Fitting and testing vast dimensional time-varying covariance models

Robert F. Engle
Stern Business School, New York University,
44 West Fourth Street, New York, NY 10012-1126, USA
rengle@stern.nyu.edu

Neil Shephard
Oxford-Man Institute, University of Oxford,
Blue Board Court, Alfred Road, Oxford OX1 4ED, UK
&
Economics Department, University of Oxford
neil.shephard@economics.ox.ac.uk

Kevin Sheppard
Department of Economics, University of Oxford,
Manor Road Building, Manor Road, Oxford, OX1 3UQ, UK
&
Oxford-Man Institute, University of Oxford
kevin.sheppard@economics.ox.ac.uk

October 24, 2007

Abstract

Building models for high dimensional portfolios is important in risk management and asset allocation. Here we propose a novel way of estimating models of time-varying covariances that overcome some of the computational problems which have troubled existing methods when applied to 1,000s of assets. The theory of this new strategy is developed in some detail, allowing formal hypothesis testing to be carried out on these models. Simulations are used to explore the performance of this inference strategy while empirical examples are reported which show the strength of this method.

Keywords: ARCH models; composite likelihood; psuedo-likelihood; quasi-likelihood; time-varying covariances; correlation; DCC.

1 Introduction

The estimation of time-varying covariances between the returns on thousands of assets is a key input in modern risk management. Typically this is carried out by calculating the sample covariance matrix based on the last 100 days of data or through the RiskMetrics exponential smoother. When these covariances are allowed to vary through time using ARCH-type models the computational burden of likelihood based fitting is overwhelming in very large dimensions, even for very simple models. In this paper we introduce novel econometric methods which sidestep this issue allowing richly parameterised ARCH models to be fit in vast dimensions.

Early work on time-varying covariances in large dimensions was carried out by Bollerslev (1990) in his constant correlation model, where the volatilities of each asset were allowed to vary through time but the correlations were time invariant. This has been shown to be empirically problematic by, for example, Tse (2000) and Tsui and Yu (1999).

The only econometric work that we know of which allows correlations to change through time in vast dimensions is that on the DECO model of Engle and Kelly (2007) and the MacGyver estimation method of Engle (2007). Engle and Kelly (2007) assume that the correlation amongst assets changes through time but is constant amongst N assets. This cross-sectional invariance means they can compute the log-likelihood for their models in $O(N)$ calculations, which is highly convenient. However, this equicorrelation model is quite restrictive since the diversity of correlations is often the key to risk management. Our estimation methods an be implemented in $O(N)$ but allow a much richer model structure.

An alternative method was suggested by Engle (2007) where he fit many pairs of bivariate estimators, governed by simple dynamics, and then took a median of these estimators. This method is known as the MyGyver estimation strategy strategy, but it requires $O(N^2)$ calculations and formalising this method in order to conduct inference is difficult.

The structure of the paper is as follows. In Section 2 we outline the model we use and discuss various general ways of fitting time-varying covariance models. In Section 3 we discuss the core of the paper, where we average in different ways the results from many small dimensional models in order to carry out inference on a large dimensional model. This section has both theoretical and Monte Carlo comparisons of our methods with full Maximum Likelihood estimation and the MacGyver strategy. In Section 4 we discuss in particular the fitting of the dynamic conditional correlation (DCC) models introduced by Engle (2002) and studied in detailed by Engle and Sheppard (2001), and the cDCC model suggested by Aielli (2006). In Section 5 we provide some empirical illustrations of the methods and Section 6 concludes.

2 The model and existing approaches

2.1 The model

We assume we have a database r of log-returns

$$r_{jt}, \quad j = 1, 2, ..., N, \quad t = 1, 2, ..., T,$$

where we think of t as time and j as referring to the j-th asset return. In our analysis we will think of the number of assets available N as being very large, as will the time series dimension T. It is helpful to sometimes refer to the cross section

$$r_t = (r_{1t}, r_{2t}, ..., r_{Nt})',$$

and the time series

$$r_{(j)} = (r_{j1}, r_{j2}, ..., r_{jT})'.$$

A typical risk management model of r_t given the information available at time t is to assume:

Assumption 1

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

$$\mathrm{Cov}(r_t|\mathcal{F}_{t-1}) = H_t, \tag{2}$$

where $\mathcal{F}_{t-1}$ is the information available at time $t-1$ to predict r_t.

Thus r_t is a martingale difference sequence with a time-varying covariance matrix. As econometricians we will model how H_t depends upon the past data allowing it to be indexed by some parameters $\theta \in \Theta$. We intend to estimate θ. For simplicity in our examples we have always used single lags in the dynamics, the extension to multiple lags is trivial but hardly used in multivariate empirical work.

Example 1 *Covariance tracking and scalar dynamics. This puts*

$$H_t = (1 - \alpha - \beta)\,\Sigma + \alpha r_{t-1} r_{t-1}' + \beta H_{t-1}, \quad \alpha \geq 0, \quad \beta \geq 0, \quad \alpha + \beta < 1,$$

which is a special case of Engle and Kroner (1995). Typically this model is completed by setting $H_1 = \Sigma$. Hence in this model $\theta = \left(\psi', vech(\Sigma)'\right)'$, where $\psi = (\alpha, \beta)'$.

Example 2 *Nonstationary covariances with scalar dynamics:*

$$H_t = (1 - \beta)\, r_{t-1} r_{t-1}' + \beta H_{t-1}, \quad \beta \in [0, 1).$$

A simple case of this is Riskmetrics, which puts $\beta = 0.94$. Inference is usually made conditional on H_j for $j \leq 0$ where these matrices are set to some values determined by presample data.

Example 3 *NOT FINISHED YET, DONT READ. Variance Targeting BEKK.*

$$H_t = CC' + Ar_{t-1}r_{t-1}'A' + BH_{t-1}B' \tag{3}$$

where A and B are diagonal matrices. Since the BEKK family of models are closed to rotations, it is possible to rotate the returns by the long-run covariance, which is estimated using the usual moment estimator, $\bar{H}$ to produce a modified equation that can be estimated on the rotated returns,

$$\begin{aligned}\bar{H}^{-\frac{1}{2}}H_t\bar{H}^{-\frac{1}{2}} =& \bar{H}^{-\frac{1}{2}}CC'\bar{H}^{-\frac{1}{2}} + \bar{H}^{-\frac{1}{2}}A\bar{H}^{\frac{1}{2}}\bar{H}^{-\frac{1}{2}}r_{t-1}r_{t-1}'\bar{H}^{-\frac{1}{2}}\bar{H}^{\frac{1}{2}}A'\bar{H}^{-\frac{1}{2}} &(4)\\ &+ \bar{H}^{-\frac{1}{2}}B\bar{H}^{\frac{1}{2}}\bar{H}^{-\frac{1}{2}}H_{t-1}\bar{H}^{-\frac{1}{2}}\bar{H}^{\frac{1}{2}}B'\bar{H}^{\frac{1}{2}} &(5)\end{aligned}$$

$$\tilde{H}_t = \tilde{C}\tilde{C}' + \tilde{A}u_{t-1}u_{t-1}'\tilde{A}' + \tilde{B}\tilde{H}_{t-1}\tilde{B}' \tag{6}$$

Because $E[\tilde{H}_t] = I_N$ by construction, this model can be variance targeted,

$$\tilde{H}_t = \left(I_N - \tilde{A}\tilde{A}' - \tilde{B}\tilde{B}'\right) + \tilde{A}u_{t-1}u_{t-1}'\tilde{A}' + \tilde{B}\tilde{H}_{t-1}\tilde{B}'$$

Finally it can be noted that the complete log-likelihood is not needed for identification of the parameters, and thus a subset pseudo-likelihood using some pairs can be used to estimate the values of $\tilde{A}$ and $\tilde{B}$.

A standard inference method is to construct a series of martingale difference based moment constraints using the score of a standard Gaussian quasi-likelihood

$$\log L_Q(\theta; r) = \sum_{t=1}^{T} l_t^Q(\theta), \tag{7}$$

where

$$l_t^Q(\theta) = -\frac{1}{2}\log|H_t| - \frac{1}{2}r_t{}'H_t^{-1}r_t.$$

Maximising this quasi-likelihood (7) directly is challanging as

- the parameter space is typically large;
- non-linear constraints on the parameters have to be imposed to ensure conditional covariances remain positive definite during estimation;
- the inversion of H_t takes $O(N^3)$ computations.

2.2 Covariance tracking and two-stage estimation

Many modern models of time-varying covariances employ covariance tracking, such as the model highlighted in Example 1. For such classes of problems it is easy to simplify the optimisation problem using a two-stage estimation strategy. Within the context of Example 1 we can estimate

$$\theta = \left(\xi', \psi'\right)', \quad \xi = vech(\Sigma),$$

by:

1. Using the moment estimator

$$\widehat{\Sigma} = \frac{1}{n}\sum_{t=1}^{T} r_t r_t'.$$

We write $\widehat{\xi} = vech(\widehat{\Sigma})$.

2. Compute

$$\widehat{\psi} = \underset{\psi}{\operatorname{argmax}} \sum_{t=1}^{T} \log L_t^Q(\widehat{\xi}, \psi; r),$$

where we call

$$\sum_{t=1}^{T} \log L_t^Q(\widehat{\xi}, \psi; r)$$

the *mofile likelihood*[1].

The above strategy is $O(N^3)$ — the appropriate econometric theory for this estimator will be discussed in Section 3.2.2. For now we move on to proposing methods which overcome this $O(N^3)$ problem.

3 The main idea: averaging likelihoods

3.1 Many small dimensional models

For all $j \in \{1, 2, ..., N\}$, $k \in \{j+1, 2, ..., N\}$

$$\mathrm{E}(r_{jt}|\mathcal{F}_{t-1}) = 0, \quad \mathrm{Cov}(r_{jt}, r_{kt}|\mathcal{F}_{t-1}) = h_{jkt}. \tag{8}$$

Then a valid pseudo-likelihood can be constructed for θ can be constructed off this pair:

$$\log L_{jk}(\theta) = \sum_{t=1}^{T} l_{jkt}(\theta),$$

where

$$l_{jkt}(\theta) = -\frac{1}{2}\log \begin{vmatrix} h_{jjt} & h_{jkt} \\ h_{jkt} & h_{kkt} \end{vmatrix} - \frac{1}{2} \begin{pmatrix} r_{jt} \\ r_{kt} \end{pmatrix}' \begin{pmatrix} h_{jjt} & h_{jkt} \\ h_{jkt} & h_{kkt} \end{pmatrix}^{-1} \begin{pmatrix} r_{jt} \\ r_{kt} \end{pmatrix}.$$

[1] Although at first sight $\sum_{t=1}^{T} \log L_t^Q(\widehat{\xi}, \psi; r)$ looks like a profile (or concentrated) likelihood, it is not as $\widehat{\xi}$ is not a ML estimator but an attractive moment estimator. Hence we call it a moment based profile likelihood, or mofile likelihood for short. This means $\widehat{\psi}$ is a two-step estimator which is typically less efficient than the maximum likelihood estimator.

This psuedo-likelihood will have information about θ but more information can be obtained by carrying out the same operation on all available pairs

$$\log L_t^B(\theta) = \sum_{j>k}^{N} \log L_{jkt}(\theta).$$

Again this is a valid pseudo-likelihood and yields our prefered estimator of θ: the maximum paired psuedo-likelihood (MPLE) estimator

$$\widetilde{\theta} = \underset{\theta}{\operatorname{argmax}} \sum_{t=1}^{T} \sum_{j>k}^{N} \log L_{jkt}(\theta).$$

Remark. This method never requires the inversion of the full N by N covariance matrices H_t.

Remark. Many databases of returns have significant holes, where the asset was not traded or not recorded, and this is problematic for likelihood methods based on $l_t^Q(\theta)$. Here the solution is trivial, as we through time we only count contributions to the psuedo-likelihood from pairs which were actively traded at that time. This also deals with the problem of assets entering and leaving indexes, for we can estimate θ based solely on data covering periods when the asset was inside the index.

This type of marginal analysis has appeared before in the non-time series statistics literature. An early example is Besag (1974) in his analysis of spatial processes, more recently it was used by Fearnhead (2003) in bioinformatics, deLeon (2005) on grouped data, Kuk and Nott (2000) and LeCessie and van Houwelingen (1994) for correlated binary data. This type of objective function is sometimes call composite likelihood methods, following the term introduced by Lindsay (1988) and "subsetting methods". See Varin and Vidoni (2005). Cox and Reid (2003) discusses the asymptotics of this problem in the non-time series case. Section 7.3 will discuss the parallels this work brings to our problem.

3.2 Covariance tracking and MPLE

3.2.1 Estimation strategy

The use of covariance tracking means that we can again use a two-stage estimation procedure. All that changes is

2' Compute

$$\widehat{\psi} = \underset{\psi}{\operatorname{argmax}} \sum_{t=1}^{T} \log L_t^B(\widehat{\xi}, \psi; r).$$

The above strategy is $O(N^2)$, rather than the usual $O(N^3)$ which would have resulted if we had used the Gaussian log-likelihood.

3.2.2 Econometric theory

The following subsection discusses the econometric theory of this estimator and can be skipped on first reading if desired.

From an econometric theory viewpoint, this two-stage estimator is a Pearson (1894) method of moments estimator — stacking the scores for this problem

$$\sum_{t=1}^{T} \begin{pmatrix} \xi - \frac{1}{T} vech(r_t r_t') \\ \partial \log L_t^B(\xi, \psi; r)/\partial \psi \end{pmatrix} = \sum_{t=1}^{T} m(\theta; y_t | \mathcal{F}_{t-1}).$$

The asymptotic behaviour of this estimator can be derived using standard two-stage GMM theory (Newey and McFadden (1994)). The result is that

$$\sqrt{T}\left(\widehat{\theta} - \theta_0\right) \xrightarrow{d} N(0, \mathcal{I}^{-1}\mathcal{J}\mathcal{I}^{-1\prime}),$$

where

$$\mathcal{I} = p \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \frac{\partial m(\theta_0; y_t|\mathcal{F}_{t-1})}{\partial \theta'}, \quad \mathcal{J} = \lim_{T\to\infty} \mathrm{Cov}\left\{\sqrt{T}\frac{1}{T}\sum_{t=1}^{T} m(\theta_0; y_t|\mathcal{F}_{t-1})\right\}.$$

Particular interest is in

$$\sqrt{T}\left(\widehat{\psi} - \psi_0\right) \xrightarrow{d} N(0, \mathcal{I}^{-1}\mathcal{J}\mathcal{I}^{-1\prime})$$

$\mathcal{I}$ has a block structure that is relatively sparse. We write

$$\frac{1}{T}\sum_{t=1} T \frac{\partial m\left(\theta_0; y_t|\mathcal{F}_{t-1}\right)}{\partial \theta'} = \frac{1}{T}\sum_{t=1}^{T} \begin{pmatrix} I & 0 \\ \partial^2 L_t^B(\xi,\psi)/\left(\partial\psi\partial\xi'\right) & \partial^2 L_t^B(\xi,\psi)/\left(\partial\psi\partial\psi'\right) \end{pmatrix} \tag{9}$$

$$\xrightarrow{p} \begin{pmatrix} I & 0 \\ \mathcal{I}_{\psi\xi} & \mathcal{I}_{\psi\psi} \end{pmatrix} = \mathcal{I}. \tag{10}$$

This means that

$$\mathcal{I}^{-1} = \begin{pmatrix} I & 0 \\ -\mathcal{I}_{\psi\psi}^{-1}\mathcal{I}_{\psi\xi} & \mathcal{I}_{\psi\psi}^{-1} \end{pmatrix} = \begin{pmatrix} I & 0 \\ \mathcal{I}^{\psi\xi} & \mathcal{I}^{\psi\psi} \end{pmatrix}.$$

Particular interest is in making inference on ψ. A special case of the above analysis is

$$\sqrt{T}\left(\widehat{\psi} - \psi_0\right) \xrightarrow{d} N\left(0, \mathcal{I}^{\psi\xi}\mathcal{J}_{\xi\psi}\mathcal{I}^{\psi\psi} + \mathcal{I}^{\psi\psi}\mathcal{J}_{\psi\psi}\mathcal{I}^{\psi\psi}\right).$$

The term $\mathcal{I}^{\psi\psi}\mathcal{J}_{\psi\psi}\mathcal{I}^{\psi\psi}$ is relatively easy to calculate due to its small dimension. The matrix $\mathcal{J}_{\xi\psi}$ is harder, but actually it is not needed for we actually have to work with $\mathcal{I}^{\psi\xi}\mathcal{J}_{\xi\psi}\mathcal{I}^{\psi\psi}$. If we write

$$m(\theta_0; y_t|\mathcal{F}_{t-1}) = \begin{Bmatrix} m_\xi(\theta_0; y_t|\mathcal{F}_{t-1}) \\ m_\psi(\theta_0; y_t|\mathcal{F}_{t-1}) \end{Bmatrix},$$

then the required matrix is the expected covariance between

$$\mathcal{I}^{\psi\xi} m_\xi(\theta_0; y_t|\mathcal{F}_{t-1}) \quad \text{and} \quad m_\psi(\theta_0; y_t|\mathcal{F}_{t-1})\mathcal{I}^{\psi\psi},$$

which are small dimensional. Of course computing this will be cumbersome however.

3.2.3 Simulation based inference: warp-speed bootstrap

An alternative is to use a bootstrap. Because the returns are generally dependant a moving block bootstrap or the stationary bootstrap (Künsch (1989) and Politis, Romano, and Wolf (1999)) must be used.

1. Using a vector time-series appropriate with the lag length chosen correctly (see Politis and White??), conduct a nonparametric bootstrap of the vector return series. It is crucial at the stage that the returns be sampled using time-series blocks of vectors to avoid breaking the cross-sectional dependance.

2. Using the re-sampled data, reestimate ψ as $\widehat{\psi}^{(b)}$, where (b) tracks the bootstrap iteration.

Confidence intervals and inference for parameters can be directly constructed from $\left\{\widehat{\psi}^{(b)}\right\}$.

3.3 Not every pair

Each $l_{jkt}(\lambda, \phi)$ is a valid contribution to the pseudo-likelihood and can contribute to learning about ϕ. So far we have calculated the "total pseudo-likelihood" over all possible pairs of observations

$$\log L_t^B = \sum_{j>k}^{N} \log L_{jkt},$$

but it also attractive to sum over just a subset to form the "subset pseudo-likelihood"

$$\log L_t^{\widetilde{B}} = \sum_{j=1}^{N^*} \log L_{J_j,K_j,t}.$$

Here, without replacement,

$$\{J_j, K_j\} \in \{j = 1, 2, ..., N; k = j + 1, k + 2, ..., N\}.$$

By taking only $O(N)$ pairs this method potentially has the advantage of being computationally fast, indeed the entire estimation method would be simply $O(N)$. It is tempting to randomly select N^* pairs and make inference conditional on the selected pairs as the selection is strongly exogenous. A theoretical analysis of this setup is provided in Appendix 8, while the Monte Carlo performance of this estimator is given in Table 5 with this estimator being denoted MSLE. It shows the efficiency loss compared to computing all possible pairs is extremely modest when N is moderate.

The idea of creating psuedo-likelihoods based on pairs obviously generalises to many triples or even higher dimensional log-likelihoods. We have not explored this here, but clearly there should be some efficiency gains in carrying this out.

4 The method applied to dynamic correlations

4.1 Model structures

We will have a particular interest in so called dynamic conditional correlation models — for these models allow for richer volatility dynamics than the models given in Examples 1 and 2. Here we will discuss them in some detail.

Without any loss we can always write

$$H_t = D_t R_t D_t,$$

where

$$D_t = diag(\sqrt{h_{1t}}, ..., \sqrt{h_{Nt}}), \quad R_t = \left\{\rho_{jkt}\right\},$$

and

$$h_{jt} = \mathrm{Var}(r_{jt}|\mathcal{F}_{t-1}), \quad \rho_{jkt} = \mathrm{Cor}(r_{jt}, r_{kt}|\mathcal{F}_{t-1}).$$

The dynamic conditional correlation models are based on the following crucial assumption. The parameters

$$\theta = \left(\lambda', \phi'\right)' \in \Theta, \quad \lambda = \left(\lambda'_{(1)}, \lambda'_{(2)}, ..., \lambda'_{(N)}\right)',$$

have the property

$$\Theta = \left(\bigcup_{j=1}^{n} \Lambda_{(j)}\right) \cup \Phi,$$

and

$$\lambda_{(j)} \in \Lambda_{(j)}, \quad j = 1, 2..., N; \quad \phi \in \Phi.$$

The $\lambda_{(j)}$ solely influences the conditional variances h_{jt} of the j-th asset and ϕ solely influences the time-varying correlations R_t.

Example 4 *Dynamic conditional correlation (DCC) model (Engle (2002) and Engle and Sheppard (2001)). For $j = 1, 2, ..., N$ let*

$$h_{jt} = \pi_j^2(1 - \alpha_j - \beta_j) + \alpha_j r_{jt-1}^2 + \beta_j h_{jt-1}, \quad \pi_j^2 \geq 0, \quad \alpha_j \geq 0, \quad \beta_j \geq 0, \quad \alpha_j + \beta_j < 1.$$

The parameters are, for each asset, $\lambda_{(j)} = \left(\pi_j^2, \alpha_j, \beta_j\right)'$. Calculate the "devolatilised returns" $s_t = (s_{1t}, ..., s_{Nt})'$ where

$$s_{jt} = \frac{r_{jt}}{\sqrt{h_{jt}}}.$$

Define another set of parameters $\omega = (\gamma, \delta)$. *Then we have*

$$Q_t = \Psi(1-\gamma-\delta) + \gamma s_{t-1} s_{t-1}' + \delta Q_{t-1}, \quad \gamma \geq 0, \quad \delta \geq 0, \quad \gamma + \delta < 1$$

$$\rho_{jkt} = \frac{q_{jkt}}{\sqrt{q_{jjt} q_{kkt}}}. \tag{11}$$

Typically we assume Ψ *is positive semidefinite with ones on its leading diagonal.*

Remark. The assumption that h_{jt} depends solely on its past squared returns can be relaxed to allow for leverage effects (e.g. through the threshold or GJR ARCH models Glosten, Jagannathan, and Runkle (1993)) without changing the principle. In general this structure is quite restrictive since it has assumed that the conditional volatility of asset j is not effected by the past of other assets. It maybe useful to include effects such as the past average volatility of other series or the squared market return to generalise this structure and then test for the significance of these effects.

Example 5 *cDCC model (Aielli (2006)). This is the same as the DCC except that the "devolatilisation" is carried out as*

$$s_{jt}^* = \frac{r_{jt}\sqrt{q_{jjt}}}{\sqrt{h_{jt}}},$$

while the structure of

$$Q_t = \Psi(1-\gamma-\delta) + \gamma s_{t-1}^* s_{t-1}^{*\prime} + \delta Q_{t-1},$$

remains the same. The virtue of this setup is that $\mathrm{E}(s_t^* s_t^{*\prime} | \mathcal{F}_{t-1}) = Q_{jkt}$, *which means the recursion in* Q *has a martingale difference representation*

$$Q_t = \Psi(1-\gamma-\delta) + \gamma\left\{s_{t-1}^* s_{t-1}^{*\prime} - \mathrm{E}\left(s_{t-1}^* s_{t-1}^{*\prime} | \mathcal{F}_{t-2}\right)\right\} + (\gamma+\delta) Q_{t-1},$$

which implies $\frac{1}{T}\sum_{t=1}^{T} s_t^* s_t^{*\prime} \xrightarrow{p} \Psi$.

Remark. Changing the devolatilisation in this way is rather minor as we would expect q_{jjt} to be very close to one, however it makes the theoretical analysis and computational implementation of the model much easier.

4.2 Existing two-stage approach

Writing the devolatilised returns

$$\begin{aligned} s_t(\lambda) &= D_t^{-1} r_t, \\ H_t^{-1} &= D_t^{-1} R_t^{-1} D_t^{-1} = D_t^{-2} + D_t^{-1}\left(R_t^{-1} - I\right) D_t^{-1}, \end{aligned}$$

we can express

$$\begin{aligned} l_t^Q &= \left\{-\frac{1}{2}\log\left|D_t^2\right| - \frac{1}{2}r_t' D_t^{-2} r_t\right\} + \left\{-\frac{1}{2}\log|R_t| - \frac{1}{2}s_t(\lambda)' R_t^{-1} s_t(\lambda)\right\} + s_t(\lambda)' s_t(\lambda) \\ &= \left\{\sum_{j=1}^{N} l_t^{A_j}(\lambda_{(j)})\right\} + l_t^{\widetilde{Q}}(\lambda,\phi) + C_t(\lambda). \end{aligned}$$

Here

$$\begin{aligned} l_t^{A_j}(\lambda_{(j)}) &= -\frac{1}{2}\log h_{jt} - \frac{1}{2}r_{jt}^2/h_{jt} \\ l_t^{\widetilde{Q}}(\lambda,\phi) &= -\frac{1}{2}\log|R_t| - \frac{1}{2}s_t(\lambda)' R_t^{-1} s_t(\lambda) \\ C_t(\lambda) &= s_t(\lambda)' s_t(\lambda). \end{aligned} \tag{12}$$

We can think of $l_t^{A_j}(\lambda_{(j)})$ as a Gaussian quasi-likelihood (e.g. Bollerslev and Wooldridge (1992)). Likewise $l_t^{\widetilde{Q}}(\lambda,\phi)$ is the Gaussian quasi-likelihood from running a multivariate time-varying correlation model on some vector of returns which we have tried to devolatilise.

The last term $C_t(\lambda)$ does not depend upon ϕ. It reflects the fact that the ARCH models for individual assets are not independent of one another in this multivariate setting and so exploiting this information could improve the efficiency of the estimation procedure compared to an asset by asset estimation method. This is exactly like running individual regressions rather than a joint regression in a seemingly unrelated regression model (see Zellner (1962)).

Engle and Sheppard (2001) develop a two stage quasi-likelihood estimation strategy to avoid the task of maximising (7). Their approach is to ignore the information in $C_t(\lambda)$ and maximise instead

$$\sum_{j=1}^{N}\left\{l_t^{A_j}(\lambda_{(j)}) + l_t^{\widetilde{Q}}(\lambda,\phi)\right\}. \tag{13}$$

This is inefficient but still yields valid martingale difference based moment constraints and so typically consistent estimators. It has the virtue that the optimisation can be carried out in two steps.

1. Compute

$$\widehat{\lambda}_{(j)} = \underset{\lambda_{(j)}}{\operatorname{argmax}} \log L_{A_j}(\lambda_{(j)}; r),$$

where

$$\log L_{A_j}(\lambda_{(j)}; r) = \sum_{t=1}^{T} l_t^{A_j}(\lambda_{(j)}).$$

2. Compute

$$\widehat{\phi} = \underset{\phi}{\operatorname{argmax}} \log L_{\widetilde{Q}}(\widehat{\lambda}, \phi; r),$$

where

$$\log L_{\widetilde{Q}}(\lambda, \phi; r) = \sum_{t=1}^{T} l_t^{\widetilde{Q}}(\lambda, \phi).$$

The first stage separately fits an ARCH-type model to each univariate return sequence. The second stage treats λ as known at $\widehat{\lambda}$ and solely maximises over ϕ. Of course it yields estimators which differ from those resulting in the maximisation of (12) but taken together it does maximise (13).

5 Fitting dynamic conditional correlation models

5.1 Block structure

Recall from Example 4 the cDCC model. Engle (2002) and Engle and Sheppard (2001) advocate the use of a blocking strategy to estimate these types of model and here we slightly adapt it to the use of likelihood or pseudo-likelihood methods. It also draws on the insights of Aielli (2006) on his cDCC model. It will be convenient to write

$$\tau_j = \left(\alpha_j', \beta_j'\right)', \quad \psi = \operatorname{vecl}(\Psi), \quad \text{and} \quad \xi = (\gamma, \delta)'.$$

Here $\operatorname{vecl}(X)$ takes the lower triangular elements of the matrix X, ignoring the leading diagonal. The two blocks are as follows:

1. For $j = 1, 2, ..., N$ compute

$$\widehat{\pi}_j^2 = \frac{1}{T} \sum_{t=1}^{T} r_{jt}^2,$$

$$\tau_j = \underset{\tau_j}{\operatorname{argmax}} \log L_{A_j}(\widehat{\pi}_j^2, \tau_j; r).$$

2. Use a flip-flop algorithm to simulatenously solve the moment constraints

$$T^{-1} \sum_{t=1}^{T} \left(\psi - \operatorname{vecl}\left(s_t s_t'\right)\right) = 0,$$

and maximising the mofile objective function

$$W(\widehat{\psi}, \xi; r),$$

where W could be the log-likelihood, a total pseudo-likelihood or a subset pseudo-likelihood. This flip-flop algorithm has three steps.

(a) Given ξ compute

$$q_{j,j,t} = (1 - \gamma - \delta) + \gamma s_{j,t-1}^{*2} + \delta q_{j,j,t-1},$$

where $s_{jt}^* = r_{jt}\sqrt{q_{j,j,t}}/\sqrt{h_{jt}}$ with $q_{j,j,0} = 1$.

(b) Calculate a moment based estimator (note only a subset of them is need if the pseudo-likelihood is being used)

$$\widehat{\Psi} = \frac{1}{T}\sum_{t=1}^{T} s_t^* s_t^{*\prime}.$$

(c) Optimise

$$\hat{\xi} = \underset{\xi}{\text{argmax}} W(\widehat{\psi}, \xi; s).$$

(d) Return to 2a until $\hat{\xi}$ has converged.

In the DCC model Engle (2002) and Engle and Sheppard (2001) advocate using the same blocking structure but without step 2a and using $s_{jt} = r_{jt}/\sqrt{h_{jt}}$ rather than s_{jt}^*. This means stage 2 does not need to be iterated, which saves considerably computationally. However, it is hard to formally establish that the estimator of this model has good statistical properties.

Remark. The use of moment estimators to estimate π_j^2 is often called variance tracking in the literature and is discussed by Engle and Mezrich (1996). It can be thought of as replacing one element of the score vector for the quasi-likelihood by a moment based estimator

$$\frac{1}{T}\sum_{t=1}^{T} \mathrm{E}\left(\pi_j^2 - r_{jt}^2\right) = 0.$$

The use of the same type of estimator on the correlation matrix was advocated by Engle and Sheppard (2001), but this is somewhat problematic for it is easy to see that

$$\frac{1}{T}\sum_{t=1}^{T} \mathrm{E}\left\{\mathrm{vecl}(\Psi) - \mathrm{vecl}(s_t s_t')\right\} \neq 0,$$

due to the presence of the transform (11). This leads to an inconsistent estimator. It is known though that the impact of this is very small (see Engle and Sheppard (2001)). In the cDCC approach this problem disappears for then

$$\frac{1}{T}\sum_{t=1}^{T} \mathrm{E}\left\{\mathrm{vecl}(\Psi) - \mathrm{vecl}(s_t^* s_t^{*\prime})\right\} = 0.$$

This point was made well by Aielli (2006) as his motivation for the cDCC adjustment.

Remark. The calculation of R_t^* is an $O(N^2)$ operation, but inference on its dynamics can be carried out in $O(N)$, $O(N^2)$ or $O(N^3)$ calculations.

The asymptotic theory for this type of estimator is derived in Appendix 9.

5.2 Engle's MacGyver method

Engle (2007) proposed a new method for estimating large dimensional models. He called it the MacGyver strategy. Again this is based on pairs. But instead of averaging the log-likelihoods of pairs of observations, the log-likelihoods were seperately maximised and then the resulting estimators were robustly averaged using medians. This overcomes the difficulty of inverting H, but has the difficulty that it is not clear that the pooled estimators should have equal weight nor what are the asymptotic properties of the resulting robust average.

Engle's MacGyver method has some similarities, but is distinct, with the Ledoit, Santa-Clara, and Wolf (2003) flexible multivariate GARCH estimation procedure which also fits models to many pairs of observations. The distinctive feature is that Engle's approach is based on the devolatilised series, rather than the original returns, and is focused entirely on estimating a small number of DCC parameters.

5.3 Monte Carlo based inference: warp-speed bootstrap

An alternative is to use a bootstrap. Because the returns are generally dependant in their squares and cross-products a moving block bootstrap (CITATION) or the stationary bootstrap must be used.

1. Using a vector time-series appropriate with the lag length chosen correctly (see Politis and White??), conduct a nonparametric bootstrap of the return series and estimate univariate volatility models for each asset. It is crucial at the stage that the returns be sampled using time-series blocks of vectors to avoid breaking the cross-sectional dependance.

2. Using the re-sampled data and the initial estimates of $\hat{\lambda}_j^{(b)}$, where (b) tracks the bootstrap iteration, estimate the MSLE or MPLE.

Confidence intervals and inference for parameters can be directly constructed from $\left\{\hat{\xi^{(b)}}\right\}$.

6 Monte Carlo

6.1 Simulation design

Here we explore the effectiveness of the following:

- likelihood based estimator;
- pseudo-likelihood based estimator;
- subset pseudo-likelihood estimators;

- Engle's MacGyver estimator;

A small Monte Carlo study based on 1,000 replications has been conducted assuming away the ARCH effects by setting throughout $\sigma_{jt}^2 = 1$ and not estimating them. Throughout we used $T = 2,000$ and the returns were was simulated according to a DCC model given in Example 4 three choices of temporal dependence in the Q process

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 0.02 \\ 0.97 \end{pmatrix}, \quad \begin{pmatrix} 0.05 \\ 0.93 \end{pmatrix}, \quad \text{or} \quad \begin{pmatrix} 0.10 \\ 0.87 \end{pmatrix}.$$

The intercept Ψ was chosen to be the unconditional correlations of a set of N observations from a cross-sectional $AR(2)$ of the form

$$y_j = 1.2y_{j-1} + .7y_{j-2} + \nu_j, \quad \nu_j \overset{i.i.d.}{\sim} N(0,1). \tag{14}$$

Example 6 *When* $N = 5$

$$\Psi = \begin{pmatrix} 1.00 & 0.71 & 0.15 & -0.32 & -0.48 \\ 0.71 & 1.00 & 0.71 & 0.15 & -0.32 \\ 0.15 & 0.71 & 1.00 & 0.71 & 0.15 \\ -0.32 & 0.15 & 0.71 & 1.00 & 0.71 \\ -0.48 & -0.32 & 0.15 & 0.71 & 1.00 \end{pmatrix}.$$

6.2 DCC Estimation

The DCC estimation differs in just Step 2c where we use a variety of objective functions W to maximise. They all have the structure

$$W = \sum_{t=1}^{T} W_t,$$

where

likelihood	$W_t = -\frac{1}{2}\log\lvert R_t\rvert - \frac{1}{2}s_t' R_t^{-1} s_t$
pseudo-likelihood	$W_t = \sum_{j>k}^{N} \log L_{j,k,t}$
subset pseudo-likelihood	$W_t = \sum_{j=1}^{N-1} \log L_{j,j+1,t}$

where

$$l_{j,k,t}(\lambda,\phi) = -\frac{1}{2}\log\left(1-\rho_{jkt}^2\right) - \frac{s_{j,t}^2 + s_{k,t}^2 - 2\rho_{jkt}s_{j,t}s_{k,t}}{2\left(1-\rho_{jkt}^2\right)}.$$

Engle's MacGyver estimator for these models performs $N(N-1)/2$ ML estimations of the bivariate DCC model and then computes the median of the resulting estimators.

6.3 Particulars to this run

- The intercept parameters were *not* estimated. Instead population values were used
- The regular DCC was used, not the cDCC. *Note: Need more on the cDCC probably and it can be used with any of these estimation methods. Replicate the results for cDCC to be placed in an appendix.*
- $T = 1000$ in all runs, N is one of $\{3, 5, 10, 25, 50, 75\}$
- A parameters were estimated using a constraint that $0 \leq \alpha \leq .9998$, $0 \leq \beta \leq .9998$, $\alpha + \beta < .9998$.

Tables 1, 2 and 3 contains the bias, standard deviation and root mean square error of the estimates over the $1,000$ runs. In all runs the bias is negligible relative to the standard deviation; as a result, the RMSE is essentially equal to the standard deviation of the parameters. Finally, Table 4 contains the average run times for each of the four methods across all runs of that method ($3 \times 1,000$ each) for a fixed N.

- The N^2-pseudolikelihood estimator has better RMSE for all cross-section sizes and parameter configurations.
- The FFMLE appears to be approximately N-consistent, as the standard deviations of the $N = 75$ case is about 10 times smaller those of the $N = 3$ case. This is likely due to the $O(N^2)$ correlations.
- The gains from increasing the cross section in the other estimators appear to be approximately $\sqrt{N}$, although calling them $\sqrt{N}$-consistent doesn't seem wise.
- The run time for the N^2-pseudo-likelihood estimator is probably about $4\times$ higher than it *could be*.

Remark. The maximum mofile likelihood (MMLE) method seems to develop a significant bias in estimating α as N increases and increases as α increases.

Remark. An interesting feature is that our subset pseudo-likelihood based inference procedure is both much faster to compute and, in our Monte Carlo analysis, more precise than the conventional flip-flop method at estimating α and less precise at estimating β. The improvement for estimating α, because of the fall in the bias, is likely to be due to the approximations used in the flip-flop, which are removed by the pair based procedures.

	Bias											
	MMLE		MSLE		MPLE		Engle		MMLE	MSLE	MPLE	Engle
N	γ	δ	γ	δ	γ	δ	γ	δ	$\delta+\gamma$	$\delta+\gamma$	$\delta+\gamma$	$\delta+\gamma$
$\gamma = .02, \delta = .97$												
3	.000	-.004	.001	-.009	.001	-.009	.001	-.011	-.004	-.008	-.008	-.010
10	-.000	-.003	-.000	-.004	-.000	-.005	.000	-.008	-.003	-.004	-.005	-.008
50	-.002	-.003	-.000	-.003	-.000	-.005	.000	-.008	-.005	-.003	-.005	-.008
100	-.004	-.004	-.000	-.003	-.000	-.005	.000	-.008	-.008	-.003	-.005	-.008
$\gamma = .05, \delta = .93$												
3	-.001	-.002	-.001	-.004	-.001	-.004	-.001	-.006	-.002	-.005	-.005	-.007
10	-.002	-.000	-.001	-.002	-.001	-.003	-.001	-.005	-.002	-.003	-.004	-.006
50	-.007	.002	-.001	-.002	-.001	-.003	-.001	-.006	-.005	-.002	-.004	-.006
100	-.010	-.002	-.000	-.002	-.001	-.003	-.001	-.006	-.012	-.002	-.004	-.007
$\gamma = .10, \delta = .87$												
3	-.003	.001	-.002	-.003	-.002	-.003	-.001	-.004	-.002	-.004	-.004	-.006
10	-.007	.005	-.003	.001	-.003	.000	-.003	-.003	-.002	-.002	-.003	-.005
50	-.017	.007	-.004	.002	-.004	.000	-.003	-.003	-.010	-.002	-.003	-.006
100	-.019	-.003	-.002	.000	-.003	-.001	-.002	-.004	-.022	-.002	-.003	-.006

Table 1: *Root-mean-square error results from a simulation study for the dynamic correlation estimators of the DCC model. We only report the estimates of γ and δ and their sum. The estimators include the subset psuedo-likelihood (MSLE), the full pseudo-likelihood (MPLE), Engle's MacGyver strategy (Engle) and the mofile likelihood (MMLE) estimator. All results based on $1,000$ replications and $T = 2,000$.*

	Standard Deviation											
	MMLE		MSLE		MPLE		Engle		MMLE	MSLE	MPLE	Engle
N	γ	δ	γ	δ	γ	δ	γ	δ	$\delta+\gamma$	$\delta+\gamma$	$\delta+\gamma$	$\delta+\gamma$
					$\gamma = .02, \delta = .97$							
3	.004	.009	.007	.021	.007	.021	.008	.024	.006	.018	.018	.020
10	.001	.002	.003	.005	.003	.005	.003	.006	.002	.004	.004	.005
50	.000	.001	.001	.002	.001	.002	.001	.002	.001	.002	.001	.002
100	.000	.000	.001	.002	.001	.001	.001	.001	.000	.001	.001	.001
					$\gamma = .05, \delta = .93$							
3	.006	.010	.009	.020	.009	.020	.011	.019	.006	.015	.015	.013
10	.002	.003	.004	.007	.005	.007	.005	.008	.002	.004	.004	.005
50	.001	.001	.002	.003	.002	.003	.002	.003	.001	.002	.002	.002
100	.000	.001	.002	.002	.002	.002	.002	.003	.001	.002	.001	.002
					$\gamma = .10, \delta = .87$							
3	.008	.012	.013	.019	.013	.019	.015	.022	.007	.011	.011	.013
10	.003	.005	.007	.009	.007	.009	.007	.010	.003	.005	.005	.006
50	.002	.002	.003	.005	.004	.005	.004	.005	.001	.003	.003	.003
100	.001	.002	.003	.004	.003	.004	.003	.004	.001	.002	.002	.003

Table 2: *Root-mean-square error results from a simulation study for the dynamic correlation estimators of the DCC model. We only report the estimates of γ and δ and their sum. The estimators include the subset psuedo-likelihood (MSLE), the full pseudo-likelihood (MPLE), Engle's MacGyver strategy (Engle) and the mofile likelihood (MMLE) estimator. All results based on $1,000$ replications and $T = 2,000$.*

	RSME											
	MMLE		MSLE		MPLE		Engle		MMLE	MSLE	MPLE	Engle
N	γ	δ	γ	δ	γ	δ	γ	δ	$\delta+\gamma$	$\delta+\gamma$	$\delta+\gamma$	$\delta+\gamma$
$\gamma = .02, \delta = .97$												
3	.004	.010	.007	.023	.007	.023	.008	.026	.007	.020	.020	.023
10	.001	.004	.003	.007	.003	.007	.003	.010	.004	.006	.006	.009
50	.002	.003	.001	.004	.001	.005	.001	.008	.005	.004	.005	.008
100	.004	.004	.001	.004	.001	.005	.001	.008	.008	.004	.005	.008
$\gamma = .05, \delta = .93$												
3	.006	.010	.009	.021	.009	.021	.011	.020	.007	.016	.016	.015
10	.003	.003	.005	.007	.005	.008	.005	.010	.003	.005	.006	.008
50	.007	.002	.002	.004	.002	.004	.002	.006	.005	.003	.004	.007
100	.010	.002	.002	.003	.002	.004	.002	.006	.012	.003	.004	.007
$\gamma = .10, \delta = .87$												
3	.009	.012	.013	.019	.013	.019	.015	.022	.007	.011	.011	.014
10	.008	.007	.008	.009	.008	.009	.008	.010	.003	.006	.006	.008
50	.017	.007	.005	.005	.005	.005	.005	.006	.010	.004	.004	.007
100	.019	.003	.004	.004	.004	.004	.004	.006	.022	.003	.004	.007

Table 3: *Root-mean-square error results from a simulation study for the dynamic correlation estimators of the DCC model. We only report the estimates of γ and δ and their sum. The estimators include the subset psuedo-likelihood (MSLE), the full pseudo-likelihood (MPLE), Engle's MacGyver strategy (Engle) and the mofile likelihood (MMLE) estimator. All results based on $1,000$ replications and $T = 2,000$.*

7 Additional remarks

7.1 Imposing structure on Ψ

The unconditional mean of the Q_t process, denoted Ψ, is assumed to be positive semidefinite and have unity on its leading diagonal. It may make sense to impose some more structure on it. A leading candidate would be that Ψ obeys a factor structure, which would mean that in the long run the correlations in the model obey a factor structure but in the short run their can be departures from it. This is simple to carry out for DCC or cDCC for it involves replacing the estimation of Ψ by the average outer product of the s_t or s_t^*, respectively, with a ML estimation step on a factor model.

7.2 Parametric modelling on the innovations

The model is incomplete without a assumption on the distribution of $r_t|\mathcal{F}_{t-1}$, for so far we have just assumed a zero conditional mean and time-varying covariance matrix H_t. A simple assumption is that

$$\varepsilon_t = R_t^{-1/2} D_t^{-1} r_t | \mathcal{F}_{t-1} \sim N(0, I),$$

which is obviously parameter free. An alternative would be to estimate the marginal distributions of the ε_t using their empirical distribution functions and then estimating their copula using a parametric form such as a Gaussian or student-t copula. Again it is possible to estimate these parametric structures using the pseudo-likelihood approach based on pairs of observations.

One non-parametric approach is to employ a bootstrap off the multivariate empirical distribution of the

$$\varepsilon_1, \varepsilon_2, ..., \varepsilon_T,$$

simply sampling from these sample points with replacement. This is certainly the easiest viable approach.

N	MMLE	MSLE	MPLE	Engle
3	1.68	.02	.02	.04
10	2.46	.06	.25	.63
50	17.6	.35	7.51	17.4
100	70.8	.76	35.7	67.8
250	2.12	268	409	6928

Table 4: *Mean run time in seconds for the 4 estimation strategies for the DCC model. Throughout* $T = 2,000$. *All based on* $1,000$ *replications except the* $N = 250$ *case which was based on* 20.

Throughout all these methods need the researcher to compute

$$R_t^{-1/2} x_t,$$

where

$$x_t = D_t^{-1} r_t.$$

This is computationally demanding, although it is not as hard as computing the entire inverse of R_t and only has to be carried out once rather than many times in a ML calculation.

7.3 Composite likelihoods

To show both the usefulness and the limitations of this approach, suppose that $Var(r_{jt}) = 1$, $\text{Cor}(r_{jt}, r_{kt}) = \rho$ and our sole task is to estimate ρ from the cross-sectional and time-series dimension. For this extreme model the cross-sectional theoretical properties of some of these estimators are easy to find when the r_t are serially independent, as discussed in some detail in Appendix 8. Simulation results, under this assumption, for the estimators are given in Table 5. The MPLE performs well in the highly and weakly correlated case and less well in the moderate case. MLE can estimate ρ solely off the cross-section so works even when $T = 1$. When $T = 2$ and the r_t are multivariate but temporally independent, the poor moderately correlated MPLE cases are much improved and the bias and efficiency losses are very small for weakly and highly correlated data and the bias in the moderately correlated data is modest.

This Example shows that the psuedo-likelihood approach is not without costs, but that it is able to extract useful information from the cross-section. Its biases will be averaged away when T is moderately large an so we are left with an expectation that it will perform well in practice for more interesting problems such as estimating the memory parameters in dynamic models.

8 Conclusions

References

Aielli, G. P. (2006). Consistent estimation of large scale dynamic conditional correlations. Unpublished paper: Department of Statistics, University of Florence.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*, 817–858.

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

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.

Bollerslev, T. and J. M. Wooldridge (1992). Quasi maximum likelihood estimation and inference in dynamic models with time varying covariances. *Econometric Reviews 11*, 143–172.

Cox, D. R. and N. Reid (2003). 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. (2007). Forecasting correlations: new time series methods. Unpublished paper: Stern Business School, NYU.

Engle, R. F. and B. Kelly (2007). Dynamic equicorrelation. Unpublished paper, Stern Business School, NYU.

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

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

Engle, R. F. and K. Sheppard (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. Unpublished paper: UCSD.

Fearnhead, P. (2003). Consistency of estimators of the population-scaled recombination rate. *Theoretical Population Biology 64*, 67–79.

Glosten, L. R., R. Jagannathan, and D. Runkle (1993). Relationship between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance 48*, 1779–1802.

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.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics 17*, 1217–1241.

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

Ledoit, O., P. Santa-Clara, and M. Wolf (2003). Flexible multivariate GARCH modeling with an application to international stock markets. *The Review of Economics and Statistics 85*, 735–747.

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

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. McFadden (Eds.), *The Handbook of Econometrics, Volume 4*, pp. 2111–2245. North-Holland.

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

Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society, Series A 185*, 71–110.

Politis, D., J. P. Romano, and M. Wolf (1999). *Subsampling*. New York: Springer.

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

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

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

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association 57*, 348–368.

Appendix A: equicorrelation case

8.1 All pairs

Theoretical analsysis of the equicorrelation case is interesting. This follows the discussion given in Section 7.3, where

$$\mathrm{Cor}(r_{jt}, r_{kt}) = \rho, \quad R = \rho\iota\iota' + (1-\rho)\, I.$$

Focus on the $T = 1$ case and ignore the t subscript.

Noting that $2\sum_{j>k}^{N} r_j r_k = \left(\sum_{j=1}^{N} r_j\right)^2 - \sum_{j=1}^{N} r_j^2$ we have

$$S = \frac{1}{N}\sum_{j=1}^{N} r_j^2 \xrightarrow{p} 1 + \rho_*(f^2 - 1), \quad U = \frac{2}{N(N-1)}\sum_{j>k}^{N} r_j r_k \xrightarrow{p} \rho_* f^2.$$

Then we have

$$\begin{aligned}
\frac{2}{N(N-1)}\log L_1^B(\rho) &= -\frac{1}{2}\log\left(1-\rho^2\right) - \frac{S-\rho U}{(1-\rho^2)}, \\
\frac{2}{N(N-1)}\frac{\partial \log L_1^B(\rho)}{\partial \rho} &= \frac{\rho + U}{(1-\rho^2)} - \frac{2\rho\,(S-\rho U)}{(1-\rho^2)^2}, \quad \text{so} \\
U + \widehat{\rho}\,(1-2S) + \widehat{\rho}^2 U - \widehat{\rho}^3 &= 0, \quad \text{so} \\
\left(\rho_* f^2 - p\lim\widehat{\rho}\right)\left(1 + p\lim\widehat{\rho}^2\right) + 2\rho_* p\lim\widehat{\rho} &= 0.
\end{aligned}$$

Figure 1 plots the true value ρ_* against $p\lim\widehat{\rho}$ for a variety of values of f. This shows inconsistency and is due to the fact that the score equation has a zero expectation when averaged over repeated samples of f and ε_i, but when $T = 1$ we only have a single draw from f. This method can be compared to the ML estimator. Now (e.g. Engle and Kelly (2007))

$$R^{-1} = \frac{1}{1-\rho}\left\{I - \frac{\rho}{1+(N-1)\rho}\iota\iota'\right\}, \quad |R| = (1-\rho)^{N-1}\left\{1 + (N-1)\rho\right\},$$

so

$$\begin{aligned}
\log L(\rho) &= -\frac{1}{2}\left[(N-1)\log(1-\rho) + \log\left\{1 + (N-1)\rho\right\}\right] \\
&\quad -\frac{1}{2(1-\rho)}\left[\sum_{j=1}^{N} r_j^2 - \frac{\rho}{1+(N-1)\rho}\left(\sum_{j=1}^{N} r_j\right)^2\right] \\
&= -\frac{1}{2}\left[(N-1)\log(1-\rho) + \log\left\{1 + (N-1)\rho\right\}\right] \\
&\quad -\frac{1}{2(1-\rho)\left\{1+(N-1)\rho\right\}}\left\{\left\{1 + (N-2)\rho\right\}\sum_{j=1}^{N} r_j^2 - 2\rho\sum_{j>k}^{N} r_j r_k\right\}.
\end{aligned}$$

When N is large

$$\frac{1}{N}\log L(\rho) \xrightarrow{p} -\frac{1}{2}\log(1-\rho) - \frac{1}{2(1-\rho)}\left(p\lim U - p\lim S\right) = -\frac{1}{2}\log(1-\rho) - \frac{1}{2(1-\rho)}\left(1-\rho_*\right),$$

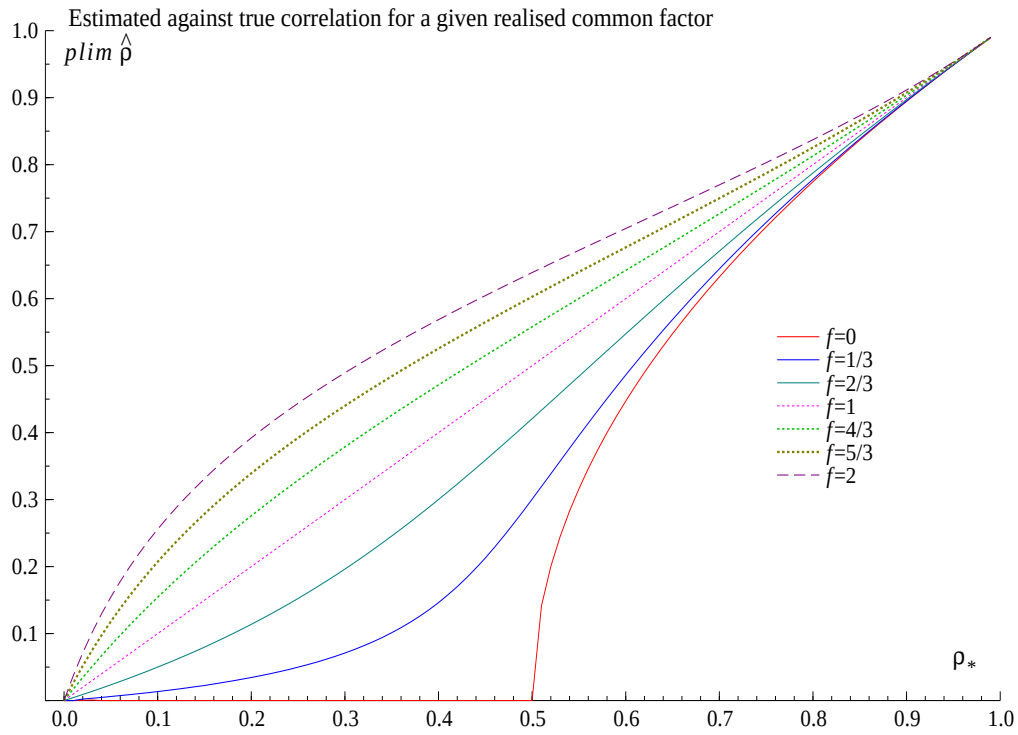

Figure 1: *In the equicorrelation model we can consistently estimate ρ even in the $T = 1$ case using the cross-sectional information and ML estimation. How does the pseudo-likelihood do when $T = 1$? It is inconsistent and this figure shows the resulting pseudo-true value. It demonstrates the necessity of time series information for the pseudo-likelihood approach.*

$$\frac{1}{N}\frac{\partial \log L(\rho)}{\partial \rho} \xrightarrow{p} \frac{1}{2(1-\rho)} - \frac{1}{2(1-\rho)^2}(1-\rho_*),$$

so $p\lim \widehat{\rho} = \rho_*$. Simulation results are given in Table 5 and discussed in the main test.

8.2 Subset estimator

Suppose $N^* = N - 1$ and

$$J_j = K_j - 1 = j,$$

then the most basic subset pseudo-likelihood estimator is

$$\begin{aligned}
\log L_t^B(\lambda, \phi) &= \sum_{j=1}^{N-1} \log L_{j,j+1,t}(\lambda, \phi), \\
S &= \frac{1}{N-1}\sum_{j=1}^{N-1} r_j^2 \xrightarrow{p} 1 + \rho_*(f^2 - 1), \quad U = \frac{1}{N-1}\sum_{j=1}^{N-1} r_j r_{j+1} \xrightarrow{p} \rho_* f^2 \\
\frac{1}{(N-1)}\log L_1^B(\rho) &= -\frac{1}{2}\log\left(1-\rho^2\right) - \frac{S - \rho U}{(1-\rho^2)},
\end{aligned}$$

which means the resulting estimator has the same limit as the paired pseudo-likelihood. The Monte Carlo performance of this estimator is given in Table 5 with this estimator being denoted MSLE.

It shows the efficiency loss compared to computing all possible pairs is extremely modest when N is moderate.

9 Appendix B: asymptotic theory for DCC

The asymptotic theory behind this is a special case of the GMM estimator. In particular the moment conditions

$$\frac{1}{T}\sum_{t=1}^{T} m_t(\theta; y_t|\mathcal{F}_{t-1}) = 0$$

are based around

$$m_t(\theta; y_t|\mathcal{F}_{t-1}) = \begin{pmatrix} \frac{1}{T}\left(\pi_1^2 - r_{1t}^2\right) \\ \partial l_t^{A_1}(\lambda_{(1)})/\partial\tau_1 \\ \vdots \\ \frac{1}{T}\left(\pi_N^2 - r_{Nt}^2\right) \\ \partial l_t^{A_N}(\lambda_{(N)})/\partial\tau_N \\ \frac{1}{T}\left\{vecl(\Psi) - vecl(s_t s_t')\right\}' \\ \partial W_t(\Psi,\xi)/\partial\xi \end{pmatrix},$$

where W_t is the contribution from the t-th observation from the log-likelihood, a total pseudo-likelihood or a subset pseudo-likelihood.

The usual way to perform asymptotics is via a Taylor expansion:

$$\sum_{t=1}^{T} m(\widehat{\theta}; y_t|\mathcal{F}_{t-1}) = 0 \simeq \sum_{t=1}^{T} m(\theta_0; y_t|\mathcal{F}_{t-1}) + \sum_{t=1}^{T} \frac{\partial m(\theta_0; y_t|\mathcal{F}_{t-1})}{\partial\theta'}\left(\widehat{\theta} - \theta_0\right),$$

so

$$\begin{aligned} \sqrt{T}\left(\widehat{\theta} - \theta_0\right) &\simeq \left\{-\frac{1}{T}\sum_{t=1}^{T} \frac{\partial m(\theta_0; y_t|\mathcal{F}_{t-1})}{\partial\theta'}\right\}^{-1} \left\{\sqrt{T}\frac{1}{T}\sum_{t=1}^{T} m(\theta_0; y_t|\mathcal{F}_{t-1})\right\} \\ &\xrightarrow{d} N(0, \mathcal{I}^{-1}\mathcal{J}\mathcal{I}^{-1}), \end{aligned}$$

where

$$\begin{aligned} \mathcal{I} &= p\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \frac{\partial m(\theta_0; y_t|\mathcal{F}_{t-1})}{\partial\theta'} \\ \mathcal{J} &= \lim_{T\to\infty} \mathrm{Cov}\left\{\sqrt{T}\frac{1}{T}\sum_{t=1}^{T} m(\theta_0; y_t|\mathcal{F}_{t-1})\right\}. \end{aligned}$$

Partition $\theta = (\lambda_1, \ldots, \lambda_N, \psi, \xi)$ where $\psi = \mathrm{vecl}\,(\Psi)$ are the intercept parameters and ξ contains the parameters that determine the DCC dynamics.

$\mathcal{I}$ has a block structure that is realtively sparse. We write

$$\frac{1}{T}\sum_{t=1}T\frac{\partial m\left(\theta_0;y_t|\mathcal{F}_{t-1}\right)}{\partial\theta'}=\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix} diag\left(\mathcal{I}_t\left(\lambda_1,\lambda_1\right),\ldots,\mathcal{I}_t\left(\lambda_1,\lambda_N\right)\right) & 0 & 0\\ \mathcal{I}_t\left(\psi,\lambda_1\right),\ldots,\mathcal{I}_t\left(\psi,\lambda_{1N}\right) & \mathcal{I}_t\left(\psi,\psi\right) & \mathcal{I}_t\left(\psi,\xi\right)\\ \mathcal{I}_t\left(\xi,\lambda_1\right),\ldots,\mathcal{I}_t\left(\xi,\lambda_{1N}\right) & \mathcal{I}_t\left(\xi,\psi\right) & \mathcal{I}_t\left(\xi,\xi\right)\end{pmatrix} \tag{15}$$

$$\xrightarrow{p}\begin{pmatrix}\mathcal{I}_{\lambda\lambda} & 0 & 0\\ \mathcal{I}_{\psi\lambda} & \mathcal{I}_{\psi\psi} & \mathcal{I}_{\psi\xi}\\ \mathcal{I}_{\xi\lambda} & \mathcal{I}_{\xi\psi} & \mathcal{I}_{\xi\xi}\end{pmatrix} \tag{16}$$

$$=\mathcal{I}, \tag{17}$$

where

$$\begin{aligned}
\mathcal{I}_t(\lambda_j\lambda_j) &= \begin{pmatrix} 1 & 0\\ \partial^2 l_t^{Aj}(\lambda_{(j)})/\partial\tau_j\partial\pi_j & \partial^2 l_t^{Aj}(\lambda_{(j)})/\partial\tau_j\partial\tau_j'\end{pmatrix},\\
\mathcal{I}_t(\psi,\lambda_j) &= -\frac{1}{T}\partial vecl(s_t^*s_t^{*\prime})/\partial\lambda_{(j)}'\\
\mathcal{I}_t(\psi,\psi) &= I_{N(N-1)/2}\\
\mathcal{I}_t(\psi,\xi) &= -\frac{1}{T}\partial vecl(s_t^*s_t^{*\prime})/\partial\xi'\\
\mathcal{I}_t(\xi,\lambda_{(j)}) &= \partial^2 W_t(\Psi,\xi)/\left(\partial\xi\partial\lambda_{(j)}'\right)\\
\mathcal{I}_t(\xi,\psi) &= \partial^2 W_t(\Psi,\xi)/\left(\partial\xi\partial\psi'\right)S'\\
\mathcal{I}_t(\xi,\xi) &= \partial^2 W_t(\Psi,\xi)/\left(\partial\xi\partial\xi'\right)
\end{aligned}$$

and where S an M by $N(N-1)/2$ selection matrix which will select the M elements of $\text{vecl}(s_t^*s_t^{*\prime})$ that are used in the subset estimator. Typically $A_j = B_j = C = D = 1$ which would mean the dimensions of these matrices are:

$\mathcal{I}$	$\left(3N+2+\frac{N(N-1)}{2}\right)\times\left(3N+2+\frac{N(N-1)}{2}\right)$
$\mathcal{I}_{\lambda\lambda}$	$3N\times 3N$
$\mathcal{I}_{\psi\lambda}$	$N(N-1)/2\times 3N$
$\mathcal{I}_{\xi\lambda}$	$2\times 3N$
$\mathcal{I}_{\psi\psi}$	$N(N-1)/2\times N(N-1)/2$
$\mathcal{I}_{\psi\xi}$	$N(N-1)/2\times 2$
$\mathcal{I}_{\xi\psi}$	$2\times N(N-1)/2$
$\mathcal{I}_{\xi\xi}$	2×2

Remark: In the complete DCC or cDCC model, or when using all $N(N-1)/2$ pairs in the pseudo-likelihood estimator, the Jacobian between the correlation intercepts and the parameters of the correlation dynamics, $\mathcal{I}_t(\xi,\psi)$ is dense and S is simply an identity matrix. However, if using a subset pseudo-likelihood estimator this block will generally be sparse and it is often substantially faster to only compute the columns of $\mathcal{I}_t(\xi,\phi)$ that correspond to the pairs used in estimation of the parameters of the correlation dynamics.

Following the usual method of moments approach, in practice one estimates $\mathcal{I}$ by

$$\widehat{\mathcal{I}}=\frac{1}{T}\sum_{t=1}^{T}\frac{\partial m(\widehat{\theta};y_t|\mathcal{F}_{t-1})}{\partial\theta'}$$

and $\mathcal{J}$ by

$$\widehat{\mathcal{J}}=\frac{1}{T}\sum_{t=1}^{T} m(\widehat{\theta};y_t|\mathcal{F}_{t-1})m(\widehat{\theta};y_t|\mathcal{F}_{t-1})'.$$

However, the moment conditions corresponding to both the variance intercepts, π_j^2, $j=1,\ldots,N$ and the the correlation intercepts, ϕ_{ij}, $i=1,\ldots,N$, $j=i+1,\ldots,N$ are not martingales when the data are conditionally heteroskedastic. As a result a HAC estimator such that of Newey and West (1987) or Andrews (1991) must be used.

N	$\rho = 0.2$				$\rho = 0.5$				$\rho = 0.9$			
	MLE	MPLE	MSLE	Engle	MLE	MPLE	MSLE	Engle	MLE	MPLE	MSLE	Engle
						T=1						
2	**0.099**	**0.099**	**0.099**	**0.099**	**0.271**	**0.271**	**0.271**	**0.269**	**0.650**	**0.650**	**0.650**	**0.650**
	0.783	0.783	0.783	0.783	0.771	0.771	0.771	0.772	0.660	0.660	0.660	0.660
10	**0.198**	**0.180**	**0.230**	**0.120**	**0.450**	**0.431**	**0.445**	**0.351**	**0.900**	**0.899**	**0.898**	**0.898**
	0.273	0.266	0.299	0.430	0.288	0.297	0.318	0.514	0.055	0.049	0.064	0.064
50	**0.187**	**0.148**	**0.175**	**0.210**	**0.496**	**0.416**	**0.415**	**0.472**	**0.900**	**0.899**	**0.899**	**0.899**
	0.151	0.151	0.168	0.238	0.110	0.198	0.220	0.315	0.020	0.022	0.026	0.026
100	**0.191**	**0.146**	**0.163**	**0.218**	**0.500**	**0.410**	**0.408**	**0.480**	**0.900**	**0.899**	**0.899**	**0.899**
	0.118	0.137	0.146	0.215	0.071	0.182	0.202	0.291	0.014	0.016	0.019	0.019
1000	**0.200**	**0.145**	**0.149**	**0.225**	**0.500**	**0.406**	**0.402**	**0.494**	**0.900**	**0.899**	**0.899**	**0.899**
	0.036	0.129	0.129	0.198	0.022	0.162	0.175	0.272	0.004	0.007	0.008	0.008
						T=2						
2	**0.123**	**0.123**	**0.123**	**0.123**	**0.351**	**0.351**	**0.351**	**0.351**	**0.814**	**0.814**	**0.814**	**0.814**
	0.649	0.649	0.649	0.649	0.618	0.618	0.618	0.618	0.404	0.404	0.404	0.404
10	**0.196**	**0.177**	**0.203**	**0.187**	**0.476**	**0.449**	**0.449**	**0.468**	**0.900**	**0.899**	**0.899**	**0.899**
	0.209	0.203	0.238	0.292	0.202	0.224	0.256	0.323	0.033	0.034	0.042	0.042
50	**0.193**	**0.167**	**0.178**	**0.222**	**0.499**	**0.454**	**0.452**	**0.525**	**0.900**	**0.900**	**0.900**	**0.900**
	0.111	0.121	0.137	0.165	0.072	0.139	0.160	0.193	0.014	0.015	0.018	0.018
100	**0.196**	**0.167**	**0.173**	**0.224**	**0.500**	**0.450**	**0.448**	**0.526**	**0.900**	**0.900**	**0.900**	**0.900**
	0.083	0.113	0.122	0.152	0.050	0.128	0.146	0.180	0.010	0.011	0.013	0.013
1000	**0.200**	**0.166**	**0.167**	**0.226**	**0.500**	**0.451**	**0.450**	**0.531**	**0.900**	**0.900**	**0.900**	**0.900**
	0.025	0.105	0.106	0.140	0.016	0.115	0.120	0.168	0.003	0.005	0.006	0.006
						T=10						
2	**0.189**	**0.189**	**0.189**	**0.189**	**0.486**	**0.486**	**0.486**	**0.486**	**0.900**	**0.900**	**0.900**	**0.900**
	0.329	0.329	0.329	0.329	0.252	0.252	0.252	0.252	0.045	0.045	0.045	0.045
10	**0.198**	**0.192**	**0.192**	**0.208**	**0.501**	**0.493**	**0.492**	**0.516**	**0.900**	**0.900**	**0.900**	**0.900**
	0.097	0.095	0.125	0.110	0.072	0.083	0.102	0.086	0.015	0.015	0.018	0.018
50	**0.199**	**0.192**	**0.193**	**0.212**	**0.500**	**0.491**	**0.491**	**0.519**	**0.900**	**0.900**	**0.900**	**0.900**
	0.047	0.061	0.072	0.066	0.031	0.055	0.060	0.052	0.006	0.007	0.008	0.008
100	**0.200**	**0.192**	**0.192**	**0.212**	**0.500**	**0.490**	**0.491**	**0.519**	**0.900**	**0.900**	**0.900**	**0.900**
	0.034	0.057	0.063	0.061	0.022	0.050	0.053	0.047	0.004	0.005	0.006	0.006
1000	**0.200**	**0.191**	**0.191**	**0.212**	**0.500**	**0.489**	**0.489**	**0.518**	**0.900**	**0.900**	**0.900**	**0.900**
	0.011	0.054	0.054	0.057	0.007	0.047	0.048	0.044	0.001	0.003	0.003	0.003

Table 5: *How efficiently do these methods use the cross-sectional information? Simulation study for the equicorrelation model based on multivariate temporally independent data where we are solely estimating ρ. Estimators studied: MLE, psuedo-likelihood, the single pair method and Engle's MacGyver method. Data is based on $T = 1, 2, 10$ and a variety of values of N. Figures in normal font are the standard deviation, the bold font are the mean of the estimator. All results based on 10,000 replications.*