) where  $v, v > 0$ , is the parameter being estimated (the multivariate asymmetric  $t$ -distribution, MAt, model).

The three special cases do not share the flat likelihood problem of the fully general MGHyp distribution, and so are considerably faster and numerically more reliable to estimate, but still retain the flexibility required for modeling asset returns by allowing for individual asset asymmetry parameters and also (semi-)heavy tail behavior. Section 4.2.1 details the derivation and some properties of the MALap (or VG) density, while Sections 4.2.2 and 4.2.3 examine the MNIG and MAt distributions, respectively.

### 4.2.1 Laplace

The multivariate asymmetric Laplace, or MALap, density can be derived by evaluating the integral

$$f_{\mathbf{Y}}(\mathbf{y}; \lambda, \boldsymbol{\mu}, \mathbf{H}, \boldsymbol{\gamma}) = \int_0^\infty f_{\mathbf{Y}|G}(\mathbf{y} | g; \boldsymbol{\mu}, \mathbf{H}, \boldsymbol{\gamma}) f_G(g; \lambda) dg; \quad (9)$$

where  $(\mathbf{Y} | G) \sim N(\boldsymbol{\mu} + \boldsymbol{\gamma}G, \sqrt{G}\mathbf{H}^{1/2})$  and  $G \sim \text{Gam}(\lambda, 1)$ . Alternatively, one can simplify the general MGHyp density with  $\lambda > 0$ ,  $\chi = 0$  and  $\psi = 2$ , by use of the limiting relation

$\mathcal{K}_\lambda(\sqrt{(2\chi)}) \simeq \Gamma(\lambda)2^{\lambda-1}(\sqrt{2\chi})^{-\lambda}$  for  $\chi \downarrow 0$ ,  $\lambda > 0$  (Paoletta, 2007, Eq. 9.6), where  $\mathcal{K}_\lambda(x)$  is Bessel function (4) and  $\Gamma$  is the gamma function. Both approaches result in (9) being given by

$$\frac{2 \exp\{(\mathbf{y} - \boldsymbol{\mu})' \mathbf{H}^{-1} \boldsymbol{\gamma}\}}{(2\pi)^{K/2} \Gamma(\lambda) |\mathbf{H}|^{1/2}} \left( \frac{m}{2 + \boldsymbol{\gamma}' \mathbf{H}^{-1} \boldsymbol{\gamma}} \right)^{\lambda/2-K/4} \mathcal{K}_{\lambda-K/2} \left( \sqrt{m(2 + \boldsymbol{\gamma}' \mathbf{H}^{-1} \boldsymbol{\gamma})} \right), \quad (10)$$

where  $\mathbf{H}$  is the dispersion matrix and  $m = (\mathbf{y} - \boldsymbol{\mu})' \mathbf{H}^{-1} (\mathbf{y} - \boldsymbol{\mu})$ . It generalizes the distribution proposed and used in Paoletta (2015), which is (10) but without the asymmetry term  $\boldsymbol{\gamma}$ . It turns out that our discovery is not new: a symmetric and univariate version of this distribution was introduced by McKay (1932); it was extended to the multivariate case, called variance-gamma and used in finance by Madan and Seneta (1990); it was also mentioned in Kotz et al. (2000, 2001) and Podgórski and Kozubowski (2001) as a generalized Laplace, but they concentrated on further special cases of it; for the univariate asymmetric version of it, see Paoletta (2007, Ch. 9).

Use of the MALap distribution in dynamic conditional correlation models is not new and appears to have been first proposed by Cajigas and Urga (2007), though our model differs significantly from their setup. In particular, Cajigas and Urga (2007) fixed all the parameters of the mixing random variable  $G_t$  prior to estimation. See Rombouts et al. (2014) and the references therein for the use of the Laplace distribution in the context of multivariate volatility models for option pricing.

#### 4.2.2 Normal Inverse Gaussian

The multivariate normal inverse Gaussian (MNIG) distribution arises from the MGHyp for  $\lambda = -1/2$ , with density

$$f_{\mathbf{Y}}(\mathbf{y}; \boldsymbol{\mu}, \mathbf{H}, \boldsymbol{\gamma}, \chi, \psi) = Cd^{-(K+1)/2} \mathcal{K}_{(K+1)/2}(d) \exp\{(\mathbf{y} - \boldsymbol{\mu})' \mathbf{H}^{-1} \boldsymbol{\gamma}\}, \quad (11)$$

for  $d = \sqrt{(\chi + (\mathbf{y} - \boldsymbol{\mu})' \mathbf{H}^{-1} (\mathbf{y} - \boldsymbol{\mu}))(\psi + \boldsymbol{\gamma}' \mathbf{H}^{-1} \boldsymbol{\gamma})}$  and the normalizing constant

$$C = \frac{(\chi/\psi)^{1/4} (\psi + \boldsymbol{\gamma}' \mathbf{H}^{-1} \boldsymbol{\gamma})^{(K+1)/2}}{(2\pi)^{K/2} |\mathbf{H}|^{1/2} \mathcal{K}_{1/2}(\sqrt{\chi\psi})}.$$

This distribution has been used in the independent and identically distributed (iid) case (McNeil et al., 2015) and also for building a multivariate predictive distribution via univariate NIG-GARCH components (Broda and Paoletta, 2009). In our model, for identification purposes, we additionally fix  $\psi = 1$ . In our forecasting study below, we show that, while use of the MNIG is far better than use of the normal, the model using the MALap distribution and the hybrid dynamics results in better performance.

#### 4.2.3 Asymmetric $t$ Distribution

The multivariate asymmetric (Student)  $t$ , or MAt, is a limiting case of the MGHyp, with  $\lambda = -1/2v$ ,  $\chi = v$  and  $\psi = 0$ , for some  $v > 0$ . Evaluating the limit of the MGHyp probability density function as  $\psi \rightarrow 0$  gives the density

$$f_{\mathbf{Y}}(\mathbf{y}; \boldsymbol{\mu}, \mathbf{H}, \boldsymbol{\gamma}, v) = C \frac{\mathcal{K}_{(v+K)/2}(\sqrt{(v+m)\boldsymbol{\gamma}' \mathbf{H}^{-1} \boldsymbol{\gamma}}) \exp((\mathbf{y} - \boldsymbol{\mu})' \mathbf{H}^{-1} \boldsymbol{\gamma})}{\left(\sqrt{(v+m)\boldsymbol{\gamma}' \mathbf{H}^{-1} \boldsymbol{\gamma}}\right)^{-(v+K)/2} (1+m/v)^{(v+K)/2}}, \quad (12)$$

where  $m = (\mathbf{y} - \boldsymbol{\mu})' \mathbf{H}^{-1} (\mathbf{y} - \boldsymbol{\mu})$ , and normalizing constant

$$C = \frac{2^{1-(v+K)/2}}{\Gamma(v/2) (\pi v)^{K/2} |\mathbf{H}|^{1/2}}.$$

This density reduces to the standard multivariate  $t$  with  $v/2$  degrees of freedom, as  $\gamma \rightarrow 0$ .

McNeil et al. (2015, Ch. 3) mention this distribution as having potential applications in finance, while Aas and Haff (2006) work with the univariate version of (12), and show that it is the only subclass of the GHyp family that has the property of different asymptotic left and right tail behavior. The heavier tail has a power decay and the lighter tail is a product of a power and an exponential function. Aas and Haff (2006) show that, in the univariate case and under an iid assumption, it leads to superior data fit than some other competitors, including the NIG distribution. Jondeau (2012) modifies this distribution and uses it to show the importance of asymmetry in the tail dependence of equity portfolios.

## 5 Model Formulation

The  $K$ -variate return vector  $\mathbf{Y}_t = (Y_{t,1}, \dots, Y_{t,K})'$  is assumed to follow a conditional MGhyp distribution with representation

$$\mathbf{Y}_t = \boldsymbol{\mu} + \gamma G_t + \boldsymbol{\varepsilon}_t, \quad \text{with} \tag{13a}$$

$$\boldsymbol{\varepsilon}_t = \mathbf{H}_t^{1/2} \sqrt{G_t} \mathbf{Z}_t, \tag{13b}$$

where  $\boldsymbol{\mu}, \gamma \in \mathbb{R}^K$ ;  $\mathbf{H}_t$  is a positive definite, symmetric, dispersion matrix of order  $K$  for all  $t$ ;  $\mathbf{Z}_t \stackrel{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_K)$  is a sequence of independent and identically distributed (iid) normal random variables; and  $(G_t | \Phi_{t-1}) \sim \text{GIG}(\lambda_t, \chi_t, \psi_t)$  are mixing random variables,  $t = 1, 2, \dots, T$ , independent of  $\mathbf{Z}_t$ , with information set  $\Phi_{t-1} = \{\mathbf{Y}_1, \dots, \mathbf{Y}_{t-1}\}$ . We assume  $\mathbf{Y} = [\mathbf{Y}_1 | \mathbf{Y}_2 | \dots | \mathbf{Y}_T]$  are equally spaced (for daily, ignoring the weekend effect) random variables representing realization of the return vector.

For the GIG parameters, PP14 propose two specifications:

1. the iid case, where  $G_t$  are iid with time-invariant parameters, i.e.,  $\lambda_t = \lambda$ ,  $\chi_t = \chi$  and  $\psi_t = \psi$ ; and
2. the stochastic volatility case, where  $G_t | \Phi_{t-1}$  has time dependent parameters with the dynamics described by a system of conditional moment equations

$$\mathbb{E}[G_t^r | \Phi_{t-1}] = c_r + \rho_r \mathbb{E}[G_{t-1}^r | \Phi_{t-2}] + \sigma_r \zeta_{r,t}, \tag{14}$$

for a set of positive integer values of  $r$ ;  $\zeta_{r,t} = \mathbb{E}[G_t^r | \Phi_t] - \mathbb{E}[G_t^r | \Phi_{t-1}]$ ; and  $c_r$ ,  $\rho_r$  and  $\sigma_r$  are parameters to be estimated.

As discussed in PP14,  $\zeta_{r,t}$  represents the unpredictable component affecting inference about the  $r$ th moment of the mixing variable  $G_t$ . It is a martingale difference sequence (MDS) with respect to  $\Phi_{t-1}$ , thus  $\mathbb{E}[\zeta_{r,t}] = 0$ ,  $\text{Cov}(\zeta_{r,t}, \zeta_{r,t-s}) = 0$ ,  $s = 1, 2, \dots$ , and it can be used as a driver of the dynamics in (14). The model with formulation (14) is called a hybrid

GARCH-Stochastic Volatility (GARCH-SV) extension. The link between this and the SV model of Taylor (1982) is detailed in PP14.

Due to the MDS property of the  $\zeta_{r,t}$  innovations, the conditional forecasts of the future conditional moments of  $G_t$  are given by

$$\mathbb{E}[G_{t+s}^r | \Phi_t] = c_r \sum_{i=0}^{s-1} \rho_r^i + \rho_r^s \mathbb{E}[G_t^r | \Phi_{t-1}], \quad s \geq 1, \quad (15)$$

where  $\mathbb{E}[G_t^r | \Phi_{t-1}]$  is measurable with respect to the information up to time  $t-1$  and is given by

$$\mathbb{E}[G_t^r | \Phi_{t-1}] = \left( \frac{\chi_t}{\psi_t} \right)^{r/2} \frac{\mathcal{K}_{\lambda_t+r}(\sqrt{\chi_t \psi_t})}{\mathcal{K}_{\lambda_t}(\sqrt{\chi_t \psi_t})}, \quad r \in \mathbb{R}, \quad (16)$$

which involves a ratio of Bessel functions  $\mathcal{K}_v(\cdot)$ . If  $|\rho_r| < 1$ , then the process in (14) is mean-reverting, and for  $s \rightarrow \infty$ , the forecast approaches the unconditional mean value  $c_r / (1 - \rho_r)$  of  $G_t^r$ .

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

$$\mathbf{H}_t \equiv \mathbf{S}_t \boldsymbol{\Gamma} \mathbf{S}_t, \quad (17)$$

where  $\mathbf{S}_t$  is a diagonal matrix composed of the strictly positive conditional scale terms  $s_{k,t}$ ,  $k = 1, \dots, K$ , and  $\boldsymbol{\Gamma}$  is a dependency matrix. The univariate scale terms  $s_{k,t}$  are modeled by a GARCH-type process, e.g., the GARCH(1,1) model

$$s_{k,t}^2 = \omega_k + \alpha_k \varepsilon_{k,t-1}^2 + \beta_k s_{k,t-1}^2, \quad (18)$$

where  $\varepsilon_{k,t} = y_{k,t} - \mu_k - \gamma_k G_t$  is the  $k$ th element of the  $\boldsymbol{\varepsilon}_t$  vector in (13). More general formulations could be used. In particular, we consider also the GJR-GARCH(1,1) model of Glosten et al. (1993), which can capture an asymmetry effect, with the dynamics given by

$$s_{k,t}^2 = \omega_k + \alpha_k \varepsilon_{k,t-1}^2 + \eta_k \varepsilon_{k,t-1}^2 \mathbf{1}_{(-\infty, 0)}(\varepsilon_{k,t-1}) + \beta_k s_{k,t-1}^2, \quad (19)$$

where the indicator function  $\mathbf{1}_A(x)$  is equal to one if  $x \in A$  and zero otherwise, and  $\eta_k$  captures asymmetry in the scale-term response to the last period innovation. The constant in the GARCH (and GJR-GARCH) recursion is enforced to be positive, and all other parameters are constrained to be non-negative (including the asymmetry parameter in the GJR-GARCH case); for further details about the GJR-GARCH model and its relationship to asymmetry and leverage, see McAleer (2014).

The conditional mean and covariance matrix of  $\mathbf{Y}_t$ , implied by the model (13), are given by

$$\mathbb{E}[\mathbf{Y}_t | \Phi_{t-1}] = \boldsymbol{\mu} + \mathbb{E}[G_t | \Phi_{t-1}] \boldsymbol{\gamma} \quad (20)$$

and

$$\text{Cov}(\mathbf{Y}_t | \Phi_{t-1}) = \mathbb{E}[G_t | \Phi_{t-1}] \mathbf{H}_t + \mathbb{V}(G_t | \Phi_{t-1}) \boldsymbol{\gamma} \boldsymbol{\gamma}', \quad (21)$$

respectively, where  $\mathbb{V}(G_t | \Phi_{t-1}) = \mathbb{E}[G_t^2 | \Phi_{t-1}] - (\mathbb{E}[G_t | \Phi_{t-1}])^2$ .

For the evaluation of risk measures such as Value at Risk, the generalized hyperbolic class (obviously including its special cases, but also the limiting case of the MAT) is closed under linear

operations, so that the density of the portfolio  $P_t = \mathbf{w}'\mathbf{Y}_t$  is given by (6) and (7). Given the conditional density of the portfolio  $f_{P_t}$ , the VaR, at the level  $\alpha$ , can be computed as

$$\text{VaR}_{\alpha}^{t|t-1}(P_t) = -Q_{\alpha}(P_t | \Phi_{t-1}) = -\inf \{x \in \mathbb{R} : \mathbb{P}(P_t \leq x | \Phi_{t-1}) \geq \alpha\}, \quad (22)$$

where  $Q_{\alpha}$  denotes the  $\alpha$ -quantile function.

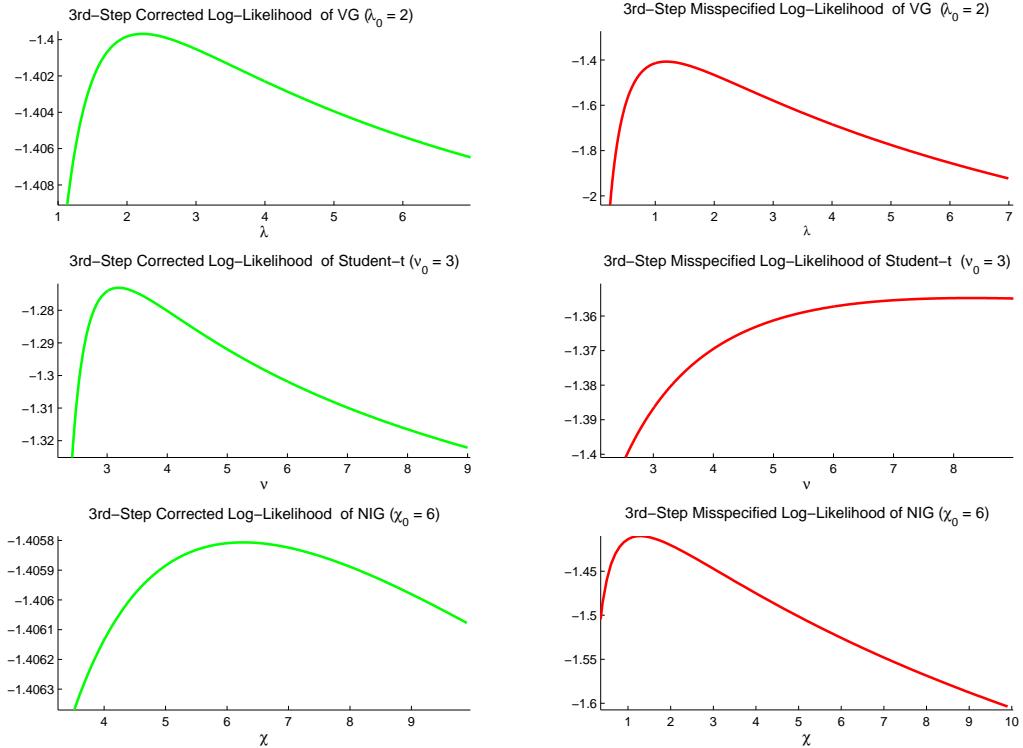
Estimation of the COMFORT model is detailed in PP14. It is an easily programmed EM algorithm, and thus exhibits reasonable estimation time for even hundreds of assets, and results in the jointly estimated maximum likelihood estimator, which, under standard regularity conditions, enjoys the usual desirable asymptotic properties of the MLE.

## 6 Multi-Step Estimation Procedures

An obvious, albeit ad hoc way commonly used to fit a non-Gaussian DCC-GARCH model is to use a three-step estimation method. First, apply the usual two-step Gaussian quasi-MLE from Bollerslev (1990) and Engle (2002) for the DCC-GARCH parameters, and thus obtain the mean vector and conditional predictive covariance matrix. Next, from the filtered residuals, estimate the remaining shape parameters of the multivariate distribution. Examples include Orskaug (2009), who consider the DCC model with multivariate standard and skewed Student's  $t$ , and Santos et al. (2013), who compare Value at Risk prediction methods for an equally weighted portfolio setting across and between univariate and multivariate models. In one multivariate model, the latter authors use the aforementioned two-step Gaussian-based procedure, and, from the filtered residuals, estimate the degrees of freedom parameter,  $d > 0$ , for the standard multivariate Student's  $t$  distribution. Observe in this special case the well known result, for  $d > 2$ , the conditional covariance matrix  $\Sigma_t$  is equal to  $d/(d-2)\mathbf{H}_t$ . This needs to be accounted for in order to estimate  $\mathbf{H}_t$  consistently. More generally, for the special, elliptical case of  $\gamma = \mathbf{0}$ , (21) implies that  $\Sigma_t = \mathbb{E}[G_t]\mathbf{H}_t$ . In the more relevant case of nonzero  $\gamma$ , the conditional mean and covariance matrix, given in (20) and (21), are linear combinations of location and skewness vectors, and dispersion and  $\gamma\gamma'$  matrices, respectively, and such a three-step estimation method is no longer applicable, because the likelihood function cannot be directly expressed; see PP14 for discussion.

We now demonstrate the effects of ignoring this correction factor in the  $\gamma = \mathbf{0}$  case. Observe that the first two steps provide the estimates of the mean vector and the (conditional) covariance matrix. If  $\mathbf{Y}_t$  has the dynamics given in (13) with  $\gamma = \mathbf{0}$ , then we can estimate the mean  $\mu$ , and the conditional covariance  $\Sigma_t$ , by a Gaussian quasi-MLE, and in the third step estimate the remaining distributional shape parameters from the residuals, which are given by  $\widehat{\Sigma}_t^{-1/2}(\mathbf{Y}_t - \widehat{\mu})$ . What we refer to as the third-best approach is to assume that these residuals have the dynamics (13) with  $\mu = \mathbf{0}$ ,  $\gamma = \mathbf{0}$ , and  $\mathbf{H}_t = \mathbf{I}_K$ . However, in general  $\mathbf{H}_t \neq \Sigma_t$ , and this approach leads to biased estimates. As mentioned, for  $\gamma = \mathbf{0}$ , we have  $\Sigma_t = \mathbb{E}[G_t]\mathbf{H}_t$ . Based on this, the residuals have the dynamics in (13) with  $\mu = \mathbf{0}$ ,  $\gamma = \mathbf{0}$ , and  $\mathbf{H}_t = \mathbf{I}_K\mathbb{E}[G_t]^{-1}$ , where  $\mathbb{E}[G_t]$ , as a special or limiting case of (16), is a function of the GIG parameters only. In the three special cases of Section 4.2, it is equal to  $\lambda$ ,  $\sqrt{(\chi/\psi)}$ , and  $v/(v-4)$ , respectively. Because of this simple structure, it can be easily incorporated into the third estimation step.

Ignoring the correction factor,  $\mathbb{E}[G_t]^{-1}$ , in the third step of estimation leads to biased estimates. Figure 1 shows that the maximum of the log-likelihood function can be strikingly different,

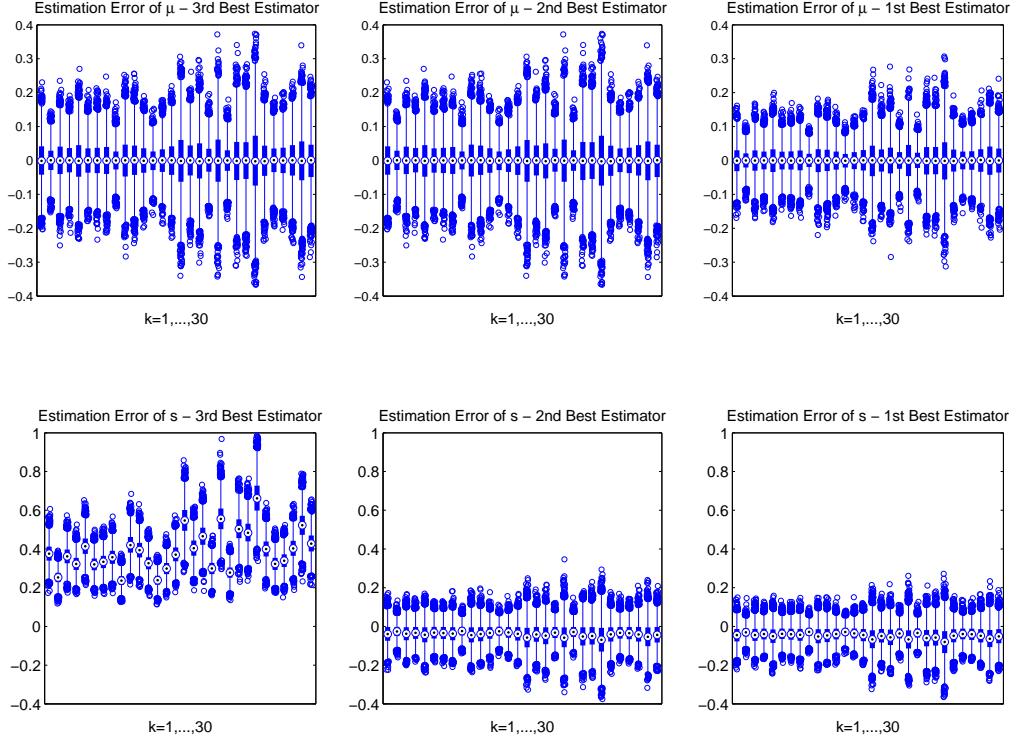


**Figure 1:** Third step log-likelihood plots for three special cases of the GHyp distribution considered in Section 4.2. The plots are generated based on a random sample of 10,000 observations from the given distribution. **Left column:** Corrected log-likelihood. **Right column:** Misspecified log-likelihood.

if the correction is not introduced. We call the three-step estimator, with the correction factor, the second-best approach. The first-best approach is to conduct joint ML parameter estimation via the EM algorithm. Under suitable regularity conditions, this approach enjoys the usual asymptotic optimality properties of the MLE. This benefit also translates into actual performance improvement: In Section 7, we empirically compare the first-best and second-best approaches on real data, and demonstrate that use of the former leads to much better in-sample and out-of-sample performance in terms of not only density forecasting measures, but also in terms of portfolio VaR forecasts.

Figures 2 and 3 illustrate the performance of the three estimation approaches for the elliptical variance-gamma (or multivariate Laplace) iid model. We compare the estimation errors, based on 10,000 simulations of  $T = 1,000$  observations with dimension  $K = 30$ , with the true mean vector and covariance matrix taken to be the sample estimates from the returns data considered in Section 7, and the  $\lambda$  parameter set to 2, this being a value very close to its estimate from the same returns data.

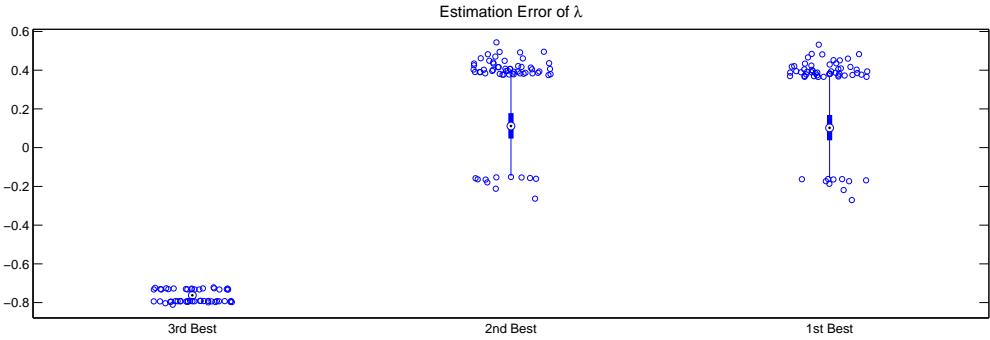
The first row in Figure 2 compares the mean estimates. All the approaches yield practically unbiased results, with the third-best and the second-best approach being identical. The estimates from the first-best approach, however, have a lower variance. The second row in Figure 2 compares the scale estimates. The third-best approach gives highly biased values, while the bias correction



**Figure 2:** Boxplots of the estimation error of parameter estimates for three estimators. The boxplots are generated based on simulations with 10,000 draws of random samples from MLap distribution with  $T = 1,000$  and  $K = 30$ . **First Row:** Estimates of the mean  $\mu$ . **Second Row:** Estimates of the scale terms  $s_k$  for  $k = 1, \dots, 30$ . **First column:** Third-best approach with the biased estimates in the third step and the scales estimated by standard deviation from the Gaussian likelihood scaled by the biased third step estimates. **Second column:** Second-best approach with the unbiased third step and the scales estimated by standard deviation from the Gaussian likelihood scaled by unbiased third step estimates. **Third column:** First-best approach with unbiased estimates and lower variance.

in the second-best approach considerably improves the estimates and the first-best approach is almost identical to the second-best. The estimates of  $\Gamma$  are very similar for all the three estimators, and we omit them.

The estimates of  $\lambda$  are compared in Figure 3. The third-best estimator leads to biased estimates, while the second-best and the first-best are essentially unbiased and have almost identical variance. The results rule out the third-best estimator and suggest that the remaining two approaches are very competitive. However, in models with dynamic volatilities and correlations, the second-best estimator has some disadvantages over the first-best approach. In particular, (i) it uses, in the first two steps, only a Gaussian quasi-MLE that ignores the impact of the latent component on the returns, and on the innovations driving the GARCH; (ii) because it lacks the iteration sequence in the estimation, it does not capture the feedback effect between the latent component and the dynamics in the consecutive steps of estimation; and (iii) it does not lend itself to allow for the non-ellipticity in the data.



**Figure 3:** Boxplots of the estimation error of the GIG parameter for three estimators. The boxplots are generated based on simulations with 10,000 draws of random samples from MLap distribution with  $T = 1,000$  and  $K = 30$ . **First column:** Third-best approach with the biased estimates. **Second column:** Second-best approach with the unbiased third step. **Third column:** First-best approach. Observe the estimates are unbiased and also exhibit a slightly lower variance compared to the second-best approach.

## 7 Empirical Results

To demonstrate the applicability and competitiveness of the model, we use the data set consisting of the 2,767 daily returns of  $K = 30$  components of the Dow Jones Industrial Index (DJ-30) from January 2nd, 2001, to December 30th, 2011 (based on the DJ-30 composition as of June 8th, 2009). 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 price of asset  $k$  at time  $t$ .

We first compare the in-sample fit of the usual, multivariate normal CCC (MN-CCC) model to the new MALap-CCC model, and show that the latter provides a much better fit to the tails of the return distribution. Next, we discuss the impact of the common market factor on the conditional volatilities. We conclude with the implications of the hybrid GARCH-SV extensions of our model.

We then compare the density forecasting performance across different models in Section 7.2. Summarizing the results, the hybrid GARCH(1,1)-SV COMFORT models, the elliptical (estimated with full MLE via the EM algorithm) models and the MAT-CCC GARCH(1,1) model deliver the best density predictions among all considered models. In particular, we find (i) a large improvement moving from the Gaussian to any of the distributions discussed in Section 4.2; and (ii) that the introduction of the dynamics in the  $G_t$  parameters (hybrid GARCH-SV model) further improves the forecasting performance. Both of these improvements are statistically significant. The overall best performing model is MALap-CCC hybrid GARCH(1,1)-SV.

Section 7.3 compares the implied mean estimates from the model with simpler estimators such as sample mean and sample median. Finally, Section 7.4 investigates the VaR performance of the models for an equally weighted portfolio of assets and compares it with the multivariate density predictions. Briefly, we find that that MNIG models with GJR-GARCH(1,1) dynamics perform the best in VaR forecasting.

## 7.1 In-Sample Performance

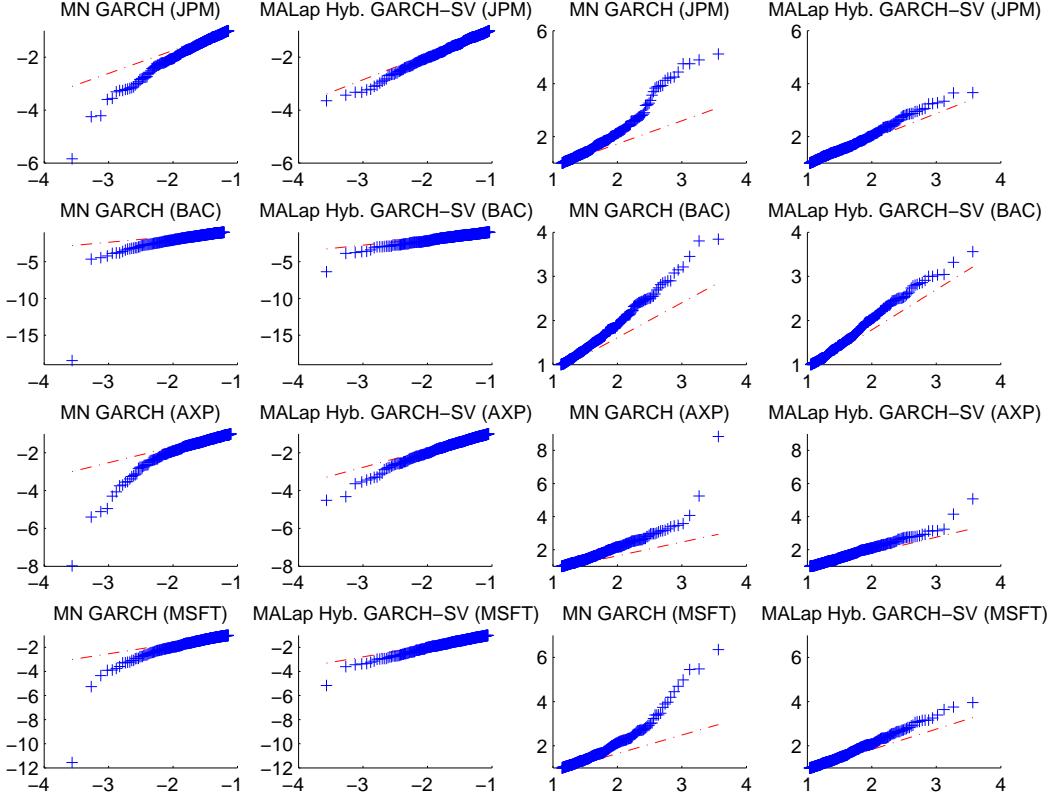
We estimate the MN-CCC and the MALap-CCC models for the whole data sample and compare the in-sample fit by inspecting the Q-Q-plots (the sample quantiles of the standardized residuals versus the theoretical quantiles from a normal distribution) of the resulting estimated standardized residuals given by

$$\widehat{\mathbf{H}}_t^{-1/2}(\mathbf{Y}_t - \widehat{\boldsymbol{\mu}}) \quad \text{and} \quad \widehat{G}_t^{-1/2}\widehat{\mathbf{H}}_t^{-1/2}(\mathbf{Y}_t - \widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\gamma}}\widehat{G}_t), \quad (23)$$

respectively; where  $\widehat{G}_t$  are the imputed values of  $G_t$  returned from the EM algorithm,  $\widehat{\mathbf{H}}_t$  are fitted conditional dispersion matrices, and other hatted entries denote parameter estimates. Observe that both sets of residuals in (23), in particular the latter, are assumed to be Gaussian under each of the assumed models. In Figures 4 and 5 we provide the Q-Q plots of the residuals of three (Gaussian, MALap, and MAt based) competitive models, for JPMorgan Chase & Co, Bank of America, American Express, and Microsoft Corp, based on the entire sample of  $T = 2,767$  observations. From Figure 4, it is apparent that the MALap-CCC Hybrid GARCH(1, 1)-SV model provides a markedly better (albeit not perfect) fit for the tail probabilities than the MN-CCC GARCH(1, 1) model. In Figure 5 we compare the fit of the MALap-CCC Hybrid GARCH(1, 1)-SV model and the MAt-CCC GARCH(1, 1) model. The latter model in the univariate, symmetric case, usually called  $t$ -GARCH, is well known for providing an excellent model fit. Here, from Figure 5, we see that the results are comparable to those of the MALap-CCC Hybrid GARCH(1, 1)-SV model, though for very extreme events, the MAt performs slightly better.

Now consider the filtered  $G_t$  sequence. Figure 6 illustrates its impact on one of the assets, Merck & Co. The top panel gives the returns, while the second panel shows the filtered  $\widehat{G}_t$  values from the EM algorithm. In the third panel, the scale-term,  $s_{k,t}$ , for the same asset, implied by the estimates of the GARCH(1, 1) dynamics from (18), are plotted over time. The panel in the last row combines the above factors and plots the  $Y_{k,t}$  volatilities, which are the square roots of the diagonal elements of the conditional covariance matrix defined in (21) computed based on the parameter estimates. A very negative spike in the second quarter of the data is synchronous with a large spike in the  $\widehat{G}_t$  sequence in the second panel, which corresponds to the spike in the volatility in the last panel (especially when compared with the scale-term dynamics from the third panel). This illustrates the role of the common market factor as a stochastic latent filter. Interestingly, in the periods of high volatility (e.g., around the crisis of 2008) there are no strong market shocks. Although volatilities are very high, their magnitude is adequately accounted for by the GARCH(1, 1) dynamics, and the  $G_t$  factor is instead responsible for sharper volatility moves.

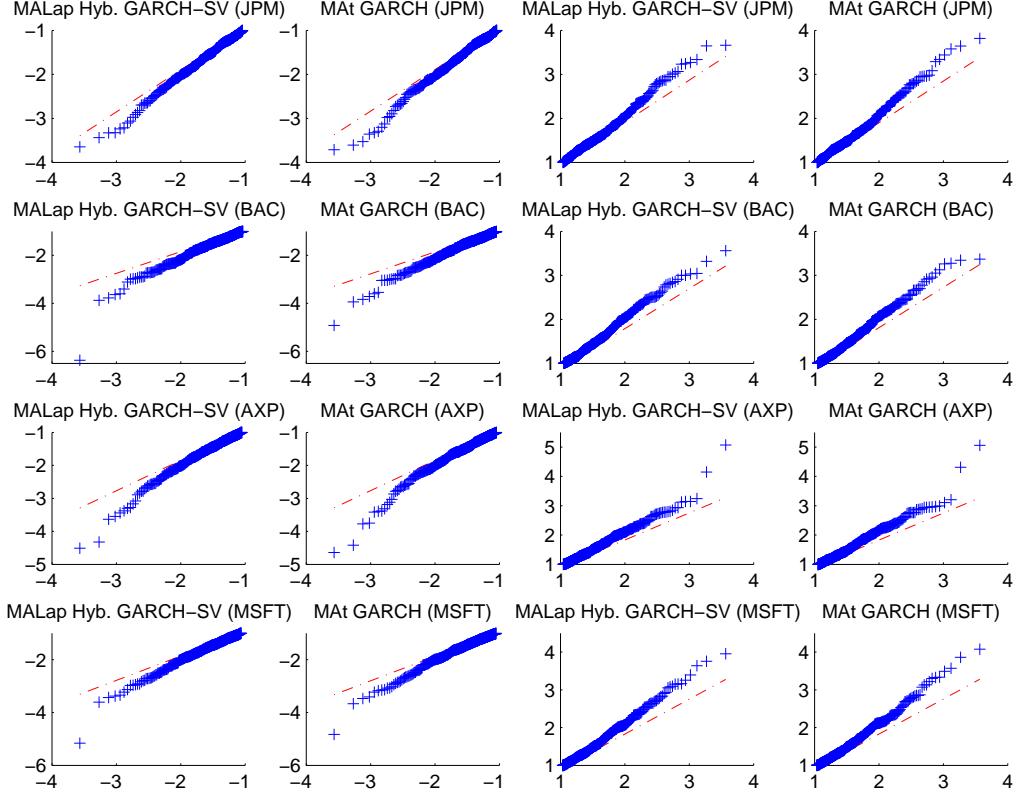
The effect of  $G_t$  across assets is not equal. From (21), each asset volatility is a sum of two terms. The first term is a product of the scale-term,  $s_{k,t}$ , and the conditional expected value of the common market factor, so the impact of the  $G_t$  term on each asset volatility depends on the level of the corresponding scale-term. The second term is a product of the conditional variance of  $G_t$  and the square of the asymmetry coefficient in the vector  $\boldsymbol{\gamma}$ . Hence, the impact of  $G_t$  on volatilities differs across assets. Figure 7 illustrates this fact. It is a multivariate analogue of Figure 6 and explains the contribution of the  $G_t$  factor in the conditional volatilities from the MALap-CCC hybrid GARCH(1, 1)-SV model. Clearly, the  $G_t$  spikes have different impacts on volatilities of different assets.



**Figure 4:** Tails of the quantile plots of the conditional distribution of innovations based on the 2,767 observations. **Rows:** From top to bottom JPMorgan Chase & Co. (JPM); Bank of America (BAC); American Express (AXP); Microsoft Corp. (MSFT). **First column:** The left tail of the MN-CCC GARCH(1,1) model. **Second column:** The left tail of the MALap-CCC Hybrid GARCH(1,1)-SV model. **Third column:** The right tail of the MN-CCC GARCH(1,1) model. **Fourth column:** The right tail of the MALap-CCC Hybrid GARCH(1,1)-SV model.

We now discuss the consequences of the hybrid GARCH(1,1)-SV extension. The sequence of unobserved mixing random variables  $G_t$  implies the non-normality of the model and, in general, cannot be predicted. The role of the SV extension is to filter, through the dynamics in (14), a possible persistence in  $G_t$ . We model only the dynamics in the parameters of the conditional distribution of  $G_t$ , and not the dynamics of  $G_t$  itself. The consequence of this is that we need to distinguish between  $\mathbb{E}[G_t | \Phi_{t-1}]$  and  $\mathbb{E}[G_t | \Phi_t]$ . The former are either constant over time (when  $G_t$  are iid) or, time-varying (with  $G_t | \Phi_{t-1}$  having time-varying parameters). The latter,  $\mathbb{E}[G_t | \Phi_t]$ , are filtered from the E-step update of the ECME algorithm. They condition on the observed data up to *and including* time  $t$  and, obviously, cannot be used for prediction, but instead they serve as a natural benchmark to judge the in-sample-fit of  $\mathbb{E}[G_t | \Phi_{t-1}]$ .

In the first two panels of Figure 8, we compare  $\mathbb{E}[G_t | \Phi_t]$  and  $\mathbb{E}[G_t | \Phi_{t-1}]$  based on the MALap-CCC models. The case with iid  $G_t$  is given in the first panel. In the iid case,  $\mathbb{E}[G_t | \Phi_{t-1}] = \mathbb{E}[G_t]$ , and we see that they result in a relatively poor fit. The second panel is for the hybrid extension of the model, where the dynamics of the  $\mathbb{E}[G_t | \Phi_{t-1}]$  are described by (14). The latter model clearly results in better fit of the common market factor. The  $\mathbb{E}[G_t | \Phi_{t-1}]$  estimates match



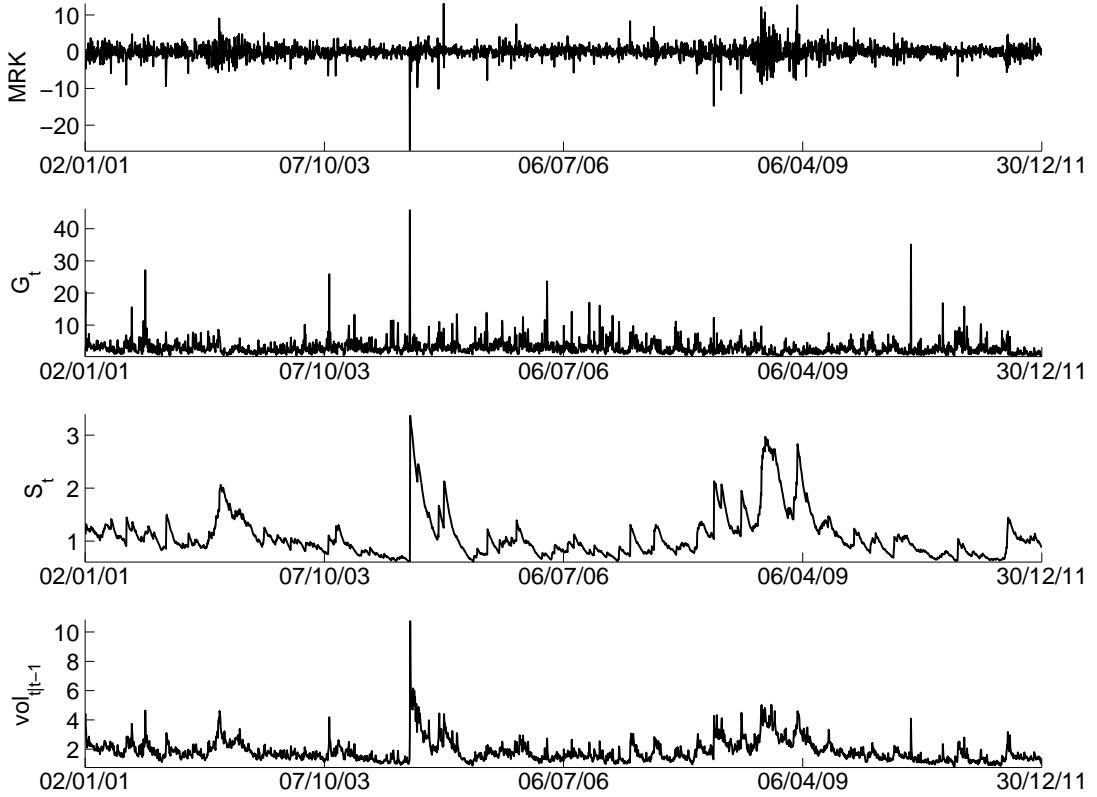
**Figure 5:** Tails of the quantile plots of the conditional distribution of innovations based on the 2,767 observations, with rows analogous to Figure 4. **First column:** The left tail of the MALap-CCC Hybrid GARCH(1,1)-SV model. **Second column:** The left tail of the MAT-CCC GARCH(1,1) model. **Third column:** The right tail of the MALap-CCC Hybrid GARCH(1,1)-SV model. **Fourth column:** The right tail of the MAT-CCC GARCH(1,1) model.

the filtered values and even the largest spikes (which could be interpreted as describing highly unexpected news) are well-accommodated.

The last two panels in Figure 8 compare the resulting conditional volatilities from the two models. Again, the conditional volatilities from the hybrid extension (computed from (21) with use of  $\mathbb{E}[G_t | \Phi_{t-1}]$ ) lie much closer to the filtered values (computed from (21) with use of  $\mathbb{E}[G_t | \Phi_t]$ ).

What is common to all assets is that the  $G_t$  factor explains a large fraction of the volatility. Based on the whole sample estimates, Figure 9 displays the correlations between  $\mathbb{E}[G_t | \Phi_{t-1}]$  and the conditional volatilities of the assets filtered from the ECME algorithm. Remarkably, for 26 out of 30 assets, the univariate common market factor accounts, on average, for more than 40% of the conditional volatility dynamics, and the lowest 4 are (for JPM, MCD, MSFT, and WMT) around 10% to 20%. This is a consequence of separating the GARCH dynamics from the volatility shock dynamics. The former are responsible for modeling the volatility persistence. The latter are modeled by the SV dynamics of the common market factor, and capture the sharp changes in the volatility.

Figure 10 displays the higher-order dynamics implied by the MALap-CCC hybrid GARCH(1,1)-

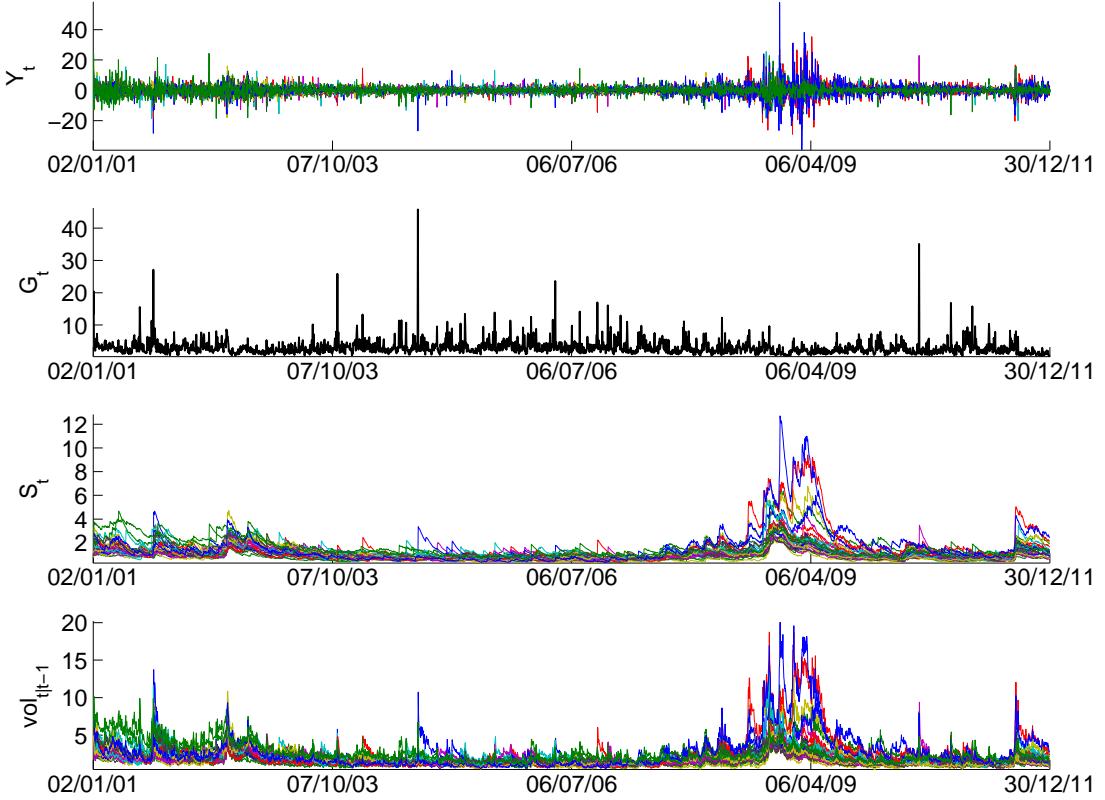


**Figure 6:** The impact of the common market factor on one of the assets (Merck & Co). **First row:** Returns  $Y_{k,t}$  of Merck & Co. **Second row:** Values of the filtered common market factor  $\hat{G}_t$  from the ECME algorithm. **Third row:** The scale-term,  $s_{k,t}$ , for the same asset, implied by the estimates of the GARCH(1,1) model. **Fourth row:** The conditional volatility of  $Y_{k,t}$ , computed as the square root of the  $k$ th element on the diagonal of matrix (21) and based on the parameter estimates.

SV model and computed as in Scott et al. (2011). The first panel plots the conditional skewness. Depending on the sign of the  $\gamma_k$  for  $k = 1, \dots, 30$ , the corresponding asset exhibits either a positive or a negative skewness and its dynamics are driven by the dynamics of the  $G_t$  parameters (the correlation between  $\mathbb{E}[G_t | \Phi_{t-1}]$  and the conditional skewness is  $\pm 0.87$ ). The second panel displays the conditional kurtosis. It is common for all the assets because, as opposed to the conditional skewness, there is no vector which would differentiate the impact of  $\mathbb{E}[G_t | \Phi_{t-1}]$ . From this panel and the second panel in Figure 8, one can note that the kurtosis and  $\mathbb{E}[G_t | \Phi_{t-1}]$  are inversely related, i.e., the lower the value of  $\mathbb{E}[G_t | \Phi_{t-1}]$ , the higher the value of the kurtosis. In fact, the correlation between  $\mathbb{E}[G_t | \Phi_{t-1}]$  and the conditional kurtosis is equal to  $-0.81$ .

## 7.2 Density Forecasting Performance Comparison

Now turning to out-of-sample forecasting performance, this section compares a number of special cases of model (13) with the elliptical ( $\gamma = \mathbf{0}$ ) versions of these models (i) estimated via the EM algorithm and (ii) via the corrected (second-best) three-step estimation method, as well as (iii) the Gaussian based models such as the CCC model of Bollerslev (1990), the DCC model of Engle



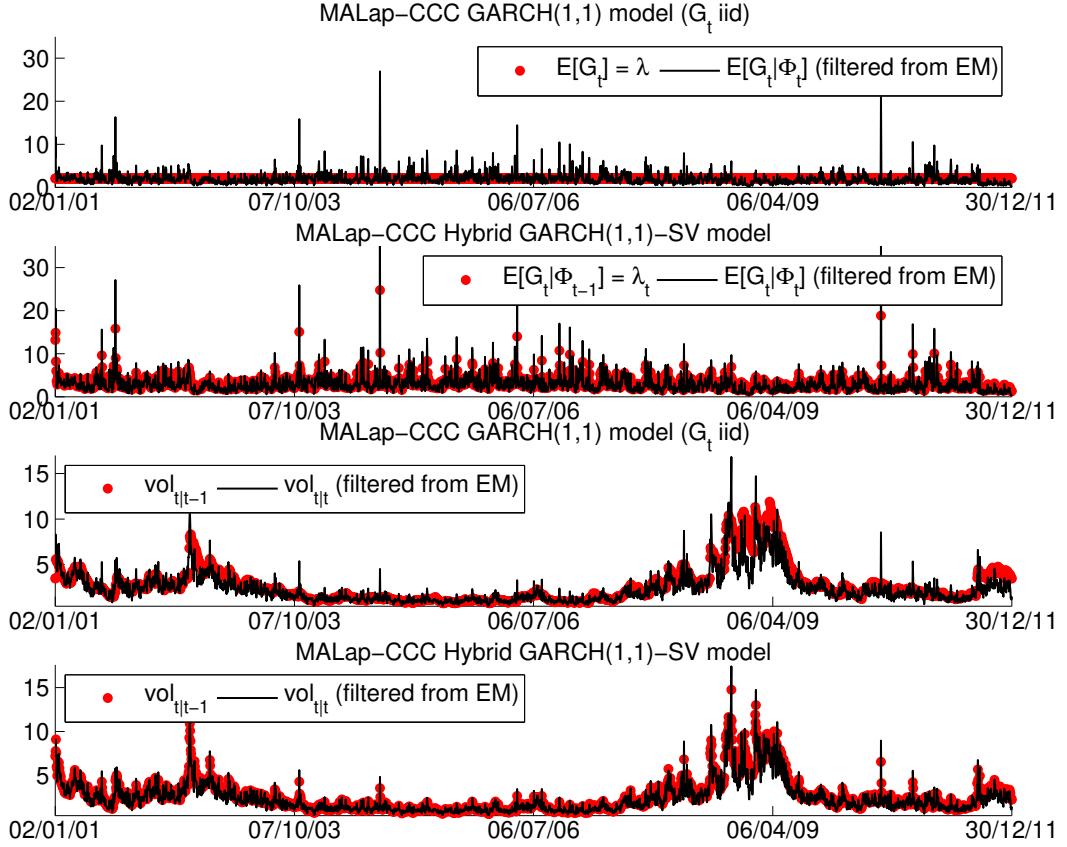
**Figure 7:** The impact of the common market factor on all of the assets. **First row:** All 30 return series. **Second row:** Values of the filtered common market factor  $\hat{G}_t$  from the ECME algorithm. **Third row:** The scale-term,  $s_{k,t}$ , for  $k = 1, \dots, K$ , implied by the estimates of the GARCH(1,1) models. **Fourth row:** The conditional volatilities of  $\mathbf{Y}_t$ , computed as the square root of the elements on the diagonal of matrix (21) and based on the parameter estimates.

(2002), the cDCC model of Aielli (2011), and the VC model of Tse and Tsui (2002), the latter all denoted with a prefix MN-, for multivariate normal distribution of the innovations.

Our interest centers on the quality of one-step ahead predictions of the return vector density. For this purpose, we estimate all the models using a rolling window of 1,000 observations, and, similar to Paoletta (2015), we use the normalized sum of the realized predictive log-likelihood values, which, for given model  $\mathcal{M}$ , is

$$S_T(\mathcal{M}) = \frac{1}{T} \sum_{t=1}^T \pi_t(\mathcal{M}), \quad \text{where} \quad \pi_t(\mathcal{M}) = \log f_{t+1|t}^{\mathcal{M}}(\mathbf{Y}_{t+1}). \quad (24)$$

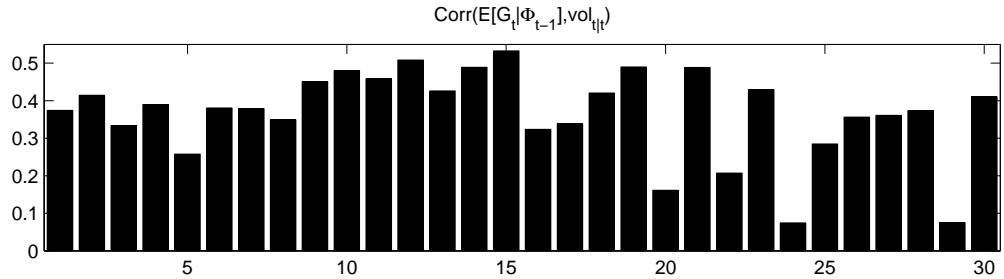
The results are given in Table 1. The hybrid MALap-CCC GARCH(1,1)-SV model performs best. It is closely followed by the MNIG-CCC GARCH(1,1)-SV model. Next in the ranking are two elliptical ( $\gamma = \mathbf{0}$ ) models estimated via the EM algorithm, then the MAT model, followed by the MNIG and MALap models (without hybrid dynamics). The Gaussian-based models perform the worst. Interestingly, even the MALap-iid model, without any GARCH dynamics, performs better than all Gaussian-based models, in particular, even with GARCH.



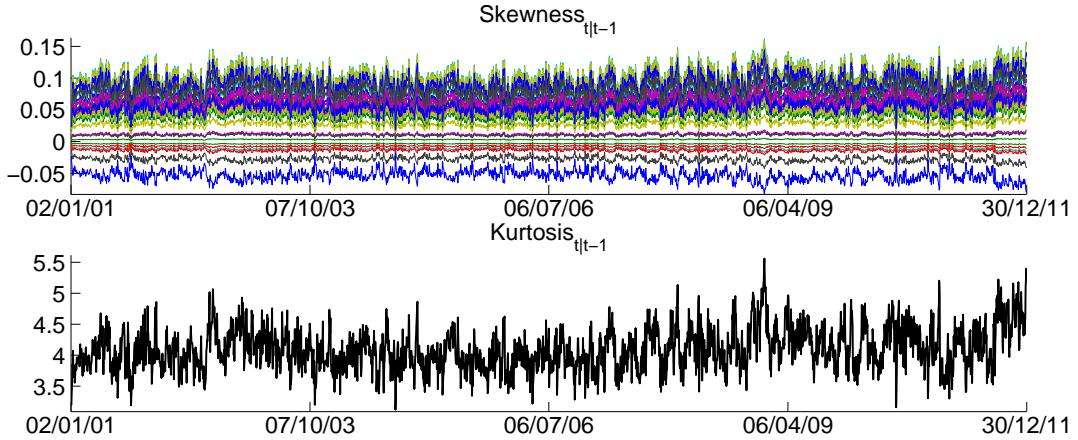
**Figure 8:** The consequences of the hybrid GARCH(1,1)-SV extension. **First row:** The filtered  $\hat{G}_t$  values from the ECME algorithm (i.e.,  $\mathbb{E}[G_t | \Phi_t]$ ) and the estimates obtained from the MALap-CCC GARCH(1,1) model. **Second row:** The filtered  $\hat{G}_t$  values from the ECME algorithm (i.e.,  $\mathbb{E}[G_t | \Phi_t]$ ) and the estimates obtained from the MALap-CCC hybrid GARCH(1,1)-SV model. **Third row:** Conditional volatilities filtered from the ECME algorithm ( $\text{vol}_{t|t}$ ) and the estimates obtained from the MALap-CCC GARCH(1,1) model ( $\text{vol}_{t|t-1}$ ). **Fourth row:** Conditional volatilities filtered from the ECME algorithm ( $\text{vol}_{t|t}$ ) and the estimates obtained from the MALap-CCC hybrid GARCH(1,1)-SV model ( $\text{vol}_{t|t-1}$ ).

Regarding the GJR-GARCH(1,1) dynamics, according to the results in Table 1, its use does not lead to better forecasting performance in any of the models. Figure 11 plots two tail quantiles, the means, and the medians of the estimates of the  $\eta_k$ ,  $k = 1 \dots, 30$ , from (19), across the moving window of 1,000 observations, for the MN-CCC GJR-GARCH model and the MALap-CCC GJR-GARCH(1,1) model. The latter model exhibits smoother  $\eta_k$  estimates, and it is clear that, in periods of high volatility such as the crisis in 2008, there was a large increase in the asymmetry effect. It thus appears that the use of GJR dynamics is enhanced, in terms of clarity and effect, when using a distribution which accounts for skewness and heavy tails.

In order to further investigate this, we check the forecasting performance of our models with the GJR-GARCH(1,1) dynamics for the data windows when the  $\eta_k$  parameters are all larger than a small threshold (we use  $\hat{\eta}_k > 0.01$  for  $k = 1, \dots, 30$ ). It turns out that, for those windows, and for all the distributions considered (MN, MALap, MNIG, and MAt), the models with GJR-GARCH(1,1) significantly outperform their plain GARCH counterparts, but the improvement



**Figure 9:** Correlation between  $\mathbb{E}[G_t | \Phi_{t-1}] = \lambda_t$  and conditional volatility,  $\text{vol}_{t|t}$ , of each of the assets filtered from the ECME algorithm (MALap-CCC hybrid GARCH(1,1)-SV model).

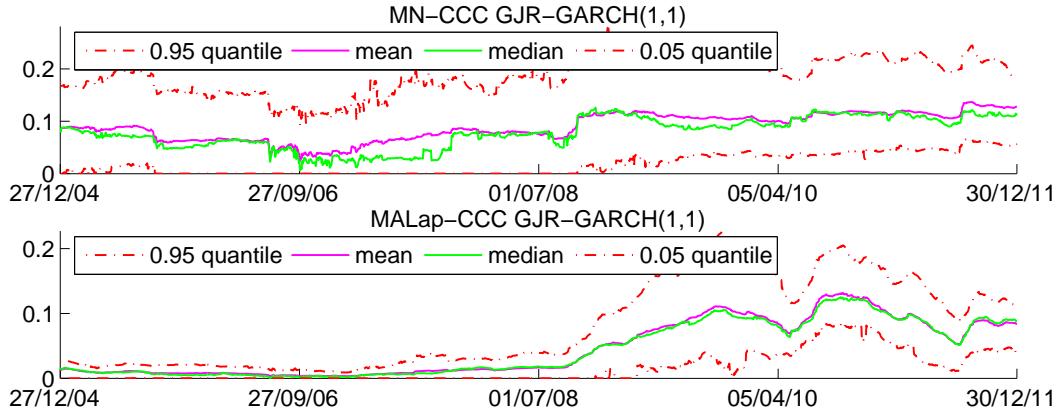


**Figure 10:** Dynamics of higher conditional moments of the returns implied by the MALap-CCC GARCH(1,1)-SV model, computed as in Scott et al. (2011). **Upper panel:** Conditional skewness. **Bottom panel:** Conditional kurtosis.

is much smaller than the gains obtained from relaxing the normality assumption, and from the gains associated with the hybrid GARCH-SV dynamics. In other words, the asymmetry in the volatility, captured by GJR-GARCH(1,1), improves the forecasting only if it is sufficiently strongly supported by the data, and then, the improvement is small, relative to the improvements obtained by use of non-normality and the SV extension.

The most pronounced improvement in forecasting performance is obtained when moving from the Gaussian-based models to any of the new models. The gap in forecasting performance between the new models (first panel in Table 1) and the Gaussian-based models (third panel in Table 1) is much larger than the gap between any models in a given panel.

In order to statistically test the forecasting results from Table 1, we use the test for unconditional predictive ability of Diebold and Mariano (1995) (see also Giacomini and White, 2006). We use a one sided test ( $\mathcal{M}_1 \succ \mathcal{M}_2$ ) and compare each model,  $\mathcal{M}_1$ , in Table 1, with models  $\mathcal{M}_2$  which resulted in a worse-than-model- $\mathcal{M}_1$  forecast. Summarizing, the first six models from Table 1 are very competitive and, according to the test results, there is no significant difference in forecasting performance between them. The first significant improvement (at the 5% level) occurs when moving from the MALap-CCC GARCH(1,1)-SV model to the MNIG-CCC GARCH(1,1) model. The MALap-CCC GARCH(1,1) and MNIG-CCC GARCH(1,1) models perform significantly worse than the analogous hybrid models. In particular, the extension to hybrid dynamics



**Figure 11:** Two tail quantiles, mean, and median of  $\eta_k$ ,  $k = 1 \dots, 30$  parameters from GJR-GARCH(1, 1) dynamics across the moving estimation window of 1,000 observations. **Upper panel:** The MN-CCC GJR-GARCH(1, 1) model. **Bottom panel:** The MALap-CCC GJR-GARCH(1, 1) model.

places the MALap-CCC GARCH(1, 1)-SV model on top. Importantly, the difference between any GARCH-type model and a corresponding hybrid GARCH(1, 1)-SV extension is highly significant.

When moving from the Gaussian-based models to any of the new models, the  $t$ -statistic ranges from 62 to 83. In comparison, moving from a very simple MN-CCC GARCH(1, 1) model to the very popular and best-performing among Gaussian-based models, the MN-DCC GARCH(1, 1) model, results in a  $t$ -statistic of only 4.2. This illustrates that, even with a reasonable law of motion for the conditional volatility, the use of Gaussian innovations in such a conditional model is blatantly inferior to use of just an iid model but with a more suitable distribution. (This is not the first occurrence of such a result: It was also found using an iid model based on a two-component discrete mixture of normals, in conjunction with short estimation windows and use of shrinkage estimation; see Paoletta, 2015.) In turn, using the superior distribution, in this case, the MALap, in conjunction with a GARCH structure, yields further improvement in the forecasts. In particular, comparing the MALap-CCC GARCH(1, 1)-SV to the MALap-iid model results in a  $t$ -statistic of 33.

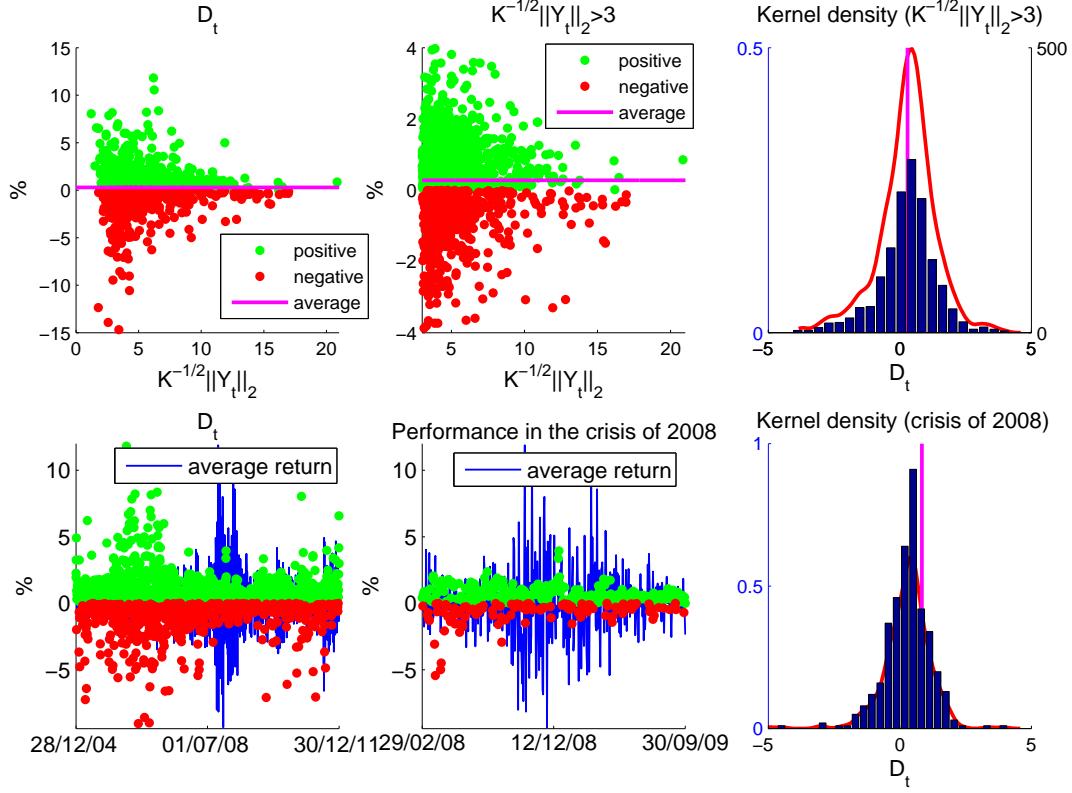
In order to further investigate the forecasting gains from the SV extension of our model for each forecast, we use the percentage measure (defined for  $\pi_t(\mathcal{M}_1)\pi_t(\mathcal{M}_2) > 0$ )

$$D_t(\mathcal{M}_1, \mathcal{M}_2) = 100(|\pi_t(\mathcal{M}_1)| - |\pi_t(\mathcal{M}_2)|) / |\pi_t(\mathcal{M}_2)|. \quad (25)$$

In Figure 12, we plot  $D_t$  for the MALap-CCC hybrid GARCH(1, 1)-SV and the MALap-CCC GARCH(1, 1). We find that (i) on average, the SV extension results in only a minor improvement in forecasting performance even when we consider only periods of large average absolute returns; (ii) but when compared across time, forecasts during the period of the 2008 crisis display a systematic improvement from the SV extension.

### 7.3 Mean Forecasts

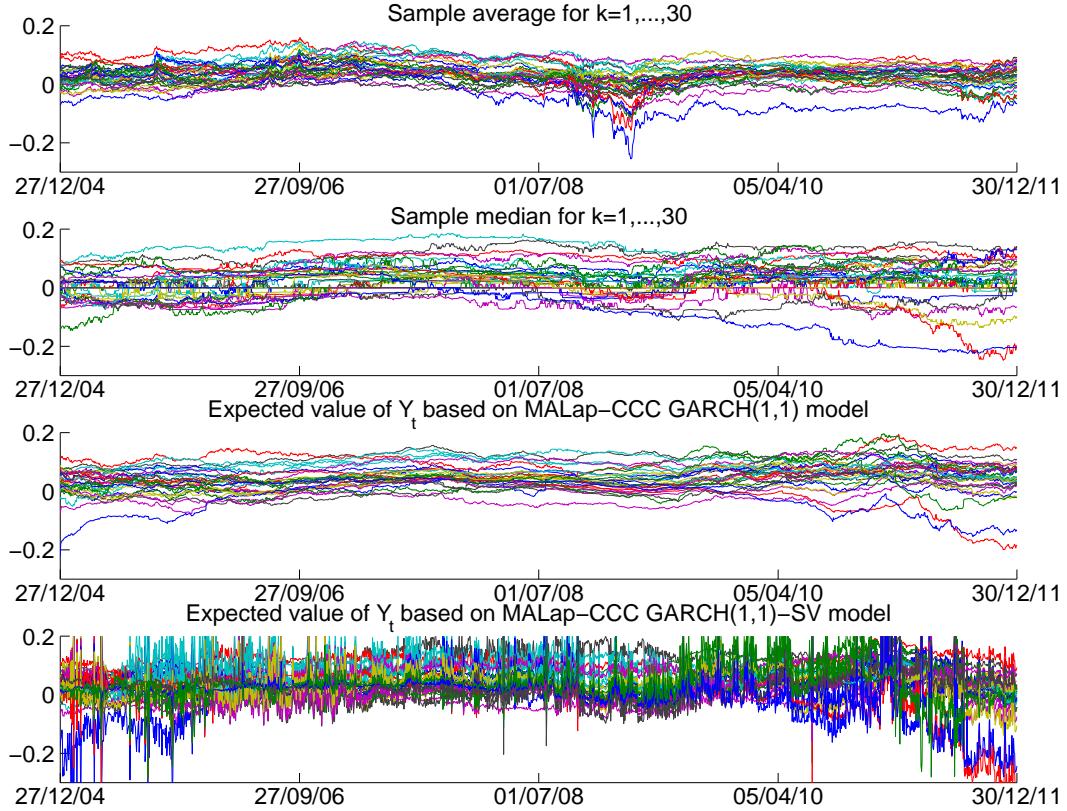
We consider here the forecast of the mean; this being, for example, of utmost importance in a portfolio selection context; see, e.g., Chopra and Ziemba (1993). Figure 13 compares the forecasts of the conditional means based on (i) the sample mean and median from a rolling window of



**Figure 12:** The forecasting gains from the hybrid GARCH(1,1)-SV extension. **First row, column-wise:** Percentage gains  $D_t$  from (25) as a function of average absolute return. Using MALap-CCC GJR-GARCH(1,1) as  $\mathcal{M}_1$  and MALap-CCC GARCH(1,1) as  $\mathcal{M}_2$ . Same, but for large average absolute returns. Histogram of percentage gains for large average absolute returns. **Second row, column-wise:** Percentage gains  $D_t$  from (25) as a function of time. Same, but for crisis of 2008. Histogram of percentage gains during the 2008 crisis.

1,000 observations; (ii) the model-based mean from the MALap-CCC GARCH(1,1) model; and (iii) that from the MALap-CCC GARCH(1,1)-SV extension. Around the 2008 crisis, the sample mean estimates are strongly influenced by negative returns, and, in general, with heavy-tailed data, the sample mean is not the optimal estimator. The sample median is more robust and, as the thickness of the tail increases, it becomes a more efficient estimator. Indeed, the MALap-CCC GARCH(1,1) mean forecasts are more similar to the median estimates. This exercise helps confirm that the model-based forecasts of the mean are accurate.

A potential drawback of the hybrid GARCH(1,1)-SV model is that the dynamics in (14) have an impact on mean dynamics. The forecasts based on the MALap-CCC GARCH(1,1)-SV model, given in the last panel of Figure 13, are more varying, because the conditional mean is a function of the  $G_t \mid \Phi_{t-1}$  parameters as in (20). One could consider more general SV dynamics incorporating the moving average component into (14). This would smooth the forecasts in the last panel of Figure 13 and result in further improvement of the forecasting performance of the hybrid GARCH(1,1)-SV models. To investigate this, we modified the forecast conditional density by scaling the estimate of  $\gamma$  with the factor  $\{\hat{c}/(1-\hat{\rho})\}/\mathbb{E}[G_t \mid \Phi_{t-1}]$ , where the hatted values come from estimation. This has the effect of removing the impact of the spikes in  $\mathbb{E}[G_t \mid \Phi_{t-1}]$  in the mean equation (20). This results in improved forecasting performance, but it was not



**Figure 13:** Conditional mean forecasts from a rolling window of 1,000 observations. **First row:** Sample mean. **Second row:** Sample median. **Third row:** The MALap-CCC GARCH(1,1) model. **Fourth row:** The MALap-CCC GARCH(1,1)-SV model.

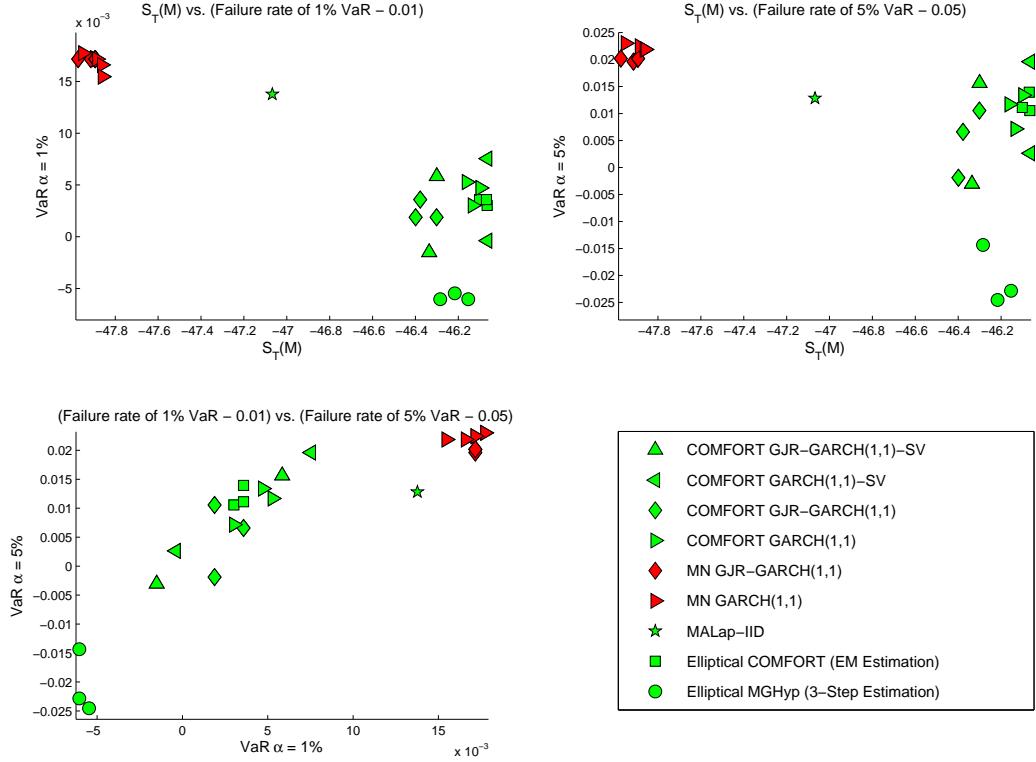
statistically significant at the 5% level.

#### 7.4 Value at Risk Forecasting

We now examine the accuracy of the VaR predictions implied by the proposed models as defined in (22). For this purpose, and following the lead of Santos et al. (2013), we compute daily one-day-ahead out-of-sample VaR forecasts at the 1%, 5% and 10% levels, for an equally weighted portfolio of the 30 stocks under consideration. The empirical coverage levels of the VaR forecasts obtained for various models are reported in Table 2. The models are ordered according to the absolute distance of the failure rate for 1% VaR, from the nominal value of 0.01. While there is no complete best performer, it is noteworthy that, for the 1% and 5% VaR results, the MNIG distribution is the best, independently of the assumed dynamics, and hybrid GARCH-SV dynamics tend to improve these VaR forecasts, supporting the fact that the SV dynamics in the  $G_t$  parameters are capturing dynamics in the tail behavior. Moreover, all the MGHyp models outperform the Gaussian based models, and the MALap-IID model again separates the two groups.

Figure 14 compares the density prediction results with the VaR backtesting results. The MGHyp models estimated with the EM algorithm perform better than the models estimated via the ad hoc three-step method. Also, non-elliptical models tend to improve VaR backtest results.

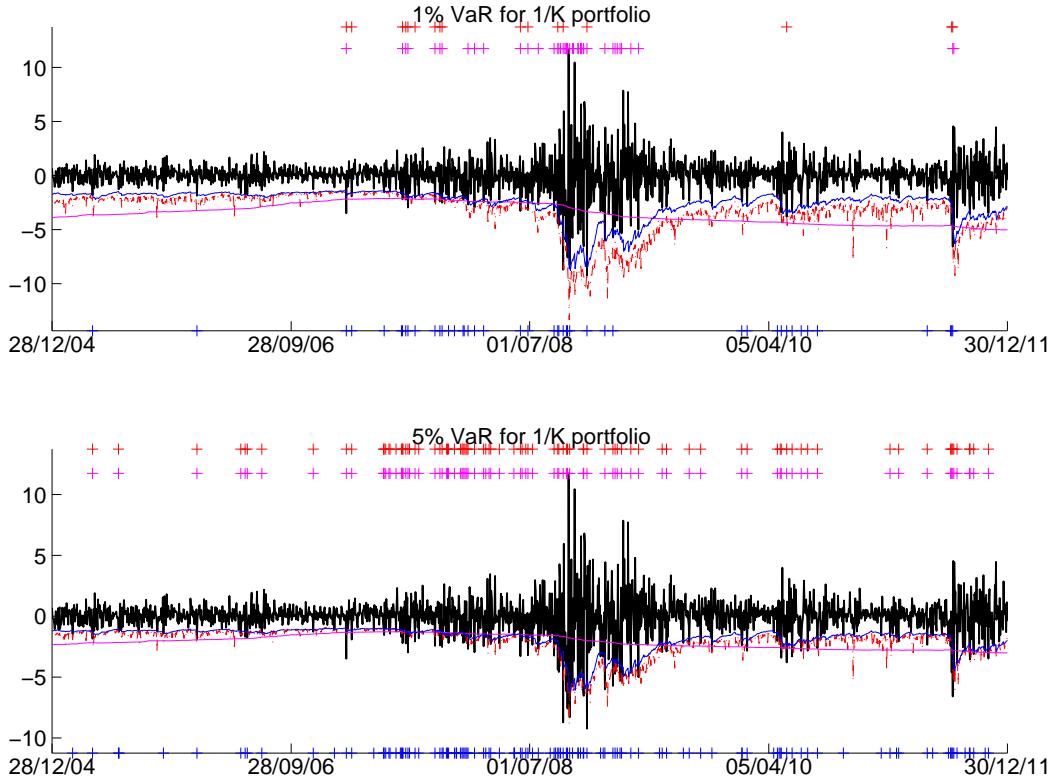
Importantly, all the MGHyp models provide better density forecasts than the Gaussian models, and also lead to better VaR predictions. The MALap-IID model performs somewhere between Gaussian and MGHyp models, but note that the empirical violation frequency (or, in short, failure rate) is not a unique criteria to evaluate the backtesting results. As discussed next, the independence of violations across time is also an important feature.



**Figure 14:** Cumulative predictive log-likelihood of various models vs. their VaR backtesting performance. **Upper left panel:**  $S_T$  vs. the dispersion from the optimal value of the 1% VaR failure rate. **Upper right panel:**  $S_T$  vs. the dispersion from the optimal value of the 5% VaR failure rate. **Bottom left panel:** The dispersion from the optimal value of the 1% VaR failure rate vs. the dispersion from the optimal value of the 5% VaR failure rate.

Figure 15 illustrates 1% and 5% VaR forecasts across time, for an equally weighted portfolio of DJ-30 stocks, using three models—the best Gaussian model (MN-cDCC GARCH(1,1)), the “group separation” MALap-IID, and the best 1% VaR model (MNIG-CCC GARCH(1,1)-SV). The Gaussian model has clearly too many violations, as confirmed in Table 2, while the MALap-IID model has a much better failure rate. However, from Figure 15, we see that almost all violations for the MALap-IID take place during the 2008 crisis, and hence they are strongly dependent. Moreover, outside of the crisis period, the predicted VaR values are much higher than those implied by the conditional models. Thus, during calm periods, the MALap-IID model predicts unnecessarily high capital requirements, and, during market downturns, it predicts insufficient capital holdings. These results are in line with other broad studies comparing unconditional and various conditional GARCH-type models, e.g., Mitnik and Paoletta (2000) and Kuester et al. (2006).

It is important to emphasize that all the COMFORT models considered herein have important



**Figure 15:** Returns, one-day-ahead 1% (5%) VaR forecasts, and VaR violations, for an equally weighted portfolio of DJ-30 stocks, using MNIG-CCC GARCH(1, 1)-SV (dashes), MALap-IID (solid, magenta), and MN-cDCC GARCH(1, 1) (solid, blue) models. The VaR violations are depicted by + signs on top and in the bottom of the figures (with the same colors as VaR predictions).

computational benefits: (i) fast estimation because of the proposed two-step estimation procedure; (ii) the possibility of using parallel computing for further computational speed; (iii) the predictive distribution is a conditional MGHyp (or one of various special cases of it), so that, recalling that the sums of the margins of a MGHyp is itself a GHyp, portfolio construction and risk forecasting is straightforward. The ability to parallelize the code, together with the univariate GARCH computations, make the model estimation feasible even in large dimensions. Since all the best performing models include GARCH dynamics, a longer window of 1000 daily observations is recommended to estimate the parameters. As long as the number of assets is less than the number of observations, estimation of the dependency matrix  $\Gamma$  can be done using a sample correlation matrix. In case, when the number of assets would be larger than the number of observations, some form of shrinkage is necessary to maintain positive definiteness of the estimated correlation matrix.

## 8 Conclusions

The COMFORT models of PP14 yield significantly better density and VaR predictions than their Gaussian counterparts. Moreover, the COMFORT full MLE estimation approach is shown to

be superior to the ad hoc three-step estimation methods currently in use. For the latter, we nevertheless show the required condition to ensure its correct application.

Furthermore, characterizing results of the forecasting benefits of using (univariate) asymmetric GARCH models (in particular, GJR-GARCH) compared to standard GARCH become available. In particular, use of GJR does indeed enhance forecasting performance, though only when the asymmetry is pronounced enough, which tends to happen in high volatility periods. Otherwise, its use tends to degrade performance, similar to classical statistical models in which usual likelihood-based estimation, as opposed to use of shrinkage towards zero, decreases model effectiveness in terms of forecasting ability.

We show that the estimated asymmetry coefficient in the GJR extension is smoother over time, and the marginal improvement in forecasting performance is relatively stronger, when using a non-Gaussian distribution, as compared to use of Gaussian. We conjecture that this is due to the model being (far) less misspecified in the non-Gaussian case. Finally, extensive forecasting comparisons reveal that, while the use of GJR does lead to improvement when the signal is strong enough, this improvement is relatively small compared to the gains achieved by changing the Gaussian distributional assumption, and is also smaller than the gains attributed to use of the SV model extension.

The CCC version of the model, proposed in PP14, assumes that the dependency matrix is time invariant, i.e.,  $\boldsymbol{\Gamma}_t = \boldsymbol{\Gamma}$ . Work in progress by the authors considers extending the COMFORT framework to support the DCC case, and computational methods relevant for the computing expected shortfall of the predictive portfolio distribution. Preliminary results therein agree with those contained herein, namely that very large gains in terms of density prediction and risk prediction are obtained when moving from the Gaussian to the non-Gaussian case.

| $\mathcal{M}$                           | $S_T(\mathcal{M})$ |
|---|--------------------|
| MALap-CCC GARCH(1,1)-SV                 | -46.064            |
| MNIG-CCC GARCH(1,1)-SV                  | -46.064            |
| MNIG (EM elliptical)-CCC-GARCH(1,1)     | -46.066            |
| Mt (EM elliptical)-CCC-GARCH(1,1)       | -46.069            |
| MA $t$ -CCC GARCH(1,1)                  | -46.097            |
| MLap (EM elliptical)-CCC-GARCH(1,1)     | -46.101            |
| MNIG-CCC GARCH(1,1)                     | -46.133            |
| MNIG (3-step elliptical)-CCC-GARCH(1,1) | -46.154            |
| MALap-CCC GARCH(1,1)                    | -46.161            |
| MLap (3-step elliptical)-CCC-GARCH(1,1) | -46.217            |
| Mt (3-step elliptical)-CCC-GARCH(1,1)   | -46.284            |
| MALap-CCC GJR-GARCH(1,1)-SV             | -46.300            |
| MA $t$ -CCC GJR-GARCH(1,1)              | -46.301            |
| MNIG-CCC GJR-GARCH(1,1)-SV              | -46.336            |
| MALap-CCC GJR-GARCH(1,1)                | -46.378            |
| MNIG-CCC GJR-GARCH(1,1)                 | -46.399            |
| MALap-iid                               | -47.067            |
| MN-cDCC GARCH(1,1)                      | -47.859            |
| MN-DCC GARCH(1,1)                       | -47.863            |
| MN-VC GARCH(1,1)                        | -47.885            |
| MN-cDCC GJR-GARCH(1,1)                  | -47.892            |
| MN-DCC GJR-GARCH(1,1)                   | -47.895            |
| MN-VC GJR-GARCH(1,1)                    | -47.914            |
| MN-CCC GARCH(1,1)                       | -47.951            |
| MN-CCC GJR-GARCH(1,1)                   | -47.973            |

**Table 1:** Performance of the one-step ahead predictions of the return vector density for different models,  $\mathcal{M}$ , and measured by  $S_T(\mathcal{M})$ , in (24). **First panel:** Hybrid GARCH-SV and GARCH-type models proposed in this paper. **Second panel:** MALap model under iid assumption. **Third panel:** Gaussian-based models.

| $\mathcal{M}$                           | VaR 1%        | VaR 5%        | VaR 10%       |
|---|---------------|---------------|---------------|
| MNIG-CCC GARCH(1,1)-SV                  | <b>0.0096</b> | <b>0.0526</b> | 0.0905        |
| MNIG-CCC GJR-GARCH(1,1)-SV              | <b>0.0084</b> | <b>0.0469</b> | 0.0837        |
| MNIG-CCC GJR-GARCH(1,1)                 | <b>0.0118</b> | <b>0.0481</b> | 0.0888        |
| MAT-CCC GJR-GARCH(1,1)                  | 0.0118        | 0.0605        | 0.0984        |
| MNIG-CCC GARCH(1,1)                     | 0.0130        | 0.0571        | 0.0962        |
| MNIG (EM-elliptical)-CCC GARCH(1,1)     | 0.0130        | 0.0605        | 0.1018        |
| MALap-CCC GJR-GARCH(1,1)                | 0.0135        | 0.0565        | 0.0945        |
| Mlap (EM-elliptical)-CCC GARCH(1,1)     | 0.0135        | 0.0611        | 0.1013        |
| Mt (EM-elliptical)-CCC GARCH(1,1)       | 0.0135        | 0.0639        | 0.1052        |
| MAT-CCC GARCH(1,1)                      | 0.0147        | 0.0633        | 0.1058        |
| MALap-CCC GARCH(1,1)                    | 0.0152        | 0.0616        | <b>0.1001</b> |
| MLap (3-step elliptical)-CCC-GARCH(1,1) | 0.0045        | 0.0255        | 0.0617        |
| MALap-CCC GJR-GARCH(1,1)-SV             | 0.0158        | 0.0656        | 0.1041        |
| Mt (3-step elliptical)-CCC-GARCH(1,1)   | 0.0040        | 0.0357        | 0.0775        |
| MNIG (3-step elliptical)-CCC-GARCH(1,1) | 0.0040        | 0.0272        | 0.0623        |
| MALap-CCC GARCH(1,1)-SV                 | 0.0175        | 0.0696        | 0.1052        |
| MALap-IID                               | 0.0237        | 0.0628        | <b>0.1007</b> |
| MN-cDCC GARCH(1,1)                      | 0.0254        | 0.0718        | 0.1041        |
| MN-DCC GARCH(1,1)                       | 0.0265        | 0.0718        | 0.1041        |
| MN-VC GJR-GARCH(1,1)                    | 0.0271        | 0.0696        | 0.1013        |
| MN-CCC GJR-GARCH(1,1)                   | 0.0271        | 0.0701        | 0.1013        |
| MN-DCC GJR-GARCH(1,1)                   | 0.0271        | 0.0701        | <b>0.1007</b> |
| MN-cDCC GJR-GARCH(1,1)                  | 0.0271        | 0.0701        | 0.1007        |
| MN-VC GARCH(1,1)                        | 0.0271        | 0.0724        | 0.1041        |
| MN-CCC GARCH(1,1)                       | 0.0277        | 0.0730        | 0.1041        |

**Table 2:** One-day-ahead 1%, 5%, and 10% VaR forecasts failure rates for different models for an equally weighted portfolio of DJ-30 stocks, using various models. The models are ordered according to absolute distance of the failure rate for 1% VaR from 0.01, and in each category the best three results are in bold.

## References

- Aas, K. and Haff, I. H. (2006). The Generalized Hyperbolic Skew Student's  $t$ -Distribution. *Journal of Financial Econometrics*, 4(2):275–309.
- Aielli, G. P. (2011). Dynamic Conditional Correlation: On Properties and Estimation. *Journal of Business and Economic Statistics*.
- Barndorff-Nielsen, O. E. (1978). Hyperbolic Distributions and Distributions on Hyperbolae. *Scandinavian Journal of Statistics*, 5:151–157.
- Barndorff-Nielsen, O. E. (1997). Processes of Normal Inverse Gaussian Type. *Finance and Stochastics*, 2:41–68.
- Bauwens, L., Laurent, S., and Rombouts, J. K. V. (2006). Multivariate GARCH Models: A Survey. *Journal of Applied Econometrics*, 21(1):79–109.
- Bianchi, M. L., Hitaj, A., and Tassinari, G. L. (2020). Multivariate Non-Gaussian Models for Financial Applications. *arXiv preprint arXiv:2005.06390*.
- Bickel, P. J. and Levina, E. (2008). Regularized Estimation of Large Covariance Matrices. *The Annals of Statistics*, 36(1):199–227.
- Blæsid, P. and Sørensen, M. (1992). 'HYP'-a Computer Program for Analyzing Data by Means of the Hyperbolic Distribution. Technical report, Research Report 248, Department of Theoretical Statistics, Aarhus University.
- Bollerslev, T. (1990). Modeling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Approach. *The Review of Economics and Statistics*, 72:498–505.
- Broda, S. A., Haas, M., Krause, J., Paolella, M. S., and Steude, S. C. (2012). Stable Mixture GARCH Models. *Journal of Econometrics*, 172(2):292–306.
- Broda, S. A. and Paolella, M. S. (2009). CHICAGO: A Fast and Accurate Method for Portfolio Risk Calculation. *Journal of Financial Econometrics*, 1:1–25.
- Cajigas, J.-P. and Urga, G. (2007). Dynamic Conditional Correlation Models with Asymmetric Multivariate Laplace Innovations. Centre for Econometric Analysis, Cass Business School, WP-CEA-08-2007.
- Caporin, M. and McAleer, M. (2013). Ten Things You Should Know about the Dynamic Conditional Correlation Representation. *Econometrics*, 1(1):115–126.

- Chopra, V. and Ziemba, W. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management*, 19:6–12.
- Diebold, F. X. and Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 13:253–263.
- Doganoglu, T., Hartz, C., and Mittnik, S. (2007). Portfolio Optimization when Risk Factors are Conditionally Varying and Heavy Tailed. *Computational Economics*, 29(3-4):333–354.
- Eberlein, E. and Keller, U. (1995). Hyperbolic Distributions in Finance. *Bernoulli*, 1:281–299.
- Eberlein, E., Keller, U., and Prause, K. (1998). New Insights into Smile, Mispricing, and Value at Risk: the Hyperbolic Model. *Journal of Business*, 38:371–405.
- 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.
- Fan, J., Fan, Y., and Lv, J. (2008). High Dimensional Covariance Matrix Estimation Using a Factor Model. *Journal of Econometrics*, 147:186–197.
- Fotopoulos, S. B., Paparas, A., and Jandhyala, V. K. (2020). Multivariate Generalized hyperbolic laws for modeling financial log-returns: Empirical and theoretical considerations. *Applied Stochastic Models in Business and Industry*, 36(5):757–775.
- Francq, C. and Zakoïan, J.-M. (2004). Maximum Likelihood estimation of Pure GARCH and ARMA-GARCH Processes. *Bernoulli*, 10(4):605–637.
- Francq, C. and Zakoïan, J.-M. (2010). *GARCH Models Structure, Statistical Inference and Financial Applications*. John Wiley & Sons Ltd.
- Giacometti, R., Bertocchi, M. I., Rachev, S. T., and Fabozzi, F. J. (2007). Stable Distributions in the Black-Litterman Approach to Asset Allocation. *Quantitative Finance*, 7(4):423–433.
- Giacomini, R. and White, H. (2006). Tests of Conditional Predictive Ability. *Econometrica*, 74(6):1545–1578.
- Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *The Journal of Finance*, 48(5):pp. 1779–1801.
- Greenspan, A. (1997). Maintaining Financial Stability in a Global Economy. Discussion at Federal Reserve Bank of Kansas City symposium.
- Hansen, J. V., McDonald, J. B., and Turley, R. S. (2006). Partially Adaptive Robust Estimation of Regression Models and Applications. *European Journal of Operational Research*, 170:132–143.
- Jondeau, E. (2012). Asymmetry in Tail Dependence of Equity Portfolios. Working Paper, HEC, University of Lausanne.

- Kotz, S., Podgórski, K., and Kozubowski, T. (2001). *The Laplace Distribution and Generalizations: A Review with Application to Communication, Economics, Engineering and Finance*. Massachusetts: Birkhäuser.
- Kotz, S., Podgórski, T. J., and Kozubowski, T. (2000). An Asymmetric Multivariate Laplace Distribution. Technical report, Department of Statistics and Applied Probability, University of California at Santa Barbara.
- Kuester, K., Mitnik, S., and Paoletta, M. S. (2006). Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics*, 4:53–89.
- Madan, D. B. and Seneta, E. (1990). The Variance Gamma (V.G.) Model for Share Market Returns. *The Journal of Business*, 63(4):pp. 511–524.
- McAleer, M. (2014). Asymmetry and Leverage in Conditional Volatility Models. *Econometrics*, 2(3):145–150.
- McDonald, J. B. (1991). Parametric Models for Partially Adaptive Estimation with Skewed and Leptokurtic Residuals. *Economics Letters*, 37:273–278.
- McDonald, J. B. (1997). Probability Distributions for Financial Models. In Maddala, G. S. and Rao, C. R., editors, *Handbook of Statistics*, volume 14. Elsevier Science.
- McDonald, J. B. and Newey, W. K. (1988). Partially Adaptive Estimation of Regression Models via the Generalized  $t$  Distribution. *Econometric Theory*, 4:428–457.
- McKay, A. T. (1932). A bessel function distribution. *Biometrika*, 24(1/2):39–44.
- McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton University Press, Princeton, revised edition.
- Mitnik, S. and Paoletta, M. S. (2000). Conditional Density and Value-at-Risk Prediction of Asian Currency Exchange Rates. *Journal of Forecasting*, 19:313–333.
- Näf, J., Paoletta, M. S., and Polak, P. (2019). Heterogeneous Tail Generalized COMFORT Modeling via Cholesky Decomposition. *Journal of Multivariate Analysis*, 172(C):84–106.
- Orskaug, E. (2009). Multivariate DCC-GARCH Model With Various Error Distributions. Working Paper 19/09, Norwegian Computing Center.
- Paoletta, M. S. (2007). *Intermediate Probability: A Computational Approach*. John Wiley & Sons, Hoboken, New Jersey.
- Paoletta, M. S. (2014). Fast Methods For Large-Scale Non-Elliptical Portfolio Optimization. *Annals of Financial Economics*, 09(02):1440001.
- Paoletta, M. S. (2015). Multivariate Asset Return Prediction with Mixture Models. *European Journal of Finance*, 21(13–14):1214–1252.
- Paoletta, M. S. (2018). *Fundamental Statistical Inference: A Computational Approach*. John Wiley & Sons, Hoboken, New Jersey.

- Paoletta, M. S. and Polak, P. (2015). COMFORT: A Common Market Factor Non-Gaussian Returns Model. *Journal of Econometrics*, 187(2):593–605.
- Paoletta, M. S. and Polak, P. (2018). COBra: Copula-Based Portfolio Optimization. In Vladik Kreinovich, Songsak Sriboonchitta, N. C., editor, *Studies in Computational Intelligence: Predictive Econometrics and Big Data*. Springer.
- Paoletta, M. S., Polak, P., and Walker, P. S. (2019). Regime Switching Dynamic Correlations for Asymmetric and Fat-Tailed Conditional Returns. *Journal of Econometrics*, 213(2):493–515.
- Phillips, R. F. (1994). Partially Adaptive Estimation via a Normal Mixture. *Journal of Econometrics*, 64:123–144.
- Podgórski, T. J. and Kozubowski, T. (2001). Asymmetric Laplace Laws and Modeling Financial Data. *Mathematical and Computer Modelling*, 34:1003–1021.
- 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.
- Rombouts, J., Stentoft, L., and Violante, F. (2014). The Value of Multivariate Model Sophistication: An Application to Pricing Dow Jones Industrial Average Options. *International Journal of Forecasting*, 30(1):78–98.
- Saliha, S. A. and Aboudi, E. H. (2021). Bayes Estimators of a Multivariate Generalized Hyperbolic Partial Regression Model. *Int. J. Nonlinear Anal. Appl.*, 12(2):961–975.
- 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.
- Schmidt, R., Hrycej, T., and Stützle, E. (2006). Multivariate Distribution Models with Generalized Hyperbolic Margins. *Computational Statistics & Data Analysis*, 50(8):2065–2096.
- Scott, D. J., Würtz, D., Dong, C., and Tran, T. T. (2011). Moments of the Generalized Hyperbolic Distribution. *Computational Statistics*, 26:459–476.
- Taylor, S. J. (1982). Financial Returns Modelled by the Product of Two Stochastic Processes - a Study of Daily Sugar Prices 1961-79. In *Time Series Analysis: Theory and Practice*, 1, pages 203–226. Amsterdam: North-Holland.
- 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.
- Wang, J. (2009). *The Multivariate Variance Gamma Process and Its Applications in Multi-asset Option Pricing*. PhD thesis, University of Maryland.

## **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