Fitting Vast Dimensional Time-Varying Covariance Models

Abstract

Estimation of time-varying covariances is a key input in risk management and asset allocation. ARCH-type multivariate models are used widely for this purpose. Estimation of such models is computationally costly and parameter estimates are meaningfully biased when applied to a moderately large number of assets. Here we propose a novel estimation approach that suffers from neither of these issues, even when the number of assets is in the hundreds. The theory of this new method is developed in some detail. The performance of the proposed method is investigated using extensive simulation studies and empirical examples.

Keywords: Composite likelihood, dynamic conditional correlations, multivariate ARCH models, volatility.

1 Introduction

The estimation of conditional covariances between the returns on hundreds of assets is a key input in modern risk management and asset allocation. A popular approach is to use ARCH-type models (e.g., Bollerslev (1990), Engle and Kroner (1995) and Engle (2002)). This literature has been quite successful in dealing with cases where the number of assets is moderate, i.e., less than 25. However, estimation becomes computationally burdensome for larger datasets. Moreover, the standard estimator is substantially biased when the number of assets is not small, even with thousands of time-series observations.

The main cause of these issues is that most widely-used estimation methods rely on a simple-to-implement estimator of the unconditional covariance/correlation matrix. This involves estimation of $O(L^2)$ parameters using $O(LT)$ data points, where T is the time-series length and L is the number of assets. Unless T is substantially larger than L, this will be subject to significant estimation error. Consequently, the likelihood function will be contaminated, leading to biased estimates. In addition, estimation is also subject to a computational issue: Gaussian (quasi-) likelihood estimation requires inversion of the $(L \times L)$ conditional variance matrix. Since estimation relies on numerical optimisation, many matrix inversions are needed until estimator convergence. As such, the computational cost of fitting the model increases rapidly in L.

We propose a novel estimation approach that is computationally fast and does not suffer from the aforementioned statistical issues, even when the number of assets is in the hundreds. In particular, we utilise the composite likelihood approach, which uses the average of bivariate log-likelihood functions for a large selection of asset-pairs as the objective function. This side-steps both of the discussed problems as each pairwise likelihood involves inversion of a (2×2) conditional variance matrix only. Intuitively, each bivariate likelihood contains some information on the full-dimensional problem; taking their average allows us to pool information without having to invert large dimensional matrices. We investigate the theory and application of our method for the BEKK and DCC models of Engle and Kroner (1995) and Engle (2002). Our simulations confirm that the proposed method does not suffer from bias as L increases; indeed, bias remains small even when L is of the same order as T. We demonstrate the usefulness of our method in a series of empirical examples.

The structure of the paper is as follows. Section 2 provides a brief overview of multivariate volatility modelling. Section 3 introduces the composite likelihood approach. In Section 4, we focus on the specific cases of BEKK and DCC. Simulation and empirical studies are conducted in Sections 5 and 6, respectively. Section 7 concludes. Proofs and additional material are in the Supplementary Appendix.

2 Multivariate volatility modelling

2.1 Background

Throughout, we use the following notational convention: $0_{a\times b}$, 0_a and I_a denote an $(a \times b)$ matrix of zeros, an $(a \times 1)$ vector of zeros and an $(a \times a)$ identity matrix, respectively. For any matrix A, $||A|| = \sqrt{tr(A'A)}$ is the Euclidian norm. For any square matrix B, $|B|$ is the determinant. Furthermore, *vec* is the vectorisation operator.

Let r_{it}, be the (log) return on asset i at time t, and let $X_t = (r_{1t}, ..., r_{Lt})'$, where $i = 1, ..., L$ and $t = 1, ..., T$. Let $\mathcal{F}_t$ be the information set at time t, based on the history of all returns. Also, let $\mathcal{F}_{it}$ be the information set at t for asset i. We make the standard simplifying assumption that $E[X_t|\mathcal{F}_{t-1}] = 0_L$, which is reasonable for daily returns. Let Z_t be some $(L \times 1)$ vector of independently and identically distributed (iid) random variables, with $E[Z_t] = 0_L$ and $Cov(Z_t) = I_L$. Then, the standard structure for X_t is given by $X_t = H_t^{1/2} Z_t$ where $H_t = Cov(X_t|\mathcal{F}_{t-1})$ is the conditional variance matrix.

The literature on models for H_t is vast; for surveys, see Bauwens, Laurent, and Rombouts (2006), Silvennoinen and Teräsvirta (2009) and Engle (2009). The typical approach in this literature is to model H_t parametrically where H_t is indexed by some parameter vector θ, the estimation of which relies on the (quasi-) maximum likelihood (ML) method. Two of the most popular models are the BEKK and DCC/cDCC models.

The BEKK model (after Baba, Engle, Kraft, and Kroner) is due to Engle and Kroner (1995). While there are different variants, a common version is

$$H_t = C + AX_{t-1}X_{t-1}'A' + BH_{t-1}B', \tag{1}$$

where A and B are $(L \times L)$ parameter matrices and C is some positive definite unconditional covariance matrix. It is possible to add further lags of $X_t X_t'$ and H_t but this greatly increases the computational burden, which is already high for moderate values of L. A computationally less demanding variant was proposed by Engle and Mezrich (1996) and studied theoretically in Pedersen and Rahbek (2014) (referred to as PR henceforth). This is based on the commonly-used "variance-tracking" approach (Francq, Horváth, and Zakoïan (2011)). Specifically, let $\Gamma = E\left[X_t X_t'\right]$. Then, it can be shown that $\Gamma = C + A\Gamma A' + B\Gamma B'$, and the variance-tracking version of (1) is given by

$$H_t = \Gamma - A\Gamma A' - B\Gamma B' + AX_{t-1}X_{t-1}'A' + BH_{t-1}B'. \tag{2}$$

This simplifies estimation by allowing sequential estimation of Γ and (A, B). First Γ is estimated by a computationally-simple moment estimator, $\hat{\Gamma} = T^{-1}\sum_{t=1}^{T} X_t X_t'$, and then (A, B) is estimated via ML, using (2) with Γ replaced by $\hat{\Gamma}$. We will mainly focus on the more parsimonious scalar version of (2):

$$H_t = \Gamma\left(1 - \alpha - \beta\right) + \alpha X_{t-1}X_{t-1}' + \beta H_{t-1} \tag{3}$$

where α and β are scalar and Γ is defined as before.

The Dynamic Conditional Correlation (DCC) model of Engle (2002) provides additional flexibility for specifying the dynamics of the time-varying conditional covariance. Here, we consider a recent version, the cDCC model, proposed by Aielli (2013). Specifically,

$$H_t = D_t^{1/2} R_t D_t^{1/2}, \tag{4}$$

where D_t is an $(L \times L)$ diagonal matrix, with the diagonal entries given by $h_{it} = Var\left(r_{it}|\mathcal{F}_{i,t-1}\right)$ which can be based on any univariate volatility model (the typical choice is the GARCH model of Bollerslev (1986)). R_t, on the other hand, is the time-varying conditional correlation matrix. Notice that for the devolatilised returns $\varepsilon_t = D_t^{-1/2} X_t$, we also have $Cov\left(\varepsilon_t|\mathcal{F}_{t-1}\right) = R_t$. Then, R_t is modelled as follows:

$$R_t = Q_t^{*-1/2} Q_t Q_t^{*-1/2}, \tag{5}$$

$$Q_t = (1-\alpha-\beta) S + \alpha Q_{t-1}^{*1/2} \varepsilon_{t-1} \varepsilon_{t-1}' Q_{t-1}^{*1/2} + \beta Q_{t-1}, \tag{6}$$

where Q_t^* is a diagonal matrix containing the diagonal entries of Q_t. The parameters of this model are (S, α, β). This structure has the virtue that for $\varepsilon_t^* = Q_t^{*1/2} \varepsilon_t$, we have $Cov\left(\varepsilon_t^* | \mathcal{F}_{t-1}\right) = Q_t^{*1/2} R_t Q_t^{*1/2} = Q_t$, so that $T^{-1} \sum_{t=1}^{T} \varepsilon_t^* \varepsilon_t^{*\prime} \xrightarrow{p} S$ (see Aielli (2013) for further details on the DCC and cDCC models). This allows for the joint estimation of S and (α, β) using a concentrated likelihood-like moment-based estimator. The computational cost of this estimator is similar to the variance-tracking approach.

2.2 Issues in large-dimensional modelling

We now discuss some important issues in the estimation of large-scale multivariate volatility models. The problems we point out here generally hold for moderately large (or larger) numbers of assets. To illustrate, suppose $Z_t \overset{iid}{\sim} N(0_L, I_L)$. Then, the (negative) log-likelihood function for the scalar BEKK model in (3) is given by

$$l_T(\gamma, \lambda) = \frac{1}{T} \sum_{t=1}^{T} \left[\log |H_t(\gamma, \lambda)| + X_t' H_t^{-1}(\gamma, \lambda) X_t\right], \tag{7}$$

where the dependence of H_t on $\gamma = vec(\Gamma)$ and $\lambda = (\alpha, \beta)'$ is made explicit. Using the variance-tracking approach, plugging $\hat{\Gamma} = T^{-1} \sum_{t=1}^{T} X_t X_t'$ into (7) yields

$$l_T(\hat{\gamma}, \lambda) = \frac{1}{T} \sum_{t=1}^{T} \left(\log |H_t(\hat{\gamma}, \lambda)| + X_t' H_t^{-1}(\hat{\gamma}, \lambda) X_t\right). \tag{8}$$

Then, the estimator of λ is given by $\tilde{\lambda} = (\tilde{\alpha}, \tilde{\beta}) = \arg\min_\lambda l_T(\hat{\gamma}, \lambda)$. Obtaining $\tilde{\lambda}$ requires numerical optimisation since the solution to the optimisation problem (independent of whether Γ is estimated in an initial step or not) does not exist in closed form. Evaluating the likelihood requires inverting $H_t(\hat{\gamma}, \lambda)$ for each t in every iteration until the numerical optimiser converges. The computational cost of this is $O(L^3)$, and so full likelihood evaluation is slow even for moderate L.

While the computational issue is an important one, the salient problems are statistical. In particular, $\hat{\Gamma}$ involves estimation of $O(L^2)$ parameters with $O(LT)$ data points. There-

	S&P 100 Components				S&P 500 Components				
	Scalar BEKK		cDCC			Scalar BEKK		cDCC	
L	α	β	α	β	L	α	β	α	β
5	.0189	.9794	.0141	.9757	5	.0261	.9715	.0101	.9823
10	.0125	.9865	.0063	.9895	25	.0080	.9909	.0030	.9908
25	.0081	.9909	.0036	.9887	50	.0055	.9932	.0018	.9882
50	.0056	.9926	.0022	.9867	100	.0034	.9934	.0015	.9524
96	.0041	.9932	.0017	.9711	250	.0015	.9842	.0020	.5561
					480	.0032	.5630	.0013	.2556

Table 1: Parameter estimates from a covariance tracking scalar BEKK and cDCC using the two-step maximum likelihood method outlined in Section 2. The left panel is based on the daily returns for 95 companies plus the index from the S&P100, from 1997 until 2006. The right panel is based on the same for 480 companies from the S&P 500.

fore, even for moderately large L, $\hat{\Gamma}$ is a noisy estimator of Γ. This is also reflected in the eigenvalues of $\hat{\Gamma}$, in the sense that the largest eigenvalues tend to be overestimated and, more importantly, the smallest ones tend to be underestimated. In a simulation study in Section 5.5 we further investigate this and show that the same occurs for the eigenvalues of $H_t(\hat{\gamma}, \lambda)$, as well. Moreover, the discrepancy between the eigenvalues increases rapidly with L. While this is not formally an issue for fixed L, the sample size required to achieve approximately unbiased parameter estimates is larger than what is available for most empirically realistic portfolio sizes. All these features combine to produce substantially biased $\hat{\lambda}$. Notice that the same story holds for the cDCC model in (4)-(6), where $\lambda = (\alpha, \beta)'$, while γ consists of $vec\,(S)$ and the parameters of the univariate volatility models $(h_{1t}, ..., h_{Lt})$. The relevant likelihood functions are, again, (7) and (8). Similar issues would arise in many multivariate covariance models.

We finish this part with an empirical illustration of the outlined issues for the models in (3) and (4)-(6). Table 1 presents estimates of (α, β) for an expanding cross-section of assets from S&P 100 (left panel) and S&P 500 (right panel); see Section 6 for details of the dataset. Estimates of α fall dramatically with L. The same is not observed for β in the first instance. However, when the analysis is extended to S&P 500, which allows for larger L, we observe that estimates of both α and β are severely biased towards zero for the largest values of L. Clearly, L has a systematic effect on the estimation of λ, independent of the model under consideration.

3 The composite likelihood method

In this part, we propose a method for estimating multivariate volatility models, which does not suffer from the computational and statistical issues outlined in Section 2.2. To achieve this aim, we use the composite likelihood (CL) method (Lindsay (1988), Cox and Reid (2004), Varin and Vidoni (2005) and Varin, Reid, and Firth (2011)). This method is based on approximating the full-dimensional joint likelihood function using combinations of lower dimensional marginal densities. To give an example, let $\dot{l}_{iT}(\cdot)$ be the marginal log-likelihood for the i^{th} asset. Then, the most basic CL function is

$$\frac{1}{L}\sum_{i=1}^{L}\dot{l}_{iT}(\theta),$$

which approximates the joint log-likelihood function using marginal univariate likelihoods (this type of marginal analysis has appeared before outside the time-series literature, e.g. Besag (1974) in spatial processes, LeCessie and van Houwelingen (1994) and Kuk and Nott (2000) for correlated binary data, and deLeon (2005) on grouped data). It is also possible to assign different weights to marginal likelihoods and/or consider higher dimensional marginals such as bivariate or trivariate. The CL approach would be particularly desirable if the full dimensional likelihood is difficult to work with (as in our case), or not straightforward to specify or compute.

We will utilise CL functions constructed from bivariate densities (see also Wang, Iglesias, and Wooldridge (2013) who propose a similar approach for estimation of spatial probit models which they call the "partial maximum likelihood estimation"). Specifically, suppose we pick N distinct asset-pairs from the initial collection of L assets. Let $X_{jt} = (r_{j_1 t}, r_{j_2 t})'$ be the vector of returns on the j^{th} pair at time t, where $j = 1, ..., N$, $(j_1, j_2) \in \{1, 2, ..., L\}^2$ and $j_1 \neq j_2$. Likewise, let $H_{jt}(\gamma_j, \lambda)$ be the conditional variance for the j^{th} pair, where the definitions of γ_j and λ depend on the particular model at hand. Then, assuming Gaussian innovations as before, the (negative) log-likelihood for pair j is given by

$$l_{jT}(\gamma_j, \lambda) = \frac{1}{T}\sum_{t=1}^{T}\left[\log\left|H_{jt}(\gamma_j, \lambda)\right| + X_{jt}' H_{jt}^{-1}(\gamma_j, \lambda) X_{jt}\right], \tag{9}$$

Implicit in this structure is the assumption that λ is common to all pairs, whereas γ_j is allowed to be pair-specific. The CL function based on all pairwise likelihoods is given by

$$l_{NT}\left(\gamma_1, ..., \gamma_N, \lambda\right) = \frac{1}{N}\sum_{j=1}^{N} l_{jT}\left(\gamma_j, \lambda\right). \tag{10}$$

Intuitively, (10) pools information on the common parameter λ across all pairs.

As before, γ_j can be estimated separately and plugged into (10) to yield $l_{NT}\left(\hat{\gamma}_1, ..., \hat{\gamma}_N, \lambda\right)$. The CL estimator for λ is then given by

$$\hat{\lambda} = \arg\min_{\lambda} l_{NT}\left(\hat{\gamma}_1, ..., \hat{\gamma}_N, \lambda\right). \tag{11}$$

In Section 4, we will discuss the application of this method to BEKK and cDCC in detail.

Remark 1 *In comparison to (7), the CL function (10) is free of (i) the computational burden of inverting large matrices, and (ii) the statistical issue of estimating large dimensional unconditional covariance/correlation matrices. In a closely related contribution, Engle, Ledoit, and Wolf (2019) address the latter problem by shrinkage methods (see also Hafner and Reznikova (2012)). In contrast, our approach is based on reducing the dimensionality of the problem by considering many low-dimensional bivariate models.*

Remark 2 *Although numerical estimation of λ by (11) involves inversion of a large number of 2-dimensional $H_{jt}\left(\hat{\gamma}_j, \lambda\right)$, the computational gains are still substantial. Evaluation of the objective function in (11) costs $O(N)$ calculations. When CL is based on all pairs, this means $O(L^2)$ calculations. This cost is favourable compared with the $O(L^3)$ calculations required by the full likelihood in (7). One can also use a selection of pairs, e.g. the contiguous pairs given by $X_{1t} = (r_{1t}, r_{2t})', X_{2t} = (r_{2t}, r_{3t})', X_{3t} = (r_{3t}, r_{4t})', \ldots, X_{Nt} = (r_{L-1,t}, r_{L,t})'$; the computational cost of this option is $O(L)$. The simulation analysis of Section 5 indicates that the contiguous- and all-pairs options perform similarly when L is moderately large.*

Remark 3 *The CL method is subject to efficiency loss, as it ignores potential dependence between the bivariate likelihoods. The exact magnitude of this loss will depend on the model under consideration; see, e.g., Mardia, Kent, Hughes, and Taylor (2009) and Kenne Pagui,* *Salvan, and Sartori (2015). While a theoretical analysis of this issue is beyond our scope, our experiments not reported here indicate that the loss is almost negligible when N is large.*

Remark 4 *Our approach can also be used in the context of more structured models, which impose stronger a priori constraints on the model. Factor models with time-varying volatility are the leading examples of this, e.g. King, Sentana, and Wadhwani (1994), Fiorentini, Sentana, and Shephard (2004) and Rangel and Engle (2012).*

4 Examples

4.1 BEKK

We consider CL estimation based on the variance-tracking scalar BEKK model given in (3). Let $\lambda = (\alpha, \beta)'$ and $\gamma_j = vec(\Gamma_j)$, where α and β are scalar parameters, and Γ_j is a (2×2) matrix. The conditional variance process for the j^{th} pair is given by

$$H_{jt}(\gamma_j, \lambda) = \Gamma_j (1 - \alpha - \beta) + \alpha X_{j,t-1} X'_{j,t-1} + \beta H_{j,t-1}(\gamma_j, \lambda). \tag{12}$$

Here $X_{jt} = H_{jt}^{1/2} Z_{jt}$, where Z_{jt} is the (2×1) version of Z_t defined in Section 2. Let the true parameter values be $\Gamma_{j0} = E[X_{jt} X'_{jt}]$ and $\lambda_0 = (\alpha_0, \beta_0)'$. Then for $l_{jT}(\gamma_j, \lambda)$ as defined in (9) and based on (12), the corresponding estimators are

$$\hat{\gamma}_j = vec\left(\frac{1}{T}\sum_{t=1}^{T} X_{jt} X'_{jt}\right), \quad \text{and} \quad \hat{\lambda} = \arg\min_{\lambda} \frac{1}{N}\sum_{j=1}^{N} l_{jT}(\hat{\gamma}_j, \lambda). \tag{13}$$

Let $\theta = (\gamma'_1, ..., \gamma'_N, \lambda')'$, $\theta_0 = (\gamma'_{10}, ..., \gamma'_{N0}, \lambda'_0)'$ and $\hat{\theta} = (\hat{\gamma}'_1, ..., \hat{\gamma}'_N, \hat{\lambda}')'$. The assumptions and the main result are presented next.

Assumption 4.1 *For all j, the distribution of Z_{jt} is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^2$, and zero is an interior point of the support of the distribution.*

Assumption 4.2 *The parameters α and β are such that $\alpha > 0$, $\beta > 0$ and $\alpha + \beta < 1$.*

Assumption 4.3 *The process $\{X_{jt}\}_{t=1}^{T}$ is strictly stationary and ergodic for all pairs.*

Assumption 4.4 *For all j, $\gamma_{j0} \in \Theta_\gamma$ and $\lambda_0 \in \Theta_\lambda$ where Θ_γ and Θ_λ are compact subsets of $\mathbb{R}^4$ and $\mathbb{R}^2$, respectively.*

Assumption 4.5 *For $\lambda \in \Theta_\lambda$, if $\lambda \neq \lambda_0$ then for all j we have $H_{jt}(\gamma_{j0}, \lambda) \neq H_{jt}(\gamma_{j0}, \lambda_0)$ almost surely, for all $t \geq 1$.*

Assumption 4.6 $E[||X_{jt}||^6] < \infty$ *for all pairs.*

Assumption 4.7 (γ_{j0}, λ_0) *is in the interior of $\Theta = \Theta_\gamma \times \Theta_\lambda$ for all j.*

Theorem 4.1 *Suppose Assumptions 4.1-4.7 hold. Then, $\hat{\theta} - \theta_0 \overset{a.s.}{\to} 0$, and*

$$\sqrt{T}(\hat{\theta}-\theta_0) \overset{d}{\to} N\left(0, \begin{bmatrix} I_{4N} & 0_{4N\times 2} \\ -J_N^{-1}(\theta_0)K_N(\theta_0) & -J_N^{-1}(\theta_0) \end{bmatrix} \Omega_0 \begin{bmatrix} I_{4N} & 0_{4N\times 2} \\ -J_N^{-1}(\theta_0)K_N(\theta_0) & -J_N^{-1}(\theta_0) \end{bmatrix}'\right),$$

as $T \to \infty$, where $J_N(\theta_0)$, $K_N(\theta_0)$ and Ω_0 are as defined in Sections B.1 and B.3 in the Supplementary Appendix.

Remark 5 *We focus on scalar BEKK here, to keep the simulation and empirical analysis parts tractable. However, one can also consider the more flexible non-scalar BEKK, letting*

$$H_{jt}(\gamma_j, \lambda) = \Gamma_j - A\Gamma_j A' - B\Gamma_j B' + AX_{j,t-1}X'_{j,t-1}A' + BH_{j,t-1}(\gamma_j, \lambda)B'. \tag{14}$$

Here A and B are some (2×2) matrices, and so $\lambda = (vec(A)', vec(B)')'$ while γ_j remains the same as before. In Theorem C.1 in Section C of the Supplementary Appendix, we prove that the CL estimator is consistent and asymptotically normal in this case, as well.

Remark 6 *The CL estimator of Theorem 4.1 requires the same set of assumptions as the ML estimator of PR (see their Theorems 4.1 and 4.2), with the only additional requirement that they hold for all pairs (the same holds for non-scalar BEKK as Theorem C.1 reveals). This is not surprising since the CL and ML objective functions are closely related: the CL objective function, $N^{-1}\sum_{j=1}^{N} l_{jT}(\hat{\gamma}_j, \lambda)$, is the average of bivariate BEKK objective functions for ML estimation, $l_{jT}(\hat{\gamma}_j, \lambda)$. We note that this relationship is not specific to BEKK; it would apply to any volatility model with pair-specific (γ_j) and common (λ) parameters. We* *therefore expect this pattern, where CL estimation requires the same set of assumptions as ML, to hold for a range of multivariate volatility models.*

Remark 7 *Some of our assumptions follow directly if they already hold for the L-dimensional model in (3). Specifically, Assumption 4.1 holds if the same conditions hold for Z_t. Next, using Theorem 3.35 of White (2001) it is straightforward to show that Assumption 4.3 holds if X_t is stationary and ergodic. Moreover, Assumption 4.6 is trivially implied by $E[||X_t||^6] < \infty$. Assumptions 4.2, 4.4 and 4.7 would directly hold for γ_j if the same assumptions hold for γ (this is because the entries of γ_j are contained in γ). One exception in this discussion is Assumption 4.5 which does not necessarily follow from the corresponding condition for the conditional variance of the L-dimensional model. However, we note that Assumption 4.5 is used in proving that $E[\partial^2 N^{-1} \sum_{j=1}^{N} l_{jt}(\gamma_{j0}, \lambda_0) / \partial\lambda\partial\lambda']$ is non-singular and it would be possible to obtain this result even if Assumption 4.5 fails for some pairs.*

Remark 8 *The asymptotic variance of $\hat{\lambda} = (\hat{\alpha}, \hat{\beta})$ is usually of more interest, which is provided in Section B.4 of the Supplementary Appendix.*

Remark 9 *We note that the dimension of $\hat{\theta} - \theta_0$ exceeds the number of the unique parameters to be estimated. For instance, in the all-pairs case the dimension of $\hat{\theta} - \theta_0$ is $2(L^2 - L) + 2$ although there are $((L^2 + L)/2) + 2$ unique parameters to be estimated: this is because of our notation where each unconditional covariance appears twice whereas each unconditional variance appears $L - 1$ times in θ. Note that, for example, the notation of PR similarly includes all covariances twice (see also their Remark 3.2).*

4.2 cDCC

We next consider CL estimation of the cDCC model (Aielli (2013)). Let $h_{j_1,t}(\eta_{j_1})$ and $h_{j_2,t}(\eta_{j_2})$ be the univariate volatility models for assets j_1 and j_2, where η_{j_1} and η_{j_2} are the model parameters. For illustration we assume that the univariate volatility model for all assets is GARCH(1,1), and so η_{j_1} and η_{j_2} are both (3×1) for all (j_1, j_2). Let $\varepsilon_{jt}(\theta_j) = (\varepsilon_{j_1,t}(\eta_{j_1}), \varepsilon_{j_2,t}(\eta_{j_2}))'$ where $\theta_j = (\eta'_{j_1}, \eta'_{j_2})'$, $\varepsilon_{j_1,t}(\eta_{j_1}) = r_{j_1,t} h_{j_1,t}^{-1/2}(\eta_{j_1})$, and similarly for

$\varepsilon_{j_2,t}(\eta_{j_2})$. Letting $\phi = (\alpha, \beta)'$, the cDCC model for the j^{th} pair is given by

$$\begin{aligned} H_{jt}(\theta_j, S_j(\theta_j,\phi),\phi) &= D_{jt}^{1/2}(\theta_j) R_{jt}(\theta_j, S_j(\theta_j,\phi),\phi) D_{jt}^{1/2}(\theta_j), \\ R_{jt}(\theta_j, S_j(\theta_j,\phi),\phi) &= Q_{jt}^{*-1/2}(\theta_j,\phi) Q_{jt}(\theta_j, S_j(\theta_j,\phi),\phi) Q_{jt}^{*-1/2}(\theta_j,\phi), \\ Q_{jt}(\theta_j, S_j(\theta_j,\phi),\phi) &= (1-\alpha-\beta) S_j(\theta_j,\phi) \\ &\quad +\alpha\{Q_{j,t-1}^{*1/2}(\theta_j,\phi)\varepsilon_{j,t-1}(\theta_j)\varepsilon'_{j,t-1}(\theta_j) Q_{j,t-1}^{*1/2}(\theta_j,\phi)\} + \beta Q_{j,t-1}(\theta_j,\phi), \\ S_j(\theta_j,\phi) &= E[Q_{jt}^{*1/2}(\theta_j,\phi)\varepsilon_{jt}(\theta_j)\varepsilon'_{jt}(\theta_j) Q_{jt}^{*1/2}(\theta_j,\phi)], \end{aligned}$$

where $D_{jt}(\theta_j)$ is a (2×2) diagonal matrix with the diagonal entries $h_{j_1,t}(\eta_{j_1})$ and $h_{j_2,t}(\eta_{j_2})$. $Q^*_{jt}(\theta_j,\phi)$ is a (2×2) matrix consisting of the diagonal entries of $Q_{jt}(\theta_j, S_j, \phi)$.

Estimation of cDCC using composite likelihood uses the same revolatilisation step as in Aielli (2013) which allows the intercept to be targeted, conditional on the correlation dynamics. First, for each pair j we obtain $\hat{\theta}_j = (\hat{\eta}'_{j_1}, \hat{\eta}'_{j_2})' = \arg\max_{\eta_{j_1},\eta_{j_2}}(\dot{l}_{j_1T}(\eta_{j_1}), \dot{l}_{j_2T}(\eta_{j_2}))'$, where $\dot{l}_{iT}(\cdot)$ is the log-likelihood function for asset i. Next, we define $l_{jT}(\theta_j, S_j, \phi) = T^{-1}\sum_{t=1}^{T} l_{jt}(\theta_j, S_j, \phi)$ where $l_{jt}(\theta_j, S_j, \phi) = -\log|H_{jt}(\theta_j, S_j, \phi)| - X'_{jt}H_{jt}^{-1}(\theta_j, S_j, \phi)X_{jt}$. Then, the composite likelihood estimator of ϕ is given by

$$\hat{\phi} = \arg\max_{\phi} \frac{1}{N}\sum_{j=1}^{N} l_{jT}(\hat{\theta}_j, \hat{S}_j(\hat{\theta}_j,\phi),\phi),$$

where $\hat{S}_j(\theta_j,\phi) = T^{-1}\sum_{t=1}^{T} Q_{jt}^{*1/2}(\theta_j,\phi)\varepsilon_{jt}(\theta_j)\varepsilon'_{jt}(\theta_j)Q_{jt}^{*1/2}(\theta_j,\phi)$. Finally, $\hat{S}_j(\hat{\theta}_j,\hat{\phi})$ is the estimator of $S_j(\theta_{j0},\phi_0)$.

Due to their considerably more complicated structure, a complete asymptotic theory is still missing for the DCC and cDCC models (important contributions in this direction are due to Aielli (2013), Francq and Zakoïan (2016), and Fermanian and Malongo (2017)). However, a heuristic argument for cDCC is provided in Sections 3.2 and 3.3 of Aielli (2013), and our asymptotic treatment here will follow along similar lines. In particular, in Section D.1.1 of the Supplementary Appendix we provide a heuristic proof of $\hat{S}_j(\hat{\theta}_j,\hat{\phi}) \xrightarrow{p} S_j(\theta_{j0},\phi_0)$ and $\hat{\phi} \xrightarrow{p} \phi_0$ as $T\to\infty$ (primitive conditions for $\hat{\theta}_j \xrightarrow{p} \theta_{j0}$ are already generally available for many univariate volatility models). Key conditions for consistency are $\sup_{\theta_j,\phi}||\hat{S}_j(\theta_j,\phi) - S_j(\theta_j,\phi)|| \xrightarrow{p} 0$, $\sup_{\theta_j,S_j,\phi}|l_{jT}(\theta_j,S_j,\phi) - E[l_{jT}(\theta_j,S_j,\phi)]| \xrightarrow{p} 0$, and the identification condition that ϕ_0 uniquely maximises $E[(NT)^{-1} \sum_{j=1}^{N} \sum_{t=1}^{T} l_{jt}(\theta_{j0}, S_j(\theta_{j0}, \phi), \phi)]$. We note that similar conditions are also required for the arguments in Section 3.2 of Aielli (2013), and the two estimators are identical at $L = 2$.

It is straightforward to adapt the same line of reasoning as in Section 3.3 of Aielli (2013) and obtain the asymptotic distribution for the entire collection of parameters $(\hat{\theta}_1, ..., \hat{\theta}_N)$, $(\hat{S}_1(\hat{\theta}_1, \hat{\phi}), ..., \hat{S}_N(\hat{\theta}_N, \hat{\phi}))$ and $\hat{\phi}$. However, in applications the interest is often on the asymptotic distribution of $\hat{\phi}$ only. As such, in Section D.1.2 in the Supplementary Appendix we provide a detailed argument for obtaining

$$\sqrt{T}(\hat{\phi} - \phi) \xrightarrow{d} N(0_2, J_N \Sigma_N J_N') \quad \text{ as } T \to \infty. \tag{15}$$

The terms J_N and Σ_N are defined in equations (D.16) and (D.17) in Section D.1.2. The non-standard three-step estimation of the cDCC model requires careful definition of the estimating equations and the relevant notation. For space considerations, we do this in the Supplementary Appendix. Nevertheless, to provide some detail, Σ_N is the asymptotic variance of the vector that contains the estimating equations for $(\hat{\theta}_1, ..., \hat{\theta}_N)$, $(\hat{S}_1\left(\theta_1, \phi\right), ..., \hat{S}_N\left(\theta_N, \phi\right))$ and $\hat{\phi}$. J_N, on the other hand, is a $(2 \times (10N + 2))$ matrix based on the derivatives of the estimating equations for $\hat{\phi}$ and $(\hat{\theta}_1, ..., \hat{\theta}_N)$. As usual, Σ_N can be estimated by a HAC estimator (Newey and West (1987)), whereas J_N can be replaced by its sample counterpart.

5 Monte Carlo experiments

5.1 Relative performance of estimators

We compare the CL method to ML estimation of the full dimensional model, which we call 2MLE ("2" stands here to denote the two-step structure where the first-step involves estimation of a targeting matrix). We consider the CL estimator based on all pairs (2MCLE) and on contiguous pairs (2MSCLE) as explained in Remark 2. We refer to the latter option also as subset CL. Throughout we use the correctly specified volatility model for estimation (e.g. if data are generated by scalar BEKK, then all estimators are based on scalar BEKK).

The first part of the Monte Carlo study is based on 2500 replications with $T = 2000$

	Bias						RMSE					
	2MLE		2MCLE		2MSCLE		2MLE		2MCLE		2MSCLE	
L	α	β	α	β	α	β	α	β	α	β	α	β
						$\alpha = .02, \beta = .97$						
3	.000	-.005	.000	-.005	.000	-.006	.005	.009	.005	.010	.006	.012
10	-.001	-.003	.000	-.004	.000	-.004	.002	.004	.003	.006	.003	.007
50	-.005	-.000	.000	-.004	.000	-.004	.005	.001	.002	.005	.002	.005
100	-.009	-.001	.000	-.004	.000	-.004	.009	.001	.002	.005	.002	.005
						$\alpha = .05, \beta = .93$						
3	-.000	-.008	-.000	-.009	.000	-.010	.008	.023	.009	.025	.010	.029
10	-.001	-.005	-.000	-.007	-.000	-.007	.003	.009	.005	.014	.006	.015
50	-.006	-.003	-.000	-.006	-.000	-.006	.006	.004	.003	.009	.003	.009
100	-.012	-.004	-.000	-.006	-.000	-.006	.012	.004	.003	.009	.003	.009
						$\alpha = .10, \beta = .80$						
3	-.001	-.005	-.001	-.006	-.001	-.006	.013	.028	.014	.030	.015	.033
10	-.003	-.003	-.001	-.005	-.001	-.005	.006	.011	.009	.019	.009	.019
50	-.014	.001	-.001	-.005	-.001	-.005	.015	.004	.006	.012	.006	.012
100	-.026	.001	-.001	-.005	-.001	-.005	.026	.003	.006	.012	.006	.012

Table 2: Bias and RMSE of the estimators of α and β in the covariance tracking scalar BEKK model. The estimators are: CL based on all pairs (2MCLE), CL based on a subset of pairs (2MSCLE), and the full-dimensional likelihood (2MLE). Based on 2500 replications with $T = 2000$.

while $L \in \{3, 10, 50, 100\}$. The returns are simulated using the scalar BEKK and cDCC models of Sections 4.1 and 4.2, respectively (for the cDCC part we fix the conditional variance to be one, and so omit estimation of univariate GARCH model by setting $h_{it} = 1$). We consider three cases spanning a fair range of empirically relevant values of the temporal dependence structure in multivariate volatilities, $(\alpha, \beta) = (.02, .97)$, $(.05, .93)$ or $(.10, .80)$. The long-run targets were constructed to have a strong factor structure and reflect typical values found in the S&P 100. For BEKK, the unconditional covariance was generated according to $\Gamma = \omega_f \beta\beta' + \Omega$ where (i) $\omega_f = (0.2)^2$, (ii) the elements of β are given by $\beta_i = v_i/5$ where $v_i \sim \chi^2_5$, and (iii) Ω is diagonal with each element generated according to $0.1 + 0.2u_i/5$ where $u_i \sim \chi^2_5$. Here $i = 1, ..., L$. For cDCC, the unconditional correlations were generated from a single-factor model, ensuring that $S_{i,i'} = \pi_i \pi_{i'}$ for $i \neq i'$ and $S_{i,i'} = 1$ for $i = i'$, where $i, i' = 1, ..., L$. We generate π_i from a truncated normal with mean 0.5 and standard deviation 0.1, where the truncation occurs at ± 4 standard deviations. This means $\pi_i \in (0.1, 0.9)$ and the average correlation in the cross section is 0.25. This choice

	Bias						RMSE					
	2MLE		2MCLE		2MSCLE		2MLE		2MCLE		2MSCLE	
L	α	β	α	β	α	β	α	β	α	β	α	β
						$\alpha = .02, \beta = .97$						
3	.001	-.011	.001	-.012	.001	-.017	.006	.033	.007	.038	.008	.059
10	-.001	-.004	-.000	-.005	-.000	-.006	.002	.005	.002	.006	.003	.009
50	-.003	-.003	-.000	-.005	-.000	-.005	.003	.003	.001	.005	.002	.006
100	-.005	-.004	-.000	-.005	-.000	-.005	.005	.004	.001	.005	.001	.005
						$\alpha = .05, \beta = .93$						
3	-.000	-.005	-.000	-.006	-.000	-.007	.008	.015	.009	.016	.011	.022
10	-.002	-.001	-.000	-.003	-.000	-.004	.003	.004	.003	.006	.005	.009
50	-.009	.003	-.001	-.003	-.001	-.003	.009	.003	.002	.004	.003	.005
100	-.014	.002	-.001	-.003	-.001	-.003	.014	.002	.002	.004	.002	.004
						$\alpha = .10, \beta = .80$						
3	-.001	-.007	-.001	-.008	-.001	-.010	.016	.037	.017	.040	.019	.051
10	-.003	-.003	-.001	-.005	-.001	-.006	.006	.011	.007	.016	.009	.022
50	-.014	.000	-.001	-.004	-.001	-.004	.014	.004	.004	.009	.005	.011
100	-.024	-.003	-.001	-.004	-.001	-.004	.024	.004	.004	.008	.005	.010

Table 3: Bias and RMSE of the estimators of α and β in the cDCC model. The estimators are: CL based on all pairs (2MCLE), CL based on a subset of pairs (2MSCLE), and the full-dimensional likelihood (2MLE). Based on 2500 replications with $T = 2000$.

for S produces assets which are all positively correlated and ensures that the intercept is positive definite for any cross-sectional dimension L. These results are not sensitive to the choice of unconditional correlation as long as the target has a strong factor structure. In both cases, estimation is subject to the constraints $0 \leq \alpha < 1$, $0 \leq \beta < 1$, and $\alpha + \beta < 1$ (nevertheless, none of the estimators in the simulations hit the parameter space boundary).

Results for BEKK and cDCC are given in Tables 2 and 3, respectively, which contain the bias and root mean square error (RMSE) of the estimates. For both BEKK and cDCC the full-dimensional 2MLE develops a significant bias in estimating α as L increases. This is consistent with our earlier theoretical discussion. The same is not observed for 2MCLE and 2MSCLE. These results confirm that our proposed approach is effective in estimating the chosen volatility models for large collections of assets, without suffering from any bias. We also note that the direction of the MLE bias observed in simulated data (Tables 2-3) is in agreement with the results based on real data (Table 1), although the results differ in their magnitudes. Misspecification is one possible cause for this difference, as the anal-

	Bias						RMSE					
	2MLE		2MCLE		2MSCLE		2MLE		2MCLE		2MSCLE	
T	α	β	α	β	α	β	α	β	α	β	α	β
						$L = 10$						
100	-.021	-.161	-.011	-.141	-.009	-.218	.025	.237	.021	.221	.028	.347
250	-.006	-.018	-.002	-.021	-.002	-.026	.008	.021	.008	.026	.012	.042
500	-.003	-.005	-.001	-.008	-.001	-.009	.005	.008	.005	.011	.007	.016
1000	-.002	-.001	-.001	-.003	-.001	-.003	.003	.004	.004	.006	.005	.009
2000	-.001	-.000	-.000	-.002	-.000	-.002	.002	.003	.003	.004	.004	.006
						$L = 50$						
100	-.050	-.915	-.014	-.091	-.013	-.108	.050	.915	.016	.103	.018	.146
250	-.022	-.034	-.003	-.018	-.003	-.019	.022	.034	.005	.020	.006	.022
500	-.013	-.004	-.001	-.007	-.001	-.007	.013	.004	.003	.009	.004	.010
1000	-.009	.003	-.001	-.003	-.001	-.003	.009	.003	.002	.004	.003	.005
2000	-.006	.003	-.000	-.001	-.000	-.001	.006	.003	.001	.002	.002	.003
						$L = 100$						
100	–	–	-.014	-.090	-.014	-.098	–	–	.016	.103	.017	.121
250	-.037	-.108	-.003	-.019	-.003	-.019	.037	.109	.004	.020	.005	.021
500	-.021	-.013	-.001	-.007	-.001	-.007	.021	.013	.003	.008	.003	.009
1000	-.014	.001	-.001	-.003	-.001	-.003	.014	.002	.002	.004	.002	.004
2000	-.010	.004	-.000	-.001	-.000	-.001	.010	.004	.001	.002	.002	.003
						$L = 200$						
100	–	–	-.014	-.086	-.013	-.082	–	–	.016	.092	.016	.095
250	-.050	-.913	-.002	-.018	-.003	-.018	.050	.918	.004	.019	.005	.019
500	-.033	-.053	-.001	-.007	-.001	-.007	.033	.053	.002	.008	.003	.008
1000	-.021	-.006	-.000	-.003	-.001	-.003	.022	.006	.002	.004	.002	.004
2000	-.015	.003	-.000	-.002	-.000	-.001	.015	.003	.001	.002	.001	.002

Table 4: Bias and RMSE of α and β in the cDCC model, where the true values are $\alpha = .05$ and $\beta = .93$. The estimators are: CL based on all pairs (2MCLE), CL based on a subset of pairs (2MSCLE), and the full-dimensional likelihood (2MLE). Based on 2500 replications.

ysis of Table 1 can be subject to misspecification of some type: the volatility model may be misspecified and/or the volatility parameters α and β can be time-varying. It is also possible that the severity of the effect of dimensionality depends on the type of misspecification. More generally, there may be other factors at play, beyond misspecification. An investigation of this is beyond our scope, and is left for for future work.

Next, to obtain a broader picture, we consider the same analysis with $L \in \{10, 50, 100, 200\}$ and $T \in \{100, 250, 500, 1000, 2000\}$, for the specific case of cDCC. The results are presented in Table 4 (for size considerations, we report the results for $(\alpha, \beta) = (.05, .93)$ only). Not surprisingly, for any estimation method and any fixed L, bias decreases with T. For

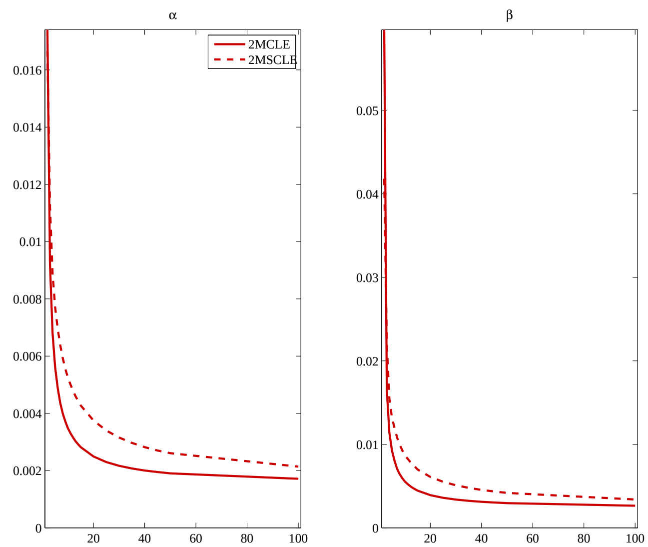

Figure 1: Standard deviation for the CL estimator based on all pairs (2MCLE) and on a subset of pairs (2MSCLE), as L varies from 2 to 100. Calculated from simulated data for the cDCC model with $\alpha = .05$, $\beta = .93$.

any given T, the bias of 2MLE clearly deteriorates with L; for example, at $L = 200$, α is biased downward by 30% even when $T = 2000$. The same is not observed for 2MCLE and 2MSCLE; in fact, for $T \geq 250$ the CL methods are not subject to any serious bias. Moreover, a systematic deterioration in bias as L increases is absent. Interestingly, when $T = 100$, $\hat{\beta}$ seems to improve with L. Finally, simulation results reveal that 2MCLE and 2MSCLE are feasible even when $T \leq L$. On the other hand, results for 2MLE in the $T \leq L$ case could not even be reported as the estimator failed to converge in most replications.

Overall, we observe that the full-dimensional 2MLE suffers from a non-vanishing bias as L increases, but CL is immune to this problem. Moreover, 2MCLE and 2MSCLE have better RMSE for all cross-section sizes and parameter configurations. Importantly, there seems to be little difference between using all or only contiguous pairs. Hence, 2MSCLE stands out as an attractive option for large dimensional problems.

5.2 Efficiency gains with increasing cross-section size

Clearly, the gap between the number of pairs used by 2MCLE and 2MSCLE widens quickly with L. For example, when $L = 50$, 2MCLE is based on 1225 pairs while 2MSCLE uses

49 pairs. Hence, an important question is whether the efficiency loss due to using only a subset of all pairs is large enough to render 2MSCLE an unattractive option.

Figure 1 plots the square root of the average estimator variance against L, for the cDCC model when estimated by 2MCLE and 2MSCLE. The standard deviations decline rapidly as the cross-section dimension grows. Of course, as 2MCLE is always based on a larger number of pairs, its standard deviation is also always below that of 2MSCLE. However, importantly, the difference between the standard deviations drops very quickly and becomes negligible by $L = 100$. The corresponding analysis for BEKK, which is reported in the Supplementary Appendix, delivers the same results except that, interestingly, the efficiency loss is even less (see Figure D.1 in Section D.2 in the Supplementary Appendix). This is a strong argument for exploiting the computational simplicity of 2MSCLE.

5.3 Performance of asymptotic standard errors

We now assess the accuracy of the asymptotic covariance estimator, using simulated data for various (α, β). These simulations are based on 1000 replications with $T = 2000$. The asymptotic variance estimator is based on (15) for cDCC and Theorem 4.1 for scalar BEKK. For cDCC, we use the HAC variance estimator of Newey and West (1987) to estimate Σ_N, while J_N is estimated by using its sample version evaluated at $(\hat{\theta}, \hat{s}(\hat{\theta}, \hat{\phi}), \hat{\phi})$. The asymptotic variance for BEKK is estimated similarly. The results for the scalar BEKK and cDCC models are virtually identical, and so we only discuss those of the cDCC.

The results are presented in Table 5, which contains the square roots of (i) the average asymptotic variance, $\bar{\sigma}^2_\alpha = \frac{1}{1000}\sum_{r=1}^{1000} \hat{\sigma}^2_{r,\alpha}$ where $\hat{\sigma}^2_{r,\alpha}$ is the estimated variance for $\hat{\alpha}$ in replication r and (ii) the sample variance of the parameter estimates across replications, $\hat{\sigma}^2_\alpha = \frac{1}{1000}\sum_{r=1}^{1000} \left(\hat{\alpha}_r - \bar{\hat{\alpha}}\right)^2$ with $\bar{\hat{\alpha}} = \frac{1}{1000}\sum_{r=1}^{1000} \hat{\alpha}_r$ where $\hat{\alpha}_r$ is $\hat{\alpha}$ in replication r, and similarly for β. The results confirm that the asymptotic covariance specification in (15) is accurate for $L \geq 10$, and therefore provides a sensible basis for inference.

5.4 Assessing CLT Accuracy

The asymptotic distribution for the CL estimator of BEKK was derived in Theorem 4.1, which requires $E[||X_{jt}||^6] < \infty$. In this section we assess the quality of this CLT, and also

	Scalar BEKK								cDCC							
	2MCLE				2MSCLE				2MCLE				2MSCLE			
L	$\bar{\sigma}_\alpha$	$\hat{\sigma}_\alpha$	$\bar{\sigma}_\beta$	$\hat{\sigma}_\beta$	$\bar{\sigma}_\alpha$	$\hat{\sigma}_\alpha$	$\bar{\sigma}_\beta$	$\hat{\sigma}_\beta$	$\bar{\sigma}_\alpha$	$\hat{\sigma}_\alpha$	$\bar{\sigma}_\beta$	$\hat{\sigma}_\beta$	$\bar{\sigma}_\alpha$	$\hat{\sigma}_\alpha$	$\bar{\sigma}_\beta$	$\hat{\sigma}_\beta$
								$\alpha = .02$, $\beta = .97$								
3	.009	.008	.121	.176	.010	.009	.139	.190	.010	.008	.261	.152	.008	.007	.052	.028
10	.001	.001	.004	.003	.002	.002	.005	.007	.002	.002	.004	.004	.003	.003	.008	.007
50	.001	.001	.002	.001	.002	.002	.002	.002	.001	.001	.002	.002	.002	.002	.003	.003
100	.001	.000	.001	.001	.001	.002	.001	.001	.001	.001	.002	.001	.001	.001	.002	.002
								$\alpha = .05$, $\beta = .93$								
3	.008	.007	.013	.013	.009	.010	.018	.020	.009	.009	.016	.015	.011	.010	.021	.019
10	.003	.004	.006	.005	.004	.006	.006	.007	.003	.003	.006	.006	.005	.005	.009	.009
50	.001	.002	.003	.004	.003	.003	.004	.004	.002	.002	.003	.003	.003	.003	.004	.004
100	.001	.002	.003	.004	.001	.001	.003	.004	.002	.002	.003	.003	.002	.002	.003	.003
								$\alpha = .10$, $\beta = .80$								
3	.015	.019	.032	.036	.019	.021	.047	.054	.017	.016	.041	.040	.020	.019	.052	.049
10	.006	.006	.013	.012	.008	.008	.020	.018	.007	.006	.015	.014	.009	.010	.022	.022
50	.003	.004	.006	.007	.004	.005	.008	.009	.004	.004	.008	.008	.005	.005	.011	.011
100	.002	.002	.006	.007	.003	.004	.008	.009	.003	.003	.007	.007	.004	.004	.009	.009

Table 5: Square root of average asymptotic variance, denoted $\bar{\sigma}_\alpha$ and $\bar{\sigma}_\beta$, and standard deviation of the Monte Carlo estimated parameters, denoted $\hat{\sigma}_\alpha$ and $\hat{\sigma}_\beta$. Based on 2500 replications for the scalar BEKK and cDCC models with $T = 2000$.

investigate its sensitivity to existence of moments. The baseline model is a scalar BEKK with ($\alpha = .05$, $\beta = .93$), which produces return series with more than 8 moments. We also simulate data from scalar BEKK processes parameterised to have 6 to 4 moments. Our theory applies to the case with 6 moments but not to the process with 4. The parameters of these additional simulated returns were generated by restricting the persistence to be $\alpha + \beta = .98$ using $\alpha = .07$ ($E[X_t^6] < \infty$) and $\alpha = .14$ ($E[X_t^4] < \infty$).

Figure 2 contains Q-Q plots of the normalised estimates by 2MCLE which have been centered and scaled according to the distribution in the CLT, where $L = 50$ and $T = 2000$. The CLT provides an accurate description of the distribution of the parameter errors when the model parameters are chosen to produce 6 or more moments. When the model does not have 6 moments, the bias in the estimated α and β is sufficient to push the distribution away from the 45-degree line. The Supplementary Appendix contains results for other configurations of L and T (see Figures D.2 and D.3). These reveal that increasing the number of series has no impact on the quality of the CLT approximation. Decreasing the sample size adversely affects the distribution due to a small increase in bias in β. The

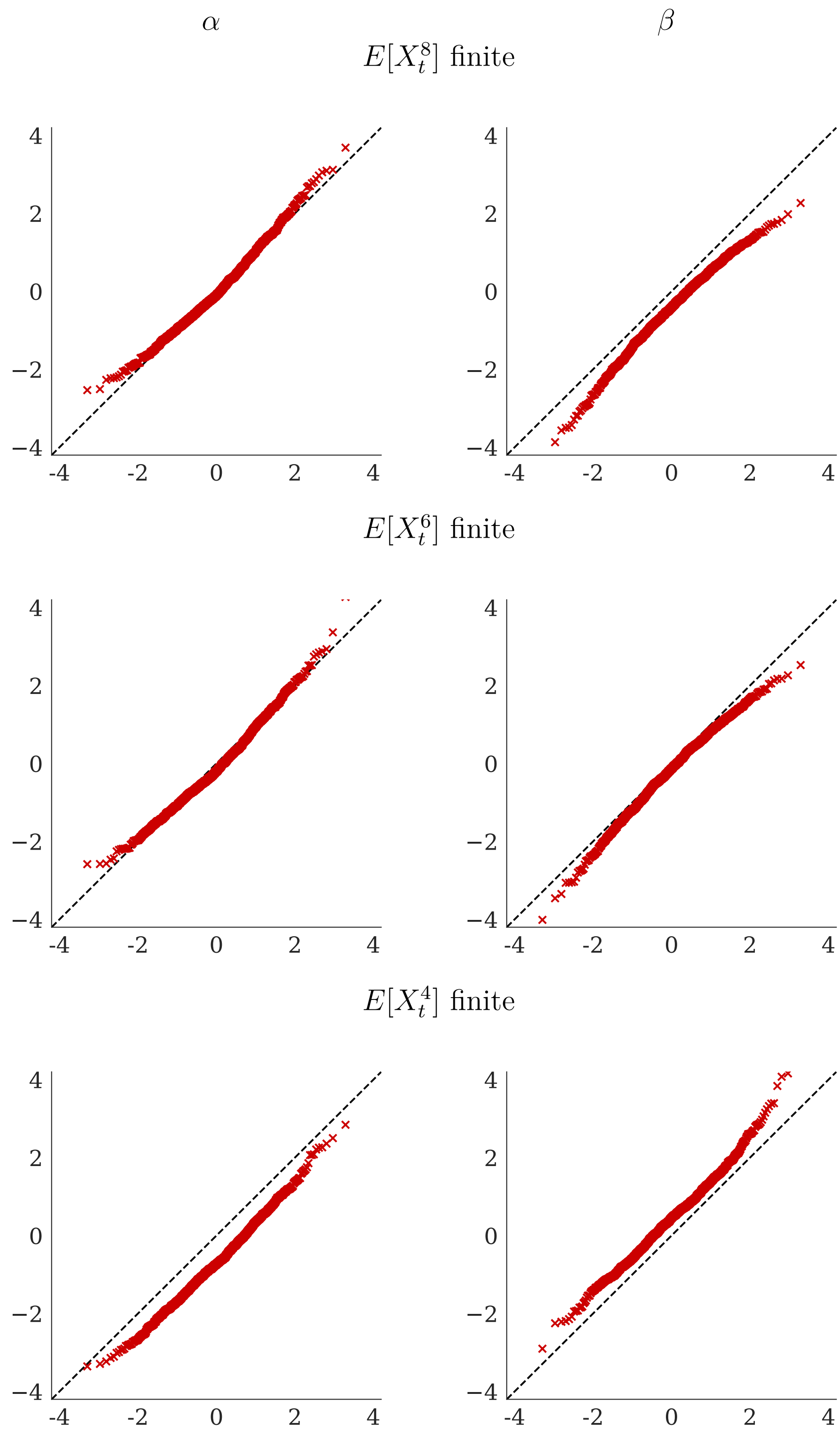

Figure 2: Q-Q plots of estimates of α and β from Scalar BEKK models parameterized to have 8, 6 and 4 finite moments. The normalized parameter errors are plotted along the y-axis. All estimates were produced from models with $L = 50$ and $T = 2000$ using all-pairs 2MCLE.

appendix also contains Q-Q plots of the normalised parameters estimated by 2MLE (Figure D.4). These estimates are so severely biased that the entire distribution is on one side of the center – $\hat{\alpha}$ is uniformly too small and $\hat{\beta}$ is always larger than the true value.

5.5 The role of covariance target uncertainty

Next, we investigate the effect of the dimensionality of the problem on estimating the long-run target matrices Γ (for BEKK) and S (for cDCC). All methods we have considered use the average outer-product of the returns to estimate these quantities. For the full-dimensional case, it is well-known that this is subject to bias since the estimation problem is $O(L^2)$ dimensional which exceeds $O(T)$ even for moderate values of L. By definition, no such issue exists for 2MCLE and 2MSCLE since Γ_j and S_j are always (2×2). As the estimates of these objects enter the likelihood function via the conditional variance specification, significant bias in these estimates would have a substantial impact on $\hat{H}_{jt}$ and $\hat{H}_t$, and consequently on the estimation of (α, β).

We explore this issue for scalar BEKK, by examining the condition number of the conditional covariance matrix for both 2MLE and 2MCLE. Eigenvalues, especially small eigenvalues, play an important role when inverting matrices or computing the log-determinant. Conditioning numbers provide a simple way to express the relative accuracy of eigenvalue estimates and the accuracy of a matrix inverse. Here we examine the accuracy of the long-run target estimator of the 2MLE and 2MCLE methods. In particular, the ratio of the condition numbers of $\hat{H}_t$ and H_t for 2MLE is given by

$$\hat{u}_{Lt} = \frac{\hat{\lambda}_{t,\max} / \hat{\lambda}_{t,\min}}{\lambda_{t,\max} / \lambda_{t,\min}},$$

where $\hat{\lambda}_{\min,t}, \hat{\lambda}_{2,t}, \ldots, \hat{\lambda}_{L-1}, \hat{\lambda}_{\max,t}$ are the ordered eigenvalues of $\hat{H}_t$, the conditional covariance based on estimated parameter values. $\lambda_{t,\min}$ and $\lambda_{t,\max}$ are defined similarly for H_t, the conditional covariance based on the true parameter values. We also compute the same for the bivariate conditional covariances for each unique pair $j = (j_1, j_2)$,

$$\hat{u}_{jt} = \frac{\hat{\lambda}_{jt,\max} / \hat{\lambda}_{jt,\min}}{\lambda_{jt,\max} / \lambda_{jt,\min}}.$$

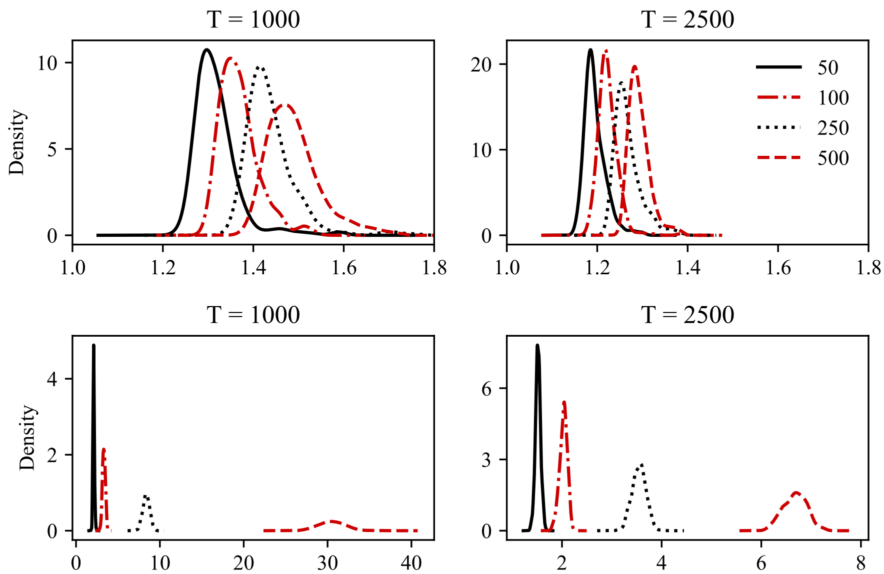

Figure 3: Sample distributions for the average ratio of realised (sample) condition number to true condition number. Based on simulated data from the scalar BEKK model with $\alpha = .05$ and $\beta = .93$. The top panel contains the ratio for the 2MCLE for $L \in \{50, 100, 250, 500\}$ for $T = 1000$ and $T = 2500$. The bottom panel contains the same for 2MLE.

Data are simulated for the scalar BEKK model using the same specification as described in Section 5.1. In particular, we let $\alpha = .05$ and $\beta = .93$. To isolate the covariance matrix estimation problem from estimation of (α, β), we compute $\hat{u}_{Lt}$ and $\hat{u}_{jt}$ using the true values of α and β. Therefore, any difference stems from the use of Γ or $\hat{\Gamma}$.

Our results are summarised in Figure 3. The top panel plots the sample density of $\bar{u}_{c,LT} = T^{-1} \sum_{t=1}^{T} \max_{1 \leq j \leq N} \hat{u}_{jt}$ for $L \in \{50, 100, 250, 500\}$, and $T = 1000$ and 2500 (we consider all distinct pairs, so $N = (L^2 - L)/2$). The bottom panel plots the same for $\bar{u}_{LT} = T^{-1} \sum_{t=1}^{T} \hat{u}_{Lt}$. Sample distributions of $\bar{u}_{LT}$ reveal that the ratio of conditioning numbers deteriorates rapidly for 2MLE as L increases, even when $T = 2500$. The worst observed ratio for 2MLE is over 40, which occurs with $L = 500$ and $T = 1000$. This contrasts heavily with 2MCLE, the worst observed ratio for which is less than 1.8. Clearly the ratio increases with L and decreases with T for both methods. However, the magnitude of these effects is quite different. For example, an increase in L does not have the same dramatic effect on 2MCLE. To further quantify this difference, in Section D.4 in the Supplementary Appendix we back out the implied magnitudes of these effects from simulated data.

	EBEE		2MCLE	
L	α	β	α	β
10	0.0165	0.0545	0.0118	0.0190
25	0.0162	0.0515	0.0116	0.0185
50	0.0163	0.0547	0.0114	0.0184

Table 6: RMSE of the EBEE and all-pairs CL (2MCLE) methods for diagonal BEKK estimation. Columns 2-3 report the results for EBEE whereas the results for 2MCLE are contained in the last two columns. The number of assets, L, is reported in the first column. The true underlying model is scalar BEKK with $\alpha = .05$ and $\beta = .93$. In all cases $T = 1000$. Based on 1000 replications.

5.6 Comparison with the EBEE method

Francq and Zakoïan (2016) have recently introduced a new method to estimate some multivariate models using only univariate estimators. This method is known as equation-by-equation estimation (EBEE). In particular, it also encompasses the diagonal and semi-diagonal BEKK models (the latter specification allows for spillovers through the shocks but limits the model to include only lags of own variance or covariance). By this approach, the parameters governing the dynamics in the model can be estimated using L univariate augmented GARCH models. In this part, we conduct a small simulation study to compare EBEE to composite likelihood estimation in the particular case of diagonal BEKK. Although we fit a diagonal BEKK model in this study, to keep the analysis tractable we use simulated data from a scalar BEKK model with $\alpha = .05$, $\beta = .93$, for $T = 1000$ and $L \in \{10, 25, 50\}$. The EBEE is implemented using L univariate GARCH(1,1) models which produce consistent estimates of the $2L$ parameters in this model. The CL estimator is based on the all-pairs specification (2MCLE) and jointly estimates the $2L$ parameters. Table 6 contains the RMSE statistics for the two methods, based on 1000 replications. In all specifications CL clearly outperforms the EBEE approach. We also note that the RMSE of the model parameters in the EBEE is invariant to the cross-section size. This is a natural consequence of estimating L univariate models since univariate time-series used in the estimators are generated using the same parameters and shock distribution. The CL estimator improves slightly as the number of assets increases despite the increase in the number of parameters. This occurs since the number of component likelihoods is $O(L^2)$ and so there are some small reductions in parameter estimation error in larger cross-sections.

	BEKK						cDCC					
	2MLE		2MCLE		2MSCLE		2MLE		2MCLE		2MSCLE	
L	α	β	α	β	α	β	α	β	α	β	α	β
5	.0189	.9794	.0287	.9692	.0284	.9696	.0141	.9757	.0143	.9829	.0099	.9885
			(.0081)	(.0092)	(.0083)	(.0094)			(.0487)	(.0846)	(.0033)	(.0045)
10	.0125	.9865	.0281	.9699	.0272	.9709	.0063	.9895	.0107	.9881	.0093	.9886
			(.0055)	(.0063)	(.0054)	(.0062)			(.0012)	(.0016)	(.0016)	(.0018)
25	.0081	.9909	.0308	.9667	.0307	.9668	.0036	.9887	.0100	.9871	.0089	.9889
			(.0047)	(.0055)	(.0049)	(.0056)			(.0009)	(.0017)	(.0011)	(.0012)
50	.0056	.9926	.0319	.9645	.0316	.9647	.0022	.9867	.0101	.9856	.0092	.9869
			(.0046)	(.0056)	(.0047)	(.0057)			(.0008)	(.0018)	(.0010)	(.0019)
96	.0041	.9932	.0334	.9636	.0335	.9634	.0017	.9711	.0103	.9846	.0094	.9860
			(.0041)	(.0049)	(.0043)	(.0051)			(.0009)	(.0019)	(.0009)	(.0014)

Table 7: Parameter estimates for the scalar BEKK and cDCC models, from the full-dimensional likelihood method (2MLE), composite likelihood method based on all pairs (2MCLE) and composite likelihood method based on only the contiguous pairs (2MSCLE). The database is built from daily returns on 95 companies plus the index from the S&P100, from January 1997 until December 2006. For 2MCLE and 2MSCLE, numbers in parentheses give the asymptotic standard errors.

6 Empirical comparison

In this part, we analyse the performances of 2MLE, 2MCLE and 2MSCLE in several in- and out-of-sample empirical exercises. Our dataset consists of all companies which were at one point listed on the S&P 100, plus the index itself. The data cover the period from January 1, 1997, to December 31, 2006, and is taken from the CRSP database. This database has 124 companies although 29 have one or more periods of non-trading, (e.g. prior to IPO or subsequent to an acquisition). Selecting only the companies that have returns throughout the sample yields a panel of returns with $T = 2516$ and $L = 96$, including the market index. We set the first asset as the market index and arrange the other assets alphabetically by ticker (for stocks that changed tickers during the sample, the ticker on the first day of the sample was used). In Section 6.3 we will extend our analysis to S&P 500 using the same choice criteria as above, yielding a collection of 480 assets (including the index).

6.1 In-sample comparison

We start with a comparison of parameter estimates for the BEKK and cDCC models obtained by the ML and CL methods. The results are given in Table 7. The numbers in parentheses are the standard errors. The results for the composite likelihood estimators

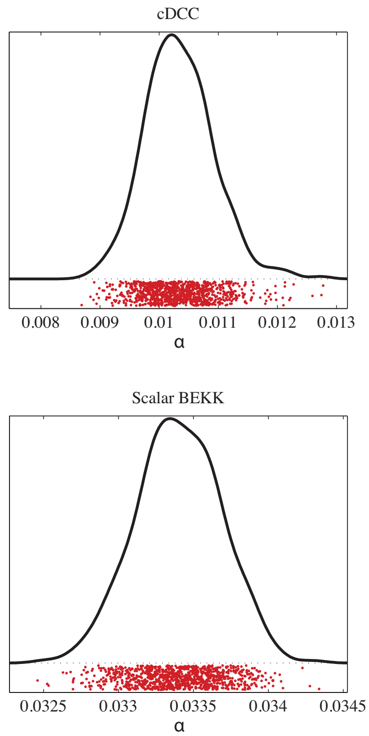

Figure 4: To investigate the sensitivity to random selection of pairs, we look at the density of the CL estimator based on $L-1$ distinct but randomly chosen pairs. The top row gives the density of estimators of the cDCC model, while the bottom row presents the same for the scalar BEKK model. For each subfigure, the lower portion provides a scatter plot of individual estimates, whereas the upper portion presents their sample distribution.

are reasonably stable with respect to L and they do not vary much as we move from using all pairs to a subset of them. Estimates from 2MLE are markedly different. The most important difference, of course, is that the CL method estimators are not subject to a bias as L grows, whereas 2MLE estimators yield extreme values for $\hat{\alpha}$ for large L — in this case close to being non-responsive to the data.

It is interesting to see how sensitive the contiguous pairs estimator is to the selection of the subset of pairs. To this end, we randomly select 1000 different sets of $L-1$ pairs and obtain parameter estimates for each selection. A scatter plot of these estimates, along with their sample distributions, are provided in Figure 4. It is clear that the estimates are hardly affected by the selection of pairs.

To examine the fit of the models, the conditional correlations with the market index are calculated for each of the 95 individual stocks from S&P 100, using estimates by the

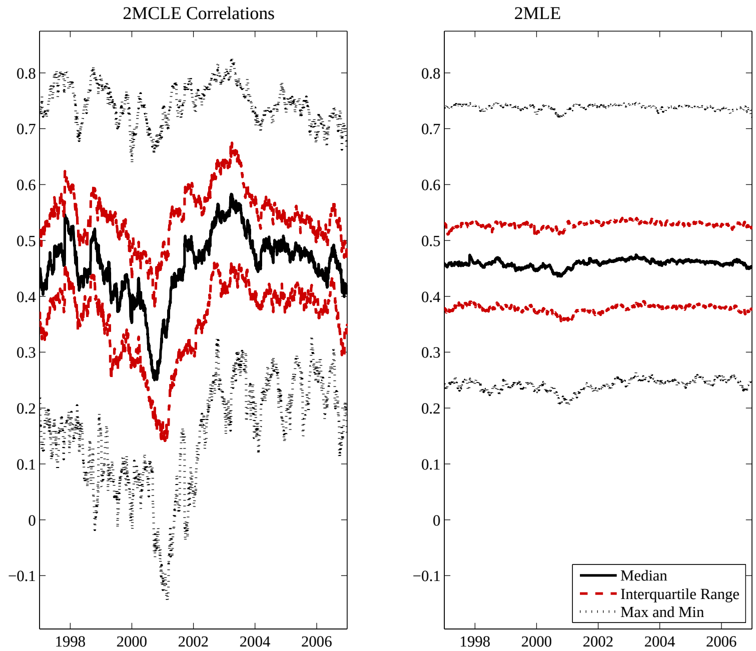

Figure 5: How do the correlations with the market change over time? Plot of the median, interquartile range, minimum, and maximum of the correlations of the 95 included S&P 100 components with the index return. For each day in the sample, these 95 correlations are obtained and sorted, to produce the daily quantiles. The left panel produces the results based on estimates by the CL method based on all pairs (2MCLE) whereas the right panel presents the same for the full-dimensional likelihood method (2MLE).

2MCLE and 2MLE methods. Figure 5 contains the median, inter-quartile range, and the maximum and minimum of these conditional correlations across the 95 stocks. The parameter estimates from the 2MCLE produce large, persistent shifts in conditional correlations with the market, including a marked decrease in the conditional correlations near the peak of the technology boom in 2001. The small estimated α for 2MLE produces conditional correlations which are nearly constant and exhibit little variation even at the height of the technology bubble in 2001.

6.2 Out of sample comparison of hedging performance

To examine the practical implications of using the different estimation methods at hand, we examine the hedging errors of a conditional CAPM model (with the S&P 100 index proxying for the market index). In particular, using one-step ahead forecasts, we compute the conditional time-varying market betas as $\widehat{\beta}_{it} = \widehat{h}_{it}^{1/2}\widehat{\rho}_{im,t}/\widehat{h}_{mt}^{1/2}$ where $h_{it} = Var(r_{it}|\mathcal{F}_{t-1})$, $h_{mt} = Var(r_{mt}|\mathcal{F}_{t-1})$, $\rho_{im,t} = Corr(r_{it}, r_{mt}|\mathcal{F}_{t-1})$ and $i = 1, 2, ..., L$. We use the same univariate models for h_{it} and h_{mt} throughout. This ensures that the differences in hedging errors are attributable to differences in correlation forecasts only, which in turn is solely attributable to the estimation method employed. The corresponding hedging errors are computed as $\widehat{\nu}_{i,t} = r_{i,t} - \widehat{\beta}_{i,t}r_{m,t}$. Here r_{it} is the return on the i^{th} asset and r_{mt} is the return on the market. This exercise is conducted using the first 75% of the sample (January 2, 1997 - July 1, 2002) as the "in-sample" period for parameter estimation, and the remaining part (July 2, 2002 - December 31, 2006) as the evaluation period.

We use the Giacomini and White (2006) (GW) test to examine the differences in hedging errors. The GW test is designed to compare forecasting methods, which incorporate such things as the forecasting model, sample period and, importantly for our purposes, the estimation method employed. Let $\widehat{\delta}_{it} = \left\{\widehat{\nu}_{it}\left(\widehat{\rho}_{it}^{2MCLE}\right)\right\}^2 - \left\{\widehat{\nu}_{it}\left(\widehat{\rho}_{it}^{2MLE}\right)\right\}^2$ be the difference in the squared hedging errors, where dependence on the estimation method is explicit. If neither estimator is superior in forecasting correlations, this difference should have 0 expectation. If the difference is significantly different from zero and negative, 2MCLE would be the preferred method, while significant positive results would indicate favor for 2MLE. The null of $H_0 : E[\widehat{\delta}_{it}] = 0$ is tested using the statistic $GW = \bar{\delta}_i/\sqrt{avar\left(\sqrt{T}\bar{\delta}_i\right)}$, where $\bar{\delta}_i$ is the average of $\hat{\delta}_{it}$ over the out-of-sample period. Under mild regularity conditions GW is asymptotically normal. See Giacomini and White (2006) for further details.

For a given comparison (e.g cDCC with 2MLE vs cDCC with 2MCLE), the test of equal hedging ability is conducted individually for each of the 95 assets. For each of these, there are three outcomes: the test favours the first method, or the second method, or it remains inconclusive. Results of these tests for all the comparisons considered here are produced in Table 8. Rows 2 and 6 provide the results for cDCC and scalar BEKK, respectively. In both cases, whenever the GW test is conclusive, it is clearly in favour of the composite likelihood method. In the case of BEKK, the picture is slightly sharper where only 1 in 95 cases is in favour of 2MLE.

We also compare cDCC 2MCLE and BEKK 2MCLE to other possible approaches to modelling multivariate volatility. First, we investigate the advantages of information pool-

Model A	Favours A	No Decision	Favours B	Model B
cDCC 2MCLE	24	63	8	cDCC 2MLE
cDCC 2MCLE	92	3	0	DECO
cDCC 2MCLE	18	68	9	Bivariate cDCC
cDCC 2MCLE	9	82	4	EWMA
BEKK 2MCLE	29	65	1	BEKK 2MLE
BEKK 2MCLE	50	44	1	Bivariate BEKK

Table 8: Which model and estimation strategy leads to smallest hedging errors? Giacomini-White test results for the null of equal out-of-sample hedging performance, at 5% level of significance. For any particular comparison, the test is conducted for each of the 95 assets. The test can favour model A, model B or be indecisive. The Table records the number of assets which fall in each of these three buckets.

ing: the CL approach is based on the assumption that all bivariate models have the same (α, β), which makes information pooling attractive. So, a question of interest is whether relaxation of this assumption leads to better hedging performance. We investigate this by comparing hedging based on a single pooled estimator against that based on different parameters for each pair. The results of this for cDCC and BEKK are reported on rows 4 and 7 of Table 8, respectively. For both models, our test results indicate that pooling delivers better performance. In the case of BEKK, the difference is substantial: assuming a common (α, β) delivers better results in 50 cases, against only one win for no-pooling.

In order to understand why the restrictive assumption of parameter homogeneity across pairs delivers better results, we look at the sample distribution of $(\hat{\alpha}_j, \hat{\beta}_j)$ for each pair j in Figure 6, for the cDCC model. There is a significant scatter. In addition, although in some cases $\hat{\alpha}_j + \hat{\beta}_j$ is quite small, in 22 cases this sum is at the unit boundary. Such unit root models (often called EWMA models) can perform poorly in terms of hedging. A simple and popular example is RiskMetrics, given by

$$H_t = \alpha r_{t-1} r_{t-1}' + (1 - \alpha) H_{t-1}, \tag{16}$$

where α is equal to 0.06 for daily and 0.03 for monthly returns.

Figure 7 shows four examples of estimated time varying correlations between a specific asset and the market, drawn for 4 pairs of returns we have chosen to reflect the variety we see in practice. The vertical dotted line indicates where we move from in-sample to out-of-sample. Top right shows a case where the estimated bivariate model and the fit from the

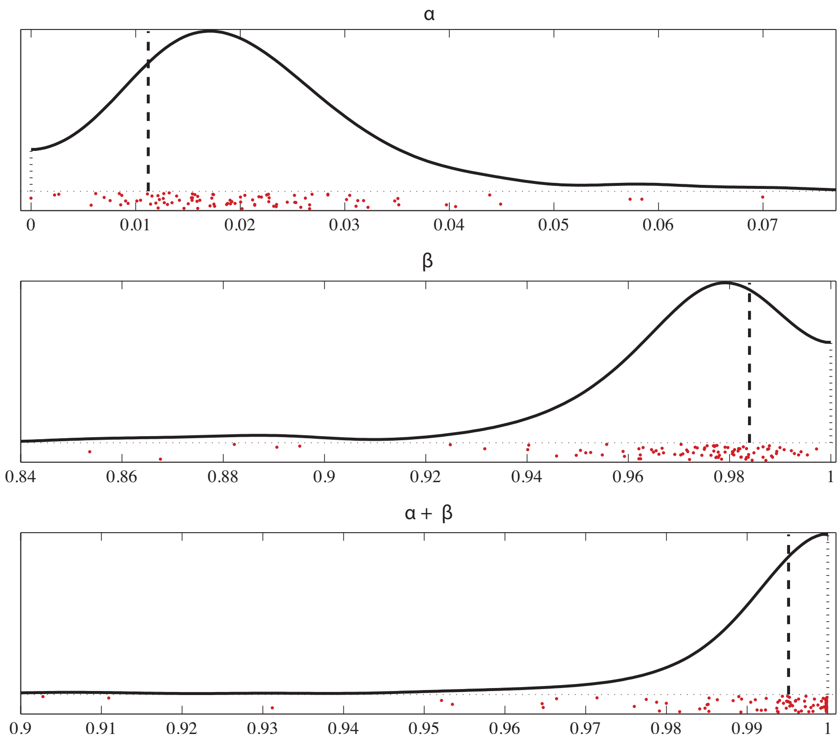

Figure 6: Should the data be pooled across pairs? Separately estimated α_j and β_j for each bivariate (cDCC) submodel for the beta-pair of the market and an individual asset. The broken vertical line marks the CL estimator obtained under the assumption of parameter homogeneity — which acts as a pooling device. For each subfigure, the lower portion presents the scatter plot of the estimates for each pair while the upper portion provides the sample distribution of these.

multivariate model are indistinguishable, both in- and out-of-sample. The top left shows a case where the fitted bivariate model has the correct dependence but the estimated α is relatively large, and so the fitted correlation is very noisy. The bottom left is the flip side of this, where the bivariate model estimates $\hat{\alpha} = 0$ and so produces a constant correlation. The bottom right is an example where the EWMA model in (16) is in effect imposed in the bivariate case, which produces a poor out-of-sample fit. Indeed, as the results in row 5 of Table 8 reveal, cDCC outperforms RiskMetrics in terms of out-of-sample hedging errors.

Finally, we consider the linear equicorrelation model of Engle and Kelly (2012), also known as DECO. This model has a similar structure to the DCC-type models, with each asset price process having its own ARCH model. However, this model assumes that asset returns have equicorrelation, in the sense that the time-changing correlation amongst assets

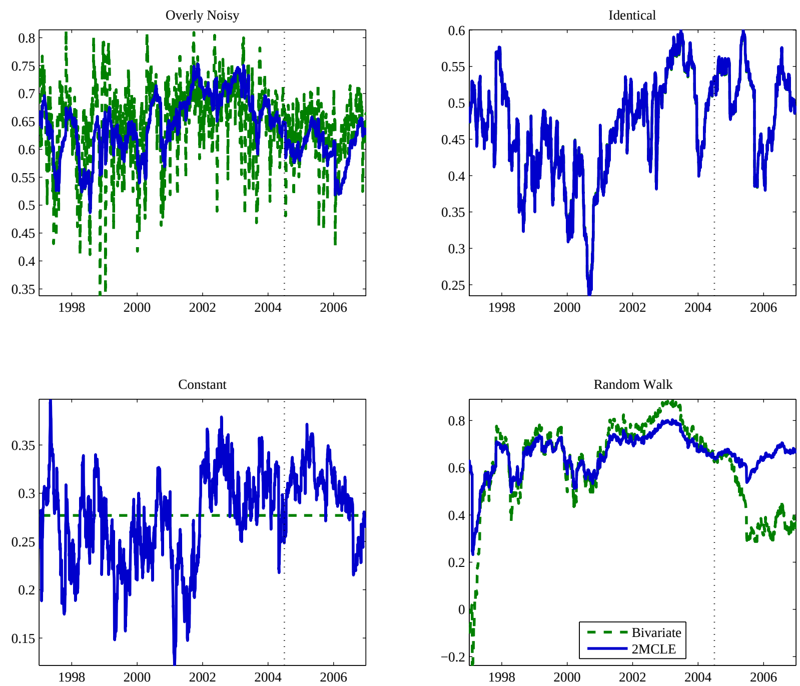

Figure 7: Comparison of estimated conditional correlations and out-of-sample projections for different selections of pairs. The comparison is between the high dimensional model (2MCLE) which assumes parameter homogeneity across pairs, and the bivariate model which allows for parameter heterogeneity. All results are based on the cDCC model. Top left suggests that the bivariate model is overly noisy. Top right presents a case where both approaches yield basically the same result. Bottom left estimates a constant correlation for the bivariate model, while the multivariate model is more responsive. Bottom right is a key example as we see it quite often. Here the bivariate model is basically estimated to be an EWMA, which has very poor out-of-sample fit.

is common across the cross-section of assets. In particular, the correlation matrix is given by $R_t = \rho_t \iota\iota' + (1 - \rho_t) I$, with $\rho_t = \omega + \gamma u_{t-1} + \beta\rho_{t-1}$, where u_{t-1} is new information about the correlation in the devolatilised r_{t-1}. A simple approach would be to take u_{t-1} as the cross-sectional MLE of the correlation based on this simple equicorrelation model. Row 3 of Table 8 compares the hedging performance of this method with the cDCC fit. We can see that cDCC is clearly uniformly superior.

We finish this part by noting that one can find other interesting out-of-sample comparisons. One example would be to focus on estimating the minimum variance portfolio, along the same lines as in Section 6.2 of Engle, Ledoit, and Wolf (2019). In the interest of space,

	BEKK						cDCC					
	2MLE		2MCLE		2MSCLE		2MLE		2MCLE		2MSCLE	
L	α	β	α	β	α	β	α	β	α	β	α	β
5	.0261	.9715	.0369 (.0057)	.9603 (.0065)	.0312 (.0053)	.9664 (.0061)	.0101	.9823	.0133 (.0041)	.9794 (.0081)	.0070 (.0033)	.9912 (.0038)
25	.0080	.9909	.0300 (.0062)	.9670 (.0075)	.0289 (.0055)	.9682 (.0067)	.0030	.9908	.0083 (.0015)	.9885 (.0031)	.0071 (.0011)	.9911 (.0016)
50	.0055	.9932	.0282 (.0051)	.9692 (.0062)	.0277 (.0049)	.9698 (.0059)	.0018	.9882	.0078 (.0010)	.9887 (.0021)	.0073 (.0010)	.9901 (.0019)
100	.0034	.9934	.0296 (.0046)	.9670 (.0057)	.0292 (.0045)	.9674 (.0056)	.0015	.9524	.0073 (.0007)	.9881 (.0015)	.0076 (.0010)	.9866 (.0028)
250	.0015	.9842	.0322 (.0049)	.9633 (.0064)	.0322 (.0048)	.9633 (.0063)	.0020	.5561	.0076 (.0007)	.9872 (.0015)	.0080 (.0016)	.9858 (.0039)
480	.0032	.5630	.0290 (.0041)	.9672 (.0054)	.0290 (.0040)	.9672 (.0053)	.0013	.2556	.0073 (.0007)	.9874 (.0016)	.0079 (.0008)	.9863 (.0020)

Table 9: Parameter estimates for the scalar BEKK and cDCC models, from the full-dimensional likelihood method (2MLE), composite likelihood method based on all pairs (2MCLE) and composite likelihood method based on only the contiguous pairs (2MSCLE). The database is built from daily returns on 480 companies of the S&P 500, from January 1997 until December 2006. For 2MCLE and 2MSCLE, numbers in parentheses give the asymptotic standard errors.

we leave the investigation of this and other relevant examples to future work.

6.3 Extending the empirical analysis

In this part we extend the analysis of Section 6.1 to up to $L = 480$. Our database consists of the returns of all equities in the S&P 500 that were continuously available from January 1, 1997 to December 31, 2006. This results in 480 unique assets, including the S&P 500 index, with 2516 observations for each. The data were extracted from CRSP and series were ordered alphabetically according to their ticker on the first day of the sample. Results are presented in Table 9.

Considering BEKK, 2MLE shows signs of bias as the cross-sectional dimension increases; indeed, for the two largest panel sizes the fitted volatilities would be virtually constant. When $L = 480$, β also shows a large downward bias. The CL estimates are very similar, all with $\hat{\alpha} \approx .03$, $\hat{\beta} \approx .96$, and the standard errors decline modestly as L increases. For large L the difference between the contiguous- and all-pairs estimators is negligible.

We next turn to cDCC. For this wider set of data the best performing volatility model was the GJR-GARCH(1,1), where $h_{it} = \omega_i + \delta_i r_{i,t-1}^2 + \gamma_i r_{i,t-1}^2 I_{[r_{i,t-1}<0]} + \kappa_i h_{i,t-1}$ for $i = 1, ..., L$. The 2MLE estimate of α exhibits strong bias as the sample size increases and for

L	2MLE	2MCLE	2MSCLE
5	24s	0.1s	0.2s
25	46s	2.1s	0.2s
50	2m 10s	10s	0.5s
100	1h 50m	39s	0.8s
250	15h 11m	4m 7s	1.6s
480	85h 33m	18m 6s	4.5s

Table 10: CPU time required to estimate a covariance tracking scalar BEKK on the assets of the S&P 500. All models were estimated on a 2.5GHz Intel Core 2 Quad.

$L > 250$ the β estimate is also badly affected. This contrasts with the estimates by 2MCLE and 2MSCLE, where $\hat{\alpha} \approx .008$ and $\hat{\alpha} + \hat{\beta} \approx .995$.

Finally, Table 10 presents the computation times for each method, for the specific case of scalar BEKK. The 2MLE method takes around 3.5 days on the $L = 480$ problem, while for $L = 25$ the time is quite modest being under a minute. This shows the impact of the $O(L^3)$ computational load. The composite methods are much more rapid than 2MLE, with 2MCLE still being quite fast for $L = 100$ - about 200 times faster than 2MLE. 2MSCLE is fast even when $L = 480$, just taking a small handful of seconds. This means it is around $68,000$ times faster than 2MLE in this vast dimensional case.

7 Conclusion

This paper has introduced a new method for estimating large dimensional time-varying covariance models, which is easy to implement, computationally fast, and free of the bias that troubles standard estimation approaches. Importantly, the proposed method works well even when the number of assets is much larger than the time dimension. We have investigated the theoretical and empirical properties of this approach in detail for BEKK and cDCC. Nevertheless, the proposed approach is generally applicable to other ARCH-type multivariate models.

Recent interest in high-dimensional multivariate volatility modelling has led to a flourishing literature (e.g. Hafner and Reznikova (2012), Bauwens, Braione, and Storti (2017) and Engle, Ledoit, and Wolf (2019)). It would be both interesting and timely to undertake a comprehensive comparison of alternative approaches, using detailed simulation studies and relevant empirical applications. We leave this idea to future work.

SUPPLEMENTARY MATERIAL

Supplementary Appendix: The Supplementary Appendix consists of four sections, A through D. Section A contains notation and general results, Sections B and C contain the proofs for the scalar- and non-scalar BEKK models, respectively, whereas Section D contains other additional material. (.pdf file)

References

Aielli, G. P. (2013). Consistent estimation of large scale dynamic conditional correlations. *Journal of Business and Economic Statistics 31*, 282–299.

Bauwens, L., M. Braione, and G. Storti (2017). A dynamic component model for forecasting high-dimensional realized covariance matrices. *Econometrics and Statistics 1*, 40–61.

Bauwens, L., S. Laurent, and J. V. K. Rombouts (2006). Multivariate GARCH models: A survey. *Journal of Applied Econometrics 21*, 79–109.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B 36*, 192–236.

Bollerslev, T. (1986). Generalised autoregressive conditional heteroskedasticity. *Journal of Econometrics 51*, 307–327.

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

Cox, D. R. and N. Reid (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika 91*, 729–737.

deLeon, A. R. (2005). Pairwise likelihood approach to grouped continuous model and its extension. *Statistics and Probability Letters 75*, 49–57.

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

Engle, R. F. (2009). *Anticipating Correlations*. Princeton University Press.

Engle, R. F. and B. Kelly (2012). Dynamic equicorrelation. *Journal of Business and Economic Statistics 30*, 212–228.

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

Engle, R. F., O. Ledoit, and M. Wolf (2019). Large dynamic covariance matrices. *Journal of Business & Economic Statistics 37*, 363–375.

Engle, R. F. and J. Mezrich (1996). GARCH for groups. *Risk 9*, 36–40.

Fermanian, J.-D. and H. Malongo (2017). On the stationarity of dynamic conditional correlation models. *Econometric Theory 33*, 636–663.

Fiorentini, G., E. Sentana, and N. Shephard (2004). Likelihood-based estimation of latent generalised ARCH structures. *Econometrica 72*, 1481–1517.

Francq, C., L. Horváth, and J. M. Zakoïan (2011). Merits and drawbacks of variance targeting in garch models. *Journal of Financial Econometrics 9*, 619–656.

Francq, C. and J.-M. Zakoïan (2016). Estimating multivariate volatility models equation by equation. *Journal of the Royal Statistical Society, Series B 78*, 613–635.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*, 1545–1578.

Hafner, C. M. and O. Reznikova (2012). On the estimation of dynamic conditional correlation models. *Computational Statistics & Data Analysis 56*, 3533–3545.

Kenne Pagui, E. C., A. Salvan, and N. Sartori (2015). On full efficiency of the maximum composite likelihood estimator. *Statistics & Probability Letters 97*, 120–124.

King, M., E. Sentana, and S. Wadhwani (1994). Volatility and links between national stock markets. *Econometrica 62*, 901–933.

Kuk, A. Y. C. and D. J. Nott (2000). A pairwise likelihood approach to analyzing correlated binary data. *Statistical and Probability Letters 47*, 329–335.

LeCessie, S. and J. C. van Houwelingen (1994). Logistic regression for correlated binary data. *Applied Statistics 43*, 95–108.

Lindsay, B. G. (1988). Composite likelihood methods. In N. U. Prabhu (Ed.), *Statistical Inference from Stochastic Processes*, pp. 221–239. Providence, RI: Amercian Mathematical Society.

Mardia, K. V., J. T. Kent, G. Hughes, and C. C. Taylor (2009). Maximum likelihood estimation using composite likelihoods for closed exponential families. *Biometrika 96*, 975–982.

Newey, W. K. and K. D. West (1987). A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

Pedersen, R. S. and A. Rahbek (2014). Multivariate variance targeting in the BEKK-GARCH model. *Econometrics Journal 17*, 24–55.

Rangel, J. G. and R. F. Engle (2012). The factor-spline-GARCH model for high and low-frequency correlations. *Journal of Business and Economic Statistics 30*, 109–124.

Silvennoinen, A. and T. Teräsvirta (2009). Multivariate GARCH models. In T. G. Andersen, R. A. Davis, J. P. Kreiss, and T. Mikosch (Eds.), *Handbook of Financial Time Series*, pp. 201–229. Springer-Verlag.

Varin, C., N. Reid, and D. Firth (2011). An overview of composite likelihood methods. *Statistica Sinica 21*, 5–42.

Varin, C. and P. Vidoni (2005). A note on composite likelihood inference and model selection. *Biometrika 92*, 519–528.

Wang, H., E. M. Iglesias, and J. M. Wooldridge (2013). Partial maximum likelihood estimation of spatial probit models. *Journal of Econometrics 172*, 77–89.

White, H. (2001). *Asymptotic Theory for Econometricians* (2 ed.). Orlando: Academic Press.