Robust Forecasting of Dynamic Conditional Correlation GARCH Models

Kris Boudt
Lessius and K.U.Leuven.
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

February 5, 2011

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. Compared to the DCC model, our method produces more precise out–of–sample covariance forecasts and, when used in portfolio allocation, it leads to portfolios with similar return characteristics but lower turnover and hence higher profits.

*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, and this is the main contribution of the paper. We are only aware of one paper doing something similar, i.e. Boudt and Croux (2010) proposing a robust estimation method for the BEKK model of Engle and Kroner (1995). Their robust model is

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

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 dynamic conditional correlation model (DCC) of Engle (2002) as our baseline model. We use the Aielli (2009) version termed cDCC, to obtain a consistent estimation of all model parameters.

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 by looking at the impact on volatility and correlations of the 50% one day drop in the stock price of Apple in 2000 due to a bad earnings announcement. Because this large event was once–off and explained by exogenous news arrival it provides an ideal test for the difference the BIP–GARCH makes compared to a standard GARCH model. We find that in the standard model the volatility shoots up sharply following the event, and conditional volatility forecasts are higher on average throughout the sample than when our procedure is used. In a portfolio of Apple and Microsoft, correlations fall sharply in October 2000 in the baseline model compared to our model, and throughout the sample remain lower. This is because an earnings surprise led to a 20% price increase for Microsoft, while Apple fell by 6% that same day. The BIP–cDCC model is designed to be robust against such cojumps triggered by once–off events.

Next, we run a Monte Carlo experiment to study the effect of jumps on the DCC and 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.

Third, 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 (2009), 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 correlation forecasting method. The BIP–cDCC forecasts rely on robust estimates of the unconditional mean, variance and correlation. We relegate the details on the computation of these estimators to the Appendix. 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 extremely large shocks caused by once–off events that cannot be expected to influence future volatility much. One example is 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.

We employed Gaussian quasi–maximum likelihood (QML) to estimate a GARCH$(1,1)$ model on the daily returns on Apple with a sample of one thousand days starting on the first day of 2000 and ending in December 2003. The results as well as the GARCH specification are reported in Table 1.

Table 1: Impact on GARCH of an extreme Apple return

Sample and method	α_1	β_1	$\alpha_1+\beta_1$	$\frac{\omega\times 10^4}{1-\alpha_1-\beta_1}$	$\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t$	VoV
GARCH full sample	0.157	0.824	0.981	29.421	15.140	1.543
GARCH after outlier	0.019	0.976	0.995	8.000	12.038	0.827
BIP–GARCH full sample	0.029	0.969	0.999	11.630	10.997	0.681

Note: The GARCH$(1,1)$ specification for the daily return series of Apple (y_t) is $y_t = \sqrt{h_t}z_t$ where $z_t \overset{i.i.d.}{\sim} N(0,1)$ and $h_t = \omega + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1}$. The robust GARCH specification is given in (2.4). VoV (volatility of volatility) is the standard deviation of $\sqrt{\hat{h}_t}$. The estimated volatilities are expressed in percentage points.

We then estimated the model using returns only after September 29, 2000 (we dummied out that day and got the same results). Including the extreme observation increases α_1 from 0.019 to 0.157, decreases β_1 from 0.979 to 0.824 and increases the long–run variance from 8×10^{-4} to about 30×10^{-4}. Including this once–off explainable event in the sample thus strongly biases the parameter estimates and, as a consequence, the out–of–sample forecasts. We denote this as *extreme bias*.

For comparison, we estimated a model taking this into account, the BIP–GARCH model discussed below, and found that the results are not affected much by the extreme observation. This can be seen in Figure 1 where we plot the daily Apple returns and the volatility forecasts obtained by the GARCH and BIP–GARCH model. We find that in the standard model the volatility shoots up sharply following the event, and conditional volatility forecasts are higher on average throughout the sample than when the BIP procedure is used.

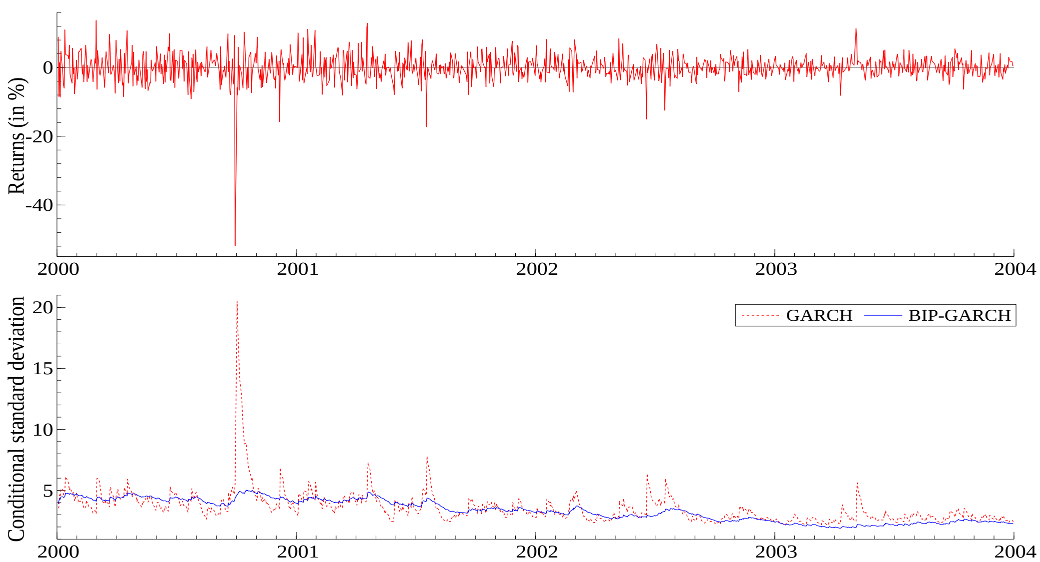

Figure 1: Daily returns in % (upper panel) for Apple and estimated conditional standard deviation for the GARCH and BIP–GARCH (lower panel) on the period 2000-2003.

Table 1 also reports two summary statistics on the estimated conditional volatilities. We see that the mean variance estimate from the GARCH model is only half of its value predicted by the model parameters, while for the BIP GARCH model, these values are very close. A final interesting observation is the difference in the estimated volatility of volatility for the two models. It is 1.543 for the GARCH model and only 0.681 for the BIP–GARCH model.

In conclusion, with this sample, the use of the BIP–GARCH model to forecast volatility lead to much more stable volatility forecasts, as is also clear from the time series plot of volatilities in Figure 1.

Other examples include the October 1987 crash caused by portfolio insurance induced automatic program trading, the extreme volatility during the downfall of LTCM in 1998, and several events during the recent crisis. It is also straightforward to demonstrate this by Monte Carlo experiments.

2.1 Proposals for Addressing the Extreme Bias

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.

Franses and Ghijsels (1999), Grossi (2004), Vlaar and Palm (1993) 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), Chan and Maheu (2002) and Benito, León, and Nave (2007). 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.

2.2 Bounded Innovation Propagation GARCH

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.

2.2.1 BIP–GARCH

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

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

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]

2.2.2 M–estimator of the BIP–GARCH

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, 2010), we recommend the one associated to the Student t_4 density

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

Table 2: 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

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 2 for $N = 1, 2, 5, 10$ and 50. Next we propose two modifications for the MY procedure that lead to more accurate volatility forecasts.

2.3 A Modified MY Procedure

To aid in forecasting the conditional variance more than one period into the future (see Section 2.4), we propose 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

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

with u a chi–square random variable with N degrees of freedom. Our simulations (available upon request) also show that this leads to a smaller bias in the GARCH parameters based on the BIP specification. Table 2 reports the correction factors $c_{\delta,1}$ for various values of δ. Note that we used this BIP–GARCH model in Table 1 with $\delta = 0.975$.

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 h) before estimating the remaining parameters. Francq, Horvath, and Zakoian (2010) 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 (2008) and described in Appendix A. In the webappendix (Boudt, Daníelsson, and Laurent, 2010), 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.

2.4 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:

$$h_{t+r|t} = \hat{\omega} + \hat{\alpha}_1 y^2_{t+r-1|t} + \hat{\beta}_1 h_{t+r-1|t}, \tag{2.11}$$

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

$$h_{t+r|t} = \hat{\omega} + \hat{\alpha}_1 wy^2_{t+r-1|t} + \hat{\beta}_1 h_{t+r-1|t}, \tag{2.12}$$

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 (2.11) and (2.12) 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. While little research has demonstrated the impact of extreme observations for correlations, it is readily demonstrated, e.g.

by adding one asset to the example with Apple in Section 2, e.g. Microsoft. Figure 2 plots the daily returns (in %) for these two series. Note the 20% return on the Microsoft stock price, triggered by the once–off event of Microsoft posting first–quarter net income of 46 cents per share, 12 percent above the mean analyst estimate of 41 cents. The same day the stock price of Apple fell by 6%.

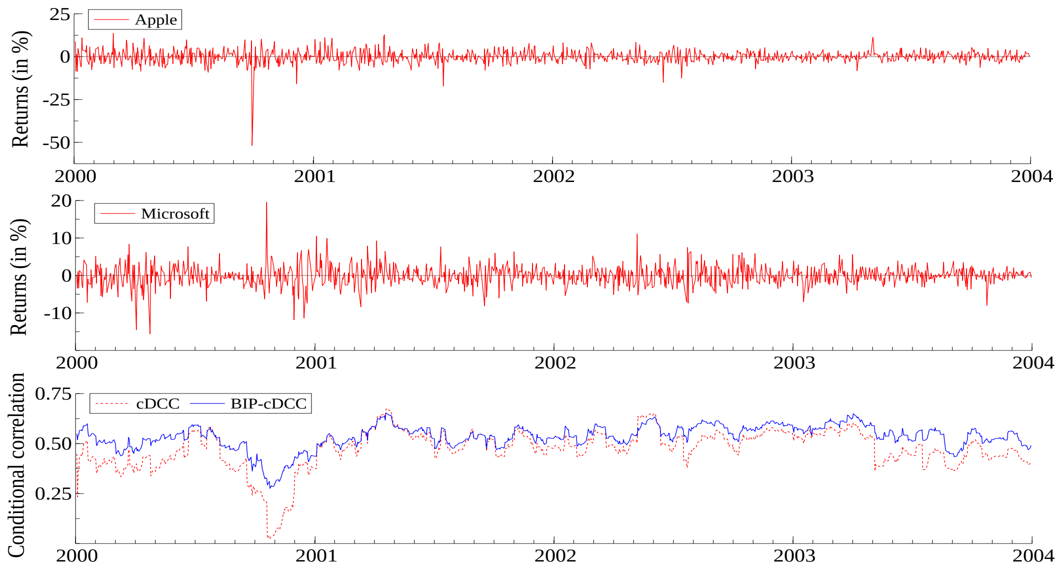

Figure 2: Daily returns in % for Apple and Microsoft (first two panels) and estimated cDCC and BIP-cDCC conditional correlation (lower panel) for the period 2000-2003.

Table 3 shows that the effect of this extreme is to cause cDCC conditional correlations to drop in one day from 21.5% to only 3.2%, while historically the average conditional correlation is around 45%. The effect of this extreme is persistent, since it takes more than a month for the estimated conditional correlation to return to its level before the once–off event.

For comparison, we estimated a BIP version of the cDCC model and found the conditional correlation only dropping by 3 percentage points, not 18 points like the baseline model. We further note the strong difference between the two unconditional correlation estimates (about 45% for the cDCC and about 55% for the BIP version) leading the BIP–cDCC correlation to be significantly higher than the cDCC corre-

Table 3: Apple/MSFT return and extremes, impact on cDCC–GARCH. Estimation on the full sample (i.e., January 2000 – December 2003).

	cDCC parameters			Correlations t_0 = Oct 19, 2000				
Method	$\overline{Q}_{12}$	α	β	$\hat{R}_{12,t_0}$	$\hat{R}_{12,t_0+1}$	$\hat{R}_{12,t_0+5}$	$\hat{R}_{12,t_0+15}$	$\hat{R}_{12,t_0+20}$
cDCC	0.449	0.026	0.958	0.215	0.032	0.041	0.091	0.167
BIP–cDCC	0.549	0.021	0.956	0.326	0.293	0.288	0.326	0.379

Note: $\hat{R}_{12,t}$ corresponds to the estimated conditional correlation (between Apple and Microsoft) of the classical and robust DCC models on day t.

lation for almost all days in the sample, as can be seen in the lower panel of Figure 2.

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:

$$S_t = Y_t + A_t \tag{3.1}$$

$$Y_t = H_t^{1/2} Z_t \quad \text{where} \quad Z_t \overset{i.i.d.}{\sim} N(0, I_N). \tag{3.2}$$

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.,

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

Then in the cDCC model the 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 $\alpha + \beta < 1$.[3] This matrix is then linked to the conditional correlation matrix as follows

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

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 $i = 1, \ldots, N$), Aielli (2009) shows that for given values of α and β

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

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] 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.

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 2.3 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.[4]

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

[4] The use of thresholds to model the conditional correlations of financial return series is also considered in Audrino and Trojani (2010). 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 (2010) 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.

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 $w_{k_{\delta,N}}(u)$ 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 2). 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 $\delta = 0.975$.

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 2.3.

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 $\tilde{S}_t$. 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 Appendix A. Details on the construction of this estimator are given in Appendix B.

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. A general class of M–estimators for MGARCH models is where one minimizes 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.,

$$\hat{\theta}^M = \text{argmin} M_T(\theta, \rho) \equiv \frac{1}{T} \sum_{t=1}^{T} \left[\log \det R_t \ + \sigma \rho \left(\tilde{S}_t' R_t^{-1} \tilde{S}_t \right) \right], \tag{3.10}$$

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:

$$\frac{\partial M_T(\theta, \rho)}{\partial \theta_j} = \frac{1}{T} \sum_{t=1}^{T} \ \text{Tr} \left[I_N - \sigma \rho'(\tilde{S}_t' R_t^{-1} \tilde{S}_t) \tilde{S}_t \tilde{S}_t' R_t^{-1} \right] \frac{\partial R_t}{\partial \theta_j} R_t^{-1} = 0, \tag{3.11}$$

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 2.

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 Q_t are obtained using the BIP version of (3.7), i.e.,

$$q_{ii,t} \;=\; (1-\alpha-\beta) + \alpha w(\tilde{s}^2_{i,t-1}) q_{ii,t-1}\tilde{s}^2_{i,t-1} + \beta q_{ii,t-1}. \tag{3.12}$$

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

$$H_{t+r|t} = D_{t+r|t} R_{t+r|t} D_{t+r|t}, \tag{3.13}$$

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 $r > 1$

$$Q_{t+r|t} = (1-\hat{\alpha}-\hat{\beta})\hat{\overline{Q}} + (\hat{\alpha}+\hat{\beta}) Q_{t+r-1|t}, \tag{3.14}$$

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. 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.

Simulation setup: We generate bivariate returns S_t^* as the sum of a standard bivariate GARCH(1,1)–cDCC process and a jump process A_t. Let $t_1, \ldots, t_l$ be the times when jumps are observed. The simulated returns are given by:

$$\begin{aligned}
S_t^* &= \begin{pmatrix} 0.05 \\ -0.05 \end{pmatrix} + \begin{cases} Y_t + A_t & \text{if } t = t_i, 1 \leq i \leq l = \varepsilon T \\ Y_t & \text{elsewhere }, \end{cases} \\
Y_t &= H_t^{1/2} Z_t, Z_t \sim N(0, I_2) \\
h_{1,t} &= 0.1 + 0.1 y_{1,t-1}^2 + 0.8 h_{1,t-1} \\
h_{2,t} &= 0.1 + 0.2 y_{2,t-1}^2 + 0.7 h_{2,t-1} \\
Q_t &= (1 - 0.1 - 0.8)\overline{Q} + 0.1 P_{t-1} \tilde{Y}_{t-1} \tilde{Y}_{t-1}' P_{t-1} + 0.8 Q_{t-1},
\end{aligned}$$

where $\overline{Q}_{1,2} = \overline{Q}_{2,1} = 0.4$ and $t = 1, \ldots, T$, with $T = 2000$. The values $t_1, \ldots, t_l$ were chosen equally spaced and $\varepsilon = 0\%$, 1% or 5%. The jump size is the conditional standard deviation of the corresponding elements of Y_t times d for the first series and negative d for the second series, with $d = 3$ or 4. The two assets have the same jump probability and 40% of jumps are cojumps. Consequently, $\varepsilon = 1\%$ (resp. 5%) corresponds to on average 0.7% (resp. 3.5%) of jumps on each series.

The nine unknown parameters are estimated using the approach described in the previous section, i.e., estimation of the 2 univariate GARCH models, followed by estimation of $\overline{Q}$, α and β. For each parameter, we compute the estimation bias and RMSE over $10,000$ replications. For the BIP–GARCH and BIP–cDCC, we consider a threshold given by $\delta = 0.975$.[5]

Results: The bias and RMSE of the BIP–cDCC and the benchmark QML estimator of the parameters of the unobserved GARCH(1,1)–cDCC process Y_t are shown in Table 4. Consider first the bias and RMSE for the parameters of the univariate

[5] 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 4: Bias and RMSE of the Gaussian QML and robust estimator for the cDCC model in presence of ε jumps of size d conditional standard deviation, with $\delta = 0.975$ and $T = 2000$.

			μ	ω	α_1	β_1	h	$\overline{Q}_{12}$	α	β
			0.050	0.100	0.100	0.800	1.000	0.400	0.100	0.800
$\varepsilon = 0\%$	QML	bias	0.001	0.008	0.000	-0.009	-0.001	-0.001	-0.001	-0.007
		RMSE	0.023	0.036	0.020	0.049	0.067	0.039	0.019	0.045
	Robust	bias	0.001	0.008	0.005	-0.013	0.003	-0.011	-0.002	-0.006
		RMSE	0.026	0.040	0.023	0.056	0.074	0.040	0.021	0.051
$\varepsilon = 1\%$	QML	bias	0.020	0.028	-0.004	-0.018	0.058	-0.055	-0.010	-0.057
$d = 3$		RMSE	0.031	0.056	0.023	0.064	0.091	0.067	0.033	0.124
	Robust	bias	0.005	0.012	0.001	-0.013	0.013	-0.016	-0.007	-0.006
		RMSE	0.026	0.044	0.022	0.058	0.076	0.041	0.022	0.055
$\varepsilon = 1\%$	QML	bias	0.027	0.060	-0.005	-0.041	0.104	-0.092	-0.008	-0.125
$d = 4$		RMSE	0.035	0.098	0.029	0.098	0.128	0.101	0.043	0.203
	Robust	bias	0.003	0.013	0.000	-0.013	0.011	-0.016	-0.007	-0.006
		RMSE	0.026	0.044	0.022	0.058	0.076	0.041	0.022	0.055
$\varepsilon = 5\%$	QML	bias	0.103	0.169	-0.030	-0.079	0.300	-0.230	-0.070	-0.166
$d = 3$		RMSE	0.106	0.252	0.046	0.186	0.311	0.233	0.080	0.277
	Robust	bias	0.022	0.031	-0.018	-0.006	0.060	-0.037	-0.034	0.015
		RMSE	0.034	0.073	0.030	0.079	0.100	0.053	0.042	0.094
$\varepsilon = 5\%$	QML	bias	0.137	0.397	-0.059	-0.164	-0.536	-0.339	-0.084	-0.219
$d = 4$		RMSE	0.106	0.252	0.046	0.186	0.311	0.341	0.090	0.284
	Robust	bias	0.013	0.029	-0.024	0.000	0.046	-0.033	-0.039	0.026
		RMSE	0.029	0.077	0.034	0.086	0.091	0.051	0.046	0.100

The bias and RMSE of the parameters underlying $h_{1,t}$ and $h_{2,t}$ are similar. To save space, we only report those for $h_{1,t}$.

GARCH model for $s^*_{1,t}$. In line with the results in MY, we find that the estimation of the GARCH parameters using the misspecified BIP–GARCH model does not seem to create any significant bias in the estimated parameter values. Of course, we see that in the absence of additive jumps (i.e., $\varepsilon = 0\%$), we pay the price of a loss of efficiency with respect to the QML estimator. But when $\varepsilon = 1$ or 5%, the QML estimator is severely biased.

The last three columns of Table 4 present the results for the multivariate case. Like in the univariate case, the estimation of the cDCC parameters using the BIP–cDCC models does not seem to create any significant bias in the estimated parameter values in the absence of jumps (i.e., $\varepsilon = 0\%$). The average of the estimated parameters is very close to the true values. Since the innovations have a conditionally Gaussian distribution, the Gaussian QML estimator based on the correctly specified GARCH model is expected to have (at least asymptotically) the lowest RMSE.

The loss of efficiency of the robust estimator in the absence of additive jumps is moderate compared to the lower bias and gain in efficiency in the presence of these jumps. For $\varepsilon = 1$ or 5% of additive jumps, we find the empirical correlation of the devolatilized returns to be a strongly biased estimate of $\overline{Q}_{1,2}$. Because jumps have the opposite sign and the true correlation is 0.4, we find a negative bias of -5.5% when $\varepsilon = 1\%$ and $d = 3$ and -33.9% when $\varepsilon = 5\%$ and $d = 4$. The persistency parameter β also is largely underestimated. Its bias is -5.7% when $\varepsilon = 1\%$ and $d = 3$ and -21.9% when $\varepsilon = 5\%$ and $d = 4$. When $\varepsilon = 1\%$, the bias in the QML estimate of α is still negligible, but for $\varepsilon = 5\%$ with $d = 4$, we find a bias of -8.4%.

Importantly, in all cases, the bias and RMSE of the estimates of the proposed robust estimator described in Section 3.2 remains small in the presence of additive jumps.

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 (2009) 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 (2009), we set the confidence 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.[6]

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.[7] 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 (corrected) 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 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 (2008). 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

[6]Implementation of this test has been done using the Ox software package MULCOM of Hansen and Lunde (2007).

[7]The EWMA model has been popularised by Riskmetrics (1996) and is widely used by practitioners.

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).[8] 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 (2009b) 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 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 $r = 1, \ldots, 10$.

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., $H_{m,t}$).[9]

[8] 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.

[9] Hansen and Lunde (2006), Laurent, Rombouts, and Violante (2009a) and Patton (2009) show

Table 5: 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																				

Results: Table 5 indicates which models belonging to the set of superior forecasting models according to the MCS test. In both the calm and turbulent market regimes, a MGARCH model with the BIP property is selected as having superior forecasting performance for all forecast horizons. Interestingly, in the calm period a BIP–model without dynamics in the conditional correlations belongs consistently to the set of superior forecasting models, i.e. the BIP–CCC. When accuracy is measured against

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 (2009a) 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.

the RCov or RBPCov, the BIP–CCC model is even the best model for all r–day ahead forecasts, with $r < 8$. Over the credit crisis (2007–2009), the BIP–cDCC model always belongs to the set of best forecasting models. It significantly beats all other models in forecasting the RBPCov and ROWCov at all forecast horizons, and the RCov for horizons $r > 2$.

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 40%.[10]

The portfolios are fully invested in equity belonging to the same sector. The initial investment universe consists of all S&P 100 stocks on July 31, 2010. Stocks with missing data at the start of the estimation period and sectors with less than five stocks in the investment universe are removed.[11] 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.[12] We split up the evaluation sample in the calm

[10]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.

[11]This leads to the following sector portfolios: Consumer Discretionary (tickers: AEP, CMCSA, ETR, F, HD, MCD, TGT, NKE, SO, TGT, TWX, DIS), Consumer Staples (MO, AVP, CL, COST, CPB, CVS, HNZ, KO, PEP, PG, SLE, WAG, WMT), Energy (BHI, COP, CVX, DVN, HAL, OXY, SLB, WMB, XOM), Financials (ALL, AXP, BAC, BK, C, JPM, L, MS, RF, USB, WFC), Healthcare (ABT, AMGN, BAX, BMY, GILD, JNJ, MDT, MRK, PFE, UNH), Industrials (BA, CAT, FDX, GD, GE, HON, LMT, MMM, NSC, RTN, UTX) and IT (AAPL, CSCO, DELL, EMC, HPQ, INTC, IBM, MSFT, ORCL, QCOM, TXN, XRX).

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

period of 2004–2006 and the turbulent 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 out–of–sample gross returns. Following Fleming, Kirby, and Ostdiek (2003), we consider the mean and standard deviation, as well as a summary measure, denoted Δ_γ, for the relative outperformance of the BIP–cDCC method versus the cDCC method. Let p_t and $p_{(\mathrm{BIP})t}$ be the portfolio returns under the classical and BIP approach, respectively ($t = 1, \ldots, T$). The value of Δ_γ is such that an investor with quadratic utility function and relative risk aversion parameter γ, is indifferent between receiving p_t and $p_{(\mathrm{BIP})t} - \Delta_\gamma$, for $t = 1, ..., T$.[13] Table 6 reports the mean, standard deviation and Δ_γ in annualized terms, with Δ_γ computed using two different values of γ, 1 and 10.

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 6 also summary statistics on the right tail distribution of the daily changes in portfolio weights, i.e. 99% quantile, maximum and kurtosis of $|w_{(i)t} - w_{(i)t-}|$, where $w_{(i)t}$ is the optimal weight of asset i on day t and $w_{(i)t-}$ is the weight of that asset before rebalancing at the end of day $t-1$ to $w_{(i)t}$.

Note first in Table 6 that the use of the BIP method has little impact on the portfolio standard deviation. For healthcare in 2004–2006, it has a negative impact on the average portfolio returns, while for industrials in 2004–2006, financials in 2007–2009 and IT in 2007–2009, it improves the portfolio return. For all other cases, the impact of the covariance forecast method on the average gross returns seems to be negligible.

Focusing on the computed values of Δ_γ, we see that for the turbulent period the robust BIP allocation outperforms slightly the classical allocation, while in the calm period, there is no evidence of dominance of one method over the other. We thus find

[13] More precisely, assuming the investor's investment budget is the same every day, Δ_γ is determined by the condition that

$$\sum_{t=1}^{T}(1 + p_{(\mathrm{BIP})t} - \Delta_\gamma) - \frac{\gamma}{2(1+\gamma)}(1 + p_{(\mathrm{BIP})t} - \Delta_\gamma)^2 = \sum_{t=1}^{T}(1 + p_t) - \frac{\gamma}{2(1+\gamma)}(1 + p_t)^2.$$

Table 6: 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 and relative performance Δ_γ) and portfolio turnover.

	Gross return performance						Turnover					
	cDCC		BIP–cDCC		Δ_γ		cDCC			BIP–cDCC		
	mean	SD	mean	SD	$\gamma = 1$	$\gamma = 10$	Q99	max	kurt	Q99	max	kurt
2004-2006												
Cons.Discret.	0.12	0.10	0.11	0.10	-1.13	-1.35	0.07	0.23	37.55	0.07	0.16	16.52
Cons.Staples	0.08	0.09	0.09	0.10	0.85	0.67	0.06	0.23	57.51	0.07	0.15	14.21
Energy	0.28	0.19	0.30	0.20	1.99	0.39	0.09	0.24	30.06	0.11	0.28	18.38
Financials	0.14	0.10	0.13	0.10	-0.81	-0.68	0.09	0.36	45.80	0.12	0.31	15.30
Healthcare	0.06	0.11	0.03	0.12	-2.89	-3.53	0.13	0.35	25.87	0.08	0.17	15.69
Industrials	0.13	0.11	0.16	0.11	2.53	2.52	0.06	0.25	48.16	0.08	0.15	10.66
IT	0.04	0.13	0.03	0.14	-1.28	-1.60	0.06	0.38	198.38	0.08	0.19	15.35
2007-2009												
Cons.Discret.	0.00	0.22	0.00	0.23	-0.00	-0.66	0.09	0.30	35.97	0.09	0.21	17.00
Cons.Staples	-0.01	0.20	0.02	0.19	2.95	4.41	0.08	0.34	57.53	0.09	0.18	13.01
Energy	-0.05	0.37	-0.04	0.37	1.63	0.97	0.12	0.39	37.42	0.14	0.32	20.94
Financials	-0.27	0.46	-0.24	0.47	3.45	-1.03	0.11	0.37	55.77	0.12	0.31	21.52
Healthcare	-0.01	0.21	-0.01	0.20	-0.63	0.84	0.13	0.39	37.99	0.07	0.19	16.63
Industrials	-0.03	0.26	-0.03	0.26	0.27	1.11	0.10	0.37	55.02	0.10	0.24	15.00
IT	0.01	0.29	0.05	0.29	4.17	4.72	0.08	0.36	62.62	0.09	0.19	11.85

that on aggregate, the returns on the portfolios constructed using the BIP method have similar unconditional first and second moments as when the baseline model is used.

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 in the last columns of Table 6 that the maximum and the kurtosis of the portfolio turnover are always significantly larger for the classical portfolio than for the robust portfolio: the percentage difference between the maximum (kurtosis) of the turnover is between 30 and 260% (80 and 900%). Therefore, even if both procedures have a similar out–of–sample mean return and standard deviation, net of transaction costs, the BIP procedure yields more profits overall.

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 model always belongs to set of superior forecasting models, and for most forecast horizons, it is identified as the best model by the model confidence set approach of Hansen, Lunde, and Nason (2009). 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.

The stability of the robust multivariate volatility forecasts should be an attractive characteristic for volatility–based economic capital determination by financial institutions. It would be interesting to compare the accuracy of downside risk estimates calculated using the BIP covariance estimates relatively to using the baseline DCC model.

We leave further work along these lines for future research.

Appendix A: Reweighted Mean and Variance Estimators

In a high-frequency data setting, Boudt, Croux, and Laurent (2008) propose to estimate the unconditional mean μ and variance h through a mean and variance estimate in which local outliers receive a zero mean. 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, 2010). This method first estimates for each observation the median absolute deviation mad_t of the returns in a window around that observation.[14] The reweighted sample mean and variance are then

$$\hat{\mu} = \frac{\sum_{t=1}^{T} s_t^* I_t}{\sum_{t=1}^{T} I_t} \quad \text{and} \quad \hat{h} = b_{0.95,1} \cdot \frac{\sum_{t=1}^{T} (s_t^* - \hat{\mu})^2 J_t}{\sum_{t=1}^{T} J_t}, \tag{7.1}$$

with

$$I_t = I\left[\frac{(s_t^* - \text{median}_t(s_t^*))^2}{mad_t^2} \leq \chi_1^2(95\%)\right] \quad \text{and} \quad J_t = I\left[\frac{(s_t^* - \hat{\mu})^2}{mad_t^2} \leq \chi_1^2(95\%)\right], \tag{7.2}$$

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 $b_{0.95,1} = 0.95 c_{0.95,1}$ is a constant adjusting for the bias due to the thresholding, with $c_{0.95,1}$ as defined in (2.10).

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

[14] 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.

models. We found that, since K affects only the weights in (7.2) 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).

Appendix B: Reweighted Correlation Estimators

As for the reweighed variance estimator in (7.2), 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:

$$SC_t = 2\sin\left(\frac{1}{6}\pi C_t\right). \tag{7.3}$$

Advantages of the Spearman correlation matrix with respect to other robust correlation estimators include its computational simplicity, high efficiency and outlier robustness, as shown by Croux and Dehon (2009). 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 quasi-reweighted correlation estimator:

$$RC \;=\; \frac{c_{0.95,N}}{\sum_{t=1}^{T} L_t} \sum_{t=1}^{T} \tilde{S}_t \tilde{S}_t' L_t, \tag{7.4}$$

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

$$\hat{\bar{Q}}_{RW} = \text{diag}\;(RC_{11}^{-1/2} \ldots RC_{NN}^{-1/2})\; RC\; \text{diag}\;(RC_{11}^{-1/2} \ldots RC_{NN}^{-1/2}). \tag{7.5}$$

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).

References

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

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 (2010): "A General Multivariate Threshold GARCH Model with Dynamic Conditional Correlations," *Forthcoming in Journal of Business and Economics Statistics.*

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

Benito, F., A. León, and J. Nave (2007): "Modeling the Euro Overnight Rate," *Journal of Empirical Finance*, 14, 756–782.

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 (2008): "Outlyingness Weighted Covariation," *Mimeo.*

——— (2010): "Robust Estimation of the Periodicity in Intraday Volatility and Intraday Jump Detection," *Forthcoming in Journal of Empirical Finance.*

Boudt, K., J. Daníelsson, and S. Laurent (2010): "Appendix to Robust Forecasting of Dynamic Conditional Correlation GARCH Models," *Available at www.econ.kuleuven.be/kris.boudt/public.*

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," *Forthcoming in Journal of Econometrics.*

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

Croux, C., and C. Dehon (2009): "Influence Functions of the Spearman and Kendall Correlation Measures," *Forthcoming in Statistical Methods and Applications.*

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.

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

FLEMING, J., C. KIRBY, AND B. OSTDIEK (2003): "The Economic Value of Volatility Timing Using Realized Volatility," *Journal of Financial Economics*, 67, 473–509.

FRANCQ, C., L. HORVATH, AND J.-M. ZAKOIAN (2010): "Merits and Drawbacks of Variance Targeting in GARCH Models," University of Lille 3 Working Paper.

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.

GROSSI, L. (2004): "Analyzing Financial Time Series through Robust Estimators," *Studies in Nonlinear Dynamics & Econometrics*, 8, Article 3.

HANSEN, P., AND A. LUNDE (2006): "Consistent Ranking of Volatility Models," *Journal of Econometrics*, 131, 97–121.

——— (2007): "MULCOM 1.00, Econometric Toolkit for Multiple Comparisons," http://www.hha.dk/∼alunde/MULCOM/MULCOM.HTM.

HANSEN, P., A. LUNDE, AND J. NASON (2009): "Model Confidence Sets," Forthcoming in Econometrica.

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.

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

Laurent, S., J. Rombouts, and F. Violante (2009a): "On Loss Functions and Ranking Forecasting Performances of Multivariate Volatility Models," Cirano discussion paper 2009-45.

——— (2009b): "On the Forecasting Accuracy of Multivariate GARCH Models," Cirano discussion paper 2010-21.

Lopuhaä, H. (1999): "Asymptotics of Reweighted Estimators of Multivariate Location and Scatter," *Annals of Statistics*, 27, 1638–1665.

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. (2009): "Volatility Forecast Comparison Using Imperfect Volatility Proxies," *Forthcoming in Journal of Econometrics.*

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.

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.