COMFORT: A Common Market Factor Non-Gaussian Returns Model

Marc S. Paolella          Paweł Polak

First version: August 2012
Current version: June 2013

This research has been carried out within the NCCR FINRISK project on “Financial Econometrics for Risk Management”
COMFORT: A Common Market Factor Non-Gaussian Returns Model*

Marc S. Paolella\textsuperscript{a,b} Pawel Polak\textsuperscript{a,b}

\textsuperscript{a}Department of Banking and Finance, University of Zurich, Switzerland

\textsuperscript{b}Swiss Finance Institute

June 25, 2013

Abstract

A new multivariate time series model with various attractive properties is motivated and studied. By extending the CCC model in several ways, it allows for all the primary stylized facts of financial asset returns, including volatility clustering, non-normality (excess kurtosis and asymmetry), and also dynamics in the dependency between assets over time. A fast EM-algorithm is developed for estimation. The predictive conditional distribution is a (possibly special case of a) multivariate generalized hyperbolic, so that sums of marginals (as required for portfolios) are tractable. Each element of the vector return at time $t$ is endowed with a common univariate shock, interpretable as a common market factor, and this stochastic process has a predictable component. This leads to the new model being a hybrid of GARCH and stochastic volatility, but without the estimation problems associated with the latter, and being applicable in the multivariate setting for potentially large numbers of assets. Its applicability to option pricing is developed. In-sample fit and out-of-sample conditional density forecasting exercises using daily returns on the 30 DJIA stocks confirm the superiority of the model to numerous competing ones.

**Keywords**: CCC; Common Jumps, Density Forecasting; EM-Algorithm; Fat Tails; GARCH; Multivariate Asymmetric Laplace Distribution; Multivariate Asymmetric $t$ Distribution; Multivariate Generalized Hyperbolic Distribution; Multivariate Normal Inverse Gaussian Distribution; Multivariate Option Pricing; Stochastic Volatility.

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

*Part of the research of Paolella has been carried out within the National Centre of Competence in Research “Financial Valuation and Risk Management” (NCCR FINRISK), which is a research program supported by the Swiss National Science Foundation. We wish to thank Tim Bollerslev, Cathy Chen, Dick van Dijk, Paul Embrechts, Matthias Fengler, Eric Jondeau, Loriano Mancini, Michael McAleer, Antonio Mele, Stefan Mittnik, Olivier Scaillet, Enrique Sentana, Mike Ka Pui So, Fabio Trojani, Michael Wolf and Philip Yu for many useful comments and suggestions.
1 Introduction

“I preferred SV models to my ARMACH models because I did not like the assumption that returns up to time \( t - 1 \) then contain all the information needed to correctly identify a conditional normal distribution for the return during period \( t \). This assumption does not permit unexpected news on day \( t \) to partially determine the variance of the conditional normal distribution, and it is also contradicted by the empirical forecasting evidence. I soon appreciated, however, that my arguments against ARCH do not apply to the more general specifications of Engle and Bollerslev (1986) and Bollerslev (1987) that permitted the \( Z \) to have a fat-tailed distribution; their Student-t choice is a mixture of normal distributions that can be motivated by uncertainty at time \( t - 1 \) about the amount of relevant news on day \( t \). There are many similarities between SV models and general ARCH models. It is very difficult, and possibly of minimal value, to say which of these two types of models provides the best description of market returns.”


We consider modeling a set of asset returns via a conditional multivariate distribution with dynamics governed by a process which has features of both GARCH and stochastic volatility (SV). These two essentially disparate paradigms for capturing volatility clustering in asset returns have their individual advantages, and also limitations. In univariate models, both can capture changes in volatility, the leverage effect (at least for more recent SV models) and the leptokurtosis and possible asymmetry of the innovations distribution. The main disadvantage of SV models is the lack of an explicit form of the likelihood function and the necessity to use moment-based methods or simulation for estimation. The former are often simple, but inefficient, while the latter attempt to achieve a close approximation of the likelihood function through computationally intensive methods, and become problematic for more than a small number of assets (see, e.g., Asai et al. 2006; Asai and McAleer 2009 and Bos 2012 and the reference therein). Alternatively, in the univariate case, a GARCH model is trivial to estimate via (conditional) maximum likelihood, but is unable to model the dynamics of an unpredictable volatility component. Both SV and many kinds of GARCH models share the same estimation infeasibility in the large-scale multivariate case, though models such as CCC, DCC, and some of their extensions, are feasible; these are considered in the empirical section below.

The recent literature has emphasized the importance of including a stochastic jump component in the volatility dynamics—a feature which can be easily incorporated into the SV structure (see, among others, Chernov et al. 2003; Eraker et al. 2003; Eraker 2004; and Todorov and Tauchen 2011) but is absent in GARCH models. (One exception is the model proposed by Chan and Maheu 2002 and Maheu and McCurdy 2004 see also Section 2 below.) Another approach is to use high-frequency returns in order to construct realized variation measures which separate the volatility jump component from the smooth continuous movement of the underlying volatility (see, e.g., Andersen et al. 2007).

In the multivariate case, jumps in individual assets can be observed to arrive simultaneously, forming what are called co-jumps. Following the arbitrage pricing theory, Bollerslev et al. (2008) distinguish the co-jumps in the idiosyncratic component, which are diversifiable, and the co-jumps in the common component, which are non-diversifiable, i.e., those which carry a risk premium. According to their statistical test, there are co-jumps which can be highly significant, but when considered on individual stocks, remain undetected. The importance of co-jumps structures for the
consumption-portfolio selection problem is discussed by Aït-Sahalia et al. (2009). More recently, Gilder et al. (2013) provide empirical evidence for co-jumps and analyse the association between jumps in the market portfolio and co-jumps in the underlying stocks. Their results suggest that news events which have market-level influence are able to generate large co-jumps in individual stocks, while Bollerslev et al. (2013) show that extreme joint dependencies observed in daily data can be implied by the diffusive volatility and co-jumps observed in the high-frequency data.

In line with this high-frequency literature, we propose a new model which also splits the dynamics of volatility; however, it is a parametric approach which does not require high-frequency data and is applicable in the multivariate setting. We propose a solution which, in a conditional setup, utilizes a flexible, fat-tailed distribution, and combines univariate GARCH-type dynamics (most of the popular variations are possible) with a relatively simple, yet flexible, SV dynamic structure, based on the seminal work of Taylor (1982), whose quotation above served as motivation for this paper. By introducing a latent component similar to that used in the SV literature, the resulting hybrid GARCH-SV model is able to capture stochastic (co-)jumps in the volatility series and across assets. To the best of our knowledge, this is the first volatility model which combines GARCH and SV paradigms—and is applicable to large-dimensional multivariate settings, owing to the proposed EM-algorithm for maximum likelihood estimation (MLE) and the possibility for parallel computing.

The model contains a univariate, latent component which is common to all \( K \) assets. We term this a common market factor. It dictates the nonlinear dependency between margins, so that, conditional upon it (as realized via the EM-algorithm), what remains to be estimated are \( K \) univariate normal GARCH models. The (conditional) MLE of each of the latter is numerically fast and reliable to obtain, and the \( K \) estimations can be conducted in parallel. We show that the latent common market factor process has a predictable component, and that this leads to (statistically significant) improvements in forecasting performance over and above the sizeable improvements obtained by relaxing the normality assumption.

Another important feature of our model is that it imposes a multiplicative structure on the volatility, and an infinitely divisible distribution which, in the iid case, generates an infinite activity jump process, as opposed to jump diffusion models, where finite activity jumps, modeled via a Poisson-distributed term, are added to Gaussian dynamics. This is in line with Aït-Sahalia and Jacod (2011), who propose two non-parametric statistical tests to discriminate between the two cases, and both tests point toward the presence of infinite-activity jumps in the data. Other examples of models which support infinite activity jumps, and are often used in the context of option pricing, are the variance gamma model of Madan and Seneta (1990) and Madan et al. (1998), the hyperbolic model of Eberlein and Keller (1995), and the CGMY model of Carr et al. (2002).

An extensive empirical exercise involving the 30 assets comprising the DJIA reveals that, in terms of in-sample fit and out-of-sample density forecasts based on the predictive log likelihood, the new model demonstrably outperforms the classical CCC and DCC models. As such, the proposed model can be used for deriving an accurate prediction of the conditional mean vector and covariance matrix of a set of asset returns, and so has direct use in market risk management. In addition, because of the tractability of the sums of margins, portfolio allocation based on minimization of downside risk measures such as the value at risk and expected shortfall can be
conducted, as well as based on the mean and variance, if desired. The general model, with parameter $\gamma$ nonzero, is non-elliptic, and the data support $\gamma \neq 0$, so that use of coherent risk measures, such as expected shortfall, in portfolio optimization, will have a positive effect on the management of portfolio risk; see Embrechts et al. (2001). Finally, by combining the equivalent martingale measure approach with Monte Carlo simulation, option pricing can also be conducted.

The remainder of the paper is as follows. The model and some of its properties are stated in Section 2. Section 3 discusses the proposed method of estimation. The MGHyp distribution being essentially too flexible, Section 4 motivates and details important special cases. Section 5 provides details on option pricing. Section 6 illustrates an exercise with real data in which in-sample fit and out-of-sample density forecasting are compared across models. Section 7 provides some concluding remarks, and an Appendix gathers various technical results.

2 Model

We consider a set of $K$ financial assets, with associated return vector at time $t$ given by $Y_t = (Y_{t,1}, Y_{t,2}, \ldots, Y_{t,K})'$, $t = 1, 2, \ldots, T$, whose conditional, time-varying distribution is taken to be multivariate generalized hyperbolic, hereafter MGHyp; see, e.g., McNeil et al. (2005). We observe a realization of $Y = [Y_1 | Y_2 | \cdots | Y_T]$, where the $Y_t$ are equally spaced in time (ignoring the weekend effect for daily data) return vectors. The information set at time $t$ is currently defined as the history of returns, $\Phi_t = \{Y_1, \ldots, Y_t\}$, though extensions to the model which include the use of exogenous variables could be straightforwardly entertained. Our focus is on the prediction of the conditional distribution of $Y_{T+h}$ given $\Phi_T$, for which we restrict our empirical demonstration in this paper to $h = 1$. The dispersion matrix of $Y_t | \Phi_{t-1}$ is decomposed as the product of scale terms and a conditional dependency matrix (a correlation matrix when the MGHyp distribution approaches the multivariate normal). For each of the univariate scales, a GARCH-type structure is imposed, while the dependency matrix is specified as being constant over time. We will see in Remark 2(ii) below that, except for some special cases, the correlations are actually time-varying. Further methods of inducing time-variation in the dependency matrix are discussed in the conclusions.

Using the mixture representation of the MGHyp (see Section 3 below; Eberlein and Keller 1995 and Eberlein et al. 1998), we can express the return vector as

$$Y_t = \mu + \gamma G_t + \varepsilon_t, \quad \text{with}$$

$$\varepsilon_t = H_t^{1/2} \sqrt{G_t} Z_t,$$

where $\mu = (\mu_1, \ldots, \mu_K)'$ and $\gamma = (\gamma_1, \ldots, \gamma_K)'$ are column vectors in $\mathbb{R}^K$; $H_t$ is a positive definite, symmetric, conditional dispersion matrix of order $K$; $Z_t \overset{iid}{\sim} N(0, I_K)$ is a sequence of independent and identically distributed (iid) normal random vectors and $(G_t | \Phi_{t-1}) \sim \text{GIG} (\lambda_t, \chi_t, \psi_t)$ are mixing random variables, $t = 1, 2, \ldots, T$, independent of $Z_t$, with typical GIG (generalized inverse Gaussian) density given by

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

4
\( K_\lambda \) is the modified Bessel function of the third kind (and not to be confused with \( K \), the number of assets), given by
\[
K_\lambda (x) = \frac{1}{2} \int_0^\infty t^{\lambda - 1} \exp \left( -\frac{x}{2} (t + t^{-1}) \right) dt, \quad x > 0;
\]
and \( \lambda > 0, \psi_t > 0 \) if \( \lambda < 0 ; \chi_t > 0, \psi_t > 0 \) if \( \lambda = 0 ; \) and \( \chi_t > 0, \psi_t > 0 \) if \( \lambda > 0 \).

We consider two specifications for the GIG parameters: (i) \( G_i \mid \Phi_{t-1} \) are iid with time-invariant parameters, i.e., \( \lambda_i = \lambda, \chi_i = \chi \) and \( \psi_i = \psi \); (ii) \( G_i \mid \Phi_{t-1} \) has time dependent, random, parameters with the dynamics described by a system of conditional moment equations
\[
\mathbb{E} [G_t^r \mid \Phi_{t-1}] = c_r + \rho_r \mathbb{E} [G_{t-1}^r \mid \Phi_{t-2}] + \zeta_{r,t},
\]
for a set of positive integer values of \( r \); \( \zeta_{r,t} = \mathbb{E} [G_t^r \mid \Phi_t] - \mathbb{E} [G_t^r \mid \Phi_{t-1}] \); and \( c_r \) and \( \rho_r \) are parameters to be estimated. The error term \( \zeta_{r,t} \) represents the unpredictable component affecting the \( r \)th moment of the mixing variable \( G_t \). It contains all new information in forming expectations about \( G_t^r \) when moving from time \( t-1 \) to \( t \). It can be used as a driver of the dynamics of \( \mathbb{E} [G_t^r \mid \Phi_{t-1}] \) in (4) because it is a martingale difference sequence (MDS) with respect to \( \Phi_{t-1}, \) implying that \( \mathbb{E} [\zeta_{r,t}] = 0 \) and \( \text{Cov} (\zeta_{r,t}, \zeta_{r,t-s}) = 0, s = 1, 2, \ldots \). From (4), the dynamics of the parameters \( \lambda_t, \chi_t, \) and \( \psi_t \) can be inferred by the expression for the moments of the GIG random variable given below in (15). However, for each of the three special cases of the MGHyp distribution we entertain, it turns out that the dynamics of only one of the three parameters associated with \( G_t \) needs to be modeled; see Remark 2(v) below, and Section 4 for details. For example, when using \( r = 1 \) in the estimation of the dynamics in (4), \( \mathbb{E} [G_t \mid \Phi_{t-1}] \) has to be positive. The dynamics in (4) can be rewritten as
\[
\mathbb{E} [G_t^r \mid \Phi_{t-1}] = c_r^r + \rho_r \mathbb{E} [G_{t-1}^r \mid \Phi_{t-2}] + \frac{1}{2} \mathbb{E} [G_t^r \mid \Phi_{t}],
\]
so that the sufficient condition for \( \mathbb{E} [G_t^r \mid \Phi_{t-1}] \) to be positive for all \( t > 0 \) is \( c_r^r > 0 \) and \( \rho_r \geq 0 \).

When the parameters of \( G_t \) are allowed to be dynamic as in (4), we call model (4) a hybrid GARCH-SV extension because it can be linked to the seminal SV model of [Taylor (1982)]; this link being detailed in Appendix A. Our model differs from it because ours is a multivariate model with GARCH dynamics in the individual scales and an SV component which describes the dynamics of the common market factor \( G_t \) (as detailed below). Moreover, the dynamics in (4) are in the same vein as [Chan and Maheu (2002)] and [Maheu and McCurdy (2004)], who model dynamics of the jump intensity of a Poisson process in individual stock returns. (One important difference is that our dynamics imply that \( G_t^r \) is not a deterministic function of the past returns.) In line with these works, we model the dynamics of the linear projections of \( G_t^r \) on past returns only, this being another difference between our approach and that of [Taylor (1982)].

Due to the MDS property of the \( \zeta_{r,t} \) innovations, the conditional forecasts of the future conditional moments are given by
\[
\mathbb{E} [G_{t+s}^r \mid \Phi_t] = c_r^s + \sum_{i=0}^{s-1} \rho_i^r \mathbb{E} [G_t^r \mid \Phi_{t-1}], \quad s \geq 1,
\]
where \( \mathbb{E} [G_t^r \mid \Phi_{t-1}] \) is measurable with respect to the information up to time \( t-1 \) and is given by (15) below. If \( |\rho_r| < 1 \), then the process in (4) is mean-reverting, and for \( s \to \infty \), the forecast approaches the unconditional mean value \( c_r / (1 - \rho_r) \) of \( G_t^r \).

The conditional dispersion matrix \( H_t \) is decomposed as

\[
H_t = S_t \Gamma S_t, \tag{7}
\]

where \( S_t \) is a diagonal matrix composed of the strictly positive conditional scale terms \( s_{k,t}, k = 1, 2, \ldots, K \), and \( \Gamma \) is a time-invariant, symmetric, with ones on the main diagonal, conditional dependency matrix, such that \( H_t \) is positive definite. The univariate scale terms \( s_{k,t} \) are modeled by a GARCH-type process. In particular, the simplest realistic choice is the GARCH(1,1) model

\[
s_{k,t}^2 = \omega_k + \alpha_k \varepsilon_{k,t-1}^2 + \beta_k s_{k,t-1}^2, \tag{8}
\]

where \( \varepsilon_{k,t} = y_{k,t} - \mu_k - \gamma_k G_t \) is the \( k \)-th element of the \( \varepsilon_t \) vector in (1), and \( \omega_k > 0, \alpha_k \geq 0, \beta_k \geq 0 \), for \( k = 1, 2, \ldots, K \). More general formulations could be used, notably those which can capture an asymmetry effect; see, e.g., Mittnik and Paolella (2000) and the references therein. For this purpose, we consider the GJR-GARCH(1,1) model of Glosten et al. (1993), given by

\[
s_{k,t}^2 = \omega_k + \alpha_k \varepsilon_{k,t-1}^2 + \eta_k \varepsilon_{k,t-1}^2 \mathbf{1}\{ \varepsilon_{k,t-1} > 0 \} + \beta_k s_{k,t-1}^2, \tag{9}
\]

where \( \mathbf{1}\{ \cdot \} \) is an indicator function, and \( \eta_k \geq 0 \) captures asymmetry in the scale-term response to the last period innovation. Engle and Ng (1993) have shown that, in the Gaussian case, the GJR-GARCH(1,1) model is the best performing parametric model for capturing the asymmetry response of the volatility to news.

Remarks:

(i) In order to maintain the news effect in future volatilities, the innovation process used in the GARCH recursions in (8) and (9) is \( \varepsilon_{k,t} = y_{k,t} - \mu_k - \gamma_k G_t \). If instead, we were to use \( \varepsilon_{k,t} / \sqrt{G_t} \), then the next period volatility would not be influenced by the current spike in \( G_t \). Hence, there would be no volatility persistence after news; the model would not capture the stylized fact of volatility clustering, and use of the GARCH-type dynamics for the scale term would be inadequate. We have also empirically confirmed that this alternative specification leads to a very low-persistence GARCH recursion (in terms of \( \alpha_k + \beta_k \) values) and that the forecasting performance of the model substantially decreases.

(ii) In model (1), \( \mu \) is the location vector and \( H_t \) is the dispersion matrix of the conditional distribution of \( Y_t \), while the mean and the covariance matrix are given by

\[
\mathbb{E} [Y_t \mid \Phi_{t-1}] = \mu + \mathbb{E} [G_t \mid \Phi_{t-1}] \gamma \tag{10}
\]

and

\[
\text{Cov} (Y_t \mid \Phi_{t-1}) = \mathbb{E} [G_t \mid \Phi_{t-1}] H_t + \nabla (G_t \mid \Phi_{t-1}) \gamma \gamma', \tag{11}
\]

respectively, where \( \nabla (G_t \mid \Phi_{t-1}) = \mathbb{E} [G_t^2 \mid \Phi_{t-1}] - (\mathbb{E} [G_t \mid \Phi_{t-1}])^2 \). Analogously, \( \Gamma \) is a
correlation matrix only conditionally on the realization of the mixing process. For this reason, we call \( \Gamma \) the dependency matrix.

While \( \Gamma \) in (7) is not dynamic, the conditional correlation matrix of \( Y_t \mid \Phi_{t-1} \) is time-varying when \( \gamma \neq 0 \) and \( \mathbb{E}[G_t \mid \Phi_{t-1}] \neq \mathbb{V}(G_t \mid \Phi_{t-1}) \). If \( \gamma = 0 \) or \( \mathbb{E}[G_t \mid \Phi_{t-1}] = \mathbb{V}(G_t \mid \Phi_{t-1}) \) (e.g., in the MALap distribution below), then \( \text{Corr}(Y_t \mid \Phi_{t-1}) = \Gamma \), or \( \text{Corr}(Y_t \mid \Phi_{t-1}) = \Gamma + \gamma \gamma' \), respectively, and the dynamics in the parameters of \( G_t \mid \Phi_{t-1} \) in (4) influence only the variances.

All the conditional moments implied by the model (including the limiting cases of the mixing law) are available in Scott et al. (2011). The co-skewness and co-kurtosis matrices are also tractable; see the impressive expressions given in Mencía and Sentana (2009).

Finally, the unconditional mean and covariance of \( Y_t \) can be expressed in terms of the unconditional mean of \( G_t \) and unconditional covariance function of \( Y_t \mid G_t \), respectively as \( \mathbb{E}[Y_t] = \mu + \mathbb{E}[G_t] \gamma \) and \( \text{Cov}(Y_t) = \mathbb{E}[\text{Cov}(Y_t \mid G_t)] + \mathbb{V}(G_t) \gamma \gamma' \).

(iii) From (11) it follows that the vector of conditional volatilities, defined as the square root of the conditional variances, and denoted by \( \text{vol}_{t \mid t-1} \), is given by

\[
\text{vol}_{t \mid t-1} = \sqrt{\left[ \mathbb{E}[G_t \mid \Phi_{t-1}] \right] S_t^2 + \mathbb{V}(G_t \mid \Phi_{t-1}) \gamma^2}.
\]

The asymmetric effect in volatility states that the effects of positive returns on volatility are different from those of negative returns of a similar magnitude. The leverage effect refers to the negative correlation between the current return and future volatility. Therefore, leverage implies asymmetry, but not all asymmetric effects display leverage (see Asai and McAleer 2011; and the reference therein). In the GJR-GARCH model, both positive and negative returns increase future volatility, but the positive returns do so by less than the negative returns. Recalling that \( \mathbb{E}[G_t \mid \Phi_{t-1}] > 0 \) and \( \mathbb{V}(G_t \mid \Phi_{t-1}) > 0 \), it follows, from (12), that the only possibility to introduce leverage into the model is to use the dynamics for \( s_{k,t} \) which support the leverage. In particular, neither the common market factor \( G_t \), nor its dynamics, can be the source of leverage in the model.

(iv) As the volatility shock of the SV component is univariate, it influences (in a multiplicative way) each of the asset-specific conditional volatilities via \( 1b \). Moreover, it drives the dynamics of higher conditional moments (e.g., skewness and kurtosis) and co-moments of the returns. One could argue that modeling volatility shocks with a univariate process is not sufficient because the reaction of the asset-specific volatilities to the common shock should vary across assets and through time. Nevertheless, because of the asset-specific conditional asymmetry coefficient \( \gamma_k \), the impact of the SV component on each volatility is not equal across assets and its conditional expected value varies over time. (See Section 6.1 for an empirical demonstration of this.)

(v) There is a minor identification problem which needs to be addressed. The same MGHyp distribution arises from the parameter constellation \( (\lambda_t, \chi_t/c, c \psi_t, \mu, c H_t, \gamma) \) for any \( c > 0 \). One way to deal with this is to constrain the determinant of \( H_t \) to some particular value when
fitting (see, e.g., McNeil et al., 2005). Alternatively, Protassov (2004) sets $\chi$ to a constant in his EM algorithm. We follow this latter approach and fix one of the GIG parameters prior to the estimation, as it is numerically simpler to implement in the iid (non-GARCH) setting, and, more crucially, because in our general model, the conditional dispersion matrix is time dependent. Moreover, this reduces the number of necessary equations to identify the parameters of $G_t | \Phi_{t-1}$.

(vi) In order to estimate the dynamics in (4), the starting values of $\zeta_{r,t}$ and $\mathbb{E}[G^r_t | \Phi_{t-1}]$ have to be set. In our empirical analysis, at each iteration in the estimation, we set them equal to the unconditional expected values, i.e., $\zeta_k,0 = 0$ for all $k$ and $\mathbb{E}[G^r_1 | \Phi_0] = c_r/(1 - \rho_r)$, respectively. In addition, we assume that the roots of the polynomials $(1 - \rho_r L)$, where $L$ is the lag operator associated with (4), lie outside the unit circle (i.e., modulus $|\rho_r| < 1$), so that the unconditional expected value of $G^r_t$ exists.

(vii) For notational convenience later, we collect the parameters of the model into three blocks (process, distribution, and correlation) as follows:

$$
\theta_P = (\mu, \gamma, \omega, \alpha, \beta, \eta), \quad \theta_D = (c, \rho) \quad \text{and} \quad \theta_C = \Gamma,
$$

where $\omega, \alpha, \beta, \eta$ are $K$-dimensional vectors of GJR-GARCH(1,1) parameters from (9) (the GARCH case (8) corresponds to $\eta_k = 0$); and $c$ and $\rho$ denote vectors of $c_r$ and $\rho_r$ parameters, respectively, though note that, in the case of constant $G_t$ parameters, $\theta_D$ reduces to just the set $(\lambda, \chi, \psi)$, of which in our three special cases, two are fixed.

3 Estimation

The explicit form of the MGHyp density and the structure of (1) implies that, if $\gamma = 0$ (necessary to make the scale shock $\varepsilon_t$ independent of the unobserved realizations of $G_t$), then the estimation of the parameters in the model can be performed by direct maximization of the corresponding likelihood function.$^1$ In particular, with $\theta_P^S = \theta_P \setminus \gamma$, vector $Y_t$ is assumed to have an MGHyp distribution with density $f_{Y_t}(y; \theta_P^S, \theta_D, \theta_C)$ given by

$$
\int_0^\infty \frac{1}{(2\pi)^{K/2} |H_t|^{1/2} g^{K/2}} \exp \left\{ -\frac{(y - \mu)'H_t^{-1}(y - \mu)}{2g} \right\} \times f_G(g; \theta_D) dg,
$$

where $f_G(\cdot)$ is the density of the unobserved mixing random variable $G$ given in (2); $H_t$ admits the decomposition (7) with constant conditional correlation; and the dynamics of the scale terms, $s_{k,t}$, are from (8) or (9). One can then evaluate this integral to yield the explicit expression

$$
f_{Y_t}(y; \theta_P^S, \theta_D, \theta_C) = C_t d_t^{-K/2+\lambda} K_{\lambda-K/2}(d_t),
$$

---

$^1$The following estimation algorithm maximizes the conditional likelihood function, to ease the notation we omit the conditioning on $\Phi_{t-1}$ across the derivation.
for \( d_t = \sqrt{(\chi + (y - \mu)' H_t^{-1} (y - \mu)) \psi} \) and the normalizing constant

\[
C_t = \frac{\left(\sqrt{\chi \psi}\right)^{-\lambda} \psi^{K/2}}{(2\pi)^{K/2} |H_t|^{1/2} K^\lambda (\sqrt{\chi \psi})^K}.
\]

Direct maximization of the likelihood function of \( Y \), denoted \( L_Y(\theta_P, \theta_D, \theta_C) \), requires estimation of all the parameters in one step. This approach works for \( K \) very small, but becomes problematic as the number of assets increases, due to the quadratic increase in the number of parameters associated with the dispersion matrix, and the linear increase due to \( \mu \) and the univariate GARCH parameters. In the case with \( K \neq 0 \) and/or for large \( K \), direct maximization is no longer feasible. However, estimation can be conducted via an Expectation-Conditional Maximization Either (ECME) algorithm from Liu and Rubin (1994). The standard ECME algorithm maximizes the likelihood function by an iterative procedure. It is a fixed-point algorithm and consists of the E-step, in which the realizations of the unobserved mixing variables \( \{G_t\}_{t=1,...,T} \) are imputed; and the CM-steps, in which all the parameters are updated by maximizing either the unconditional likelihood function \( L_Y(\theta_P, \theta_D, \theta_C) \), or the conditional one, \( L_{Y|G}(\theta_P, \theta_C) \). This is now detailed.

The complete log-likelihood function is

\[
\log L_{Y,G}(\theta_P, \theta_D, \theta_C) = \log L_{Y|G}(\theta_P, \theta_C) + \log L_G(\theta_D),
\]

where, based on the observed values \( Y_t = y_t \) and conditional on \( G_t = g_t = \hat{G}_t \) (where the hatted notation indicates that \( G_t \) is estimated, and not observed), \( t = 1,2,\ldots,T \),

\[
\log L_{Y|G}(\theta_P, \theta_C) = -\frac{1}{2} \sum_{t=1}^{T} \left[ K \log (2\pi) + \log |g_t S_t^t \Omega S_t| ight. \\
+ g_t^{-1}(y_t - \mu - \gamma g_t)' S_t^{-1} \Gamma^{-1} S_t^{-1}(y_t - \mu - \gamma g_t) + \log g_t \bigg],
\]

and \( L_G(\theta_D) \) denotes the likelihood function of \( (G_t | \Phi_{t-1}) \sim \text{GIG}(\lambda_t, \chi_t, \psi_t), t = 1,2,\ldots,T \). In a similar fashion to Bollerslev (1990), Engle and Sheppard (2002), and Pelletier (2006), we split \( \log L_{Y|G}(\theta_P, \theta_C) \) into a sum of two terms,

\[
\log L_{Y|G}(\theta_P, \theta_C) = \log L_{Y|G}^{\text{MV}}(\theta_P) + \log L_{Y|G}^{\text{Cor}}(\theta_P, \theta_C),
\]

where \( L_{Y|G}^{\text{MV}}(\theta_P) \) is the mean-volatility term given by

\[
\log L_{Y|G}^{\text{MV}}(\theta_P) = -\frac{1}{2} \sum_{t=1}^{T} \left[ K \log (2\pi) + \log |S_t|^2 ight. \\
+ g_t^{-1}(y_t - \mu - \gamma g_t)' S_t^{-1} S_t^{-1}(y_t - \mu - \gamma g_t) + \log g_t \bigg],
\]
and $L_{Y|G}^{Corr} (\theta_P, \theta_C)$ is the correlation term given by

$$
\log L_{Y|G}^{Corr} (\theta_P, \theta_C) = -\frac{1}{2} \sum_{t=1}^{T} \left[ \log |\Gamma| + e_t' \Gamma^{-1} e_t - e_t' e_t \right],
$$

where $e_t = g_t^{-1/2} S_t^{-1/2} \varepsilon_t$, and $\varepsilon_t = y_t - \mu - \gamma g_t$ from (1).

Owing to the mixture structure of the MGHyp, $L_{Y|G} (\theta_P, \theta_C)$ is a multivariate Gaussian likelihood with a GARCH-type structure for the scales and a given conditional correlation model. As such, maximization of $L_{Y|G} (\theta_P, \theta_C)$ can be done in two steps. First, with the correlation structure ignored, the GARCH parameters in (8) or (9) are estimated for each of the $K$ assets separately (and concurrently with parallel computing) by maximizing $L_{Y|G} (\theta_P \mid \theta_C = I_K)$, and, in the second step, the correlation parameters $\theta_C$ are estimated from the first-step de-volatilized residuals. This idea is not new, and is similar to the Gaussian setup in Bollerslev (1990), Engle (2002), and Pelletier (2006). Given the $\theta_P$ and $\theta_C$ estimates, denoted by $\hat{\theta}_P$ and $\hat{\theta}_C$, the mixing process parameters $\theta_D$ are estimated by maximizing $L_Y (\theta_D \mid \hat{\theta}_P, \hat{\theta}_C)$. Given all the estimates, we proceed with the next E-step update (see (20) below) of the unobserved mixing random variable $G$, and continue to iterate until convergence.

Observe that $L_{Y|G} (\theta_P, \theta_C)$ reduces to the mean-volatility component, $L_{Y|G}^{MV} (\theta_P)$, if and only if we assume zero correlations. As such, $L_{Y|G}^{MV} (\theta_P)$ corresponds to the likelihood of $Y$ conditional on $G$ for the model under zero correlations. Based on this decomposition, estimates of $\theta_P$, $\theta_C$ and $\theta_D$ can be obtained by the following ECME algorithm.

**E-step:** Calculate $E \left[ \log L_{Y,G} \mid \hat{Y}, \hat{\theta}_P, \hat{\theta}_D, \hat{\theta}_C \right]$.

The log-likelihood function (18) is linear with respect to $g_t$ and $g_t^{-1}$. Hence the E-step involves replacing unobserved realizations of $G_t$ and $G_t^{-1}$ in (15) by the imputed values, $\hat{G}_t$.

Calculation shows that (see, e.g., Paolella 2013, Eq. 35)

$$
\left( G_t \mid \Phi_t; \hat{\theta}_P, \hat{\theta}_C, \hat{\theta}_D \right) \sim \text{GIG} \left( \lambda_t^*, \chi_t^*, \psi_t^* \right),
$$

where

$$
\lambda_t^* = \lambda_t - K/2, \quad \chi_t^* = \chi_t + (y_t - \hat{\mu})' S_t^{-1} \Gamma^{-1} S_t^{-1} (y_t - \hat{\mu}) \quad \text{and} \quad \psi_t^* = \psi_t + \gamma^2 \chi_t^* \Gamma^{-1} S_t^{-1} \Gamma^{-1} \psi_t.
$$

The latent values of $g_t$ and $g_t^{-1}$ are then updated by their conditional expectations from (20), using the expression for the moments of the GIG random variable given in (45).

**CM1-step:** Update $\theta_P$ and $\theta_C$.

(P) Update $\theta_P$ by computing

$$
\arg \max_{\theta_P} \log L_{Y|G}^{MV} (\theta_P),
$$

where $L_{Y|G}^{MV} (\theta_P)$ is a Gaussian likelihood with zero correlation, so we can estimate the parameters of each asset, $(\mu_k, \gamma_k, \omega_k, \alpha_k, \beta_k)$, separately by maximizing the corresponding likelihood function.
(C) Update $\theta_C$ by computing the usual empirical correlation estimator (the MLE under normality) of the de-volatilized residuals $\hat{G}_t^{-1/2}S_t^{-1}\hat{\varepsilon}_t$, where $\hat{\varepsilon}_t = Y_t - \hat{\mu} - \hat{\gamma} \hat{G}_t$, $\hat{\mu}$ and $\hat{\gamma}$ are obtained in part (P) directly above, and $\hat{G}_t$ is obtained in the E-step.

**CM2-step:** Given the CM1-step estimates of $\theta_P$ and $\theta_C$, obtain new estimates of $\theta_D$ by maximizing the incomplete data log-likelihood function, i.e., compute

$$\arg\max_{\theta_D} \log L_Y(\theta_D | \hat{\theta}_P, \hat{\theta}_C).$$

(22)

Iterate the above steps until convergence.

**Remarks:**

(i) In contrast to the Gaussian distribution, setting $\Gamma$ equal to the identity matrix does not imply independence when assuming MGHyp returns, due to the dependence induced by the GIG mixing variable $G_t$. But via the application of the ECME algorithm, which conditions on the realizations of $G_t$, we can estimate the parameters of the GARCH equations separately for each asset. This is the key to fast, simple, joint likelihood-based estimation of a multivariate non-normal model.

(ii) In the ECME algorithm, the E-step computation of the conditional expectations of $G_t$ and $G_t^{-1}$ involves computation of ratios of Bessel functions for different arguments. In case of large arguments ($\chi_t^*$ and $\psi_t^*$ in (20) can get very large because they involve a quadratic form of the inverse of the covariance matrix), numerical computation is subject to rounding error which affects estimates of all the parameters. We propose a method which increases numerical accuracy such that estimation for all data windows used in our empirical application was successful. It is given in Appendix D.

(iii) The method which increases the accuracy of the Bessel function computation given in Appendix D is also used in the CM2-step for evaluating $\log L_Y$ for different $\theta_D$ values in (22).

(iv) Here we state the starting values used for the estimation procedure. Those of the asymmetry parameter $\gamma_k$ and the GJR-GARCH parameter $\eta_k$ are taken to be zero, and the remaining ones in $\theta_P$, $({\mu}_k, {\omega}_k, {\alpha}_k, {\beta}_k)$, to be those values obtained from the normal-based GARCH estimates, using the method in Paolella (2010) to avoid inferior local likelihood maxima. For $\theta_C$, we use the empirical correlation matrix computed from the normal-based GARCH residuals. For $\theta_D = (\lambda, \chi, \psi)$, we have confirmed that the likelihood is such that optimization is rather robust to the choice of starting values. For the special case of the multivariate Laplace distribution (MALap) used in the empirical section below, we use $\lambda = 2$ as the starting value. For the hybrid GARCH-SV model we set, $c = 0.1$ and $\rho = 0.8$. In the estimation with a rolling window, we use also as starting values the previous window estimates; and take the final estimates to be those with the higher likelihood value.

(v) If the starting values are sufficiently close to those which maximize the likelihood, or if the likelihood function is unimodal in the parameter space, and the maximum is not on the boundary of the parameter space, then monotonicity in the likelihood values of the
consecutive ECME estimates guarantees their convergence to the corresponding maximum likelihood parameters; see, e.g., McLachlan and Krishnan (2008). As such, under further standard regularity conditions on the likelihood, consistent and asymptotically normal point estimates are obtained.

(vi) For $\gamma = 0$ and $K$ small, we confirmed the accuracy of the proposed EM algorithm by comparing the EM estimates with the results based on the direct likelihood maximization of (14).

4 Special Cases of the MGHyp

Fixing some of the MGHyp distribution parameters in a judiciously chosen way can result in a distribution which 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, 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).

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. 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 higher kurtosis than the normal distribution. Section 4.1 details the derivation and properties of the MALap density, while Sections 4.2 and 4.3 examine the MNIG and MAT distributions, respectively.

4.1 Multivariate Laplace Distribution

The MALap density can be derived by evaluating the integral

$$f_Y(y; \lambda, \mu, H, \gamma) = \int f_Y|_G(y \mid g; \mu, H, \gamma) f_G(g; \lambda) dg;$$

where $G \sim \text{Gam}(\lambda, 1)$; or from the MGHyp density (14), with $\lambda > 0$, $\chi = 0$ and $\psi = 2$, by use of the limiting relation $K_\lambda(\sqrt{2\chi}) \approx \Gamma(\lambda)2^{\lambda-1}(\sqrt{2\chi})^{-\lambda}$ for $\chi \downarrow 0$, $\lambda > 0$ (Paolella, 2007, Eq. 9.6), where $\Gamma$ is the gamma function. Both approaches result in $f_{Y_t}(y; \lambda, \mu, H_t, \gamma)$ given by

$$2\exp\left\{ (y - \mu)^T H_t^{-1} \gamma \right\} \left( \frac{m_t}{2 + \gamma^T H_t^{-1} \gamma} \right)^{\lambda/2 - K/4} K_{\lambda - K/2} \left( \sqrt{m_t \left( 2 + \gamma^T H_t^{-1} \gamma \right)} \right),$$

(23)
The MNIG distribution arises from the MGHyp for \( \gamma = -1/2 \), with density

\[
 f_Y(y; \mu, H_t, \gamma, \chi, \psi) = C_t d_t^{-(K+1)/2} K_{(K+1)/2}(d_t) \exp \left\{ (y - \mu)' H_t^{-1} \gamma \right\},
\]

for \( d_t = \sqrt{\chi + (y - \mu)' H_t^{-1} (y - \mu)} \) and the normalizing constant

\[
 C_t = \left( \frac{\chi/\psi}{2\pi} \right)^{1/4} \left( \frac{\gamma' H_t^{-1} \gamma}{(2\pi)^{K/2} |H_t|^{1/2} K_{1/2} (\sqrt{\chi\psi})} \right)^{(K+1)/2}.
\]

This distribution has been used in the iid case (McNeil et al. 2005) and also for building a multivariate predictive distribution via univariate NIG-GARCH components (Broda and Paolella 2009). In our model, for identification purposes (see Remark 2(v)) 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.
4.3 Multivariate $t$-Distribution

The MA$t$ is a limiting case of the MGHyp, with $\lambda = -1/2v$, $\chi = v$ and $\psi = 0$, for some $v > 0$. Evaluating the limit of (14) as $\psi \to 0$ gives the density

$$f_Y(y; \mu, H_t, \gamma, v) = C_t \frac{K_{(v+K)/2} \left( \sqrt{(v + m_t) \gamma' H_t^{-1} \gamma} \exp \left( (y - \mu)' (y - \mu) \right) \right)}{\left( \sqrt{(v + m_t) \gamma' H_t^{-1} \gamma} \right)^{(v+K)/2} (1 + m_t/v)^{(v+K)/2}},$$

(26)

where $m_t = (y - \mu)' H_t^{-1} (y - \mu)$, and normalizing constant

$$C_t = \frac{2^{1-(v+K)/2}}{\Gamma \left( \frac{v}{2} \right) \pi^{1/2} |H_t|^{1/2}}.$$

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

McNeil et al. (2005, Ch. 3) mention this distribution as having potential applications in finance, while Aas and Haff (2006) work with the univariate version of (26), 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 Option Pricing

We present a feasible technique which allows for multivariate option pricing in the framework of our model. Since the work of Duan (1995), GARCH models have become increasingly popular in option pricing. More recent literature includes Heston and Nandi (2000), who derive a nearly closed-form pricing formula under normal return innovations and the valuation assumption from Duan (1995); Christoffersen et al. (2006), who propose a model with inverse Gaussian innovations which allows for conditional skewness; Barone-Adesi et al. (2008), who use filtered historical simulation; Christoffersen et al. (2010), who develop a theoretical framework for option valuation under very general assumptions which allow for conditional heteroskedasticity and non-normality; and Rombouts and Stentoft (2011), who consider multivariate option pricing in a model with a finite-mixture-of-normal. Our model is also multivariate and, as it allows for all the primary stylized facts of asset returns, it is expected to be a good candidate for option pricing, given a feasible calibration algorithm.

The proposed algorithm combines the equivalent martingale measure (EMM) technique in the presence of a GARCH structure as in Christoffersen et al. (2010), with an iterative estimation of the model dynamics. Like in Barone-Adesi et al. (2008), it does not focus on the analytical form of the change of measure. Barone-Adesi et al. (2008) utilize a Monte Carlo simulation based on QML estimates of model parameters under the historical measure and calibrate the EMM on the option prices. In contrast to their nonparametric method, our algorithm estimates the model
parameters via maximizing the complete data likelihood function under the historical measure \( P \), it changes the measure as if \( G_t \) were observed, and it iteratively replaces the missing information about \( G_t \) with conditional expectation, under the new measure, of \( G_t \) given \( Y_t \). Crucially for the COMFORT model, the proposed algorithm does not suffer from the curse of dimensionality, and so, it is applicable in a multivariate setup with a large number of underlying assets.

Denote by \( X_t = (X_{t,1}, X_{t,2}, \ldots, X_{t,K})' \) a vector of prices of assets \( k = 1, 2, \ldots, K \) at time \( t \). The price of an option contract at time \( t \) with maturity \( T \) and terminal payoff function \( \vartheta (X_T) \), \( \vartheta \) being the set of relevant parameters such as the strike price, can be computed as the following discounted expectation,

\[
C_t (T, \vartheta) = \exp (- (T - t) r) \mathbb{E}^P \left[ \frac{dQ^*}{dP} \bigg| \Phi_t \right],
\]

where \( \frac{dQ^*}{dP} \bigg| \Phi_t \) is the change of measure such that the discounted stock price process under \( Q^* \) is a martingale with respect to the \( \Phi_t \) filtration. This filtration is defined in Section 2 and it is associated with the incomplete conditional density function \( f_{Y_t} \). For the purpose of option pricing, we define a complete information filtration\(^2\) associated with the complete information density \( f_{Y_t,G_t} \), by \( \mathcal{F}_t = \sigma (\{G_1, Y_1, G_2, Y_2, \ldots, G_t, Y_t\}) \); and a second filtration, which includes information about the realization of \( G_{t+1} \), \( \mathcal{F}^{t+G}_t = \sigma (\{G_1, Y_1, G_2, Y_2, \ldots, G_t, Y_t, G_{t+1}\}) \). Hence, \( \Phi_t \subseteq \mathcal{F}_t \subseteq \mathcal{F}^{t+G}_t \) and, more generally, for any \( t \), the following chain property for the latter two filtrations holds,

\[
\mathcal{F}_1 \subseteq \mathcal{F}_1^{t+G} \subseteq \mathcal{F}_2 \subseteq \mathcal{F}_2^{t+G} \subseteq \ldots \subseteq \mathcal{F}_t \subseteq \mathcal{F}_t^{t+G}.
\]

The mixture property of the MGHyp distribution implies that \( Y_t \big| \mathcal{F}_t \sim N (\mu + \gamma G_t, G_t \mathbf{H}_t) \), hence the standard theory for option pricing under normality applies. In particular, following Christoffersen et al. (2010), and Rombouts and Stentoft (2011), if we impose the exponential affine form on the Radon-Nikodym derivative, with respect to \( \mathcal{F}^{t+G}_t \), then, under the corresponding measure \( Q^{t+G} \), as detailed in Appendix B, the dynamics of the returns remain Gaussian although with a shift in the mean.

Next, we define a change of measure, under \( \mathcal{F}_t \), by

\[
\frac{dQ}{dP} \bigg| \mathcal{F}_t = \mathbb{E}^P \left[ \left( \frac{dQ^{t+G}}{dP} \bigg| \mathcal{F}^{t+G}_t \right) \bigg| \mathcal{F}_t \right].
\]

This transformation defines a Radon-Nikodym derivative under \( \mathcal{F}_t \) and the resulting measure \( Q \) is an EMM, under \( \mathcal{F}_t \). Note that, from (38) the information contained in \( \mathcal{F}_t \) is sufficient to construct \( \frac{dQ^{t+G}}{dP} \bigg| \mathcal{F}^{t+G}_t \). Therefore, the change of measure (29) is available explicitly and equal to (38).

Under the measure \( Q \)

\[
Y_t \big| \mathcal{F}^{t+G}_{t-1} \sim N \left( (\mu - r) + \gamma G_t + \frac{1}{2} \text{diag} (S_t^2) G_t, G_t \mathbf{H}_t \right),
\]

where \( r = (r, \ldots, r)' \) is a \( K \times 1 \) vector of the risk free interest rates \( r \).

Under no arbitrage conditions, the derivatives which are a function of the underlying stock price

\(^2\)Note an important difference that the complete information filtration does not imply directly that the market is complete. The presence of time varying volatilities induces incompleteness.
process can be priced as the expected value, under EMM, of their future cash flows discounted using the risk free interest rate, as given in (27). So far we have derived the necessary tools for pricing the option under the condition that the realizations of the \( G_t \) sequence are observed. In particular, note that \( \mathcal{F}_{t-1}^G \not\subseteq \Phi_t \) and \( \mathcal{F}_{t-1} \not\subseteq \Phi_t \), so a more general chain relation, than (28), does not hold, and one cannot rely on a simple two step procedure to obtain the EMM under \( \Phi_t \).

The algorithm in Appendix C mitigates this problem. Instead of analytically deriving the change of measure, it estimates the parameters of the model under the EMM \( \Phi^* \). This method implicitly defines the change of measure, and it is feasible even in case of a large number of assets. Moreover, it is expected to reduce the estimation error (even if there would be a feasible and consistent two step approach) because it directly focuses on the \( Q^* \) measure parameters instead of doing this in two steps (first ML under historical measure \( P \), and then calibration of the measure change).

Having obtained, the parameter estimates under the risk neutral measure \( Q^* \) the conditional distribution of the returns is given by

\[
Y_t | \Phi_{t-1}^Q \sim \text{MGHyp}\left(\mu^Q, \gamma^Q, H_t^Q, \theta_D^Q\right),
\]

where \( H_t^Q = S_t^Q \Gamma^Q S_t^Q \) is the dispersion matrix under \( Q^* \), with the scale term dynamics, in a diagonal matrix \( S_t^Q \), with the \( \theta_P \) parameters under \( Q^* \).

Given the distribution of the returns under the risk neutral measure, one can price various options by using (27), and Monte Carlo simulation.

### 6 Empirical Application

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 forecasting performance across different models. Summarizing the results, the MALap-CCC models and MALap-CCC hybrid GARCH(1,1)-SV model deliver the best density predictions among all considered models. In particular, we find (i) a large improvement moving from the normal (even with dynamic correlations) to any of the distributions discussed in Section 4; 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.
6.1 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

\[
\hat{H}_t^{-1/2} (Y_t - \hat{\mu}) \quad \text{and} \quad \hat{G}_t^{-1/2} \hat{H}_t^{-1/2} (Y_t - \hat{\mu} - \hat{\gamma} \hat{G}_t),
\]

respectively; where \(\hat{G}_t\) are the imputed values of \(G_t\) returned from the ECME algorithm, \(\hat{H}_t\) are fitted conditional dispersion matrices, and other hatted entries denote parameter estimates. Observe that both sets of residuals in (32), in particular the latter, are assumed to be Gaussian under each of the assumed models. In Figures 1 and 2 we provide the Q-Q plots of the residuals of three (Gaussian, MALap, and MA\(t\) 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 1 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 2 we compare the fit of the MALap-CCC Hybrid GARCH(1,1)-SV model and the MA\(t\)-CCC GARCH(1,1) model. The latter model in the univariate case, usually called \(t\)-GARCH, is well known for providing an excellent model fit. Here, from Figure 2, 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 MA\(t\) performs slightly better.

Now consider the filtered \(G_t\) sequence. Figure 3 illustrates its impact on one of the assets, Merck & Co. The top panel gives the returns, while the second panel shows the filtered \(\hat{G}_t\) values from the ECME algorithm computed from (20). 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 (8), are plotted over time. The panel in the last row combines the above factors and plots the \(Y_{k;t}\) volatilities, as defined in (12) and based on the parameter estimates. A very negative spike in the second quarter of the data is synchronic with a large spike in the \(\hat{G}_t\) 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 (11), 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 \(\gamma\). Hence, the impact of \(G_t\) on volatilities differs across assets. Figure 4 illustrates this fact. It is a multivariate analogue of Figure 3 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
Figure 1: 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.

volatilities of different assets.

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 (4), a possible persistence in $G_t$. We model only the dynamics in the parameters 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 $\mathbb{E}[G_t | \Phi]$. The former are either constant over time (when $G_t$ are iid) or, with $G_t | \Phi_{t-1}$ having time-varying parameters. The latter, $\mathbb{E}[G_t | \Phi]$, 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 5 we compare $\mathbb{E}[G_t | \Phi]$ 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 (4). The latter
model clearly results in better fit of the common market factor. The $E[G_t | \Phi_{t-1}]$ estimates match 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 5 compare the resulting conditional volatilities from the two models. Again, the conditional volatilities from the hybrid extension (computed from (11) with use of $E[G_t | \Phi_{t-1}]$) lie much closer to the filtered values (computed from (11) with use of $E[G_t | \Phi_t]$).

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 6 displays the correlations between $E[G_t | \Phi_{t-1}]$ 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 3: 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 (11) and based on the parameter estimates.

Figure 7 displays the higher-order dynamics implied by the MALap-CCC hybrid GARCH(1, 1)-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 $k = 1, 2, \ldots, 30$, 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 $E[G_t | \Phi_{t-1}]$ 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 $E[G_t | \Phi_{t-1}]$. From this panel and the second panel in Figure 5, one can note that the kurtosis and $E[G_t | \Phi_{t-1}]$ are inversely related, i.e., the lower the value of $E[G_t | \Phi_{t-1}]$, the higher the value of the kurtosis. In fact, the correlation between $E[G_t | \Phi_{t-1}]$ and the conditional kurtosis is equal to $-0.81$.

### 6.2 Density Forecasting Performance Comparison

Now turning to out-of-sample forecasting performance, this section compares a number of special cases of model (1) with the CCC model of Bollerslev (1990), the DCC model of Engle (2002), the cDCC model of Aielli (2011), and the VC model of Tse and Tsui (2002), all denoted with a prefix
Figure 4: 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 \( k = 1, \ldots, K \), implied by the estimates of the GARCH(1,1) models. **Fourth row:** The conditional volatilities of \( Y_t \), computed as the square root of the elements on the diagonal of matrix (11) and based on the parameter estimates.

MN-, for multivariate normal distribution of the innovations. For each model, both GARCH(1,1) and GJR-GARCH(1,1) univariate dynamics are employed.

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 Paolella (2013), we use the normalized sum of the realized predictive log-likelihood values, which, for given model \( \mathcal{M} \), is

\[
S_T(\mathcal{M}) = \frac{1}{T} \sum_{t=1}^{T} \pi_t(\mathcal{M}), \quad \text{where} \quad \pi_t(\mathcal{M}) = \log f_{t+1|t}(Y_{t+1} | \theta).
\]  

(33)

In case of the hybrid GARCH-SV models we use, a first order approximation to \( \pi_t \), and replace random parameters of \( G_{t+1} \) with the values implied by the conditional expectations \( \mathbb{E}[G_{t+1} | \Phi_t] \).

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 is the MA\( t \) model, followed by the MNIG and MALap models (without hybrid dynamics). The
MALap−CCC GARCH(1,1) model \((G_{t} \text{iid})\)

\[ E[G_{t}] = \lambda \]  

\[ E[G_{t} | \Phi_{t}] \text{ (filtered from EM)} \]

MALap−CCC Hybrid GARCH(1,1)−SV model

\[ E[G_{t} | \Phi_{t-1}] = \lambda_{t} \]  

\[ E[G_{t} | \Phi_{t}] \text{ (filtered from EM)} \]

Figure 5: The consequences of the hybrid GARCH(1,1)-SV extension. **First row:** The filtered \(\hat{G}_{t}\) values from the ECME algorithm (i.e., \(E[G_{t} | \Phi_{t}]\)) 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., \(E[G_{t} | \Phi_{t}]\)) and the estimates obtained from the MALap-CCC hybrid GARCH(1,1)-SV model. **Third row:** Conditional volatilities filtered from the ECME algorithm (vol\(_{t} | t\)) and the estimates obtained from the MALap-CCC GARCH(1,1) model (vol\(_{t} | t-1\)). **Fourth row:** Conditional volatilities filtered from the ECME algorithm (vol\(_{t} | t\)) and the estimates obtained from the MALap-CCC hybrid GARCH(1,1)-SV model (vol\(_{t} | t-1\)).

Figure 6: Correlation between \(E[G_{t} | \Phi_{t-1}] = \lambda_{t}\) and conditional volatility, vol\(_{11}\), of each of the assets filtered from the ECME algorithm (MALap-CCC hybrid GARCH(1,1)-SV model).

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.
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 8 plots two tail quantiles, the means, and the medians of the estimates of the $\eta_k$, $k = 1, \ldots, 30$, from (9), 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 clearer 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 $k = 1, \ldots, 30$). It turns out that, for those windows, and for all the distributions considered (MN, MALap, MNIG, and $t$), the models with GJR-GARCH(1, 1) significantly outperform their plain GARCH counterparts, but the improvement 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 I) and the Gaussian-based models (third panel in Table I) is much larger than the gap between any models in a given panel.

In order to statistically test the forecasting results from Table I, we use the test for unconditional predictive ability of Diebold and Mariano (2002) (see also Giacomini and White, 2006). We use a one sided test ($M_1 \succ M_2$) and compare each model, $M_1$, in Table I with models $M_2$ which resulted in a worse-than-model-$M_1$ forecast. Tables of the test results are given in Paolella and Polak (2013).

Summarizing, the first three models from Table I are very competitive and, according to the
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 (33). 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.

<table>
<thead>
<tr>
<th>$\mathcal{M}$</th>
<th>$S_T(\mathcal{M})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MALap-CCC GARCH(1,1)-SV</td>
<td>−45.873</td>
</tr>
<tr>
<td>MNIG-CCC GARCH(1,1)-SV</td>
<td>−45.879</td>
</tr>
<tr>
<td>MALap-CCC GARCH(1,1)</td>
<td>−45.909</td>
</tr>
<tr>
<td>MNIG-CCC GARCH(1,1)</td>
<td>−45.936</td>
</tr>
<tr>
<td>MALap-CCC GARCH(1,1)</td>
<td>−45.978</td>
</tr>
<tr>
<td>MALap-CCC GJR-GARCH(1,1)-SV</td>
<td>−46.113</td>
</tr>
<tr>
<td>MALap-CCC GJR-GARCH(1,1)</td>
<td>−46.116</td>
</tr>
<tr>
<td>MNIG-CCC GJR-GARCH(1,1)-SV</td>
<td>−46.164</td>
</tr>
<tr>
<td>MALap-CCC GJR-GARCH(1,1)</td>
<td>−46.173</td>
</tr>
<tr>
<td>MNIG-CCC GJR-GARCH(1,1)</td>
<td>−46.197</td>
</tr>
<tr>
<td>MALap-IID</td>
<td>−47.097</td>
</tr>
<tr>
<td>Normal-DCC GARCH(1,1)</td>
<td>−47.670</td>
</tr>
<tr>
<td>Normal-cDCC GARCH(1,1)</td>
<td>−47.670</td>
</tr>
<tr>
<td>Normal-cDCC GJR-GARCH(1,1)</td>
<td>−47.682</td>
</tr>
<tr>
<td>Normal-DCC GJR-GARCH(1,1)</td>
<td>−47.685</td>
</tr>
<tr>
<td>Normal-VC GARCH(1,1)</td>
<td>−47.700</td>
</tr>
<tr>
<td>Normal-VC GJR-GARCH(1,1)</td>
<td>−47.709</td>
</tr>
<tr>
<td>Normal-CCC GARCH(1,1)</td>
<td>−47.787</td>
</tr>
<tr>
<td>Normal-CCC GJR-GARCH(1,1)</td>
<td>−47.796</td>
</tr>
</tbody>
</table>

Figure 8: Two tail quantiles, mean, and median of $\eta_k$, $k = 1, \ldots, 30$ 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.

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 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 Paolella, 2013.) 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 \( \pi_t (M_1) \pi_t (M_2) > 0 \))

\[
D_t (M_1, M_2) = 100 \left( \frac{|\pi_t (M_1)| - |\pi_t (M_2)|}{|\pi_t (M_2)|} \right) .
\] (34)

In Figure 9, 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.

### 6.3 Mean Forecasts

Lastly, we consider 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 10 compares the forecasts of the conditional means based on (i) the sample mean and median from a rolling window of 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 (4) 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 10, are more varying, because the conditional mean is a function of the \( G_t | \Phi_{t-1} \) parameters as in (11). One could consider more general SV dynamics incorporating the moving average component into (4). This would smooth the forecasts in the last panel of Figure 10 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 \( \hat{\gamma} / (1 - \hat{\rho}) \) \( / E [G_t | \Phi_{t-1}] \), where the hatted values come from estimation. This has the effect of removing the impact of the spikes in \( E [G_t | \Phi_{t-1}] \) in the mean equation.
Figure 9: The forecasting gains from the hybrid GARCH(1, 1)-SV extension. **First row, column-wise:** Percentage gains $D_t$ from (34) as a function of average absolute return. Using MALap-CCC GJR-GARCH(1, 1) as $M_1$ and MALap-CCC GARCH(1, 1) as $M_2$. 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 (34) as a function of time. Same, but for crisis of 2008. Histogram of percentage gains during the 2008 crisis.
Figure 10: Conditional mean forecasts from a rolling window of 1,000 observations. **First row:** Sample mean. **Second row:** Sample median. **Third row:** The MA Lap-CCC GARCH(1,1) model. **Fourth row:** The MA Lap-CCC GARCH(1,1)-SV model.

This results in improved forecasting performance, but it was not statistically significant at the 5% level.

## 7 Conclusions

We introduce a new class of models which combines GARCH-type dynamics with an SV structure (hybrid GARCH-SV class). The former captures the asset-specific volatility clustering effects and the latter is responsible for common market shocks. The proposed model also allows for a new type of dynamic in the dependency structure leading to additional dynamics in the higher-order moments. Maximum likelihood estimation is numerically reliable and fast, and can be used with a large number of assets. It yields consistent and asymptotically normal estimates of the parameters. In- and out-of-sample exercises provide justification for use of the model with real data. The model delivers a non-Gaussian predictive distribution with a tractable sum of margins, and hence can be straightforwardly applied to portfolio optimization.

The model also lends itself to multivariate option pricing by combining the equivalent martingale measure technique in the presence of a GARCH structure with an iterative estimation of the model dynamics. Future work will pursue the theoretical properties, convergence analysis, and
empirical performance of the proposed option pricing algorithm.

In future research, one could entertain the use of the multivariate noncentral $t$ distribution, as used in Jondeau (2012). This also arises as a mixture similar to (1) but such that, in (1a), the square root of $G_t$ is used. As such, the noncentral $t$ is not a special case of the MGHyp, but is still such that its sums of marginals is noncentral $t$, and has the appealing property of asymmetric tail dependence.

The dependency matrix $\mathbf{\Gamma}$ could be augmented by information in high-frequency data, possibly along the lines of Noureldin et al. (2012). Finally, the CCC structure could be modified in the same vein as in the DCC model of Engle (2002), the VC model of Tse and Tsui (2002), and the switching CCC model of Pelletier (2006). These models, particularly the latter, will entail substantially more estimation time, but might yield improved density forecasts. These ideas are currently being pursued.

Appendices

A Link to the Taylor (1982) SV model

The univariate model proposed in Taylor (1982) is given by

$$Y_t = \exp \left( \frac{Q_t}{2} \right) Z_t, \quad \text{with} \quad Q_t = c + \rho Q_{t-1} + \sigma \eta_{t-1}, \quad (35)$$

for $Z_t \overset{\text{iid}}{\sim} N(0, 1)$; $\eta_t \overset{\text{iid}}{\sim} N(0, 1)$; and $Z_t$ and $\eta_t$ are independent. Define $Q_t$ by $G_t = \exp (Q_t)$. Then (1) can be rewritten as

$$Y_t = \mu + \gamma \exp (Q_t) + H_t^{1/2} \exp \left( \frac{Q_t}{2} \right) Z_t. \quad (36)$$

Switching from $(G_t | \Phi_{t-1})$ dynamics in (4) to $(Q_t | \Phi_{t-1})$ dynamics and dropping the conditional expectations, we get

$$(Q_t | \Phi_{t-1}) = c_r + \rho_r (Q_{t-1} | \Phi_{t-2}) + \zeta_{r,t}, \quad (37)$$

where $\zeta_{r,t} = (Q'_t | \Phi_t) - (Q'_t | \Phi_{t-1})$. Setting $\mu = 0$, $\gamma = 0$, $r = 1$ and $H_t = 1$ in (36) (for $K = 1$) results in (36) reducing to (35), with $\eta_{t-1}$ replaced by $\zeta_{t,t}$ and $(Q_t | \Phi_{t-1})$ instead of $Q_t$.

Our model differs from the SV model in three additional aspects. Firstly, we replace the past shock $\eta_{t-1}$ by the current shock $\eta_t$ in (35). In the SV literature, one has to use a lag shock (together with $\text{Corr} (Z_t, \eta_t) < 0$) to obtain the key feature of SV models, the asymmetric return-volatility relation, often called a statistical leverage effect. In our setup, we can incorporate the asymmetry or the leverage effect through the scale-term dynamics.

Note that, if we were to use $\zeta_{r,t-1}$ instead of $\zeta_{r,t}$ in our model, then there would be a one-period shift between the filtered $\hat{G}_t$ values, and the moments of $G_t$ conditional on $\Phi_{t-1}$. Use of $\zeta_{r,t}$ avoids this and allows for shocks to the volatility to have an immediate impact on returns, a feature which is absent in discrete time SV models.

The second difference is: we work with $G_t | \Phi_{t-1}$ instead of $\log G_t | \Phi_{t-1}$ because the former has tractable moment expressions, and we can still guarantee positive values of $E[G_t | \Phi_{t-1}]$ by
imposing the constraints $c_r > 0$ and $\rho_r \geq 0$.

Thirdly, the dynamics in our model are written in terms of $\mathbb{E} [G_t \mid \Phi_{t-1}]$ because $G_t$ is a latent process. Here, we are able to filter them out through our ECME algorithm without an additional computational burden. More importantly, keeping the dynamics of $G_t$ only in terms of conditional expectations allows us to maintain, without any extra conditions, the monotonic increase of the incomplete-data likelihood which is a key property of the ECME algorithm.

### B Derivation of the $Q^{+G}$-Dynamics for Option Pricing

Following [Christoffersen et al. (2010)](Christoffersen2010) and a multivariate extension given in [Rombouts and Stentoft (2011)](Rombouts2011), we impose the exponential affine form on the Radon-Nikodym derivative, with respect to $\mathcal{F}_t^{+G}$. Hence, by the law of iterative expectation, we get

**Lemma B.1.** For any $K$-dimensional sequence $\mathbf{v}_s$,

$$
\frac{dQ^{+G}}{dP} \mid \mathcal{F}_t^{+G} = \exp \left( - \sum_{s=1}^{t} \left( \mathbf{v}_s' \mathbf{\varepsilon}_s + \frac{1}{2} \mathbf{v}_s' \mathbf{H}_s \mathbf{v}_s G_s \right) \right)
$$

is a Radon-Nikodym derivative with respect to $\mathcal{F}_t^{+G}$.

**Proof.** We need to show that $\frac{dQ^{+G}}{dP} \mid \mathcal{F}_t^{+G} > 0$ and $\mathbb{E}_0^P \left[ \frac{dQ^{+G}}{dP} \mid \mathcal{F}_t^{+G} \right] = 1$. Non-negativity is an immediate consequence of the exponential form. For the second condition we use the law of iterated expectations, with respect to $\mathcal{F}_t^{+G}$, to obtain

$$
\mathbb{E}_0^P \left[ \frac{dQ^{+G}}{dP} \mid \mathcal{F}_t^{+G} \right] = \exp \left( - \sum_{s=1}^{t} \left( \mathbf{v}_s' \mathbf{\varepsilon}_s + \frac{1}{2} \mathbf{v}_s' \mathbf{H}_s \mathbf{v}_s G_s \right) \right)
$$

where the second last equality follows from the normality of $\mathbf{\varepsilon}_t$ conditional on $\mathcal{F}_{t-1}^{+G}$. Iterating on this yields the required result. \qed

Having a valid candidate for the change of measure, we proceed to find conditions for the sequence $\mathbf{v}_s$ under which the proposed Radon-Nikodym derivative defines an EMM under $\mathcal{F}_t^{+G}$. Denote by $\mathbf{r} = (r, r, \ldots, r)'$ a $K$ vector of risk free interest rates, then
**Proposition B.1.** The probability measure $Q^{+G}$ defined by the Radon-Nikodym derivative in (38) is an EMM under $\mathcal{F}_t^{+G}$ if and only if

$$v_t = H_t^{-1}\left((\mu - r)G_t^{-1} + \gamma + \frac{1}{2}\text{diag}(S_t^2)\right).$$

**Proof.** We need to show that, for all $k = 1, 2, \ldots, K$, $\mathbb{E}^{Q^{+G}}\left[\frac{X_{t,k}}{X_{t-1,k}}|\mathcal{F}^{+G}_{t-1}\right] = \exp(r)$, where $r$ is the risk free interest rate, and $X_{t,k}$ is the price of stock $k$ at time $t$. We have

$$\mathbb{E}^{Q^{+G}}\left[\frac{X_{t,k}}{X_{t-1,k}}\exp(-r)|\mathcal{F}^{+G}_{t-1}\right] =$$

$$= \mathbb{E}^{P}\left[\frac{dQ^{+G}}{dP}\left|\mathcal{F}_t^{+G}\right.\right]
\left(\frac{dQ^{+G}}{dP}\left|\mathcal{F}_{t-1}\right.\right)
\frac{X_{t,k}}{X_{t-1,k}}\exp(-r)|\mathcal{F}^{+G}_{t-1}\right]$$

$$= \mathbb{E}^{P}\left[\frac{dQ^{+G}}{dP}\left|\mathcal{F}_t^{+G}\right.\right]
\frac{X_{t,k}}{X_{t-1,k}}\exp(-r)|\mathcal{F}^{+G}_{t-1}\right]$$

$$= \mathbb{E}^{P}\left[\exp\left(-\mathbf{v}'_t\mathbf{H}_t\mathbf{v}_t G_t\right)\exp\left(\mu_k + \gamma_k G_t + \varepsilon_t\right)\exp(-r)|\mathcal{F}^{+G}_{t-1}\right]$$

$$= \exp\left(-\frac{1}{2}\mathbf{v}'_t\mathbf{H}_t\mathbf{v}_t G_t + \mu_k + \gamma_k G_t - r - \mathbf{v}'_t\varepsilon_t + e'_k e_t\right)$$

$$= \exp\left(-e'_k H_t v_t G_t + \frac{1}{2} e'_k H_t e_k G_t + \mu_k + \gamma_k G_t - r\right),$$

where $e_k = (0, \ldots, 0, 1, 0, \ldots, 0)'$ a vector of zeros with one at position $k$. Thus, if we ensure that

$$-e'_k H_t v_t G_t + \frac{1}{2} e'_k H_t e_k G_t + \mu_k + \gamma_k G_t - r = 0$$

for all $k = 1, 2, \ldots, K$, by choosing the vector series $v_t$, then the probability measure $Q^{+G}$ is an EMM since it makes discounted asset prices martingales. Solving it for $v_t$, in vector form, we obtain

$$v_t = H_t^{-1}\left((\mu + \gamma G_t - r)G_t^{-1} + \frac{1}{2}\text{diag}(S_t^2)\right).$$

Thus, if we would know the realizations of $G_s$ for $s = 1, 2, \ldots, t$, then it is possible to solve explicitly for the respective $v_s$, and Proposition B.1 guarantees that the corresponding measure is an EMM under $\mathcal{F}_t^{+G}$.

Use of the $\mathcal{F}_t^{+G}$ filtration implies that, although we are working within an incomplete market framework, there is only one source of randomness. Therefore, constraining the Radon-Nikodym derivative to be of the exponential affine form, as in (38), allows us to derive a unique measure under which the discounted asset prices are $Q^{+G}$-martingales. Moreover, because $Y_t | G_t$ is Gaussian, we can characterize the change of measure and the $Q^{+G}$-dynamics corresponding to model (11) explicitly. Denote by $\Psi_t(u)$ the logarithm of the conditional moment generating function.
of $\varepsilon_t$ given $\mathcal{F}^+_t$, i.e.,

$$\Psi_t(u) = \frac{1}{2} u^T H_t u G_t. \quad (39)$$

In order to obtain the $Q^+G$-dynamics of our model, we derive the analogue of $\Psi_t$ under $Q$.

**Corollary B.1.** The logarithm of the conditional (on $\mathcal{F}^+_t$) moment generating function of $\varepsilon_t$ under $Q^+G$ is given by

$$\Psi_{Q^+G}^+(u) = \log \mathbb{E}^{Q^+G}\left[ \exp(-u'\varepsilon_t) \mid \mathcal{F}^+_t \right] = \Psi_t(v_t + u) - \Psi_t(v_t). \quad (40)$$

**Proof.** By change of measure and rearranging we get

$$\mathbb{E}^{Q^+G}\left[ \exp(-u'\varepsilon_t) \mid \mathcal{F}^+_t \right] = \mathbb{E}^P\left[ \left. \left( \frac{dQ^+G}{dP} \right)_{\mathcal{F}^+_t} \right| \mathcal{F}^+_t \right] \exp(-u'\varepsilon_t) \mid \mathcal{F}^+_t \right]$$

$$= \mathbb{E}^P\left[ \exp(-v_t'\varepsilon_t - \Psi_t(v_t)) \exp(-u'\varepsilon_t) \mid \mathcal{F}^+_t \right]$$

$$= \mathbb{E}^P\left[ \exp(-(v_t + u)'\varepsilon_t - \Psi_t(v_t)) \mid \mathcal{F}^+_t \right]$$

$$= \exp(\Psi_t(v_t + u) - \Psi_t(v_t)).$$

Taking log of both sides completes the proof. \qed

Substituting (39) into (40), we get

$$\Psi_{Q^+G}^+(u) = u'H_t v_t G_t + \frac{1}{2} u'H_t u G_t. \quad (41)$$

Now, using the expression for $v_t$ given in Proposition B.1

$$\Psi_{Q^+G}^+(u) = u'(\mu - r) + u'\gamma G_t + \frac{1}{2} u'diag(S_t^2) G_t + \frac{1}{2} u'H_t u G_t.$$ 

So, the $Q^+G$-dynamics of the returns remain Gaussian with a shift in the mean, as in (30).

### C Iterative Change of Measure (ICM) Algorithm

If the returns follow the dynamics from the COMFORT model [1], then one can construct a sequence of measures $Q^{[\ell]}$ by the following algorithm. It is an extension of the ECME algorithm, detailed in Section 3, with an additional step in which the measure is changed. The algorithm uses the historical measure $P$ estimates as the starting values, and iterates the following steps, for $\ell = 1, 2, \ldots$

**E-step:** Calculate $\mathbb{E}^{Q^{[\ell]}_P} \left[ \log L_{Y,G} \mid Y; \hat{\theta}_P^{Q^{[\ell]}}, \hat{\theta}_D^{Q^{[\ell]}}, \hat{\theta}_C^{Q^{[\ell]}}, \hat{\theta}_C^{Q^{[\ell]}} \right]$. Computationally, this step is the same as under historical measure $P$; the difference is only in the distribution parameters.

**CM1-step:** Update $\theta_P, \theta_C$, i.e., Gaussian parameters under measure $P$, with the unobserved realizations of $G_t$'s replaced by their filtered values, under $Q^{[\ell]}$, from the E-step above.
Let \( D \) Evaluation of the Bessel Function

Iterate the above steps until \( besselk.m \) the Matlab function \( K \) which involves a ratio of Bessel functions \( N (\theta) \). Let \( G \gamma \theta \) of \( \theta \) EM-step Given the estimates of \( \theta \) and \( Q \) in Appendix \( B \) and obtain the estimates of \( \theta \).

CM2-step: Given the estimates of \( \theta \), obtain new estimates of \( \theta \) by maximizing the incomplete data log-likelihood function, i.e., compute

\[
\hat{\theta}^{Q(t+1)}_{D} = \arg \max_{\theta_D} L_{Y} \left( \theta_D \mid \hat{\theta}^{Q(t+1)}_{P}, \hat{\theta}^{Q(t+1)}_{C} \right),
\]

where

\[
Y_t \mid \Phi_{t-1}^{Q(t+1)} \sim MGHyp \left( \hat{\mu}^{Q(t+1)}, \hat{\gamma}^{Q(t+1)}, \hat{H}^{Q(t+1)}, \theta_D \right),
\]

and \( \hat{H}^{Q(t+1)} = \hat{S}^{Q(t+1)} \Gamma^{Q(t+1)} \hat{S}^{Q(t+1)} \) is the dispersion matrix under \( Q(t+1) \), with the scale term dynamics, in a diagonal matrix \( \hat{S}^{Q(t+1)} \), with the \( \hat{\theta}^{Q(t+1)} \) parameters from the EMM step above.

Iterate the above steps until

\[
\left| \log L_{Y} \left( \hat{\theta}^{Q(t+1)}_{P}, \hat{\theta}^{Q(t+1)}_{C}, \hat{\theta}^{Q(t+1)}_{D} \right) \right| \leq \varepsilon,
\]

for some fixed \( \varepsilon > 0 \).

D Evaluation of the Bessel Function

Let \( G \sim GIG (\lambda, \chi, \psi) \), for \( \chi > 0 \) and \( \psi > 0 \). Then it may be shown (see, e.g., Paolella 2007 Ch. 9) that

\[
\mathbb{E} [G^\alpha] = \left( \frac{\chi}{\psi} \right)^{\alpha/2} \frac{K_{\lambda+\alpha} \left( \sqrt{\chi \psi} \right)}{K_{\lambda} \left( \sqrt{\chi \psi} \right)}, \quad \alpha \in \mathbb{R},
\]

which involves a ratio of Bessel functions \( K_{\lambda} (z) \) as given in [3]. For large \( z \) (\( z > 700 \)) and \( \nu < 55 \), the Matlab function \texttt{besselk.m} returns 0 and hence the ratio is undefined.

It is possible to compute the limit of the Bessel function ratio for some cases. Let \( Y \mid (G = g) \sim N (\mu - \gamma g, g \Sigma) \). We are interested in the expectations of \( G_{\pm} \mid (Y = y) \) which, for \( \psi \neq 0 \) or \( \gamma \neq 0 \),
are always positive and have their limits given by

\[
\lim_{m \to \infty} \left( \frac{m + \chi}{\psi + \gamma \Sigma^{-1} \gamma'} \right)^{\pm 1/2} \frac{K_{\lambda-K/2 \pm 1} \left( \sqrt{(m + \chi) \left( \psi + \gamma \Sigma^{-1} \gamma' \right)} \right)}{K_{\lambda-K/2} \left( \sqrt{(m + \chi) \left( \psi + \gamma \Sigma^{-1} \gamma' \right)} \right)} = 0,
\]

where \( m = (y - \mu)' \Sigma^{-1} (y - \mu) \).

In our model, these ratios are responsible for proper weights, in the E-step and CM1-step of the ECME algorithm. We thus require a highly accurate approximation of Bessel function ratios for large \( v \) or \( z \). This can be done by using the asymptotic expansion of \( K_v(z) \) given by

\[
K_v(z) = \sqrt{\frac{\pi}{2z}} \exp(-z) E(v, z),
\]

where

\[
E(v, z) = 1 + \sum_{k=1}^{\infty} \frac{\prod_{l=1}^{k} \left( 4v^2 - (2l - 1)^2 \right)}{k! (8z)^k}.
\]

Inspection of (46) reveals that, for a ratio of Bessel functions, a numerically problematic \( \exp(z) \) cancels out. In order to use (46), we have to truncate the series at some finite \( K \) which causes, for \( z \) small \( (z < 10) \), some loss of accuracy (compared to the Matab function \( 	ext{besselk.m} \)), but as \( z \) increases, the accuracy grows very rapidly because of \( z \) to the \( k \) power in terms of the series (47).

References


