UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Scienze Economiche "Marco Fanno"

DYNAMIC CONDITIONAL CORRELATION: ON PROPERTIES AND ESTIMATION

GIAN PIERO AIELLI

November 2011

"MARCO FANNO" WORKING PAPER N.142

Dynamic Conditional Correlation: on Properties and Estimation

Gian Piero Aielli*

July 14, 2011

Abstract

We address some issues that arise with the Dynamic Conditional Correlation (DCC) model. We prove that the DCC large system estimator (DCC estimator) can be inconsistent, and that the traditional interpretation of the DCC correlation parameters can lead to misleading conclusions. We then suggest a more tractable dynamic conditional correlation model (cDCC model). A related large system estimator (cDCC estimator) is described and heuristically proven to be consistent. Sufficient stationarity conditions for cDCC processes of interest, including the covariance-return process, are established. The DCC and cDCC estimators are compared by means of applications to simulated and real data.

KEYWORDS: *Multivariate GARCH Model, Quasi-Maximum-Likelihood, Two-step Estimation, Integrated Correlation, Generalized Profile Likelihood.*

JEL CODES: C13, C32, C51, C52, C53.

1. INTRODUCTION

During the last decade, the focus on the variance-correlation decomposition of the asset conditional covariance matrix has become one of the most popular approaches to the modeling of multivariate volatility. Seminal works in this area are the Constant Conditional Correlation (CCC) model of Bollerslev (1990), the Dynamic Conditional Correlation (DCC) model of Engle (2002), and the Varying Correlation (VC) model of Tse and Tsui (2002). Extensions of the DCC model have been proposed, among others, by Cappiello, Engle, and Sheppard (2006), Billio, Caporin, and Gobbo (2006), Pesaran and Pesaran (2007), and Franses and Hafner (2009). Related examples are the models of Silvennoinen and Teräsvirta (2005), Pelletier (2006), McAleer, Chan, Hoti, and Lieberman (2008), and Kwan, Li, and Ng (in press).

In the DCC model, the conditional variances are modeled as univariate GARCH models; the conditional correlations are then modeled as peculiar functions of the past GARCH standardized returns. In its original intentions, such a modeling approach should have been capable of providing two major advantages. First, thanks to the modular structure of the DCC conditional covariance matrix, a consistent large system estimator (DCC estimator) should have been available as a three-step procedure. Second, thanks to the peculiar parametrization of the DCC conditional correlation process, testing for correlation hypotheses, such as whether or not the correlation process is integrated, should have been easier than with other models. Aielli (2006) pointed out that the DCC model is a partial solution to the aims of the DCC modeling approach. The author demonstrated that the DCC model is less tractable than expected and that the conjecture on the consistency of the DCC

*e-mail: gian.piero.aielli@gmail.com

estimator is not substantiated. He then suggested reformulating the DCC correlation driving process as a *linear* multivariate generalized autoregressive conditional heteroskedasticity (MGARCH) process (cDCC model). The resulting model allows for an *ad hoc* profile quasi-log-likelihood estimator (cDCC estimator) which is feasible with large systems. Compared with the DCC estimator, the cDCC estimator only requires a minor additional computational effort.

This paper extends the paper of Aielli (2006) with new theoretical and empirical results. We prove — no longer only conjecture — that the DCC estimator of the location correlation parameter can be inconsistent (sec. 3.1.1). Regarding the testing of correlation hypotheses, we point out that the test of DCC integrated correlations is an inconclusive procedure because of the unknown meaning of the alternative hypothesis (sec. 2.1.3). More generally, we show that the traditional GARCH-like interpretation of the DCC correlation parameters can lead to paradoxical conclusions (sec. 2.1.1). For example, in spite of the presence of a "unit root", the DCC correlation driving process is weakly stationary.

As for the cDCC model, relying on a recent result of Boussama, Fuchs and Stelzer (in press), sufficient stationarity conditions for cDCC processes of interest are established. Such stationarity conditions are stated as a flexible stationarity principle (cDCC stationarity principle), which is capable of encompassing a wide range of possible variance specifications (sec. 2.2.1). A consistency property of the cDCC estimator of the location correlation parameter is proved, relying on which a heuristic consistency proof for the cDCC estimator as a whole is provided (sec. 3.2.1). The meaning of the test of cDCC integrated correlations is also established (sec. 2.2.3). Under the null hypothesis of integrated correlation, the squared conditional correlation is expected to increase with the forecast horizon; under the alternative hypothesis, it is expected to revert to the stationary state.

The finite sample performances of the cDCC and DCC estimators are compared by means of applications to simulated and real data. Under correctly specified model, for parameter values that are common in financial applications, the bias of the DCC location correlation parameter estimator is negligible (sec. 4.1). For less common parameter values, it can be substantial. In general, such a bias is an increasing function of the persistence of the correlation process and of the impact of the news. Such a bias disappears when the cDCC estimator is used. Some simulation experiments under misspecification are discussed (sec. 4.2), where the DCC and cDCC estimators prove to perform very similarly. For the applications to the real data (sec. 4.3), two datasets are considered, namely, a small dataset of ten equity indices, and a large dataset of 100 equities. On both datasets, the cDCC correlation forecasts perform as well as or significantly better than the DCC correlation forecasts. The remainder of the paper is organized as follows: section 2 illustrates a theoretical comparison of the DCC and cDCC models; section 3 discusses the DCC and cDCC estimators; section 4 compares the empirical performances of the two estimators, and section 5 concludes the paper. The proofs of the propositions are compiled in the Appendix.

2. STRUCTURAL PROPERTIES

2.1 The DCC model

Let $y_t \equiv [y_{1,t}, \ldots, y_{N,t}]'$ denote the vector of excess returns at time $t = 0, \pm 1, \pm 2, \ldots$ We assume that y_t is a martingale difference, or $E_{t-1}[y_t] = 0$, where $E_{t-1}[\cdot]$ denotes expectations conditional on $y_{t-1}, y_{t-2}, \ldots$ The conditional covariance matrix of the excess returns, $H_t \equiv E_{t-1}[y_t y_t']$, can be written as

$$H_t = D_t^{1/2} R_t D_t^{1/2}, \tag{1}$$

where $R_t \equiv [\rho_{ij,t}]$ is the asset conditional correlation matrix and $D_t \equiv \mathrm{diag}(h_{1,t}, \ldots, h_{N,t})$ is a diagonal matrix, with the asset conditional variances as diagonal elements. By construction, R_t is the conditional covariance matrix of the vector of the standardized returns, $\varepsilon_t \equiv [\varepsilon_{1,t}, \varepsilon_{2,t}, \ldots, \varepsilon_{N,t}]'$, where $\varepsilon_{i,t} = y_{i,t}/\sqrt{h_{i,t}}$.

In the DCC model, the variance processes are modeled as univariate GARCH models,

$$h_{i,t} = h_i(\theta_i; y_{i,t-1}, y_{i,t-2}, \ldots), \tag{2}$$

where $h_i(\,\cdot\,; \cdot\,, \cdot\,, \ldots)$ is a known function and θ_i is a vector of parameters, $i = 1, 2, \ldots, N,$. The conditional correlation process is modeled as

$$R_t = Q_t^{*-1/2} Q_t Q_t^{*-1/2}, \tag{3}$$

where

$$Q_t = (1 - \alpha - \beta)\, S + \alpha\, \varepsilon_{t-1}\varepsilon_{t-1}' + \beta\, Q_{t-1}, \tag{4}$$

where $Q_t \equiv [q_{ij,t}]$, $S \equiv [s_{ij}]$, $Q_t^* \equiv \mathrm{diag}(q_{11,t}, \ldots, q_{NN,t})$, and α and β are scalars. If Q_t is positive definite (pd), R_t is pd with unit diagonal elements, as required for R_t to be a pd correlation matrix. To ensure that Q_t is pd, it is assumed that $\alpha \geq 0$, $\beta \geq 0$, $\alpha + \beta < 1$, and S is pd. It is common (though unnecessary — see sec. 2.2) to also assume that $s_{ii} = 1$ for $i = 1, 2, \ldots, N$.

Some simulated series of $\rho_{ij,t}$ are plotted in Fig. (1). *Coeteris paribus*, the persistence of $\rho_{ij,t}$ increases with $\alpha + \beta$, the dispersion of $\rho_{ij,t}$ is an increasing function of α, and s_{ij} is essentially a location parameter for $\rho_{ij,t}$. In the financial applications, a common estimation output is $\hat{\alpha} + \hat{\beta} \geq .96$, with $\hat{\alpha} \leq .04$.

More general models can be obtained by replacing (4) with more general equations, such as

$$Q_t = (\iota\iota' - A - B) \odot S + A \odot \varepsilon_{t-1}\varepsilon_{t-1}' + B \odot Q_{t-1}, \tag{5}$$

(Engle 2002, eq. 24), where $A \equiv [\alpha_{ij}]$, $B \equiv [\beta_{ij}]$, ι denotes the $N \times 1$ vector with unit entries, and $\odot$ denotes the element-wise (Hadamard) matrix product. With this model, Q_t is pd provided that A is positive semi definite (psd), B is psd, and $(\iota\iota' - A - B) \odot S$ is pd (Ding and Engle 2001).

For illustrative purposes, it is useful to also consider the VC model of Tse and Tsui (2002),

$$R_t = (1 - \alpha - \beta) S + \alpha\, \mathcal{R}_{M,t-1} + \beta\, R_{t-1}, \tag{6}$$

where $\mathcal{R}_{M,t}$ is the sample correlation matrix of $\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-M+1}$, and $M \geq N$ is fixed arbitrary. If $\alpha \geq 0$, $\beta \geq 0$, $\alpha + \beta < 1$, and S is unit-diagonal pd, R_t is unit-diagonal pd.

2.1.1 GARCH-like interpretation of the dynamic correlation parameters.

The DCC correlation driving process, Q_t, is often treated as a *linear* MGARCH process (see, e.g., Engle 2002, eq. (18)). The way α, β, and S affect the dynamics of Q_t is then interpreted accordingly. Since the conditional covariance matrix of ε_t is R_t (not Q_t), the process Q_t is not linear MGARCH. In fact, treating Q_t as a linear MGARCH process can lead to misleading conclusions. For example, for $\alpha + \beta = 1$, the process $q_{ij,t}$, which is thought of as an integrated GARCH process (Engle 2002, eq. (17)), is weakly stationary. To see why, rewrite $q_{ij,t}$ for $i = j$ as

$$q_{ii,t} = (1-\beta) + \beta q_{ii,t-1} + \alpha(\varepsilon_{it-1}^2 - 1), \tag{7}$$

where we applied $s_{ii} = 1$. Since $\alpha(\varepsilon_{it-1}^2 - 1)$ is a martingale difference, $q_{ii,t}$ is AR(1) with autoregressive parameter β. For $\alpha + \beta \leq 1$ and $\alpha > 0$, $q_{ii,t}$ is weakly stationary. Since $|q_{ij,t}| \leq \sqrt{q_{ii,t} q_{jj,t}}$, it follows that, for $\alpha + \beta \leq 1$ and $\alpha > 0$, the second moment of $q_{ij,t}$ exists finite, in which case, if $q_{ij,t}$ is strictly stationary, it is weakly stationary.

After recognizing that Q_t is not linear MGARCH, finding a tractable representation of Q_t proves to be difficult. Specifically, the dynamic properties of $R_t(= Q_t^{*-1/2} Q_t Q_t^{*-1/2})$ turn out to be hard to study. Similar problems also arise with the VC model (see eq. (6)), in that R_t is generally not the conditional expectation of $\mathcal{R}_{M,t}$.

2.1.2 Formula and interpretation of S.

Applying a standard result on linear MGARCH processes, S is treated as the second moment of ε_t (Engle 2002, eq. (18), (23) and (24)). Accordingly, during the fitting of large systems, S is replaced by the sample second moment of the estimated standardized returns (Engle and Sheppard 2001; Cappiello et al. 2006; Billio et al. 2006; Pesaran and Pesaran 2007; Franses and Hafner 2009). Unfortunately, the equality $S = E[\varepsilon_t \varepsilon_t']$ does not hold in general. As an example, consider the model in (5), with $N = 2$, $\alpha_{11} > 0$, $\alpha_{22} = \alpha_{12} = \beta_{11} = \beta_{22} = \beta_{12} = 0$, and $0 < s_{12} < \sqrt{1-\alpha_{11}}$. As required, A is psd, B is psd, and $(\iota\iota' - A - B) \odot S$ is pd. By Jensen's inequality and $E[\varepsilon_{1,t}^2] = 1$, we get

$$\begin{aligned} E[\varepsilon_{1,t}\,\varepsilon_{2,t}] &= E[\rho_{12,t}] = E\left[s_{12}/\sqrt{(1-\alpha_{11}) + \alpha_{11}\varepsilon_{1,t-1}^2}\right] \\ &= s_{12}\, E\left[1/\sqrt{(1-\alpha_{11}) + \alpha_{11}\varepsilon_{1,t-1}^2}\right] > s_{12}/\sqrt{(1-\alpha_{11}) + \alpha_{11} E[\varepsilon_{1,t-1}^2]} \\ &= s_{12} \\ &= E[q_{12,t}]. \end{aligned}$$

Noting that $s_{12} = E[\varepsilon_{1,t}\varepsilon_{2,t}]$ if and only if $E[\varepsilon_{1,t}\varepsilon_{2,t}] = E[q_{12,t}]$ (see eq. 5), it follows that $s_{12} \neq E[\varepsilon_{1,t}\varepsilon_{2,t}]$.

Proposition 2.1. *Suppose that $\alpha + \beta < 1$ and that $E[Q_t]$ and $E[\varepsilon_t \varepsilon_t']$ are independent of t. Then,*

$$S = \frac{1-\beta}{1-\alpha-\beta} E[Q_t^{*1/2} \varepsilon_t\, \varepsilon_t'\, Q_t^{*1/2}] - \frac{\alpha}{1-\alpha-\beta} E[\varepsilon_t\, \varepsilon_t']. \tag{8}$$

Proof. All proofs are reported in the Appendix.

Prop. 2.1 shows that the interpretation of S is not immediate. Specifically, replacing S with the sample second moment of ε_t is not an obvious estimation device. Indeed, the only case in which the equality $S = E[\varepsilon_t \varepsilon_t']$ holds seems to be the case of constant conditional correlations ($\alpha = \beta = 0$). As theoretical evidence in favor of this, consider the following argument. From eq. (4), under stationarity it follows that $S = E[\varepsilon_t \varepsilon_t']$ if and only if $E[\varepsilon_t \varepsilon_t'] = E[Q_t]$. But the DCC defining equations imply that $E[\varepsilon_t \varepsilon_t'] = E[E_{t-1}[\varepsilon_t \varepsilon_t']] = E[R_t] = E[Q_t^{*-1/2} Q_t Q_t^{*-1/2}]$, where, apart from the case of constant conditional correlations, a.s. $Q_t^{*-1/2} Q_t Q_t^{*-1/2} \neq Q_t$.

Regarding the VC model, taking the expectations of both members of (6) under stationarity yields

$$S = \frac{1-\beta}{1-\alpha-\beta} E[\varepsilon_t \varepsilon_t'] - \frac{\alpha}{1-\alpha-\beta} E[\mathcal{R}_{M,t}]. \tag{9}$$

Again, S is neither easy to interpret, nor easy to estimate.

2.1.3 Testing for integrated correlations.

In analogy with $q_{ij,t}$, for $\alpha + \beta = 1$ the process $\rho_{ij,t}$ is purported to exhibit "integrated" dynamics (Engle and Sheppard 2001; Pesaran and Pesaran 2007). The sense in which $\rho_{ij,t}$ should be considered as integrated is left unclear. The next proposition provides an answer to such a question.

Proposition 2.2. *Suppose that $\alpha > 0$ and $\alpha + \beta = 1$. Then, for $i \neq j$, it holds*

$$\rho^2_{ij,t+1} = E_t\left[\rho^2_{ij,t+1}\right] < E_t\left[\rho^2_{ij,t+2}\right] < E_t\left[\rho^2_{ij,t+3}\right] < \cdots < \lim_{m \longrightarrow \infty} E_t\left[\rho^2_{ij,t+m}\right].$$

Thus, for $\alpha + \beta = 1$, the asset cross-dependence is expected to increase with the forecast horizon. Unfortunately, the behavior of $\rho_{ij,t}$ for $\alpha + \beta < 1$ is unknown. Therefore, at least from a theoretical point of view, testing for DCC integrated correlations should be considered as inconclusive.

2.2 The *c*DCC model

The tractability of the DCC model can be substantially improved by reformulating the correlation driving process as

$$Q_t = (1-\alpha-\beta)\, S + \alpha \left\{ Q^{*1/2}_{t-1} \varepsilon_{t-1} \varepsilon'_{t-1} Q^{*1/2}_{t-1} \right\} + \beta\, Q_{t-1} \tag{10}$$

(Aielli 2006). The resulting model is called *c*DCC model, where *c* stands for *corrected*. Pre- and post-multiplying both members of (10) by a diagonal matrix Z yields an analogous formula, where S, Q_t and Q^*_t are replaced by $\bar{S} \equiv ZSZ$, $\bar{Q}_t \equiv ZQ_tZ$, and $\bar{Q}^*_t \equiv ZQ^*_tZ$, respectively. Since $R_t = Q_t^{*-1/2} Q_t Q_t^{*-1/2} = \bar{Q}_t^{*-1/2} \bar{Q}_t \bar{Q}_t^{*-1/2}$, the parameters (S, α, β) and $(\bar{S}, \alpha, \beta)$ are indistinguishable from ε_t. The identifiability of (S, α, β) is guaranteed by the assumption that S is unit-diagonal. We notice that, with the *c*DCC model, such an assumption is an innocuous identification condition, whereas, with the DCC model, it is an overidentifying restriction.

Some simulated series of conditional correlations are plotted in Fig. (1). The response of the *c*DCC $\rho_{ij,t}$ to the variations of α and β is analogous to the response of the DCC $\rho_{ij,t}$.

One might argue that the *c*DCC model is not really a correlation model in that R_t is modeled implicitly, as a byproduct of the model of Q_t. Indeed, this is not the case. An explicit representation of $\rho_{ij,t}$ can in fact be obtained by dividing the numerator and denominator of the right hand side of $\rho_{ij,t} = q_{ij,t}/\sqrt{q_{ii,t} q_{jj,t}}$ by $\sqrt{q_{ii,t-1} q_{jj,t-1}}$. This yields

$$\rho_{ij,t} = \frac{\omega_{ij,t-1} + \alpha\, \varepsilon_{i,t-1} \varepsilon_{j,t-1} + \beta \rho_{ij,t-1}}{\sqrt{\left\{\omega_{ii,t-1} + \alpha\, \varepsilon^2_{i,t-1} + \beta \rho_{ii,t-1}\right\}\left\{\omega_{jj,t-1} + \alpha\, \varepsilon^2_{j,t-1} + \beta \rho_{jj,t-1}\right\}}}, \tag{11}$$

where $\omega_{ij,t} \equiv (1-\alpha-\beta) s_{ij} / \sqrt{q_{ii,t}\, q_{jj,t}}$. Since $E_{t-1}[\varepsilon_{i,t} \varepsilon_{j,t}] = \rho_{ij,t}$, the formula of $\rho_{ij,t}$ proves to combine a sort of GARCH devices for the relevant past values and innovations into a correlation-like ratio. The parameters α

and β, originally related to Q_t, prove to be the dynamic parameters of the correlation GARCH devices. The time-varying intercept, $\omega_{ij,t}$, can be seen as an *ad hoc* correction required for purposes of tractability.

The DCC analog of (11) is

$$\rho_{ij,t} = \frac{\omega_{ij,t-1} + \alpha_{t-1}\varepsilon_{i,t-1}\varepsilon_{j,t-1} + \beta\rho_{ij,t-1}}{\sqrt{\left\{\omega_{ii,t-1} + \alpha_{t-1}\varepsilon^2_{i,t-1} + \beta\rho_{ii,t-1}\right\}\left\{\omega_{jj,t-1} + \alpha_{t-1}\varepsilon^2_{j,t-1} + \beta\rho_{jj,t-1}\right\}}},$$

where $\omega_{ij,t} \equiv (1-\alpha-\beta)\, s_{ij}/\sqrt{q_{ii,t}\, q_{jj,t}}$ and $\alpha_t \equiv \alpha/\sqrt{q_{ii,t}\, q_{jj,t}}$. Compared with the cDCC formula, the DCC formula involves more time-varying parameters, none of them being supported by any apparent motivation. Regarding the VC model (see eq. (6)), $\rho_{ij,t}$ is actually modeled explicitly, but at the cost of an *ad hoc* innovation term, $\mathcal{R}_{M,t}$, which makes the model difficult to deal with (see sec. 2.1.1, 2.1.2, and 3.2).

Setting $\varepsilon^*_t \equiv Q^{*1/2}_t\varepsilon_t$, we get $E_{t-1}[\varepsilon^*_t] = 0$ and $E_{t-1}[\varepsilon^*_t\varepsilon^{*\prime}_t] = Q_t$, where $Q_t = (1-\alpha-\beta)\,S + \alpha\,\varepsilon^*_{t-1}\varepsilon^{*\prime}_{t-1} + \beta\,Q_{t-1}$. Hence, ε^*_t follows a linear MGARCH process (Engle and Kroner 1995). This argument suggests the assumption of a general linear MGARCH model for ε^*_t as a practical way to extend the cDCC model in (10). Thus, modeling ε^*_t as a Baba-Engle-Kraft-Kroner (BEKK) model (Engle and Kroner 1995), delivers

$$Q_t = C + \sum_{\mathrm{q}=1}^{\bar{\mathrm{Q}}}\sum_{\mathrm{k}=1}^{\mathrm{K}} \bar{A}_{\mathrm{q,k}}\,\{Q^{*1/2}_{t-\mathrm{q}}\,\varepsilon_{t-\mathrm{q}}\varepsilon'_{t-\mathrm{q}}Q^{*1/2}_{t-\mathrm{q}}\}\,\bar{A}'_{\mathrm{q,k}} + \sum_{\mathrm{p}=1}^{\bar{\mathrm{P}}}\sum_{\mathrm{k}=1}^{\mathrm{K}} \bar{B}_{\mathrm{p,k}}\,Q_{t-\mathrm{p}}\,\bar{B}'_{\mathrm{p,k}}, \tag{12}$$

where

$$C \equiv S - \sum_{\mathrm{q}=1}^{\bar{\mathrm{Q}}}\sum_{\mathrm{k}=1}^{\mathrm{K}} \bar{A}_{\mathrm{q,k}}\,S\,\bar{A}'_{\mathrm{q,k}} - \sum_{\mathrm{p}=1}^{\bar{\mathrm{P}}}\sum_{\mathrm{k}=1}^{\mathrm{K}} \bar{B}_{\mathrm{p,k}}\,S\,\bar{B}'_{\mathrm{p,k}}. \tag{13}$$

For this model, Q_t is pd provided that C is pd. We will refer to model (12) as the BEKK cDCC model. A special case of the BEKK cDCC model is the Diagonal Vech cDCC model (Bollerslev 1990), which can be written as

$$Q_t = (\iota\iota' - \sum_{\mathrm{q}=1}^{\mathrm{Q}} A_{\mathrm{q}} - \sum_{\mathrm{p}=1}^{\mathrm{P}} B_{\mathrm{p}}) \odot S + \sum_{\mathrm{q}=1}^{\mathrm{Q}} A_{\mathrm{q}} \odot \{Q^{*1/2}_{t-\mathrm{q}}\,\varepsilon_{t-\mathrm{q}}\,\varepsilon'_{t-\mathrm{q}}Q^{*1/2}_{t-\mathrm{q}}\} + \sum_{\mathrm{p}=1}^{\mathrm{P}} B_{\mathrm{p}} \odot Q_{t-\mathrm{p}}. \tag{14}$$

For this model, Q_t is pd provided that the intercept is pd, $A_{\mathrm{q}} \equiv [\alpha_{ij,\mathrm{q}}]$ is psd for $\mathrm{q} = 1, 2, \ldots, \mathrm{Q}$, and $B_{\mathrm{p}} \equiv [\beta_{ij,\mathrm{p}}]$ is psd for $\mathrm{p} = 1, 2, \ldots, \mathrm{P}$ (Ding and Engle 2001). As an identification condition, we set $s_{ii} = 1$ for $i = 1, 2, \ldots, N$. It can easily be seen that the representation of $\rho_{ij,t}$ in terms of correlation GARCH devices, given in (11), directly extends to the Diagonal Vech cDCC model. If $A_{\mathrm{q}} = \alpha_{\mathrm{q}}\,\iota\iota'$ for $\mathrm{q} = 1, 2, \ldots, \mathrm{Q}$, and $B_{\mathrm{p}} = \beta_{\mathrm{p}}\,\iota\iota'$ for $\mathrm{p} = 1, 2, \ldots, \mathrm{P}$, we get the Scalar cDCC model. For this model, Q_t is pd if $\alpha_{\mathrm{q}} \geq 0$ for $\mathrm{q} = 1, 2, \ldots, \mathrm{Q}$, $\beta_{\mathrm{p}} \geq 0$ for $\mathrm{p} = 1, 2, \ldots, \mathrm{P}$, $\sum_{\mathrm{q}=1}^{\mathrm{Q}} \alpha_{\mathrm{q}} + \sum_{\mathrm{p}=1}^{\mathrm{P}} \beta_{\mathrm{p}} < 1$, and S is pd. For $\mathrm{Q} = \mathrm{P} = 1$, one gets the model in (10).

2.2.1 *c*DCC stationarity principle.

The next proposition draws on the modular structure of the cDCC model to establish cDCC stationarity conditions as a modular stationarity principle (cDCC stationarity principle). Loosely speaking, if (R_t, ε_t) is stationary and $h_{i,t}$ is stationary for $i = 1, 2, \ldots, N$, then (H_t, y_t) is stationary. From a modeling perspective, a useful consequence of this is that, if the correlation-return process is stationary, any functional form of the variance processes is allowed for (e.g., GARCH, TGARCH, EGARCH,...) without prejudice for the stationarity of the covariance-return process, provided that the single variance processes are stationary under the assumption of univariate stationary standardized innovations. We notice that this property is not peculiar to the cDCC dynamic conditional correlation model, but it is true, e.g., for the DCC and VC models as well. With the cDCC model, however, it takes on a practical interest because the stationarity conditions for the cDCC specification of (R_t, ε_t) can be explicitly derived from a recent result of Boussama et al. (in press). Before stating the proposition, we need to introduce some notations. For the BEKK cDCC model in (12), set $\mathbf{A}_{\mathrm{q}} \equiv U\{\sum_{\mathrm{k}=1}^{\mathrm{K}} A_{\mathrm{q,k}} \otimes A_{\mathrm{q,k}}\}W'$ for $\mathrm{q} = 1, 2, \ldots, \bar{\mathrm{Q}}$, and $\mathbf{B}_{\mathrm{p}} \equiv U\{\sum_{\mathrm{k}=1}^{\mathrm{K}} B_{\mathrm{p,k}} \otimes B_{\mathrm{p,k}}\}W'$ for $\mathrm{p} = 1, 2, \ldots, \bar{\mathrm{P}}$, where $\otimes$ denotes the Kronecker product. The matrices U and W are the unique $N(N+1)/2 \times N^2$ matrices such that $\mathrm{vech}(Z) = U\mathrm{vec}(Z)$, $\mathrm{vec}(Z) = W'\mathrm{vech}(Z)$, and $UW' = I_{N(N+1)/2}$ for any N-dimensional symmetric matrix Z, where vec and vech, respectively, denote the operators stacking the columns and the lower triangular part of the matrix argument. The existence and uniqueness of U and W hold by linearity of vech and vec. Let

$$h_{i,t} = \mathscr{H}_i(\theta_i; \varepsilon_{i,t-1}, \varepsilon_{i,t-2}, \ldots)$$

denote the representation of $h_{i,t}$ in terms of standardized innovations, obtained by recursively substituting backward for lagged $y_{i,t} = \varepsilon_{i,t}\sqrt{h_{i,t}}$ into $h_{i,t} = h_i(\theta_i; y_{i,t-1}, y_{i,t-2}, \ldots)$. We assume that $\varepsilon_t = R_t^{1/2}\eta_t$, where $R_t^{1/2}$ is the unique psd matrix such that $R_t^{1/2}R_t^{1/2} = R_t$, and η_t is iid such that $E[\eta_t] = 0$ and $E[\eta_t\eta_t'] = I_N$.

Proposition 2.3. *(cDCC stationarity principle). For the BEKK cDCC model in (12), suppose that:*

H1) the density of η_t is absolutely continuous with respect to the Lebesgue measure, positive in a neighborhood of the origin;

H2) C is pd, and the largest eigenvalue of $\sum_{\mathrm{q}=1}^{\bar{\mathrm{Q}}} \mathbf{A}_{\mathrm{q}} + \sum_{\mathrm{p}=1}^{\bar{\mathrm{P}}} \mathbf{B}_{\mathrm{p}}$ is less than one in modulus.

Then, (i) the process $[\mathrm{vech}(R_t)', \varepsilon_t']'$ admits a non-anticipative, strictly and weakly stationary, and ergodic solution. In addition to H1-H2, suppose that:

H3) $\mathscr{H}_i(\theta_i; \varepsilon_{i,t-1}, \varepsilon_{i,t-2}, \ldots)$ is measurable for $i = 1, 2, \ldots, N$.

Then, (ii) the process $[\mathrm{vech}(H_t)', y_t', \mathrm{vech}(R_t)', \varepsilon_t']'$ admits a non-anticipative, strictly stationary, and ergodic solution. In addition to H1-H3, suppose that:

H4) $E[y_{i,t}^2] < \infty$ for $i = 1, 2, \ldots, N$.

Then, (iii) the process y_t admits a weakly stationary solution.

Regarding the correlation assumptions, namely, H1-H2, we notice that H1 is typically fulfilled in any practical application, and that H2 is easy to check. For example, with the Diagonal Vech cDCC model, H2 is equivalent to the assumption that $\left(\iota\iota' - \sum_{\mathrm{q}=1}^{\mathrm{Q}} A_{\mathrm{q}} - \sum_{\mathrm{p}=1}^{\mathrm{P}} B_{\mathrm{p}}\right) \odot S$ is pd and $\sum_{\mathrm{q}=1}^{\mathrm{Q}} \alpha_{ii,\mathrm{q}} + \sum_{\mathrm{p}=1}^{\mathrm{P}} \beta_{ii,\mathrm{p}} < 1$ for $i = 1, 2, \ldots, N$. The specific form of the variance assumptions, H3-H4, depends on the model of $h_{i,t}$ (see Francq and Zakoïan 2010 for a survey). As for H3, the point to keep in mind is that $\mathscr{H}_{i,t}$ must be measurable under strictly stationary $\varepsilon_{i,t}$, which is a less stringent assumption than the usual one of iid $\varepsilon_{i,t}$. For example, in the GARCH(1,1) model (Bollerslev 1986), $h_{i,t} = c_i + a_i\, y_{i,t-1}^2 + b_i\, h_{i,t-1}$, where $c_i > 0$, $a_i \geq 0$, and $b_i \geq 0$, $\mathscr{H}_{i,t}$ is measurable if $E[\log(a_i\,\varepsilon_{i,t}^2 + b_i)] < 0$ and $E[\max\{0, \log(a_i\,\varepsilon_{i,t}^2 + b_i)\}] < \infty$ (Brandt 1986 and Bougerol and Picard 1992). If $\varepsilon_{i,t}$ is iid (which holds if η_t is Gaussian), it suffices that $E[\log(a_i\,\varepsilon_{i,t}^2 + b_i)] < 0$. As for H4, by a well known result, it follows that $E[y_{i,t}^2] < \infty$ if $a_i + b_i < 1$.

2.2.2 Formula and interpretation of S.

Proposition 2.4. *For the BEKK* c*DCC model in (12-13), suppose that assumptions H1-H2 of prop. 2.3 hold. Then, $Q_t^{*1/2}\varepsilon_t$ is covariance stationary, and*

$$S = E[Q_t^{*1/2}\varepsilon_t\,\varepsilon_t'\,Q_t^{*1/2}]. \tag{15}$$

Noting that $E[Q_t^{*1/2}\varepsilon_t\varepsilon_t'Q_t^{*1/2}] = E[\varepsilon_t^*\varepsilon_t^{*\prime}] = E[E_{t-1}[\varepsilon_t^*\varepsilon_t^{*\prime}]] = E[Q_t]$, from eq. (15) it follows that $S = E[Q_t]$. Thus, S is the expectation of the correlation driving process. We notice that, as with the DCC and VC models, with the cDCC model S is not the second moment of ε_t. Nevertheless, relying on the simple structure of (15), a psd intuitive large system estimator can be easily constructed (see sec. 3.2).

2.2.3 Testing for integrated correlations.

Proposition 2.5. *(Test of* c*DCC integrated correlations). For the* c*DCC model in (10), suppose that: H1) the density of η_t is absolutely continuous with respect to the Lebesgue measure, positive in a neighborhood of the origin, H2) $\alpha > 0$, and H3) S is pd. Then, for $i \neq j$, it holds that*

(i) if $\alpha + \beta < 1$, then $\rho_{ij,t}^2$ is strictly and weakly stationary, and $\lim_{m\longrightarrow\infty} E_t\,[\rho_{ij,t+m}^2] = E[\rho_{ij,t}^2] < 1$;

(ii) if $\alpha + \beta = 1$, then $\rho_{ij,t+1}^2 = E_t\left[\rho_{ij,t+1}^2\right] < E_t\left[\rho_{ij,t+2}^2\right] < E_t\left[\rho_{ij,t+3}^2\right] < \cdots < \lim_{m\longrightarrow\infty} E_t\left[\rho_{ij,t+m}^2\right]$.

In the long-run, the non-integrated $\rho_{ij,t+m}^2$ is expected to either increase or decrease; the sign of the expected dynamics depends on the current state, $\rho_{ij,t+1}^2$, and the stationary state, $E[\rho_{ij,t}^2]$, which is proved to be non-degenerate. Conversely, in the long-run, the integrated $\rho_{ij,t+m}^2$ is in any case expected to increase.

We notice that, for $\alpha + \beta = 1$, the intercept of the GARCH(1,1) process $q_{ii,t}$ is zero, which implies that a.s. $q_{ii,t} \longrightarrow 0$ for $t \longrightarrow \infty$ (Nelson 1990, Proposition 1). This can cause numerical problems during the computation of $\rho_{ij,t} = q_{ij,t}/\sqrt{q_{ii,t}q_{jj,t}}$ for large t and $\alpha + \beta = 1$. The DCC model is free from such a drawback in that, for $\alpha + \beta = 1$, the DCC $q_{ii,t}$ is AR(1) stationary (see sec. 2.1.1).

3. LARGE SYSTEM ESTIMATION

3.1 The DCC estimator

Let $L_T(\theta, S, \phi) = \sum_{t=1}^T l_t(\theta, S, \phi)$ denote the error correction decomposition of the DCC quasi-log-likelihood (QLL), where $\theta \equiv [\theta_1', \ldots, \theta_N']'$ and $\phi \equiv [\alpha, \beta]'$. Thus, $l_t(\theta, S, \phi) = -(1/2)\left\{N\log(2\pi) + \log|\tilde{H}_t| + y_t'\,\tilde{H}_t^{-1}y_t\right\}$, where $\tilde{H}_t \equiv \tilde{D}_t^{1/2}\tilde{R}_t\tilde{D}_t^{1/2}$, where $\tilde{D}_t$ and $\tilde{R}_t$, respectively, stand for D_t and R_t evaluated at (θ, S, ϕ) (see eq. (1-4)). Because of the presence of S, which includes $N(N-1)/2$ distinct parameters to estimate, the joint quasi-maximum-likelihood (QML) estimation of the DCC model is infeasible for large N. As a feasible estimator, Engle (2002) suggested a three-step procedure called DCC estimator. Before introducing it, we first define an estimator of S conditional on θ.

Definition 3.1. *(DCC conditional estimator of S). For fixed θ, set $\tilde{S}_\theta \equiv T^{-1}\sum_{t=1}^T \tilde{\varepsilon}_t\tilde{\varepsilon}_t'$, where $\tilde{\varepsilon}_t \equiv [\tilde{\varepsilon}_{1,t}, \ldots, \tilde{\varepsilon}_{N,t}]'$, $\tilde{\varepsilon}_{i,t} \equiv y_{i,t}/\sqrt{\tilde{h}_{i,t}}$, and $\tilde{h}_{i,t} \equiv h_{i,t}(\theta_i; y_{i,t-1}, y_{i,t-2}, \ldots)$.*

$\tilde{S}_\theta$ is the sample second moment of the standardized returns evaluated at θ. Alternatively, $\tilde{S}_\theta$ can be defined as the sample (centered or uncentered) correlation of $\tilde{\varepsilon}_t$.

Definition 3.2. *(DCC estimator).*

Step (1): set $\hat{\theta} \equiv [\hat{\theta}'_1, \hat{\theta}'_2, \ldots, \hat{\theta}'_N]'$*, where* $\hat{\theta}_i$ *is the univariate QML estimator of* θ_i*,* $i = 1, 2, \ldots, N$*;*

Step (2): set $\hat{S} \equiv \tilde{S}_{\hat{\theta}}$*;*

Step (3): set $\hat{\phi} \equiv \operatorname{argmax}_{\phi} L_T(\hat{\theta}, \hat{S}, \phi)$*, subject to* $\alpha \geq 0$*,* $\beta \geq 0$ *and* $\alpha + \beta < 1$*.*

A common choice for the initial values of the correlation recursions is to set $\varepsilon_{i,0} = 0$ for $i = 1, 2, \ldots, N$ and $Q_0 = \hat{S}$. The choice of the initial values for the variance recursions depends on the model of $h_{i,t}$. To alleviate the bias of $\hat{\phi}$ due to the presence of the large dimensional estimated nuisance parameter $\hat{S}$, Engle et al. (2009) suggested replacing $L_T(\theta, S, \phi)$ with the so-called bivariate composite DCC QLL,

$$\sum_{i=2}^{N} L_{T,i,i-1}(\theta, S, \phi), \tag{16}$$

where $L_{T,i,i-1}(\theta, S, \phi)$ denotes the bivariate QLL of the DCC submodel of $(y_{i,t}, y_{i-1,t})$.

3.1.1 Inconsistency of the DCC estimator.

The consistency conditions for $\hat{\theta}$ depend on the GARCH models of the conditional variances (see Francq and Zakoïan 2010 for a survey). Regarding $\hat{S}$, we have proven in section 2.1.2 that the equality $S^0 = E[\varepsilon_t \varepsilon'_t]$ does not hold in general (thereafter a superscript zero will denote true values.). If $S^0 \neq E[\varepsilon_t \varepsilon'_t]$ and $\operatorname{plim} \tilde{S}_\theta$ is finite in a neighborhood of θ^0, under consistency of $\hat{\theta}$ it follows that $\operatorname{plim} \hat{S} = \operatorname{plim} \tilde{S}_{\hat{\theta}} \neq S^0$ (Wooldridge 1994, Lemma A.1). As for $\hat{\phi}$, we notice that, unless S and ϕ are proven to be orthogonal, if $\hat{S}$ is inconsistent, $\hat{\phi}$ is inconsistent in turn (Newey and McFadden 1994, sec. 6.2, p. 2179).

3.1.2 Inferences from DCC estimations.

The inconsistency of $\hat{S}$ is a potential cause of inconsistent inferences. Following Engle (2002, p. 342, eq. 33 — see, e.g., Franses and Hafner 2009), the standard errors of $\hat{\phi}$ are corrected for the estimation of (θ, S) relying on the formulas in Newey and McFadden (1994, Theorem 6.1; analogous formulas are derived by Engle and Sheppard 2001 relying on White 1996). Unfortunately, for such formulas to hold, a consistent estimator of (θ, S) is generally required.

3.2 The cDCC estimator

As a large system estimator for the cDCC model, Aielli (2006) suggested the cDCC estimator, an *ad hoc* generalized profile QLL estimator which can be directly extended to the Diagonal Vech cDCC model. In this section, $L_T(\theta, S, \phi) = \sum_{t=1}^{T} l_t(\theta, S, \phi)$ will denote the error-correction decomposition of the QLL of the Diagonal Vech cDCC model, where ϕ is the vector stacking the lower triangle of the model dynamic parameter matrices. Thus, (θ, S, ϕ) relates to $\tilde{H}_t = \tilde{D}_t^{1/2} \tilde{R}_t \tilde{D}_t^{1/2}$ according to eq. (1-3) and (14). The parameter space of ϕ and θ will be denoted by Φ and Θ, respectively. Before introducing the cDCC estimator, we first define an estimator of S conditional on (θ, ϕ).

Definition 3.3. *(cDCC conditional estimator of S). For fixed (θ, ϕ), set $\tilde{S}_{\theta,\phi} \equiv T^{-1} \sum_{t=1}^{T} \tilde{Q}_t^{*1/2}\, \tilde{\varepsilon}_t\, \tilde{\varepsilon}_t'\, \tilde{Q}_t^{*1/2}$, where $\tilde{\varepsilon}_t$ is defined as in def. 3.1, and $\tilde{Q}_t^* \equiv \mathrm{diag}(\tilde{q}_{11,t}, \ldots, \tilde{q}_{NN,t})$, where*

$$\tilde{q}_{ii,t} = (1 - \sum_{\mathrm{q}=1}^{\mathrm{Q}} \alpha_{ii,\mathrm{q}} - \sum_{\mathrm{p}=1}^{\mathrm{P}} \beta_{ii,\mathrm{p}})\, s_{ii} + \sum_{\mathrm{q}=1}^{\mathrm{Q}} \alpha_{ii,\mathrm{q}}\, \tilde{\varepsilon}_{i,t-\mathrm{q}}^2\, \tilde{q}_{ii,t-\mathrm{q}} + \sum_{\mathrm{p}=1}^{\mathrm{P}} \beta_{ii,\mathrm{p}}\, \tilde{q}_{ii,t-\mathrm{p}}. \tag{17}$$

$\tilde{S}_{\theta,\phi}$ is the sample second moment of S conditional on (θ, ϕ) (see eq. (15)). By construction, $\tilde{S}_{\theta,\phi}$ is pd. Note that $\tilde{S}_{\theta,\phi}$ is fast to compute as it simply requires running N univariate recursions for $\tilde{\varepsilon}_t$ plus N univariate recursions for $\tilde{Q}_t$. By unit variance of $\varepsilon_{i,t}\sqrt{q_{ii,t}}$ (see again (15)), we can alternativly define $\tilde{S}_{\theta,\phi}$ as the sample (centered or uncentered) correlation of $\tilde{Q}_t^{*1/2}\, \tilde{\varepsilon}_t$. In this case, $\tilde{S}_{\theta,\phi}$ will be unit-diagonal, like S.

Definition 3.4. *(cDCC estimator).*

Step (1): set $\hat{\theta}$ as in def. 3.2;

Step (2): set $\hat{\phi} \equiv \mathrm{argmax}_{\phi \in \Phi}\ L_T(\hat{\theta}, \tilde{S}_{\hat{\theta},\phi}, \phi)$;

Step (3): set $\hat{S} \equiv \tilde{S}_{\hat{\theta},\hat{\phi}}$.

The objective function for ϕ (see Step 2) is an example of *generalized profile* QLL (Severini 1998). The parameter ϕ is the parameter of interest and (θ, S) is the nuisance parameter. The estimator of the nuisance parameter conditional on the parameter of interest is defined as $(\hat{\theta}, \hat{S}_{\hat{\theta},\phi})$, where only $\hat{S}_{\hat{\theta},\phi}$ depends on ϕ. Once $\hat{\phi}$ is computed, the unconditional estimate of the nuisance parameter is computed as $(\hat{\theta}, \hat{S}_{\hat{\theta},\hat{\phi}})$, which is nothing but the value of the conditional estimator of the nuisance parameter at the end of the maximization process. Thereafter, the objective function for ϕ will be referred to as the cDCC profile QLL (cDCC PQLL for short).

For suitably restricted ϕ, the cDCC PQLL is a function of a relatively small number of parameters. For example, in the case of the Scalar Vech cDCC model, the cDCC PQLL is a function of P+Q parameters, where P and Q are typically small. More general models can be estimated assuming reduced rank Diagonal Vech dynamic parameter matrices, and/or the block partition of Billio et al. (2006). Under appropriate conditions (Newey and McFadden 1994, Theorem 3.5), a Newton-Raphson one-step iteration from the cDCC estimator will deliver an estimator with the same asymptotic efficiency as the joint QML estimator. The bivariate composite version of the cDCC estimator is obtained by profiling the cDCC analog of (16) in place of the cDCC QLL.

We notice that the DCC estimator is a generalized profile QLL estimator where the conditional estimator of the nuisance parameter, $(\hat{\theta}, \hat{S})$, does not depend on ϕ. At least in principle, a DCC estimator totally analogous to the cDCC estimator can be defined setting the DCC conditional estimator of S to the sample counterpart of (8) for fixed (θ, ϕ). Unfortunately, such an estimator is not always psd. Similar problems also arise with the VC model (see eq. 9).

3.2.1 Consistency of the *c*DCC estimator: a heuristic proof.

Proposition 3.1. *For the Diagonal Vech cDCC model in (14), suppose that assumptions H1-H2 of prop. 2.3 hold. Then,* $\mathrm{plim}\, \tilde{S}_{\theta^0,\phi^0} = S^0$.

Relying on prop. 3.1, we can provide a heuristic proof of the consistency of the cDCC estimator (see also Fig. (2)).

The cDCC estimator is the maximizer of the cDCC QLL subject to $\{\theta = \hat{\theta}, S = \tilde{S}_{\hat{\theta},\phi}, \phi \in \Phi\}$. If $\hat{\theta}$ is consistent and $\text{plim}\, \tilde{S}_{\theta,\phi}$ is finite for all (θ, ϕ), then

$$\text{plim}\, \{\theta = \hat{\theta}, S = \tilde{S}_{\hat{\theta},\phi}, \phi \in \Phi\} = \{\theta = \theta^0, S = \text{plim}\, \tilde{S}_{\theta^0,\phi}, \phi \in \Phi\}$$

(Wooldridge 1994, Lemma A.1). Since $\text{plim}\, \tilde{S}_{\theta^0,\phi^0} = S^0$ (see prop. 3.1), the limit in probability of the cDCC constraint is a *correctly* specified constraint. Therefore, if $\text{plim}\, T^{-1} L_T(\theta, S, \phi)$ is finite for all (θ, S, ϕ) and uniquely maximized in (θ^0, S^0, ϕ^0) (which is a common assumption in QML settings — see Bollerslev and Wooldridge 1992), $\text{plim}\, T^{-1} L_T(\hat{\theta}, \tilde{S}_{\hat{\theta},\phi}, \phi)$ is uniquely maximized in ϕ^0 (Wooldridge 1994, Lemma A.1). This proves the consistency of $\hat{\phi}$ provided that the convergence of the scaled cDCC PQLL to its limit is uniform (Newey and McFadden 1994, Theorem 2.1). As for $\hat{S}$, if $\text{plim}\, \tilde{S}_{\theta,\phi}$ is finite for all (θ, ϕ) and $\text{plim}(\hat{\theta}, \hat{\phi}) = (\theta^0, \phi^0)$, then $\text{plim}\, \hat{S} = \text{plim}\, \hat{S}_{\hat{\theta},\hat{\phi}} = S^0$ (Wooldridge, 1994, Lemma A.1).

We notice that a consistency proof like that above would not work with the DCC estimator, in that, if $\text{plim}\, \hat{S} \neq S^0$, the limit in probability of the DCC constraint, $\{\theta = \theta^0, S = \text{plim}\, \hat{S}, \phi \in \Phi\}$, is a misspecified constraint.

3.2.2 Inferences from cDCC estimations.

Let s, $\tilde{s}(\theta, \phi)$ and $\lambda_t(\theta, \phi)$ denote the vectors stacking the lower off-diagonal elements of S, $\tilde{S}_{\theta,\phi}$ and $\tilde{Q}_t^{*1/2}\, \tilde{\varepsilon}_t\, \tilde{\varepsilon}_t'\, \tilde{Q}_t^{*1/2}$, respectively. Let $\gamma \equiv [\theta', \phi', s']'$, $l_t(\theta, s, \phi)$, and $l_{i,t}(\theta_i) \equiv -(1/2)\{\log(2\pi) + \log \tilde{h}_{i,t} + y_{i,t}^2/\tilde{h}_{i,t}\}$, where $i = 1, 2, \ldots, N$, denote the cDCC parameter, the individual cDCC QLL at time t, and the i-th individual GARCH QLL at time t. The cDCC estimator, denoted as $\hat{\gamma} \equiv [\hat{\theta}', \hat{\phi}', \hat{s}']$, is a solution of the estimating equations $T^{-1} \sum_{t=1}^{T} g_t(\gamma) = 0$, where the individual score is defined as $g_t(\gamma) \equiv [\{u_t(\theta)\}', \{v_t(\theta, \phi)\}', \{p_t(\theta, \phi, s)\}']'$, where $p_t(\theta, \phi, s) \equiv \lambda_t(\theta, \phi) - s$, $v_t(\theta, \phi) \equiv (\partial/\partial\phi) l_t(\theta, \tilde{s}(\theta, \phi), \phi)$, and $u_t(\theta) = [\{u_{1,t}(\theta_1)\}', \ldots, \{u_{N,t}(\theta_N)\}']'$, where $u_{i,t} = (\partial/\partial\theta_i) l_{i,t}(\theta_i)$. Under the conditions in Newey and McFadden (1994, Theorem 3.2, with $\hat{W}$ replaced by the identity), it follows that $\sqrt{T}(\hat{\gamma} - \gamma^0) \overset{A}{\approx} N(0, \{\mathcal{A}^0\}^{-1} \mathcal{B}^0 \{\mathcal{A}^{0\prime}\}^{-1})$, where

$$\mathcal{A}^0 \equiv \text{plim} \left\{ T^{-1} \sum_{t=1}^{T} \frac{\partial}{\partial \gamma'}\, g_t(\gamma) \bigg|_{\gamma=\gamma^0} \right\} \;\text{ and }\; \mathcal{B}^0 \equiv \lim_{T \longrightarrow \infty} VAR \left[T^{-1/2} \sum_{t=1}^{T} g_t(\gamma^0) \right].$$

The matrix $\mathcal{A}^0$ is block lower triangular with block partition determined by the partition of γ into θ, ϕ, and s. It can easily be seen that the standard errors of $\hat{\theta}_i$ are the traditional univariate QML standard errors, and that the standard errors of $\hat{\phi}$ do not depend on derivatives with respect to the $\mathcal{O}(N^2)$ term, s. Estimates of the standard errors of $\hat{\theta}$, $\hat{\phi}$ and — if needed — $\hat{s}$, can be computed by replacing the relevant blocks of $\mathcal{A}^0$ and $\mathcal{B}^0$ with the sample counterparts evaluated at $\hat{\gamma}$. For $\mathcal{B}^0$, heteroskedastic and autocorrelation (HAC)-robust estimates (see, e.g., Newey and West 1987) are required. If $\tilde{S}_{\theta,\phi}$ is computed as a sample (centered or uncentered) correlation, the individual score of s — namely, $p_t(\theta, \phi, s)$ — will be designed accordingly.

4. EMPIRICAL APPLICATIONS

To compare the empirical performances of the DCC and cDCC estimators we use a MATLAB code based on a sequential quadratic programming optimizer. As a starting point for the estimations of the correlation parameters, we use the true value of the data generating process (DGP) in the simulations under correctly specified model, and the maximizer of the objective function over a grid in the simulations under misspecification and in the applications to the real data. We impose $\epsilon \leq \alpha \leq 1-\epsilon$, $\epsilon \leq \beta \leq 1-\epsilon$, and $\alpha+\beta \leq 1-\epsilon$, where $\epsilon = 10^{-5}$, to get more stable maximizations. The conditional estimators of S are computed as centered correlations.

4.1 Simulations under correctly specified model

For $s \equiv s_{12} = 0, \pm.3, \pm.6, \pm.9$, $\alpha = .01, .04, .16$, and $\alpha + \beta = .8, .99, .998$, we generate $M = 500$ DCC bivariate Gaussian series of length $T = 1750$, discarding a burning period of 500 observations to alleviate the effect of the initial values. Since our focus is on the correlation fitting performances, the DCC estimator is computed assuming that the standardized innovations are known. An analogous experiment is run for the cDCC estimator. The box plots of the estimation error of $\hat{s}$ are presented in the panel of Fig. (3). Regarding the DCC estimator, $\hat{s}$ exhibits a positive bias for $s^0 < 0$ and a negative bias for $s^0 > 0$. Such a shrinkage effect increases with the persistence of the correlation process, as measured by $\alpha^0 + \beta^0$, and with the impact of the news, as measured by α^0. *Coeteris paribus*, for varying s^0 the bias moves according to a sinusoidal pattern. This is particularly evident for $\alpha^0 + \beta^0 = .998$ and $\alpha = .16$ (see the lower left plot of the panel). For typical values of the dynamic parameters (see the plots for $\alpha^0 + \beta^0 > .8$ and $\alpha^0 \leq .04$) the bias is negligible. For $s^0 = 0$, the estimator appears to be unbiased whatever the value of (α^0, β^0) is. The box plots of the relative estimation error of $\hat{\alpha}$ are presented in Fig. (4). For $\alpha^0 + \beta^0 = .998$ and $\alpha^0 = .16$, the box plots are symmetric and well centered around zero. The same holds for the estimation error of $\hat{\beta}$ (see the lower left plot of Fig. (5)), suggesting that the large bias of $\hat{s}$ for $\alpha^0 + \beta^0 = .998$ and $\alpha^0 = .16$ does not affect the performances of $(\hat{\alpha}, \hat{\beta})$. Fig. (6) presents the box plots of the mean absolute error of the conditional correlation estimator, $M^{-1}T^{-1}\sum_{m=1}^{M}\sum_{t=1}^{T}\left|\hat{\rho}_t - \rho_t\right|$, where $\hat{\rho}_t \equiv \hat{\rho}_{12,t}$ and $\rho_t \equiv \rho_{12,t}$. For $\alpha^0 + \beta^0 = .998$ and $\alpha^0 = .16$, in spite of the large bias of $\hat{s}$, the estimator exhibits the best performances.

The good performances of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\rho}_t$ for $\alpha^0 + \beta^0 = .998$ and $\alpha^0 = .16$, can be explained as follows. For large t, by backward substitutions we can approximate $q_{12,t}$ as

$$q_{12,t} \approx \frac{1-\alpha^0-\beta^0}{1-\beta^0}s^0 + \frac{\alpha^0}{1-\beta^0}\left\{(1-\beta^0)\sum_{n=1}^{t-1}\{\beta^0\}^{n-1}\varepsilon_{1,t-n}\,\varepsilon_{2,t-n}\right\}.$$

The right hand side is a weighted mean of s^0 and of the term in brackets. If $\alpha^0 + \beta^0$ and α^0 are both large — such as for $\alpha^0 + \beta^0 = .998$ and $\alpha^0 = .16$ — the weight of s^0 is small. In this case, the effect on the second-step estimates, $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\rho}_t$, due to replacing s^0 with a biased first-step estimate, like $\hat{s}$, will be small in turn.

The cDCC estimator of s does not exhibit any apparent bias (see Fig. (3)). All the related box plots are symmetric and well centered around zero. Regarding $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\rho}_t$, we notice that, apart from the two plots corresponding to $\alpha^0 + \beta^0 \leq .99$ and $\alpha^0 = .16$ (see the last row of the panels of Fig. (4), (5) and (6), respectively), the DCC estimator and the cDCC estimator provide practically identical performances. For $\alpha^0 + \beta^0 \leq .99$ and $\alpha^0 = .16$, instead, the cDCC estimator tends to underperform the DCC estimator.In this case, however, a comparison of the two estimators is not appropriate because of the different DGP's. This argument is illustrated in Fig. (7), where the ACF of $\rho_{ij,t}$ is reported. For $\alpha^0 + \beta^0 \leq .99$ and $\alpha^0 = .16$, the ACF's of the two models are different, denoting different DGP's. As expected, in the remaining plots, the ACF's of the two models are very similar if not identical.

4.2 Simulations under misspecification

Following Engle (2002), we generate $M = 500$ bivariate Gaussian series of length $T = 1000$, with GARCH(1,1) variances set as $h_{1,t} = .01 + .05\, y^2_{1,t-1} + .94\, h_{1,t-1}$ and $h_{2,t} = .30 + .20\, y^2_{2,t-1} + .50\, h_{2,t-1}$. The correlation series are set as $CONSTANT \equiv \{\rho_t = .9\}$, $STEP \equiv \{\rho_t = .9 - .5(t \geq 500)\}$, $FASTSINE \equiv \{\rho_t = .5 + .4.\cos(2\pi t/20)\}$, $SINE \equiv \{\rho_t = .5 + .4.\cos(2\pi t/200)\}$, and $RAMP \equiv \{\rho_t = \mathrm{mod}(t/200)/200\}$. Such correlation processes ensure a variety of dynamics, such as rapid changes, gradual changes, and periods of constancy. For performance measures we consider the following regression-based tests computed on portfolio returns, $w'y_t$, where w is the vector of portfolio weights (we recall that the conditional variance of $w'y_t$ is $w'H_t\, w$).

• **Engle-Colacito regression.** The Engle-Colacito (E&C) regression (Engle and Colacito 2006) is defined as $\{(w'y_t)^2\ /(w'\ \hat{H}_t w\ )\} - 1 = \lambda + \xi_t$, where ξ_t is an innovation term. The test is a test of the null hypothesis that $\lambda = 0$. An HAC-robust estimator of the standard error of ξ_t is required.

• **Dynamic Quantile test.** Denote the $\tau \times 100\%$-quantile of the conditional distribution of $w'y_t$ as $VaR_t(\tau)$ (where *VaR* stands for *Value at Risk*). For fixed τ, set $HIT_t \equiv 1$ if $w'y_t < VaR_t(\tau)$, and $HIT_t \equiv 0$ otherwise. By construction, $\{HIT_t - \tau\}$ is zero-mean iid. The Dynamic Quantile (DQ) test (Engle and Manganelli 2004) is an F-test of the null hypothesis that all coefficients, as well as the intercept, are zero in a regression of $\{HIT_t - \tau\}$ on past values, $VaR_t(\tau)$ and any other variables. We set $VaR_t(\tau) = -1.96\sqrt{w'\,\hat{H}_t w}$, which corresponds to the 2.5% estimated quantile under Gaussianity. We use five lags and the current estimated VaR as regressors.

• **LM test of ARCH effects.** The LM test of ARCH effects (Engle 1982) is based on the property that the series $(w'y_t)^2/(w'H_t w)$ does not exhibit serial correlation. The test is a test of the null hypothesis that $(w'y_t)^2/(w'\hat{H}_t w)$ is serially uncorrelated. In this paper five lags are used.

If the model is correctly specified and the estimates are set to the true values, the above tests are asymptotically normal. In our experiment, because of the model misspecification and the replacement of true quantities with estimated quantities, the size of the tests will likely be different from the nominal size. As portfolio specifications, we consider the equally weighted portfolio, denoted as EW, and the minimum variance portfolios with and without short selling, denoted as MV and *MV, respectively. The vector of EW weights is known and equal to ι/N. The MV and *MV weights, which depend on the unknown H_t, are computed from the estimate of H_t. For each of the two estimators, we have $5 \times 3 \times 3\,(= 45)$ tests to compute, where 5 is the number of correlation specifications, 3 is the number of regression-based tests, and 3 is the number of considered portfolios. For each test, we compute the percentage of rejections at a 5% level on the generated series. The more rejections, the more evidence of misspecification. All the resulting percentages are close to or less than the nominal size, and they are practically the same for the two estimators (the related table is not reported here for reasons of space). For each test, the null hypothesis of equal percentage of rejections for the two estimators is not rejected at any standard levels.

4.3 Applications to real data

We consider two datasets extracted from Datastream. The first dataset includes the S&P 500 composite index and the related nine SPDR sector indices for a total of $N = 10$ assets. The sample period is May 2003 - February 2010, which results in $T = 1750$ daily returns. The second dataset includes $N = 100$ randomly selected equities from the S&P 1500 index industrial and consumer goods components. The sample period is

June 2003 - March 2010, again for a total of $T = 1750$ daily returns. The two estimators are computed in their bivariate composite versions (see sec. 3.1 and 3.2) assuming GARCH(1,1) variances. For performance measures, we consider three sets of out-of-sample forecast criteria, namely, (i) a set of Equal Predictive Ability (EPA) tests of one-step-ahead correlation forecasts, (ii) regression-based tests computed from one-step-ahead forecasts, and (iii) EPA tests of multi-step-ahead correlation forecasts.

4.3.1 EPA tests of one-step-ahead correlation forecasts.

Let $\hat{H}_{t|t-1}$ denote the DCC one-step-ahead estimate of H_t based on a rolling window of $\bar{T} < T$ excess returns. The excess returns of the rolling window, denoted as $\hat{y}_{t-j}$, $j = 1, 2, \ldots, \bar{T}$, are estimated setting $\hat{y}_{t-j} \equiv z_{t-j} - \bar{z}_{t-1}$, where z_{t-j} is the observed return at time $t-j$ and $\bar{z}_{t-1} \equiv \bar{T}^{-1} \sum_{j=1}^{\bar{T}} z_{t-j}$. We set $\bar{T} = 1250$, which results in $T - \bar{T} = 500$ out-of-sample forecasts. As a mean square error (MSE) loss (Diebold and Mariano 1996) for the DCC forecasts of the EW conditional variance we set

$$d_t \equiv EWMSE_t \equiv \left((w'\hat{y}_{t|t-1})^2 - w'\,\hat{H}_{t|t-1} w \right)^2, \tag{18}$$

where $\hat{y}_{i,t|t-1} \equiv z_{i,t} - \bar{z}_{i,t-1}$. For the DCC correlation forecasts we define the MSE loss

$$d_t \equiv CORRMSE_t \equiv \frac{1}{N(N-1)/2} \sum_{i<j=2,\ldots,N} \left(\hat{\varepsilon}_{i,t|t-1}\,\hat{\varepsilon}_{j,t|t-1} - \hat{\rho}_{ij,t|t-1} \right)^2, \tag{19}$$

where $\hat{\rho}_{ij,t|t-1}$ is the (i,j)-th correlation associated to $\hat{H}_{t|t-1}$ and $\hat{\varepsilon}_{i,t|t-1} \equiv \hat{y}_{i,t|t-1} / \sqrt{\hat{h}_{i,t|t-1}}$, where $\hat{h}_{i,t|t-1}$ is the one-step-ahead forecast of $h_{i,t}$. As Gaussian score losses (Amisano and Giacomini 2007) we set

$$d_t \equiv EWSCORE_t \equiv \log(w'\,\hat{H}_{t|t-1} w) + (w'\hat{y}_{t|t-1})^2 / (w'\,\hat{H}_{t|t-1} w) \tag{20}$$

for the DCC forecasts of the EW conditional variance, and

$$d_t \equiv CORRSCORE_t \equiv \log|\hat{R}_{t|t-1}| + \hat{\varepsilon}'_{t|t-1} \{\hat{R}_{t|t-1}\}^{-1} \hat{\varepsilon}_{t|t-1} \tag{21}$$

for the DCC correlation forecasts, where $\hat{\varepsilon}_{t|t-1} \equiv [\hat{\varepsilon}_{1,t|t-1}, \ldots, \hat{\varepsilon}_{N,t|t-1}]'$ and $\hat{R}_{t|t-1}$ is the correlation matrix associated to $\hat{H}_{t|t-1}$. For the MV and *MV portfolios, the EW weights in (18) and (20) are replaced by the appropriate weights. Analogous losses are computed for the cDCC estimator.

Let $\bar{d}$ and $\bar{d}^c$ denote the DCC and cDCC average losses, respectively. The null hypothesis, $\mathcal{H}_0 : E\left[\bar{d}^c - \bar{d}\ \right] = 0$, denotes equal predictive ability for the two estimators. Under appropriate conditions (Diebold and Mariano 1996), it holds that

$$EPA \equiv \frac{\sqrt{T-\bar{T}}(\bar{d}^c - \bar{d}\ )}{\sqrt{\widehat{VAR}[\sqrt{T-\bar{T}}(\bar{d}^c - \bar{d}\ )]}} \overset{A}{\sim} N(0,1),$$

where $\widehat{VAR}[\sqrt{T-\bar{T}}(\bar{d}^c - \bar{d}\ )]$ is an HAC-robust estimate of the variance of $\sqrt{T-\bar{T}}(\bar{d}^c - \bar{d}\ )$. Negative (resp. positive) values of EPA provide evidence in favor of cDCC (resp. DCC) forecasts. Note that the only model dependent estimation error entering the considered loss functions is that of the correlation estimator. Hence, all resulting EPA tests will essentially compare the correlation performances of the two estimators. The EPA test statistics for the considered losses are reported in Table (1).

Table 1. EPA Tests of One-step-ahead Forecasts

Loss type	Small dataset *CORR*	*EW*	*MV*	**MV*	Large dataset *CORR*	*EW*	*MV*	**MV*
MSE	-1.44	-0.17	-0.65	0.26	**-2.41**	**-2.50**	-1.82	-1.11
SCORE	-1.88	-1.31	0.28	0.12	**-2.68**	-1.43	-0.93	**-2.46**

NOTE: Negative (resp. positive) values are in favor of the *c*DCC (resp. DCC) estimator. Numbers in boldface denote significance at 5% level.

For the small dataset, all tests are insignificant at a 5% level. With the large dataset, the message is in favor of the *c*DCC estimator. The sign of the test statistic is always negative, four tests of eight are significant at a 5% level, and one among them is significant at a 1% level.

4.3.2 Regression-based tests from one-step-ahead forecasts.

The regression based tests of section 4.2, computed replacing y_t and $\hat{H}_t$ with $\hat{y}_{t|t-1}$ and $\hat{H}_{t|t-1}$, are tests of correct specification of the DCC model. Analogous tests can be computed to assess the correct specification of the *c*DCC model. If the model is correctly specified and the estimates are set to the true values of the parameters, the tests are asymptotically normal. In this paper, because of the replacement of true quantities with estimated quantities, the size of the tests can be different from the nominal size even if the model is correctly specified. Table (2) reports the test results. The performances of the DCC and *c*DCC estimators are similar, with performances slightly better for the *c*DCC estimator in the small dataset, where six *c*DCC test statistics of nine are smaller than the corresponding DCC test statistics.

Table 2. Regression-based Tests Computed From One-step-ahead Forecasts

Dataset	Estimator	E & C Test EW	MV	*MV	DQ Test EW	MV	*MV	ARCH Test EW	MV	*MV
Small	*c*DCC	**3.77**	**19.55**	**6.19**	-0.67	-0.47	-0.50	0.59	**3.72**	**3.53**
	DCC	**3.90**	**25.39**	**6.31**	-0.67	-0.48	-0.50	0.68	**4.63**	**3.53**
Large	*c*DCC	**34.63**	**83.23**	**72.11**	-0.25	-0.87	-0.77	**4.28**	**2.84**	1.32
	DCC	**33.83**	**85.66**	**73.94**	-0.23	-0.91	-0.75	**3.92**	**2.24**	1.37

NOTE: Numbers in boldface denote significance at 5% level.

4.3.3 EPA tests of multi-step-ahead correlation forecasts.

The forecast at time t of $\rho_{ij,t+m}$ is defined as $E_t[\rho_{ij,t+m}]$, where $m \geq 2$ (for $m = 1$ we get $E_t[\rho_{ij,t+1}] = \rho_{ij,t+1}$). With both the DCC model and the *c*DCC model, $E_t[\rho_{ij,t+m}]$ is infeasible. For the DCC model, Engle and Sheppard (2001) suggested replacing $E_t[\rho_{ij,t+m}]$ with two possible approximations. One is defined as,

$$\bar{\rho}_{ij,t+m|t} \equiv \frac{q_{ij,t+m|t}}{\sqrt{q_{ii,t+m|t}\, q_{jj,t+m|t}}}\,, \tag{22}$$

where

$$q_{ij,t+m|t} \equiv s_{ij}\,(1-\alpha-\beta)\sum_{n=0}^{m-2}(\alpha+\beta)^n + q_{ij,t+1}(\alpha+\beta)^{m-1}. \tag{23}$$

The second is

$$\bar{\varrho}_{ij,t+m|t} \equiv \bar{\varrho}_{ij}\,(1-\alpha-\beta)\sum_{n=0}^{m-2}(\alpha+\beta)^n + \rho_{ij,t+1}(\alpha+\beta)^{m-1}, \tag{24}$$

where $\bar{\varrho}_{ij} \equiv E[\varepsilon_{i,t}\,\varepsilon_{j,t}]$. We can adopt analogous formulas for the cDCC model. We then set

$$d_t \equiv \frac{1}{N(N-1)/2}\sum_{i<j=2,\ldots,N}\left\{\hat{\varepsilon}_{i,t+m|t+m-1}\,\hat{\varepsilon}_{i,t+m|t+m-1} - \bar{\rho}_{ij,t+m|t}\right\}^2 \tag{25}$$

as a MSE loss for $\bar{\rho}_{ij,t+m|t}$, where $t = \bar{T}+1, \bar{T}+2, \ldots, T-m$ and $m \geq 2$. An analogous loss is adopted for $\bar{\varrho}_{ij,t+m|t}$. The unknown parameters entering $\bar{\rho}_{ij,t+m|t}$ and $\bar{\varrho}_{ij,t+m|t}$ (including $\bar{\varrho}_{ij}$) are replaced by the estimates computed from the rolling window $\hat{y}_{t-1}, \hat{y}_{t-2}, \ldots, \hat{y}_{t-\bar{T}}$. The estimated standardized returns entering (25), which are based on $\hat{y}_{t+m-1}, \hat{y}_{t+m-2}, \ldots, \hat{y}_{t+m-\bar{T}}$ (see (19)), are preferred to the estimated standardized returns based on the whole sample to alleviate the effect of possible structural breaks in the mean and conditional variance of $z_{i,t}$. As forecast horizons we select the Fibonacci numbers greater than 1 and less than 250, where 250 is about the number of daily returns per year. We notice that, for varying the forecast horizon and the approximation, the EPA tests will not be independent.

As a first goal, we are interested in comparing, for a fixed estimator, the predictive ability of the two competing forecasts, denoted with $\bar{\rho}$ and $\bar{\varrho}$. The related EPA test statistics are reported in Table (3).

Table 3. EPA Tests of Multi-step-ahead Correlation Forecasts

		Forecast horizon										
Dataset	EPA Test	2	3	5	8	13	21	34	55	89	144	233
Small	$\bar{\rho}^c$ vs. $\bar{\varrho}^c$	-1.31	-1.39	-1.45	-1.30	-0.88	-0.77	-0.69	**-2.47**	**-2.72**	**-2.61**	**-2.15**
	$\bar{\rho}$ vs. $\bar{\varrho}$	-1.41	-1.61	-1.71	-1.51	-0.51	0.48	**2.41**	0.46	-1.43	0.02	-1.92
	$\bar{\rho}^c$ vs. $\bar{\rho}$	-1.49	-1.51	-1.50	-1.39	-1.33	-1.36	-1.76	**-2.69**	**-2.64**	**-2.60**	**-2.15**
	$\bar{\varrho}^c$ vs. $\bar{\varrho}$	-1.51	-1.54	-1.56	-1.49	-1.52	-1.58	**-2.11**	**-2.72**	**-2.02**	**-2.11**	-1.56
Large	$\bar{\rho}^c$ vs. $\bar{\varrho}^c$	**-2.19**	**-2.84**	**-2.88**	**-2.84**	**-2.45**	**-2.18**	-1.51	-1.17	0.42	1.14	1.44
	$\bar{\rho}$ vs. $\bar{\varrho}$	-1.86	**-2.69**	**-2.93**	**-2.99**	**-2.93**	**-2.91**	**-2.39**	**-2.29**	-0.60	-0.77	1.58
	$\bar{\rho}^c$ vs. $\bar{\rho}$	**-2.31**	**-2.21**	**-2.16**	-1.68	-1.71	-0.90	0.19	-0.07	0.79	0.98	1.63
	$\bar{\varrho}^c$ vs. $\bar{\varrho}$	**-2.25**	**-2.07**	-1.93	-1.31	-1.43	-0.47	0.62	0.58	1.09	0.19	**2.10**

NOTE: Negative (resp. positive) values of the EPA test, "X vs. Y", are in favor of X (resp. Y). ρ and ϱ denote the correlation forecasts Numbers in boldface denote significance at 5% level. A superscript c denotes cDCC forecasts.

Negative (resp. positive) values of the test statistics denote a preference for $\bar{\rho}$ (resp. $\bar{\varrho}$). With the cDCC estimator (see the first row of the table), the preference is for $\bar{\rho}$. The test statistic is always negative, and, for $m \geq 55$, it is significant at a 5% level. With the DCC estimator (see the second row of the table), the message is less conclusive. The only significance is for $m = 34$ and in favor of $\bar{\varrho}$, but most test statistics are negative, which is in favor of $\bar{\rho}$. With the large dataset, with both estimators the preference is for $\bar{\rho}$ (see the fifth and sixth rows of the table).

As a second goal, we are interested in comparing the performances of the DCC and cDCC estimators for a fixed correlation forecast. Regarding the small dataset (see the third and fourth row of the table), the test statistic is always negative, which is in favor of cDCC forecasts. For $m \geq 34$, eight tests of ten are significant at a 5% level. Regarding the large dataset (see the seventh and eighth rows of the table), the only test which is significant and in favor of the DCC estimator is for $n = 233$, when $\bar{\varrho}$ is used. In all the remaining cases, the test is either not significant, or significant and in favor of cDCC forecasts. The test statistics, however, increase with m, denoting an improvement of the DCC forecasts as long as the forecast horizon increases. In summary, for the considered datasets, there is some evidence that $\bar{\rho}$ is recommendable with respect to $\bar{\varrho}$, and that the cDCC correlation forecasts outperform the DCC correlation forecasts.

5. CONCLUSIONS

In this paper we discussed some problems which arise with the DCC model. We pointed out that the test of DCC integrated correlations is inconclusive, and that the DCC estimator of the location correlation parameter can be inconsistent. We then discussed the cDCC model as a possible tractable alternative to the DCC model. The formula of the cDCC conditional correlation has been proven to be more intuitive than the corresponding DCC formula. Sufficient conditions for the stationarity of the cDCC relavant processes have been derived, and the test of the cDCC integrated correlation has been proven to be a conclusive procedure. A large system estimator for the cDCC model, called the cDCC estimator, has been discussed in detail and heuristically proven to be consistent.

The performances of the DCC and cDCC estimators have been compared by means of applications to simulated and real data. When the persistence of the correlation process and the impact of the news are both high, the DCC estimator of the location correlation parameter has been proven to be seriously biased. The corresponding cDCC estimator has been shown to be uniformly unbiased. On two sets of real data, the cDCC multi-step-ahead correlation forecasts have been proven to perform equally or significantly better than the corresponding DCC forecasts.

ACKNOWLEDGMENTS

I would like to thank Farid Boussama, Christian T. Brownlees, Giorgio Calzolari, Gabriele Fiorentini, Giampiero M. Gallo, Matteo Guastini, Davide Perilli, Pietro Rigo, Paolo Santucci De Magistris, Nicola Sartori, Enrique Sentana and two anonymous referees for their careful and useful suggestions. I would also like to thank my brother Gian Luca for his precious support.

APPENDIX: PROOFS

Proof of prop. 2.1. Taking the expectations of both members of (4) and rearranging, yields $S = \{(1 - \beta)E[Q_t] - \alpha E[\varepsilon_t \varepsilon_t']\}/(1 - \alpha - \beta)$. Noting that $E[Q_t] = E[Q_t^{*1/2} R_t Q_t^{*1/2}] = E[Q_t^{*1/2} E_{t-1}[\varepsilon_t \varepsilon_t'] Q_t^{*1/2}] = E[E_{t-1}[Q_t^{*1/2} \varepsilon_t \varepsilon_t' Q_t^{*1/2}]] = E[Q_t^{*1/2} \varepsilon_t \varepsilon_t' Q_t^{*1/2}]$, proves the proposition. ■

Proof of prop. 2.2. Dividing the numerator and denominator of the right hand side of $\rho_{ij,t} = q_{ij,t} / \sqrt{q_{ii,t} q_{jj,t}}$

by $\sqrt{q_{ii,t-1}q_{jj,t-1}}$, and replacing β with $1-\alpha$, yields

$$\rho^2_{ij,t+m+2} = \frac{\{\alpha\,(\varepsilon_{i,t+m+1}\varepsilon_{j,t+m+1}/\sqrt{q_{ii,t+m+1}q_{jj,t+m+1}}) + (1-\alpha)\,\rho_{ij,t+m+1}\}^2}{\{\alpha\,(\varepsilon^2_{i,t+m+1}/q_{ii,t+m+1}) + (1-\alpha)\} \times \{\alpha\,(\varepsilon^2_{j,t+m+1}/q_{jj,t+m+1}) + (1-\alpha)\}},$$

where $m \geq 0$. For $\alpha = 0$ we get $E_{t+m}[\rho^2_{ij,t+m+2}] = \rho^2_{ij,t+m+1}$, and, for $\alpha = 1$, we get $E_{t+m}[\rho^2_{ij,t+m+2}] = E_{t+m}[(\varepsilon^2_{i,t+m+1}\varepsilon^2_{j,t+m+1})/\ (\varepsilon^2_{i,t+m+1}\varepsilon^2_{j,t+m+1})] = 1$. By continuity and monotonicity of $E_{t+m}[\rho^2_{ij,t+m+2}]$ with respect to α, it follows that $\rho^2_{ij,t+m+1} < E_{t+m}[\rho^2_{ij,t+m+2}]$ for all $\alpha \in (0,1)$. Taking the expectations at time t of both members of the latter inequality, yields $E_t[\rho^2_{ij,t+m+1}] < E_t[E_{t+m}[\rho^2_{ij,t+m+2}]] = E_t[\rho^2_{ij,t+m+2}]$, which, for $m = 0, 1, \ldots, r-2$ and $\lim_{m\to\infty}$, proves the proposition.

Proof of prop. 2.3. From $\varepsilon^*_t = Q^{*1/2}_t\varepsilon_t$, where $\varepsilon_t = R^{1/2}_t\eta_t$, it follows that $\varepsilon^*_t = Q^{1/2}_t\eta_t$, where $Q^{1/2}_t \equiv Q^{*1/2}_t R^{1/2}_t$ is the unique psd matrix such that $Q^{1/2}_t Q^{1/2}_t = Q_t$. This fact, together with H1-H2, ensures that the BEKK process, $[\text{vech}(Q_t)', \varepsilon^{*\prime}_t]'$, admits a non-anticipative, strictly stationary, and ergodic solution (Boussama et al., 2010, Theorem 2.3). Therefore, any time-invariant function of $[\text{vech}(Q_t)', \varepsilon^{*\prime}_t]'$, such as $[\text{vech}(R_t)', \varepsilon'_t]'$, admits a non-anticipative, strictly stationary, and ergodic solution (Billingsley 1995, Theorem 36.4). Since the elements of $[\text{vech}(R_t)', \varepsilon'_t]'$ have finite variance, by Cauchy-Schwartz inequality the second moment of $[\text{vech}(R_t)', \varepsilon'_t]'$ exists finite, which completes the proof of point (i). To prove point (ii), it suffices to note that $[\text{vech}(H_t)', y'_t, \text{vech}(R_t)', \varepsilon'_t]'$ is a time-invariant function of $[h_{1,t}, \ldots, h_{N,t}, \text{vech}(R_t)', \varepsilon'_t]'$, which, under H1-H3, is a measurable function of the non-anticipative, strictly stationary, and ergodic process $[\text{vech}(R_t)', \varepsilon'_t]'$. Under H1-H4, point (iii) follows by strict stationarity of y_t and Cauchy-Schwartz inequality. ■

Proof of prop. 2.4. Under H1-H2 of prop. 2.3, ε^*_t is weakly stationarity (Boussama et al., 2010, Theorem 2.3). If ε^*_t is weakly stationary, the second moment of ε^*_t is the matrix S in eq. (13) (Engle and Kroner 1995). Since $\varepsilon^*_t = Q^{*1/2}_t\varepsilon_t$, the proposition is proven. ■

Proof of prop. 2.5. As for point (i), under H1-H3, $\text{vech}(R_t)$ is strictly stationary and ergodic (see prop. 2.3). Therefore, under H1-H3, $\rho^2_{ij,t}$, which is a time-invariant function of $\text{vech}(R_t)$, is strictly stationary and ergodic (Billingsley 1995, Theorem 36.4). Since $\rho^2_{ij,t} \leq 1$, the expectation of $\rho^2_{ij,t}$ exists finite; hence, if $\rho^2_{ij,t}$ is ergodic, $\lim_{m\longrightarrow\infty} E_t\,[\rho^2_{ij,t+m}] = E[\rho^2_{ij,t}]$. If $E[\rho^2_{ij,t}] = 1$, a.s. $\rho_{ij,t} = \pm 1$, or, equivalently, a.s. $q_{ij,t} = \pm\sqrt{q_{ii,t}q_{jj,t}}$. Under H1-H3 this is not possible because, under H1-H3, Q_t is p.d. This completes the proof of point (i). As for point (ii), divide the numerator and denominator of the right hand side of $\rho_{ij,t} = q_{ij,t}/\sqrt{q_{ii,t}q_{jj,t}}$ by $\sqrt{q_{ii,t-1}q_{jj,t-1}}$. Replacing β with $1-\alpha$, yields

$$\rho^2_{ij,t+m+2} = \frac{\{\alpha\,\varepsilon_{i,t+m+1}\varepsilon_{j,t+m+1} + (1-\alpha)\,\rho_{ij,t+m+1}\}^2}{\{\alpha\,\varepsilon^2_{i,t+m+1} + (1-\alpha)\} \times \{\alpha\,\varepsilon^2_{j,t+m+1} + (1-\alpha)\}}\,,$$

where $m \geq 0$. The proof is then analogous to the proof of prop. 2.2. ■

Proof of prop. 3.1. Under H1-H2 of prop. 2.3, ε^*_t is weakly stationary and ergodic with stationary second moment S^0 (see prop. 2.4). Hence, under H1-H2 of prop. 2.3, the matrix process $\varepsilon^*_t\varepsilon^{*\prime}_t$, which is a time-invariant functions of ε^*_t, is ergodic with finite first moment S^0. This yields $\text{plim}\,T^{-1}\sum_{t=1}^T \varepsilon^*_t\varepsilon^{*\prime}_t = S^0$, which proves the proposition in that, for $(\theta,\phi) = (\theta^0,\phi^0)$, we have that $\tilde{Q}^{*1/2}_t\,\tilde{\varepsilon}_t = \varepsilon^*_t$. ■

REFERENCES

[1] Aielli, G. P. (2006), "Consistent Estimation of Large Scale Dynamic Conditional Correlations," unpublished manuscript, University of Florence.

[2] Alexander, C. (2002), "Principal Component Models for Generating Large GARCH Covariance Matrices," *Economic Notes*, 31, pp. 337–359.

[3] Amisano, G., and Giacomini, R. (2007), "Comparing Density Forecasts via Weighted Likelihood Ratio Tests," *Journal of Business and Economic Statistics*, 25, pp. 177–190.

[4] Billio, M., Caporin, M., and Gobbo, M. (2006), "modular Dynamic Conditional Correlation Multivariate GARCH for Asset Allocation," *Applied Financial Economics Letters*, 2, pp. 123–130.

[5] Billingsley, P. (1995), *Probability and Measure*, New York, John Wiley & Sons Inc.

[6] Bollerlsev, T., (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, pp. 307–327.

[7] Bollerlsev, T. (1990), "Modeling the Coherence in Short Run Nominal Exchange Rates: A Multivariate generalized ARCH Model," *The Review of Economics and Statistics*, 72, pp. 498 – 505.

[8] Bollerslev, T., and Wooldridge, J.M. (1992), "Quasi Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances," *Econometric Reviews*, 11, pp. 143–172.

[9] Bougerol, P., and Picard, N. (1992), "Strict Stationarity of Generalized Autoregressive Processes," *Annals of Probability*, 20, pp. 1714–1729.

[10] Boussama, F., Fuchs, F., and Stelzer, R. (in press), "Stationarity and Geometric Ergodicity of BEKK Multivariate GARCH Models," *Stochastic Processes and their Applications*.

[11] Brandt, A. (1986), "The Stochastic Equation $Y_{n+1} = A_n Y_n + B_n$ With Stationary Coefficients," *Advance in Applied Probability*, 18, pp. 221–254.

[12] Cappiello, L., Engle R.F., and Sheppard, K. (2006), "Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns," *Journal of Financial Econometrics*, 4, pp. 537–572.

[13] Diebold, F. X., and Mariano, R. S. (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, pp. 253–263.

[14] Ding, Z., and Engle, R.F. (2001), "Large Scale Conditional Covariance Matrix Modeling, Estimation and Testing," *Academia Economic Papers*, 29, pp. 157–184.

[15] Engle, R.F. (2002), "Dynamic Conditional Correlation — a Simple Class of Multivariate GARCH Models," *Journal of Business and Economic Statistics*, 20, pp. 339–350.

[16] Engle, R.F., and Colacito, R. (2006), "Testing and Valuing Dynamic Correlations for Asset Allocation," *Journal of Business and Economic Statistics*, 24, pp. 238–253.

[17] Engle, R.F., and Kroner, K. (1995), "Multivariate Simultaneous Generalized ARCH," *Econometric Theory*, 11, pp. 122–150.

[18] Engle, R.F., and Manganelli, S. (2004), "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles," *Journal of Business & Economic Statistics*, 22, pp. 367–381.

[19] Engle, R.F., Shephard, N., and Sheppard, K. (2009), "Fitting Vast Dimensional Time-Varying Covariance Models," Working Paper FIN-08-009, NYU Stern School of Business, Department of Finance.

[20] Engle, R.F., and Sheppard, K. (2001), "Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH," Working Paper FIN-01-027, NYU Stern School of Business, Department of Finance.

[21] Francq C., and Zakoian, M. (2010), *GARCH Models: Structure, Statistical Inference and Financial Applications*, Chichester, UK, John Wiley & Sons Ltd.

[22] Franses, P.H., and Hafner, C.M. (2009), "A Generalized Dynamic Conditional Correlation Model for Many Asset Returns," *Econometric Reviews*, 28, pp. 612–631.

[23] Kwan, C. K., Li, W. K., and Ng, K. (in press), "A Multivariate Threshold GARCH Model With Time-Varying Correlations," *Econometric Reviews*.

[24] McAleer, M., Chan, F., Hoti, S., and Lieberman, O. (2008), "Generalized Autoregressive Conditional Correlation," *Econometric Theory*, 24, pp. 1554–1583.

[25] Newey, W., and McFadden D. (1994), "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics* (Vol. 4) , eds. R. Engle and D. McFadden, New York, Elsevier Science, pp. 2113–2245.

[26] Newey, W.K., and West, K.D. (1987), "A Simple Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, pp. 703–708.

[27] Pelletier, D. (2006), "Regime Switching for Dynamic Correlations," *Journal of Econometrics*, 131, pp. 445–473.

[28] Pesaran B., and Pesaran, M.H. (2007), "Modeling Volatilities and Conditional Correlations in Futures Markets with a Multivariate t Distribution," IZA Discussion Papers No. 2906, Institute for the Study of Labor.

[29] Severini, T. A. (1998), "Likelihood Functions for Inference in the Presence of a Nuisance Parameter," *Biometrika*, 85, pp. 507–522.

[30] Silvennoinen, A., and Teräsvirta, T. (2009), "Modeling Multivariate Autoregressive Conditional Heteroskedasticity with the Double Smooth Transition Conditional Correlation GARCH Model," *Journal of Financial Econometrics*, 7, pp. 373–411.

[31] Tse, Y. K., and Tsui, K. C. (2002), "A Multivariate Generalized Autoregressive Conditional Heteroskedasticity Model With Time-Varying Correlation," *Journal of Business and Economic Statistics*, 20, pp. 351–362.

[32] White, H. (1996), *Estimation, Inference, and Specification Analysis*, Cambridge, Cambridge University Press.

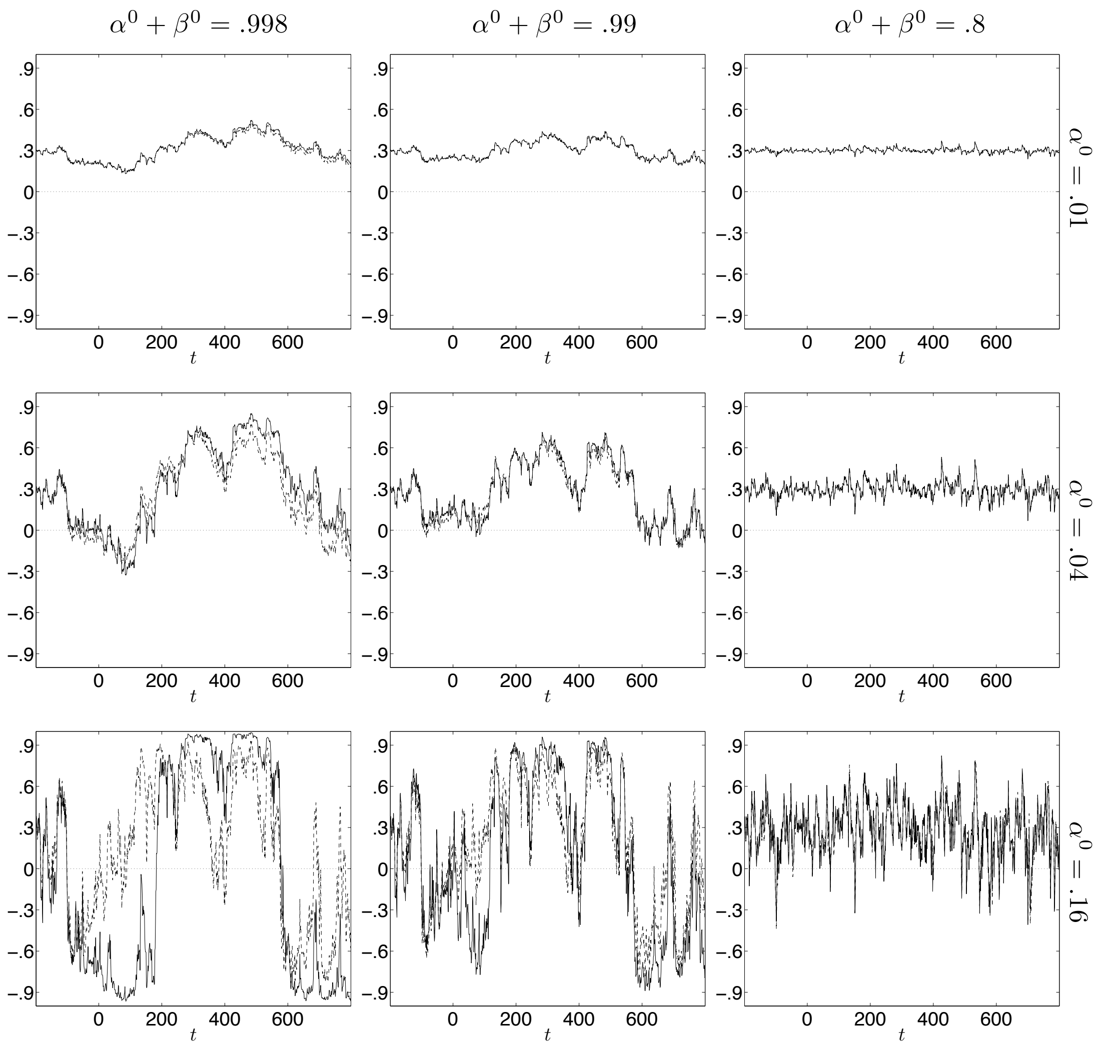

Figure 1. DCC and cDCC Conditional Correlations. Simulated series of $\rho_{ij,t}$. The DGP parameter values are reported at the top of the panel for $\alpha^0 + \beta^0$, and on the right hand side of the panel for α^0. The location parameter is set as $s_{ij}^0 = .3$. DCC in straight line and cDCC in dashed line.

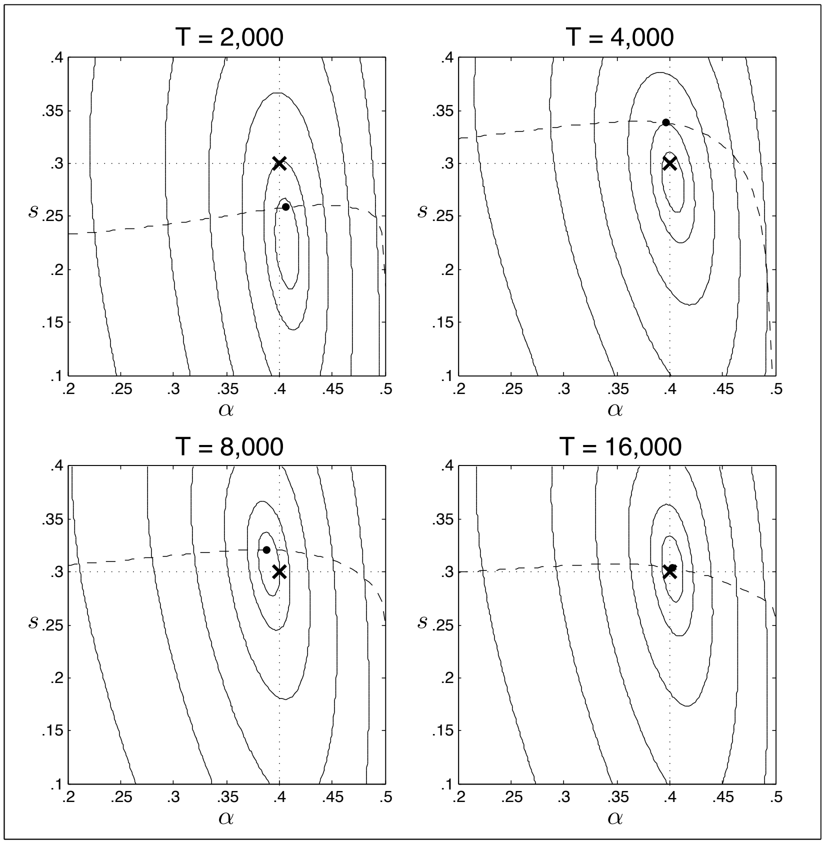

Figure 2. Performance of the cDCC Estimator for Increasing T. The figure provides an example of the behavior of the cDCC estimator for increasing the sample size. For illustrative purposes, it is assumed that θ^0 and β^0 are known. The DGP is bivariate Gaussian with parameters set as $(s^0, \alpha^0) = (.3, .4)$, where $s^0 \equiv s^0_{12}$. The contour plots refer to the scaled cDCC QLL, that is, $T^{-1}L_T(s, \alpha)$, where $s \equiv s_{12}$. The dashed line refers to the constraint $\{s = \tilde{s}_\alpha, \alpha \in [0, 1)\}$, under which the cDCC QLL is maximized by the cDCC estimator (see section 3.2.1). With the considered model the constraint is a curve of the plane. For varying α, the scaled cDCC PQLL, that is, $T^{-1}L_T(\tilde{s}_\alpha, \alpha)$, describes the value assumed by the scaled cDCC QLL along the curve of the plane. The cDCC estimator, $(\hat{\alpha}, \hat{s})$, is denoted with a bullet, and the true value with a cross and dotted lines. The more observations, the more the scaled cDCC QLL centers on the true value of the parameters, the more the cDCC constraint approaches a correctly specified constraint.

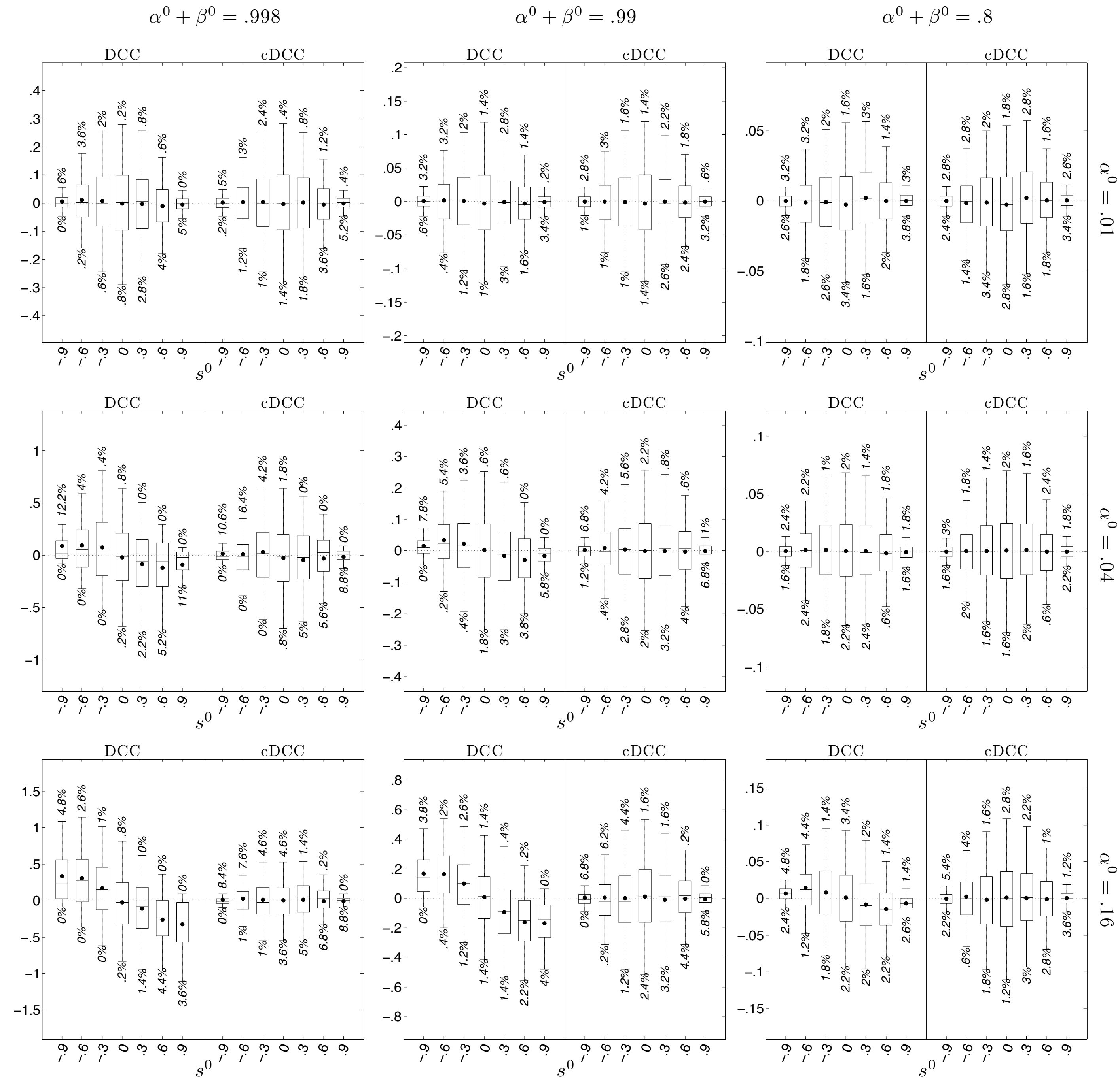

Figure 3. Estimation Error of $\hat{s}$. Box plots of $\hat{s} - s^0$. The DGP parameter values are reported at the top of the panel for $\alpha^0 + \beta^0$, on the right hand side of the panel for α^0, and along the x-axis of each plot for s^0. For each box plot, the box, the line inside the box, and the bullet, denote, respectively, the interquartile range, the median, and the average of the estimates. The end of the upper whisker is computed as the greatest observation less than or equal to the sum of the third quartile and the interquartile range. The percentage at the end of the upper whisker refers to the estimates greater than the end of the whisker. Symmetric definitions hold for the lower whisker.

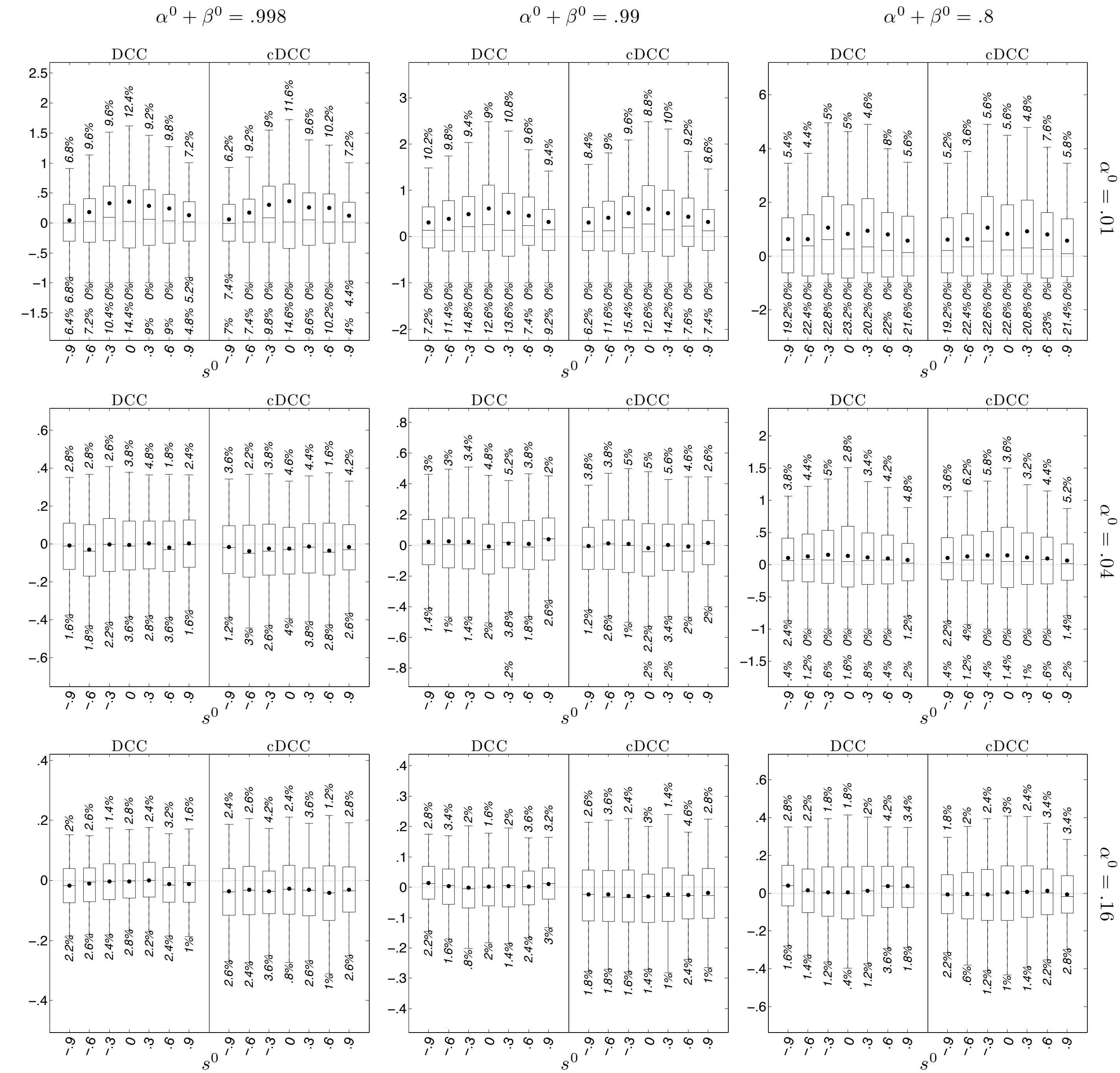

Figure 4. Estimation Error of $\hat{\alpha}$. Box plots of $(\hat{\alpha}-\alpha^0)/\alpha^0$ (for the general layout of the panel, the construction of the box plots, and the percentages at the end of the whiskers of the box plots, see the caption of Fig. 3). If there are estimates on the upper boundary, the related percentage is reported at the top of the plot. If many estimates fall on the upper boundary, the end of the upper whisker can coincide with the boundary, which results in a percentage of observations greater than the end of the whisker equal to zero. Symmetric definitions hold for the percentages below the lower whisker.

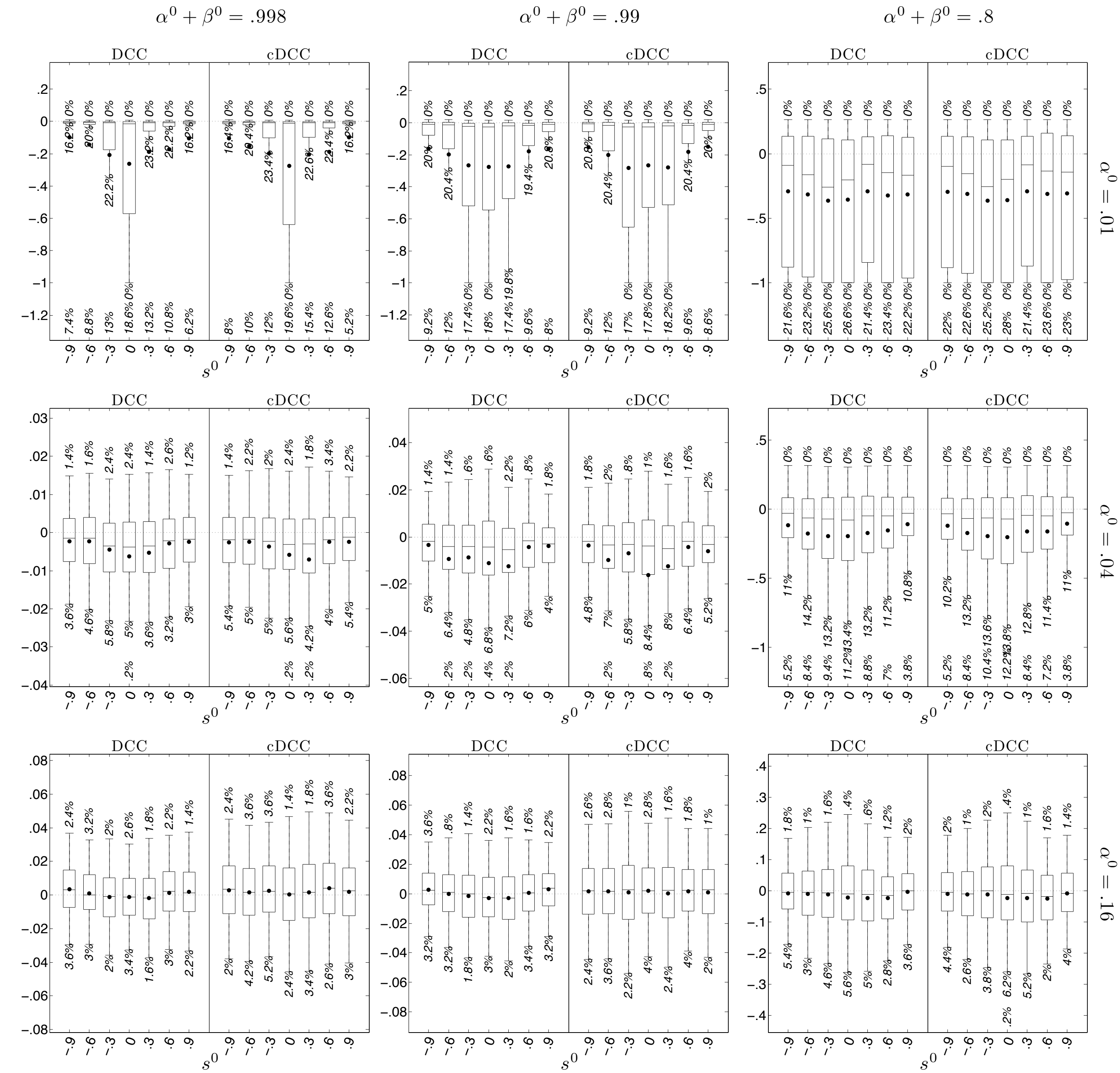

Figure 5. Estimation Error of $\hat{\beta}$. Box plots of $(\hat{\beta} - \beta^0)/\beta^0$ (for the general layout of the panel, the construction of the box plots, and the percentages appearing in the fugure, see the caption of Fig. 4).

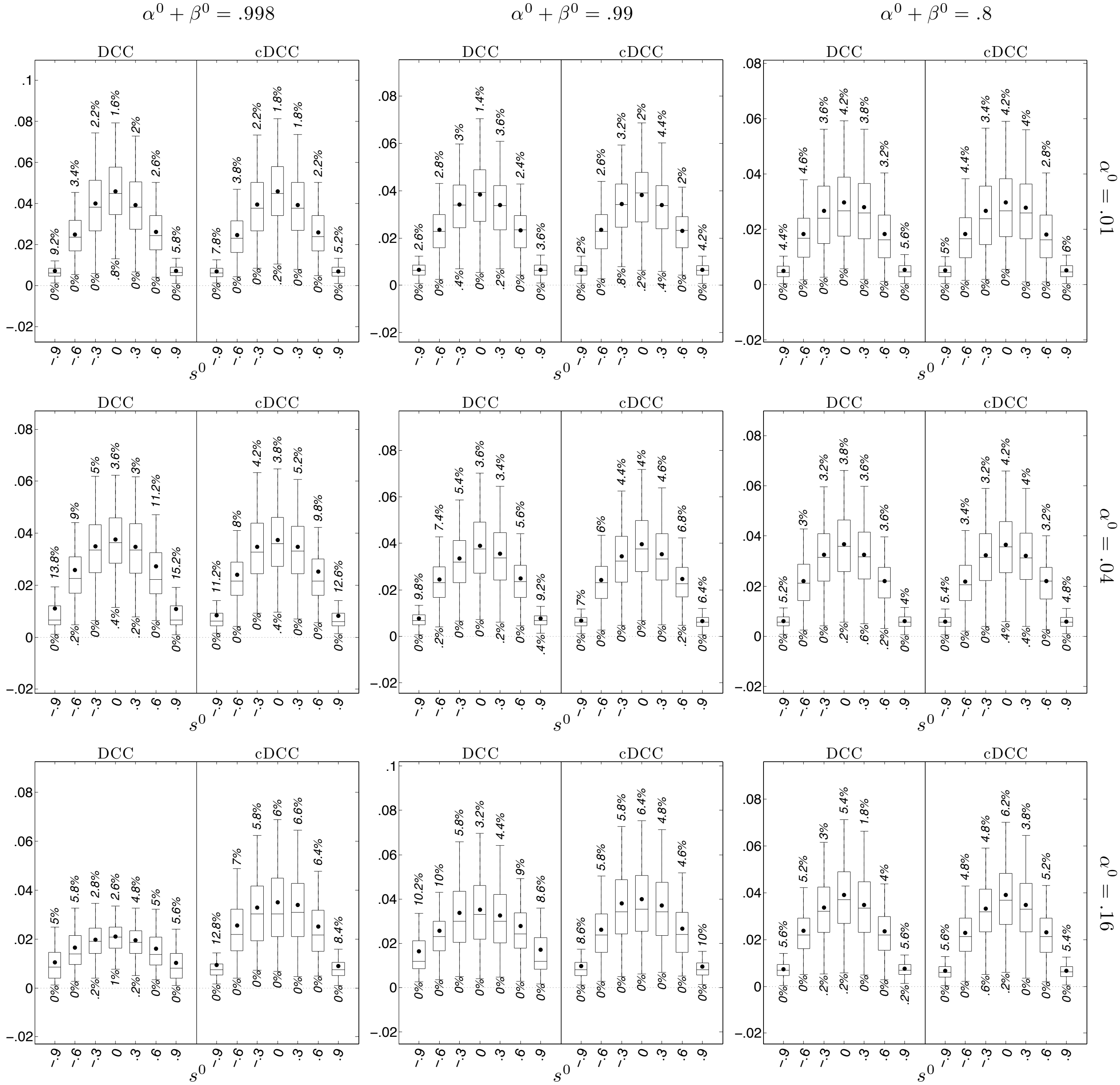

Figure 6. Mean Absolute Error of $\hat{\rho}_t$. (for the general layout of the panel, the construction of the box plots, and the percentages appearing in the figure, see the caption of Fig. 4).

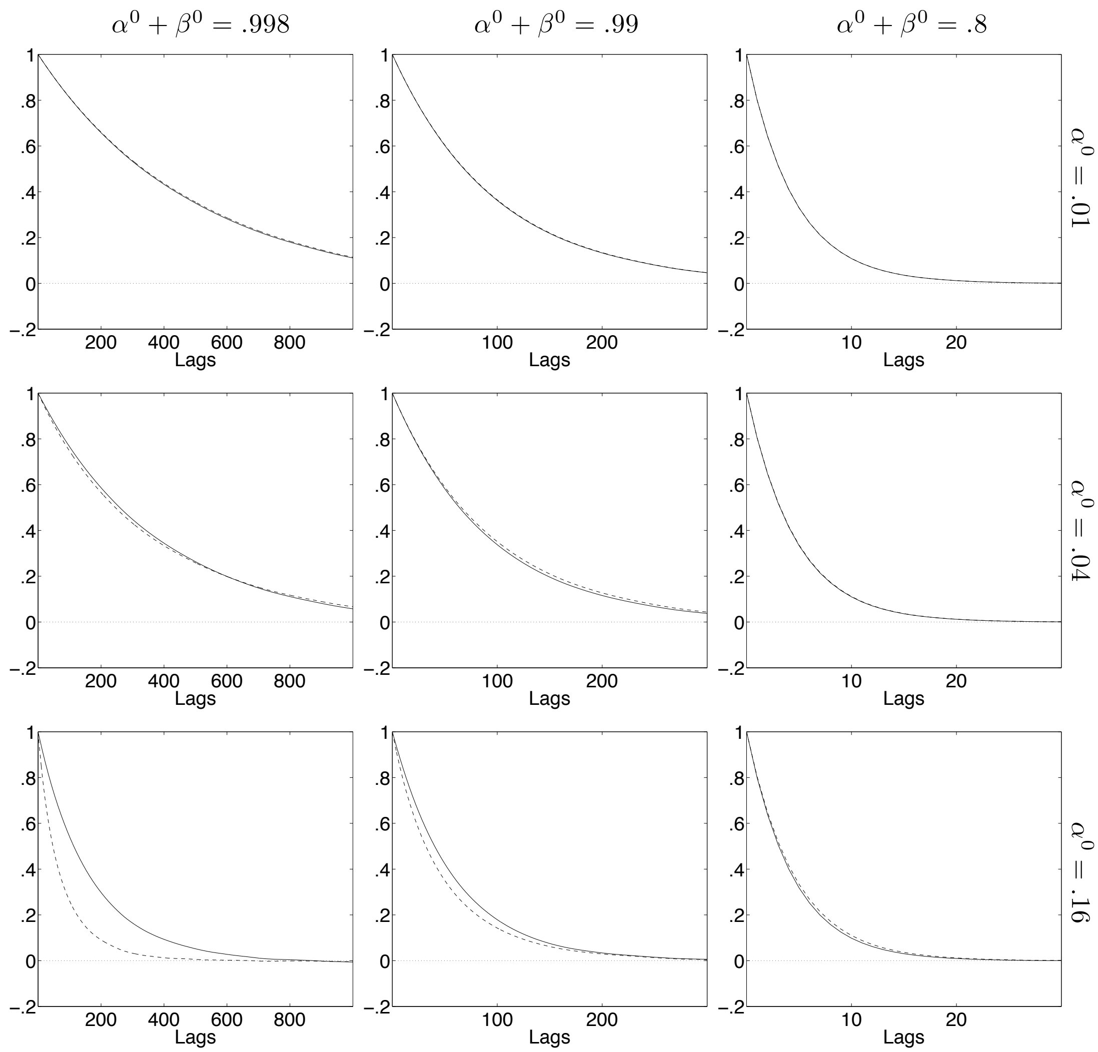

Figure 7. Autocorrelation Function of $\rho_{ij,t}$. ACF computed as the mean of the sample ACF's of 100 series of length $T = 100,000$. The DGP parameter values are reported at the top of the panel for $\alpha^0 + \beta^0$, and on the right hand side of the panel for α^0. The location parameter is set as $s^0_{ij} = .3$. DCC in straight line and cDCC in dashed line.