

National Centre of Competence in Research  
Financial Valuation and Risk Management

Working Paper No. 137

Semiparametric Multivariate GARCH Models for  
Volatility Asymmetries and Dynamic Correlations

Francesco Audrino

Giovanni Barone-Adesi

First version: October 2003  
Current version: October 2003

This research has been carried out within the NCCR FINRISK project on  
“Interest Rate and Volatility Risk”.

---

# Semiparametric Multivariate GARCH Models for Volatility Asymmetries and Dynamic Correlations

Francesco Audrino\* and Giovanni Barone-Adesi†

Institute of Finance  
University of Southern Switzerland

October 2003

## Abstract

We propose a simple class of semiparametric multivariate GARCH models, allowing for asymmetric volatilities and time-varying conditional correlations. Estimates for time-varying conditional correlations are constructed by means of a convex combination of estimates for averaged correlations (across all assets) and dynamic realized (historical) correlations. Our model is very parsimonious. Estimation based on Functional Gradient Descent is computationally feasible also in very large dimensions without resorting to any variance reduction technique. We back-test the model on a six-dimensional exchange-rate time series and collect empirical evidence of its strong predictive power compared to other related existing procedures.

*Keywords:* Multivariate GARCH models; Asymmetric volatility; Dynamic conditional correlations; Functional Gradient Descent (FGD) estimation; Tree-structured GARCH models.

*JEL codes:* C12, C13, C14, C51, C61

---

\*Corresponding author. Address: Institute of Finance, Via Buffi 13, Centro Civico, CH-6900 Lugano, Switzerland. E-mail: francesco.audrino@lu.unisi.ch. Phone: 0041919124677. Financial support 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, Centro Civico, CH-6900 Lugano, Switzerland. E-mail: giovanni.barone-adesi@lu.unisi.ch.

# 1 Introduction

We propose a simple class of semiparametric multivariate GARCH models to estimate conditional covariance matrices which allows for asymmetric volatilities and time-varying conditional correlations. Moreover, it takes into account the possible non-linear dependence structure across individual series. 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. Since the covariance matrix is an essential ingredient for many issues in financial econometrics – such as the computation of risk measure estimates for portfolios of assets, asset allocation or tests of asset pricing models – modelling and forecasting truly conditional covariance matrices is an important and central problem in modern empirical finance.

It has already been shown that 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 when working with a separate univariate model for each individual financial instrument<sup>1</sup>. 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  $\{P_t; t = 1, \dots, n\}$ . 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 Criterion (AIC) or the Schwarz Bayesian Information Criterion (BIC). This problem is well illustrated by a standard BEKK(1,1) model with 10 assets. When using AIC, we have to check and to 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 semiparametric 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 through 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 a limited dynamic in the correlations, introducing a GARCH-type structure. The DCC model guarantees the positivity of the conditional

correlation matrix under simple conditions on the parameters and has become very popular. However, the dynamic is 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 completely investigated and compared.

Another stylized fact that has been widely studied over the past twenty years is the so-called “asymmetric volatility phenomenon”, i.e. the fact that volatility increases more after a negative shock than after a positive shock of the same magnitude. Two main explanations have been advocated for this phenomenon: the leverage hypothesis (Black, 1976, and Christie, 1982) and the volatility feedback effect (Campbell and Hentschel, 1992, and Wu, 2001). Asymmetric effects have also been recently found in conditional correlations (see, for instance, Kroner and Ng, 1998; Beekaert and Wu, 2000, and Errunza and Hung, 1999) and have been analyzed by Cappiello et al. (2003). They investigated whether, in addition to stocks, government fixed income securities also exhibit asymmetry in conditional second moments and explored the dynamics and changes in the correlations of international asset markets. In their empirical investigations, they used a generalization of the DCC-GARCH model (see, for instance, Sheppard, 2002) which allows for asymmetries in both volatilities and conditional correlations. However, the focus of that paper was on understanding and estimating asymmetries of in-sample conditional covariances and correlations across equity and bond returns, and not on the forecasting power of the model.

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 for 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. The estimation procedure is similar to the two-stage one used for the DCC model. In a second step, we improve the parametric estimates for the individual volatility functions from the classical two-stage estimation procedure by means of the nonparametric Functional Gradient Descent (FGD) technique of Audrino and Bühlmann (2003). Such technique is computationally feasible in large dimensions and yields reliable out-of-sample performances.

We test the model on a six-dimensional time series of exchange-rate data. We compare its out-of-sample forecasting power with the ones of the CCC-GARCH and the DCC-GARCH models, both at the multivariate and univariate portfolio level. 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.

We confront the results from our models with those from the CCC-GARCH and the DCC-GARCH models also allowing for general, nonparametric FGD volatilities. We find that FGD is particular important and improves substantially the performances in the CCC context. It seems that in this particular setting, where the conditional covariances are modelled as a product of the constant conditional correlations and the individual volatilities, a FGD estimation of the latter can also incorporate some of the (non-linear and asymmetric) time-varying behavior of the former, improving the final covariance forecasts.

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 and nonparametric, asymmetric volatilities. The estimation procedure is presented 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 presented in Section 4. Section 5 summarizes and concludes.

## 2 Semiparametric multivariate GARCH models

This Section describes our class of semiparametric multivariate GARCH models with dynamic conditional correlation and asymmetric non-linear volatilities.

### 2.1 Starting point

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

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

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 accurate 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  $V_t = \text{Cov}_{d \times d}(\mathbf{X}_t | \mathcal{F}_{t-1})$  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

$$\mathbf{X}_t = \mu_t + \Sigma_t \mathbf{Z}_t. \quad (2.2)$$

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  $\text{Cov}(\mathbf{Z}_t) = I_d$ .

(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  $v_{t,ij} = \rho_{t,ij}(v_{t,ii}v_{t,jj})^{1/2}$  ( $i, j = 1, \dots, d$ ). In this model  $\rho_{t,ij} = \text{Corr}(x_{t,i}, x_{t,j} | \mathcal{F}_{t-1})$  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

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

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

(A4) (conditional mean)<sup>2</sup> The conditional mean  $\mu_t$  is of the form

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

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

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

$$\begin{aligned} V_t &= \Sigma_t \Sigma_t^T = D_t R_t D_t, \\ D_t &= \text{diag}(\sigma_{t,1}, \dots, \sigma_{t,d}), \quad R_t = [\rho_{t,ij}]_{i,j=1}^d. \end{aligned}$$

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  $\sigma_{t,i}$ ,  $i = 1, \dots, d$  on  $\mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \dots$ , 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:

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

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

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

In this model,  $\varepsilon_t$  is a standardized error term,  $\varepsilon_t^T = ((X_{t,1} - \mu_{t,1})/\sigma_{t,1}, \dots, (X_{t,d} - \mu_{t,d})/\sigma_{t,d})$ , and  $\bar{Q}$  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 semiparametric models (2.2) 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.<sup>3</sup> Then, the conditional variance of the portfolio return can be computed as

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

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

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

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:

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

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  $0 < \bar{\rho}_t \leq 1$ . 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

We first describe a new parametric class of multivariate GARCH models in the general context given in (2.2). In Section 2.4 we are going to allow also for more general nonparametric volatility functions  $F_i$  in (A3).

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

$$\begin{aligned} \sigma_{t,i}^2 &= \alpha_{0,i} + \alpha_{1,i} (X_{t-1,i} - \mu_{t-1,i})^2 + \beta_i \sigma_{t-1,i}^2, \text{ where} \\ \alpha_{0,i} &> 0, \alpha_{1,i} \geq 0, \beta_i \geq 0, \alpha_{1,i} + \beta_i < 1, \quad i = 1, \dots, d. \end{aligned} \quad (2.8)$$

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

$$R_t = (1 - \lambda)\overline{Q}_{t-p}^{t-1} + \lambda\overline{R}_t, \quad \lambda \in [0, 1]; \quad (2.9)$$

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

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  $\varepsilon_t$  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  $\bar{r}_t = (d\overline{\rho}_t - 1)/(d - 1) \leq 1$ , 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  $\text{mean}(\overline{R}_t) = \overline{\rho}_t$ ). The model (2.9) is a convex combination of realized dynamic conditional correlations and averaged conditional correlations. Clearly, when the parameter  $\lambda$  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 still being a convex combination of two terms, and 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  $(\overline{\rho}_{t-1}, \mathbf{X}_{t-1})^T$ :

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

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  $\mathcal{R}_k$  whose edges are 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 zero, the data are uncorrelated. Otherwise, some regime-dependent weight is also given to the historical dynamic conditional correlations.

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  $i = 1, \dots, d$ , 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  $\bar{\rho}_t$  in (2.7) satisfies  $\bar{\rho}_t \geq 1/d \forall t$ ;
- 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  $\sigma_{t,i}^2$  is strictly positive since  $\alpha_{0,i} > 0$ . Moreover:

- i) from the restriction  $\bar{\rho}_t \geq 1/d \forall t$  it follows that  $0 \leq \bar{r}_t \leq 1, \forall t$ . Then, we can write the matrix  $\bar{R}_t$  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:

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

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

- ii)  $R_t$  is positive definite for all  $t$  as it is a weighted average of two positive definite matrices  $\overline{Q}_{t-p}^{t-1}$  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  $\bar{\rho}_t \geq 1/d$  is satisfied in most of the practical applications, in particular when the dimension  $d$  of the problem is high.

## 2.4 Asymmetric volatilities

In this Section, we generalize the time-varying dynamics of the GARCH(1,1) conditional variances (2.8) by allowing for more general nonparametric individual volatility functions  $F_i$  in the model (2.2). The time-varying behavior of the conditional correlations follows the RW-ACC model (2.9) or the RW-TACC model (2.10).

This general methodology includes important non-linear effects like asymmetric volatility and is fitted using Functional Gradient Descent (FGD), which is an optimization technique in function space. This technique is the same proposed and tested by Audrino and Bühlmann (2003) or Audrino and Barone-Adesi (2002), with very satisfactory out-of-sample results. In particular, we use FGD for estimating the (squared) individual volatility functions  $F_i(\cdot)$  in (A3), where we restrict  $F_i(\cdot) : \mathbb{R}^{qd} \rightarrow \mathbb{R}^+$ , to depend on the first  $q$  lagged multivariate observations which are used as predictor data. The main task of FGD is to find the estimates for the functions  $F_i(\cdot)$  which minimize a suitable loss function  $\lambda$ . Estimation of such  $F_i(\cdot)$  from the data can be done via a constrained minimization of a suitable empirical criterion by applying functional gradient descent. The minimizer  $\widehat{F}_i(\cdot)$  is imposed to satisfy a “smoothness” constraint in terms of an additive expansion of “simple estimates”. These “simple estimates” are given from a statistical procedure called the base learner that we denote by

$$\mathcal{S}(\cdot, \gamma)_{U_i, X}, \quad x \in \mathbb{R}^{qd}, \quad (2.12)$$

where  $\gamma$  is a finite or infinite-dimensional parameter which is estimated from the data  $(U_i, X)$ , where  $U_i$  denote some generalized residuals or pseudo-response variables (see Section 3.2). The base learner  $\mathcal{S}$  is often constructed from a (constrained or penalized) least squares fitting; common examples of base learners are regression trees, projection pursuit, splines or neural nets. For

more details about FGD, see Friedman et al. (2000), Friedman (2001), Audrino and Bühlmann (2003) and Bühlmann and Yu (2003).

Summarizing, we start from simple parametric initial functions  $F_{i,0}(\cdot)$  and then we add nonparametric terms to minimize the empirical risk. The final (squared) volatility functions have the form

$$F_i(\cdot) = F_{i,0}(\cdot) + \sum_{m=1}^{M_i} \mathcal{S}(\cdot, \gamma_{i,m})_{\mathbf{U}, \mathbf{X}}, \quad i = 1, \dots, d, \quad (2.13)$$

where  $M_i$  is the number of terms needed to minimize the empirical criterion. As starting volatility functions  $F_{i,0}(\cdot)$ , we use the simple univariate parametric GARCH(1,1) volatility estimates. We choose as base learner  $\mathcal{S}$  regression trees, because particular in high dimensions they have the ability to do variable selection by choosing just a few of the explanatory variables for prediction. In this particular case, the parameter  $\gamma$  in (2.12) describes the axis to be split, the split points and the fitted values for every terminal node (the constants in the piecewise constant fitted function).

We call such models allowing for non-linear and nonparametric volatilities RW-ACC-FGD model and RW-TACC-FGD model, respectively. For comparison purposes, we also apply in Section 4 the FGD estimation technique to the classical CCC model (2.3) and the DCC(1,1) model (2.4). Similarly, we call the resulting models CCC-FGD and DCC(1,1)-FGD, respectively.

### 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.

#### 3.1 Classical two-stage estimation

The parameters  $\phi = (a_{0,i}, a_{1,i}, \alpha_{0,i}, \alpha_{1,i}, \beta_i, \quad i = 1, \dots, d)$  and the parameter(s)  $\psi = \lambda$  or  $\psi = (\lambda_k, \quad k = 1, \dots, N)$  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

$$\begin{aligned} l(\phi, \psi; \mathbf{X}_2^n) &= \sum_{t=1}^n \log \left( (2\pi)^{-d/2} \det(V_t)^{-1/2} \exp \left( - (\mathbf{X}_t - \mu_t)^T V_t^{-1} (\mathbf{X}_t - \mu_t) / 2 \right) \right) \\ &= -\frac{1}{2} \sum_{t=1}^n \left( d \log(2\pi) + 2 \log(\det(D_t)) + \log(\det(R_t)) + \varepsilon_t^T R_t^{-1} \varepsilon_t \right), \end{aligned} \quad (3.1)$$

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)  $\psi$  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

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

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  $\bar{\rho}_t$  defined in (2.7). This can be easily achieved from the first-stage estimates for the individual volatilities  $\hat{\sigma}_{t,i}$ ,  $i = 1, \dots, d$  and estimating univariate AR(1)-GARCH(1,1) volatilities  $\hat{\sigma}_{t,P}$  of the equally weighted portfolio  $\Delta_t$ , 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

$$\hat{\rho}_t = \hat{\sigma}_{t,P}^2 / \left( \sum_{i=1}^d \hat{\sigma}_{t,i}^2 \right). \quad (3.3)$$

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

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

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

Note that the only portion of the second stage likelihood (3.4) that will influence the parameter selection for  $\psi$  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  $d$ .

### 3.2 FGD volatilities

The estimation of the RW-ACC- and RW-TACC-FGD models introduced in Section 2.4 is based on the FGD technique. After having computed two-step estimates for the model parameters, we perform a final step using FGD for the individual volatility functions. The estimated time-varying conditional correlation dynamics are kept fixed.

The FGD algorithm used in this paper is the one recently proposed by Audrino and Bühlmann (2003) and is described in Appendix A. In our empirical investigations we found that the optimal value for the number  $q$  of lagged variables to be used as predictors was  $q = 2$ . We chose as a suitable loss function the (multivariate) negative log-likelihood function (A.1) and, as base learner, regression trees with three final nodes. For the estimation, we use a shrinkage factor  $\nu = 0.25$  in (A.3).

## 4 Empirical results

This Section presents the results of our estimations of RW-(T)ACC-GARCH(1,1) and RW-(T)ACC-FGD models 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. We also allow the CCC model and the DCC(1,1) model to incorporate nonparametric FGD volatility

functions. We do that in order to assess the contribution given by FGD to the different modelling approaches.

#### 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  $\Delta_t$  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<sup>th</sup> 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  $\alpha_1$ 's and  $\beta$ '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  $\bar{Q}_{t-p}^{t-1}$ . 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, multivariate and univariate performance results 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<sup>th</sup> 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. The effect produced by the possibility of

estimating non-linear, asymmetric FGD volatilities appears to be important with respect to such a metric. The percentage of failures is in all cases reduced by FGD. 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):

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

$$\text{OS-NL: } -\log\text{-likelihood}\left(\tilde{\mathbf{X}}_1^{n_{\text{out}}}; \hat{\phi}, \hat{\psi}\right)$$

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

$$\text{OS-MAE: } \frac{1}{d^2} \sum_{i,j=1}^d \frac{1}{n_{\text{out}}} \sum_{t=1}^{n_{\text{out}}} \left| \hat{v}_{t,ij} - (\tilde{X}_{t,i} - \hat{\mu}_{t,i})(\tilde{X}_{t,j} - \hat{\mu}_{t,j}) \right|$$

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

$$\text{OS-RMSE: } \left( \frac{1}{d^2} \sum_{i,j=1}^d \frac{1}{n_{\text{out}}} \sum_{t=1}^{n_{\text{out}}} \left| \hat{v}_{t,ij} - (\tilde{X}_{t,i} - \hat{\mu}_{t,i})(\tilde{X}_{t,j} - \hat{\mu}_{t,j}) \right|^2 \right)^{1/2}$$

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

$$\text{OS-R}^2: \frac{1}{d^2} \sum_{i,j=1}^d \left| \text{Cor}\left(\hat{v}_{t,ij}, (\tilde{X}_{t,i} - \hat{\mu}_{t,i})(\tilde{X}_{t,j} - \hat{\mu}_{t,j})\right) \right|,$$

where  $\tilde{\mathbf{X}}_1^{n_{\text{out}}} = \mathbf{X}_{n+1}, \dots, \mathbf{X}_{n+n_{\text{out}}}$  are the test data and the parameter estimates equipped with hats have been constructed from the training sample  $\mathbf{X}_1^n = \mathbf{X}_1, \dots, \mathbf{X}_n$ . 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 one time by the CCC-FGD model, three times by the DCC(1,1)-FGD model, four times by the RW-ACC-FGD model and one time by the RW-TACC-FGD model. In particular, when focusing on the most important OS statistics, we see that the optimal values are reached by the RW-(T)ACC-FGD models. As expected, the CCC models are clearly beaten by models allowing for dynamic conditional correlations with respect to most of the OS statistics. Moreover, we observe over-fitting problems when using DCC(1,1) models: they reach the optimal values in-sample, but they do not seem to be as good as the RW-(T)ACC models for prediction. Finally, in all four types of models, introducing FGD volatilities improves conditional covariance matrix estimation and prediction with respect to most performance measures.

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 in large 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<sup>5</sup>,

$$\widehat{D}_t = \widetilde{U}_{t;\text{model}_1} - \widetilde{U}_{t;\text{model}_2}, \quad t = 1, \dots, n_{\text{out}},$$

where

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

We test the null hypothesis that differences  $\widehat{D}_t$  have mean zero against the alternative of mean less than zero, i.e. the estimates from model<sub>1</sub> are better than the ones from model<sub>2</sub>. Moreover, we also test the null hypothesis that the number of negative differences has mean 1/2 against the

alternative of mean bigger than  $1/2$ . This allows us to investigate whether there is a systematic difference between the estimates from the two models. For this purpose, we use versions of the t-test and sign-test, adapted to the case of dependent observations. The exact definition of the tests is presented in Appendix B following Audrino and Bühlmann (2003). Results of the tests for the real data example under investigation are summarized in Table 5.

**TABLE 5 ABOUT HERE.**

The upper part of Table 5 clearly shows that, as expected, models allowing for dynamic conditional correlations are preferred to the classical CCC-GARCH(1,1) model. Such significant results seem to disappear when considering differences of OS-NL performance terms between models allowing for dynamic conditional correlations with respect to the t-type tests. This finding may be just a fact of a low power of the test due to non-Gaussian observations. On the other hand, the sign-type tests, which are robust against deviations from Gaussianity, yield very significant results: 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 some significant advantage also over the ones from the RW-ACC-GARCH(1,1) model.

The tests in the central part of Table 5 investigate the impact of FGD volatilities in both the constant and the dynamic conditional correlations context. Both t-type and sign-type tests are significant at the 5% level only in the CCC setting, meaning that, with respect to the OS-NL measure, the use of FGD in this case yields important improvements in estimation and forecasting and can not be neglected. However, the final tests performed in Table 5 show that the mis-specification given by the assumption of constant conditional correlations can not be completely corrected with the use of more general non-parametric FGD volatility functions.

#### **4.5 Portfolio performance results**

We test in this Section the accuracy of volatility estimates and predictions for the equally weighted portfolio  $\Delta_t$  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 6.

## TABLE 6 ABOUT HERE.

Table 6 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. It is also interesting to observe that the performance results from a CCC-FGD fit are at least equally good as those from a DCC(1,1)-GARCH(1,1) fit with respect to most statistics. This was not the case for the tests at the multivariate level.

Similarly to Table 4, differences between the models are in general small. Hence, we consider differences of each term in the OS-MAE<sup>6</sup> statistic and use again the concept of hypothesis testing. Results of the t-type and sign-type tests described in the Appendix B for the equally weighted portfolio  $\Delta_t$  are summarized in Table 7.

## TABLE 7 ABOUT HERE.

As expected, multivariate GARCH(1,1) models with time-varying conditional correlations are preferred to the classical CCC-GARCH(1,1) approach. Moreover, from Table 7 the RW-TACC-GARCH(1,1) model is also clearly preferred to the RW-ACC-GARCH(1,1) model. No significant difference was found between our multivariate models and the DCC(1,1)-GARCH(1,1) model. Moreover, no significant difference was found between the CCC-FGD model and the DCC(1,1)-GARCH(1,1) model, too. This result highlights the improvements in the estimation allowed by more general FGD volatilities in the CCC context. In contrast, our RW-(T)ACC-GARCH(1,1) models significantly improve predictions from the CCC-FGD model. This can also be seen as an indirect evidence of the better predictive power of our multivariate models over a DCC(1,1)-GARCH(1,1) approach.

### 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  $\Delta_t$  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  $(\Delta_t - \hat{\mu}_{t,P})/\hat{\sigma}_{t,P}$  are fitted using a univariate scaled  $t_\xi$  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

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

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

Let

$$\text{Hit}_t = I_{\{\Delta_t < \widehat{\text{VaR}}_t\}}, \quad t = 1, \dots, n_{\text{out}}, \quad (4.2)$$

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<sup>th</sup> order for the 5% and 1% confidence levels are summarized in Table 8. 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 8 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 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 are constructed using the Functional Gradient Descent (FGD) technique 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. In particular, we considered forecast accuracy at the multivariate and at the portfolio, univariate level, persistence of multivariate standardized residuals and precision of portfolio Value-at-Risk estimates.

## A Appendix: FGD for multivariate volatility

Let the loss function  $\lambda$  be defined as

$$\begin{aligned}\lambda_R(\mathbf{Y}, \mathbf{f}) &= \log(\det(D(\mathbf{f}))) + \frac{1}{2}(D(\mathbf{f})^{-1}\mathbf{Y})^T R^{-1}(D(\mathbf{f})^{-1}\mathbf{Y}) + \frac{1}{2}\log(\det(R)) + \frac{d}{2}\log(2\pi), \\ D(\mathbf{f}) &= \text{diag}(f_1, \dots, f_d).\end{aligned}\tag{A.1}$$

As pointed out with the subscript, the loss function depends on the already estimated dynamic conditional correlation matrix  $R$ , which will be kept fixed during the algorithm.

The partial derivatives of the loss function with respect to the squared volatilities  $f_i$  are

$$\frac{\partial \lambda_R(\mathbf{Y}, \mathbf{f})}{\partial f_i} = (f_i - \sum_{j=1}^d \frac{\gamma_{ij} y_i y_j}{f_i^{3/2} f_j^{1/2}})/2, \quad i = 1, \dots, d,\tag{A.2}$$

where  $[\gamma_{ij}]_{i,j=1}^d = R^{-1}$ . This will be used when computing negative gradients (see Step 2 in the following FGD algorithm) for every component  $i = 1, \dots, d$ .

*Step 1 (initialization).*

Choose the starting function  $\hat{F}_{i,0}(\cdot)$  and denote by  $\hat{F}_{i,0}(t) = \hat{F}_{i,0}(\mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \dots)$  ( $i = 1, \dots, d$ ).

Denote as  $\hat{\mu}_t$  the first stage estimates from (3.2) for the conditional mean. Set  $m = 1$ .

For every component  $i = 1, \dots, d$ , do the following.

*Step 2<sub>i</sub> (projection of component gradients to base learner).*

Compute the negative gradient

$$U_{t,i} = - \frac{\partial \lambda_{\hat{R}_t}(\mathbf{X}_t - \hat{\mu}_t, \mathbf{F})}{\partial F_i} \Big|_{\mathbf{F}=\hat{\mathbf{F}}_{m-1}(t)}, \quad t = q+1, \dots, n.$$

This is explicitly given in (A.2). Then, fit the negative gradient vector  $U_i = (U_{q+1,i}, \dots, U_{n,i})^T$  with a base learner, using always the first  $q$  time-lagged predictor variables (i.e.  $\mathbf{X}_{t-q}^{t-1}$  is the predictor for  $U_{t,i}$ )

$$\hat{f}_{m,i}(\cdot) = \mathcal{S}(\cdot, \hat{\gamma})_{U_i, X},$$

where  $\mathcal{S}(x, \hat{\gamma})_{U_i, X}$  denotes the predicted value at  $x$  from the base learner  $\mathcal{S}$  using the response vector  $U_i$  and predictor variables  $X = \mathbf{X}_{t-q}^{t-1}$ ,  $t = q+1, \dots, n$ .  $\hat{\gamma}$  is a finite or infinite-dimensional parameter which is estimated from the data.

*Step 3<sub>i</sub> (line search).* Perform one-dimensional optimization for the step-length,

$$\hat{w}_{m,i} = \operatorname{argmin} \sum_{t=q+1}^n \lambda_{\hat{R}_t}(\mathbf{X}_t - \hat{\mu}_t, \hat{\mathbf{F}}_{m-1}(t) + w \hat{f}_{m,i}(\mathbf{X}_{t-q}^{t-1})).$$

( $\hat{\mathbf{F}}_{m-1}(t) + w\hat{f}_{m,i}(\cdot)$  is defined as the function which is constructed by adding in the  $i$ th component only). This can be expressed more explicitly by using (A.1).

*Step 4 (up-date).* Select the best component as

$$i_m^* = \operatorname{argmin}_i \sum_{t=q+1}^n \lambda_{\hat{R}_t}(\mathbf{X}_t - \hat{\mu}_t, \hat{\mathbf{F}}_{m-1}(t) + \hat{w}_{m,i} \hat{f}_{m,i}(\mathbf{X}_{t-q}^{t-1})).$$

Up-date

$$\hat{\mathbf{F}}_m(\cdot) = \hat{\mathbf{F}}_{m-1}(\cdot) + \hat{w}_{m,i_m^*} \hat{f}_{m,i_m^*}(\cdot).$$

*Step 5 (iteration).* Increase  $m$  by one and iterate Steps 2–4 until stopping with  $m = M$ . This produces the FGD estimate

$$\hat{\mathbf{F}}_M(\cdot) = \hat{\mathbf{F}}_0(\cdot) + \sum_{m=1}^M \hat{w}_{m,i_m^*} \hat{f}_{m,i_m^*}(\cdot).$$

The stopping value  $M$  is chosen with the following scheme: split the (in-sample) estimation period into two sets, the first of size  $0.7 \cdot n$  used as training set and the second of size  $0.3 \cdot n$  used as test set (this can also be used when the data are dependent). The optimal value of  $M$  is then chosen to optimize the cross-validated log-likelihood.

Note that it is often useful to reduce the complexity of the base learner to avoid over-fitting. A simple but effective way to do it is via shrinkage towards zero. The up-date in Step 4 of the FGD algorithm is then replaced by

$$\hat{\mathbf{F}}_m(\cdot) = \hat{\mathbf{F}}_{m-1}(\cdot) + \nu \cdot \hat{w}_{m,i_m^*} \hat{f}_{m,i_m^*}(\cdot), \quad 0 < \nu \leq 1. \quad (\text{A.3})$$

Obviously, this reduces the variance of the base learner by the factor  $\nu^2$ .

## B Appendix: t-type and sign-type tests

Consider differences  $\widehat{D}_t$ ,  $t = 1, \dots, n_{\text{out}}$ , of performance terms. The t-type test statistic in the case of dependent observations introduced in Section 4.4 is

$$\sqrt{n_{\text{out}}} \frac{\overline{D}}{\widehat{\sigma}_{D;\infty}}, \quad \text{where } \overline{D} = \frac{1}{n_{\text{out}}} \sum_{t=1}^{n_{\text{out}}} \widehat{D}_t. \quad (\text{B.1})$$

In (B.1),  $\widehat{\sigma}_{D;\infty}^2 = (2\pi)\widehat{f}_D(0)$ , where  $\widehat{f}_D(0)$  is a smoothed periodogram estimate at frequency zero, based on  $\widehat{D}_1, \dots, \widehat{D}_{n_{\text{out}}}$ ; see for example Brockwell and Davis (1991). The motivation for this estimate is based on the assumption that  $\{\widehat{D}_t\}_t$  is stationary (conditional on the training data) and satisfies suitable dependence conditions, e.g. mixing. Then, conditional on the training data,

$$\begin{aligned} \sqrt{n_{\text{out}}}(\overline{D} - E[\widehat{D}_t]) &\implies \mathcal{N}(0, \sigma_{D;\infty}^2) \quad (n_{\text{out}} \rightarrow \infty), \\ \sigma_{D;\infty}^2 &= \sum_{k=-\infty}^{+\infty} \text{Cov}[\widehat{D}_0, \widehat{D}_k] = (2\pi)\widehat{f}_D(0), \end{aligned} \quad (\text{B.2})$$

where  $\widehat{f}_D(0)$  is the spectral density at zero of  $\{\widehat{D}_t\}_t$ .

Thus, using (B.2) for the test statistic in (B.1), and conditional on the training data,

$$\sqrt{n_{\text{out}}} \frac{\overline{D}}{\widehat{\sigma}_{D;\infty}} \implies \mathcal{N}(0, 1) \quad (n_{\text{out}} \rightarrow \infty) \quad (\text{B.3})$$

under the null-hypothesis.

Analogously, the version of the sign test in the case of dependent observations introduced in Section 4.4 is based on the number of negative differences

$$\widehat{W}_t = I_{\{\widehat{D}_t \leq 0\}}, \quad t = 1, \dots, n_{\text{out}},$$

for the null hypothesis that the negative differences  $\widehat{W}_t$  have mean  $\frac{1}{2}$  against the alternative of mean bigger than  $\frac{1}{2}$ . The test statistic is given by

$$\sqrt{n_{\text{out}}} \frac{\overline{W} - \frac{1}{2}}{\widehat{\sigma}_{W;\infty}}, \quad \text{where } \overline{W} = \frac{1}{n_{\text{out}}} \sum_{t=1}^{n_{\text{out}}} \widehat{W}_t \quad (\text{B.4})$$

and  $\widehat{\sigma}_{W;\infty}^2$  as in (B.1) but based on  $\widehat{W}_1, \dots, \widehat{W}_T$ . As in the derivation of the t-type test above, we have, conditional on the training data,

$$\sqrt{n_{\text{out}}} \frac{\overline{W} - \frac{1}{2}}{\widehat{\sigma}_{W;\infty}} \implies \mathcal{N}(0, 1) \quad (n_{\text{out}} \rightarrow \infty) \quad (\text{B.5})$$

under the null-hypothesis.

## Notes

<sup>1</sup>See, for example, Audrino and Bühlmann (2002) for an application to the measurement of risk in global stock markets.

<sup>2</sup>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.

<sup>3</sup>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.

<sup>4</sup>The estimation of the optimal partition  $\hat{\mathcal{P}}$  (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).

<sup>5</sup>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.

<sup>6</sup>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 Barone-Adesi, G. (2002). A multivariate FGD technique to improve VaR computation in equity markets. To appear in *Computational Management Science*.
- 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. (2002). Synchronizing multivariate financial time series. Manuscript USI, Lugano, Switzerland.
- 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 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.
- Beckaert, G. and Wu, G. (2000). Asymmetric volatility and risk in equity markets. *The Review of Financial Studies*, **13**, 1-42.
- Billio, M., Caporin, M. and Gobbo, M. (2003). Block Dynamic Conditional Correlation Multivariate GARCH models. Working Paper, Greta, Italy.

- Black, F. (1976). Studies in stock price volatility changes. *Proceedings of The American Statistical Association, Business and Economic Statistics Section*, 177-181.
- 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.
- Bühlmann, P. and Yu, B. (2003). Boosting with the  $L_2$ -loss: regression and classification. *Journal of the American Statistical Association* **98**, 324-339.
- Campbell, J.Y. and Hentschel, L. (1992). No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, **31**, 281-318.
- Cappiello, L., Engle, R.F. and Sheppard, K. (2003). Asymmetric dynamics in the correlations of global equity and bond returns. Mimeo, University of California, San Diego.
- Christie, A.A. (1982). The stochastic behavior of common stock variances- value, leverage and interest rates effects. *Journal of Financial Economics*, **10**, 407-432.
- 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.
- Errunza, V.K.H. and Hung, M.-W. (1999). Can the gains from international diversification be achieved without trading abroad? *Journal of Finance* **54**, 2075-2107.

- Freedman, D. (1984). On bootstrapping two-stage least-squares estimates in stationary linear models. *Annals of Statistics* **12**, 827–842.
- Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics* **29**, 1189-1232.
- Friedman, J.H., Hastie, T. and Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting. *Annals of Statistics* **28**, 337-407 (with discussion).
- Hansen, P.R. and Lunde, A. (2002). A forecast comparison of volatility models: does anything beat a GARCH(1,1)? Manuscript, Brown University.
- Kroner, K.F. and Ng, V.K. (1998). Modeling asymmetric comovements of asset returns. *Review of Financial Studies* **11**, No. 4, 817-844.
- 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.

Wu, G. (2001). The determinants of asymmetric volatility. *The Review of Financial Studies*, **14**, 837-859.

### Averaged conditional correlations

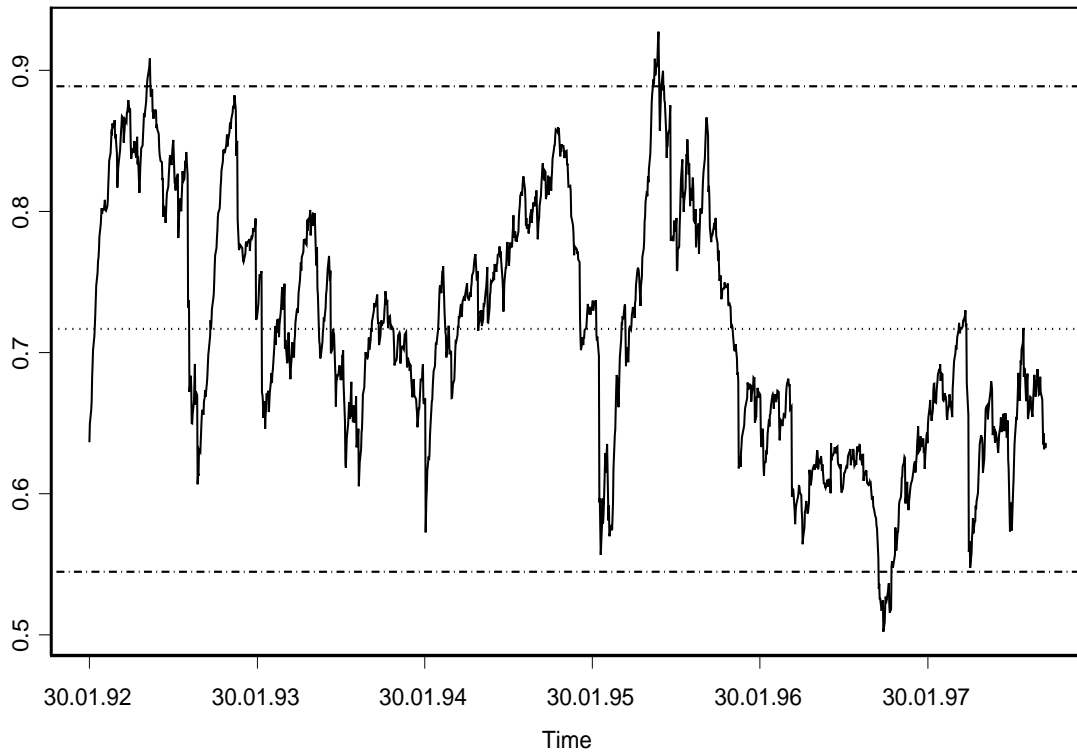


Figure 1: Estimated averaged conditional correlation series during the in-sample period between January 2, 1992 and October 14, 1997 for a total of 1500 trading days. The dotted line and the dashed lines indicate the mean of estimated averaged conditional correlations and a classical two standard deviations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

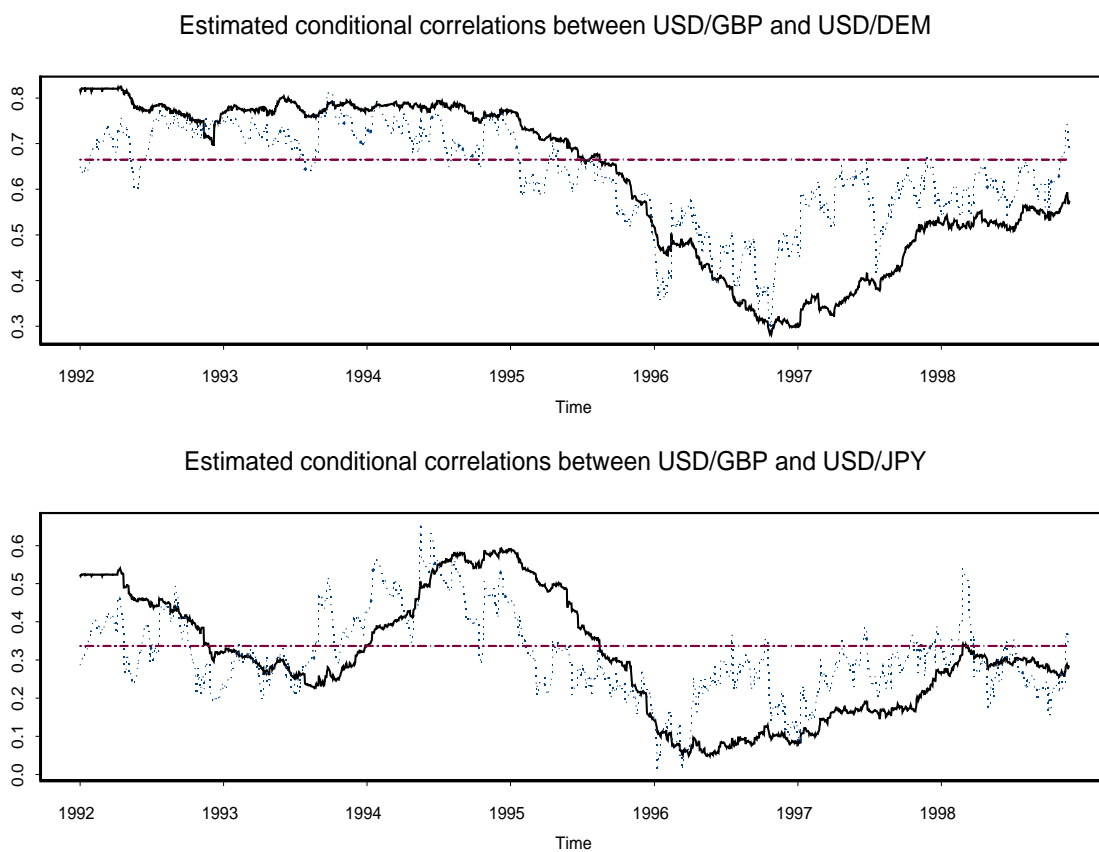


Figure 2: Conditional correlation dynamics between USD/GBP and USD/DEM (top) and between USD/GBP and USD/JPY (bottom) during the entire sample beginning January 2, 1992 and ending September 13, 1999. Conditional correlations are estimated using the RW-TACC-GARCH(1,1) model (solid line), the DCC(1,1)-GARCH(1,1) 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. $\Delta$	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  $\Delta_t$  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<sup>th</sup> 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	
	$\alpha_0$	$\alpha_1$	$\beta$	$a_0$	$a_1$
USD/GBP	0.0022 (0.0032)	0.0359* (0.0124)	0.9570* (0.0190)	-0.0093 (0.0135)	0.0129 (0.0290)
USD/DEM	0.0054 (0.0065)	0.0347* (0.0129)	0.9516* (0.0248)	0.0087 (0.0158)	-0.0025 (0.0266)
USD/JPY	0.0141 (0.0097)	0.0639* (0.0204)	0.9034* (0.0363)	0.0049 (0.0162)	0.0229 (0.0283)
USD/ITL	0.0028 (0.0036)	0.0596* (0.0177)	0.9350* (0.0197)	0.0007 (0.0134)	-0.0032 (0.0279)
USD/FRF	0.0037 (0.0071)	0.0345* (0.0118)	0.9548* (0.0275)	0.0043 (0.0152)	0.0060 (0.0283)
USD/NLG	0.0060 (0.0089)	0.0355* (0.0141)	0.9491* (0.0332)	0.0085 (0.0142)	-0.0089 (0.0278)
Eq. weighted Portf. $\Delta$	0.0016 (0.0048)	0.0311* (0.0107)	0.9633* (0.0290)	0.0059 (0.0127)	-0.0027 (0.0276)
Model	Cond. corr. structure		Cond. corr. parameters		
	$\mathcal{R}_k$		$\lambda_k$		
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)
CCC-FGD	100 (6/6)	22.222 (8/36)
DCC(1,1)-GARCH(1,1)	100 (6/6)	16.667 (6/36)
DCC(1,1)-FGD	83.3 (5/6)	11.111 (4/36)
RW-ACC-GARCH(1,1)	83.3 (5/6)	22.222 (8/36)
RW-ACC-FGD	83.3 (5/6)	11.111 (4/36)
RW-TACC-GARCH(1,1)	83.3 (5/6)	22.222 (8/36)
RW-TACC-FGD	83.3 (5/6)	11.111 (4/36)

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<sup>th</sup> 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
CCC-FGD	1499.1	0.3543	0.7391	0.0339	569.74	0.3102	0.5708	0.0090
DCC(1,1)-GARCH(1,1)	1266.4	0.3549	0.7381	0.0376	191.13	0.3079	0.5707	0.0066
DCC(1,1)-FGD	1184.7	0.3548	0.7380	0.0378	185.06	0.3077	0.5706	0.0069
RW-ACC-GARCH(1,1)	1551.8	0.3592	0.7397	0.0371	-50.354	0.3061	0.5702	0.0097
RW-ACC-FGD	1458.7	0.3593	0.7395	0.0376	-68.499	0.3057	0.5689	0.0105
RW-TACC-GARCH(1,1)	1522.5	0.3589	0.7397	0.0372	-2.6861	0.3059	0.5702	0.0096
RW-TACC-FGD	1423.9	0.3590	0.7395	0.0377	-10.795	0.3061	0.5699	0.0105

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	Type of test	
		t-type	sign-type
DCC(1,1)-GARCH(1,1)	CCC-GARCH(1,1)	-2.3566 (0.0092)	8.2609 (0)
RW-ACC-GARCH(1,1)	CCC-GARCH(1,1)	-2.3886 (0.0085)	17.009 (0)
RW-TACC-GARCH(1,1)	CCC-GARCH(1,1)	-2.0545 (0.0199)	21.492 (0)
RW-ACC-GARCH(1,1)	DCC(1,1)-GARCH(1,1)	-2.1897 (0.0143)	14.246 (0)
RW-TACC-GARCH(1,1)	DCC(1,1)-GARCH(1,1)	-0.9175 (0.1794)	10.074 (0)
RW-ACC-GARCH(1,1)	RW-TACC-GARCH(1,1)	-1.0258 (0.1525)	-2.2162 (0.0133)

Model 1	Model 2	Type of test	
		t-type	sign-type
CCC-FGD	CCC-GARCH(1,1)	-1.8201 (0.0344)	2.4652 (0.0068)
DCC(1,1)-FGD	DCC(1,1)-GARCH(1,1)	-1.3303 (0.0917)	-1.0482 (0.1473)
RW-ACC-FGD	RW-ACC-GARCH(1,1)	-0.5451 (0.2928)	-1.9991 (0.0228)
RW-TACC-FGD	RW-TACC-GARCH(1,1)	-0.3494 (0.3634)	-1.0782 (0.1405)

Model 1	Model 2	Type of test	
		t-type	sign-type
DCC(1,1)-GARCH(1,1)	CCC-FGD	-1.6205 (0.0526)	6.3941 (0)
RW-ACC-GARCH(1,1)	CCC-FGD	-2.3487 (0.0094)	13.168 (0)
RW-TACC-GARCH(1,1)	CCC-FGD	-2.0684 (0.0193)	17.064 (0)

Table 5: Testing differences of multivariate OS-NL performance terms between Model 1 and Model 2. Values of t-type and sign-type test statistics adapted to the case of dependent observations and corresponding  $P$ -values (between parentheses) are shown.

Model	IS-			OS-		
	MAE	RMSE	R <sup>2</sup>	MAE	RMSE	R <sup>2</sup>
CCC-GARCH(1,1)	0.3052	0.6127	0.0431	0.2500	0.4373	0.0210
CCC-FGD	0.3025	0.6067	0.0439	0.2459	0.4330	0.0208
DCC(1,1)-GARCH(1,1)	0.3044	0.6067	0.0443	0.2437	0.4341	0.0169
DCC(1,1)-FGD	0.3044	0.6066	0.0449	0.2440	0.4339	0.0175
RW-ACC-GARCH(1,1)	0.3083	0.6075	0.0446	0.2422	0.4336	0.0199
RW-ACC-FGD	0.3083	0.6074	0.0451	0.2426	0.4334	0.0213
RW-TACC-GARCH(1,1)	0.3081	0.6075	0.0447	0.2421	0.4337	0.0206
RW-TACC-FGD	0.3081	0.6073	0.0452	0.2424	0.4335	0.0214

Table 6: In-sample and out-of-sample goodness of fit results of the different models for the equally weighted portfolio  $\Delta$  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	Type of test	
		t-type	sign-type
DCC(1,1)-GARCH(1,1)	CCC-GARCH(1,1)	-2.2369 (0.0127)	2.4519 (0.0071)
RW-ACC-GARCH(1,1)	CCC-GARCH(1,1)	-2.5227 (0.0058)	3.1374 (0.0009)
RW-TACC-GARCH(1,1)	CCC-GARCH(1,1)	-3.8844 ( $\approx 10^{-5}$ )	3.3581 ( $\approx 10^{-4}$ )
RW-ACC-GARCH(1,1)	DCC(1,1)-GARCH(1,1)	-0.8662 (0.1932)	0.2481 (0.4020)
RW-TACC-GARCH(1,1)	DCC(1,1)-GARCH(1,1)	-0.9819 (0.1631)	0.4477 (0.3272)
RW-TACC-GARCH(1,1)	RW-ACC-GARCH(1,1)	-2.0599 (0.0197)	5.8092 (0)

Model 1	Model 2	Type of test	
		t-type	sign-type
CCC-FGD	CCC-GARCH(1,1)	-1.3902 (0.0822)	2.0193 (0.0217)
DCC(1,1)-GARCH(1,1)	CCC-FGD	-0.8623 (0.1943)	1.0345 (0.1505)
RW-ACC-GARCH(1,1)	CCC-FGD	-1.8189 (0.0345)	2.2758 (0.0114)
RW-TACC-GARCH(1,1)	CCC-FGD	-1.8536 (0.0319)	2.7266 (0.0032)

Table 7: Testing differences of univariate OS-MAE performance terms between Model 1 and Model 2 for the equally weighted portfolio  $\Delta$ . Values of t-type and sign-type test statistics adapted to the case of dependent observations and the corresponding  $P$ -values (between parentheses) are shown.

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
CCC-FGD	0.0607	0.0142	0.1687	0.2425
DCC(1,1)-GARCH(1,1)	0.0648	0.0182*	0.0754	0.1887
DCC(1,1)-FGD	0.0628	0.0162	0.1724	0.5329
RW-ACC-GARCH(1,1)	0.0688	0.0162	0.1700	0.5329
RW-ACC-FGD	0.0668	0.0142	0.0753	0.3175
RW-TACC-GARCH(1,1)	0.0688	0.0162	0.1700	0.5329
RW-TACC-FGD	0.0628	0.0162	0.1724	0.5329

Table 8: 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  $\Delta$  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.