Efficient Estimation of a Multivariate Multiplicative Volatility Model

Christian M. Hafner* Oliver B. Linton†

Université catholique de Louvain London School of Economics

January 26, 2009

Abstract

We propose a multivariate generalization of the multiplicative volatility model of Engle and Rangel (2008), which has a nonparametric long run component and a unit multivariate GARCH short run dynamic component. We suggest various kernel-based estimation procedures for the parametric and nonparametric components, and derive the asymptotic properties thereof. For the parametric part of the model, we obtain the semiparametric efficiency bound. Our method is applied to a bivariate stock index series.

Some key words: GARCH; Kernel Estimation; Semiparametric

JEL Classification Number: C12, C13, C14.

*Institut de statistique and CORE, Université catholique de Louvain, Voie du Roman Pays 20, B-1348. Louvain-la-Neuve, Belgium, `christian.hafner@uclouvain.be`. Financial support from the contract *Projet d'Actions de Recherche Concertées* nr. 07/12/002 of the *Communauté française de Belgique*, granted by the *Académie universitaire Louvain*, is gratefully acknowledged.

†Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail address: `o.linton@lse.ac.uk`. Thanks to the ESRC and Leverhulme foundations for financial support. This paper was partly written while I was a Universidad Carlos III de Madrid-Banco Santander Chair of Excellence, and I thank them for financial support.

1 Introduction

Modelling multivariate volatility is now a big area with many important contributions, see Bauwens, Laurent and Rombouts (2006). We propose a new semiparametric multivariate volatility model that has several advantages. It captures a slowly changing unconditional covariance matrix (low frequency volatility) in a nonparametric way, but also allows for dynamic evolution of the conditional covariance matrix (high frequency volatility) in a more standard fashion. Our model can be viewed as a generalization of the univariate multiplicative model of Engle and Rangel (2008) to the multivariate case as well as a generalization of Rodriguez-Poo and Linton (2001) to allow for short run dynamics. Engle and Rangel (2008) estimate their model by a spline methodology and apply this to a cross-country panel dataset.

We propose several estimation methods for the unknown parameters of low and high frequency volatility based on kernel methods combined with maximum likelihood. The advantage of kernel methods is that one can provide a rigorous asymptotic distribution theory both for the finite dimensional parameters and the nonparametric functions under quite weak conditions, and thereby conduct valid inference about the parameters. We establish the asymptotic properties of our procedures under a semi-strong form specification of the errors. Under the strong form Gaussian distributional specification our procedures are semiparametrically efficient, and we characterize this efficiency bound. We apply our methods to the study of several empirical problems. We should point out that the generalization from the univariate case to the multivariate case is not straightforward and some of the features exploited by Engle and Rangel (2008) do not carry over to the multivariate case.

Throughout the paper we use the following notation: $A \otimes B$ is the Kronecker product of two matrices A and B. vec() is the operator that stacks the columns of a matrix in a column vector, while vech() stacks only the lower triangular including the diagonal into a column vector. D_N is the $N^2 \times N(N+1)/2$ duplication matrix defined by the property $D_N \text{vech}(A) = \text{vec}(A)$ for any symmetric matrix A, and D_N^+ is its generalized inverse.

Proofs of the theorems as well as lemmata are delegated to an appendix.

2 The Model

We observe a vector time series $y_t \in \mathbb{R}^N$ for $t = 1, \ldots, T$. We suppose that y_t satisfies the model

$$y_t = \Sigma(t/T)^{1/2} u_t = \Sigma(t/T)^{1/2} G_t^{1/2} \varepsilon_t, \tag{1}$$

where: ε_t is (at least) a unit conditional variance martingale difference sequence, i.e., $E(\varepsilon_t|\mathcal{F}_{t-1}) = E(\varepsilon_t\varepsilon_t^\top - I_N|\mathcal{F}_{t-1}) = 0$, where $\mathcal{F}_{t-1}$ is the sigma field generated by $\{y_{t-1}, y_{t-2}, \ldots\}$ and I_N is the identity matrix, $\Sigma(t/T)$ is a deterministic covariance matrix, while $G_t \in \mathcal{F}_{t-1}$ is a stationary stochastic covariance matrix process with $EG_t = I_N$. We model G_t parametrically so that $G_t = G_t(\phi)$ for $\phi \in \mathbb{R}^p$. Specifically, we shall assume that G_t is a unit BEKK process

$$G_t = I_N - AA^\top - BB^\top + A\Sigma(t-1/T)^{-1/2} y_{t-1} y_{t-1}^\top \Sigma(t-1/T)^{-1/2} A^\top + BG_{t-1}B^\top \tag{2}$$

in which case $\phi = (\text{vec}(A)^\top, \text{vec}(B)^\top)^\top$ denote the free parameters of G_t. However, other models for G_t can be considered. The matrix function $\Sigma(u)$ is assumed to be of unknown functional form, either smooth or having a finite number of jumps in any compact interval. In the main part of the paper we restrict attention to smooth Σ, but we discuss later the important extension to allow for breaks. The model allows slowly varying unconditional variance matrix Σ along with short run dynamics through the process G_t.

This model is a multivariate generalization of the scalar multiplicative volatility model of Engle and Rangel (2008) where $y_t = \sigma(t/T) g_t^{1/2} \varepsilon_t$ with g_t a unit stationary GARCH process and ε_t is i.i.d. with mean zero and variance one. Actually, they considered a more general model with observed covariates x_t also entering the unconditional variance function $\sigma^2(\cdot)$, we shall discuss this generalization later.[1]

We remark on some properties of the stochastic process y_t. For the univariate process, the local autocorrelation function (LACF) of the squared return series is actually time invariant, i.e.,

$$\begin{aligned}\rho_{y^2}(t,j) &= \frac{\text{cov}(y_t^2, y_{t-j}^2)}{\sqrt{\text{var}(y_t^2)\text{var}(y_{t-j}^2)}} = \frac{\sigma^2(t/T)\sigma^2(t-j/T)\text{cov}(g_t\varepsilon_t^2, g_{t-j}\varepsilon_{t-j}^2)}{\sqrt{\sigma^4(t/T)\sigma^4(t-j/T)\text{var}(g_t\varepsilon_t^2)\text{var}(g_{t-j}\varepsilon_{t-j}^2)}} \\ &= \frac{\text{cov}(g_t\varepsilon_t^2, g_{t-j}\varepsilon_{t-j}^2)}{\text{var}(g_t\varepsilon_t^2)} = \rho_{y^{*2}}(t,j) = \rho_{y^{*2}}(j),\end{aligned}$$

where $y_t^{*2} = g_t\varepsilon_t^2$, and $\rho_{y^{*2}}(t,j)$ is time invariant because of the stationarity of $g_t\varepsilon_t^2$. The correlation structure of the data is just driven by the stationary process g_t. This suggests a simple specification test of the model by looking at the local (in time) correlogram of squared returns and testing whether it is constant.

In the multivariate case this time invariance does not follow, even approximately, which follows from well known properties of the multivariate autocorrelation matrices. That is, they are not invariant to affine transformations of the data $x_t \mapsto b + Ax_t$ for nonsingular matrix A. Specifically,

[1]Their model for long run volatility was of the form

$$\sigma_t^2 = c\exp\left(w_0 t + \sum_{i=1}^{k} w_i (t - t_{i-1})_+^2 + x_t^\top \delta\right),$$

where x_t are observed covariates, while $w_0, \ldots, w_k, \delta$ are unknown parameters.

consider $\eta_t = \text{vech}(y_t y_t^\top)$ and define:

$$z_t = \text{vech}(G_t^{1/2} \varepsilon_t \varepsilon_t^\top G_t^{1/2} - I_N), \tag{3}$$

$M_j = E[z_t z_{t-j}^\top]$, and $W(t,j) = \Sigma(t - j/T)^{1/2} \otimes \Sigma(t - j/T)^{1/2}$. Then

$$\Gamma(t,0) = \text{var}(\eta_t) = D_N^+ W(t,0) M_0 W(t,0) D_N^{+\top}$$

$$\Gamma(t,j) = \text{cov}(\eta_t, \eta_{t-j}) = D_N^+ W(t,0) M_j W(t,j) D_N^{+\top}.$$

Then the local autocorrelation matrix is

$$\Psi(t,j) = \text{diag}[\Gamma(t,0)]^{-1/2} \Gamma(t,j) \text{diag}[\Gamma(t,0)]^{-1/2}.$$

In this case there is generally no cancellation even when one just takes the leading term in $\Gamma(t,j)$, which is $D_N^+ W(t,0) M_j W(t,0) D_N^{+\top}$ for small j. If $\Sigma(u)$ is diagonal, then there is an invariance property that can be exploited [specifically the invariance of autocorrelation to transformations $x_t \mapsto b + A x_t$, where A is diagonal] to show that $\Psi(t,j) \simeq \Psi(j)$ for any fixed j.

3 Estimation

In the sequel we propose an estimation method for the parameters ϕ along with the function $\Sigma(.)$. The estimation method is designed to be efficient under the assumption that ε_t is i.i.d. normal with mean zero and covariance matrix I_N but to be consistent and asymptotically normal for a much broader range of circumstances.

The estimation strategy is in several steps. First, we obtain consistent initial estimators of the unknown quantities, then we improve these using the (Gaussian) likelihood that takes full account of the dependence and non-stationarity structure.

3.1 Step 1 Initial Estimation of Σ

Under the model assumptions,

$$E[y_t y_t^\top] = \Sigma(t/T)$$

for all t with $t = 1, \ldots, T$. Therefore, one can estimate $\Sigma(u)$ by the estimator of Rodriguez-Poo and Linton (2001)

$$\widetilde{\Sigma}(u) = \frac{\sum_{t=1}^T K_h(u - t/T) y_t y_t^\top}{\sum_{t=1}^T K_h(u - t/T)}, \tag{4}$$

where K is a kernel function, h is a bandwidth, and $K_h(.) = K(./h)/h$. Rodriguez-Poo and Linton (2001) established the consistency and asymptotic normality of $\widetilde{\Sigma}(u)$ under general conditions on $\{y_t\}$.

This estimator can be interpreted as the minimizer of the local log-likelihood (upto constants) criterion

$$L_T(\Omega; u) = \sum_{t=1}^{T} K_h(u - t/T) l(\Omega; y_t),$$
$$l(\Omega; y_t) = -\log\det(\Omega) - y_t^\top \Omega^{-1} y_t.$$

Letting $\omega = \mathrm{vech}(\Omega)$, we have

$$\begin{aligned}\frac{\partial l}{\partial \omega} &= -D_N^\top \mathrm{vec}(\Omega^{-1} - \Omega^{-1} y_t y_t^\top \Omega^{-1}) \\ &= -D_N^\top (\Omega^{-1} \otimes \Omega^{-1}) D_N \mathrm{vech}(\Omega - y_t y_t^\top),\end{aligned}$$

which, solving for Ω, yields (4) exactly.

3.2 Step 2 Initial Estimation of ϕ

First, one computes the profiled G process, i.e., for each ϕ, let

$$\widetilde{G}_t(\phi) = I_N - AA^\top - BB^\top + A\widetilde{\Sigma}(t - 1/T)^{-1/2} y_{t-1} y_{t-1}^\top \widetilde{\Sigma}(t - 1/T)^{-1/2} A^\top + B\widetilde{G}_{t-1}(\phi) B^\top \quad (5)$$

for $t = 2, \ldots, T$, where some initialization $\widetilde{G}_1(\phi)$ is chosen. One then computes the profiled global likelihood function

$$\widetilde{\ell}_T(\phi) = \sum_{t=1}^{T} l(\widetilde{\Omega}_t(\phi); y_t)$$
$$\widetilde{\Omega}_t(\phi) = \widetilde{\Sigma}(t/T)^{1/2} \widetilde{G}_t(\phi) \widetilde{\Sigma}(t/T)^{1/2}.$$

Minimize $\widetilde{\ell}_T(\phi)$ with respect to ϕ to give $\widetilde{\phi}$. Actually, since $\widetilde{\Sigma}(t/T)$ does not depend on ϕ, we can replace $\widetilde{\ell}_T(\phi)$ by

$$\widetilde{\ell}_T(\phi) = \sum_{t=1}^{T} l(\widetilde{G}_t(\phi); \widetilde{y}_t),$$

where $\widetilde{y}_t = \widetilde{\Sigma}(t/T)^{-1/2} y_t$. This estimator is expected to be consistent and asymptotically normal but inefficient.

3.3 Step 3 Improved Estimation

Suppose that one knew the random variable G_t, how would you proceed to improve the estimate of $\Sigma(t/T)$ and hence of ϕ? In the scalar case considered by Engle and Rangel (2008), one can just divide through by g_t, using $y_t/g_t^{1/2} = \sigma(t/T)\varepsilon_t$ and then form local averages of $y_t^2/\widetilde{g}_t$. However, in the multivariate case one cannot just "divide through" by $G_t^{1/2}$, since

$$G_t^{-1/2}y_t = G_t^{-1/2}\Sigma(t/T)^{1/2}G_t^{1/2}\varepsilon_t \neq \Sigma(t/T)^{1/2}\varepsilon_t.$$

Our approach instead is to treat G_t as fixed known numbers inside the local likelihood. In particular, suppose that ε_t is normally distributed (this is not maintained in the distribution theory) with mean zero and identity covariance matrix. Then we have conditional on G_t that y_t is normally distributed with conditional mean zero and conditional variance matrix

$$\Omega_t = E\left[y_t y_t^\top | \mathcal{F}_{t-1}\right] = \Sigma(t/T)^{1/2}G_t\Sigma(t/T)^{1/2}.$$

In the sequel we treat $\Sigma(t/T)^{1/2}$ as an unknown parameter and replace it by Θ. Let $\theta = \text{vech}(\Theta) \in \mathbb{R}^{N(N+1)/2}$ be the unique elements of Θ. Consider the local likelihood function

$$L_T(\theta; u) = \sum_{t=1}^{T} K_h(u - t/T)l(\Omega_t(\theta); y_t),$$

as before but where $\Omega_t(\theta) = \Theta G_t \Theta$. Then minimize $L_T(\theta; u)$ with respect to θ. In practice we apply this with the estimated G, i.e.,

$$\widehat{L}_T(\theta; u) = \sum_{t=1}^{T} K_h(u - t/T)l(\widehat{\Omega}_t(\theta, \widetilde{\phi}); y_t), \tag{6}$$

where $\widehat{\Omega}_t(\theta, \widetilde{\phi}) = \Theta\widetilde{G}_t(\widetilde{\phi})\Theta$ and for any ϕ, $\widetilde{G}_t(\phi)$ is given in (5). The resulting estimator is denoted $\widehat{\theta}_{\widetilde{\phi}}(u)$ and hence $\widehat{\Sigma}(u) = \widehat{\Theta}^2_{\widetilde{\phi}}(u)$. We expect that $\widehat{\Sigma}(u)$ is more efficient than $\widetilde{\Sigma}(u)$, at least in the case where ε_t is i.i.d. normal.

Next one computes a new profiled G, i.e., for each ϕ, let

$$\widehat{G}_t(\phi) = I_N - AA^\top - BB^\top + A\widehat{\Sigma}(t-1/T)^{-1/2}y_{t-1}y_{t-1}^\top\widehat{\Sigma}(t-1/T)^{-1/2}A^\top + B\widehat{G}_{t-1}(\phi)B^\top,$$

where some initialization $\widehat{G}_0(\phi)$ is chosen, and then compute the profiled global likelihood function

$$\widehat{\ell}_T(\phi) = -\sum_{t=1}^{T} l(\widehat{\Omega}_t(\phi); y_t)$$

$$\widehat{\Omega}_t(\phi) = \widehat{\Theta}_{\widetilde{\phi}}(t/T)\widehat{G}_t(\phi)\widehat{\Theta}_{\widetilde{\phi}}(t/T),$$

and minimize it with respect to ϕ to give $\widehat{\phi}$ and $\widehat{\theta}_{\widehat{\phi}}$ and hence $\widehat{\Sigma}(u) = \widehat{\Theta}^2_{\widehat{\phi}}(u)$.

One can iterate this procedure by updating now the local likelihood using the new estimator of ϕ and so on, but asymptotically this will not affect the variances of the procedures. We have chosen to describe an iterative method of estimator construction for practical advantages, because profiling is not an attractive option here due to the presence of $\Sigma(1/T), \ldots, \Sigma(1)$ inside the definition of $G_t(\phi)$.

4 Efficient Estimation

We discuss here the question of efficient estimation of the nonparametric part and the parametric part.

4.1 The Function $\Sigma(.)$

We now discuss efficient estimation of $\Sigma(.)$. Efficient estimation of nonparametric functions is not as clear cut as in the parametric case since mean squared error typically only induces a partial ordering on different estimators. However, one can make some comparisons according to variance as we shall see.

Consider the following two nonparametric regressions:

$$\widetilde{y}_t^2 = \frac{y_t^2}{g_t} = \sigma^2(t/T) + \sigma^2(t/T)(\varepsilon_t^2 - 1) \tag{7}$$

$$y_t^2 = \sigma^2(t/T) + \sigma^2(t/T)(g_t\varepsilon_t^2 - 1). \tag{8}$$

In both cases the error term is mean zero so that smoothing on time will yield consistent estimates of $\sigma^2(t/T)$ in both cases. However, in the first case, the error is a martingale difference sequence, while in the second it is not. Also, the variance of the error term in the second equation is larger than in the first. Specifically, $\text{var}(g_t\varepsilon_t^2 - 1) = E[(g_t\varepsilon_t^2 - 1)^2] = E(g_t^2\varepsilon_t^4) - 1 = E(\varepsilon_t^4)Eg_t^2 - 1$. Since $\text{var}(g_t) = Eg_t^2 - 1$, we have $Eg_t^2 \geq 1$ and so $\text{var}(g_t\varepsilon_t^2 - 1) \geq \text{var}(\varepsilon_t^2 - 1)$. In fact, using $g_t = 1 - \alpha - \beta + \alpha g_{t-1}\varepsilon_{t-1}^2 + \beta g_{t-1}$, we have

$$\begin{aligned} Eg_t^2 &= (1-\alpha-\beta)^2 + \alpha^2 Eg_{t-1}^2 E\varepsilon_{t-1}^4 + \beta^2 Eg_{t-1}^2 + 2(1-\alpha-\beta)\alpha + 2(1-\alpha-\beta)\beta + 2\alpha\beta Eg_{t-1}^2 \\ &= \frac{(1-\alpha-\beta)^2 + 2(1-\alpha-\beta)\alpha + 2(1-\alpha-\beta)\beta}{1 - ((\alpha^2 E\varepsilon_{t-1}^4 + \beta^2 + 2\alpha\beta))}. \end{aligned}$$

The process $\varepsilon_t^2 - 1$ is uncorrelated but $g_t\varepsilon_t^2 - 1$ is autocorrelated, and in particular $\text{cov}(g_t\varepsilon_t^2 - 1, g_{t-j}\varepsilon_{t-j}^2 - 1) = E(g_t g_{t-j}\varepsilon_{t-j}^2) - 1 \neq 0$. This is one intuition why the improved estimator of $\Sigma(.)$ is likely to be more efficient than the original one. Another way of seeing the value of this transformation is to observe that the local likelihood function for y_t^2 with known g_t and unknown $\sigma^2(t/T)$ can be written

$$L_T(\tau;u) = \sum_{t=1}^{T} K_h(u - t/T)\left[\ln\tau + \frac{y_t^2}{g_t\tau}\right],$$

which yields the estimator

$$\widehat{\tau}(u) = \frac{\sum_{t=1}^{T} K_h(u-t/T)\frac{y_t^2}{g_t}}{\sum_{t=1}^{T} K_h(u-t/T)},$$

which corresponds to a standard regression smoother from (7).

4.2 The Parameters ϕ

Here we consider the question of semiparametric efficiency, Bickel, Klaassen, Ritov, and Wellner (1993). Consider the univariate case where

$$\sigma_t^2(\phi,\theta) = h(t/T)g_t(\phi)$$

for some unknown function $h(.)$. We suppose that

$$\log h(t/T) = \sum_{j=0}^{\infty} \theta_j\psi_j(t/T)$$

for some orthonormal basis $\{\psi_j\}_{j=0}^{\infty}$ with $\psi_0(u) = 1$, and

$$\frac{1}{T}\sum_{t=1}^{T} \psi_j(t/T)\psi_k(t/T) \to \delta_{jk},$$

where $\delta_{jk} = 1$ if $j = k$ and 0 if $j \neq k$. Then consider some finite order approximation $h_\theta(t/T)$ with $J < \infty$. We have:

$$\frac{\partial \log h_\theta(t/T)}{\partial\theta_j} = \psi_j(t/T)$$

$$I_{\theta\theta} = \lim_{T\to\infty} E\left[\frac{\partial \ell_T(\phi_0,\theta_0)}{\partial\theta}\frac{\partial \ell_T(\phi_0,\theta_0)}{\partial\theta^\top}\right] = m_4 \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\left[\psi_j(t/T)\psi_k(t/T)\right]_{j,k} = m_4 I_J$$

$$I_{\phi\theta} = m_4 E\left[\frac{\partial \log g_t}{\partial\phi}\right]\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\left[\psi_j(t/T)\right]_j = m_4 E\left[\frac{\partial \log g_t}{\partial\phi}\right](1,0,\ldots,0)^\top$$

$$\sqrt{T}\frac{\partial \ell_T(\phi_0,\theta_0)}{\partial\theta_j} = -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\varepsilon_t^2 - 1)\psi_j(t/T),$$

because $\int_0^1 \psi_j(u)du = 0$ for all $j \geq 1$, where $m_4 = E[(\varepsilon_t^2 - 1)^2]$. Therefore, the semiparametric efficient score function is

$$\frac{\partial \ell_T^*(\phi_0)}{\partial\phi} = \frac{1}{T}\sum_{t=1}^{T}(\varepsilon_t^2 - 1)\left[\frac{\partial \log g_t}{\partial\phi} - E\left(\frac{\partial \log g_t}{\partial\phi}\right)\right].$$

Following the approach explained in Bickel, Klaassen, Ritov, and Wellner (1993), we can construct an efficient estimator of ϕ from the one-step estimator

$$\widehat{\phi}^{eff2-step} = \widetilde{\phi} - \left[\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial \log \widetilde{g}_t}{\partial \phi} - \frac{1}{T}\sum_{t=1}^{T}\frac{\partial \log \widetilde{g}_t}{\partial \phi}\right]\left[\frac{\partial \log \widetilde{g}_t}{\partial \phi} - \frac{1}{T}\sum_{t=1}^{T}\frac{\partial \log \widetilde{g}_t}{\partial \phi}\right]^{\top}\right]^{-1} \times\frac{1}{T}\sum_{t=1}^{T}(\widetilde{\varepsilon}_t^2 - 1)\left[\frac{\partial \log \widetilde{g}_t}{\partial \phi} - \frac{1}{T}\sum_{t=1}^{T}\frac{\partial \log \widetilde{g}_t}{\partial \phi}\right]. \quad (9)$$

In the multivariate case, the semiparametric efficient score function is given by

$$\frac{\partial \ell_T^*(\phi_0)}{\partial \phi_i} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\text{vec}\left(G_t^{-1/2}\frac{\partial G_t}{\partial \phi_i}G_t^{-1/2} - E\left[G_t^{-1/2}\frac{\partial G_t}{\partial \phi_i}G_t^{-1/2}\right]\right)^{\top}\text{vec}\left(\varepsilon_t\varepsilon_t^{\top} - I_N\right)$$

as we next explain. With this one can compute the sample equivalent and implement the two-step efficient estimator as in the univariate case. In the following we motivate a proof for the efficient score in the multivariate case.

Consider the following likelihood function

$$\ell_T(\phi) = \sum_{t=1}^{T} l(\Omega_t(\phi); y_t)$$
$$y_t = \Sigma(t/T)^{1/2}G_t^{1/2}(\phi)\varepsilon_t$$
$$\Omega_t(\phi) = \Sigma(t/T)^{1/2}G_t(\phi)\Sigma(t/T)^{1/2}.$$

where $\phi \subset \Phi \in \mathbb{R}^K$. The score with respect to ϕ is given by

$$\frac{\partial l_t}{\partial \phi} = W_t\text{vec}(\varepsilon_t\varepsilon_t^{\top} - I_N)$$

where

$$W_t = \frac{\partial \text{vec}(G_t)^{\top}}{\partial \phi}(G_t^{-1/2} \otimes G_t^{-1/2})$$

The matrix function $\Sigma(t/T)$ is nonparametric. Consider a parametric submodel, $\Sigma_\theta(t/T)$, where $\theta \subset \Theta \in \mathbb{R}^J$ is the nuisance parameter. The tangent set $\mathcal{T}$ is defined as the mean square closure of $A(\partial l_t/\partial\theta)$, where $A \in \mathbb{R}^{K\times J}$. The score with respect to the nuisance parameter θ is given by

$$\frac{\partial l_t}{\partial \theta} = -\frac{\partial \text{vech}(\Sigma_\theta^{1/2})^{\top}}{\partial \theta}D_N^{\top}\left(\Sigma(t/T)^{-1/2}G_t^{-1/2} \otimes G_t^{1/2}\right)\text{vec}(\varepsilon_t\varepsilon_t^{\top} - I_N).$$

Due to the properties of $\partial l_t/\partial\theta$, the tangent set can be defined as

$$\mathcal{T} = \{f : \mathbb{R}^N \to \mathbb{R}^K \mid \text{E}[f(x)] = 0, \text{E}[f(x)f(x)^{\top}] < \infty\}$$

The projection of $\partial l_t/\partial\phi$ on $\mathcal{T}$ is given by

$$P_t = \mathrm{E}\left[W_t\right]\mathrm{vec}(\varepsilon_t\varepsilon_t^\top - I_N)$$

To see this, note first that $P_t \in \mathcal{T}$ as it has mean zero and finite variance, due to the independence of W_t and $\mathrm{vec}(\varepsilon_t\varepsilon_t^\top - I_N)$. Then, the orthogonal complement of the projection,

$$\frac{\partial l_t^*}{\partial\phi} = \frac{\partial l_t}{\partial\phi} - P_t = (W_t - \mathrm{E}[W_t])\mathrm{vec}(\varepsilon_t\varepsilon_t^\top - I_N)$$

is orthogonal to all elements of $\mathcal{T}$, since $(W_t - \mathrm{E}[W_t])$ has mean zero and is independent of $\mathrm{vec}(\varepsilon_t\varepsilon_t^\top - I_N)$ and, hence, of all elements of $\mathcal{T}$. The uniqueness of the projection completes the proof. □

The score $\partial l_t^*/\partial\phi$ is the efficient score function, see Bickel, Klaassen, Ritov, and Wellner (1993)

5 Distribution Theory

In this section we give the asymptotic distribution theory of the various estimators considered above. We first introduce some notation. Consider z_t defined in (3), which is a stationary mixing process with unconditional mean zero, and let

$$\Gamma_j = E[z_t z_{t-j}^\top], \quad j = 0, 1, \ldots \tag{10}$$

Then let

$$V_\sigma(u) = \|K\|_2^2 D_N \left[\Sigma(u)^{1/2} \otimes \Sigma(u)^{1/2}\right] D_N M_\infty D_N \left[\Sigma(u)^{1/2} \otimes \Sigma(u)^{1/2}\right] D_N \tag{11}$$

$$M_\infty = \mathrm{lrvar}(z_t) = \Gamma_0 + \sum_{j=1}^{\infty}(\Gamma_j + \Gamma_j^\top).$$

$$V_\phi = J^{-1} Q J^{-1} \tag{12}$$

$$J = E\left[\rho_t\rho_t^\top\right],$$

where

$$\rho_t = \frac{\partial \mathrm{vec}(G_t)^\top}{\partial\phi}(G_t^{-1/2} \otimes G_t^{-1/2}), \tag{13}$$

and where the matrix Q is defined in (35) in the appendix. Let $\sigma(u) = \mathrm{vech}(\Sigma(u))$ and $\widetilde{\sigma}(u) = \mathrm{vech}(\widetilde{\Sigma}(u))$.

Theorem 1. *Under our conditions, there exist bounded continuous functions $b_\sigma(u)$ such that*

$$\sqrt{Th}(\widetilde{\sigma}(u) - \sigma(u) - h^2 b_\sigma(u)) \Longrightarrow N(0, V_\sigma(u)) \tag{14}$$

$$\sqrt{T}(\widetilde{\phi} - \phi) \Longrightarrow N(0, V_\phi). \tag{15}$$

The result in (14) corrects the asymptotic variance of Rodriguez-Poo and Linton (2001). In particular, $V_\sigma(u)$ depends on the correlation structure of the error term z_t. The bias function is proportional to $\sigma''(u)$.

Define

$$\zeta_t = \mathrm{vec}(\varepsilon_t \varepsilon_t^\top - I_N), \tag{16}$$

which is a vector martingale difference sequence, and let

$$\Xi_t = E[\zeta_t \zeta_t^\top | \mathcal{F}_{t-1}]. \tag{17}$$

Then let

$$V_\sigma^e(u) = \|K\|_2^2 \, \Delta(u)\Lambda(u)^{-1}\Psi(u)\Lambda(u)^{-1}\Delta(u)^\top \tag{18}$$

$$\Delta(u) = D_N^+ \left[\left(I_N \otimes \Sigma^{1/2}(u)\right) + \left(\Sigma^{1/2}(u) \otimes I_N\right) \right] D_N$$

$$\Lambda(u) = E\left[W_t(u) W_t(u)^\top\right] = 4D_N^\top \left(\Sigma^{-1/2}(u) \otimes I_N\right) E(G_t^{-1} \otimes G_t) \left(\Sigma^{-1/2}(u) \otimes I_N\right) D_N$$

$$\Psi(u) = E\left[W_t^\top(u) \Xi_t W_t(u)\right]$$

$$W_t(u) = 2D_N^\top (\Sigma(u)^{-1/2} G_t^{-1/2} \otimes G_t^{1/2}).$$

Note that $E(G_t^{-1} \otimes G_t) \neq I_N$ except in the scalar case.

Let $\widehat{\sigma}(u) = \mathrm{vech}(\widehat{\Sigma}(u))$.

THEOREM 2. *Under our conditions, there exist bounded continuous functions $b_\sigma^e(u)$ such that*

$$\sqrt{Th}(\widehat{\sigma}(u) - \sigma(u) - h^2 b_\sigma^e(u)) \Longrightarrow N(0, V_\sigma^e(u)) \tag{19}$$

When ε_t are i.i.d. standard normal, $E[\Xi_t] = 2D_N D_N^+$ and one can show that $V_\sigma^e(u) \leq V_\sigma(u)$.

Let

$$V_\phi^e = J^{e-1} Q^e J^{e-1} \tag{20}$$

$$J^e = E\left[(\rho_t - E[\rho_t])(\rho_t - E[\rho_t])^\top\right]$$

$$Q^e = E\left[(\rho_t - E[\rho_t]) \Xi_t (\rho_t - E[\rho_t])^\top\right].$$

THEOREM 3. *Under our conditions*

$$\sqrt{T}(\widehat{\phi} - \phi) \Longrightarrow N(0, V_\phi^e) \tag{21}$$

When ε_t are i.i.d. standard normal, $E[\Xi_t] = 2D_N D_N^+$ and $V_\phi^e = 2J^{e-1}$. In this case, $V_\phi^e \leq V_\phi$.

Our distribution theory can be used to conduct inference and to select bandwidth. Specifically, let $a(\sigma)$ be a scalar function of Σ such as the trace or determinant, and let $a_0(u) = \partial a(\sigma(u))/\partial\sigma$. Then the pointwise mean squared error is

$$s(u) = \frac{1}{Th} a_0^\top(u) V_\sigma^e(u) a_0(u) + h^4 a_0^\top(u) b_\sigma^e(u) b_\sigma^e(u)^\top a_0(u)$$

and the integrated mean squared error is $\int s(u)w(u)du$ for some non-negative weighting function w. The optimal global bandwidth sequence is

$$h_{opt}(T) = \left[\frac{\int a_0^\top(u) V_\sigma^e(u) a_0(u) w(u) du}{4 \int a_0^\top(u) b_\sigma^e(u) b_\sigma^e(u)^\top a_0(u) w(u) du} \right] T^{-1/5} \tag{22}$$

and likewise for the optimal pointwise bandwidth. In practice we should estimate the unknown quantities consistently.

6 Extensions

In this section we discuss three possible extensions of the model.

6.1 Discontinuities

One can allow Σ to have a finite number of discontinuities by using only one sided kernels. Suppose that our model is that for some known union of intervals $U = \cup_{\ell=1}^L [u_-^\ell, u_+^\ell] \subset [0,1]$,

$$\Sigma(u) = \Sigma_c(u) + \Sigma_d 1(u \in U),$$

where $\Sigma_c(\cdot)$ is a smooth unknown function and Σ_d is an unknown matrix. This model is potentially useful for studying the effect of business cycles on volatility in which case U might correspond to recession periods. The continuous part $\Sigma_c(\cdot)$ is estimated as before. We now show how to estimate Σ_d. Let

$$\begin{aligned} \widetilde{\Sigma}_-(u) &= \frac{\sum_{t=1}^T K_h^-(u - t/T) y_t y_t^\top}{\sum_{t=1}^T K_h^-(u - t/T)} \\ \widetilde{\Sigma}_+(u) &= \frac{\sum_{t=1}^T K_h^+(u - t/T) y_t y_t^\top}{\sum_{t=1}^T K_h^+(u - t/T)}, \end{aligned}$$

where K^-, K^+ are respectively left and right sided kernels defined on $[-1,0]$ and $[0,1]$ respectively, say. We then propose the estimator

$$\widetilde{\Sigma}_d = \sum_{\ell=1}^L w_{\ell-} \left(\widetilde{\Sigma}_+(u_-) - \widetilde{\Sigma}_-(u_-) \right) + w_{\ell+} \left(\widetilde{\Sigma}_-(u_+) - \widetilde{\Sigma}_+(u_+) \right)$$

for some weighting sequence $\{w_{\ell-}, w_{\ell+}\}_{\ell=1}^L$ with $\sum_{\ell=1}^L w_{\ell-} + w_{\ell+} = 1$.

6.2 Exogenous covariates

One could suppose also that Σ depends on strictly exogenous covariates. For example, suppose that

$$\Sigma(t/T, X_t) = \Psi^{1/2}(t/T)H_\eta(X_t)\Psi^{1/2}(t/T),$$

where $H_\eta(X_t)$ is a unit covariance matrix determined by unknown parameters η. This is like in the multiplicative model of Engle and Rangel (2008). It is straightforward to modify the estimation algorithms to accomodate this case.

6.3 Reduced rank

One could also introduce reduced rank assumptions into $\Sigma(t/T)$ as in Rodriguez-Poo and Linton (2001). Since $\Sigma(t/T)$ is a real symmetric matrix we have the decomposition

$$\Sigma(t/T) = Q(t/T)\Lambda(t/T)Q(t/T)^\top,$$

where $Q(t/T)Q(t/T)^\top = I$ and $\Lambda(t/T) = \text{diag}\{\lambda_1(t/T), \ldots, \lambda_N(t/T)\}$. Now suppose that $\lambda_j(.) \equiv 0$ for $j = K+1, \ldots, N$, where $K \leq N$. When $K < N$ there is a reduction in the effective dimensionality of the long run covariance matrix. One may be interested in identifying and testing restrictions on the rank K. Such issues are discussed in detail in Rodriguez-Poo and Linton (2001).

7 Application

We apply the proposed estimator to the bivariate series of daily Dow Jones and NASDAQ index returns, January 2, 1990 to January 7, 2009, giving a sample size of $T = 4795$. A shorter series has been analysed in Engle (2002) and Boswijk and van der Weide (2006).

For the nonparametric estimation of $\Sigma(u)$ we use estimator of Rodriguez-Poo and Linton (2001) with quartic kernel function. The bandwidth of the first stage estimator is set to 0.05, such that about 5 % of the data are used for local averaging. The second stage bandwidth is chosen according to (22), where the unknown quantities $V_\sigma^e(u), a_0(u)$ and $b_\sigma^e(u)$ are estimated using the first stage estimates and the weight function $w(u)$ is set to one, which gives a bandwidth of 0.056.

Table 1 shows parameter estimates for G_t using the first stage and the efficient estimator. The first four parameters are the elements of A, the remaining those of B. We see that all parameters are very similar, certainly due to the large sample and the fact the first stage estimator is consistent. In the BEKK model, a measure of persistence of volatilities and correlations is given by the largest eigenvalue of the matrix $A \otimes A + B \otimes B$, which is 0.9295 for the first stage estimator and 0.9322 for the efficient one. Note that these measures of persistence are substantially lower than those of models

Parameter	First stage	(std.err.)	Efficient	(std.err.)
a11	0.2115	(0.0233)	0.2163	(0.0211)
a12	0.0339	(0.0219)	0.0348	(0.0204)
a21	-0.0177	(0.0328)	-0.0117	(0.0151)
a22	0.2488	(0.0205)	0.2619	(0.0202)
b11	0.9489	(0.0136)	0.9522	(0.0135)
b12	-0.0236	(0.0133)	-0.0266	(0.0105)
b21	0.0199	(0.0288)	0.0183	(0.0157)
b22	0.9228	(0.0179)	0.9181	(0.0163)

Table 1: Estimated Parameters of G_t using the first stage and the efficient estimator.

neglecting changes in unconditional volatilities and correlations, where persistence is typically very close to one.

The estimated conditional and unconditional standard deviation and correlation plots are shown in Figures 1 to 3. There are no substantial differences between the first stage and efficient estimates, and hence we show only the efficient estimates to economize on space. The decline in correlations around the year 2000, due to the decoupling of technology and brick and mortar stocks during the new economy boom, is more pronounced in our case than it is using DCC or OGARCH models. Note the steep increase in volatilities and correlations towards the end of the sample, due to the financial crisis. The unconditional volatility of the NASDAQ is about as high as around the new economy boom, whereas the Dow Jones, although at the same level as the NASDAQ, shows a much higher unconditional volatility than in 2000.

The eigenvalues of the efficient estimator of $\Sigma(u)$ are shown in Figure 4. We see that especially at the beginning of the sample, the smaller eigenvalue is close to zero. In higher dimensions this may occur for a number of eigenvalues, in which case one may want to use tests for zero eigenvalues as in Rodriguez-Poo and Linton (2001) and impose factor-type restrictions as discussed in Section 6.3. The largest eigenvalue explodes towards the end of 2008 reflecting the big increase in volatility.

8 Conclusions

We have introduced a new multivariate semiparametric volatility model that combines the idea of a long term smoothly evolving component with a short term, more erratic one that fluctuates around the smooth component. This generalizes the model of Engle and Rangel (2008) to the multivariate case. We provide estimation theory and suggest a semiparametric efficient estimator of the parametric part.

We have mentioned several extensions of the basic model, including exogenous variables, discontinuities of the nonparametric functions and reduced rank of the parametric part of the model. Future work consists of extending the parametric part, for which we assumed BEKK, to other model classes such as DCC or factor-type GARCH models. Also, more practical experience is needed to demonstrate the usefulness of the model.

Appendix

A Assumptions

(A1) The matrix function $\Sigma(u)$ is uniformly positive definite and twice continuously differentiable on $(0, 1)$.

(A2) The centered random vectors $\{\varepsilon_t\}$ have a positive lower semi-continuous density w.r.t. the Lebesgue measure on the set $\{\varepsilon_t \in \mathbb{R}^N : \|\varepsilon_t\| \leq \eta\}$, for some $\eta > 0$. The initial condition x_0 is independent of $\{\varepsilon_t\}$.

(A3) $\det(A) \neq 0$ and $\rho(B) < 1$, where $\rho(B)$ is the spectral radius of B.

(A4) The parameter space Φ is compact.

(A5) The sequence $\{u_t\}$ is strictly stationary and ergodic and $\mathrm{E}\|u_t\|^6 < \infty$.

(A6) $\mathrm{E}\|\varepsilon_t\|^4 < \infty$ and $\mathrm{var}(\varepsilon_t) = I_N$

(A7) The BEKK model is identifiable: If for any $\phi, \phi_0 \in \Phi$, $G_t(\phi) = G_t(\phi_0)$ a.s., then $\phi = \phi_0$.

(A8) The parameter ϕ_0 is an interior point of Φ.

(A9) The function K is symmetric about zero with compact support and satisfies $\int sK(s)ds = 0$. Let $\|K\|_2^2 = \int K(s)^2 ds$.

(A10)

(a) $h(T) \to 0$ as $T \to \infty$ such that $Th \to \infty$ and $Th^4 \to 0$.

(b) $h(T) = c_T T^{-1/5}$ *with* $0 < \liminf_{T\to\infty} c_T \leq \limsup_{T\to\infty} c_T < \infty$.

The assumptions concerning the BEKK model are similar to those of Jeantheau (1998) and Comte and Lieberman (2003). The assumptions A10(a) are used to derive the properties of the estimators of ϕ, while assumptions A10(b) are used to derive the properties of the estimators of $\sigma(u)$.

Define

$$\ell(\phi) = T^{-1} E\ell_T(\phi)$$
$$\ell_T(\phi) = -\sum_{t=1}^{T} \log\det G_t(\phi) - \sum_{t=1}^{T} y_t^\top G_t^{-1}(\phi) y_t.$$

We suppose that ϕ_0 is the unique minimizer of $\ell(\phi)$.

B Proof of Theorem 1

Let $U_t(u) = \Sigma(u)^{1/2}[G_t^{1/2}\varepsilon_t\varepsilon_t^\top G_t^{1/2} - I_N]\Sigma(u)^{1/2}$. To establish (14) we use the following lemma.

Lemma 1. *For some bounded continuous function $B(u)$,*

$$\sup_{u\in[0,1]} \left\| \widetilde{\Sigma}(u) - \Sigma(u) - \frac{1}{T}\sum_{t=1}^{T} K_h(u-t/T)U_t(u) - h^2 B(u) \right\| = O\left(\frac{\log T}{Th}\right) + o\left(h^2\right) \text{ a.s.}$$

$$\sup_{u\in[0,1]} \left\| \frac{1}{T}\sum_{t=1}^{T} K_h(u-t/T)U_t(u) \right\| = O\left(\sqrt{\frac{\log T}{Th}}\right) \text{ a.s.}$$

It then follows by CLT for mixing processes that

$$\sqrt{Th}\frac{1}{T}\sum_{t=1}^{T} K_h(u-t/T)u_t(u) \Longrightarrow N(0, V_\sigma(u)),$$

where $u_t(u) = \mathrm{vech}(U_t(u)) = D_N\left[\Sigma(u)^{1/2}\otimes\Sigma(u)^{1/2}\right] D_N^+ z_t$ where $z_t = \mathrm{vech}(G_t^{1/2}\varepsilon_t\varepsilon_t^\top G_t^{1/2} - I_N)$ is stationary and mixing. The bias function $B(u)$ is $\mu_2(K)\Sigma''(u)/2$.

The proof of (15) is obtained using two lemmas that are proved below.

Lemma 2. *As $T\to\infty$,*

$$\widetilde{\phi} \xrightarrow{P} \phi_0. \tag{23}$$

Let

$$Q = \lim_{T\to\infty} \mathrm{var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \aleph_t\right),$$

where $\aleph_t = (\aleph_t^1, \ldots, \aleph_t^p)^\top$ defined below is a (locally) stationary mixing process.

Lemma 3. *As $T\to\infty$, the following expansion holds*

$$\sqrt{T}\left(\widetilde{\phi} - \phi\right) = -\left[\frac{\partial^2\widetilde{\ell}_T(\phi_0)}{\partial\phi\partial\phi^\top}\right]^{-1} \sqrt{T}\frac{\partial\widetilde{\ell}_T(\phi_0)}{\partial\phi} + o_p(1), \tag{24}$$

$$\sqrt{T}\frac{\partial \widetilde{\ell}_T(\phi_0)}{\partial \phi} \Longrightarrow N(0,Q) \tag{25}$$

$$\frac{\partial^2 \widetilde{\ell}_T(\phi_0)}{\partial \phi \partial \phi^\top} \xrightarrow{P} J. \tag{26}$$

■

C Proof of Theorem 2

Since $\widetilde{\phi} = \phi_0 + O_p(T^{-1/2})$ we can replace $\widetilde{\phi}$ by ϕ_0. We then show that one can equally assume that we can replace $\widetilde{G}_t(\widetilde{\phi})$ by $G_t(\phi_0)$. It follows that by a Taylor series expansion we have

$$\widehat{\theta} - \theta = -\left[\frac{\partial^2 \widehat{L}_T(\theta;u)}{\partial \theta \partial \theta^\top}\right]^{-1} \frac{\partial \widehat{L}_T(\theta;u)}{\partial \theta} + o_p(T^{-1/2}) + o_p\left(\left\|\widehat{\theta} - \theta\right\|\right).$$

Using $\theta = \text{vech}(\Sigma^{1/2}(t/T))$, we can obtain for the local score,

$$\frac{\partial L_T(\theta;u)}{\partial \theta} = -\sum_{t=1}^T K_h(u - t/T)\frac{\partial \text{vec}(\Omega_t)^\top}{\partial \theta}\left(\Omega_t^{-1/2} \otimes \Omega_t^{-1/2}\right) D_N \text{vech}\left(\varepsilon_t \varepsilon_t^\top - I_N\right).$$

This score function is a kernel weighted sum of martingale difference sequence errors. Therefore, it satisfies a CLT. As G_t and $\Sigma(t/T)$ are symmetric, we have (Lütkepohl (1996), p.190, (5)) $\frac{\partial \text{vec}(\Omega_t)^\top}{\partial \theta} = 2D_N^\top (I_N \otimes \Sigma(t/T)^{1/2} G_t) D_N D_N^+$. Furthermore,

$$\begin{aligned}
\frac{\partial \text{vec}(\Omega_t)^\top}{\partial \theta}\left(\Omega_t^{-1/2} \otimes \Omega_t^{-1/2}\right) D_N &= 2D_N^\top (I_N \otimes \Sigma(t/T)^{1/2} G_t) D_N D_N^+ \left(\Omega_t^{-1/2} \otimes \Omega_t^{-1/2}\right) D_N \\
&= 2D_N^\top (I_N \otimes \Sigma(t/T)^{1/2} G_t) \left(\Omega_t^{-1/2} \otimes \Omega_t^{-1/2}\right) D_N \\
&= 2D_N^\top (\Omega_t^{-1/2} \otimes \Sigma(t/T)^{1/2} G_t \Omega_t^{-1/2}) D_N \\
&= 2D_N^\top (\Sigma(t/T)^{-1/2} G_t^{-1/2} \otimes G_t^{1/2}) D_N =: W_t(t/T) D_N.
\end{aligned}$$

The second equality follows since for any A, $D_N D_N^+ (A \otimes A) D_N = (A \otimes A) D_N$ by Lütkepohl (1996), p. 124, 9.5.4.(1d). Thus, we can write the score as

$$\frac{\partial L_T(\theta;u)}{\partial \theta} = -\sum_{t=1}^T K_h(u - t/T) W_t(u) \text{vec}\left(\varepsilon_t \varepsilon_t^\top - I_N\right).$$

We calculate the variance matrix.

$$\begin{aligned}
E\left[\frac{1}{T}\frac{\partial L_T(\theta;u)}{\partial \theta}\frac{\partial L_T(\theta;u)}{\partial \theta^\top}\right] &= \frac{1}{T}\sum_{t=1}^T K_h(u - t/T) E\left[W_t(u)\text{vec}\left(\varepsilon_t \varepsilon_t^\top - I_N\right)\text{vec}\left(\varepsilon_t \varepsilon_t^\top - I_N\right)^\top W_t(u)^\top\right] \\
&= \frac{1}{T}\sum_{t=1}^T K_h(u - t/T) E\left[W_t(u) \Xi_t W_t(u)^\top\right] \\
&= E\left[W_t(u) \Xi_t W_t(u)^\top\right] + o_p(1)
\end{aligned}$$

At the true local parameter $\Theta_0 = \Sigma^{1/2}(u)$ we have apart from smoothing biases

$$\begin{aligned}
E\left[\frac{1}{T}\frac{\partial^2 L_T(\theta;u)}{\partial\theta\partial\theta^\top}\right]_{\Theta_0=\Sigma^{1/2}(u)} &= \frac{1}{T}\sum_{t=1}^{T} K_h(u-t/T) E\left[\frac{\partial \mathrm{vec}(\Omega_t)^\top}{\partial\theta}\left(\Omega_t^{-1}\otimes\Omega_t^{-1}\right)\frac{\partial \mathrm{vec}(\Omega_t)}{\partial\theta^\top}\right] \\
&= \frac{1}{T}\sum_{t=1}^{T} K_h(u-t/T) E\left[W_t(u)W_t(u)^\top\right] \\
&= E\left[W_t(u)W_t(u)^\top\right] + o_p(1) \\
&= 4D_N^\top E(\Omega_t^{-1}\otimes G_t)D_N + o_p(1) \\
&= \Lambda(u) + o_p(1),
\end{aligned}$$

In conclusion,

$$\sqrt{Th}\left(\widehat{\theta}(u)-\theta(u)\right) \Longrightarrow N(0, \|K\|_2^2\,\Lambda(u)^{-1}\Psi(u)\Lambda(u)^{-1}).$$

Then note that

$$\mathrm{vec}(\Sigma(u)) = (I_N\otimes\Theta(u))\,\mathrm{vec}(\Theta(u)) = (\Theta(u)\otimes I_N)\,\mathrm{vec}(\Theta(u)).$$

Therefore,

$$\sqrt{Th}(\widehat{\sigma}(u)-\sigma(u)) = \Delta(u)\sqrt{Th}\left(\widehat{\theta}(u)-\theta(u)\right) + o_p(1)$$

and so (19) follows. ■

D Proof of Theorem 3

Define

$$\widehat{L}_T(\theta;u,\phi) = \sum_{t=1}^{T} K_h(u-t/T)\,l(\widehat{\Omega}_t(\theta,\phi);y_t),$$

where $\widehat{\Omega}_t(\theta,,\phi) = \Theta\widehat{G}_t(\phi)\Theta$ and

$$\widehat{G}_t(\phi) = I_N - AA^\top - BB^\top + A\widetilde{\Sigma}(t-1/T)^{-1/2}y_{t-1}y_{t-1}^\top\widetilde{\Sigma}(t-1/T)^{-1/2}A^\top + B\widehat{G}_{t-1}(\phi)B^\top,$$

where $\widetilde{\Sigma}(.)$ is the preliminary estimator and $\widehat{G}_1(\phi)$ is given. As before we show that there exists a deterministic function $L(\theta;u,\phi)$ such that

$$\sup_{u\in[0,1]}\sup_{\theta,\phi}\left|\widehat{L}_T(\theta;u,\phi) - L(\theta;u,\phi)\right| = o_p(1).$$

Then the profiled estimator $\widehat{\theta}_\phi(u)$ satisfies

$$\sup_{u\in[0,1]}\sup_{\phi\in\Phi}\left|\widehat{\theta}_\phi(u)-\theta_\phi(u)\right| = o_p(1).$$

We can show that

$$\sup_{\phi\in\Phi}\left|\widehat{\ell}_T(\phi)-\ell(\phi)\right| = o_p(1),$$

so that $\widehat{\phi}$ is consistent. Then we obtain an expansion for $\widehat{\Sigma}_\phi(u)$ around its probability limit $\Sigma_\phi(u)$.

Let

$$\overline{u}_t(u) = H_t(u)\text{vec}\left(\varepsilon_t\varepsilon_t^\top - I_N\right),$$

where the matrix $H_t(u)$ is measurable with respect to $\mathcal{F}_{t-1}$.

Lemma 4. *There exists a bounded continuous function $\overline{B}_\phi(u)$ such that for a sequence $\delta_T \to 0$ we have*

$$\sup_{||\phi-\phi_0||\le\delta_T}\sup_{u\in[0,1]}\left\|\widehat{\sigma}(u)-\sigma(u)-\frac{1}{T}\sum_{t=1}^T K_h(u-t/T)\overline{u}_t(u)-h^2\overline{B}_\phi(u)\right\| = O\left(\frac{\log T}{Th}\right)+o(h^2) \text{ a.s.}$$

Lemma 5. *As $T\to\infty$, the following expansion holds*

$$\sqrt{T}\left(\widehat{\phi}-\phi\right) = -\left[\frac{\partial^2\widehat{\ell}_T(\phi_0)}{\partial\phi\partial\phi^\top}\right]^{-1}\sqrt{T}\frac{\partial\widehat{\ell}_T(\phi_0)}{\partial\phi}+o_p(1) \tag{27}$$

$$\sqrt{T}\frac{\partial\widehat{\ell}_T(\phi_0)}{\partial\phi} \Longrightarrow N(0,Q^e) \tag{28}$$

$$\frac{\partial^2\widehat{\ell}_T(\phi_0)}{\partial\phi\partial\phi^\top} \xrightarrow{P} J^e. \tag{29}$$

■

E Proof of Lemmas

Here we give sketch proofs of Lemmas 1-5.

Proof of Lemma 1. We have

$$\begin{aligned}
\widetilde{\Sigma}(u)-\Sigma(u) &= \frac{\sum_{t=1}^T K_h(u-t/T)\left[y_ty_t^\top-\Sigma(t/T)\right]}{\sum_{t=1}^T K_h(u-t/T)}+\frac{\sum_{t=1}^T K_h(u-t/T)\left[\Sigma(t/T)-\Sigma(u)\right]}{\sum_{t=1}^T K_h(u-t/T)} \\
&= \frac{1}{Th}\sum_{t=1}^T K_h(u-t/T)U_{tT}+\frac{h^2}{2}\Sigma''(u)\int s^2K(s)ds+o(h^2)+o_p(T^{-1/2}h^{-1/2}) \\
& \quad \frac{1}{Th}\sum_{t=1}^T K_h(u-t/T)U_t(u)+\frac{h^2}{2}\Sigma''(u)\int s^2K(s)ds+o(h^2)+o_p(T^{-1/2}h^{-1/2})
\end{aligned}$$

by standard kernel arguments using the smoothness of $\Sigma(.)$ and the fact that

$$\frac{1}{Th}\sum_{t=1}^{T} K_h(u-t/T) = 1 + O(T^{-1}h^{-1}).$$

■

Proof of Lemma 2. By the triangle inequality

$$\sup_{\phi\in\Phi}\left|T^{-1}\widetilde{\ell}_T(\phi)-\ell(\phi)\right| \le \sup_{\phi\in\Phi}\left|T^{-1}\widetilde{\ell}_T(\phi)-T^{-1}\ell_T(\phi)\right| + \sup_{\phi\in\Phi}\left|T^{-1}\ell_T(\phi)-\ell(\phi)\right|.$$

It follows from standard results that

$$\sup_{\phi\in\Phi}\left|T^{-1}\ell_T(\phi)-\ell(\phi)\right| = o_p(1). \tag{30}$$

We shall show that

$$\sup_{\phi\in\Phi}\left|T^{-1}\widetilde{\ell}_T(\phi)-T^{-1}\ell_T(\phi)\right| = o_p(1). \tag{31}$$

This then implies consistency of $\widetilde{\phi}$ by the identifiability condition.

Write

$$\begin{aligned}
\widetilde{G}_t(\phi)-G_t(\phi) &= A\Delta_{t-1}A^\top + B\left[\widetilde{G}_{t-1}-G_{t-1}\right]B^\top \\
&= A\Delta_{t-1}A^\top + B\left[A\Delta_{t-2}A^\top\right]B^\top + B^2\left[\widetilde{G}_{t-2}-G_{t-2}\right]B^{\top 2} \\
&= \sum_{j=1}^{t-1} B^{j-1}\left[A\Delta_{t-j}A^\top\right](B^\top)^{j-1},
\end{aligned}$$

$$\Delta_{t-j} = \widetilde{\Sigma}(t-j/T)^{-1/2}y_{t-j}y_{t-j}^\top\widetilde{\Sigma}(t-j/T)^{-1/2} - \Sigma(t-j/T)^{-1/2}y_{t-j}y_{t-j}^\top\Sigma(t-j/T)^{-1/2}.$$

Then, since

$$\begin{aligned}
\Delta_{t-j} &= \left[\widetilde{\Sigma}(t-j/T)^{-1/2}-\Sigma(t-j/T)^{-1/2}\right]y_{t-j}y_{t-j}^\top\Sigma(t-j/T)^{-1/2} \\
&\quad +\Sigma(t-j/T)^{-1/2}y_{t-j}y_{t-j}^\top\left[\widetilde{\Sigma}(t-j/T)^{-1/2}-\Sigma(t-j/T)^{-1/2}\right] \\
&\quad +\left[\widetilde{\Sigma}(t-j/T)^{-1/2}-\Sigma(t-j/T)^{-1/2}\right]y_{t-j}y_{t-j}^\top\left[\widetilde{\Sigma}(t-j/T)^{-1/2}-\Sigma(t-j/T)^{-1/2}\right],
\end{aligned}$$

we have

$$\begin{aligned}
&\widetilde{G}_t(\phi)-G_t(\phi) \\
=\ &\sum_{j=1}^{t-1} B^{j-1}\left[\widetilde{\Sigma}(t-j/T)^{-1/2}-\Sigma(t-j/T)^{-1/2}\right]y_{t-j}y_{t-j}^\top\Sigma(t-j/T)^{-1/2}(B^\top)^{j-1} \\
&+\sum_{j=1}^{t-1} B^{j-1}\Sigma(t-j/T)^{-1/2}y_{t-j}y_{t-j}^\top\left[\widetilde{\Sigma}(t-j/T)^{-1/2}-\Sigma(t-j/T)^{-1/2}\right](B^\top)^{j-1} \\
&+\sum_{j=1}^{t-1} B^{j-1}\left[\widetilde{\Sigma}(t-j/T)^{-1/2}-\Sigma(t-j/T)^{-1/2}\right]y_{t-j}y_{t-j}^\top\left[\widetilde{\Sigma}(t-j/T)^{-1/2}-\Sigma(t-j/T)^{-1/2}\right](B^\top)^{j-1}.
\end{aligned}$$

Note that

$$\frac{1}{T}\sum_{t=1}^{T} y_t^\top \widetilde{G}_t^{-1}(\phi) y_t - \frac{1}{T}\sum_{t=1}^{T} y_t^\top G_t^{-1}(\phi) y_t = -\frac{1}{T}\sum_{t=1}^{T} y_t^\top G_t^{-1}(\phi)\left[\widetilde{G}_t(\phi) - G_t(\phi)\right]\widetilde{G}_t^{-1}(\phi) y_t,$$

whence it suffices to show that

$$\max_{1\le t\le T}\sup_{\phi\in\Phi}\left\|\widetilde{G}_t(\phi) - G_t(\phi)\right\| = o_p(1)$$

$$\min_{1\le t\le T}\inf_{\phi\in\Phi}\lambda_{\min}(G_t(\phi)) > 0.$$

The first property follows from Lemma 1 and the mapping $\phi \mapsto G_t(\phi)$, and the second follows by assumption on Φ. ■

Proof of Lemma 3. We will consider the quantities:

$$\begin{aligned}
\sqrt{T}\frac{\partial \ell_T(\phi_0)}{\partial \phi_i} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathrm{Tr}\left[(I_N - u_t u_t^\top G_t^{-1})\frac{\partial G_t}{\partial \phi_i} G_t^{-1}\right] \\
\sqrt{T}\frac{\partial \widetilde{\ell}_T(\phi_0)}{\partial \phi_i} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathrm{Tr}\left[(I_N - \widetilde{u}_t \widetilde{u}_t^\top \widetilde{G}_t^{-1})\frac{\partial \widetilde{G}_t}{\partial \phi_i} \widetilde{G}_t^{-1}\right].
\end{aligned}$$

Then, $\widetilde{G}_t^{-1} = G_t^{-1} - G_t^{-1}(\widetilde{G}_t - G_t)G_t^{-1} + o_p(T^{-1/2})$. Also,

$$\widetilde{\Sigma}(u) = \Sigma^{1/2}(u)\left(I + \Sigma^{-1/2}(u)(\widetilde{\Sigma}(u) - \Sigma(u))\Sigma^{-1/2}(u)\right)\Sigma^{1/2}(u) + o_p(T^{-1/2})$$

for any $u \in [0,1]$, so that

$$\begin{aligned}
\widetilde{\Sigma}^{-1/2}(u) &= \Sigma^{-1/4}(u)\left(I + \Sigma^{-1/2}(u)(\widetilde{\Sigma}(u) - \Sigma(u))\Sigma^{-1/2}(u)\right)^{-1/2}\Sigma^{-1/4}(u) + o_p(T^{-1/2}) \\
&= \Sigma^{-1/4}(u)\left(I - \frac{1}{2}\Sigma^{-1/2}(u)(\widetilde{\Sigma}(u) - \Sigma(u))\Sigma^{-1/2}(u)\right)\Sigma^{-1/4}(u) + o_p(T^{-1/2}) \\
&= \Sigma^{-1/2}(u) - \frac{1}{2}\Sigma^{-3/4}(u)(\widetilde{\Sigma}(u) - \Sigma(u))\Sigma^{-3/4}(u) + o_p(T^{-1/2}).
\end{aligned}$$

Therefore,

$$\begin{aligned}
& \widetilde{u}_t\widetilde{u}_t^\top - u_t u_t^\top \\
=\ & \widetilde{\Sigma}^{-1/2}(t/T) y_t y_t^\top \widetilde{\Sigma}^{-1/2}(t/T) - \Sigma^{-1/2}(t/T) y_t y_t^\top \Sigma^{-1/2}(t/T) \\
=\ & -\frac{1}{2}\Sigma^{-3/4}(\widetilde{\Sigma} - \Sigma)\Sigma^{-3/4}(t/T) y_t y_t^\top \Sigma^{-1/2}(t/T) - \frac{1}{2}\Sigma^{-1/2}(t/T) y_t y_t^\top \Sigma^{-3/4}(\widetilde{\Sigma} - \Sigma)\Sigma^{-3/4}(t/T) + o_p(T^{-1/2}) \\
=\ & -\frac{1}{2}\Sigma^{-3/4}(\widetilde{\Sigma} - \Sigma)\Sigma^{-1/4}(t/T) G_t^{1/2}\varepsilon_t\varepsilon_t^\top G_t^{1/2} - \frac{1}{2}G_t\varepsilon_t\varepsilon_t^\top \Sigma^{-1/4}(\widetilde{\Sigma} - \Sigma)\Sigma^{-3/4}(t/T) + o_p(T^{-1/2}) \\
=\ & -\frac{1}{2}\Sigma^{-3/4}(\widetilde{\Sigma} - \Sigma)\Sigma^{-1/4}(t/T) G_t - \frac{1}{2}G_t\Sigma^{-1/4}(\widetilde{\Sigma} - \Sigma)\Sigma^{-3/4}(t/T) \\
& -\frac{1}{2}\Sigma^{-3/4}(\widetilde{\Sigma} - \Sigma)\Sigma^{-1/4}(t/T) G_t^{1/2}\left[\varepsilon_t\varepsilon_t^\top - I_N\right] G_t^{1/2} - \frac{1}{2}G_t\left[\varepsilon_t\varepsilon_t^\top - I_N\right]\Sigma^{-1/4}(\widetilde{\Sigma} - \Sigma)\Sigma^{-3/4}(t/T) \\
& + o_p(T^{-1/2})
\end{aligned}$$

We have

$$
\begin{aligned}
\sqrt{T}\frac{\partial\widetilde{\ell}_T(\phi_0)}{\partial\phi_i} &= \sqrt{T}\frac{\partial\ell_T(\phi_0)}{\partial\phi_i} \\
&\quad+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathrm{Tr}\left[(I_N-u_tu_t^\top G_t^{-1})\left(\frac{\partial\widetilde{G}_t}{\partial\phi_i}-\frac{\partial G_t}{\partial\phi_i}\right)G_t^{-1}\right] \\
&\quad-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathrm{Tr}\left[(I_N-u_tu_t^\top G_t^{-1})\frac{\partial G_t}{\partial\phi_i}\left(G_t^{-1}(\widetilde{G}_t-G_t)G_t^{-1}\right)\right] \\
&\quad+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathrm{Tr}\left[u_tu_t^\top\left(G_t^{-1}(\widetilde{G}_t-G_t)G_t^{-1}\right)\frac{\partial G_t}{\partial\phi_i}G_t^{-1}\right] \\
&\quad-\frac{1}{2\sqrt{T}}\sum_{t=1}^{T}\mathrm{Tr}\left[\Sigma^{-3/4}(\widetilde{\Sigma}-\Sigma)\Sigma^{-1/4}(t/T)\frac{\partial G_t}{\partial\phi_i}G_t^{-1}\right] \\
&\quad-\frac{1}{2\sqrt{T}}\sum_{t=1}^{T}\mathrm{Tr}\left[\Sigma^{-1/4}(\widetilde{\Sigma}-\Sigma)\Sigma^{-3/4}(t/T)G_t^{-1}\frac{\partial G_t}{\partial\phi_i}\right]+o_p(1) \\
&= \sqrt{T}\frac{\partial\ell_T(\phi_0)}{\partial\phi_i} \\
&\quad+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathrm{Tr}\left[G_t\left(G_t^{-1}(\widetilde{G}_t-G_t)G_t^{-1}\right)\frac{\partial G_t}{\partial\phi_i}G_t^{-1}\right] \\
&\quad-\frac{1}{2\sqrt{T}}\sum_{t=1}^{T}\mathrm{Tr}\left[\Sigma^{-3/4}(\widetilde{\Sigma}-\Sigma)\Sigma^{-1/4}(t/T)\frac{\partial G_t}{\partial\phi_i}G_t^{-1}\right] \\
&\quad-\frac{1}{2\sqrt{T}}\sum_{t=1}^{T}\mathrm{Tr}\left[\Sigma^{-1/4}(\widetilde{\Sigma}-\Sigma)\Sigma^{-3/4}(t/T)G_t^{-1}\frac{\partial G_t}{\partial\phi_i}\right] \\
&\quad+o_p(1),
\end{aligned}
$$

because $(I_N-u_tu_t^\top G_t^{-1})$ is a martingale difference sequence. Actually, the above steps need some justification. Then note that

$$
\mathrm{Tr}\left[\Sigma^{-1/4}(\widetilde{\Sigma}-\Sigma)\Sigma^{-3/4}(t/T)G_t^{-1}\frac{\partial G_t}{\partial\phi_i}\right]=\mathrm{Tr}\left[\Sigma^{-3/4}(\widetilde{\Sigma}-\Sigma)\Sigma^{-1/4}(t/T)\frac{\partial G_t}{\partial\phi_i}G_t^{-1}\right]
$$

to obtain that

$$
\begin{aligned}
\sqrt{T}\frac{\partial\widetilde{\ell}_T(\phi_0)}{\partial\phi_i} &= \sqrt{T}\frac{\partial\ell_T(\phi_0)}{\partial\phi_i} \\
&\quad+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathrm{Tr}\left[G_t^{-1}\frac{\partial G_t}{\partial\phi_i}G_t^{-1}(\widetilde{G}_t-G_t)\right] \\
&\quad-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathrm{Tr}\left[\Sigma^{-3/4}(\widetilde{\Sigma}-\Sigma)\Sigma^{-1/4}(t/T)\frac{\partial G_t}{\partial\phi_i}G_t^{-1}\right] \\
&\quad+o_p(1),
\end{aligned}
$$

Consider

$$
\begin{aligned}
& \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathrm{Tr}\left[\Sigma^{-3/4}(\widetilde{\Sigma}-\Sigma)\Sigma^{-1/4}(t/T)\frac{\partial G_t}{\partial\phi_i}G_t^{-1}\right] \\
= & \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathrm{Tr}\left[\Sigma^{-3/4}\frac{1}{T}\sum_{s=1}^{T}K_h((t-s)/T)U_{sT}\Sigma^{-1/4}(t/T)\frac{\partial G_t}{\partial\phi_i}G_t^{-1}\right]+o_p(1) \\
= & \frac{1}{\sqrt{T}}\sum_{s=1}^{T}\mathrm{Tr}\left[\frac{1}{T}\sum_{t=1}^{T}K_h((t-s)/T)\Sigma^{-3/4}(t/T)U_{sT}\Sigma^{-1/4}(t/T)\frac{\partial G_t}{\partial\phi_i}G_t^{-1}\right]+o_p(1) \\
= & \frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=1}^{T}K_h((t-s)/T)\mathrm{vec}\left(G_t^{-1}\frac{\partial G_t}{\partial\phi_i}\right)^{\top}\left[\Sigma^{-1/4}(t/T)\otimes\Sigma^{-3/4}(t/T)\right]\mathrm{vec}\left(U_{sT}\right)+o_p(1) \\
= & \frac{1}{\sqrt{T}}\sum_{s=1}^{T}\mathrm{vec}\left(E\left[G_t^{-1}\frac{\partial G_t}{\partial\phi_i}\right]\right)^{\top}\left[\Sigma^{-1/4}(s/T)\otimes\Sigma^{-3/4}(s/T)\right]\mathrm{vec}\left(U_{sT}\right)+o_p(1) \\
= & \mathrm{vec}\left(E\left[G_t^{-1}\frac{\partial G_t}{\partial\phi_i}\right]\right)^{\top}\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\left[\Sigma^{1/4}(s/T)\otimes\Sigma^{-1/4}(s/T)\right]\mathrm{vec}\left[(G_s^{1/2}\varepsilon_s\varepsilon_s^{\top}G_s^{1/2}-I_N)\right]+o_p(1) \qquad (32)
\end{aligned}
$$

using $\mathrm{Tr}(ABCD)=\mathrm{vec}(D^{\top})^{\top}(C^{\top}\otimes A)\mathrm{vec}(B)$, and

$$
\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}K_h((t-s)/T)\mathrm{vec}\left(G_t^{-1}\frac{\partial G_t}{\partial\phi_i}\right)^{\top}\left[\Sigma^{-1/4}(t/T)\otimes\Sigma^{-3/4}(t/T)\right] \\
= & \frac{1}{T}\sum_{t=1}^{T}K_h((t-s)/T)\mathrm{vec}\left(E\left[G_t^{-1}\frac{\partial G_t}{\partial\phi_i}\right]\right)^{\top}\left[\Sigma^{-1/4}(t/T)\otimes\Sigma^{-3/4}(t/T)\right]+o_p(1) \qquad (33) \\
= & \mathrm{vec}\left(E\left[G_t^{-1}\frac{\partial G_t}{\partial\phi_i}\right]\right)^{\top}\left[\Sigma^{-1/4}(s/T)\otimes\Sigma^{-3/4}(s/T)\right]+o_p(1). \qquad (34)
\end{aligned}
$$

The term (32) is asymptotically normal. Now consider

$$\widetilde{G}_t-G_t=\sum_{j=1}^{t-1}B^{j-1}A\Delta_{t-j}A^{\top}(B^{\top})^{j-1},$$

where

$$\begin{aligned}
\Delta_{t-j} &= \widetilde{u}_{t-j}\widetilde{u}_{t-j}^\top - u_{t-j}u_{t-j}^\top \\
&= \widetilde{\Sigma}(t-j/T)^{-1/2} y_{t-j}y_{t-j}^\top \widetilde{\Sigma}(t-j/T)^{-1/2} - \Sigma(t-j/T)^{-1/2} y_{t-j}y_{t-j}^\top \Sigma(t-j/T)^{-1/2} \\
&\quad -\frac{1}{2}\Sigma^{-3/4}(\widetilde{\Sigma}-\Sigma)\Sigma^{-1/4}(t-j/T)G_{t-j}^{1/2}\varepsilon_{t-j}\varepsilon_{t-j}^\top G_{t-j}^{1/2} \\
&\quad -\frac{1}{2}G_{t-j}^{1/2}\varepsilon_{t-j}\varepsilon_{t-j}^\top G_{t-j}^{1/2}\Sigma^{-1/4}(\widetilde{\Sigma}-\Sigma)\Sigma^{-3/4}(t-j/T) + o_p(T^{-1/2}) \\
&= -\frac{1}{2}\Sigma^{-3/4}(\widetilde{\Sigma}-\Sigma)\Sigma^{-1/4}(t-j/T)G_{t-j} - \frac{1}{2}G_{t-j}\Sigma^{-1/4}(\widetilde{\Sigma}-\Sigma)\Sigma^{-3/4}(t-j/T) \\
&\quad -\frac{1}{2}\Sigma^{-3/4}(\widetilde{\Sigma}-\Sigma)\Sigma^{-1/4}(t-j/T)G_{t-j}^{1/2}\left[\varepsilon_{t-j}\varepsilon_{t-j}^\top - I_N\right]G_{t-j}^{1/2} \\
&\quad -\frac{1}{2}G_{t-j}^{1/2}\left[\varepsilon_{t-j}\varepsilon_{t-j}^\top - I_N\right]G_{t-j}^{1/2}\Sigma^{-1/4}(\widetilde{\Sigma}-\Sigma)\Sigma^{-3/4}(t-j/T) + o_p(T^{-1/2}).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\widetilde{G}_t - G_t &= -\frac{1}{2}\sum_{j=1}^{t-1} B^{j-1}A\Sigma^{-3/4}(\widetilde{\Sigma}-\Sigma)\Sigma^{-1/4}(t-j/T)G_{t-j}A^\top(B^\top)^{j-1} \\
&\quad -\frac{1}{2}\sum_{j=1}^{t-1} B^{j-1}AG_{t-j}\Sigma^{-1/4}(\widetilde{\Sigma}-\Sigma)\Sigma^{-3/4}(t-j/T)A^\top(B^\top)^{j-1} + o_p(T^{-1/2})
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathrm{Tr}\left[G_t^{-1}(\widetilde{G}_t - G_t)G_t^{-1}\frac{\partial G_t}{\partial \phi_i}\right] \\
=\ & \frac{1}{2T\sqrt{T}}\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{j=1}^{t-1}K_h\left(\frac{t-j-s}{T}\right)\mathrm{Tr}\left[B^{j-1}A\Sigma^{-3/4}U_{sT}\Sigma^{-1/4}(t-j/T)G_{t-j}A^\top(B^\top)^{j-1}G_t^{-1}\frac{\partial G_t}{\partial \phi_i}G_t^{-1}\right] \\
&+\frac{1}{2T\sqrt{T}}\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{j=1}^{t-1}K_h\left(\frac{t-j-s}{T}\right)\mathrm{Tr}\left[B^{j-1}AG_{t-j}\Sigma^{-1/4}U_{sT}\Sigma^{-3/4}(t-j/T)A^\top(B^\top)^{j-1}G_t^{-1}\frac{\partial G_t}{\partial \phi_i}G_t^{-1}\right] \\
&+o_p(1).
\end{aligned}$$

Then

$$\begin{aligned}
&\mathrm{Tr}\left[G_t^{-1}\frac{\partial G_t}{\partial \phi_i}G_t^{-1}B^{j-1}A\Sigma^{-3/4}(t-j/T)U_{sT}\Sigma^{-1/4}(t-j/T)G_{t-j}A^\top(B^\top)^{j-1}\right] \\
=\ & \mathrm{vec}\left(B^{j-1}A\right)^\top\left[G_{t-j}\otimes G_t^{-1}\frac{\partial G_t}{\partial \phi_i}G_t^{-1}\right]\mathrm{vec}(B^{j-1}A\Sigma^{-3/4}(t-j/T)U_{sT}\Sigma^{-1/4}(t-j/T)) \\
=\ & \mathrm{vec}\left(B^{j-1}A\right)^\top\left[G_{t-j}\otimes G_t^{-1}\frac{\partial G_t}{\partial \phi_i}G_t^{-1}\right]\left[\Sigma^{-1/4}(t-j/T)\otimes B^{j-1}A\Sigma^{-3/4}(t-j/T)\right]\mathrm{vec}(U_{sT}),
\end{aligned}$$

$$
\begin{aligned}
& \operatorname{Tr}\left[G_{t-j}\Sigma^{-1/4}U_{sT}\Sigma^{-3/4}(t-j/T)A^{\top}(B^{\top})^{j-1}G_t^{-1}\frac{\partial G_t}{\partial \phi_i}G_t^{-1}B^{j-1}A\right] \\
= \; & \operatorname{vec}\left(B^{\top j-1}A^{\top}\right)^{\top}\left[G_t^{-1}\frac{\partial G_t}{\partial \phi_i}G_t^{-1}\otimes G_{t-j}\right]\operatorname{vec}(\Sigma^{-1/4}U_{sT}\Sigma^{-3/4}(t-j/T)A^{\top}(B^{\top})^{j-1}) \\
= \; & \operatorname{vec}\left(B^{\top j-1}A^{\top}\right)^{\top}\left[G_t^{-1}\frac{\partial G_t}{\partial \phi_i}G_t^{-1}\otimes G_{t-j}\right]\left[B^{j-1}A\Sigma^{-3/4}(t-j/T)\otimes\Sigma^{-1/4}\right]\operatorname{vec}(U_{sT}).
\end{aligned}
$$

So we need the probability limit of

$$
\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t-1}K_h\left(\frac{t-j-s}{T}\right)\operatorname{vec}\left(B^{j-1}A\right)^{\top}\left[G_{t-j}\otimes G_t^{-1}\frac{\partial G_t}{\partial \phi_i}G_t^{-1}\right] \\
& \times\left[\Sigma^{-1/4}\left(\frac{t-j}{T}\right)\otimes B^{j-1}A\Sigma^{-3/4}\left(\frac{t-j}{T}\right)\right] \\
= \; & \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t-1}K_h\left(\frac{t-j-s}{T}\right)\operatorname{vec}\left(B^{j-1}A\right)^{\top}\mathcal{M}_j^i\left[\Sigma^{-1/4}\left(\frac{t-j}{T}\right)\otimes B^{j-1}A\Sigma^{-3/4}\left(\frac{t-j}{T}\right)\right]+o_p(1) \\
= \; & \sum_{j=1}^{\infty}\operatorname{vec}\left(B^{j-1}A\right)^{\top}\mathcal{M}_j^i\left[\Sigma^{-1/4}(s/T)\otimes B^{j-1}A\Sigma^{-3/4}(s/T)\right]+o_p(1),
\end{aligned}
$$

where

$$
\mathcal{M}_j^i = E\left[G_{t-j}\otimes G_t^{-1}\frac{\partial G_t}{\partial \phi_i}G_t^{-1}\right].
$$

Then recalling that $\operatorname{vec}(U_{tT}) = [\Sigma(t/T)^{1/2}\otimes\Sigma(t/T)^{1/2}]\operatorname{vec}(G_t^{1/2}\varepsilon_t\varepsilon_t^{\top}G_t^{1/2}-I_N)$, we have

$$
\begin{aligned}
& \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\operatorname{Tr}\left[(\widetilde{G}_t-G_t)G_t^{-1}\frac{\partial G_t}{\partial \phi_i}G_t^{-1}\right] \\
= \; & \frac{1}{\sqrt{T}}\sum_{t=1}^{T}C_i(t/T)\operatorname{vec}(G_t^{1/2}\varepsilon_t\varepsilon_t^{\top}G_t^{1/2}-I_N),
\end{aligned}
$$

where

$$
\begin{aligned}
C_i(t/T) \;=\; & \sum_{j=1}^{\infty}\operatorname{vec}\left(B^{j-1}A\right)^{\top}\mathcal{M}_j^i\left[I_N\otimes B^{j-1}A\right]\left[\Sigma^{1/4}(s/T)\otimes\Sigma^{-1/4}(s/T)\right] \\
& +\sum_{j=1}^{\infty}\operatorname{vec}\left(B^{\top j-1}A^{\top}\right)^{\top}\mathcal{M}_j^{*i}\left[B^{j-1}A\otimes I_N\right]\left[\Sigma^{-1/4}(s/T)\otimes\Sigma^{1/4}(s/T)\right],
\end{aligned}
$$

where

$$
\mathcal{M}_j^{*i} = E\left[G_t^{-1}\frac{\partial G_t}{\partial \phi_i}G_t^{-1}\otimes G_{t-j}\right].
$$

This term is also asymptotically normal. In conclusion we have

$$\begin{aligned}
\sqrt{T}\frac{\partial \widetilde{\ell}_T(\phi_0)}{\partial \phi_i} &= -\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \text{vec}\left(G_t^{-1/2}\frac{\partial G_t}{\partial \phi_i}G_t^{-1/2}\right)^{\top} \text{vec}\left(\varepsilon_t\varepsilon_t^{\top} - I_N\right) \\
&+\frac{1}{\sqrt{T}}\sum_{t=1}^{T} C_i(t/T) \times \text{vec}(G_t^{1/2}\varepsilon_t\varepsilon_t^{\top}G_t^{1/2} - I_N) \\
&-\text{vec}\left(E\left[G_t^{-1}\frac{\partial G_t}{\partial \phi_i}\right]\right)^{\top} \frac{1}{\sqrt{T}}\sum_{s=1}^{T} \left[\Sigma^{1/4}(s/T) \otimes \Sigma^{-1/4}(s/T)\right] \text{vec}\left[(G_s^{1/2}\varepsilon_s\varepsilon_s^{\top}G_s^{1/2} - I_N)\right] \\
&+o_p(1)
\end{aligned}$$

and

$$\begin{aligned}
\sqrt{T}\frac{\partial \widetilde{\ell}_T(\phi_0)}{\partial \phi} &= -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{\zeta}_t + \frac{1}{\sqrt{T}}\sum_{t=1}^{T} C(t/T)z_t \\
&-E[\rho_t]\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\left[\Sigma^{1/4}(s/T) \otimes \Sigma^{-1/4}(s/T)\right] z_s + o_p(1) \\
&\equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \aleph_t + o_p(1),
\end{aligned}$$

where $\tilde{\zeta}_t = \rho_t \text{vec}\left(\varepsilon_t\varepsilon_t^{\top} - I_N\right)$ is a martingale difference sequence. Furthermore, ρ_t is defined in (13), z_t in (3), $C(t/T) = (C_1(t/T)^{\top}, \ldots, C_p(t/T)^{\top})^{\top}$, and

$$\begin{aligned}
\aleph_t &= \tilde{\zeta}_t + \Upsilon(t/T)z_t \\
\Upsilon(t/T) &= C(t/T) - E[\rho_t]\left[\Sigma^{1/4}(s/T) \otimes \Sigma^{-1/4}(s/T)\right].
\end{aligned}$$

We have

$$\frac{1}{\sqrt{T}}\sum_{s=1}^{T} \Upsilon(t/T)z_t \Longrightarrow N(0, \Psi),$$

Letting $\Gamma_u = E[z_s z_{s+u}^\top]$ as defined in (10),

$$\begin{aligned}
\Psi &= \lim_{T\to\infty}\frac{1}{T}E\left[\left(\sum_{s=1}^{T}\Upsilon(s/T)z_s\right)\left(\sum_{s=1}^{T}\Upsilon(s/T)z_s\right)^\top\right] \\
&= \lim_{T\to\infty}\frac{1}{T}\left[\sum_{s=1}^{T}\Upsilon(s/T)E\left[z_s z_s^\top\right]\Upsilon(s/T)^\top + \sum_{\substack{s=1\\ s\neq s'}}^{T}\sum_{s'=1}^{T}\Upsilon(s/T)E\left[z_s z_{s'}^\top\right]\Upsilon(s'/T)^\top\right] \\
&= \lim_{T\to\infty}\frac{1}{T}\left[\sum_{s=1}^{T}\Upsilon(s/T)\Gamma_0\Upsilon(s/T)^\top + \sum_{s=1}^{T-1}\sum_{s'=s+1}^{T}\Upsilon(s/T)(\Gamma_{s'-s}+\Gamma_{s-s'}^\top)\Upsilon(s'/T)^\top\right] \\
&= \lim_{T\to\infty}\frac{1}{T}\left[\sum_{s=1}^{T}\Upsilon(s/T)\Gamma_0\Upsilon(s/T)^\top + \sum_{s=1}^{T-1}\Upsilon(s/T)\sum_{u=1}^{T-s}(\Gamma_u+\Gamma_{-u}^\top)\Upsilon((s+u)/T)^\top\right] \\
&= \lim_{T\to\infty}\frac{1}{T}\sum_{s=1}^{T}\Upsilon(s/T)\Gamma_0\Upsilon(s/T)^\top + \lim_{T\to\infty}\frac{1}{T}\sum_{s=1}^{T}\Upsilon(s/T)\sum_{u=\pm 1}^{\pm T-1}\Gamma_u\Upsilon(s/T)^\top
\end{aligned}$$

by a Taylor expansion provided $\sum_{u=-\infty}^{\infty} u||\Gamma_u|| < \infty$, so that

$$\mathrm{vec}(\Psi) = \int_0^1 [\Upsilon(u)\otimes\Upsilon(u)]\,du\sum_{u=-\infty}^{\infty}\mathrm{vec}(\Gamma_u).$$

Now define $\tilde{\Gamma}_u = E[\tilde{\zeta}_t z_{t+u}^\top]$ and note that $\tilde{\Gamma}_u = 0$ for $u < 0$. Then,

$$\begin{aligned}
H &= \lim_{T\to\infty}\frac{1}{T}E\left[\left(\sum_{s=1}^{T}\tilde{\zeta}_s\right)\left(\sum_{s=1}^{T}\Upsilon(s/T)z_s\right)^\top\right] \\
&= \lim_{T\to\infty}\frac{1}{T}\left[\sum_{s=1}^{T}E\left[\tilde{\zeta}_s z_s^\top\right]\Upsilon(s/T)^\top + \sum_{s=1}^{T-1}\sum_{s'=s+1}^{T}E\left[\tilde{\zeta}_s z_{s'}^\top\right]\Upsilon(s'/T)^\top\right] \\
&= \lim_{T\to\infty}\frac{1}{T}\left[\sum_{s=1}^{T}\tilde{\Gamma}_0\Upsilon(s/T)^\top + \sum_{s=1}^{T-1}\sum_{s'=s+1}^{T}\tilde{\Gamma}_{s'-s}\Upsilon(s'/T)^\top\right] \\
&= \lim_{T\to\infty}\frac{1}{T}\left[\sum_{s=1}^{T}\tilde{\Gamma}_0\Upsilon(s/T)^\top + \sum_{s=1}^{T-1}\sum_{u=1}^{T-s}\tilde{\Gamma}_u\Upsilon((s+u)/T)^\top\right] \\
&= \lim_{T\to\infty}\frac{1}{T}\sum_{s=1}^{T}\tilde{\Gamma}_0\Upsilon(s/T)^\top + \lim_{T\to\infty}\frac{1}{T}\sum_{s=1}^{T}\sum_{u=\pm 1}^{\pm T-1}\tilde{\Gamma}_u\Upsilon(s/T)^\top
\end{aligned}$$

by a Taylor expansion provided $\sum_{u=-\infty}^{\infty} u||\tilde{\Gamma}_u|| < \infty$, so that

$$\mathrm{vec}(H) = \int_0^1 [\Upsilon(u)\otimes I_p]\,du\sum_{u=-\infty}^{\infty}\mathrm{vec}(\tilde{\Gamma}_u).$$

Finally, we obtain the asymptotic covariance matrix,

$$Q = \lim_{T\to\infty} \text{var}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \aleph_t = \Phi + H + H^\top + E[\rho_t \Xi_t \rho_t^\top] \tag{35}$$

where Ξ_t is defined in (17).

Note also that

$$\frac{\partial^2 \widehat{\ell}_T(\phi_0)}{\partial\phi\partial\phi^\top} = \frac{1}{T}\sum_{t=1}^{T} E\left[\frac{\partial^2 \ell_t(\phi_0)}{\partial\phi\partial\phi^\top}\right] + o_p(1),$$

where

$$\begin{aligned}\frac{\partial^2 l_t}{\partial\phi_i\partial\phi_j} &= \text{Tr}\left[\frac{\partial^2 G_t}{\partial\phi_i\partial\phi_j}G_t^{-1} - u_t u_t^\top G_t^{-1}\frac{\partial^2 G_t}{\partial\phi_i\partial\phi_j}G_t^{-1} - \frac{\partial G_t}{\partial\phi_i}G_t^{-1}\frac{\partial G_t}{\partial\phi_j}G_t^{-1}\right.\\ &\quad \left. + u_t u_t^\top G_t^{-1}\frac{\partial G_t}{\partial\phi_j}G_t^{-1}\frac{\partial G_t}{\partial\phi_i}G_t^{-1} + u_t u_t^\top G_t^{-1}\frac{\partial G_t}{\partial\phi_i}G_t^{-1}\frac{\partial G_t}{\partial\phi_j}G_t^{-1}\right]\end{aligned}$$

with

$$\begin{aligned}E\left[\frac{\partial^2 l_t}{\partial\phi_i\partial\phi_j}(\phi_0)\right] &= \text{Tr}\left[E\left(G_t^{-1}\frac{\partial G_t}{\partial\phi_j}G_t^{-1}\frac{\partial G_t}{\partial\phi_i}\right)\right]\\ &= E\left[\text{vec}\left(\frac{\partial G_t}{\partial\phi_i}\right)^\top (G_t^{-1}\otimes G_t^{-1})\,\text{vec}\left(\frac{\partial G_t}{\partial\phi_j}\right)\right]\end{aligned}$$

after cancellation. ■

Proof of Lemma 4. The proof of Lemma 4 follows from the arguments given in Theorem 2, which hold uniformly over the required arguments. ■

Proof of Lemma 5. This follows by lengthy arguments similar to those given above.

■

F Miscellanea

F.1 Derivatives w.r.t. ϕ and σ

$$\begin{aligned}\ell_t(\phi) &= \log|G_t(\phi)| + u_t^\top G_t^{-1}(\phi)u_t, \quad \phi = (\text{vec}(A)^\top, \text{vec}(B)^\top)^\top\\ u_t &= \Sigma(t/T)^{-1/2}y_t\\ G_t(\phi) &= \sum_{j=0}^{t-1} B^j\left(I_N - AA^\top - BB^\top + Au_{t-j}u_{t-j}^\top A^\top\right)B^j\\ g_t = \text{vech}(G_t) &= \sum_{j=0}^{t-1} D_N^\top (B\otimes B)^j D_N \text{vech}(I_N - AA^\top - BB^\top + Au_{t-j}u_{t-j}^\top A^\top)\end{aligned} \tag{36}$$

F.1.1 First derivatives

Notation: ϕ_i is the i-th element of $\phi = (\text{vec}(A)^\top, \text{vec}(B)^\top)^\top$, A_{ij} and B_{kl} are the ij-th and kl-th elements of A and B, respectively. Then:

$$\begin{aligned}
\frac{\partial \ell_t}{\partial \phi_i} &= \text{Tr}\left[(I_N - u_t u_t^\top G_t^{-1})\frac{\partial G_t}{\partial \phi_i} G_t^{-1}\right] \\
&= \text{vec}\left(\frac{\partial G_t}{\partial \phi_i}\right)^\top \text{vec}\left(G_t^{-1} - G_t^{-1} u_t u_t^\top G_t^{-1}\right) \\
&= -\text{vec}\left(G_t^{-1/2}\frac{\partial G_t}{\partial \phi_i} G_t^{-1/2}\right)^\top \text{vec}\left(\varepsilon_t \varepsilon_t^\top - I_N\right) \\
\frac{\partial G_t}{\partial A_{ij}} &= \sum_{m=0}^{t-1} B^m \left\{J_{ij}(u_{t-m}u_{t-m}^\top - I_N)A^\top + A(u_{t-m}u_{t-m}^\top - I_N)J_{ji}\right\}(B^\top)^m \\
\frac{\partial G_t}{\partial B_{ij}} &= \sum_{m=0}^{t-1} \frac{\partial B^m}{\partial B_{ij}}\left(I_N - AA^\top - BB^\top + Au_{t-m}u_{t-m}^\top A^\top\right) B^m \\
&+ B^m\left(I_N - AA^\top - BB^\top + Au_{t-m}u_{t-m}^\top A^\top\right)\frac{\partial (B^\top)^m}{\partial B_{ij}} \\
&- B^m J_{ij}(B^\top)^{m+1} - B^{m+1} J_{ji}(B^\top)^m \\
\frac{\partial B^m}{\partial B_{ij}} &= \sum_{n=0}^{m-1} B^n J_{ij} B^{m-1-n},
\end{aligned}$$

where J_{ij} is an $N \times N$ matrix with zeros everywhere except for a one at the ij-th position.

F.1.2 Second derivatives

Notation: σ_j is the j-th element of $\text{vech}\{\Sigma(t/T)\}$.

$$
\begin{aligned}
\frac{\partial^2 l_t}{\partial\phi_i\partial\phi_j} &= \mathrm{Tr}\left[\frac{\partial^2 G_t}{\partial\phi_i\partial\phi_j}G_t^{-1} - u_t u_t^\top G_t^{-1}\frac{\partial^2 G_t}{\partial\phi_i\partial\phi_j}G_t^{-1} - \frac{\partial G_t}{\partial\phi_i}G_t^{-1}\frac{\partial G_t}{\partial\phi_j}G_t^{-1}\right.\\
&+ \left. u_t u_t^\top G_t^{-1}\frac{\partial G_t}{\partial\phi_j}G_t^{-1}\frac{\partial G_t}{\partial\phi_i}G_t^{-1} + u_t u_t^\top G_t^{-1}\frac{\partial G_t}{\partial\phi_i}G_t^{-1}\frac{\partial G_t}{\partial\phi_j}G_t^{-1}\right]\\
\frac{\partial^2 l_t}{\partial\phi_i\partial\sigma_j} &= \mathrm{Tr}\left[\frac{\partial^2 G_t}{\partial\phi_i\partial\sigma_j}G_t^{-1} - u_t u_t^\top G_t^{-1}\frac{\partial^2 G_t}{\partial\phi_i\partial\sigma_j}G_t^{-1} - \frac{\partial G_t}{\partial\phi_i}G_t^{-1}\frac{\partial G_t}{\partial\sigma_j}G_t^{-1}\right.\\
&+ u_t u_t^\top G_t^{-1}\frac{\partial G_t}{\partial\sigma_j}G_t^{-1}\frac{\partial G_t}{\partial\phi_i}G_t^{-1} + u_t u_t^\top G_t^{-1}\frac{\partial G_t}{\partial\phi_i}G_t^{-1}\frac{\partial G_t}{\partial\sigma_j}G_t^{-1}\\
&- \left.\frac{\partial\Sigma^{-1/2}(t/T)}{\partial\sigma_j}y_t y_t^\top\Sigma^{-1/2}(t/T)G_t^{-1}\frac{\partial G_t}{\partial\phi_i}G_t^{-1} - \Sigma^{-1/2}(t/T)y_t y_t^\top\frac{\partial\Sigma^{-1/2}(t/T)}{\partial\sigma_j}G_t^{-1}\frac{\partial G_t}{\partial\phi_i}G_t^{-1}\right]\\
\frac{\partial^2 G_t}{\partial A_{ij}\partial A_{kl}} &= \sum_{m=0}^{t-1}B^m\left\{J_{ij}(u_{t-m}u_{t-m}^\top - I_N)J_{lk} + J_{kl}(u_{t-m}u_{t-m}^\top - I_N)J_{ji}\right\}(B^\top)^m\\
\frac{\partial^2 G_t}{\partial A_{ij}\partial B_{kl}} &= \sum_{m=0}^{t-1}\frac{\partial B^m}{\partial B_{kl}}\left\{J_{ij}(u_{t-m}u_{t-m}^\top - I_N)A^\top + A(u_{t-m}u_{t-m}^\top - I_N)J_{ji}\right\}(B^\top)^m\\
&+ \sum_{m=0}^{t-1}B^m\left\{J_{ij}(u_{t-m}u_{t-m}^\top - I_N)A^\top + A(u_{t-m}u_{t-m}^\top - I_N)J_{ji}\right\}\frac{\partial(B^\top)^m}{\partial B_{ij}}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 G_t}{\partial B_{ij}\partial B_{kl}} &= \sum_{q=1}^{5}(I_q + I_q^\top)\\
I_1 &= \sum_{m=0}^{t-1}\frac{\partial^2 B^m}{\partial B_{ij}\partial B_{kl}}\left(I_N - AA^\top - BB^\top + Au_{t-m}u_{t-m}^\top A^\top\right)(B^\top)^m\\
I_2 &= \sum_{m=0}^{t-1}\frac{\partial B^m}{\partial B_{ij}}\left(I_N - AA^\top - BB^\top + Au_{t-m}u_{t-m}^\top A^\top\right)\frac{\partial(B^\top)^m}{\partial B_{kl}}\\
I_3 &= \sum_{m=0}^{t-1}\frac{\partial B^m}{\partial B_{ij}}\left(-J_{kl}B^\top - BJ_{lk}\right)(B^\top)^m\\
I_4 &= \sum_{m=0}^{t-1}\frac{\partial B^m}{\partial B_{kl}}\left(-J_{ij}B^\top - BJ_{ji}\right)(B^\top)^m\\
I_5 &= -B^m J_{ij}J_{lk}(B^\top)^m
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 G_t}{\partial A_{ij}\partial\sigma_k} &= \sum_{m=0}^{t-1} B^m \left\{ J_{ij}\frac{\partial\Sigma^{-1/2}(t-m/T)}{\partial\sigma_k} y_{t-m}u_{t-m}^\top A^\top + J_{ij}u_{t-m}y_{t-m}^\top \frac{\partial\Sigma^{-1/2}(t-m/T)}{\partial\sigma_k}A^\top \right. \\
&+ \left. A\frac{\partial\Sigma^{-1/2}(t-m/T)}{\partial\sigma_k} y_{t-m}u_{t-m}^\top J_{ji} + Au_{t-m}y_{t-m}^\top \frac{\partial\Sigma^{-1/2}(t-m/T)}{\partial\sigma_k} J_{ji} \right\} (B^\top)^m \\
\frac{\partial^2 G_t}{\partial B_{ij}\partial\sigma_k} &= \sum_{m=0}^{t-1} \frac{\partial B^m}{\partial B_{ij}} \left(A\frac{\partial\Sigma^{-1/2}(t-m/T)}{\partial\sigma_k} y_{t-m}u_{t-m}^\top A^\top + Au_{t-m}y_{t-m}^\top \frac{\partial\Sigma^{-1/2}(t-m/T)}{\partial\sigma_k}A^\top \right) B^m \\
&+ B^m \left(A\frac{\partial\Sigma^{-1/2}(t-m/T)}{\partial\sigma_k} y_{t-m}u_{t-m}^\top A^\top + Au_{t-m}y_{t-m}^\top \frac{\partial\Sigma^{-1/2}(t-m/T)}{\partial\sigma_k}A^\top \right) \frac{\partial (B^\top)^m}{\partial B_{ij}},
\end{aligned}
$$

where:

$$
\frac{\partial^2 B^m}{\partial B_{ij}\partial B_{kl}} = \sum_{n=0}^{m-1} \frac{\partial B^n}{\partial B_{kl}} J_{ij} B^{m-1-n} + B^n J_{ij} \frac{\partial B^{m-1-n}}{\partial B_{ij}},
$$

$$
\frac{\partial \mathrm{vech}\Sigma^{-1/2}(t/T)}{\partial\sigma_k} = -D_N^+ \left\{ \Sigma^{-1/2}(t/T) \otimes \Sigma^{-1/2}(t/T) \right\} D_N \frac{\partial \mathrm{vech}\Sigma^{1/2}(t/T)}{\partial\sigma_k},
$$

and where $\partial\mathrm{vech}\Sigma^{1/2}(t/T)/\partial\sigma_k$ depends on the particular definition used for the matrix square root.

References

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

[2] Bickel, P.J., Klaassen, C. A. J., Ritov, Y. and J. A. Wellner (1993), Efficient and adaptive estimation for semiparametric models (The John Hopkins University Press, Baltimore and London).

[3] Boswijk, P. and van der Weide, R. (2006), Wake me up before you GO-GARCH, discussion paper, UvA Amsterdam.

[4] Comte F. and O. Lieberman (2003). Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis*, 84, 61-84.

[5] Engle, R.F. (2002), Dynamic conditional correlation - A simple class of multivariate generalized autoregressive conditional heteroskedasticity models, *Journal of Business and Economic Statistics* 20, 339–350.

[6] Engle, R.F. and Rangel (2008), The spline GARCH model for unconditional volatility and its global macroeconomic causes, *Review of Financial Studies*, forthcoming.

[7] Jeantheau. T. (1998). Strong consistency for multivariate ARCH models. *Econometric Theory*, 14, 70-86.

[8] Lütkepohl, H. (1996). *Handbook of Matrices.* Wiley, New York.

[9] Rodriguez-Poo, J.R. and Linton, O.B. (2001), Nonparametric Factor Analysis for Residual Time Series. *TEST* 10, 161-182.

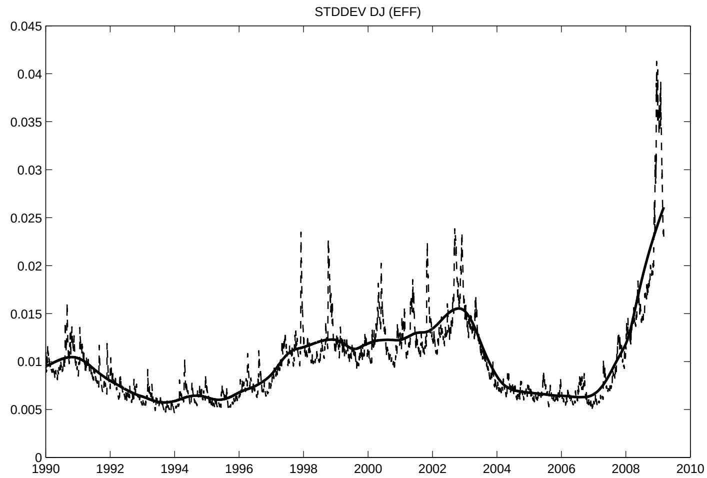

Figure 1: *Volatility (standard deviation) of Dow Jones daily returns, January 1990 1990 to January 2009, using the efficient estimator. Solid line: unconditional volatility, dashed line: conditional volatility.*

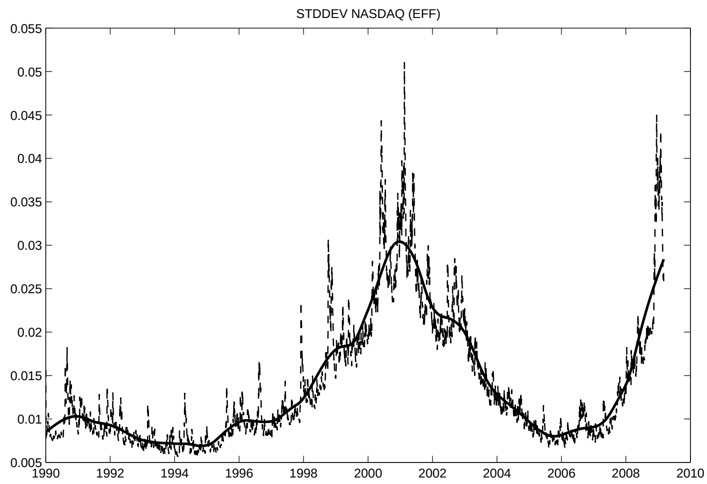

Figure 2: *Volatility (standard deviation) of NASDAQ daily returns, January 1990 1990 to January 2009, using the efficient estimator. Solid line: unconditional volatility, dashed line: conditional volatility.*

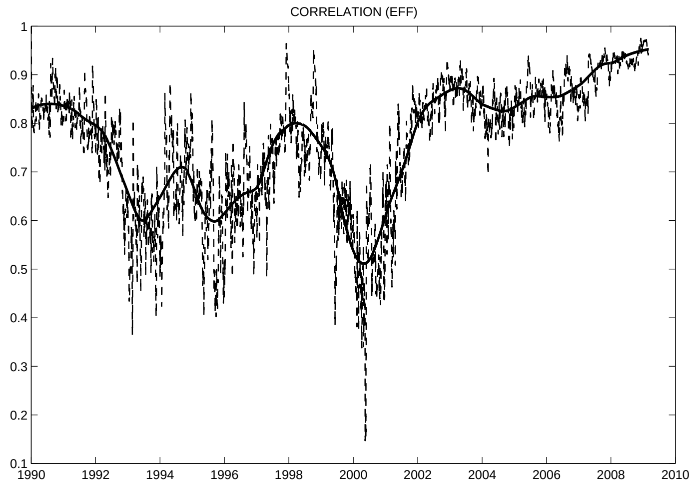

Figure 3: *Correlation of Dow Jones and NASDAQ daily returns, January 1990 1990 to January 2009, using the efficient estimator. Solid line: unconditional correlation, dashed line: conditional correlation.*

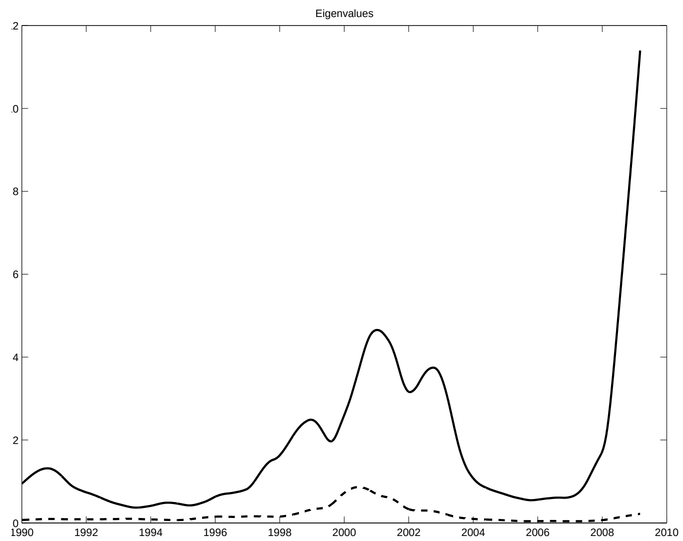

Figure 4: *Eigenvalues of the efficient estimator of* $\Sigma(u)$.