Estimating and forecasting volatility with large scale models: theoretical appraisal of professionals' practice

Paolo Zaffaroni *

Imperial College London

This draft: 27th April 2006

*Address correspondence to:
Tanaka Business School, Imperial College London, South Kensington Campus, SW7 2AZ London, tel. + 44 207 594 9186, email p.zaffaroni@imperial.ac.uk

Abstract

This paper examines the way in which GARCH models are estimated and used for forecasting by practitioners. Although it permits sizable computational gains and provide a simple way to impose positive semi-definitiveness of multivariate version of the model, we show that this approach delivers non-consistent parameter' estimates. The novel theoretical result is corroborated by a set of Montecarlo exercises. Various empirical applications suggest that this could cause, in general, unreliable forecasts of conditional volatilities and correlations.

Keywords: GARCH, *Riskmetrics*^{TM}, estimation, forecasting, multivariate volatility models.

1 Introduction

Accurate forecasting of volatility and correlations of financial asset returns is essential for optimal asset allocation, managing portfolio risk, derivative pricing and dynamic hedging. Volatility and correlations are not directly observable but can only be estimated using historical data on asset returns. Financial institutions typically face the problem of estimating time-varying conditional volatilities and correlations for a large number of assets. Moreover, fast and computationally efficient methods are required. Therefore, parametric models of changing volatility are those most commonly used, rather than semi- and non-parametric methods. In particular, the autoregressive conditional heteroskedasticity (ARCH) model of Engle (1982), and the generalized ARCH (GARCH) of Bollerslev (1986), represent the most relevant paradigms. GARCH models are easy to estimate and fit financial data remarkably well (see Andersen and Bollerslev (1998)). In fact, GARCH models do account for several of the empirical regularities of asset returns (see Bollerslev, Engle, and Nelson (1994)), in particular dynamic conditional heteroskedasticity.

The popularity of GARCH models among practitioners in part stems from their close analogies with linear time series models such as autoregressive integrated moving average models (ARIMA), as well as with other, a-theoretical, models such as the exponentially weighted moving average (EWMA) model. Precisely by exploiting such analogies has permitted a feasible and computationally fast method for evaluating the conditional time-varying covariance matrix for a large number of assets, of the order of the hundreds. This method, which can be viewed as a highly restricted multivariate GARCH, has been popularized under the name of *RiskMetrics* ^{TM} approach (see J.P.Morgan/Reuters (1996)). In this paper we shall call this the *common* approach, acknowledging that it has been the dominant paradigm used by most financial analysts in the last years (see J.P.Morgan/Reuters (1996) and Litterman and Winkelmann (1998)).

Given the widespread evidence of practitioners using the *common* approach, this paper examines its theoretical underpinnings and effective performance. Surprisingly, very little theoretical research has been carried out on this topic. A notable exception is Cheng, Fan, and Spokoiny (2003) who nest the *common* approach within a wide class of filtering problems. They show that, under mild conditions, the filtering performance of the *common* approach does not depend on whether one uses (observed) square returns rather than the (unobserved) volatility process. This paper focuses instead on the estimation part of the filtering. Its main contribution is to show how the estimation method, embedded within the *common* approach, delivers non consistent estimates of the model parameters. A Montecarlo exercise describes the finite-sample properties of the estimator, indicating that its poor performance does not only arise asymptotically. Consequently, misleading forecasts are likely to occur. More importantly, conditional cross-covariances and correlations are poorly estimated, possibly leading to unexpected risk exposure when the estimated conditional covariance matrix is used to calculate dynamic hedge-ratios, Value-at-Risk performance and mean-variance efficient portfolios.

The *common* approach is frequently carried out without preliminary estimation, with parameters fixed a priori. When a change in the dynamic pattern of the data is likely to occur, calibrated parameters values must change accordingly. New estimates are needed in such circumstances and this is troublesome for the *common* approach. The impossibility to estimate parameters' model all depends on the particular, albeit attractively simple, specification that characterizes the *common* approach.

Adopting the *common* approach contrasts with the use of correctly specified GARCH models which we will be referring to as the *correct* approach. Practical applications of the *correct* approach for large scale problems (involving a large number of assets) is limited by the large number of parameters involved. As a consequence, the several proposed versions of multivariate GARCH models entail strong forms of parametric simplification, in order to achieve computational feasibility. Recent advances include the orthogonal GARCH model of Alexander (2001), the dynamic conditional correlation (DCC) model of Engle (2002), which generalizes the constant conditional correlation model of Bollerslev (1990), the regime-switching DCC of Pelletier (2002) and the averaged conditional correlations of Audrino and Barone-Adesi (2004). Bauwens, Laurent, and Rombouts (2003) provides a complete survey of this literature.

This paper proceeds as follows. Section 2 presents both the univariate *common* and *correct* approach. In particular, Section 2.1 describes the way in which univariate GARCH models are routinely specified and estimated by practitioners. Section 2.2 recalls the correct specification of GARCH models and related estimation issue. Section 2.3 looks at the small-sample properties of the estimation nested within the *common* approach by means of a set of Montecarlo exercises. A comparison of the predictive ability of the two approaches is described in Section 2.4, based on the Olsen's data set of the spot Mark/Dollar foreign exchange rate. Multivariate models are examined in Section 3, which also proposes two further empirical illustrations based on the Olsen's data set and on the Standard & Poor's 500 industry indexes. Concluding remarks are in Section 4. Section 5 contains a mathematical appendix.

2 Univariate case

2.1 Common approach

Let P_t be the speculative price of a generic asset at date t and define the continuously compounded one-period rate of return as $r_t = \ln(P_t/P_{t-1})$. To focus on the volatility dynamics, assume for the sake of simplicity that the r_t are martingale differences:

$$E(r_t \mid \mathcal{F}_{t-1}) = 0, \tag{1}$$

where $\mathcal{F}_t$ defines the sigma-algebra induced by the r_s, $s \leq t$. The simplest estimator of the conditional variance $E(r_t^2 \mid \mathcal{F}_{t-1}) = \sigma_t^2$ is the weighted rolling estimator, with window of length n:

$$\hat{\sigma}_t^2 = \sum_{s=1}^{n} w_{s(n)} r_{t-s}^2 \tag{2}$$

where the weights $w_{s(n)}$ satisfy

$$w_{s(n)} \geq 0, \quad \lim_{n\to\infty} \sum_{s=1}^{n} w_{s(n)} = 1,$$

(see J.P.Morgan/Reuters (1996, Table 5.1) and Litterman and Winkelmann (1998, eq.(1)) among others). $\hat{\sigma}_t^2$ is a function of n but we will not make this explicit for simplicity's sake. Important particular cases of (2) are the equally weighted estimator, for $w_{s(n)} = 1/n$, and the exponentially weighted estimator, for

$$w_{s(n)} = (1 - \lambda_0)\, \lambda_0^{s-1}, \tag{3}$$

for constant $0 < \lambda_0 < 1$, known as the *decay factor*. The weights (3) yield the popular EWMA estimator

$$\hat{\sigma}_t^2 = (1 - \lambda_0) \sum_{s=1}^{n} \lambda_0^{s-1} r_{t-s}^2. \tag{4}$$

The practical appeal of the EWMA estimator (4) lies in the fact that, by suitably choosing λ_0, the estimate will be more sensitive to newer observations than to older observations (cf. J.P.Morgan/Reuters (1996, p.80) and Litterman and Winkelmann (1998, p.15)). Computationally, EWMA is not more burdensome than simple averages, thanks to the recursion

$$\hat{\sigma}_t^2 = \lambda_0 \, \hat{\sigma}_{t-1}^2 + (1 - \lambda_0) r_{t-1}^2,$$

where the initial condition $\hat{\sigma}_{t-n}^2$ implies another term $\lambda_0^n \hat{\sigma}_{t-n}^2$ on the right hand side of (4). For practical implementation,

$$\hat{\sigma}_t^2 = \frac{(1 - \lambda_0)}{(1 - \lambda_0^n)} \sum_{s=1}^{n} \lambda_0^{s-1} r_{t-s}^2. \tag{5}$$

is used, rather than (4), ensuring that $\sum_{s=1}^{n} w_{s(n)} = 1$ for any finite n. (5) implies the recursion

$$\hat{\sigma}_t^2 = \lambda_0 \, \hat{\sigma}_{t-1}^2 + \frac{(1 - \lambda_0)}{(1 - \lambda_0^n)} r_{t-1}^2 - \frac{(1 - \lambda_0)}{(1 - \lambda_0^n)} \lambda_0^{n-1} r_{t-n-1}^2.$$

Typically, the initialization of such recursion is based on the sample variance of a pre-sample of data. Alternatively, assuming to observe a sample $r_1, ..., r_T$ of data and for a given n, one can evaluate (5) for $t \leq n$ with an expanding window (replacing n with $t - 1$), and returning to (5) when $t > n$.

When the weights $w_{s(n)}$ vary suitably with n, the rolling estimator, though a-theoretical, can be justified as a non-parametric estimator of the conditional variance. This rules out the possibility that $\hat{\sigma}_t^2$ nests the EWMA (4) which, however, is closely related to parametric time series models, such as GARCH. The weights of the (rolling) EWMA (23) vary with n but without converging towards zero as n grows to infinity, again ruling out the non-parametric interpretation.

The r_t are said to obey the GARCH$(1, 1)$ model when satisfying

$$r_t = z_t \sigma_t, \tag{6}$$

$$\sigma_t^2 = \omega_0 + \alpha_0 r_{t-1}^2 + \beta_0 \sigma_{t-1}^2 \;\; a.s. \tag{7}$$

where $a.s.$ stands for almost surely. The $\{z_t\}$ represent an independent and identically distributed ($i.i.d.$) sequence with mean zero and unit variance. Equation (6) then satisfies (1). The coefficients ω_0, α_0 and β_0 must be non-negative for σ_t^2 to be well defined (cf. Bollerslev (1986)). When

$$\alpha_0 + \beta_0 = 1, \tag{8}$$

one obtains the integrated GARCH$(1,1)$ (henceforth IGARCH$(1,1)$), by replacing (7) with

$$\sigma_t^2 = \omega_0 + \alpha_0 r_{t-1}^2 + (1-\alpha_0)\sigma_{t-1}^2 \ \ a.s. \tag{9}$$

Recursive substitution (n times) in (9) yields

$$\sigma_t^2 = \omega_0 \left(\frac{1-(1-\alpha_0)^{n-1}}{\alpha_0} \right) + (1-\alpha_0)^n \sigma_{t-n}^2 + \alpha_0 \sum_{s=1}^{n} (1-\alpha_0)^{s-1} r_{t-s}^2. \tag{10}$$

Imposing

$$\omega_0 = 0 \tag{11}$$

IGARCH$(1,1)$ (10) and EWMA (4) coincide.

GARCH have, moreover, a close connection with ARMA (cf. Bollerslev (1986)). In fact, setting

$$\nu_t = r_t^2 - \sigma_t^2 = (z_t^2 - 1)\sigma_t^2,$$

GARCH$(1,1)$ has an ARMA$(1,1)$ representation

$$r_t^2 = \omega_0 + (\alpha_0 + \beta_0) r_{t-1}^2 + \nu_t - \beta_0 \nu_{t-1}.$$

Note that the ν_t are martingale differences (not necessarily with finite variance). Under (8) one gets an ARIMA$(0,1,1)$

$$r_t^2 = \omega_0 + r_{t-1}^2 + \nu_t - (1-\alpha_0)\nu_{t-1}. \tag{12}$$

Therefore, the squares r_t^2 display a unit root with non-negative drift. It is well known that for standard, linear, unit root models, with positive drift, the series diverges $a.s.$ to infinity. This suggested to focus on IGARCH$(1,1)$ models satisfying (11) (cf. J.P.Morgan/Reuters (1996, eq. (5.37)) and Litterman and Winkelmann (1998, eq. (4))).

For estimation, let us introduce the parameterized model

$$\sigma_t^2(\alpha) = \alpha r_{t-1}^2 + (1-\alpha)\sigma_{t-1}^2(\alpha), \;\; \sigma_0^2(\alpha) = 0.$$

with $\alpha \in (0,1)$. Given the emphasis on forecasting, the parameter α_0 is routinely estimated by least squares (LS), i.e. minimizing the mean square error (MSE) of the predictions:

$$MSE_T(\alpha) = \frac{1}{T}\sum_{t=2}^{T}\left(r_t^2 - \sigma_t^2(\alpha)\right)^2 \tag{13}$$

yielding

$$\hat{\alpha}_T^{lse} = \text{argmin}_{\alpha\in[\underline{\alpha},\bar{\alpha}]} MSE_T(\alpha), \tag{14}$$

for given $0 < \underline{\alpha} < \bar{\alpha} < 1$ (cf. J.P.Morgan/Reuters (1996, section 5.3.2.1)). The $\underline{\alpha}, \bar{\alpha}$ can be chosen arbitrarily, yet the minimization in (14) necessarily requires the definition of a compact interval, bounded away from both zero and unity.

We now show that specifying GARCH models by imposing (11) has dramatic implications. Therefore, exploiting the analogies of GARCH models with both EWMA and ARIMA, could lead to misleading inferences and forecasting results.

The crucial condition, ensuring strict stationarity and ergodicity, for GARCH$(1,1)$ is

$$E\,\ln(\beta_0 + \alpha_0 z_t^2) < 0. \tag{15}$$

(see Nelson (1990, Theorem 1 and 2)), independently of ω_0. It is easy to see that IGARCH$(1,1)$, with $E\,z_t^2 = 1$, satisfies (15) (see Nelson (1990, p.321))). Therefore, in contrast to standard linear unit root models, such as the classical random walk, the r_t^2 are strictly stationary and do not exhibit any type of explosive behaviour. This outcome is linked to the fact that, unlike the standard linear framework, the innovations ν_t in (12) are not independent of the r_s^2 $(s < t)$.

Consider now representation (10) with $n = t$:

$$\sigma_t^2 = \omega_0\left(\frac{1-(1-\alpha_0)^{t-1}}{\alpha_0}\right) + (1-\alpha_0)^t\sigma_0^2 + \alpha_0\sum_{s=1}^{t}(1-\alpha_0)^{s-1}r_{t-s}^2.$$

The most dramatic effect of imposing (11) on IGARCH$(1,1)$ is that

$$\sigma_t^2 \to 0 \;\; a.s. \text{ for } t \to \infty. \tag{16}$$

when (11) holds (see Nelson (1990, Theorem 1)). The impact of this result can be viewed by means of a simulation, displayed in Figure 1 (left panel), based on setting $\alpha = 0.95$. It turns out that the smaller α is, the sooner the conditional variance will converge towards zero. Note that once $\sigma_t^2 = 0$ for some t, then $\sigma_s^2 = 0$, and thus $r_{s+1} = 0$ for any $s \geq t$, as from (9)

$$\sigma_t^2 = \sigma_{t-1}^2 \left(\alpha_0 z_{t-1}^2 + (1 - \alpha_0)\right),$$

so zero represents an *absorption state* for the process.

This asymptotic degenerateness might, nevertheless, not be important for estimation and forecasting over short horizon. It turns out that this statement is false, as indicated by the following result, whose proof is in the Appendix.

Theorem 1 *For IGARCH*$(1,1)$*, when (11) holds,*
(a)

$$\sup_{\alpha \in [\underline{\alpha}, \bar{\alpha}]} MSE_T(\alpha) \to 0 \;\; a.s. \quad for\ T \to \infty \tag{17}$$

for any $0 < \underline{\alpha} < \bar{\alpha} < 1$.
(b) Under the same conditions

$$\inf_{\alpha \in [\underline{\alpha}, \bar{\alpha}]} E\left(MSE_T(\alpha)|\sigma_0^2\right) \to \infty \quad for\ T \to \infty. \tag{18}$$

Remarks (i) The first part of Theorem 1 implies that the LS estimator $\hat{\alpha}_T^{lse}$ is non-consistent for α_0 and is therefore meaningless. In fact (17) says that the MSE, non negative by construction, is (asymptotically) minimized for any value of $\alpha \in (0,1)$. It follows that $\hat{\alpha}_T^{lse}$ is globally (asymptotically) unidentified. One might wonder whether this is merely the symptom of a different rate of convergence, namely whether $T^b MSE_T(\alpha)$ would converge to a non-random expression for some $b \neq 0$, uniquely minimized at α_0. Simple inspection of the proof of Theorem 1 shows that $T\, MSE_T(\alpha)$ is bounded $a.s.$ and non zero but the limit is random and model identification fails.
(ii) Cheng, Fan, and Spokoiny (2003) show that no differences arises when estimating the *common* approach through (13) with respect to the ideal, but unfeasible, case when the true volatility is observed and one minimizes $T^{-1}\sum_{t=2}^{T}\left(\sigma_t^2 - \sigma_t^2(\alpha)\right)^2$.
(iii) Establishing almost sure convergence plays an important role. In fact, taking the (conditional) expectation of $MSE_T(\alpha)$ yields (18) which seems

to contradict Theorem 1. This (apparent) contradiction is a by-product of the non stationarity of IGARCH. Basically, the asymptotic behaviour of the average of moments does not reflect the asymptotic behaviour of the average of the underlying random variables. Recall that for IGARCH$(1,1)$ condition (15) holds.

2.2 Correct approach

Under (15) IGARCH$(1,1)$ are strictly stationary and ergodic. This implies that, despite the ARIMA$(0,1,1)$ representation, there is no harm in imposing

$$\omega_0 > 0. \tag{19}$$

Indeed, Nelson (1990, Theorem 2) has shown that under (15) and (19)

$$\sigma_t^2 - {}_u\sigma_t^2 \to 0 \quad a.s. \quad \text{for } t \to \infty,$$

setting

$${}_u\sigma_t^2 = \frac{\omega_0}{\alpha_0} + \alpha_0 \sum_{s=1}^{\infty} (1-\alpha_0)^{s-1} r_{t-s}^2.$$

${}_u\sigma_t^2$ defines the *unconditional* process, in contrast to σ_t^2 which defines the *conditional* process as it depends on the initial condition σ_{t-n}^2. Moreover, it has been shown that the ${}_u\sigma_t^2$ are strictly stationary and ergodic, with a well-defined non degenerate probability measure on $[\omega_0/\alpha_0, \infty)$. Figure 1 (middle panel), based on setting $\alpha = 0.95$, provides a typical sample path for the process. Now the model conditional variance is always bounded away from zero and the process is never degenerate. IGARCH$(1,1)$ are, however, not covariance stationary. In fact, under (8) the r_t have infinite variance although the sample path will not be explosive, thanks to the strict stationarity and ergodicity. For this reason, the IGARCH$(1,1)$ parameters ω_0 and α_0 cannot be estimated by LS. However, the model is well specified and can be estimated in various ways, the most common of which is by pseudo maximum likelihood (PML). The PML estimator (PMLE) is characterized by standard asymptotic statistical properties (see Lee and Hansen (1994) and Lumsdaine (1996)), and it is given by

$$(\hat{\omega}_T^{pmle}, \hat{\alpha}_T^{pmle}) = \text{argmin}_{\omega,\alpha \in [\underline{\alpha},\bar{\alpha}]\times[\underline{\omega},\bar{\omega}]} L_T(\omega,\alpha)$$

for constants $0 < \underline{\alpha} < \bar{\alpha} < 1$ and $0 < \underline{\omega} < \bar{\omega} < \infty$, setting

$$L_T(\omega, \alpha) = \frac{1}{T} \sum_{t=1}^{T} \ln \sigma_t^2(\omega, \alpha) + \frac{1}{T} \sum_{t=1}^{T} \frac{r_t^2}{\sigma_t^2(\omega, \alpha)} \tag{20}$$

with the parameterized conditional variance

$$\sigma_t^2(\omega, \alpha) = \omega + \alpha r_{t-1}^2 + (1 - \alpha)\sigma_{t-1}^2(\omega, \alpha).$$

2.3 Small-sample performance

Table 1 reports a Montecarlo exercise in order to evaluate the performance of the LS estimator $\hat{\alpha}_T^{lse}$ and of the PMLE $\hat{\alpha}_T^{pmle}$ used, respectively, to estimate the *common* and the *correct* approach. One can also compare the *common* and the *correct* approaches using the same estimator, in particular the PMLE. However, we feel that it is more relevant to compare the two models using the corresponding estimation procedure most frequently used by practitioners. Moreover, the asymptotic degeneratedness (16) would make the implementation of the PMLE problematic for the *common* approach.

Concerning the former, we simulated IGARCH$(1, 1)$ imposing (11), with $\alpha_0 = 0.05, 0.75$ and 0.95. We consider samples of length $15,000$ and estimate the model considering the first $1,000$ observations, the first $5,000$ observations and, lastly, all $15,000$ observations. The MSE is minimized by numerical methods with a MatLab code, starting from an arbitrary value equal to 0.5. Such choice should be completely irrelevant for a well specified model. The non-consistency of $\hat{\alpha}_T^{lse}$ clearly emerges when comparing the Montecarlo variances (column three) for different sample sizes, which do not vary with T. On the other hand, when looking at the Montecarlo means (column two), it seems that the LS estimator is reliable for large values of α_0. (For case $\alpha_0 = 0.05$ also the mean indicates the non consistency). These numerical results can be better evaluated by looking at the Montecarlo frequency distributions of the LS estimator, reported in Figure 2 (top three panels). It clearly appears that the LS estimator is both non-consistent and non-centered (biased) with respect to the true value. Interestingly, note how the behaviour of the LS estimator for case $\alpha_0 = 0.95$ (top right panel) is heavily influenced by the initial, very persistent, observations. As a result, the estimate is close to the true value although in reality this is independent of its asymptotic statistical properties. It is thus possible that a mis-specified model achieves a good forecasting performance. This possibility is investigated in section 2.4.

Columns seven to twelve of Table 1 report the small-sample properties of the PMLE, based on samples of size $T = 1,000,\ 5,000$ and $15,000$. Again the optimization started from the same arbitrary value of 0.5 for α. The estimates are all centered around the true value and their variance decreases as the sample size increases, at rate $1/T$. The average of (20) across replications, evaluated at the PMLE (column twelve of Table 1), does not significantly change with T. The empirical distribution of the estimates are reported in Figure 2 (bottom three panels).

2.4 Comparing forecasting performances (univariate)

Our theoretical result indicates the inherent difficulties for estimation of the *common* approach. We now present a simple empirical application aimed at providing some evidence on the forecasting implications of the theoretical result.

We compare the predictive capability of IGARCH$(1,1)$ models described in section 2.1 and 2.2, that is with and without condition (11). Hereafter, we denote the former as model 2 and the latter as model 1. We consider the Diebold and Mariano (1995) test of predictive accuracy

$$DM_S(L) = \frac{1}{\sqrt{A_S}} \sum_{s=1}^{S} \left(L(\sigma_s^2, {}^{(1)}\sigma^2_{s|s-p}) - L(\sigma_s^2, {}^{(2)}\sigma^2_{s|s-p}) \right) \tag{21}$$

where S defines the number of p-step ahead forecasts employed and $L(a,b)$ defines a generic loss function. ${}^{(1)}\sigma^2_{s|s-p}$ and ${}^{(2)}\sigma^2_{s|s-p}$ express the two, competing, p-step ahead forecasts of the true conditional volatility σ_s^2. The normalizing quantity A_S is a consistent estimate of the variance of the numerator of (21), robust to autocorrelation of unknown form (see Diebold and Mariano (1995, section 1.1) for details), such that under suitable regularity conditions $DM_S(L)$ converges in distribution to a standard normal (for $S \to \infty$), under the null hypothesis of equal forecasting performance

$$E\left(L(\sigma_s^2, {}^{(1)}\sigma^2_{s|s-p}) - L(\sigma_s^2, {}^{(2)}\sigma^2_{s|s-p}) \right) = 0.$$

We employ the well-known data set of Olsen & Ass., frequently used to compare predictive performance of volatility models since Andersen and Bollerslev (1998). In particular, the data consist of daily and intra-daily returns for the Mark/Dollar spot exchange rate (henceforth Mark/Dollar returns), from July 1, 1974, through September 30, 1992 - $4,573$ observations - for daily data, and from October 1, 1992, through September 30, 1993 - $74,880$ observations - for intra-daily data (five-minute returns). We then obtain 260 observations of the so-called realized volatility, obtained by summing squared intra-daily returns (288 intra-daily observations per day), as indicated by Andersen and Bollerslev (1998). In this exercise we assume that the realized volatility expresses true (daily) volatility, denoted by σ_t^2, with a negligible error.

The results are reported in Table 2. The data in levels (returns) have been preliminary filtered with an AR(1) model. The two volatility models are then estimated, first, using $4,573$ daily observations, then $4,573 + 1$ observations, and so on up to $4,573 + 259$ observations. West (1996) noted that the asymptotic distribution of $DM_S(L)$ might depend on the sample variability of parameter estimates. However he established that no effect arises when the length of the estimation sample dominates the length of the evaluation sample. Since in our case the former $(4,573)$ is nearly twenty times the latter (260), we proceed ignoring the effect of parameter uncertainty. The averages of the 260 different point estimates of α_0 for both models (estimates of ω_0 for model 1 are not reported for the sake of simplicity) are reported in column two and column three for model 2 and model 1 respectively. For both models estimation starts from an initial value $\alpha = 0.95$.

We consider two different types of loss function, the square-rooted absolute difference $L(a, b) =\mid a - b \mid^{\frac{1}{2}}$ and the loss function implicit in the Gaussian log likelihood $L(a, b) = \ln(b) + a/b$ (see Bollerslev, Engle, and Nelson (1994, section 7)), whose results are reported in columns $4-7$ and $8-11$ respectively. Four different forecasting horizons are considered: at 1 day, 1 week, 1 month and 6 months ahead. The forecasting function for model 1 (correctly specified IGARCH$(1, 1)$, without imposing (11)), is

$$^{(1)}\sigma^2_{t+p|t} = E(^{(1)}\sigma^2_{t+p}|\mathcal{F}_t) = p\omega_0 + \alpha_0 r_t^2 + (1 - \alpha_0)\sigma_t^2.$$

For model 2 (degenerate IGARCH$(1, 1)$, imposing (11)) the forecast function is constant and equal to

$$^{(2)}\sigma^2_{t+p|t} = E(^{(2)}\sigma^2_{t+p}|\mathcal{F}_t) = \alpha_0 r_t^2 + (1 - \alpha_0)\sigma_t^2,$$

for any $p \geq 1$. Obviously the preceding expressions are in practice evaluated at the estimated, rather than true, parameter values. The choice made for the loss functions reflects the estimation methods employed for model 1 and model 2 respectively. The square-rooted absolute difference loss function should potentially favor model 2 whereas the Gaussian likelihood loss function should favor model 1. Comparing the results for the two loss functions should avoid biases when assessing the forecasting performance of the competing models.

There are two main findings. First, the forecasting performance of model 2 is significantly worse than that of model 1 in most cases, or at most not significantly different from it, for short and medium-run horizons (1 day, 1 week and 1 month). By contrast, for longer horizon (6 months), model 2 outperforms model 1 for the square-root loss function and it is not significantly different from model 2 for the Gaussian loss function. This outcome is not completely surprising. In fact, the asymptotic degenerateness of model 2 implies a well-behaved forecasting function for all horizons, whereas for (the correctly specified) model 1 the forecasting function diverges to infinity, as the forecast horizon grows to infinity.

Second, the forecasting performance of model 2 is extremely dependent on the initial value for $\sigma_0^2(\alpha)$ chosen when estimating the model. The first four rows of Table 2 refer to different values for $\sigma_0^2(\alpha)$, all arbitrarily chosen, except for row four for which $\sigma_0^2(\alpha)$ has been estimated jointly with α. Nevertheless, even for this last case, the poor forecasting performance of model 2 emerges (except for the long horizon). In fact, different initial conditions imply extremely different point estimates for α, as indicated in the second column reporting the average of the $\hat{\alpha}_T^{lse}$, and thus different forecasts. The performance of the two models is comparable when setting $\sigma_0^2(\alpha) = 0.1$ in the estimation of model 2, yielding an average of the $\hat{\alpha}_T^{lse}$ equal to 0.881, not surprisingly close to the average of the $\hat{\alpha}_T^{pmle}$ of 0.851. Clearly, finding ex-ante a *good* value for $\sigma_0^2(\alpha)$, to estimate model 2, is not an attainable task in general. This should not, and in fact is not, an issue for the *correct* approach, with the effect of initial conditions being asymptotically negligible. The last three rows report the result obtained setting $\sigma_0^2(\alpha)$ equal to the sample variance of the data, and setting $\alpha = 0.94$. This is the value suggested by J.P.Morgan/Reuters (1996, p.100) and, presumably, close by to any other values used by practitioners. The last two rows consider the rolling EWMA (eq.(5)) with $n = 15, 30$ whereas row five considers the expanding window $n = t - 1$. For these case as well, no significative difference arises between the two competing models for all but the 6 months horizon.

3 Multivariate case

Typical situations faced by practitioners, such as optimal asset allocations and portfolio risk diversification, involve many assets. For instance, J.P.Morgan/Reuters (1996, p.97) considers the problem of estimating the conditional covariance matrix of 480 time series. In such circumstances, as discussed in Section 1, use of correctly specified (unrestricted) multivariate GARCH models is not a possibility and some form of restrictions are required in order to achieve computational feasibility. The *common* approach, instead, is always applicable in such situations, which motivates its widespread use for real time applications. Note that implementation and estimation of the multivariate *common* approach heavily relies on the results derived for univariate case, fully described in Section 2.1.

3.1 Common approach

Assuming that (1) holds for each asset and that there are m assets, the rolling estimator for the conditional covariance between the returns of asset i and asset j is

$$\hat{\sigma}_{ij,t} = \sum_{s=1}^{n} w_{s(n)} r_{i,t-s} r_{j,t-s}, \quad i,j = 1,...,m, \tag{22}$$

where $r_{i,t} = \ln(P_{i,t}/P_{i,t-1})$ and $P_{i,t}$ denotes the speculative price of asset i (see J.P.Morgan/Reuters (1996, Table 5.4) and Litterman and Winkelmann (1998, eq. (2))). Likewise, the EWMA is

$$\hat{\sigma}_{ij,t} = (1-\lambda_0) \sum_{s=1}^{n} \lambda_0^{s-1} r_{i,t-s} r_{j,t-s}, \tag{23}$$

whose recursive form is

$$\hat{\sigma}_{ij,t} = \lambda_0 \hat{\sigma}_{ij,t-1} + (1-\lambda_0) r_{i,t-1} r_{j,t-1}, \ \hat{\sigma}_{ij,t-n} = 0. \tag{24}$$

Note that the parameter λ_0 does not vary with i, j and, indeed, this choice guarantees that the $m \times m$ matrix $\hat{\mathbf{\Sigma}}_{\mathbf{t}} = [\hat{\sigma}_{ij,t}]$ $(1 \le i, j \le m)$, solution of

$$\hat{\mathbf{\Sigma}}_{\mathbf{t}} = \lambda_0 \hat{\mathbf{\Sigma}}_{\mathbf{t-1}} + (1-\lambda_0) \mathbf{r}_{\mathbf{t-1}} \mathbf{r}'_{\mathbf{t-1}}, \ \hat{\mathbf{\Sigma}}_{\mathbf{0}} = 0,$$

is positive semi-definite, setting $\mathbf{r_t} = (r_{1,t}, ..., r_{m,t})'$ (cf. J.P.Morgan/Reuters (1996, p. 97)). Practical implementation of (23) is often based on

$$\hat{\sigma}_{ij,t} = \frac{(1-\lambda_0)}{(1-\lambda_0^n)} \sum_{s=1}^{n} \lambda_0^{s-1} r_{i,t-s} r_{j,t-s}, \tag{25}$$

with the recursion (in matrix notation)

$$\hat{\mathbf{\Sigma}}_\mathbf{t} = \lambda_0 \hat{\mathbf{\Sigma}}_\mathbf{t-1} + \frac{(1-\lambda_0)}{(1-\lambda_0^n)} \mathbf{r_{t-1} r'_{t-1}} - \frac{(1-\lambda_0)}{(1-\lambda_0^n)} \lambda_0^{n-1} \mathbf{r_{t-n-1} r'_{t-n-1}}, \tag{26}$$

with the same initialization issues of the univariate case as discussed after eq. (5).

First we establish the analogy of EWMA (23) with multivariate GARCH(1, 1). The latter, in its most general specification, is

$$\mathbf{r_t} = \mathbf{\Sigma}_\mathbf{t}^{\frac{1}{2}} \mathbf{z_t}, \tag{27}$$

$$\text{vech}(\mathbf{\Sigma_t}) = \mathbf{\Omega_0} + \mathbf{A_0} \text{vech}(\mathbf{\Sigma_{t-1}}) + \mathbf{B_0} \text{vech}(\mathbf{r_{t-1} r'_{t-1}}), \tag{28}$$

where vech(.) denotes the column stacking operator of the lower portion of a symmetric matrix, $\mathbf{z_t} = (z_{1,t}, ..., z_{m,t})'$ is an m-valued $i.i.d.$ sequence with $E\mathbf{z_t} = 0$ and $E\mathbf{z_t z'_t} = \mathbf{I_m}$ where $\mathbf{I_m}$ is the identity matrix of dimension $m \times m$, $\mathbf{\Omega_0}$ is an $m(m+1)/2 \times 1$ vector and $\mathbf{A_0}$, $\mathbf{B_0}$ are $m(m+1)/2 \times m(m+1)/2$ matrices of coefficients (see Bollerslev, Engle, and Wooldridge (1988, eq. 4)). An appealing feature of (28) is that it produces time-varying conditional correlations, given by (at the one-step-ahead horizon)

$$\rho_{ij,t} = \frac{E(r_{i,t} r_{j,t} \mid \mathcal{F}_{t-1})}{\sqrt{E(r_{i,t}^2 \mid \mathcal{F}_{t-1}) E(r_{j,t}^2 \mid \mathcal{F}_{t-1})}} = \frac{\sigma_{ij,t}}{\sigma_{i,t} \sigma_{j,t}}, \tag{29}$$

assuming that (1) holds.

Imposing diagonality and constancy of the diagonal terms of $\mathbf{A_0}$, $\mathbf{B_0}$ yields

$$\mathbf{A_0} = \alpha_\mathbf{0} \mathbf{I_{m(m+1)/2}}, \quad \mathbf{B_0} = \beta_\mathbf{0} \mathbf{I_{m(m+1)/2}}.$$

Further imposing (8), one obtains the multivariate IGARCH(1, 1):

$$\sigma_{ij,t} = \omega_{ij,0} + \alpha_0 r_{i,t-1} r_{j,t-1} + (1-\alpha_0) \sigma_{ij,t-1}, \quad i, j = 1, ..., m. \tag{30}$$

Despite the $r_{i,t}$ are not covariance stationary, the conditional correlations $\rho_{ij,t}$ are well defined. Finally, imposing

$$\omega_{ij,0} = 0, \quad i, j = 1, ..., m, \tag{31}$$

and

$$\sigma_{ij,t-n} = 0, \quad i, j = 1, ..., m, \tag{32}$$

the EWMA (24) and the multivariate IGARCH(1, 1) (30) coincide. In matrix notation, the latter is

$$\mathbf{r_t} = \mathbf{\Sigma_t^{\frac{1}{2}} z_t}, \tag{33}$$

$$\mathbf{\Sigma_t} = \alpha_0 \mathbf{r_{t-1} r'_{t-1}} + (1 - \alpha_0)\mathbf{\Sigma_{t-1}}. \tag{34}$$

Substituting (33) into (34) yields

$$\mathbf{\Sigma_t} = \mathbf{\Sigma_{t-1}^{\frac{1}{2}} A_{t-1} \Sigma_{t-1}^{\frac{1}{2}}}, \tag{35}$$

setting $\mathbf{A_t} = (\alpha_0 \mathbf{I_m} + (1 - \alpha_0)\mathbf{z_t z'_t})$. From (35) it follows that when $\mathbf{\Sigma_t} = 0$ then $\mathbf{\Sigma_s} = 0$ and $\mathbf{r_s} = 0$ for any $s > t$, in analogy with the univariate case, suggesting that

$$\text{tr}(\mathbf{\Sigma_t}) \to 0 \;\; a.s. \text{ for } t \to \infty, \tag{36}$$

where tr(·) is the trace operator. The right panel of Figure 1 reports the results of a simulation exercise showing that (36) really does occur.

As for the univariate case, the asymptotic degenerateness of the process causes numerous problems. First, note that when (31) is imposed, the conditional correlations (29) are no longer well-defined. For estimation purposes, one can generalize the procedure of section 2.1, and estimate α_0 by minimizing the multivariate MSE

$$\hat{\alpha}_T^{lse} = \text{argmin}_{\alpha \in [\underline{\alpha}, \bar{\alpha}]} \frac{1}{T} \sum_{t=2}^{T} \| \mathbf{r_t r'_t} - \mathbf{\Sigma_t}(\alpha) \|^2$$

setting

$$\mathbf{\Sigma_t}(\alpha) = (1 - \alpha)\mathbf{\Sigma_{t-1}}(\alpha) + \alpha \mathbf{r_{t-1} r'_{t-1}}, \;\; \mathbf{\Sigma_0}(\alpha) = 0,$$

where $\| \cdot \|$ indicates the Euclidean norm.

However, rather than using the LS estimator $\hat{\alpha}_T^{lse}$, practitioners adopt a two-stage approach. First, they estimate univariate EWMA for each asset by minimizing the univariate MSE (cf. (13)), yielding $\hat{\alpha}_T^{lse,(i)}$ $(i = 1, ..., m)$. They then estimate α_0 by a weighted average of the former, yielding

$$\tilde{\alpha}_T = \sum_{i=1}^{m} \phi_T^{(i)} \, \hat{\alpha}_T^{lse,(i)},$$

with weights that penalize assets whose estimated coefficients have a large MSE:

$$\phi_T^{(i)} = \frac{(\theta_T^{(i)})^{-1}}{\sum_{j=1}^{m} (\theta_T^{(j)})^{-1}}, \quad i = 1, ..., m,$$

setting

$$\theta_T^{(i)} = \frac{\sqrt{MSE_T(\hat{\alpha}_T^{(i)})}}{\sqrt{MSE_T(\hat{\alpha}_T^{(1)})} + ... + \sqrt{MSE_T(\hat{\alpha}_T^{(m)})}}, \quad i = 1, ..., m,$$

(see J.P.Morgan/Reuters (1996, section 5.3.2.2)).

As for the univariate case, it turns out that (31) implies that neither $\hat{\alpha}_T^{lse}$ nor the frequently used $\tilde{\alpha}_T$ is consistent for α_0. Concerning the latter, Theorem 1 applies directly since $\tilde{\alpha}_T$ is a weighted average of m estimates of α_0, each of which being non-consistent, and whose weights ϕ_T^i converge to a random limit as $T \to \infty$. An additional drawback of $\tilde{\alpha}_T$ is that it is, by construction, completely independent of the information stemming from the conditional cross-correlations characterizing the data.

3.2 Comparing forecasting performances (multivariate)

This section compares the predictive performance of the *common* approach versus the *correct* approach in a multivariate setting. We illustrate a bivariate $(m = 2)$ and a medium-scale multivariate $(m = 22)$ exercise. Given our emphasis on the theoretical result, we mirror section 2.4 and adopt a statistical approach rather than a decision-theoretic approach to forecast evaluation, such as in Pesaran and Zaffaroni (2004).

Regarding the bivariate specification, we employ the Mark/Dollar spot exchange rate series described in section 2.4 as well as the time series of the Yen/Dollar spot exchange rate. (The data are described in Andersen and Bollerslev (1998).) Considering only the observations which share the same trading periods yields $4,573$ daily observations - from July 1, 1974, through September 30, 1992 - and 74, 304 intra-daily (five-minute returns) observations - from October 1, 1992, through September 30, 1993. Table 3 describes the results. The return data have been first filtered with an AR(1) model. For this bivariate exercise we consider the loss functions

$$
\begin{aligned}
L_1(A,B) &= \mid a_{11} - b_{11} \mid^{\frac{1}{2}} + \mid a_{22} - b_{22} \mid^{\frac{1}{2}} + \mid a_{12} - b_{12} \mid^{\frac{1}{2}}, \\
L_2(A,B) &= \mid a_{12}/\sqrt{a_{11}a_{22}} - b_{12}/\sqrt{b_{11}b_{22}} \mid,
\end{aligned}
$$

for any pair of 2×2 symmetric matrices

$$
A = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}, B = \begin{pmatrix} b_{11} & b_{12} \\ b_{12} & b_{22} \end{pmatrix}.
$$

The first type of loss function generalizes the square-rooted absolute difference used in section 2.4. The second type of loss function compares (conditional) cross-correlations. The forecasting function for model 1 is, for $p \geq 1$,

$$
{}^{(1)}\mathbf{\Sigma_{t+p|t}} = p\tilde{\mathbf{\Omega}}_\mathbf{0} + \alpha_\mathbf{0}\mathbf{r_t}\mathbf{r'_t} + (\mathbf{1} - \alpha_\mathbf{0})\mathbf{\Sigma_t}.
$$

As for the univariate case, the forecast function of model 2 is constant and equal to

$$
{}^{(2)}\mathbf{\Sigma_{t+p|t}} = \alpha_\mathbf{0}\mathbf{r_t}\,\mathbf{r'_t} + (\mathbf{1} - \alpha_\mathbf{0})\mathbf{\Sigma_t},
$$

for any p. The results of Table 3 confirm that model 1 significantly outperforms model 2 in most cases. The results are less conclusive for longer horizons. The better performance of model 1 is particularly evident when comparing conditional cross-correlations (columns $8 - 11$). This has implications when one uses multivariate GARCH models to construct optimal (dynamic) hedge-ratios, so as to minimize portfolio risk. Second, we find the extreme sensibility of the estimates and forecasting performance of model 2 from initial conditions $\tilde{\mathbf{\Sigma}}_\mathbf{0}(\alpha) = \tilde{\sigma}_0^2(\alpha)\mathbf{I_2}$. Again, the performance of the two models is never significantly different when $\tilde{\sigma}_0^2(\alpha) = 0.1$. When deriving $\tilde{\sigma}_0(\alpha)$ jointly with $\tilde{\alpha}_T$, the performance of model 2 is significantly worse than that of model 1 for most cases. Finally, note that the estimates $\tilde{\alpha}_T$ are only marginally different from $\hat{\alpha}_T^{lse}$, as reported in Table 2. (In both cases, estimation started from $\alpha = 0.95$.) This is, again, a by-product of the poor statistical properties of the LS estimator. No significative differences arise when calibrating model 2 with $\alpha = 0.94$ and setting $\tilde{\mathbf{\Sigma}}_\mathbf{0}(\alpha)$ equal to the sample covariance matrix of the data. Summarizing, when comparing variances and the covariance, the results are very much similar to the univariate case of Table 2. However, when comparing conditional correlations, the *common* approach never outperforms the *correct* when λ_0 is calibrated. The *common* approach is markedly inferior, at all horizons, when the smoothing parameter is estimated.

For the multivariate exercise, we shall consider the 22 main industry indices of the Standard & Poor's 500 (source: Datastream) extracted from the S&P 500 industry price indices defined according to the Global Industry Classification Standard. Our data set covers the industry indices from 2nd January 1995 to 13th October 2003 ($T = 2291$ observation). Daily returns are computed as $r_{jt} = 100 \ln (P_{jt}/P_{jt-1})$, $j = 1, ..., 22$, where P_{jt} is the j^{th} price index. (For a description of the data, and of their statistical properties, we refer to Pesaran and Zaffaroni (2004).) We considered the generalization of the loss function $L_1(A, B)$ in (37) to 22×22 matrices, comparing $r_{it+p} r_{jt+p}$, $i, j = 1, ..., 22$, with their forecast based on the (i, j)th entry of ${}^{(\mathbf{1})}\mathbf{\Sigma_{t+p|t}}$. Returns have not been preliminary filtered with an AR(1) model since negligible time variation of the conditional mean is documented. Unlike the univariate and bivariate examples, it is unfeasible to estimate a well-specified (unrestricted) GARCH model of dimension 22×22 such as (27)-(28). The dynamic conditional correlation (DCC) of Engle (2002) and its asymmetric variation, namely the asymmetric DCC (ADCC) of Cappiello, Engle, and Sheppard (2002), appear superior at describing the dynamic properties of this data set, when compared with many other multivariate GARCH-type specifications (see Pesaran and Zaffaroni (2004)). We recall that the DCC(r,s,R,S) implies ${}^{(\mathbf{1})}\mathbf{\Sigma_t} = \mathbf{D_t R_t D_t}$ for a diagonal matrix $\mathbf{D_t}$, with a (square-rooted) GARCH(r, s) on the $(i, i)th$ entry, and, considering a simple specification ($R = S = 1$), $\mathbf{R_t}$ has $q_{hjt}/\sqrt{q_{hht}\, q_{jjt}}$ in its $(h, j)th$ position, setting $\mathbf{Q}_t = [q_{hjt}]_{h,j=1}^{22}$ for

$$\mathbf{Q}_t = \overline{\mathbf{Q}}\, (1 - \gamma_0 - \delta_0) + \gamma_0 \tilde{\mathbf{r}}_{t-1} \tilde{\mathbf{r}}'_{t-1} + \delta_0 \mathbf{Q}_{t-1},$$

$\tilde{\mathbf{r}}_t = \mathbf{D}_t^{-1} \mathbf{r}_t$, a positive definite matrix $\overline{\mathbf{Q}}$ and positive parameters satisfying $0 < \gamma_0 + \delta_0 < 1$. The DCC$(r, s, 1, 1)$ and the ADCC$(r, s, 1, 1)$, for $1 \leq r, s \leq 2$, are compared with the *common* approach method (26) with $\lambda_0 = 0.94$, over the horizon $p = 1, 5, 20, 120$. Derivation of ${}^{(\mathbf{1})}\mathbf{\Sigma_{t+p|t}}$ for DCC and ADCC is straightforward and details are skipped for sake of simplicity. All models were estimated recursively using an expanding window starting with 1784 observations as the first estimation sample. The evaluation sample covers the last two years of data (from November 2, 2001 to October 13, 2003, inclusive) yielding 507–p values for the loss function. However, the parameter values are updated at monthly intervals, yielding twenty-four set of estimates rather than 507–p, in order to alleviate the already sizeable computational burden. The estimates are nor reported for sake of simplicity. All the computations have been carried out in MatLab and the codes are available upon request. Compared with the bivariate case, the results, reported in Table 4, now show more clearly that the *correct* approach (here represented by DCC and ADCC models), is markedly superior to the *common* approach ($Riskmetrics^{TM}$) for all cases and all horizon.

4 Concluding remarks

Practitioners face the problem of estimating in real time conditional volatilities and cross-correlations for a large set of asset returns. The so-called *common* approach of specifying, estimating and forecasting GARCH models provides a simple and feasible way to achieve this task. Unfortunately, as this paper shows, the estimates obtained in this way lack of the usual (asymptotic) statistical properties. The Monte-Carlo experiments indicates that such estimation procedure is invalid even in small-sample. The empirical applications here presented suggest that this can have a non trivial effect on the forecasting performance of conditional variances and correlations.

5 Appendix

Proof of Theorem 1. (a) Impose (11). Substituting for t times, recursively, $r_t^2 = z_t^2 \sigma_t^2$ into (9) (cf. Nelson (1990, eq.(6))) yields

$$\sigma_t^2 = \sigma_0^2 \prod_{i=1}^{t} (1 - \alpha_0 + \alpha_0 z_{t-i}^2). \tag{37}$$

which is bounded $a.s.$ for any $t < \infty$ whenever $\sigma_0^2 < \infty$. Moreover, there exists a random integer $K < \infty$ $a.s.$ such that (cf. Nelson (1990, p.320))

$$\prod_{s=1}^{t} (1 - \alpha_0 + \alpha_0 z_{t-i}^2) \le C^t \quad a.s. \text{ for any } t > K, \tag{38}$$

setting

$$C = e^{\frac{\gamma}{2}} < 1,$$

with $\gamma = E \ln z_0^2 < 0$. Next, by Markov's inequality, for arbitrary $\epsilon > 0$,

$$Pr(t^{-2}\, z_t^2 \geq \epsilon) \leq \frac{1}{\epsilon t^2}$$

yielding by Borel-Cantelli lemma

$$z_t^2 = o(t^2) \;\; a.s. \text{ for } t \to \infty. \tag{39}$$

Finally, using $(a-b)^2 \leq 2(a^2+b^2)$ for any real a, b yields (assume $T > K$ with no loss of generality)

$$MSE_T(\alpha) \leq \frac{2}{T}\sum_{t=2}^{T} r_t^4 + \frac{2}{T}\sum_{t=2}^{T} \sigma_t^4(\alpha). \tag{40}$$

For the first term on the right-hand side of (40), using (38) and (39),

$$\begin{aligned}
&\sum_{t=2}^{T} r_t^4 = \sum_{t=2}^{K} r_t^4 + \sum_{t=K+1}^{T} r_t^4 \\
&\leq \sum_{t=2}^{K} r_t^4 + \sigma_0^4 \sum_{t=K+1}^{T} C^{2t} z_t^4 \leq \sum_{t=2}^{K} r_t^4 + \sigma_0^4 \sum_{t=1}^{\infty} C^{2t} z_t^4 < \infty \;\; a.s.,
\end{aligned}$$

yielding

$$\frac{1}{T}\sum_{t=2}^{T} r_t^4 = O\left(\frac{1}{T}\right) \;\; a.s. \text{ for } T \to \infty.$$

Note that this term is independent of α.

For the second term on the right-hand side of (40)

$$\sum_{t=2}^{T} \sigma_t^4(\alpha) = \sum_{t=2}^{K+1} \sigma_t^4(\alpha) + \sum_{t=K+2}^{T} \sigma_t^4(\alpha)$$

and

$$\sum_{t=K+2}^{T} \sigma_t^4(\alpha) \leq 2\alpha^2 \sum_{t=K+2}^{T} \left(\sum_{s=1}^{t-K-1} (1-\alpha)^{s-1} r_{t-s}^2\right)^2 + 2\alpha^2 \sum_{t=K+2}^{T} \left(\sum_{s=t-K}^{t-1} (1-\alpha)^{s-1} r_{t-s}^2\right)^2 \tag{41}$$

using

$$\sigma_t^2(\alpha) = \alpha \sum_{s=1}^{t-1} (1-\alpha)^{s-1} r_{t-s}^2.$$

For the first term on the right hand side of (41), using the c_r inequality, viz. $(\sum_{i=1}^m a_i)^2 \leq m\,(\sum_{i=1}^m a_i^2)$ for any sequence $\{a_i\}$,

$$\alpha^2 \sum_{t=K+2}^{T} \left(\sum_{s=1}^{t-K-1} (1-\alpha)^{s-1} r_{t-s}^2 \right)^2 \leq \sigma_0^4 \alpha^2 \sum_{t=K+2}^{T} t \left(\sum_{s=1}^{t-K-1} (1-\alpha)^{2(s-1)} C^{2(t-s)} z_{t-s}^4 \right)$$
$$\leq \sigma_0^4 \alpha^2 \sum_{t=2}^{T} t \left(\sum_{s=1}^{t-1} (1-\alpha)^{2(s-1)} C^{2(t-s)} z_{t-s}^4 \right)$$
$$\leq \sigma_0^4 \alpha^2 \sum_{t=2}^{T} t \left(\sum_{s=1}^{|t/2|} (1-\alpha)^{2(s-1)} C^{2(t-s)} z_{t-s}^4 \right) + \sigma_0^4 \alpha^2 \sum_{t=2}^{T} t \left(\sum_{s=|t/2|+1}^{t-1} (1-\alpha)^{2(s-1)} C^{2(t-s)} z_{t-s}^4 \right)$$
$$\leq \sigma_0^4 \alpha^2 \sum_{t=2}^{T} t\, C^t \left(\sum_{s=1}^{|t/2|} (1-\alpha)^{2(s-1)} z_{t-s}^4 \right) + \sigma_0^4 \alpha^2 \sum_{t=2}^{T} t\,(1-\alpha)^t \left(\sum_{s=|t/2|+1}^{t-1} C^{2(t-s)} z_{t-s}^4 \right)$$
$$\leq \sigma_0^4 \alpha^2 \sum_{t=2}^{T} t\, C^t \left(\sum_{s=1}^{t-1} z_{t-s}^4 \right) + \sigma_0^4 \alpha^2 \sum_{t=2}^{T} t\,(1-\alpha)^t \left(\sum_{s=1}^{t-1} z_{t-s}^4 \right) \leq 2\sigma_0^4 \bar{\alpha}^2 \sum_{t=2}^{T} B^t \left(\sum_{s=1}^{t-1} z_{t-s}^4 \right)$$
$$= 2\sigma_0^4 \bar{\alpha}^2 \sum_{t=1}^{T-1} z_t^4 \left(\sum_{s=t+1}^{T} B^s \right) \leq \frac{2\sigma_0^4 \bar{\alpha}^2}{1-B} \sum_{t=1}^{T-1} z_t^4 B^{t+1} \leq \frac{2\sigma_0^4 \bar{\alpha}^2}{1-B} \sum_{t=1}^{\infty} z_t^4 B^{t+1} < \infty \;\; a.s.$$

for a constant B satisfying

$$sup\,[\,(1-\underline{\alpha}),\, C] < B < 1$$

For the second term on the right-hand side of (41)

$$2\alpha^2 \sum_{t=K+2}^{T} \left(\sum_{s=t-K}^{t-1} (1-\alpha)^{s-1} r_{t-s}^2 \right)^2 \leq \bar{\alpha}^2\, K\, max_{1\leq s\leq K}\, r_i^4 \sum_{t=K+2}^{T} \left(\sum_{s=t-K}^{t-1} (1-\underline{\alpha})^{2(s-1)} \right)$$
$$= \bar{\alpha}^2\, K\, max_{1\leq s\leq K} r_i^4 \sum_{t=K+2}^{T} (1-\underline{\alpha})^{2(t-K-1)} \left(\sum_{s=0}^{K-1} (1-\underline{\alpha})^{2s} \right)$$
$$\leq \bar{\alpha}^2\, K\, max_{1\leq s\leq K}\, r_i^4 \left(\sum_{t=0}^{\infty} (1-\underline{\alpha})^{2t} \right)^2 = K\, max_{1\leq s\leq K}\, r_i^4 \left(\frac{\bar{\alpha}}{1-(1-\underline{\alpha})^2} \right)^2 < \infty \;\; a.s.$$

Finally, collecting terms

$$\sup_{\alpha\in[\underline{\alpha},\bar{\alpha}]} \frac{2}{T}\sum_{t=2}^{T}\sigma_t^4(\alpha) = O\left(\frac{1}{T}\right) \quad a.s. \quad \text{for } T\to\infty. \ \square$$

(b)

$$E(MSE_T(\alpha)\mid\sigma_0^2) = T^{-1}\sum_{t=1}^{T}\left(E(r_t^4\mid\sigma_0^2) + E(\sigma_t^4(\alpha)\mid\sigma_0^2) - 2E(r_t^2\sigma_t^2(\alpha)\mid\sigma_0^2)\right).$$

Easy calculations yield

$$E(r_t^4\mid\sigma_0^2) = \sigma_0^4\delta_0^t$$

setting

$$\delta_0 := E(1-\alpha_0+\alpha_0 z_t^2)^2.$$

Next

$$\begin{aligned}
&E(\sigma_t^4(\alpha)\mid\sigma_0^2)\\
&= \alpha^2\sum_{j=1}^{t-1}(1-\alpha)^{2(j-1)}E(r_{t-j}^4\mid\sigma_0^2) + 2\alpha^2\sum_{j_2=1}^{t-2}(1-\alpha)^{j_2-1}\sum_{j_1=j_2+1}^{t-1}(1-\alpha)^{j_1-1}E(r_{t-j_1}^2 r_{t-j_2}^2\mid\sigma_0^2)\\
&= \sigma_0^4\alpha^2\sum_{j=1}^{t-1}(1-\alpha)^{2(j-1)}\delta_0^{t-j} + 2\sigma_0^4\alpha^2\kappa_0\sum_{j_2=1}^{t-2}(1-\alpha)^{j_2-1}\sum_{j_1=j_2+1}^{t-1}(1-\alpha)^{j_1-1}\delta_0^{t-j_1}
\end{aligned}$$

setting

$$\kappa_0 := E(1-\alpha_0+\alpha_0 z_t^2)z_t^2,$$

and

$$E(r_t^2\sigma_t^2(\alpha)\mid\sigma_0^2) = \kappa_0\alpha\sum_{j=1}^{t-1}(1-\alpha)^{(j-1)}\delta_0^{t-j}.$$

Tedious calculations yield

$$E(r_t^4\mid\sigma_0^2) + E(\sigma_t^4(\alpha)\mid\sigma_0^2) - 2E(r_t^2\sigma_t^2(\alpha)\mid\sigma_0^2) = \delta_0^t c_t(\alpha),$$

where the sequence of positive constants $c_t(\alpha)$ satisfy $c_t(\alpha)\to c(\alpha)<\infty \ \ as \ \ t\to\infty$, setting

$$c(\alpha) = \left(1+\frac{\alpha^2}{(\delta_0-(1-\alpha)^2)} + \frac{2\alpha^2\kappa_0(1-\alpha)}{(\delta_0-(1-\alpha))(\delta_0-(1-\alpha)^2)} - \frac{2\alpha\kappa_0}{(\delta_0-(1-\alpha))}\right).$$

Note that $c_t(\alpha), c(\alpha)$ depend also on $\delta_0,\ \kappa_0$ but we are not making this explicit for simplicity. By simple manipulations, noting that $\delta_0 = 1+\alpha_0^2(\mu_4-1),\ \ \kappa_0 = 1+\alpha_0(\mu_4-1)$ for $\mu_4 := Ez_0^4$, one gets

$$c(\alpha) = \frac{\alpha(\mu_4-1)^2}{(\delta_0-(1-\alpha)^2)(\delta_0-(1-\alpha))},$$

implying

$$\inf_{\alpha\in[\underline{\alpha},\bar{\alpha}]} c(\alpha) = \underline{c} > 0.$$

Hence (18) easily follows since $\delta_0 > 1$. □

References

Alexander, C. (2001): “Orthogonal GARCH,” in *Mastering risk*, ed. by C. Alexander, vol. 2, pp. 21–38. London: Financial Times - Prentice Hall.

Andersen, A., and T. Bollerslev (1998): “Answering the skeptics: Yes, standard volatility models do provide accurate forecasts,” *International Economic Review*, 39, 885–905.

Audrino, F., and G. Barone-Adesi (2004): “Average conditional correlation and tree structures for multivariate GARCH models,” *University of Lugano*, Preprint.

Bauwens, L., S. Laurent, and J. Rombouts (2003): “Multivariate GARCH models: a survey,” *CORE* , Preprint.

Bollerslev, T. (1986): “Generalized autoregressive conditional heteroskedasticity,” *Journal of Econometrics*, 31, 302–327.

——— (1990): “Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH,” *Review of Economics and Statistics*, 72, 498–505.

Bollerslev, T., R. Engle, and D. Nelson (1994): “ARCH models,” in *Handbook of Econometrics vol. IV*, ed. by R. Engle, and D. McFadden. Amsterdam: North Holland.

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

Cappiello, L., R. Engle, and K. Sheppard (2002): "Asymmetric dynamics in the conditional correlations of global equity and bond returns," *NYU*, Preprint.

Cheng, M., J. Fan, and V. Spokoiny (2003): "Dynamic nonparametric filtering with application to finance," *Weiesestrass Insitute and Humboldt University Berlin*, Preprint.

Diebold, F., and R. Mariano (1995): "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, 13, 253–263.

Engle, R. (2002): "Dynamic conditional correlation - a simple class of multivariate generalized autoregressive conditional heteroskedasticity models," *Journal of Business & Economic Statistics*, 20, 339–350.

Engle, R. F. (1982): "Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom," *Econometrica*, 50, 987–1007.

J.P.Morgan/Reuters (1996): "RiskMetrics^{TM} - Technical Document, Fourth Edition," .

Lee, S., and B. Hansen (1994): "Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator," *Econometric Theory*, 10, 29–52.

Litterman, R., and K. Winkelmann (1998): "Estimating Covariance Matrices," Risk management series, Goldman Sachs & Co.

Lumsdaine, R. (1996): "Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models," *Econometrica*, 64/3, 575–596.

Nelson, D. (1990): "Stationarity and persistence in the GARCH(1,1) model," *Econometric Theory*, 6, 318–334.

Pelletier, D. (2002): "Regime switching for dynamic correlations," *Journal of Econometrics*, forthcoming.

Pesaran, M., and P. Zaffaroni (2004): “Model Averaging and Value-at-Risk based Evaluation of Large Multi Asset Volatility Models for Risk Management,” *University of Cambridge*, Preprint.

West, K. (1996): “Asymptotic inference about predictive ability,” *Econometrica*, 64, 1067–1084.

Table 1: univariate estimation of α_0

$\omega_0 = 0$ (LSE)						$\omega_0 = 1$ (PMLE)					
Value	Mean	Stand. Dev.	Maximum	Minimum	Obj. fun.	Value	Mean	Stand. Dev.	Maximum	Minimum	Obj. fun.
$T = 1,000$											
0.05	0.2121	0.2275	0.9616	0	2.5017	0.05	0.0503	0.0197	0.1091	0	2.2072
0.75	0.7419	0.1660	0.9387	0.0251	101.137	0.75	0.7494	0.0202	0.8077	0.6864	4.5885
0.95	0.9525	0.0283	1	0.5	19.78	0.95	0.9510	0.0108	0.9956	0.9185	6.5560
$T = 5,000$											
0.05	0.3608	0.1617	0.9616	0	0.5004	0.05	0.0495	0.0091	0.0791	0	2.2076
0.75	0.7474	0.1482	0.9388	0.0635	20.2274	0.75	0.7495	0.0093	0.7768	0.7192	4.6318
0.95	0.9519	0.0300	0.9809	0.5	69.5787	0.95	0.9499	0.0044	0.9678	0.9332	7.3530
$T = 15,000$											
0.05	0.5116	0.1081	0.9616	0	0.1668	0.05	0.0500	0.0056	0.0643	0	2.2078
0.75	0.7347	0.1524	0.9388	0.2377	6.7082	0.75	0.7500	0.0052	0.7640	0.7297	4.6430
0.95	0.9528	0.0226	0.9809	0.5	28.4408	0.95	0.9501	0.0025	0.9580	0.9422	7.5519

Obj. func. is the average (across 1000 replications) of MSE (13) for LS estimator and of Gaussian log likelihood (20) for PMLE.

Table 2: Testing predictive accuracy (univariate)

			$\frac{1}{260}\sum_{t=p+1}^{260}\left\|\sigma_t^2-\hat{\sigma}_{t\|t-p}^2\right\|^{\frac{1}{2}}$				$\frac{1}{260}\sum_{t=p+1}^{260}\left(\ln(\hat{\sigma}_{t\|t-p}^2)+\frac{\sigma_t^2}{\hat{\sigma}_{t\|t-p}^2}\right)$			
$\hat{\sigma}_0^{2,lse}(\alpha)$	$\hat{\alpha}_T^{lse}$	$\hat{\alpha}_T^{pmle}$	p: 1	5	20	120	1	5	20	120
0.1	0.881	0.851	0.829	1.476	1.958	8.885	-0.913	-0.957	-0.389	2.645
0.5	0.512	”	-5.103	-2.319	-0.257	4.334	-4.275	-3.595	-2.239	-0.219
1	0.307	”	-4.478	-7.474	-0.895	3.906	-2.916	-3.041	-3.408	-1.087
0.226	0.491	”	-5.067	-2.502	-0.313	4.207	-4.455	-3.446	-2.443	-0.313
s^2	0.94	”	0.067	0.527	1.824	9.791	1.288	0.387	0.549	4.913
s^2, $n = 15$	0.94	”	0.684	0.684	0.816	5.227	-0.496	-0.551	-0.583	-0.369
s^2, $n = 30$	0.94	”	0.783	2.026	1.980	6.415	-0.553	-1.339	0.106	2.459

Columns 4 to 11 report the Diebold and Mariano (21) test statistic. Columns 1 and 2 report the average (across 260 replications) of the LS estimates of σ_0^2, α_0. Column 3 reports the average of the PMLE of α_0 (the average of the PMLE of ω_0 is 9×10^{-7}).
s^2 indicates the sample covariance matrix of the data and n is the time-window (in days).

Table 3: Testing predictive accuracy (bivariate)

			$\frac{1}{258}\sum_{t=p+1}^{258}\left(e_{(11),t}+e_{(22),t}+e_{(12),t}\right)$ $e_{(i),t}=\left\|\sigma^2_{t(i)}-\hat{\sigma}^2_{t(i)\|t-p}\right\|^{\frac{1}{2}}$				$\frac{1}{258}\sum_{t=p+1}^{258}\left\|\rho_{12,t}-\hat{\rho}_{12,t\|t-p}\right\|$ $\hat{\rho}_{12,t\|t-p}=\frac{\hat{\sigma}^2_{t(12)\|t-p}}{\hat{\sigma}_{t(11)\|t-p}\hat{\sigma}_{t(22)\|t-p}}$			
$\tilde{\sigma}^2_0(\alpha)$	$\tilde{\alpha}_T$	$\hat{\alpha}^{pmle}_T$	p: 1	5	20	120	1	5	20	120
0.1	0.881	0.859	0.487	0.816	4.125	8.501	1.01	0.877	-0.125	0.674
0.5	0.512	"	-5.281	-2.662	-0.216	4.810	-6.543	-7.308	-5.979	-8.938
1	0.307	"	-7.269	-4.939	-0.688	4.283	-8.244	-8.605	-7.732	-8.171
0.226	0.491	"	-5.431	-2.812	-0.287	4.689	-6.675	-7.303	-6.141	-8.861
s^2	0.94	"	0.635	1.023	4.245	8.645	1.078	0.731	-0.341	0.402
s^2, $n=15$	0.94	"	-0.490	1.209	1.316	7.097	-3.563	-2.710	-1.873	-0.838
s^2, $n=30$	0.94	"	0.648	1.351	2.832	8.347	0.271	-0.177	-0.857	-0.048

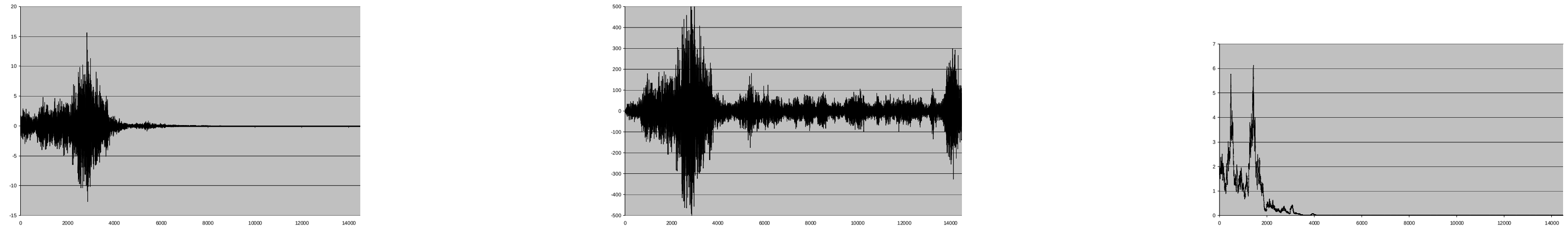
Columns 4 to 11 report the Diebold and Mariano (21) test statistic. Columns 1 and 2 report the average (across 258 replications) of the LS estimates of σ^2_0, α_0. Column 3 reports the average of the PMLE of α_0 (the average of the PMLE of $\Omega_0 = (\omega_{0,1}, \omega_{0,12}, \omega_{0,2})$ is $(9\times 10^{-7}, 3\times 10^{-7}, 10\times 10^{-7})$). s^2 indicates the sample covariance matrix of the data and n is the time-window (in days).

Table 4: Testing predictive accuracy (multivariate)

	DCC$(r,s,1,1)$				ADCC$(r,s,1,1)$			
	p: 1	5	20	120	1	5	20	120
$r=1, s=1$	-6.547	-4.939	-5.510	-5.710	-5.549	-5.568	-6.2549	-5.954
$r=2, s=1$	-6.151	-4.3277	-5.005	-5.033	-4.078	-5.641	-6.318	-5.973
$r=1, s=2$	-6.880	-4.313	-4.987	-5.099	-6.423	-5.919	-6.566	-6.242
$r=2, s=2$	-6.201	-4.280	-4.972	-4.911	-5.073	-6.556	-7.125	-6.485

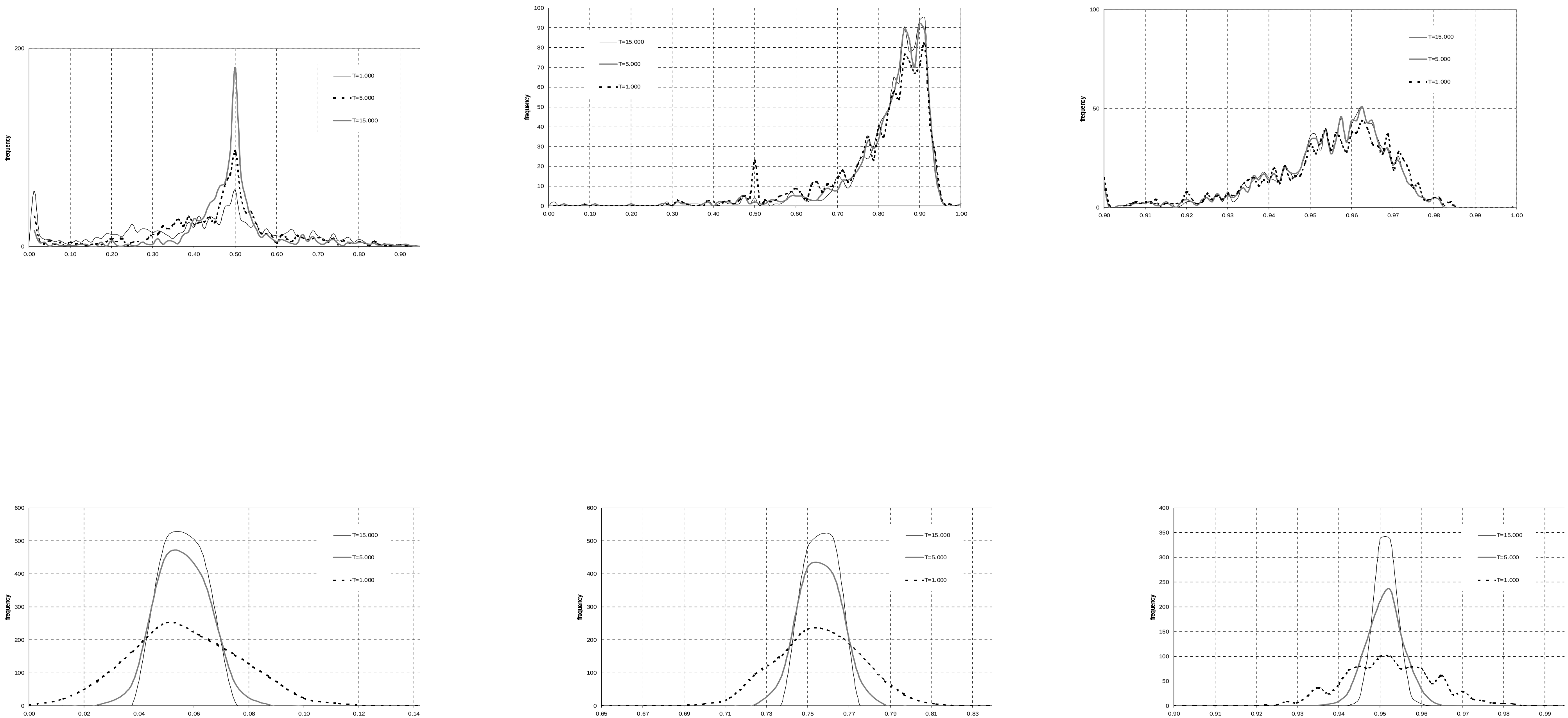
Columns 2 to 9 report the Diebold and Mariano (21) test statistic based on the loss function $\frac{1}{507-p}\sum_{t=p+1}^{507} \iota' \mid vech(\mathbf{\Sigma}_{t|t-p} - \mathbf{r_t r_t'}) \mid^{\frac{1}{2}}, \;\; \iota = (1, ..., 1)'$.
Columns 2-5 report the result for DCC$(p,q,1,1)$ **and columns 6-9 for ADCC**$(p,q,1,1)$**,**
both with respect to the common approach with $\lambda_0 = 0.94$, $n = 250$ **and initialized at** s^2**.**
All models are estimated recursively with an expanding window starting with 1784 **observations as the first estimation sample, with the parameter values updated at monthly intervals.**

Figure 1



Note: Simulated path of GARCH$(1,1)$ r_t with $\alpha_0 = 0.95$ and $\omega_0 = 0$ (first panel) and $\omega_0 = 1$ (second panel) and of $tr(\Sigma_t)$ (cf. (34)) with $\Omega_0 = 0$ and $\alpha_0 = 0.95$ (third panel).

Figure 2



Note: Montecarlo distribution (1,000 replications) of the LS estimator $\hat{\alpha}_T^{lse}$ with true values $\omega_0 = 0$ and $\alpha_0 = 0.05$ (first panel), $\alpha_0 = 0.75$ (second panel), $\alpha_0 = 0.95$ (third panel) and of of the PML estimator $\hat{\alpha}_T^{pmle}$ with true values $\omega_0 = 1$ and $\alpha_0 = 0.05$ (fourth panel), $\alpha_0 = 0.75$ (fifth panel), $\alpha_0 = 0.95$ (sixth panel). Estimation starts from an initial value of $\alpha = 0.5$. The dotted line refers to samples of $T = 1,000$ observations, the grey line to samples of $T = 5,000$ observations and the black line to samples of $T = 15,000$ observations.