Robust Forecasting of Dynamic Conditional Correlation GARCH Models

Kris Boudt
KU Leuven/Lessius
and V.U. University Amsterdam
kris.boudt@econ.kuleuven.be

Jón Daníelsson
London School of Economics.
j.danielsson@lse.ac.uk

Sébastien Laurent*
Maastricht University, Department of Quantitative Economics
and CORE, Université catholique de Louvain.
s.laurent@maastrichtuniversity.nl

April 18, 2012

Abstract

Large once–off events cause large changes in prices but may not affect volatility and correlation dynamics as much as smaller events. Standard volatility models may deliver biased covariance forecasts in this case. We propose a multivariate volatility forecasting model that is accurate in the presence of large once–off events. The model is an extension of the dynamic conditional correlation model (DCC) model. In our empirical application to forecasting the covariance matrix of the daily EUR/USD and Yen/USD return series, we find that, compared to the DCC model, our method produces more precise out–of–sample covariance forecasts. Furthermore, when used in portfolio allocation, it leads to portfolios with similar return characteristics but lower turnover and hence higher profits.

*The authors thank the Dutch science foundation for financial support, as well as Esther Ruiz (the Editor) and two anonymous referees for their helpful comments and advice. Correspondence to: Sébastien Laurent, Department of Quantitative Economics, Maastricht University, School of Business and Economics, P.O. Box 616, 6200 MD Maastricht, The Netherlands. Tel.: +31 43 388 38 43. Fax.: +31 43 388 48 74.

1 Introduction

Prices of financial assets sometimes exhibit large jumps caused by once–off events, such as news announcements and it is often found that such extreme returns affect volatility less than a standard GARCH model would predict.[1] Using standard GARCH in such cases therefore leads not only to an overestimation of volatility for the days following the event, but, since the unconditional volatility forecast is upward biased, all volatility forecasts tend to be larger than they otherwise would be. A similar argument applies for correlation estimates. If only one of the stocks is subject to a large jump in prices, it biases the correlation estimates towards zero. In the case of co–jumps of the same (opposite) sign, the correlations are biased toward (minus) one.

Our objective in this paper is the development of a multivariate volatility forecasting model that is accurate in the presence of once–off events causing large changes in prices whilst not affecting volatility dynamics. There are two main directions one could take in the development of such a model. Either by explicitly modeling a jump process within a standard volatility model or by employing a robust estimation procedure for a standard volatility model. The former approach is necessary in applications where the properties of the jumps are of interest. However, jumps in daily returns are rare events and estimates of the jump process have large confidence bands. If the ultimate objective is forecasting volatility, a robust approach may therefore be a better choice, and this is what we do in this paper.

Our starting point is the univariate procedure proposed by Muler and Yohai (2008) for the estimation of GARCH models, whereby the impact of returns on volatility forecasts is bounded. They term this procedure as "bounded innovation propagation" (BIP) GARCH. We adjust their procedure to make it suitable for multivariate volatility forecasting when the underlying assets are subject to once–off shocks. This is the main contribution of the paper.

The most widely used models for forecasting conditional covariances and correlations

[1] See e.g. Andersen, Bollerslev, and Diebold (2007), Bauwens and Storti (2009) and Carnero, Peña, and Ruiz (2012).

are the BEKK model of Engle and Kroner (1995) and the dynamic conditional correlation model (DCC) of Engle (2002). Boudt and Croux (2010) proposed a robust estimation method for the BEKK model. Their robust model bounds the impact of jumps on the conditional variance of each asset based on the extremeness of the complete lagged return vector. This makes their model not suited for volatility forecasting since on the days following a jump in one of the assets, it underestimates the volatility of the assets that have not jumped. To avoid this and preserve the positive definiteness of the covariance forecasts, we choose to disentangle the robust forecasting of univariate volatilities and correlations and take the Aielli (2009) version of the dynamic conditional correlation model (cDCC) of Engle (2002) as our baseline model.

We make three extensions to the cDCC model. First, we use BIP–GARCH for the univariate volatilities instead of a standard GARCH. Second, we bound the impact of large innovations on the correlation matrix, through a BIP procedure in the update equation of the conditional correlation, along with a robust procedure to estimate the unconditional correlation. Finally, we propose a robust M–estimator for the parameters of the correlation dynamics. These three extensions ensure that extreme once–off events have little influence on the covariance predictions made by the proposed BIP–cDCC model.

We compare our model with the baseline models (standard GARCH and cDCC) by a variety of means. First, we run a Monte Carlo experiment to study the effect of jumps on the cDCC parameter estimates. The results indicate that estimates from the baseline model can result in a large bias when the data has additive jumps, while the BIP procedure provides accurate estimates of the parameters.

Second, the conditional covariance forecasts are compared with ex post covariance estimates based on high–frequency data. Using the model confidence set methodology, proposed by Hansen, Lunde, and Nason (2011), we find that the BIP models always belong to the set of superior forecasting models. Moreover, for most forecast horizons, their covariance forecasts are significantly better than all other models considered.

Our key empirical conclusion follows from applying the BIP model to the problem of optimal portfolio allocation where we study an investor who adopts a volatility timing strategy where the changes in ex ante optimal portfolio weights are solely determined by the forecasts of the conditional covariance matrix. We find that the portfolio returns when using the BIP method have similar unconditional first and second moments as when the baseline model is employed. However, by using the BIP procedure we increase profits because the BIP conditional covariance matrices are more stable, resulting in lower portfolio churn and thus lower transaction costs. Therefore, even if both procedures have same mean return and standard deviation, net of transaction costs, the BIP procedure yields more profits overall.

The structure of the paper is as follows. Section 2 describes the univariate BIP–GARCH volatility forecast model. Section 3 then proposes the robust conditional correlation forecasting method. These forecasts rely on robust M-estimates of the dependence parameters, a BIP–cDCC conditional correlation model and a robust estimate of the unconditional correlation. Section 4 reports the results of the Monte Carlo study on the effect of jumps on parameter estimates of the cDCC and BIP–cDCC models. Section 5 evaluates the forecasting precision of the models, using high–frequency data covariance estimates as proxies for the true covariance. Section 6 analyzes the economic consequences of using the BIP–cDCC model for a minimum variance investor. Finally, Section 7 summarizes our main findings and points out some directions for future research.

2 Univariate Volatility Forecasting in the Presence of Extremes

Many volatility models, such as GARCH, are based on the assumption that each return observation has the same relative impact on future volatility, regardless of the magnitude of the return. This assumption is at odds with an increasing body of evidence indicating that the largest return observations have a relatively smaller effect on future volatility than smaller shocks (see for instance Andersen, Bollerslev, and Diebold, 2007).

One reason is the occurrence of extremely large shocks caused by once–off events that cannot be expected to influence future volatility much. In the webappendix, this is illustrated with the example of the stock price of Apple, which fell 52% on September 29, 2000 after it warned its fourth–quarter profit would fall well short of Wall Street forecasts.

Several proposals for explicitly addressing how extreme returns affect volatility have been made, e.g. Andersen, Bollerslev, and Diebold (2007) and Corsi, Pirino, and Renó (2008) who use a simple restricted autoregressive model to forecast the realized volatility. They show that decomposing volatility into a jump component and a continuous component results in the jump component being considerably less persistent than the continuous component.

Carnero, Peña, and Ruiz (2006), Franses and Ghijsels (1999), and Muler and Yohai (2008), among others, propose new methods designed to estimate the parameters of a GARCH(1,1) model in the presence of additive, but once–off, jumps. After subtracting the mean μ, the observed return series s_t^* has a standard normal GARCH component y_t and a jump component a_t, i.e.

$$
\begin{aligned}
s_t &= s_t^* - \mu = y_t + a_t && (2.1)\\
y_t &= \sqrt{h_t} z_t \text{ where } z_t \overset{i.i.d.}{\sim} N(0,1) && (2.2)\\
h_t &= \omega + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1}. && (2.3)
\end{aligned}
$$

The assumption of normal innovations could by replaced by another distributional hypothesis.

In the economic literature, the occurrence of additive jumps is mostly modeled by means of a Poisson distribution, such as in Vlaar and Palm (1993) and Chan and Maheu (2002). Such a parametric approach requires one to further specify and estimate a model governing the time–varying jump intensity and also the jump size. This leads to an increased complexity of the estimation method (especially in a multivariate framework) and a potential specification bias. Moreover, because of the low frequency of extremes in the sample, these estimates often have wide confidence bands.

If the ultimate objective is forecasting volatility, inference on the jump process is not needed to produce accurate volatility forecasts. In the "robust" approach, jumps are automatically detected in the estimation step and their effect on parameter estimation and volatility forecasts is bounded.

In absence of jumps (i.e., when $a_t = 0 \ \forall t$), model (2.1)-(2.3) reduces to a standard GARCH(1,1) with normal innovations. This model is usually estimated by (Q)ML. When $a_t \neq 0$ for some $t \in \{1, \ldots, T\}$, y_t and a_t are not directly distinguishable from s_t^*. In this case the Gaussian QML is not appropriate because a_{t-1} has no impact on h_t while assuming a GARCH(1,1) for s_t^* would imply $h_t = \omega + \alpha_1 (y_{t-1} + a_{t-1})^2 + \beta_1 h_{t-1}$, i.e., a large and slowly decaying effect of a_{t-1} on future volatility predictions. Furthermore, if $\mathrm{E}(a_t) \neq 0$, μ is no longer the unconditional mean of s_t^* and thus both its QML estimate and the empirical mean are expected to be strongly biased.

2.1 BIP–GARCH and robust M–estimator

Muler and Yohai (2008) (MY) show that one can limit the effect of a_t on the estimation of the parameters of the GARCH model by using a modified GARCH specification downweighing the effect of past jumps (a_{t-1}) on time t conditional variance. Their approach estimates the GARCH parameters and detects jumps jointly, by identifying returns as jumps when a return is an extreme outlier under the estimated GARCH model. Because of the time–varying volatility, extremes need to be identified by the squared devolatilized return s_{t-1}^2/h_{t-1} rather than the squared return itself. Otherwise, jumps would be overdetected on days with high volatility and underdetected on days with low volatility. This leads to an auxiliary GARCH(1,1) model with weights on extremes:

$$h_t = \omega + \alpha_1 w\left(\frac{s_{t-1}^2}{h_{t-1}}\right) s_{t-1}^2 + \beta_1 h_{t-1}, \tag{2.4}$$

where $w(\cdot)$ is a weight function. Since s_{t-1}^2/h_{t-1} is chi–square distributed with one degree of freedom if there is no jump a time $t-1$, it is natural to detect a jump occurrence in s_{t-1}^2 if s_{t-1}^2/h_{t-1} exceeds $k_{\delta,1}$, the δ quantile of the chi–square distribution with one degree of freedom. The weight function used by MY is given by

$$w^{\mathrm{MY}}_{k_{\delta,1}}(u) = \min\left(1, \frac{k_{\delta,1}}{u}\right). \tag{2.5}$$

Model (2.4) with weight function (2.5) is called *Bounded Innovation Propagation* (BIP)–GARCH since the effect of past shocks on future volatility is bounded. MY show that the combination of a BIP–GARCH with an outlier robust M–estimator considerably reduces the root mean squared error (RMSE) of the parameter estimates in presence of additive jumps.[2]

For the estimation of the BIP–GARCH model, MY recommend using a M–estimator that minimizes the average value of an objective function $\rho(\cdot)$, evaluated at the log–transform of squared devolatilized returns, i.e.

$$\hat{\theta}^{\mathrm{M}} = \mathrm{argmin}_{\theta\in\Theta}\frac{1}{T}\sum_{t=1}^{T}\rho\left(\log\frac{s_t^2}{h_t}\right). \tag{2.6}$$

For robustness, this ρ–function needs to downweight the extreme observations and hence the jumps. The choice of $\rho(\cdot)$ trades off robustness vs. efficiency. Based on a comparison of several candidate ρ–function in the webappendix (Boudt, Daníelsson, and Laurent, 2012), we recommend the one associated to the Student t_4 density function:

$$\rho_2(z) = -z + \sigma_{1,4}\rho_{t_{1,4}}(\exp(z)),$$

where

$$\rho_{t_{N,\nu}}(u) = (N+\nu)\log\left(1 + \frac{u}{\nu-2}\right) \tag{2.7}$$

and

$$\sigma_{N,\nu} = \frac{N}{E[\rho'_{t_{N,\nu}}(u)u]}, \tag{2.8}$$

with u a chi–squared random variable with N degrees of freedom. $\sigma_{N,4}$ is reported in Table 1 for $N = 1, 2, 5, 10$ and 50 and is needed to ensure Fisher consistency, as shown in Boudt and Croux (2010). Carnero, Peña, and Ruiz (2012) document the

[2]Harvey and Chakravarty (2008) propose an alternative weight function based on the score of the t distribution with ν degrees of freedom $w^{\mathrm{HC}}_{\nu}(u) = \frac{\nu+1}{\nu-2+u}$.

Table 1: Correction factor $c_{\delta,N}$ for the weighted variance estimator, the BIP–GARCH model and for the M–estimator of (c)DCC models with N–dimensional Gaussian innovations

	$c_{\delta,N}$					$\sigma_{N,4}$
N / δ	1	0.99	0.975	0.95	0.90	
1	1	1.0185	1.0465	1.0953	1.2030	0.8260
2	1	1.0101	1.0257	1.0526	1.1111	0.8258
5	1	1.0050	1.0122	1.0255	1.0542	0.8467
10	1	1.0028	1.0073	1.0154	1.0330	0.8835
50	1	1.0009	1.0025	1.0053	1.0118	0.9644

good properties of the MY procedure in forecasting volatilities in the presence of additive outliers. We propose two modifications for the MY procedure that further improve the accuracy of the volatility forecasts.

2.2 A Modified MY Procedure

In the BIP–GARCH model of (2.4) also returns that are not affected by jumps are downweighted. Our simulations (available upon request) show that a smaller bias in the GARCH parameters based on the BIP specification are obtained when modifying the MY weighting scheme to ensure that the conditional expectation of the weighted squared unexpected shocks is still the conditional variance in absence of jumps, i.e.

$$w_{k_{\delta,1}}(u) = c_{\delta,1} w^{\mathrm{MY}}_{k_{\delta,1}}(u), \tag{2.9}$$

where

$$c_{\delta,N} = \frac{E[u]}{E[w^{\mathrm{MY}}_{k_{\delta,N}}(u)u]} = \frac{1}{F_{\chi^2_{N+2}}(\chi^2_N(\delta)) + (1-\delta)\chi^2_N(\delta)}, \tag{2.10}$$

with u a chi–square random variable with N degrees of freedom.[3] Table 1 reports the correction factors $c_{\delta,1}$ for various values of δ.

[3]The multivariate notation is used here, since the correction factor will also be useful for robust estimation of correlation parameters in Section 3.

A second modification of the MY procedure is that we integrate *reweighted* estimates of the mean and variance in the forecasting procedure. The above definitions of the BIP–GARCH model and M–estimators are for $s_t = s_t^* - \mu$. MY assume that $\mu = 0$ and thus only focus on the conditional variance. Unfortunately, this assumption may not hold in practice and a jump robust estimator of μ is therefore needed. Furthermore, MY estimate the intercept ω jointly with the parameters α and β. As noted by Engle and Mezrich (1996), this is especially difficult if α and β add up to a number very close to one, as the intercept will be very small but must remain positive.

Engle and Mezrich (1996) propose variance targeting as an estimation procedure where ω is reparameterized as $\hat{h}(1 - \alpha_1 - \beta_1)$ (with $\hat{h}$ a consistent estimator of the unconditional variance h) before estimating the remaining parameters. Francq, Horvath, and Zakoian (2011) show that when the model is misspecified, the variance targeting estimator can be superior to the QMLE for long–term prediction or Value–at–Risk calculations.

In absence of outliers, natural choices for $\hat{\mu}$ and $\hat{h}$ are the sample mean and the sample variance of the returns. However, these estimators are known to be very sensitive to outliers (e.g. outliers causing a large upward bias in the sample variance). We therefore propose to use robust reweighted mean and variance estimators proposed by Boudt, Croux, and Laurent (2011a) and described in the next subsection. In the webappendix (Boudt, Daníelsson, and Laurent, 2012), we verify the accuracy of the BIP M–estimator with Student t_4 loss function and targeting towards the robust reweighed mean and variance, relatively to the QML estimator and the estimators considered in MY.

2.3 Reweighted Mean and Variance Estimators

In a high-frequency data setting, Boudt, Croux, and Laurent (2011a) propose to estimate the unconditional mean μ and variance h through a mean and variance estimate in which local outliers receive a zero weight. The locality of the outlier detection method is needed in order to avoid an overdetection of outliers at times of high volatility and an underdetection when volatility is low (Boudt, Croux, and Laurent, 2011b). This method first estimates for each observation the median absolute deviation mad_t of the returns in a window around that observation.[4] The reweighted sample mean and variance are then

$$\hat{\mu} = \frac{\sum_{t=1}^{T} s_t^* I_t}{\sum_{t=1}^{T} I_t} \quad \text{and} \quad \hat{h} = 1.318 \cdot \frac{\sum_{t=1}^{T} (s_t^* - \hat{\mu})^2 J_t}{\sum_{t=1}^{T} J_t}, \tag{2.11}$$

with

$$I_t = I\left[\frac{(s_t^* - \text{median}_t(s_t^*))^2}{mad_t^2} \leq \chi_1^2(95\%)\right] \quad \text{and} \quad J_t = I\left[\frac{(s_t^* - \hat{\mu})^2}{mad_t^2} \leq \chi_1^2(95\%)\right], \tag{2.12}$$

and $\chi_1^2(\delta)$ is the δ quantile of the χ^2 distribution with 1 degree of freedom and 0 otherwise. The correction factor 1.318 is a constant adjusting for the bias due to the thresholding.

Practically, the local window around every observation s_t^* is the one that spans the interval $[t - K/2, t + K/2]$. At the borders, when $t < K/2$, the interval is given by $[1, K+1]$ or when $t > T - K/2$, the interval is given by $[T - K, T]$. The choice of K must be such that the local window contains as many observations as possible while still satisfying the condition that, approximately, the returns in the local window that are not affected by outliers come from the same normal distribution. Ideally, the choice of K should thus depend on the persistence of the underlying GARCH model. Through simulation, we tried several values of K for common GARCH models. We found that, since K affects only the weights in (2.12) it is not a critical tuning parameter. Throughout the paper, we set $K = 30$. A topic for further research is to create a data-driven method for optimally selecting the length of the local window, as in e.g. Mercurio and Spokoiny (2004).

[4]The mad of a sequence of observations $x_1, \ldots, x_n$ is defined as $1.486 \cdot \text{median}_i(|x_i - \text{median}_j(x_j)|)$, where 1.486 is a correction factor to guarantee that the mad is a consistent scale estimator at the normal distribution.

2.4 Forecasting with Univariate BIP–GARCH Models

In the GARCH$(1,1)$ model in (2.3), the optimal r–step–ahead forecast of the conditional variance, $h_{t+r|t} \equiv \mathrm{E}_t(h_{t+r})$ is given by:

$$h_{t+r|t} = \hat{\omega} + \hat{\alpha}_1 y^2_{t+r-1|t} + \hat{\beta}_1 h_{t+r-1|t}, \tag{2.13}$$

where $y^2_{t+r-1|t} \equiv \mathrm{E}_t[y^2_{t+r-1}]$, which is $h_{t+r-1|t}$ for $r = 2, 3, \ldots$ and y_t^2 for $r = 1$. Similarly, r–step–ahead forecasts of the conditional variance of the BIP–GARCH$(1,1)$ are obtained as follows

$$h_{t+r|t} = \hat{\omega} + \hat{\alpha}_1 wy^2_{t+r-1|t} + \hat{\beta}_1 h_{t+r-1|t}, \tag{2.14}$$

where $wy^2_{t+r-1|t} \equiv \mathrm{E}_t\left[w_{k_{\delta,1}}\left(y^2_{t+r-1}/h_{t+r-1}\right) y^2_{t+r-1}\right]$, which is $h_{t+r-1|t}$ for $r = 2, 3, \ldots$ and $w_{k_{\delta,1}}\left(y_t^2/h_t\right) y_t^2$ for $r = 1$, because $w_{k_{\delta,1}}$ in (2.9) is chosen such that $\mathrm{E}[w_{k_{\delta,1}}(u)u] = 1$ for u a χ^2_1 random variable (which does not hold for the original specification of MY).

Note that (2.13) and (2.14) are based on expectations of future squared returns under the assumption of a model without price jumps. In the presence of jumps, y_t is not observed and is naturally replaced by s_t in the above formulas. Extremes thus affect the forecasts not only through a potential bias in the parameter estimation, but also because lagged returns are used to forecast future variances. The effect of these outliers on future variances is unbounded under the GARCH model, but limited under the BIP–GARCH model.

3 Extremes and Multivariate Volatility Forecasting

The effect of extremes on univariate volatility forecasting can equally be expected to be present in the forecasting of correlations. In contrast with the large literature on robust estimation of univariate GARCH models, discussed above, little work exists on estimation of multivariate GARCH models in the presence of once–off jumps. We are aware of only one paper, i.e. Boudt and Croux (2010), proposing a method for robust estimation of the BEKK covariance matrix (Engle and Kroner, 1995).

3.1 Baseline Model

Our baseline model is a multivariate version of (2.1)–(2.3), where the vector of demeaned observed returns $S_t = (s_{1,t}, \ldots, s_{N,t})'$ is composed of two non–observable components, a GARCH(1,1)–cDCC process $Y_t = (y_{1,t}, \ldots, y_{N,t})'$ and an N–dimensional additive jump process A_t:

$$\begin{aligned} S_t &= Y_t + A_t && (3.1)\\ Y_t &= H_t^{1/2} Z_t \quad \text{where} \quad Z_t \overset{i.i.d.}{\sim} N(0, I_N). && (3.2) \end{aligned}$$

Let R_t be the conditional correlation matrix with $R_{ij,t}$ its $(i,j)^{\text{th}}$ element and define D_t as the diagonal matrix containing the conditional variances $h_{ii,t}$, i.e.,

$$D_t = \text{diag}\ (h_{11,t}^{1/2} \ldots h_{NN,t}^{1/2}). \tag{3.3}$$

The cDCC conditional covariance matrix H_t is given by:

$$H_t = D_t R_t D_t = \left(R_{ij,t} \sqrt{h_{ii,t} h_{jj,t}} \right). \tag{3.4}$$

The conditional correlation R_t is based on the matrix process Q_t, whose time–variation is driven by the devolatilized returns $\tilde{Y}_t = (\tilde{y}_{1,t}, \ldots, \tilde{y}_{N,t})' = D_t^{-1} Y_t$, i.e.

$$Q_t = (1 - \alpha - \beta)\overline{Q} + \alpha P_{t-1} \tilde{Y}_{t-1} \tilde{Y}_{t-1}' P_{t-1} + \beta Q_{t-1}, \tag{3.5}$$

with $P_t = \text{diag}\ (q_{11,t}^{1/2} \ldots q_{NN,t}^{1/2})$, $\overline{Q}$ the unconditional correlation matrix of $P_t \tilde{Y}_t$ and α and β are nonnegative scalar parameters satisfying $\alpha + \beta < 1$.[5] This matrix is

[5] The extension of the scalar–cDCC model (and its robust version discussed in the next section) to the more general cases where α and β are $N \times N$ matrices (Engle, 2002) or block-diagonal (Billio, Caporin, and Gobbo, 2006) is straightforward but not investigated here for the sake of simplicity.

then linked to the conditional correlation matrix as follows

$$R_t = \text{diag}\ (q_{11,t}^{-1/2} \dots q_{NN,t}^{-1/2})\ Q_t\ \text{diag}\ (q_{11,t}^{-1/2} \dots q_{NN,t}^{-1/2}). \tag{3.6}$$

Under the cDCC model, the estimation of the matrix $\overline{Q}$ and the parameters α and β are intertwined, since $\overline{Q}$ is estimated sequentially as the correlation matrix of $P_t\tilde{Y}_t$, where P_t also depends on α and β. However since P_t only involves the diagonal elements of Q_t, the diagonal elements of which do not depend on $\overline{Q}$ (because $\overline{Q}_{ii} = 1$ for $i = 1, \dots, N$), Aielli (2009) shows that for given values of α and β

$$q_{ii,t} \;\; = \;\; (1 - \alpha - \beta) + \alpha q_{ii,t-1}\tilde{y}_{i,t-1}^2 + \beta q_{ii,t-1},\ \ i = 1, \dots, N. \tag{3.7}$$

The remainder of the estimation procedure of Aielli (2009) is an iteration until convergence of (i) estimation of $\overline{Q}$ as the sample correlation of $P_t\tilde{Y}_t$ and (ii) multivariate Gaussian QML estimation of α and β using the cDCC specification.

3.2 A cDCC Model with Weights on Extremes

In the presence of additive jumps, i.e. A_t in (3.2), the cDCC procedure in Section 3.1 is likely to deliver biased parameter estimates and hence covariance forecasts. To remedy this, we propose three modifications to the original cDCC, i.e.,

$i)$ the replacement of GARCH with the BIP–GARCH model on $S_t = (s_{1,t}^* - \hat{\mu}_1, ..., s_{N,t}^* - \hat{\mu}_N)$ as described in Section 2.2 and Appendix A to compute the devolatilized returns $\tilde{S}_t = D_t^{-1}S_t$;

$ii)$ estimation of $\overline{Q}$ with a robust correlation estimator;

$iii)$ and replacement of the cDCC model with a BIP–cDCC specification.

In the BIP–cDCC model, the effect of $\tilde{S}_{t-1}\tilde{S}_{t-1}'$ on Q_t is bounded. The BIP–cDCC model must only bound the most extreme returns and still yield positive semidefinite covariance matrices. The latter condition is verified if the weights are positive and identical for all elements in $\tilde{S}_{t-1}\tilde{S}'_{t-1}$. This calls for a scalar measure of the extremeness of $\tilde{S}_{t-1}$, and to apply the bounding, the distribution of this statistic should be known.[6]

Under the standard cDCC model without jumps, $\tilde{S}_{t-1}$ is conditionally normally distributed with mean zero and covariance matrix R_{t-1}. Cox (1968) and Healy (1968) proposed to use the squared Mahalanobis Distance (MD) to detect extremes in multivariate normal data, i.e.,

$$d_{t-1} = \tilde{S}'_{t-1} R_{t-1}^{-1} \tilde{S}_{t-1}. \tag{3.8}$$

The MD is conditionally distributed as a chi–square random variable with N degrees of freedom. If any of the components in $\tilde{S}_{t-1}$ is an extreme or $\tilde{S}_{t-1}$ is a correlation outlier, the MD will be inflated. Hence, if d_{t-1} exceeds a high quantile of the $\chi^2(N)$ distribution (denoted $k_{\delta,N}$), it is likely that S_{t-1} is an extreme return and should be downweighted. We can thus use a similar weight function as in the univariate BIP–GARCH model. The BIP–cDCC then takes the form

$$Q_t = (1-\alpha-\beta)\overline{Q} + \alpha w_{k_{\delta,N}}(d_{t-1}) P_{t-1}\tilde{S}_{t-1}\tilde{S}'_{t-1}P_{t-1} + \beta Q_{t-1}, \tag{3.9}$$

where $w_{k_{\delta,N}}(u) = c_{\delta,N}\min\left(1, \frac{k_{\delta,N}}{u}\right)$ is as in (2.9) but the threshold $k_{\delta,N}$ and correction factor $c_{\delta,N}$ are now computed under the $\chi^2(N)$ distribution (see Table 1). The correction factor $c_{\delta,N}$ is such that $\mathrm{E}[w_{k_{\delta,N}}(Z'Z)ZZ'] = I_N$ if $Z \overset{i.i.d.}{\sim} N(0, I_N)$. The choice of δ is based on an efficiency versus robustness trade–off. If all return observations follow the cDCC model, then the BIP–cDCC model induces a bias and larger mean squared errors in the estimated parameters. The higher the values of $k_{\delta,N}$, the closer the BIP–cDCC model is to the cDCC model and hence the smaller the misspecification bias is. But large values of $k_{\delta,N}$ also imply that the effect of extremes on correlations becomes larger. In the remainder of the paper, we take

[6] The use of thresholds to model the conditional correlations of financial return series is also considered in Audrino and Trojani (2011). Their threshold is however based on the average cross-product of the components of $\tilde{S}_{t-1}$. In the presence of time–varying conditional correlations, the distribution of this statistic is unknown. Audrino and Trojani (2011) use a data–driven method to estimate a fixed threshold. Because of the time–variation in the distribution of their statistic, this approach may not be optimal.

$\delta = 0.975$.

Note that in the univariate case, the MD is just the squared devolatilized return. Our multivariate technique for bounding is thus analogous with the approach proposed in Section 2.2.

3.3 Reweighted Unconditional Correlation Estimator

In the cDCC model the intercept of Q_t is an explicit function of $\overline{Q}$, the long run correlation matrix of $\tilde{S}_t$, typically estimated by the sample correlation of $\tilde{S}_t$.[7] The presence of additive jumps is likely to bias this estimate, just like it does for volatilities. It is thus desirable to replace the simple correlation matrix with an estimator that is more robust in the presence of such extremes.

We propose using a robustly reweighted correlation estimator that is proportional to the sample correlation of the observations for which no extreme has been detected using a multivariate test statistic based on local correlation estimates. It is analogous to the reweighted mean and variance estimators of Subsection 2.3.

As for the reweighed variance estimator in (2.12), the reweighted correlation estimator proceeds in two steps. First the Spearman correlation matrix RC_t is computed in a local window around every observation $\tilde{S}_t$. More precisely, write $\tilde{S}_{t:1}, \ldots, \tilde{S}_{t:K+1}$ as the time ordered observations in the window t. Then rank each component series within the local window. Denote the vector series containing these ranks $L_{t:1}, \ldots, L_{t:K+1}$. The raw Spearman correlation matrix C_t for the window around $\tilde{S}_t$ is the sample correlation matrix of $L_{t:1}, \ldots, L_{t:K+1}$. Moran (1948) showed that this correlation matrix needs to be corrected as follows to ensure consistency:

$$SC_t = 2 \sin \left(\frac{1}{6} \pi C_t \right). \tag{3.10}$$

Advantages of the Spearman correlation matrix with respect to other robust correlation estimators include its computational simplicity, high efficiency and outlier

[7] We suppose that N/T is not close to one. If N/T is close to one, the correlation matrix is ill–conditioned and other methods, such as shrinkage estimators, are preferred (see Hafner and Reznikova 2011).

robustness, as shown by Croux and Dehon (2010). The locality is needed such that $\tilde{S}_t' SC_t^{-1} \tilde{S}_t$ is approximately chi-square distributed with N degrees of freedom (see Subsection 3.2).

The second step is then to compute the robust reweighted correlation estimator:

$$RC \;=\; \frac{c_{0.95,N}}{\sum_{t=1}^{T} L_t} \sum_{t=1}^{T} \tilde{S}_t \tilde{S}_t' L_t, \tag{3.11}$$

with weights $L_t = I[\tilde{S}_t' RC_t^{-1} \tilde{S}_t \leq \chi^2_N(0.95)]$. The scalar $c_{0.95,N}$ is as defined in (2.10). The reweighted (RW) correlation estimator of $\bar{Q}$, denoted $\hat{\bar{Q}}_{RW}$ is given by

$$\hat{\bar{Q}}_{RW} = \text{diag}\ (RC_{11}^{-1/2} \ldots RC_{NN}^{-1/2})\ RC\ \text{diag}\ (RC_{11}^{-1/2} \ldots RC_{NN}^{-1/2}). \tag{3.12}$$

This correlation estimate is positive semidefinite and inherits the good robustness properties from the first step correlation estimate used to compute the weights (Lopuhaä, 1999).

3.4 Estimation of the BIP–cDCC Model

Estimation of the cDCC model is straightforward, even for moderately large N, and we aim to keep estimation of the model we propose straightforward as well.

Like in the univariate approach, we use M–estimators to estimate α and β. We define the M–estimators for conditional correlation models as the minimizers of the sum of the average value of a ρ–function, evaluated at the squared Mahalanobis distances and the average value of the log of the determinant of the correlation matrices, i.e.,

$$\hat{\theta}^M = \text{argmin} M_T(\theta, \rho) \equiv \frac{1}{T} \sum_{t=1}^{T} \left[\log\det R_t \ + \sigma\rho\left(\tilde{S}_t' R_t^{-1} \tilde{S}_t\right)\right], \tag{3.13}$$

where σ is a correction factor. If $\rho(z) = z$ and $\sigma = 1$, we obtain the Gaussian QML estimator as a special case which corresponds to the original estimation method advocated by Engle (2002). From the first order condition of the M–estimator, it is clear that the influence of extremes on the M–estimate depends strongly on the derivative of the ρ–function used:

$$\frac{\partial M_T(\theta,\rho)}{\partial\theta_j} = \frac{1}{T}\sum_{t=1}^{T} \operatorname{Tr}\left[I_N - \sigma\rho'(\tilde{S}_t' R_t^{-1}\tilde{S}_t)\tilde{S}_t\tilde{S}_t' R_t^{-1}\right]\frac{\partial R_t}{\partial\theta_j} R_t^{-1} = 0, \tag{3.14}$$

where Tr is the trace operator. Since each return $\tilde{S}_t$ is weighted by the derivative of the ρ–function evaluated at the squared Mahalanobis distance of $\tilde{S}_t$ in terms of R_t, the extreme bias on the estimate will be lower for M-estimators with decreasing ρ–functions. The Gaussian QML is very sensitive to additive jumps because in this case $\rho'(\tilde{S}_t' R_t^{-1}\tilde{S}_t) = 1\ \forall t$, irrespective of the Mahalanobis distance. Conversely, the M-estimator with $\rho_{t_{N,4}}$ as defined in (2.7) is less sensitive to extremes, since its derivative $\rho'_{t_{N,4}}(z) = \frac{N+4}{2+z}$ decreases at the rate $1/z$ to zero. The correction factor $\sigma_{N,4}$ in (2.8) makes this estimator consistent in the absence of extremes. For typical values of N, we tabulate the correction factor $\sigma_{N,4}$ in Table 1.

The M–estimation procedure of the BIP–cDCC model is similar to the QML estimation of the cDCC since $\overline{Q}$, α and β are estimated in an iterative way. It starts with the estimation of $\overline{Q}$ as the reweighted unconditional correlation matrix of $P_1\tilde{S}_1, \ldots, P_T\tilde{S}_T$ where the diagonal elements of P_t are obtained using the BIP version of (3.7), i.e.,

$$q_{ii,t} \quad = \quad (1-\alpha-\beta) + \alpha w(\tilde{s}^2_{i,t-1}) q_{ii,t-1}\tilde{s}^2_{i,t-1} + \beta q_{ii,t-1}. \tag{3.15}$$

The estimator then iterates until convergence through estimation of $\overline{Q}$ as the reweighted correlation of $P_1\tilde{S}_1, \ldots, P_T\tilde{S}_T$ and Student t_4 M-estimation of α and β using the BIP–cDCC specification.

3.5 Forecasting with BIP-cDCC models

We now have all building blocks to construct robust multivariate volatility forecasts. Unfortunately, multistep forecasts of the covariance matrix cannot be made analytically, because the model is not linear in squares and crossproducts of the data. As an approximation, we follow Engle and Sheppard (2001) by constructing the BIP–cDCC r–step–ahead volatility forecasts as

$$H_{t+r|t} = D_{t+r|t} R_{t+r|t} D_{t+r|t}, \tag{3.16}$$

where $D_{t+r|t}$ is the diagonal matrix holding the r-step ahead conditional variance forecasts as described in Section (2.4). The correlation forecast $R_{t+r|t}$ is the standardized version of $Q_{t+r|t}$, where the 1–step-ahead forecast is obtained by projecting (3.9) one step into the future and for $r > 1$

$$Q_{t+r|t} = (1 - \hat{\alpha} - \hat{\beta})\hat{\overline{Q}} + (\hat{\alpha} + \hat{\beta}) Q_{t+r-1|t}, \tag{3.17}$$

since $\mathrm{E}[w_{k_{\delta,N}}(Z'Z)ZZ'] = I_N$ if $Z \overset{i.i.d.}{\sim} N(0, I_N)$.

4 Simulation Study of Estimation Precision

Robust estimation of the model parameters is a key feature of the proposed robust covariance forecasting method. In the webappendix we confirm the good finite sample properties (bias and RMSE) of the estimator with a Monte Carlo study and show that, in contrast with the QML estimator, it is not much influenced by jumps in the data.

As noted by a referee, M–estimation of the BIP–cDCC model may become unfeasible if the dimension is very large, mainly due to the inversion of the matrix R_t in (3.8) and (3.13) and/or because the unconditional correlation estimate used in the correlation targeting is ill-conditioned when the cross section size is close to the sample size. A similar observation was made by Engle, Shephard, and Sheppard (2008) and Hafner and Reznikova (2011) for the QML–estimator of the cDCC model. In those cases, we recommend to use a robust version of the MacGyver method proposed by Engle (2009) to obtain accurate estimates of α and β. When the cross section size is close to the sample size, shrinkage methods should be used to estimate $\overline{Q}$.

The main intuition of the MacGyver method is that, since the dependence structure of the cDCC model is the same for any submatrix of the covariance matrix, accurate estimates of α and β in the N-dimensional cDCC model can be obtained as the median of a large number of pairwise estimates of the cDCC-model.[8] In Table 2 we compare the MacGyver QML and BIP M-estimates for $N = 2$, 10 and 50. The cDCC parameters α and β are 0.1 and 0.8, and the intercept $\bar{Q}$ was chosen to match the unconditional correlation of the daily returns 50 stocks belonging to the S&P 500 index for the period 1991-2011.

Since we are only interested in the estimation precision of the correlation parameters, we follow Engle and Sheppard (2001) and fix the conditional variances to one and only do the second step estimation of the cDCC parameters. Additive jumps are generated as a combination of $0.4\varepsilon T$ equispaced cojumps and a Bernouilli process generating on average $0.6\varepsilon T$ individual jumps. We simulated either 1% or 5% of jumps of size $d = 3$ or 4 (recall that the conditional variances are set to 1) and their sign is set equal to the sign of Y_t. For each parameter, we compute the estimation bias and RMSE over $10,000$ replications. For the BIP–cDCC, we consider a threshold given by $\delta = 0.975$.[9]

Consider first the panel with the results for the no jumps case ($\varepsilon = 0\%$) and $N = 2$. We see that the RMSE of the robust estimator is for all settings only slightly higher than the RMSE of the maximum likelihood estimator. In the next panels it is shown that this small loss of efficiency in the absence of jumps is compensated with a smaller bias and lower RMSE in the presence of jumps. Additive jumps lead to a downward bias in the QML estimates of the unconditional correlation $\overline{Q}$ and the dependence parameters α and β. The larger the jumps and the higher the percentage of jumps, the larger the bias is. Jumps have almost no impact on the bias and RMSE of the BIP M–estimator.

For $N > 2$, the estimates of α and β are the median of pairwise estimates of the cDCC-model. We see that, when N increases, the RMSE of the QML and BIP

[8]When the total number of possible combinations of 2 assets exceeds 200, we take the median of 200 randomly chosen subsets.

[9]We repeated the simulation for $\delta = 0.95$, but the resulting bias and RMSE were similar as for $\delta = 0.975$. Results reported in this paper are based on programs written by the authors using Ox version 6.0 (Doornik, 2009) and G@RCH version 6.0 (Laurent, 2009).

Table 2: Bias and RMSE of the MacGyver Gaussian QML and robust estimator for the dependence parameters α and β of the N-dimensional cDCC model in presence of ε jumps of size d times the conditional standard deviation, with $\delta = 0.975$ and $T = 2000$. For $\overline{Q}$ the average bias and RMSE of the lower diagonal elements are reported.

			$N=2$			$N=10$			$N=50$		
			$\overline{Q}$	α	β	$\overline{Q}$	α	β	$\overline{Q}$	α	β
$\varepsilon = 0\%$	QML	Bias	-0.003	0.001	-0.008	0.001	-0.001	-0.004	0.000	0.000	-0.004
		RMSE	0.039	0.018	0.043	0.042	0.005	0.012	0.043	0.003	0.007
	Robust	Bias	-0.010	0.010	-0.009	-0.013	0.001	-0.005	-0.010	0.001	0.004
		RMSE	0.038	0.021	0.051	0.042	0.006	0.014	0.045	0.004	0.009
$\varepsilon = 1\%$	QML	Bias	-0.019	-0.011	-0.014	-0.001	-0.013	-0.001	-0.008	-0.013	0.000
$d = 3$		RMSE	0.045	0.029	0.084	0.045	0.016	0.018	0.046	0.015	0.010
	Robust	Bias	-0.012	-0.004	-0.008	-0.013	-0.005	-0.004	-0.014	-0.005	-0.004
		RMSE	0.038	0.021	0.051	0.042	0.008	0.013	0.048	0.006	0.008
$\varepsilon = 1\%$	QML	Bias	-0.031	-0.019	-0.030	-0.017	-0.019	-0.001	-0.014	-0.020	0.004
$d = 4$		RMSE	0.054	0.039	0.130	0.051	0.022	0.026	0.050	0.022	0.013
	Robust	Bias	-0.012	-0.004	0.010	-0.013	-0.005	-0.003	-0.017	-0.005	-0.004
		RMSE	0.039	0.020	0.051	0.042	0.008	0.013	0.052	0.006	0.008
$\varepsilon = 5\%$	QML	Bias	-0.058	-0.033	-0.029	-0.035	-0.037	0.017	-0.028	-0.038	0.020
$d = 3$		RMSE	0.072	0.046	0.150	0.058	0.038	0.031	0.055	0.038	0.024
	Robust	Bias	-0.017	-0.021	-0.005	-0.016	-0.023	0.003	-0.026	-0.023	0.002
		RMSE	0.042	0.028	0.056	0.044	0.024	0.015	0.058	0.023	0.008
$\varepsilon = 5\%$	QML	Bias	-0.084	-0.042	-0.054	-0.052	-0.049	0.014	-0.041	-0.049	0.024
$d = 4$		RMSE	0.097	0.057	0.190	0.074	0.050	0.043	0.066	0.050	0.028
	Robust	Bias	-0.017	-0.022	-0.006	-0.015	-0.024	0.004	-0.036	-0.023	0.003
		RMSE	0.042	0.028	0.055	0.044	0.024	0.015	0.067	0.024	0.009

M-estimators of α and β decreases substantially. The number of pairs is 45 and 200, for $N = 10$ and $N = 50$, respectively. This explains why the decrease in RMSE is relatively more important for $N = 10$ vs $N = 2$, than $N = 50$ vs $N = 10$. Interestingly, since jumps have a different impact on each pairwise estimate and because of the aggregation using the median, the bias induced by the jumps also reduces when the cross-section becomes larger. In all cases of jump contamination, the RMSE of the robust estimator of the dependence parameters is substantially smaller than the RMSE of the QML estimator.

For $N > 2$, $\overline{Q}$ is estimated as the (reweighted) unconditional correlation matrix of $P_1\tilde{S}_1, \ldots, P_T\tilde{S}_T$ with $P_1, \ldots, P_T$ constructed using the MacGyver estimates of α and β (see Subsections 3.2-3.4). In Table 2 we report the average bias and RMSE computed on the lower diagonal elements of the estimated unconditional correlation matrix. Since an increase in N reduces the RMSE of the QML estimate of α and β, we also find a reduction in the RMSE of the classical estimator of $\overline{Q}$ in the presence of jumps. For $N = 50$, the RMSE of the robust and classical estimator are similar, due to the fact that in this simulation study, 60% of the jumps are component-wise leading to a propagation of outliers when the dimension is high. This is consistent with Alqallaf and Zamar (2009) showing that the reweighted correlation estimators have poor breakdown properties in case of independent contamination. In such cases, we recommend to use cellwise weights, as proposed by Van Aelst and Willems (2011).

The results obtained above are for a specific set of dependence parameters, namely $\alpha = 0.1$ and $\beta = 0.8$. In Figures 1 and 2 we report the bias and RMSE of the QML and BIP M-estimators of the bivariate cDCC model when α ranges between 0.03 and 1 and $\alpha + \beta$ is fixed at 0.95. The full line shows the bias and RMSE in case of no jumps. The dotted and dashed lines correspond to the case of 5% of additive jumps of size $d = 3$ and $d = 4$, respectively. In the left panel, we see that jumps induce a large increase in the bias and RMSE of the QML estimator and this increase tends to be larger when the jump size increases. The bias and RMSE of the QML estimate of α (resp. β) increases for larger values of α (resp. β). Comparing the left and right panels of Figures 1 and 2, we see that, for all parameter combinations considered, the robust BIP M-estimator has a smaller bias and RMSE than the Gaussian QML-estimator in the presence of jumps. In the webappendix we repeated the same analysis for $\alpha + \beta = 0.98$ and obtained similar results.

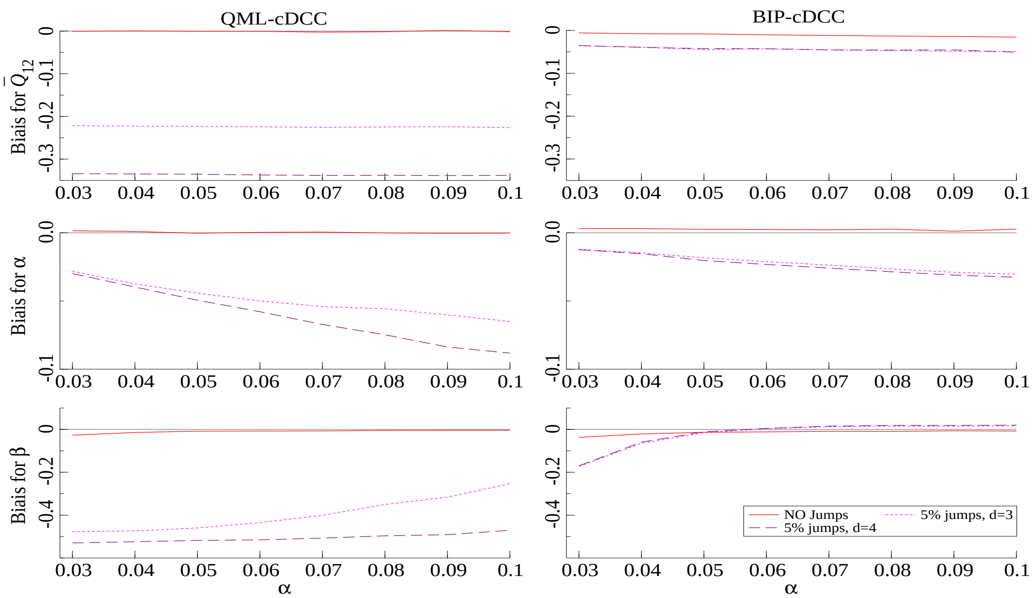

Figure 1: BIAS of the Gaussian QML and robust BIP M-estimator for unconditional correlation $\overline{Q}$ and the dependence parameters α and β of the 2-dimensional cDCC model in presence of ε jumps of size d times the conditional standard deviation, with $\delta = 0.975$ and $T = 2000$ and $\alpha + \beta = 0.95$.

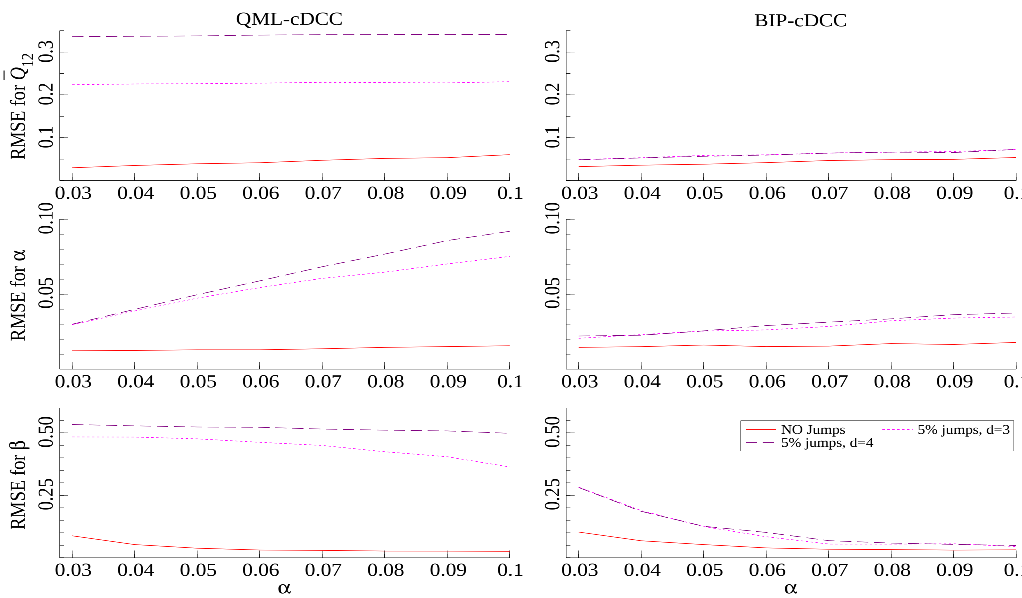

Figure 2: RMSE of the Gaussian QML and robust BIP M-estimator for unconditional correlation $\overline{Q}$ and the dependence parameters α and β of the 2-dimensional cDCC model in presence of ε jumps of size d times the conditional standard deviation, with $\delta = 0.975$ and $T = 2000$ and $\alpha + \beta = 0.95$.

5 Forecast Evaluation using the Model Confidence Set

Our first application is on forecasting the r–step ahead daily conditional covariance matrix of the EUR/USD and Yen/USD exchange rates over the period 2004–2009. The model confidence set (MCS) approach of Hansen, Lunde, and Nason (2011) is used to compare the forecasts. Given a universe of model based forecasts, the MCS allows us to identify the subset of models that are equivalent in terms of forecasting ability, but outperform all the other competing models. The accuracy of the forecasts is evaluated by comparing the forecast with a high–frequency based ex post measure of the daily covariance matrix and a robust loss function.

The next subsections present the set of competing models, the proxy of the true but unobserved covariance, the data, the loss function, and finally the results of our application. Following Hansen, Lunde, and Nason (2011), we set the significance level for the MCS to $\alpha = 0.25$ and 10,000 bootstrap samples were used to obtain the distribution under the null of equal forecasting performance.[10]

Set of competing models: We consider eight MGARCH models with a constant conditional mean. The first three models belong to the class of BEKK models proposed by Engle and Kroner (1995). We consider the diagonal BEKK$(1,1)$, the scalar-BEKK$(1,1)$ and the multivariate exponentially weighted moving average (EWMA) model.[11] The first two models are estimated by Gaussian QML while the EWMA model does not require any parameter estimation (apart from the mean, set here to the empirical mean). We also consider the constant conditional correlation (CCC) model of Bollerslev (1990) and the (consistent) DCC model, with GARCH$(1,1)$ specifications for the conditional variances. The CCC and cDCC models are estimated by Gaussian QML using the three step–approach described in Section 3.1 (or two–step for the CCC). The last two models are the BIP versions of the CCC

[10]Implementation of this test has been done using the Ox software package MULCOM 2.0 of Hansen and Lunde (2011).

[11]The EWMA model has been popularised by Riskmetrics (1996) and is widely used by practitioners. See e.g. Bauwens, Laurent, and Rombouts (2006) for the exact specification of these models.

and cDCC models. The first step is common to all three models and consists of the estimation of N BIP–GARCH$(1,1)$ models with variance targeting. The second step corresponds to the estimation of the correlation matrix as described in Section 3.3. We choose $\delta = 0.975$ for the BIP weight functions.

Proxy: We judge the performance of the competing models through the use of a statistical loss function. The evaluation of the forecasting performance of volatility models is challenging since the variable of interest (i.e., the covariance) is unobservable and therefore the evaluation of the loss function has to rely on a proxy. We consider three proxies based on the theory of quadratic variation of Brownian semimartingale with jumps processes, i.e., the realized covariance (RCov) of Andersen, Bollerslev, Diebold, and Labys (2003), the realized bipower covariation (RBPCov) of Barndorff-Nielsen and Shephard (2004) and the realized outlyingness weighted covariation (ROWCov) of Boudt, Croux, and Laurent (2011a). The motivation for the choice of these proxies is that RCov estimates the total quadratic variation, while RBPCov and ROWCov only estimate the continuous component of the quadratic variation.

Data: Our data consists of daily and 30–minute log–returns computed from the indicative quotes provided by Olsen & Associates, on the Euro (Deutsche Mark before 1999) and the Yen exchange rates expressed in US dollars (EUR and YEN).[12] The sample period goes from January 3, 1995 to December 31, 2009. We removed days with too many missing values and/or constant prices, as well as weekends and holidays when trading is infrequent. The cleaned dataset spans 3819 trading days. One trading day extends from 21.00 GMT on day $t-1$ to 21.00 GMT on date t. On equity data, Laurent, Rombouts, and Violante (2011) find that the relative performance of MGARCH models depends strongly on the state of the market. We therefore distinguish between the calm market period of 2004–2006 and the turbulent period of the credit crisis in 2007–2009. The standard deviation of the daily EUR and YPY return

[12] The 30–minute returns are needed to compute the proxies discussed above. Muthuswamy, Sarkar, Low, and Terry (2001), among others, show that, because of non–synchronicity of trading, correlations are biased toward zero if returns on foreign exchange prices are calculated at ultra high frequencies such as five minutes. For our dataset, the correlation in the 5, 10, 15, 30 and daily EUR and YEN returns is 26%, 30%, 31%, 32% and 34%, respectively.

is 59% and 56% in the calm period, but increases to 73% and 87% in the turbulent period. From the daily returns, rolling estimation samples of 2303 observations are used to produce the out–of–sample r–step ahead daily covariance forecasts, with $r = 1, \ldots, 10$. In the webappendix we report the time series and scatter plots of the forecasted cDCC and BIP–cDCC variances and covariances. Overall, we find that the difference in the cDCC and BIP–cDCC variances and covariances tends to be small, except for the period October-November 2008, where the cDCC variance forecast for the Yen/USD is almost double the BIP–cDCC forecast. Carnero, Peña, and Ruiz (2012) document that robust variances of stock returns tend to be smaller than standard GARCH(1,1) variance forecasts. We find a similar result for the exchange rate data: the 1 (10) days ahead GARCH variance forecast is on 63% (58%) and 94% (98%) of the days higher that the BIP-GARCH variance forecast for the EUR/USD and Yen/USD, respectively.

Loss function: In the presence of outliers (or jumps), Preminger and Franck (2007) recommend using forecast performance evaluation criteria that are less sensitive to extreme observations. For this reason we rely on the Entrywise 1 – (matrix) norm, defined as follows:

$$L_{m,t} = \sum_{1 \leq i,j \leq N} \left| \sigma_{i,j,t} - h_{m,i,j,t} \right|, \tag{5.1}$$

where $L_{m,t}$ is the Entrywise 1 - (matrix) norm of model m (for $m = 1, \ldots, 9$) and day t, $\sigma_{i,j,t}$ and $h_{m,i,j,t}$, indexed by $i, j = 1, ..., N$, refer respectively to the elements of the covariance matrix proxy for day t (i.e., Σ_t) and covariance forecast of model m (i.e., $H_{m,t}$).[13]

Results: Table 5 indicates which models belong to the set of superior forecasting models according to the MCS test. In both the calm and turbulent market regimes,

[13] Hansen and Lunde (2006), Laurent, Rombouts, and Violante (2009) and Patton (2011) show that the substitution of the underlying volatility by a proxy may induce a distortion in the ranking i.e., the evaluation based on the proxy might differ from the ranking that would be obtained if the true target was used. However, such distortion can be avoided if the loss function has a particular functional form or when the proxy is accurate enough. Monte Carlo simulation results reported in Laurent, Rombouts, and Violante (2009) suggest that when the proxy of the daily covariance is computed from high frequency returns (e.g., 30–minute returns like in our case), all loss functions deliver the expected ranking (i.e., the one based on the true covariance) which justifies our choice.

Table 3: Models that have superior forecasting performance for the RCov, RPBCov and ROWCov of EUR/USD and Yen/USD returns in 2004–2006 and 2007–2009, as indicated by the model confidence set approach.

	Calm period: 2004-2006										Turbulent period: 2007-2009									
r	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
											RCov									
BIP-cDCC	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
BIP-CCC		+	+	+	+	+	+	+												
cDCC											+	+								
DCC											+	+								
CCC																				
RM																				
Scalar-BEKK											+	+								
Diag-BEKK											+	+								
											RBPCov									
BIP-cDCC	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
BIP-CCC		+	+	+	+	+	+	+												
cDCC											+									
DCC																				
CCC																				
RM											+									
Scalar-BEKK											+									
Diag-BEKK											+									
											ROWCov									
BIP-cDCC	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
BIP-CCC	+	+	+																	
cDCC																				
DCC																				
CCC																				
RM	+																			
Scalar-BEKK																				
Diag-BEKK																				

the BIP–cDCC model is selected as having superior forecasting performance for all forecast horizons. Moreover, it always has the smallest value for the loss function. Interestingly, in the calm period a BIP–model without dynamics in the conditional correlations (i.e. the BIP–CCC) often belongs to the set of superior forecasting models. In the turbulent market regime, the RCov and RBPCov are clearly less informative about the best 1-day ahead forecast, since not only the BIP-cDCC, but also the cDCC, DCC, scalar-BEKK and diag-BEKK belong to the model confidence set. But for the ROWCov and for longer horizons (RCov: horizons of at least three days, RBPCov: horizons of at least two days), the BIP-cDCC is the best model according the the MCS procedure.

6 Economic Gains in Portfolio Allocation

Our key empirical application follows from applying the BIP model to the problem of optimal portfolio allocation. We study an investor who adopts a volatility timing strategy where the changes in the ex ante optimal portfolio weights are solely determined by the forecasts of the conditional covariance matrix. The optimal portfolio allocation is supposed to be the minimum variance portfolio; the minimum variance portfolio being the only portfolio on the efficient frontier that is independent of the mean forecast. To make the portfolio allocation more realistic, portfolio weights are required to be nonnegative and less than 10%.[14]

The portfolios are fully invested in equity belonging to the same sector. The initial investment universe consists of all S&P 500 stocks on July 31, 2010. Stocks with missing data at the start of the estimation period and sectors with less than ten stocks in the investment universe are removed.[15] We study the portfolio performance gains obtained by allocating the sector portfolios based on the BIP–cDCC covariance forecasts instead of the forecasts from the baseline cDCC model, over the period January 2004 – July 2009.[16]

Because of the high dimensionality of the covariance estimates, the dependence parameters in the correlation process are estimated using the MacGyver method. Summary statistics of the rolling estimates of α and β are reported in Table 4. We see that in all cases, the range of estimates of α and β is much more narrow for the BIP–cDCC method than for the cDCC method. Furthermore, the average estimate of β is for all sectors higher using the BIP–cDCC method than when using the cDCC estimator. Finally note the higher stability of the BIP–cDCC covariance and variance forecasts compared to the cDCC forecasts, as indicated by the first order autocorrelation of the forecasted series.

We split up the evaluation sample in the calm period of 2004–2006 and the turbulent

[14] In the webappendix we verify that relaxing such bound constraints on the portfolio weights has little impact on the conclusions regarding the relative profitability of using the BIP–cDCC vs cDCC covariance forecasts.

[15] See the webappendix for the list with tickers.

[16] We repeated the analysis for the BIP–CCC vs CCC covariance forecasts and obtained similar conclusions.

Table 4: Summary statistics on rolling estimates of the correlation dependence parameters, α and β, obtained using the cDCC and BIP-cDCC model. The last two columns report the average first order autocorrelation of the covariance ($h_{i,j}$) and variance (h_{ii}) forecasts.

		$\hat{\alpha}$			$\hat{\beta}$			Mean AC(1)	
		min	mean	max	min	mean	max	$\hat{h}_{i,j}$	$\hat{h}_{i,i}$
Cons.Discret.	cDCC	0.005	0.009	0.018	0.933	0.980	0.994	0.933	0.877
	BIP-cDCC	0.006	0.008	0.028	0.978	0.990	0.997	0.978	0.990
Cons.Staples	cDCC	0.006	0.009	0.012	0.852	0.931	0.989	0.929	0.844
	BIP-cDCC	0.006	0.008	0.010	0.989	0.991	0.993	0.983	0.989
Energy	cDCC	0.007	0.013	0.025	0.959	0.983	0.990	0.956	0.944
	BIP-cDCC	0.010	0.015	0.020	0.979	0.984	0.990	0.986	0.992
Financials	cDCC	0.011	0.017	0.029	0.900	0.948	0.980	0.917	0.912
	BIP-cDCC	0.012	0.015	0.021	0.974	0.981	0.985	0.976	0.987
Healthcare	cDCC	0.006	0.010	0.014	0.899	0.941	0.993	0.901	0.770
	BIP-cDCC	0.006	0.007	0.009	0.990	0.992	0.994	0.976	0.985
Industrials	cDCC	0.005	0.009	0.027	0.933	0.985	0.994	0.933	0.877
	BIP-cDCC	0.009	0.012	0.020	0.978	0.987	0.990	0.978	0.990
IT	cDCC	0.005	0.008	0.017	0.952	0.983	0.995	0.953	0.885
	BIP-cDCC	0.010	0.012	0.016	0.983	0.987	0.989	0.981	0.993
Materials	cDCC	0.007	0.013	0.025	0.959	0.983	0.990	0.955	0.925
	BIP-cDCC	0.009	0.011	0.014	0.986	0.989	0.991	0.987	0.990

For all sectors, the cDCC and BIP-cDCC estimates of α and β are significantly different at a 99% confidence level, using the t-test and Wilcoxon tests.

years 2007–2009. Forecasts of the conditional covariance matrix at time t are based on the data from January 1994 up to $t-1$. We measure the portfolio performance in two ways. First we look at the mean and standard deviation of the out–of–sample daily gross returns. The net return for an investor corresponds to these gross return performance measures, from which the transaction costs are deducted. As an indicator for the transaction costs, we report in Table 5 also the average portfolio turnover. The daily portfolio turnover is defined as the percentage of wealth traded that day: $\kappa_t = |w_t - w_{t+}|'\iota$, where w_t is the vector of weights at the rebalancing period t and w_{t+} is the vector of weights before rebalancing at t and ι is a vector of ones. Finally, we report in Table 5 the difference in Sharpe ratios of the minimum variance portfolios constructed using the proposed robust method versus the classical method in the presence of a proportional transaction cost f. Then, the daily evolution of the wealth invested is given by

$$W_{t+1} = W_t(1 - f \cdot \kappa_t)(1 + p_t), \tag{6.1}$$

with p_t the portfolio return on trading day t and $W_0 = 1000$ USD. Following Balduzzi and Lynch (1999), the transaction cost per dollar of portfolio value f is set to either 0, 1e-4 or 1e-3 USD.

Similarly as in Engle and Colacito (2005), we test for significant differences between the daily portfolio returns, squared returns and turnover using a Diebold and Mariano (1995) type test. It regresses the daily differences between the performance measures of two portfolio methods on a constant and tests whether the estimated constant is significantly different from zero using a Newey-West standard error. The significance of the difference in Sharpe ratio is evaluated using the test of Jobson and Korkie (1981) and Memmel (2003), which uses also Newey-West standard errors. Similar results were obtained when testing using the bootstrap procedure of Ledoit and Wolf (2008).

Note first in Table 5 that the use of the BIP method has no significant impact on the average gross portfolio return, but for most sectors it reduces significantly the portfolio standard deviation. For 11 out of the 16 portfolios considered, the BIP method has a positive impact on the Sharpe ratio, but the impact is not statistically significant. This does not mean that there is no profit for the investor in using the BIP procedure, since the BIP conditional covariance matrices are more stable, resulting in lower portfolio churn and thus lower transaction costs. Indeed, we see that, except for the portfolio invested in the IT sector in 2004-2006, the portfolio turnover is always significantly larger for the classical portfolio than for the robust portfolio. Therefore, even if both procedures have a similar out–of–sample mean return and standard deviation, net of transaction costs, the BIP procedure potentially yields more profits overall. This is illustrated in the last two columns of Table 5. When the Sharpe ratio is computed on the returns net of a 0.1% proportion transaction costs, the difference in Sharpe ratio between the robust and classical allocation method is (except for the IT sector) always positive, and for some sectors

Table 5: Summary statistics on out–of–sample performance of minimum variance portfolios based on the BIP–cDCC vs cDCC model: gross returns (annualized mean, standard deviation), portfolio turnover and difference in annualized Sharpe ratio when the proportional trading cost is κ.

	cDCC			BIP–cDCC			Δ SR BIP–cDCC vs cDCC		
	mean	SD	Turn	mean	SD	Turn	$\kappa = 0$	$1e-4$	$1e-3$
2004-2006									
Cons.Discret.	0.101	0.098***	0.151***	0.114	0.102***	0.244***	0.047	0.063	0.237
Cons.Staples	0.090	0.089*	0.289***	0.089	0.086*	0.156***	0.032	0.063	0.379*
Energy	0.325	0.233***	0.334***	0.283	0.223***	0.173***	-0.126	-0.111	0.032
Financials	0.209	0.118***	0.474***	0.135	0.104***	0.321***	-0.474	-0.459	-0.237
Healthcare	0.076	0.114	0.330***	0.090	0.111	0.205***	0.142	0.174	0.395
Industrials	0.129	0.106***	0.346***	0.151	0.102***	0.195***	0.269	0.316	0.601***
IT	0.111	0.132	0.165***	0.123	0.132	0.180***	0.095	0.079	0.063
Materials	0.165	0.146***	0.270***	0.131	0.134***	0.169***	-0.142	-0.126	0.000
2007-2009									
Cons.Discret.	-0.179	0.262***	0.186***	-0.149	0.251***	0.215***	0.032	0.063	0.379*
Cons.Staples	-0.018	0.188***	0.314***	0.005	0.176***	0.171***	0.126	0.142	0.300**
Energy	-0.036	0.439***	0.305***	0.001	0.429**	0.187***	0.079	0.095	0.142*
Financials	-0.124	0.346*	0.359***	-0.108	0.339*	0.210***	0.047	0.047	0.142
Healthcare	0.009	0.214***	0.350***	0.016	0.201***	0.193***	0.032	0.047	0.206
Industrials	-0.068	0.264***	0.373***	-0.059	0.251***	0.220***	0.016	0.032	0.158
IT	-0.028	0.282***	0.264***	-0.048	0.273***	0.257***	-0.079	-0.079	-0.079
Materials	-0.045	0.311***	0.243***	-0.017	0.303***	0.150***	0.095	0.095	0.158*

***, ** and * indicate significant differences between BIP–cDCC and cDDC at the 1%, 5% and 10% level respectively.

even significantly positive.

7 Conclusion

We propose the BIP–cDCC model for multivariate volatility forecasting in the presence of once–off events causing large changes in prices whilst not affecting volatility dynamics. Under this model, extremes have a bounded impact both on the parameter estimates and the volatility forecasts. In an application to forecasting the covariance matrix of the daily returns on the EUR/USD and Yen/USD exchange rates, the BIP–cDCC model always belongs to set of superior forecasting models. Furthermore, for most forecast horizons, it is identified as the best model by the model confidence set approach of Hansen, Lunde, and Nason (2011). We also show that for minimum variance allocation of sector portfolios on US stocks, the BIP covariance forecasts reduce significantly the portfolio turnover, while not deteriorating the out–of–sample portfolio return performance.

Throughout the paper we focused on using the BIP–GARCH(1,1) model for univariate volatility forecasting. A natural extension is to consider volatility models with leverage effects, such as the asymmetric power ARCH model of Ding, Granger, and Engle (1993) or the GJR model of Glosten, Jagannathan, and Runkle (1993), provided the models are adapted such that lagged returns have a bounded impact on future volatility.

A further limitation of the paper is that we only considered one type of outliers, namely those corresponding to price jumps caused by large once–of events. As noted e.g. in Tsay, Peña, and Pankratz (2000), there exist other definitions of outliers in multivariate time series.

We leave further work along these lines for future research.

References

Aielli, G. (2009): "Dynamic Conditional Correlations: on Properties and Estimation," *Unpublished paper: Department of Statistics, University of Florence.*

Alqallaf, F., V. A. S. Y. V., and R. Zamar (2009): "Propagation of Outliers in Multivariate Data," *Annals of Statistics*, 37, 311–331.

Andersen, T., T. Bollerslev, F. Diebold, and P. Labys (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

Andersen, T. G., T. Bollerslev, and F. Diebold (2007): "Roughing it up: Including Jump Components in the Measurement, Modelling and Forecasting of Return Volatility," *Review of Economics and Statistics*, 89, 701–720.

Audrino, F., and F. Trojani (2011): "A General Multivariate Threshold

GARCH Model with Dynamic Conditional Correlations," *Journal of Business and Economics Statistics*, 29, 138–149.

Balduzzi, P., and A. Lynch (1999): "Transaction costs and predictability: some utility cost calculations," *Journal of Financial Economics*, 52, 47–78.

Barndorff-Nielsen, O. E., and N. Shephard (2004): "Measuring the Impact of Jumps in Multivariate Price Processes Using Bipower Covariation," *Discussion paper, Nuffield College, Oxford University.*

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

Bauwens, L., and G. Storti (2009): "A Component GARCH Model with Time Varying Weights," *Studies in Nonlinear Dynamics & Econometrics*, 13/2, article 1.

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

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

Boudt, K., and C. Croux (2010): "Robust M-Estimation of Multivariate GARCH Models," *Computational Statistics and Data Analysis*, 54, 2459–2469.

Boudt, K., C. Croux, and S. Laurent (2011a): "Outlyingness Weighted Covariation," *Journal of Financial Econometrics*, 9, 657–684.

——— (2011b): "Robust Estimation of the Periodicity in Intraday Volatility and Intraday Jump Detection," *Journal of Empirical Finance*, 18, 353–367.

Boudt, K., J. Daníelsson, and S. Laurent (2012): "Appendix to Robust Forecasting of Dynamic Conditional Correlation GARCH Models," *Available at www.econ.kuleuven.be/kris.boudt/public.*

CARNERO, M., D. PEÑA, AND E. RUIZ (2006): "Effects of outliers on the identification and estimation of GARCH models," *Journal of time series analysis*, 28, 471–497.

——— (2012): "Estimating GARCH volatility in the presence of outliers," *Economics Letters*, 114, 86–90.

CHAN, W. H., AND J. M. MAHEU (2002): "Conditional Jump Dynamics in Stock Market Returns," *Journal of Business and Economic Statistics*, 20, 377–389.

CORSI, F., D. PIRINO, AND R. RENÓ (2008): "Threshold Bipower Variation and the Impact of Jumps on Volatility Forecasting," *Journal of Econometrics*, 159, 276–288.

COX, D. (1968): "Notes on Some Aspects of Regression Analysis," *Journal of Royal Statistical Society A*, 131, 265–279.

CROUX, C., AND C. DEHON (2010): "Influence Functions of the Spearman and Kendall Correlation Measures," *Statistical Methods and Applications*, 19, 497–515.

DIEBOLD, F., AND R. MARIANO (1995): "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253–263.

DING, Z., C. W. J. GRANGER, AND R. F. ENGLE (1993): "A Long Memory Property of Stock Market Returns and a New Model," *Journal of Empirical Finance*, 1, 83–106.

DOORNIK, J. (2009): *Object-Oriented Matrix Programming Using Ox*. Timberlake Consultants Press.

ENGLE, R. (2002): "Dynamic Conditional Correlation - a Simple Class of Multivariate GARCH Models," *Journal of Business and Economic Statistics*, 20, 339–350.

——— (2009): *The Methodology and Practice of Econometrics: Festschrift for David Hendry*chap. High dimension dynamic correlations, pp. 122–148. Oxford University Press.

Engle, R., and F. Kroner (1995): "Multivariate Simultaneous Generalized ARCH," *Econometric Theory*, 11, 122–150.

Engle, R., and J. Mezrich (1996): "GARCH for Groups," *RISK*, 9, 36–40.

Engle, R., N. Shephard, and K. Sheppard (2008): "Fitting vast dimensional time-varying covariance models," *Mimeo.*

Engle, R., and K. Sheppard (2001): "Theorical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH," Mimeo, UCSD.

Engle, R. F., and R. Colacito (2005): "Testing and Valuing Dynamic Correlations for Asset Allocation," *Journal of Business and Economic Statistics*, 24, 238–253.

Francq, C., L. Horvath, and J.-M. Zakoian (2011): "Merits and Drawbacks of Variance Targeting in GARCH Models," *Journal of Financial Econometrics*, 9, 619–656.

Franses, P., and H. Ghijsels (1999): "Additive Outliers, GARCH and Forecasting Volatility," *International Journal of Forecasting*, 15, 1–9.

Glosten, L., R. Jagannathan, and D. Runkle (1993): "On the Relation Between Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 48, 1779–1801.

Hafner, C., and O. Reznikova (2011): "On the Estimation of Dynamic Conditional Correlation Models," *Forthcoming in Computational Statistics & Data Analysis.*

Hansen, P., and A. Lunde (2006): "Consistent Ranking of Volatility Models," *Journal of Econometrics*, 131, 97–121.

——— (2011): "MULCOM 2.00, Econometric Toolkit for Multiple Comparisons," http://www.hha.dk/∼alunde/MULCOM/MULCOM.HTM.

Hansen, P., A. Lunde, and J. Nason (2011): "Model Confidence Sets," *Econometrica*, 79, 453–497.

HARVEY, A., AND T. CHAKRAVARTY (2008): "Beta-t-(E)GARCH," Working Paper series, University of Cambridge.

HEALY, M. (1968): "Multivariate Normal Plotting," *Applied Statistics*, 17, 157–16.

JOBSON, J., AND B. KORKIE (1981): "Performance Hypothesis Testing with the Sharpe and Treynor Measures," *Journal of Finance*, 36, 889–908.

LAURENT, S. (2009): *G@RCH 6. Estimating and Forecasting GARCH Models*. Timberlake Consultants Ltd.

LAURENT, S., J. ROMBOUTS, AND F. VIOLANTE (2009): "On Loss Functions and Ranking Forecasting Performances of Multivariate Volatility Models," Cirano discussion paper 2009-45.

——— (2011): "On the Forecasting Accuracy of Multivariate GARCH Models," *Forthcoming in Journal of Applied Econometrics*.

LEDOIT, O., AND M. WOLF (2008): "Robust performance hypothesis testing with the Sharpe ratio," *Journal of Empirical Finance*, 15, 850–859.

LOPUHAÄ, H. (1999): "Asymptotics of reweighted estimators of multivariate location and scatter," *Annals of Statistics*, 27, 1638–1665.

MEMMEL, C. (2003): "Performance hypothesis testing with the Sharpe Ratio," *Finance Letters*, 1, 21–23.

MERCURIO, D., AND V. SPOKOINY (2004): "Statistical Inference for Time-inhomogenous Volatility Models," *Annals of Statistics*, 32, 577–602.

MORAN, P. (1948): "Rank Correlation and Permutation Distributions," *Proceedings of the Cambridge Philosophical Society*, 44, 142–144.

MULER, N., AND V. YOHAI (2008): "Robust Estimates for GARCH Models," *Journal of Statistical Planning and Inference*, 138, 2918–2940.

Muthuswamy, J., S. Sarkar, A. Low, and E. Terry (2001): "Time Variation in the Correlation Structure of Exchange Rates: High Frequency Analyses," *Journal of Futures Markets*, 21, 127–144.

Patton, A. (2011): "Volatility Forecast Comparison Using Imperfect Volatility Proxies," *Journal of Econometrics*, 160, 246–256.

Preminger, A., and R. Franck (2007): "Forecasting Exchange Rates: A Robust Regression Approach," *International Journal of Forecasting*, 23, 7184.

Riskmetrics (1996): *Riskmetrics Technical Document, 4th ed.* J.P. Morgan, New York.

Tsay, R., D. Peña, and A. Pankratz (2000): "Outliers in multivariate time series," *Biometrika*, 87, 789–804.

Van Aelst, S., V. E., and G. Willems (2011): "Stahel-Donoho Estimators with Cellwise Weights," *Journal of Statistical Computation and Simulation*, 81, 1–27.

Vlaar, P., and F. Palm (1993): "The Message in Weekly Exchange Rates in the European Monetary System: Mean Reversion, Conditional Heteroskedasticity and Jumps," *Journal of Business and Economic Statistics*, 11, 351–360.