Average Conditional Correlation and Tree Structures for Multivariate GARCH Models

Francesco Audrino* and Giovanni Barone-Adesi†

Institute of Finance
University of Lugano

First version: May 2004

Final version: May 2006

Abstract

We propose a simple class of multivariate GARCH models, allowing for time-varying conditional correlations. Estimates for time-varying conditional correlations are constructed by means of a convex combination of averaged correlations (across all series) and dynamic realized (historical) correlations. Our model is very parsimonious. Estimation is computationally feasible in very large dimensions without resorting to any variance reduction technique. We back-test the models on a six-dimensional exchange-rate time series using different goodness-of-fit criteria and statistical tests. We collect empirical evidence of their strong predictive power, also in comparison to alternative benchmark procedures.

Keywords: Multivariate GARCH models; Dynamic conditional correlations; Tree-structured GARCH models; Model confidence set approach.

JEL codes: C12, C13, C51, C53, C61

*Corresponding author. Address: Institute of Finance, Via Buffi 13, Centrocivico, CH-6900 Lugano, Switzerland. E-mail: francesco.audrino@lu.unisi.ch. Phone: 0041586664789. Financial support by the Foundation for Research and Development of the University of Lugano and by the National Centre of Competence in Research "Financial Valuation and Risk Management" (NCCR FINRISK) is gratefully acknowledged.

†Address: Institute of Finance, Via Buffi 13, Centrocivico, CH-6900 Lugano, Switzerland. E-mail: giovanni.barone-adesi@lu.unisi.ch.

1 Introduction

We propose a simple class of multivariate GARCH models to estimate conditional covariance matrices with time-varying conditional correlations. Our class takes into account possible non-linear dependence structures across individual series. Modelling and forecasting conditional covariance matrices is an important and central problem in modern empirical finance, since the covariance matrix is an essential element of many problems in financial econometrics such as the computation of risk measure estimates for portfolios of assets, asset allocation or tests of asset pricing models. In the last decade there has been a tremendous number of studies focusing on the time-varying behavior of correlations and covariances of financial instruments.

It is now widely accepted that financial volatilities and correlations move together over time across assets and markets. In most financial applications modelling the covariance matrix dynamics by a suitable multivariate approach yields more appropriate empirical models and allows for better decisions than using a separate univariate model for each individual financial instrument[1]. Moreover for many relevant problems it is not possible to reduce complexity by working with univariate models. Consider for example a portfolio with price $P_t = \sum_i w_i P_{t,i}$ and portfolio weights w_i. A naive approach may suggest that for predicting volatility of the portfolio returns the multivariate problem can be bypassed to a large extent by just looking at the univariate portfolio price process $\{P_t; t = 1, \ldots, n\}$. Proceeding in this way however, a substantial information loss has typically to be paid resulting in less accurate volatility predictions for portfolio returns. But more important is the fact that for *time-changing* portfolio weights – which is most often the case in practice – portfolio returns become typically non-stationary. We then have to model the multivariate time series of asset returns in order to obtain accurate volatility predictions.

In this paper, we focus on multivariate extensions of the simple univariate GARCH(1,1) model firstly introduced by Bollerslev (1986). This model is often used as a benchmark in practice (see Andersen et al., 1999; Lee and Saltoglu, 2001 or Hansen and Lunde, 2002 among others). When estimating time-varying conditional covariance matrices using multivariate GARCH-type models (for a recent survey, see Bauwens et al., 2003), we have to face additional problems especially when the number of individual instruments is in the order of several dozens or hundreds, as it is the case in most practical applications.

In particular, when the dimension of the problem is high, it can be almost unfeasible to estimate general multivariate GARCH models, such as the VEC (Bollerslev et al., 1988) or the BEKK models (Engle and Kroner, 1995), due to the well-known curse of dimensionality. Moreover, further restrictions have to be imposed on the parameters to ensure positivity of the covariance matrix and to avoid over-fitting. A final issue is related to the selection of the optimal model based on standard criteria such as the Akaike Information Criterium (AIC) or the Schwarz Bayesian Information Criterium (BIC). This problem is well illustrated by a standard BEKK(1,1) model with 10 assets. When using AIC, we have to check and fit more than 10^{73} models. This is clearly too expensive and not computationally feasible.

For these reasons, researchers have been often constrained to estimate models for time-varying covariances and correlations under considerable restrictions. Engle et al. (1990) proposed some Factor models, where the co-movements of the different instruments are driven by a small number of common factors. Alexander and Chibumba (1997) and Alexander (2001) have recently introduced a particular class of Factor models, called Orthogonal GARCH (O-GARCH) models. In such models the time-varying covariance matrix is generated by a small number of orthogonal univariate GARCH models, identified using principal components analysis. In contrast, in this paper we propose to estimate the dynamics of time-varying covariances and correlations using full multivariate GARCH models. This avoids the use of variance reduction or similar techniques which can yield very poor forecasts in some practical applications.

Bollerslev (1990) introduced a new class of multivariate GARCH models: the Constant Conditional Correlation (CCC) GARCH models. In such models univariate GARCH processes are estimated for each financial instrument. The correlation matrix is then computed using the standard MLE correlation estimator applied to a sequence of standardized residuals. This constant conditional correlation structure ensures also in large dimensions the feasibility of the model estimation and the positivity of the covariance matrix. However, conditional correlations seem not to be constant trough time for many empirical applications (see Tsui and Yu, 1999 and Tse, 2000, among others).

Therefore a lot of work has been recently devoted to develop models allowing also correlations to change over time. Tse and Tsui (2002), Engle (2002) and Engle and Sheppard (2001) proposed a generalization of the CCC-GARCH model where the conditional correlation matrix is time dependent. The multivariate Dynamic Conditional Correlation (DCC) GARCH model introduced by Engle (2002) added to the CCC model some dynamics in the correlations, introducing a GARCH-type structure. The DCC model, which is now very popular, guarantees the positivity of the conditional correlation matrix under simple conditions on the parameters.

However, the dynamics are constrained to be equal for all correlations. In the last year, Billio et al. (2003) generalize the DCC model constraining the dynamics of the conditional correlations to be equal only among groups of variables. Other models allowing conditional correlations to change over time have been recently proposed by Ledoit et al. (2003), Pelletier (2002) and Baur (2003) using different approaches and techniques. However, the forecasting power of such models has not been yet investigated and compared extensively.

Similarly to the DCC-GARCH model, our approach preserves the ease of estimation of Bollerslev's CCC-GARCH model while allowing correlations to change over time. In our model, estimates and forecasts for time-varying conditional correlations are constructed by means of a convex combination of realized (historical) correlations and estimates of averaged correlations (across all series). The estimation of the averaged correlations involves only univariate GARCH volatility processes for each financial instrument and for the corresponding equally weighted portfolio and is therefore computationally feasible also in large dimensions. The estimation procedure is similar to the two-stage one used in the DCC model.

We test our model on a six-dimensional time series of exchange-rate data. We compare its out-of-sample forecasting power with the CCC-GARCH and the DCC-GARCH models, both at the multivariate and univariate portfolio level. Moreover, we also use the idea of statistical hypothesis testing on differences of performance terms across the models to eliminate the strong noise component. It is generally difficult to answer the question "Which is the best model?" because asset returns do not contain sufficient information to identify a single volatility model as "best". For this reason, we apply the Model Confidence Set (MCS) method proposed by Hansen et al. (2003) to characterize the set of models that significantly dominate others. In this exercise, we collect empirical evidence of the strong predictive potential of our model and show that in the most cases it improves both on the CCC-GARCH and the DCC-GARCH models.

Finally, in a practical application for Value-at-Risk (VaR) computation of an equally weighted portfolio (similar to Ledoit et al., 2003), we find that our approach yields accurate VaR estimates.

The remainder of the paper is structured as follows. Section 2 presents our model for time-varying conditional correlations. The estimation procedure is in Section 3. Empirical goodness-of-fit and forecasting results for a six-dimensional exchange-rate time series both at the multivariate and at the univariate portfolio level are in Section 4. Section 5 summarizes and concludes.

2 The models

This Section describes the new proposed class of multivariate GARCH models for dynamic conditional correlations.

2.1 Starting point

Let the multivariate time series of daily log-returns (in percentages) of d assets be denoted by

$$\mathbf{X}_t = 100 \cdot \begin{pmatrix} \log\left(\frac{P_{t,1}}{P_{t-1,1}}\right) \\ \vdots \\ \log\left(\frac{P_{t,d}}{P_{t-1,d}}\right) \end{pmatrix} = 100 \cdot \Big(\log\left(\mathbf{P}_t\right) - \log\left(\mathbf{P}_{t-1}\right)\Big), \tag{2.1}$$

where $P_{t,i}$ is the value of the asset i at day t. We assume stationarity of this series. Our goal is to find in-sample and out-of-sample estimates for the time-varying conditional covariance matrix of the returns $\mathbf{X}_t$. To this purpose, we consider a multivariate approach to model the conditional covariance matrix $V_t = \mathrm{Cov}_{d\times d}(\mathbf{X}_t|\mathcal{F}_{t-1})$ of $\mathbf{X}_t$, where $\mathcal{F}_{t-1}$ denotes the information available up to time $t-1$.

For exposition purposes it is useful to start by a general semiparametric model for $\mathbf{X}_t$ of the form

$$\mathbf{X}_t = \mu_t + \Sigma_t \mathbf{Z}_t. \tag{2.2}$$

The following assumptions on the process (2.2) are imposed.

(A1) (innovations) $\{\mathbf{Z}_t\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. zero mean multivariate innovations having covariance matrix $\mathrm{Cov}(\mathbf{Z}_t) = I_d$.

(A2) (conditional correlation construction) The conditional covariance matrix $V_t = \Sigma_t\Sigma_t^T$ is almost surely positive definite for all t. The typical element of V_t is $v_{t,ij} = \rho_{t,ij}(v_{t,ii}v_{t,jj})^{1/2}$ $(i,j = 1,\ldots,d)$. In this model $\rho_{t,ij} = \mathrm{Corr}(x_{t,i}, x_{t,j}|\mathcal{F}_{t-1})$ equals the conditional correlation at time t. Hence, $-1 \leq \rho_{t,ij} \leq 1$, $\rho_{t,ii} = 1$.

(A3) (functional nonparametric form for conditional variance) The conditional variances are functions of the form

$$v_{t,ii} = \sigma_{t,i}^2 = \mathrm{Var}(X_{t,i}|\mathcal{F}_{t-1}) = F_i(\{X_{t-j,k};\ j = 1,2,\ldots,k = 1,\ldots,d\})$$

where F_i is a function that takes values in $\mathbb{R}^+$.

(A4) (conditional mean)[2] The conditional mean μ_t is of the form

$$\mu_t = (\mu_{t,1}, \ldots, \mu_{t,d})^T = A_0 + A_1 \mathbf{X}_{t-1}$$

with both $A_0 = \mathrm{diag}(a_{0,1}, \ldots, a_{0,d})$ and $A_1 = \mathrm{diag}(a_{1,1}, \ldots, a_{1,d})$ diagonal $d \times d$ matrices.

Note that (A2) can be also rewritten in matrix form as

$$V_t = \Sigma_t \Sigma_t^T = D_t R_t D_t,$$
$$D_t = \mathrm{diag}(\sigma_{t,1}, \ldots, \sigma_{t,d}), \; R_t = [\rho_{t,ij}]_{i,j=1}^d.$$

The functional form (A3) allows for cross-dependence across the different components, since the conditional variance of all components depends on past multivariate observations. This is one of the nice features of such a multivariate GARCH-type model and is motivated by the fact that generally time series of asset returns are highly cross-correlated. The dependence of $\sigma_{t,i}, \; i = 1, \ldots, d$ on $\mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \ldots,$ allows for a broad variety of asymmetric and non-linear volatility patterns in response to past multivariate market information.

Several models in the literature are special cases of the above general setting. For instance, the parametric CCC-GARCH(1,1) model of Bollerslev (1990) is encompassed by (2.2) if we impose the further constraints:

$$\text{(constant conditional correlations)} \quad R_t \equiv R \text{ for all } t;$$
$$\text{(GARCH(1,1) volatilities)} \quad \sigma_{t,i}^2 = \alpha_{0,i} + \alpha_{1,i}(X_{t-1,i} - \mu_{t-1,i})^2 + \beta_i \sigma_{t-1,i}^2,$$
$$\alpha_{0,i} > 0, \alpha_{1,i} \geq 0, \beta_i \geq 0, \alpha_{1,i} + \beta_i < 1, \; i = 1, \ldots, d. \tag{2.3}$$

In this model, the correlations are constant over time.

Similarly, the DCC(1,1)-GARCH(1,1) model in Engle (2002) and Engle and Sheppard (2001) is encompassed by (2.2) if we impose the restrictions

$$\text{(dynamic conditional correlations)} \quad R_t = (\mathrm{diag}\; Q_t)^{-1/2}\; Q_t \; (\mathrm{diag}\; Q_t)^{-1/2}, \text{ where}$$
$$Q_t = (1 - \phi - \lambda)\overline{Q} + \phi \varepsilon_{t-1} \varepsilon_{t-1}^T + \lambda Q_{t-1}, \; \phi, \lambda \geq 0, \phi + \lambda < 1;$$
$$\text{(GARCH(1,1) volatilities)} \quad \sigma_{t,i}^2 = \alpha_{0,i} + \alpha_{1,i}(X_{t-1,i} - \mu_{t-1,i})^2 + \beta_i \sigma_{t-1,i}^2,$$
$$\alpha_{0,i} > 0, \alpha_{1,i} \geq 0, \beta_i \geq 0, \alpha_{1,i} + \beta_i < 1, \; i = 1, \ldots, d. \tag{2.4}$$

In this model, ε_t is a standardized error term, $\varepsilon_t^T = \big((X_{t,1} - \mu_{t,1})/\sigma_{t,1}, \ldots, (X_{t,d} - \mu_{t,d})/\sigma_{t,d}\big)$, and $\overline{Q}$ is the unconditional covariance matrix of the standardized residuals. In particular, conditional correlations are allowed to change over time. However, such dynamics must satisfy strong restrictions to ensure positivity of the conditional covariance matrix and computational feasibility of the model.

We propose a class of models in (2.2) for dynamic correlations which allows to reach a good trade-off between parameter parsimony and flexibility. To this purpose, we introduce in the next Section the concept of *averaged conditional correlation* across all d assets.

2.2 Averaging conditional correlations

For a date t we define the "averaged conditional correlation" as a weighted sum of all elements in the conditional correlation matrix R_t. The time-varying weights are constructed as follows. Let $\Delta_t = \frac{1}{d}\sum_{i=1}^{d} X_{t,i}$ be the equally weighted portfolio returns on day t constructed from the d individual assets.[3] Then, the conditional variance of the portfolio return can be computed as

$$\sigma_{t,P}^2 = \mathrm{Var}(\Delta_t|\mathcal{F}_{t-1}) = \frac{1}{d^2}\sum_{i=1}^{d}\sum_{j=1}^{d}\sigma_{t,i}\sigma_{t,j}\rho_{t,ij}. \qquad (2.5)$$

Consider now the particular case where all assets are perfectly correlated, i.e. $\rho_{t,ij} = \rho_{ij} \equiv 1$, for all $i, j = 1, \dots, d$. In this case, the portfolio conditional variance is

$$(\sigma_{t,P}^2)' = \mathrm{Var}(\Delta_t|\mathcal{F}_{t-1}) = \frac{1}{d^2}\big(\sum_{i=1}^{d}\sigma_{t,i}\big)^2. \qquad (2.6)$$

The averaged conditional correlation is constructed as the quotient of the portfolio conditional variance (2.5) in the general case and the portfolio conditional variance (2.6) in the case of perfect correlation among all assets:

$$\overline{\rho}_t = \sigma_{t,P}^2/(\sigma_{t,P}^2)' = \sum_{i=1}^{d}\sum_{j=1}^{d} w_{t,ij}\rho_{t,ij}, \qquad (2.7)$$

with weights given by $w_{t,ij} = (\sigma_{t,i}\sigma_{t,j})/(\sum_{k=1}^{d}\sigma_{t,k})^2$. Note that by construction we have that $\sum_{i=1}^{d}\sum_{j=1}^{d} w_{t,ij} = 1$ and $0 < \overline{\rho}_t \leq 1$. As we will see in Section 3, simple estimates for the time-varying averaged conditional correlation can be easily computed from the univariate volatility estimates of each individual asset and from those for the equally weighted portfolio.

At this point, the averaged conditional correlations (2.7) can be used to model the dynamics of the conditional correlation matrix R_t in (2.2).

2.3 The RW-ACC model and the RW-TACC model

Analogously to the CCC-GARCH(1,1) model (2.3) and the DCC(1,1)-GARCH(1,1) model (2.4), we assume that the time-varying dynamics of the individual volatilities in (2.2) follow a GARCH(1,1) model

$$\sigma_{t,i}^2 = \alpha_{0,i} + \alpha_{1,i}(X_{t-1,i} - \mu_{t-1,i})^2 + \beta_i \sigma_{t-1,i}^2, \text{ where}$$
$$\alpha_{0,i} > 0, \alpha_{1,i} \geq 0, \beta_i \geq 0, \alpha_{1,i} + \beta_i < 1, \; i = 1, \ldots, d. \tag{2.8}$$

The conditional correlations in (2.2) can have one of the following two forms:

$$R_t = (1-\lambda)\overline{Q}_{t-p}^{t-1} + \lambda \overline{R}_t, \quad \lambda \in [0,1[; \tag{2.9}$$
$$R_t = \Big(1 - \sum_{k=1}^{N} \lambda_k I_{[(\overline{\rho}_{t-1}, \mathbf{X}_{t-1}) \in \mathcal{R}_k]}\Big)\overline{Q}_{t-p}^{t-1} + \Big(\sum_{k=1}^{N} \lambda_k I_{[(\overline{\rho}_{t-1}, \mathbf{X}_{t-1}) \in \mathcal{R}_k]}\Big) I_d, \; \lambda_k \in [0,1] \forall k, \tag{2.10}$$

where I_d is a rank d identity matrix, $I_{[\cdot]}$ is the indicator function, $\overline{Q}_{t-p}^{t-1}$ is defined as the unconditional correlation matrix of the standardized residuals ε_t over the past p days similarly to (2.4), and $\overline{R}_t$ is a matrix with ones on the diagonal and all other elements equal to $\overline{r}_t = (d\overline{\rho}_t - 1)/(d-1) \leq 1$, with $\overline{\rho}_t$ defined in (2.7) (note that this particular choice of the off-diagonal elements of $\overline{R}_t$ is such that mean$(\overline{R}_t) = \overline{\rho}_t$). The model (2.9) is a convex combination of realized dynamic conditional correlations and averaged conditional correlations. Clearly, when the parameter λ is zero all weight is given to the historical term, meaning that the averaged conditional correlations are not able to improve the estimation. As we will see, this is not the case.

The model (2.10) for the dynamics of the conditional correlation matrix R_t is more structured, although it is still a convex combination of two terms. It can be seen as a model with different regimes. In such a model, the estimation of the optimal number and type of regimes involves a partition $\mathcal{P}$ of the predictor space $G =]0,1] \times \mathbb{R}^d$ of $(\overline{\rho}_{t-1}, \mathbf{X}_{t-1})^T$:

$$\mathcal{P} = \{\mathcal{R}_1, \ldots \mathcal{R}_N\}, \; G = \cup_{k=1}^{N} \mathcal{R}_k, \; \mathcal{R}_k \cap \mathcal{R}_h = \emptyset \; (k \neq h). \tag{2.11}$$

The construction of the optimal partition is based exactly on the tree-structured AR-GARCH methodology recently proposed by Audrino and Bühlmann (2001) and generalized by Audrino and Trojani (2003). Such methodology is applied to the univariate time series of averaged conditional correlations $\overline{\rho}_t$ defined in (2.7). In such a model, the partition $\mathcal{P}$ is constructed on a binary tree where every terminal node represents a rectangular partition cell $\mathcal{R}_k$ with edges determined by thresholds for the predictor variables $(\overline{\rho}_{t-1}, \mathbf{X}_{t-1})^T$. Given a partition cell $\mathcal{R}_k$, the dynamics of $\overline{\rho}_t$ on this cell are described by a local AR-GARCH model. Note that regimes for the conditional correlations are in this case determined by multivariate thresholds. Consequently, tail-dependence effects already in the next period can be described by our model.

In general, the optimal number N of partition cells is small, i.e. $N \leq 4$, keeping the complexity of the model (2.10) reasonable. When $N = 1$, clearly we have no partition of the predictor space. If all the parameters are equal to one, the data are uncorrelated. Otherwise, some regime-dependent weight is also given to the historical term, based on the information derived from the analysis of the averaged conditional correlation series.

The models proposed in (2.8), (2.9) or (2.10) are very simple, involving only a small number of parameters, and are computationally feasible also in large dimensions d. Analogously to the DCC(1,1)-GARCH(1,1) model (2.4), they preserve the ease of estimation of the CCC-GARCH(1,1) model (2.3) yet allowing correlations to change over time. We call the model (2.8)-(2.9) rolling window, averaged conditional correlation (RW-ACC) GARCH(1,1) model and the model (2.8)-(2.10) rolling window, tree-structured averaged conditional correlation (RW-TACC) GARCH(1,1) model.

The following proposition gives us the sufficient conditions to guarantee positive definiteness of the conditional covariance matrix V_t for both the RW-ACC-GARCH(1,1) model and the RW-TACC-GARCH(1,1) model.

Proposition 1. *(Positive Definiteness)*
Let the univariate GARCH(1,1) parameter restrictions given in (2.8) be satisfied for all asset series $i = 1, \ldots, d$, and let the parameters involved in the conditional correlation dynamics satisfy the restrictions given in (2.9) and (2.10), respectively. Then:

i) the conditional covariance matrix V_t in the RW-ACC-GARCH(1,1) model is positive definite for all t, if in addition the averaged conditional correlation $\overline{\rho}_t$ in (2.7) satisfies $\overline{\rho}_t \geq 1/d \ \forall t$;

ii) the conditional covariance matrix V_t in the RW-TACC-GARCH(1,1) model is positive definite for all t.

Proof. To ensure positivity of the matrix V_t in the general setting (2.2) we have to ensure that the individual volatilities are strictly positive for each asset and that the conditional correlation matrix R_t is positive definite for all t. From the parameter restrictions in (2.8), we have that each individual conditional variance $\sigma^2_{t,i}$ is strictly positive since $\alpha_{0,i} > 0$. Moreover:

i) from the restriction $\overline{\rho}_t \geq 1/d$ $\forall t$ it follows that $0 \leq \overline{r}_t \leq 1$, $\forall t$. Then, we can write the matrix $\overline{R}_t$ as a weighted average of a positive definite matrix I_d and a positive semi-definite $d \times d$ matrix C with all coefficients equal to one:

$$\overline{R}_t = (1 - \overline{r}_t)I_d + \overline{r}_t C.$$

Consequently, R_t is positive definite for all t as it is a weighted average of a positive definite matrix $\overline{Q}^{t-1}_{t-p}$ and a positive (semi-) definite matrix $\overline{R}_t$.

ii) R_t is positive definite for all t as it is a weighted average of two positive definite matrices $\overline{Q}^{t-1}_{t-p}$ and I_d.

□

The restriction on the parameters in Proposition 1 are not necessary, but only sufficient to guarantee positive definiteness for V_t. The additional restriction on the averaged conditional correlations $\overline{\rho}_t \geq 1/d$ is satisfied in most of the practical applications, in particular when the dimension d of the problem is high.

3 The estimation procedure

We describe in this Section the procedure which is applied to estimate the multivariate GARCH models introduced in the last Section.

The parameters $\phi = (a_{0,i}, a_{1,i}, \alpha_{0,i}, \alpha_{1,i}, \beta_i,\ i = 1, \cdots, d)$ and the parameter(s) $\psi = \lambda$ or $\psi = (\lambda_k,\ k = 1, \ldots, N)$ of the RW-ACC-GARCH(1,1) model (2.8)-(2.9) and of the RW-TACC-GARCH(1,1) model (2.8)-(2.10), respectively, can be estimated with the pseudo maximum likelihood method. To this purpose, we assume the innovations $\mathbf{Z}_t$ in (2.2) to be multivariate standard normally distributed. The quasi log-likelihood (conditional on the first observation) in the general setting (2.2) is then given by

$$\begin{aligned} l(\phi, \psi; \mathbf{X}^n_2) &= \sum_{t=1}^{n} \log \Big((2\pi)^{-d/2} \det(V_t)^{-1/2} \exp\big(-(\mathbf{X}_t - \mu_t)^T V_t^{-1} (\mathbf{X}_t - \mu_t)/2 \big) \Big) \\ &= -\frac{1}{2} \sum_{t=1}^{n} \big(d \log(2\pi) + 2\log(\det(D_t)) + \log(\det(R_t)) + \varepsilon_t^T R_t^{-1} \varepsilon_t \big)\,, \end{aligned} \tag{3.1}$$

where $\varepsilon_t = D_t^{-1}(\mathbf{X}_t - \mu_t)$ as before. Our class of models, similarly to the CCC and DCC ones, was designed to allow for a two-stage estimation. In the first stage univariate GARCH(1,1) models are estimated for each series. In the second stage, residuals, standardized using the volatilities estimated in the first stage, are used to estimate the parameter(s) ψ of the dynamic correlation structure. The likelihood of the first stage is computed by replacing the conditional correlation matrix R_t for all t with the constant $d \times d$ identity matrix I_d. The resulting first stage quasi log-likelihood from (3.1) is

$$\begin{aligned} l_1(\phi; \mathbf{X}_2^n) &= -\frac{1}{2}\sum_{t=1}^{n}\left(d\log(2\pi) + 2\log(\det(D_t)) + \log(\det(I_d)) + \varepsilon_t^T I_d^{-1}\varepsilon_t\right) \\ &= -\frac{1}{2}\sum_{t=1}^{n}\left(d\log(2\pi) + \sum_{i=1}^{d}\left(\log(\sigma_{t,i}^2) + \varepsilon_{t,i}^2\right)\right) \\ &= -\frac{1}{2}\sum_{i=1}^{d}\left(n\log(2\pi) + \sum_{t=1}^{n}\left(\log(\sigma_{t,i}^2) + \varepsilon_{t,i}^2\right)\right). \end{aligned} \tag{3.2}$$

Note that (3.2) is simply the sum of the log-likelihoods of individual AR(1)-GARCH(1,1) models for each asset.

Before performing the second stage, we have to construct an estimate for the averaged conditional correlation $\overline{\rho}_t$ defined in (2.7). This can be easily achieved from the first-stage estimates for the individual volatilities $\widehat{\sigma}_{t,i},\ i = 1, \ldots, d$ and estimating univariate AR(1)-GARCH(1,1) volatilities $\widehat{\sigma}_{t,P}$ of the equally weighted portfolio Δ_t, based on the parameters $\phi_P = (a_{0,P}, a_{1,P}, \alpha_{0,P}, \alpha_{1,P}, \beta_P)$. The averaged conditional correlation estimates can then be constructed as

$$\widehat{\overline{\rho}}_t = \widehat{\sigma}_{t,P}^2 / \Big(\sum_{i=1}^{d}\widehat{\sigma}_{t,i}\Big)^2. \tag{3.3}$$

Based on the estimates (3.3) we can construct the optimal partition $\widehat{\mathcal{P}}$ (2.11) necessary for the second stage estimation of our RW-TACC-GARCH(1,1) model.[4]

The second-stage parameters for the conditonal correlations dynamics are estimated using correctly specified likelihood from (3.1), conditioning on first-stage parameters $\widehat{\phi}$, $\widehat{\phi}_P$ and $\widehat{\mathcal{P}}$

$$l_2\left(\psi; \mathbf{X}_2^n, \widehat{\phi}, \widehat{\phi}_P, \widehat{\mathcal{P}}\right) = -\frac{1}{2}\sum_{t=1}^{n}\left(d\log(2\pi) + 2\log\left(\det\big(\widehat{D}_t\big)\right) + \log(\det(R_t)) + \widehat{\varepsilon}_t^T R_t^{-1}\widehat{\varepsilon}_t\right). \tag{3.4}$$

Note that the only portion of the second stage likelihood (3.4) that will influence the parameter selection for ψ is $\log(\det(R_t)) + \widehat{\varepsilon}_t^T R_t^{-1}\widehat{\varepsilon}_t$.

Consistency and asymptotic normality of our two-step estimates $(\widehat{\phi}, \widehat{\psi})$ can be derived in the usual way under standard regularity conditions for the validity of the quasi-likelihood functions

(3.2)-(3.4); cf. Newey and McFadden (1994) and Engle and Sheppard (2001). Efficient estimates can be obtained under the same regularity conditions by applying one step of a Newton-Raphson estimation of the full likelihood (3.1) using as starting parameters the two-step estimates; for all details, see Pagan (1986). However, note that the computation of these estimates can be computationally expensive when dealing with large dimensions d.

4 Empirical tests

This Section presents the results of our estimations of RW-(T)ACC-GARCH(1,1) for a six-dimensional exchange-rate return time series. Our models are estimated with a rolling-window of about one year of daily data, i.e. $p = 265$ in (2.9) and (2.10).

We always compare the in-sample and out-of-sample performance of our models to those from (i) the classical CCC-GARCH(1,1) model (2.3) and (ii) the DCC(1,1)-GARCH(1,1) model (2.4). The second comparison is particular useful, because it highlights the exact contribution of our models relatively to a "benchmark" model allowing for time-varying conditional correlations.

4.1 Data

We consider a six-dimensional multivariate time series of daily log-returns for the following exchange rates against the U.S. Dollar: the British Pound USD/GBP, the German Deutschmark USD/DEM, the Japanese Yen USD/JPY, the Italian Lira USD/ITL, the French Franc USD/FRF and the Dutch Pound USD/NLG. The data span the time-period between January 2, 1992 and September 13, 1999, for a total of 1994 observations, and have been downloaded from the Olsen&Associates Database. We split our sample in a back-testing period used to test the predictive accuracy of our models and an in-sample estimation period used to initialize the model parameter estimates. The back-testing period goes from October 15, 1997 to September 13, 1999, for a total of 494 trading days. Summary statistics of in-sample daily returns for the above exchange rates and the corresponding equally weighted portfolio Δ_t are presented in Table 1.

TABLE 1 ABOUT HERE.

Sample means for the different exchange rates are very similar. The USD/JPY exchange rate shows a negative mean return that is attributable to a strong Japanese Yen during the considered in-sample period. The sample standard deviation exhibited by all exchange rates are similar. As expected, the sample standard deviation is reduced by constructing the equally weighted portfolio. The Ljung-Box statistics LB(10) testing for autocorrelations in the level of returns up to the 10^{th} order show in all cases except for the USD/GBP exchange rates no significant presence of autocorrelation in daily exchange rate returns. The |LB(10)| statistics for testing the null hypothesis of dependency of the absolute exchange rate and portfolio returns are all highly significant. The USD/DEM, USD/FRF and USD/NLG exchange rate returns exhibit the highest sample correlations with each other, indicating a strong dependence structure among the exchange rates of these markets, whereas the lowest correlations are those with the USD/JPY exchange rate returns.

4.2 Estimation of the models

This Section presents the estimated multivariate RW-ACC and RW-TACC models for the exchange rate data example under scrutiny. Estimated parameters from the two-stage procedure described in Section 3 for the RW-ACC- and RW-TACC-GARCH(1,1) models are summarized in Table 2. Standard errors are computed using the sub-sampling model-based bootstrap methodology (see Freedman, 1984, or Efron and Tibshirani, 1993). Figure 1 plots the corresponding estimated averaged conditional correlation series (3.3) in our in-sample period.

TABLE 2 AND FIGURE 1 ABOUT HERE.

As we expect from Table 1, Table 2 shows that all α_1's and β's parameters in the individual GARCH(1,1) models are highly significant. Moreover, no significant parameter is found for the conditional mean functions. The sum $\alpha_1 + \beta$ is for all series near to one, implying strong persistence in the conditional variances. The dynamic behavior of the averaged conditional correlations is well illustrated in Figure 1. A constant conditional correlations hypothesis is clearly rejected based on the averaged correlation series. We can identify at least three different short time-periods with estimated averaged correlations outside a classical two standard deviation confidence interval implied by the constant conditional correlations hypothesis.

The estimated parameters for the conditional correlations are in the most cases significantly different from zero, although they are mostly around zero. This implies that most weight in the conditional correlations dynamics (2.9) and (2.10) is given to the historical term $\overline{Q}_{t-p}^{t-1}$. However, the results in Table 2 and Figure 1 show that information coming from averaged conditional correlations is important and can not be neglected in the model specification.

In our particular example, Table 2 shows that past values of the USD/JPY individual return series completely characterize the regime structure of averaged conditional correlations and consequently the type of regimes for the conditional correlations in the RW-TACC model. In particular, we found three different regimes: the first one characterized by high negative (i.e. smaller than the estimated threshold) past USD/JPY returns, a second one by bounded past USD/JPY returns and a third one by high positive past USD/JPY returns.

Figure 2 shows our sample period conditional correlation dynamics for two representative examples estimated using the RW-TACC-GARCH(1,1) model, the DCC(1,1)-GARCH(1,1) model and the CCC-GARCH(1,1) model.

FIGURE 2 ABOUT HERE.

The constant conditional correlations approach yields clearly only a rough approximation of the conditional correlations dynamics, in particular for our back-testing period (last 494 days). Our RW-TACC model yields conditional correlation estimates and predictions which change more slowly and exhibit more small scale fluctuations than those from a DCC(1,1) approach. As we will see in the next Sections, results of multivariate and univariate performance tests favor this behavior of conditional correlations.

4.3 Standardized residuals

We also analyze the goodness of the standardized residuals estimated using the different multivariate models introduced so far. The comparison is performed using the same two goodness of fit criteria already proposed by Engle and Sheppard (2001).

Consider the standardized residuals $\mathbf{Z}_t = \Sigma_t^{-1}(\mathbf{X}_t - \mu_t)$ in (2.2). From assumption (A1) they have constant conditional covariance matrix equal the identity. Moreover, cross products $\mathbf{Z}_t\mathbf{Z}_t^T$ are uncorrelated over time. It is therefore natural to test whether (i) the multivariate standardized residual estimated with the different models have unit variance and (ii) the estimated cross products are uncorrelated over time.

The first criteria are the percentage of multivariate standardized residuals which have variance in a confidence interval of one. The second criteria are the percentages of rejected classical Ljung-Box tests investigating whether there is excess serial correlation in the squares and cross products of standardized residuals up to the 15^{th} lag at a confidence level of 5%. The results of such tests on the in-sample standardized residuals estimated using the different multivariate models proposed in the paper are summarized in Table 3.

TABLE 3 ABOUT HERE.

All the models considered perform well with respect to the percentage of standardized residuals with conditional variance in a confidence interval of one. Models with time-varying conditional correlations perform better than models with constant conditional correlations with respect to the percentage of failing Ljung-Box tests. More than 20% of the CCC models standardized residuals fail the test. In contrast, when allowing for dynamic conditional correlations the percentage of failures is substantially reduced. Similarly to Engle and Sheppard (2001), we find that the percentage failing is always greater than the 5% which would have been expected.

4.4 Multivariate performance results

To measure and compare precision of the conditional covariance matrix estimates and forecasts from the different models we use several in-sample and out-of-sample statistics. The multivariate negative log-likelihood statistics (NL), a multivariate version of the classical mean absolute error (MAE), a multivariate version of the root mean squared error (RMSE) and the mean of absolute empirical correlations (R^2) between actual values and one-step ahead predicted values of the conditional covariance, averaged over all possible components. More specifically, the following statistics are used (where IS and OS denote in-sample and out-of-sample, respectively):

$$\text{IS-NL: } -2 \text{ log-likelihood } (3.1)$$

$$\text{OS-NL: } -\text{log-likelihood}\Big(\widetilde{\mathbf{X}}_1^{n_{\text{out}}}; \widehat{\phi}, \widehat{\psi}\Big)$$

$$\text{IS-MAE: } \frac{1}{d^2}\sum_{i,j=1}^{d}\frac{1}{n}\sum_{t=1}^{n} \mid \widehat{v}_{t,ij} - (X_{t,i} - \widehat{\mu}_{t,i})(X_{t,j} - \widehat{\mu}_{t,j}) \mid$$

$$\text{OS-MAE: } \frac{1}{d^2}\sum_{i,j=1}^{d}\frac{1}{n_{\text{out}}}\sum_{t=1}^{n_{\text{out}}} \mid \widehat{v}_{t,ij} - (\widetilde{X}_{t,i} - \widehat{\mu}_{t,i})(\widetilde{X}_{t,j} - \widehat{\mu}_{t,j}) \mid$$

$$\text{IS-RMSE: } \Big(\frac{1}{d^2}\sum_{i,j=1}^{d}\frac{1}{n}\sum_{t=1}^{n} \mid \widehat{v}_{t,ij} - (X_{t,i} - \widehat{\mu}_{t,i})(X_{t,j} - \widehat{\mu}_{t,j}) \mid^2 \Big)^{1/2}$$

$$\text{OS-RMSE: } \Big(\frac{1}{d^2}\sum_{i,j=1}^{d}\frac{1}{n_{\text{out}}}\sum_{t=1}^{n_{\text{out}}} \mid \widehat{v}_{t,ij} - (\widetilde{X}_{t,i} - \widehat{\mu}_{t,i})(\widetilde{X}_{t,j} - \widehat{\mu}_{t,j}) \mid^2 \Big)^{1/2}$$

$$\text{IS-R}^2\text{: } \frac{1}{d^2}\sum_{i,j=1}^{d} \mid \text{Cor}\Big(\widehat{v}_{t,ij}, (X_{t,i} - \widehat{\mu}_{t,i})(X_{t,j} - \widehat{\mu}_{t,j})\Big) \mid$$

$$\text{OS-R}^2\text{: } \frac{1}{d^2}\sum_{i,j=1}^{d} \mid \text{Cor}\Big(\widehat{v}_{t,ij}, (\widetilde{X}_{t,i} - \widehat{\mu}_{t,i})(\widetilde{X}_{t,j} - \widehat{\mu}_{t,j})\Big) \mid,$$

where $\widetilde{\mathbf{X}}_1^{n_{\text{out}}} = \mathbf{X}_{n+1}, \ldots, \mathbf{X}_{n+n_{\text{out}}}$ are the test data and the parameter estimates equipped with hats have been constructed from the training sample $\mathbf{X}_1^n = \mathbf{X}_1, \ldots \mathbf{X}_n$. Clearly we see the OS statistics as the most important ones to judge the predictive potential of the different models.

The goodness of fit results of the different models are summarized in Table 4. Note that "low is better" for all goodness of fit statistics except for the R^2 measures.

TABLE 4 ABOUT HERE.

The optimal values with respect to the different statistics are reached in-sample by the DCC(1,1) model with respect to all performance measures. In contrast, when focusing on the most important out-of-sample statistics, we see that the optimal values are reached by the RW-(T)ACC-FGD models. As expected, the CCC model is clearly beaten by models allowing for dynamic conditional correlations with respect to most of the out-of-sample statistics. Moreover, we observe over-fitting problems when using the DCC(1,1) model: it reaches the optimal values in-sample, but it does not seem to be as good as the RW-(T)ACC models for prediction.

Table 4 shows that differences between the models are in general small, except for the multivariate NL statistics. Such small differences with respect of goodness of fit measures like

MAE or RMSE could be obscured by a low signal to noise ratio when replacing the unobservable conditional covariances by their corresponding actual return values which are noisy estimates. It is well known that in real data examples the noise component is often dominant and differences in the conditional covariance estimates may be masked. Thus, such criteria typically allows only to discriminate between forecasts whose performance is different by orders of magnitude.

One possible solution to avoid this problem is to construct estimates for the actual unobserved conditional covariances which are less noisy, for example by using the integrated volatility approach (see, among others, Andersen et al., 1999 or 2001). An alternative is to consider differences of performance terms and to use the concept of hypothesis testing. This is our approach in this Section and in Section 4.5.

We consider differences of each term in the OS-NL statistic[5],

$$\widehat{D}_{t,ij} = \widetilde{U}_{t;\text{model}_i} - \widetilde{U}_{t;\text{model}_j},\ t = 1,\ldots,n_{\text{out}},\ i,j = 1,\ldots,4, i < j,$$

where

$$\sum_{t=1}^{n_{\text{out}}} \widetilde{U}_{t;\text{model}} = \text{OS-NL}.$$

Moreover, similarly to Audrino and Bühlmann (2003), we also consider the "direction" of the differences of each term in the OS-NL statistic

$$\widehat{W}_{t,ij} = \begin{cases} 1\ , & \text{if } \widehat{D}_{t,ij} \leq 0 \\ -1, & \text{else} \end{cases}\ ,\ t = 1,\ldots,n_{\text{out}},\ i,j = 1,\ldots,4, i < j.$$

These type of tests allow us to investigate whether there is a systematic difference between the estimates from the models. We denote the first and second class of tests as t-type and sign-type tests, respectively. Since it is difficult to identify a single model as the "best" model due to the fact that returns do not contain generally sufficient information, we apply the Model Confidence Set (MCS) method proposed by Hansen et al. (2003) to characterize the multivariate GARCH models that significantly dominate others.

The MCS is determined after sequentially trimming the set of candidate models (in our case the CCC, DCC(1,1), RW-ACC and RW-TACC specifications). At each step, the null-hypothesis of equal predictive ability (EPA) $\mathcal{H}_0 : \mathbb{E}[D_{t,ij}] = 0,\ \forall i,j \in \mathcal{M}$ (respectively $\mathcal{H}_0 : \mathbb{E}[W_{t,ij}] = 0$) is tested for a set of models $\mathcal{M}$. The first test is for the full set of candidate models. If $\mathcal{H}_0$ is rejected, the worst performing model is eliminated from $\mathcal{M}$. This trimming is repeated until the first non-rejection occurs, an the set of surviving models is the model confidence set $\widehat{\mathcal{M}}_\alpha$, for a fixed confidence level α.

Our tests of EPA employ the range statistic T_R and the less conservative semi-quadratic statistic T_{SQ}

$$T_R = \max_{i,j\in\mathcal{M}} \frac{|\overline{D}_{ij}|}{\sqrt{\widehat{\text{var}}(\overline{D}_{ij})}} \quad \text{and} \quad T_{SQ} = \sum_{i<j} \frac{\overline{D}_{ij}^2}{\widehat{\text{var}}(\overline{D}_{ij})},$$

where the sum is taken over the models in $\mathcal{M}$, $\overline{D}_{ij} = 1/n_{\text{out}} \sum_{t=1}^{n_{\text{out}}} \widehat{D}_{t,ij}$, and $\widehat{\text{var}}(\overline{D}_{ij})$ is an estimate of $\text{var}(\overline{D}_{ij})$ that is obtained from a block-bootstrap implementation of the series $\widehat{D}_{t,ij}$, $t = 1, \ldots, n_{\text{out}}$. Estimates of the (asymptotic) distributions of T_R and T_{SQ} to test for EPA under the null hypothesis can also be consistently derived under mild regularity conditions from the bootstrap. For more details and for a complete description of the procedure we remand the reader to Hansen et al. (2003).

Results of the t-type and sign-type tests introduced above as well as resulting model confidence sets for the real data example under investigation are summarized in Tables 5 and 6.

TABLES 5 AND 6 ABOUT HERE.

Table 5 clearly shows that, as expected, models allowing for dynamic conditional correlations are preferred by t-type tests to the classical CCC-GARCH(1,1) model. Both 95% and 90% confidence sets $\widehat{\mathcal{M}}_{0.05}$ and $\widehat{\mathcal{M}}_{0.1}$ consist in fact only of dynamic conditional correlation models. Moreover, significant differences can also be seen when considering differences of OS-NL performance terms between models allowing for dynamic conditional correlations. The proposed RW-(T)ACC models belong in the most cases to the 95% and 90% confidence sets with respect to both the range and semi-quadratic statistics. On the contrary, the DCC(1,1) model is always eliminated from the model confidence set, except for the more conservative range statistic at the 95% confidence level.

The results in Table 6 are even more significant. This finding may be just a fact of a low power of the t-test due to non-Gaussian observations. On the other hand, the sign-type tests are robust against deviations from Gaussianity. Dynamic conditional correlations models are better than the standard CCC-GARCH(1,1) model, and our RW-(T)ACC-GARCH(1,1) models are better than the DCC(1,1)-GARCH(1,1) model. Moreover, predictions estimated using the RW-TACC-GARCH(1,1) model show significant advantages also over the ones from the RW-ACC-GARCH(1,1) model. The model confidence set at both the 95% and 99% confidence levels consists only of the RW-TACC-GARCH(1,1) model, which is clearly the best model for prediction purposes with respect to the sign-type test.

4.5 Portfolio performance results

We test in this Section the accuracy of volatility estimates and predictions for the equally weighted portfolio Δ_t constructed on the six-dimensional exchange-rate data introduced in Section 4.1. To measure and compare goodness of fit from the different models we use standard univariate versions of the in-sample and out-of-sample MAE, RMSE and R^2 measures introduced in Section 4.4. Results are summarized in Table 7.

TABLE 7 ABOUT HERE.

Table 7 shows similar results to those found at the multivariate level and summarized in Table 4. The RW-(T)ACC models yield more accurate volatility predictions than both standard CCC and DCC(1,1) models. Similarly to Table 4, differences between the models are in general small. Hence, we consider differences of each term in the OS-MAE^6 statistic and use again the concept of EPA hypothesis testing to construct model confidence sets. Results based on t-type tests for the equally weighted portfolio Δ_t are summarized in Table 8.

TABLE 8 ABOUT HERE.

As expected, multivariate GARCH(1,1) models with time-varying conditional correlations are preferred to the classical CCC-GARCH(1,1) approach and belong to the 95% model confidence set with respect to both range and semi-quadratic statistics. Moreover, from Table 8 the RW-(T)ACC-GARCH(1,1) models belong also to the 90% model confidence set with respect to the range statistic. On the contrary the DCC(1,1)-GARCH(1,1) model is eliminated from the 90% model confidence sets. With respect to the less conservative EPA hypothesis based on the semi-quadratic statistic, the 90% model confidence set consists only of the RW-TACC-GARCH(1,1) model. All these results support the already collected empirical evidence at the multivariate level of the better predictive power of our multivariate models over a DCC(1,1)-GARCH(1,1) approach. Results of statistical tests for the EPA hypothesis based on sign-type differences of performance terms are similar and therefore not reported here.

4.6 A practical application: Value-at-Risk computation

As a practical application, we investigate the forecasting power of volatility predictions from the different models in computing 1-day ahead Value-at-Risk (VaR) estimates for the univariate equally weighted portfolio Δ_t at the 5% and 1% confidence levels.

To construct daily VaR estimates for the equally weighted portfolio, we use the same strategy recently proposed by Ledoit et al. (2003). Once that portfolio conditional means and volatilities are estimated (using the different multivariate approaches), portfolio standardized residuals $(\Delta_t - \widehat{\mu}_{t,P})/\widehat{\sigma}_{t,P}$ are fitted using a univariate scaled t_ξ distribution in order to allow for fat tails. The optimal degrees of freedom parameter $\widehat{\xi}$ is estimated by maximum likelihood. The 1-day VaR estimates for our back-testing period at the confidence level x are then given by

$$\widehat{\mathrm{VaR}}_t = \widehat{\mu}_{t,P} + \widehat{\sigma}_{t,P}\sqrt{(\widehat{\xi}-2)/\widehat{\xi}}\, t_{\widehat{\xi};x}, \; t = 1, \ldots, n_{\mathrm{out}}, \tag{4.1}$$

where $t_{\widehat{\xi};x}$ denotes the x-quantile of the standard $t_{\widehat{\xi}}$ distribution.

Let

$$\mathrm{Hit}_t = I_{\{\Delta_t < \widehat{\mathrm{VaR}}_t\}}, \; t = 1, \ldots, n_{\mathrm{out}}, \tag{4.2}$$

be the sequence of hit indicator variables. If the model is correctly specified, the hit series should be uncorrelated over time and have expected value equal to the desired confidence level.

Results of classical binomial tests on the number of hits and of Ljung-Box tests for autocorrelation in the hit sequence up to the 12^{th} order for the 5% and 1% confidence levels are summarized in Table 9. Our back-testing period goes from October 15, 1997, to September 13, 1999, for a total of 494 trading days. Asterisks denote significance at the 5% confidence level or better.

TABLE 9 ABOUT HERE.

The hit rates are all reasonably close to the target levels, although they tend to be larger, with the only exception of the DCC(1,1)-GARCH(1,1) model at the 1% coverage rate, which is the only loser in terms of hit rates. Differences between the models with respect to the Ljung-Box tests are also small. The only rejection at the 5% confidence level was recorded by the CCC-GARCH(1,1) model at the 5% coverage rate.

5 Conclusions

We proposed a simple class of semiparametric multivariate GARCH models. Our models are more flexible and accurate for the estimation and prediction of conditional variance-covariance matrices than two popular alternative multivariate GARCH models, namely the CCC and the

DCC models. Analogously to the DCC-GARCH model proposed by Engle (2002), our multivariate GARCH models preserve the ease of estimation of Bollerslev's CCC-GARCH model while allowing for possible asymmetric non-linear individual volatilities and time-varying conditional correlations.

Our models can be easily estimated using a classical two-stage procedure. Non-parametric estimates for the individual volatility functions can be easily constructed using the Functional Gradient Descent (FGD) technique introduced in Audrino and Bühlmann (2003).

Testing the models on real exchange-rate data we collect empirical evidence of the strong forecasting power of our multivariate GARCH models with respect to various goodness-of-fit criteria and statistical tests for the equal predictive ability hypothesis. In particular, we considered forecasting accuracy at the multivariate and at the portfolio, univariate level, persistence of multivariate standardized residuals and precision of portfolio Value-at-Risk estimates.

Notes

[1]See, for example, Audrino and Bühlmann (2004) for an application to the measurement of risk in global stock markets.

[2]The model for the conditional mean is kept very simple, because our primary focus is on the covariance matrix. Moreover, in the empirical investigations of Section 4 we found that all conditional mean parameters were not significantly different from zero.

[3]The choice of the equally weighted portfolio is not restrictive. One can also apply the same strategy for portfolios with non-equal weights. However, the explanation and computations are in the particular case of equal weights straightforward.

[4]The estimation of the optimal partition $\widehat{\mathcal{P}}$ (2.11) is performed by applying to the series (3.3) of estimated averaged conditional correlations the tree-structured AR(1)-GARCH(1,1) model and using the same methodology already introduced in Audrino and Bühlmann (2001).

[5]The choice of the OS-NL statistic is clearly not restrictive. However, we think that, at the multivariate level, OS-NL is the most interesting measure.

[6]We choose differences of OS-MAE terms because they are more robust and less affected by a few large outliers. Tests on difference of OS-MSE terms (defined as the square of the OS-RMSE statistic) yield similar results.

References

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association* **96**, 42-55.

Andersen, T.G., Bollerslev, T. and Lange, S. (1999). Forecasting financial market volatility: sample frequency vis-a-vis forecast horizon. *Journal of Empirical Finance* **6**, 457-477.

Alexander, C. (2001). A primer on the orthogonal GARCH model. ISMA Centre, Mimeo.

Alexander, C. and Chibumba, A. (1997). Multivariate orthogonal factor GARCH. University of Sussex, Mimeo.

Audrino, F. and Bühlmann, P. (2001). Tree-Structured GARCH Models. *Journal of the Royal Statistical Society, Series B*, **63**, No. 4, 727-744.

Audrino, F. and Bühlmann, P. (2003). Volatility estimation with functional gradient descent for very high-dimensional financial time series. *Journal of Computational Finance* **6**, No. 3, 65-89.

Audrino, F. and Bühlmann, P. (2004). Synchronizing multivariate financial time series. *Journal of Risk* **6**, No. 2, 81-106.

Audrino, F. and Trojani, F. (2003). Estimating and predicting multivariate volatility regimes in global stock markets. Manuscript USI, Lugano, Switzerland.

Baur, D. (2003). A flexible dynamic correlation model. Mimeo, Joint Research Center, Ispra, Italy.

Bauwens, L., Laurent, S. and Rombouts, J.V.K. (2003). Multivariate GARCH models: a survey. CORE Discussion Paper, Université catholique de Louvain.

Billio, M., Caporin, M. and Gobbo, M. (2003). Block Dynamic Conditional Correlation Multivariate GARCH models. Working Paper, Greta, Italy.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, **31**, 307-327.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *The Review of Economics and Statistics* **72**, 498-505.

Bollerslev, T., Engle R.F. and Wooldridge, J. (1988). A capital asset pricing model with time varying covariances. *Journal of Political Economy* **96**, 116-131.

Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap.* Chapman & Hall, London.

Engle, R.F. (2002). Dynamic conditional correlation - a simple class of multivariate GARCH models. *Journal of Business and Economic Statistics* **20**, 339-350.

Engle, R.F. and Kroner, K.F. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory* **11**, 122-150.

Engle, R.F., Ng, V.K. and Rothschild, M. (1990). Asset pricing with a factor ARCH covariance structure: empirical estimates for treasury bills. *Journal of Econometrics* **45**, 231-238.

Engle, R.F. and Sheppard, K. (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. Mimeo, University of California, San Diego.

Freedman, D. (1984). On bootstrapping two-stage least-squares estimates in stationary linear models. Annals of Statistics **12**, 827–842.

Hansen, P.R. and Lunde, A. (2002). A forecast comparison of volatility models: does anything beat a GARCH(1,1)? Manuscript, Brown University.

Hansen, P.R., Lunde, A. and Nason, J.M. (2003). Choosing the best volatility models: The model confidence set approach. *Oxford Bulletin of Economics and Statistics* **65**, 839-861.

Ledoit, O., Santa-Clara, P. and Wolf, M (2003). Flexible multivariate GARCH modeling with an application to international stock markets. Forthcoming in *The Review of Economics and Statistics.*

Lee, T. and Saltoglu, B. Evaluating the predictive performance of Value-at-Risk models in emerging markets: a reality check. Mimeo, University of California, Riverside.

Newey, W.K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics*, Vol. **4**, 2111-2245. North-Holland, Amsterdam.

Pagan, A. (1986). Two stage and related estimators and their applications. *The Review of Economic Studies* **53**, No. 4, 517-538.

Pelletier, D. (2002). Regime switching for dynamic correlations. Mimeo, Université de Montréal.

Sheppard, K. (2002). Understanding the dynamics of equity covariance. Manuscript, University of California, San Diego.

Tse, Y.K. (2000). A test for constant correlations in a multivariate GARCH model. *Journal of Econometrics* **98**, 107-127.

Tse, Y. and Tsui, A. (2002). A multivariate GARCH model with time-varying correlations. *Journal of Business and Economic Statistics* **20**, 351-362.

Tsui, A.K. and Yu, Q. (1999). Constant conditional correlation in a bivariate GARCH model: Evidence from the stock market in China. *Mathematics and Computers in Simulation* **48**, 503-509.

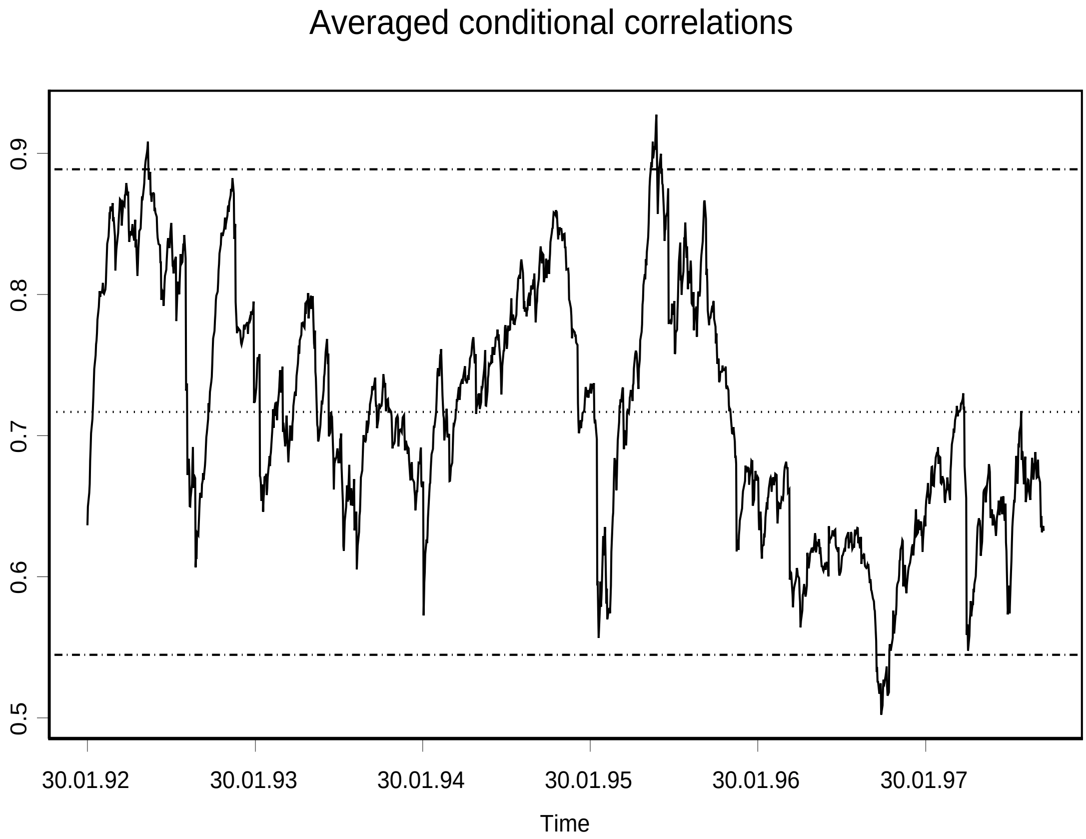

Figure 1: Estimated averaged conditional correlation series during the in-sample period between January 2, 1992 and October 14, 1997 for a total of 1500 trading days. The dotted line and the dashed lines indicate the mean of estimated averaged conditional correlations and a classical two standard deviations confidence interval for a constant mean averaged conditional correlation hypothesis, respectively.

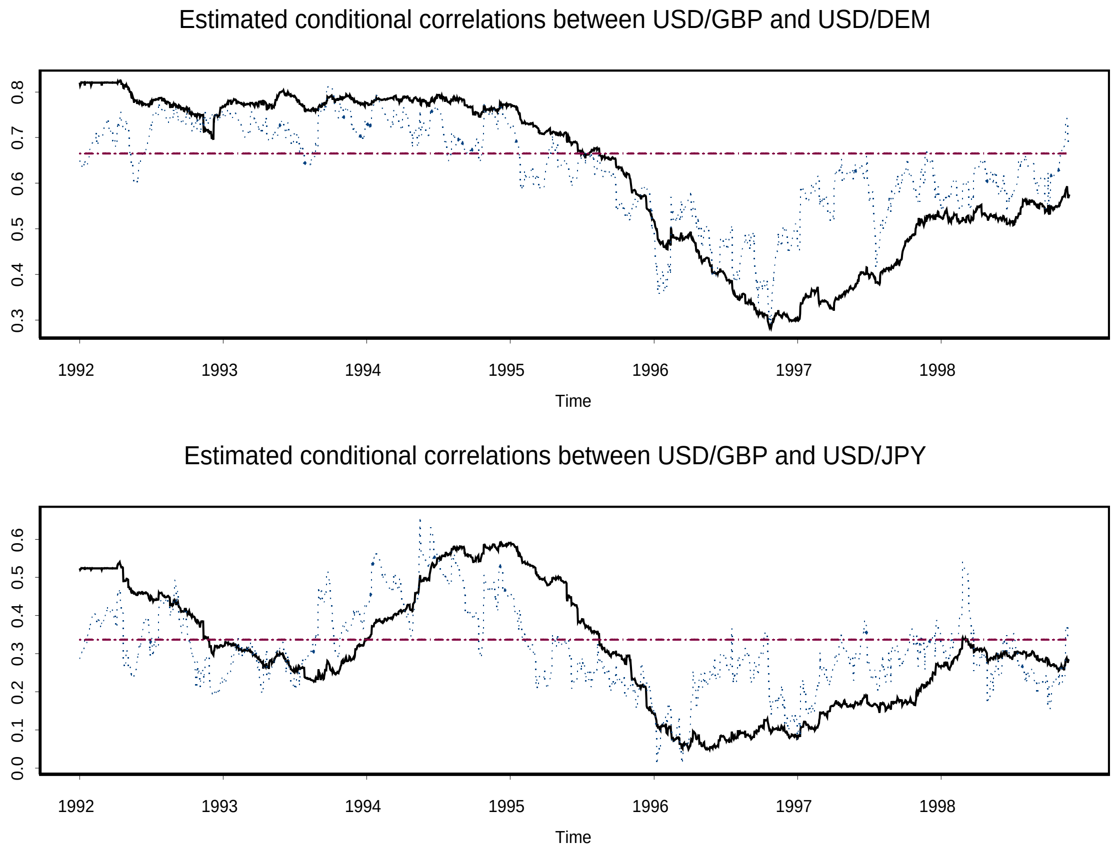

Figure 2: Conditional correlation dynamics between USD/GBP and USD/DEM (top) and between USD/GBP and USD/JPY (bottom) during the entire sample beginning January 2, 1992 and ending September 13, 1999. Conditional correlations are estimated using the RW-TACC-GARCH(1,1) model (solid line), the DCC(1,1)-GARCH(1,1) model (dotted line) and the CCC-GARCH(1,1) model (dashed line).

Exchange rate	Sample mean	Sample sdev	LB(10)	|LB(10)|
USD/GBP	0.0095	0.5947	24.615*	304.30*
USD/DEM	0.0096	0.6534	8.9746	177.80*
USD/JPY	-0.0016	0.6608	15.244	149.74*
USD/ITL	0.0267	0.6547	21.472	326.34*
USD/FRF	0.0084	0.6161	9.8358	190.94*
USD/NLG	0.0096	0.6492	10.086	183.64*
Eq. weighted Portf. Δ	0.0066	0.5295	10.167	136.09*

	USD/GBP	USD/DEM	USD/JPY	USD/ITL	USD/FRF	USD/NLG
USD/GBP	1	0.67207	0.32942	0.57703	0.67617	0.66925
USD/DEM	0.67207	1	0.54209	0.68017	0.95554	0.99377
USD/JPY	0.32942	0.54209	1	0.34134	0.50855	0.54195
USD/ITL	0.57703	0.68017	0.34134	1	0.72297	0.67890
USD/FRF	0.67617	0.95554	0.50855	0.72297	1	0.95360
USD/NLG	0.66925	0.99377	0.54195	0.67890	0.95360	1

Table 1: Summary statistics on log-returns of six exchange rates against the U.S. dollar and the corresponding equally weighted portfolio Δ_t for the time period between January 2, 1992 and October 14, 1997, for a total of 1500 in-sample observations. Sample sdev, LB(10) and |LB(10)| are the sample standard deviations and the Ljung-Box statistics testing for autocorrelation in the level of returns and the level of absolute returns, respectively, up to the 10^{th} lag. Asterisks indicate statistical significance at the 1% level or better. Instantaneous empirical correlations among the exchange rates are given in the second table.

Exchange rate	GARCH(1,1) parameters			AR(1) parameters	
	α_0	α_1	β	a_0	a_1
USD/GBP	0.0022	0.0359*	0.9570*	-0.0093	0.0129
	(0.0032)	(0.0124)	(0.0190)	(0.0135)	(0.0290)
USD/DEM	0.0054	0.0347*	0.9516*	0.0087	-0.0025
	(0.0065)	(0.0129)	(0.0248)	(0.0158)	(0.0266)
USD/JPY	0.0141	0.0639*	0.9034*	0.0049	0.0229
	(0.0097)	(0.0204)	(0.0363)	(0.0162)	(0.0283)
USD/ITL	0.0028	0.0596*	0.9350*	0.0007	-0.0032
	(0.0036)	(0.0177)	(0.0197)	(0.0134)	(0.0279)
USD/FRF	0.0037	0.0345*	0.9548*	0.0043	0.0060
	(0.0071)	(0.0118)	(0.0275)	(0.0152)	(0.0283)
USD/NLG	0.0060	0.0355*	0.9491*	0.0085	-0.0089
	(0.0089)	(0.0141)	(0.0332)	(0.0142)	(0.0278)
Eq. weighted Portf. Δ	0.0016	0.0311*	0.9633*	0.0059	-0.0027
	(0.0048)	(0.0107)	(0.0290)	(0.0127)	(0.0276)

Model	Cond. corr. structure	Cond. corr. parameters
	$\mathcal{R}_k$	λ_k
RW-ACC	–	0.0042*
		(0.0021)
RW-TACC	$X_{t-1,\text{usd/jpy}} \leq -0.6084$	0.0088*
		(0.0015)
	$-0.6084 \leq X_{t-1,\text{usd/jpy}} \leq 0.3486$	0.0013*
		(0.0005)
	$X_{t-1,\text{usd/jpy}} \geq 0.3486$	0.0005
		(0.0026)

Table 2: Estimated parameters of the RW-ACC- and the RW-TACC-GARCH(1,1) models from the two-stage procedure described in Section 3 for the six-dimensional real data example under scrutiny. Asterisks denote significance at the 5% level or better.

Model	% variance of stand. res. in CI	% Ljung-Box rejected
CCC-GARCH(1,1)	100 (6/6)	27.667 (10/36)
DCC(1,1)-GARCH(1,1)	100 (6/6)	16.667 (6/36)
RW-ACC-GARCH(1,1)	83.3 (5/6)	19.444 (7/36)
RW-TACC-GARCH(1,1)	83.3 (5/6)	19.444 (7/36)

Table 3: Multivariate tests on standardized residuals using different models. Percentages of in-sample multivariate standardized residuals having variance in a confidence interval of one and percentages of rejected classical Ljung-Box tests investigating whether there is excess serial correlation in the squares and cross products of standardized residuals up to the 15^{th} lag at a confidence level of 5%. Results are computed for our six-dimensional real data example.

Model	IS-				OS-			
	NL	MAE	RMSE	R^2	NL	MAE	RMSE	R^2
CCC-GARCH(1,1)	1583.4	0.3566	0.7465	0.0331	610.39	0.3149	0.5768	0.0088
DCC(1,1)-GARCH(1,1)	1266.4	0.3549	0.7381	0.0376	191.13	0.3079	0.5707	0.0066
RW-ACC-GARCH(1,1)	1551.8	0.3592	0.7397	0.0371	-50.354	0.3061	0.5702	0.0097
RW-TACC-GARCH(1,1)	1522.5	0.3589	0.7397	0.0372	-2.6861	0.3059	0.5702	0.0096

Table 4: Multivariate in-sample and out-of-sample goodness of fit results of the different models for our six-dimensional real data example. NL, MAE, RMSE and R^2 are multivariate versions of the standard univariate negative log-likelihood statistic, the mean absolute error, the root mean squared error and the R^2 statistics, respectively.

Model 1	Model 2	Statistic	
		range	semi-quadratic
DCC(1,1)-GARCH(1,1)	CCC-GARCH(1,1)	3.7272	13.892
RW-ACC-GARCH(1,1)	CCC-GARCH(1,1)	4.7137	22.219
RW-TACC-GARCH(1,1)	CCC-GARCH(1,1)	3.9684	15.748
RW-ACC-GARCH(1,1)	DCC(1,1)-GARCH(1,1)	1.9566	3.8284
RW-TACC-GARCH(1,1)	DCC(1,1)-GARCH(1,1)	1.4416	2.0781
RW-TACC-GARCH(1,1)	RW-ACC-GARCH(1,1)	1.4876	3.1955
Global value for EPA test:	1^{st} step	4.7137 (0.0010)	60.961 (0.0015)
Global value for EPA test:	2^{nd} step	1.9566 (0.0925)	9.1019 (0.0492)
Global value for EPA test:	3^{rd} step	1.4876 (0.1115)	3.1955 (0.0715)

Model	Worst performing index		
	1^{st} step	2^{nd} step	3^{rd} step
CCC-GARCH(1,1)	4.6281	–	–
DCC(1,1)-GARCH(1,1)	0.0707	1.6956	–
RW-ACC-GARCH(1,1)	-4.1429	-0.4056	-1.4876
RW-TACC-GARCH(1,1)	-3.3064	-0.9772	1.4876

Table 5: Testing differences of multivariate OS-NL performance terms among different multivariate GARCH models. Upper panel: Values of pairwise *t-type* test statistics, values of the range statistic T_R and the semi-quadratic statistic T_{SQ} for global test of equal predictive ability (EPA). Corresponding P-values are given between parentheses. Lower panel: Worst performing index results for the construction of the confidence model sets. If the null hypothesis of EPA is rejected, the model with the largest worst performing index value is eliminated.

Model 1	Model 2	Statistic	
		range	semi-quadratic
DCC(1,1)-GARCH(1,1)	CCC-GARCH(1,1)	10.843	117.56
RW-ACC-GARCH(1,1)	CCC-GARCH(1,1)	33.502	1122.4
RW-TACC-GARCH(1,1)	CCC-GARCH(1,1)	32.936	1084.9
RW-ACC-GARCH(1,1)	DCC(1,1)-GARCH(1,1)	21.104	445.37
RW-TACC-GARCH(1,1)	DCC(1,1)-GARCH(1,1)	19.798	391.95
RW-TACC-GARCH(1,1)	RW-ACC-GARCH(1,1)	2.7151	22.232
Global value for EPA test:	1^{st} step	33.502 (0)	3184.3 (0)
Global value for EPA test:	2^{nd} step	21.104 (0)	859.56 (0)
Global value for EPA test:	3^{rd} step	2.7151 (0.0064)	22.232 (0)

Model	Worst performing index		
	1^{st} step	2^{nd} step	3^{rd} step
CCC-GARCH(1,1)	26.583	–	–
DCC(1,1)-GARCH(1,1)	8.7035	20.941	–
RW-ACC-GARCH(1,1)	-19.597	-8.7309	2.7151
RW-TACC-GARCH(1,1)	-20.748	-13.419	-2.7151

Table 6: Testing differences of multivariate OS-NL performance terms among different multivariate GARCH models. Upper panel: Values of pairwise *sign-type* test statistics, values of the range statistic T_R and the semi-quadratic statistic T_{SQ} for global test of equal predictive ability (EPA). Corresponding P-values are given between parentheses. Lower panel: Worst performing index results for the construction of the confidence model sets. If the null hypothesis of EPA is rejected, the model with the largest worst performing index value is eliminated.

Model	IS-			OS-		
	MAE	RMSE	R^2	MAE	RMSE	R^2
CCC-GARCH(1,1)	0.3052	0.6127	0.0431	0.2500	0.4373	0.0210
DCC(1,1)-GARCH(1,1)	0.3044	0.6067	0.0443	0.2437	0.4341	0.0169
RW-ACC-GARCH(1,1)	0.3083	0.6075	0.0446	0.2422	0.4336	0.0199
RW-TACC-GARCH(1,1)	0.3081	0.6075	0.0447	0.2421	0.4337	0.0206

Table 7: In-sample and out-of-sample goodness of fit results of the different models for the equally weighted portfolio Δ constructed on the six-dimensional exchange-rate return series introduced in Section 4. MAE, RMSE are the standard univariate mean absolute errors and root mean squared errors.

Model 1	Model 2	Statistic	
		range	semi-quadratic
DCC(1,1)-GARCH(1,1)	CCC-GARCH(1,1)	3.7726	14.233
RW-ACC-GARCH(1,1)	CCC-GARCH(1,1)	4.4346	19.666
RW-TACC-GARCH(1,1)	CCC-GARCH(1,1)	4.5190	20.421
RW-ACC-GARCH(1,1)	DCC(1,1)-GARCH(1,1)	1.5570	2.4241
RW-TACC-GARCH(1,1)	DCC(1,1)-GARCH(1,1)	1.7142	2.9383
RW-TACC-GARCH(1,1)	RW-ACC-GARCH(1,1)	1.4954	2.2363
Global value for EPA test:	1^{st} step	4.5190 (0.0012)	61.919 (0.0008)
Global value for EPA test:	2^{nd} step	1.7142 (0.092)	7.5987 (0.0555)
Global value for EPA test:	3^{rd} step	1.4954 (0.1340)	2.2363 (0.0570)

Model	Worst performing index		
	1^{st} step	2^{nd} step	3^{rd} step
CCC-GARCH(1,1)	4.4019	–	–
DCC(1,1)-GARCH(1,1)	-1.1403	1.6360	–
RW-ACC-GARCH(1,1)	-4.0013	-1.3963	1.4954
RW-TACC-GARCH(1,1)	-4.2501	-1.8674	-1.4954

Table 8: Testing differences of univariate OS-MAE performance terms among different multivariate GARCH models. Upper panel: Values of pairwise *t-type* test statistics, values of the range statistic T_R and the semi-quadratic statistic T_{SQ} for global test of equal predictive ability (EPA). Corresponding P-values are given between parentheses. Lower panel: Worst performing index results for the construction of the confidence model sets. If the null hypothesis of EPA is rejected, the model with the largest worst performing index value is eliminated.

Model	Hit rate		Ljung-Box P-value	
	$x = 0.05$	$x = 0.01$	$x = 0.05$	$x = 0.01$
CCC-GARCH(1,1)	0.0668	0.0162	0.0193*	0.5329
DCC(1,1)-GARCH(1,1)	0.0648	0.0182*	0.0754	0.1887
RW-ACC-GARCH(1,1)	0.0688	0.0162	0.1700	0.5329
RW-TACC-GARCH(1,1)	0.0688	0.0162	0.1700	0.5329

Table 9: Value-at-Risk application: results of classical binomial test on the total number of hits and of Ljung-Box tests for autocorrelation in the hit sequence. VaR predictions for the equally weighted portfolio Δ constructed on the six-dimensional exchange-rate data described in Section 4.1 are estimated using the different multivariate models described in the paper. The back-testing period goes from October 15, 1997, to September 13, 1999, for a total of 494 trading days. Asterisks denote significance at the 5% confidence level or better.