



**HAL**  
open science

## Efficient estimation of a multivariate multiplicative volatility model

Christian M. Hafner, Oliver Linton

► **To cite this version:**

Christian M. Hafner, Oliver Linton. Efficient estimation of a multivariate multiplicative volatility model. *Econometrics*, 2010, 159 (1), pp.55. 10.1016/j.jeconom.2010.04.007 . hal-00732539

**HAL Id: hal-00732539**

**<https://hal.science/hal-00732539>**

Submitted on 15 Sep 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Accepted Manuscript

Efficient estimation of a multivariate multiplicative volatility model

Christian M. Hafner, Oliver Linton

PII: S0304-4076(10)00113-2  
DOI: 10.1016/j.jeconom.2010.04.007  
Reference: ECONOM 3376

To appear in: *Journal of Econometrics*

Received date: 16 October 2009  
Revised date: 16 October 2009  
Accepted date: 29 April 2010



Please cite this article as: Hafner, C.M., Linton, O., Efficient estimation of a multivariate multiplicative volatility model. *Journal of Econometrics* (2010), doi:10.1016/j.jeconom.2010.04.007

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Efficient Estimation of a Multivariate Multiplicative Volatility Model\*

Christian M. Hafner<sup>†</sup>

Université catholique de Louvain

Oliver Linton<sup>‡</sup>

London School of Economics

April 14, 2010

## 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. We find that the univariate model of Engle and Rangel (2008) appears to be violated in the data whereas our multivariate model is more consistent with the data.

*Some key words:* GARCH; Kernel Estimation; Local Stationarity; Semiparametric

*JEL Classification Number:* C12, C13, C14.

---

\*We would like to thank A. Ron Gallant and two referees for helpful comments.

<sup>†</sup>Corresponding author. Institut de statistique and CORE, Université catholique de Louvain, Voie du Roman Pays 20, B-1348. Louvain-la-Neuve, Belgium, [christian.hafner@uclouvain.be](mailto: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.

<sup>‡</sup>Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail address: [o.linton@lse.ac.uk](mailto:o.linton@lse.ac.uk). Thanks to the ESRC, the ERC, and the Leverhulme foundation 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. Thanks to Yang Yan for research assistance.

# 1 Introduction

Modelling volatility is a big industry after the seminal work of Engle (1982) and Bollerslev (1986). One of the issues that has recently come to the fore is that of stationarity. Specifically, a number of authors, Drees and Starica (2003), Pohlzehl and Spokoiny (2004), Pesaran and Timmerman (2004), Engle and Rangel (2008) have presented evidence against global stationarity and proposed modifications of standard models. In Table 1 we present our own evidence based on a quite general parametric volatility model applied to S&P500 daily returns data. Evidently, assuming constant parameters for these data is problematic. Engle and Rangel (2008) have proposed a nonstationary model of volatility that makes the unconditional variance time varying in a deterministic but smooth fashion, but the short run volatility to be driven by a univariate GARCH model. Engle and Rangel (2008) estimate their model by a spline methodology and apply this to a cross-country panel dataset. This model can be viewed as a special case of the class of locally stationary processes introduced by Dahlhaus (1997) (see also Robinson (1989)). Recently, Dahlhaus and Subba Rao (2006) have investigated a class of time varying ARCH( $\infty$ ) processes that allows all the parameters to vary smoothly over time. Pohlzehl and Spokoiny (2004) have proposed a methodology for estimation of time varying volatility processes that is quite general and allows the process to have structural breaks at many unknown locations.

Modelling multivariate volatility is now a big area with many important contributions, see Engle and Sheppard (2001), Alexander (2001), Tse and Tsui (2002), van der Weide (2002), Vrontos, Dellaportas and Politis (2003), Lanne and Saikkonen (2005), Boswijk and van der Weide (2006), Patton (2006), Jondeau and Rockinger (2006), Kawatsu (2006), Bauwens, Laurent and Rombouts (2006) and Silvennoinen and Teräsvirta (2009b). We propose a new semiparametric multivariate volatility model that allows for nonstationarity. 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. Silvennoinen and Teräsvirta (2009a) propose a related methodology based on smooth transition between a finite number of states, where the transition is determined by a speed parameter and a c.d.f. Rangel and Engle (2009) consider a specialized multivariate version of Engle and Rangel (2008).

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 spec-

ification 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. This provides additional motivation for studying the multivariate model.

Throughout the paper we use the following notation:  $A \otimes B$  is the Kronecker product of two matrices  $A$  and  $B$ .  $\text{vec}()$  is the operator that stacks the columns of a matrix in a column vector, while  $\text{vech}()$  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. For any matrix  $A$ , let  $\|A\| = \text{Tr}(A^\top A)^{1/2}$  be the Euclidean norm.

Assumptions, proofs of the theorems as well as lemmata are delegated to an appendix.

## 2 The Model and its Properties

We observe a vector time series  $y_t \in \mathbb{R}^N$  for  $t = 1, \dots, T$ . We shall ignore mean effects in the main development and suppose that  $y_t$  satisfies the model

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

where:  $\varepsilon_t$  is (at least) a strictly stationary 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  $\{y_{t-1}, y_{t-2}, \dots\}$  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 strictly stationary stochastic covariance matrix process. 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  $\Sigma$ , but we discuss later the important extension to allow for breaks. We model  $G_t$  parametrically so that  $G_t = G_t(\phi)$  for  $\phi \in \Phi \subset \mathbb{R}^p$  and in fact we shall assume that

$$G_t(\phi) = \Gamma(u_{t-1}, u_{t-2}, \dots; \phi) \quad (2)$$

for some fixed known function  $\Gamma$ . This is a very general class of processes, and includes many of the models reviewed in Bauwens, Laurent and Rombouts (2006). As in the univariate case, there is an identification issue, and we shall assume that  $EG_t(\phi) = I_N$ . It is perhaps too general to understand stationarity and mixing conditions and to obtain detailed analytical results for distribution theory so we shall in some cases specialise to the BEKK process that imposes the normalization condition  $EG_t(\phi) = I_N$ ,

$$G_t(\phi) = I_N - AA^\top - BB^\top + Au_{t-1}u_{t-1}^\top A^\top + BG_{t-1}B^\top \quad (3)$$

in which case  $\phi = (\text{vec}(A)^\top, \text{vec}(B)^\top)^\top$  denote the free parameters of  $G_t$ .<sup>1</sup> However, other models for  $G_t$  can be considered.<sup>2</sup> The model allows slowly varying unconditional variance matrix  $\Sigma$  along with short run dynamics through the process  $G_t$ . Under some conditions, one can approximate this process by a family of locally stationary processes. Specifically, let  $\tilde{y}_t(u) = \Sigma(u)^{1/2} G_t^{1/2} \varepsilon_t$  for each  $u \in [0, 1]$ . Then one can approximate  $y_t$  by the stationary process  $\tilde{y}_t(u)$  for  $t/T$  in a small neighborhood of  $u$ . See Dahlhaus (1997).

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  $\varepsilon_t$  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.<sup>3</sup>

We remark on some properties of the stochastic process  $y_t$ . For the univariate process, the local autocorrelation function (LACF) of any power of absolute returns in the Engle and Rangel (2008) model is time invariant, i.e.,

$$\begin{aligned} \rho_{|y|^\alpha}(t, j) &= \frac{\text{cov}(|y_t|^\alpha, |y_{t-j}|^\alpha)}{\sqrt{\text{var}(|y_t|^\alpha) \text{var}(|y_{t-j}|^\alpha)}} \\ &= \frac{\sigma_t^\alpha \sigma_{t-j}^\alpha \text{cov}(g_t^{\alpha/2} |\varepsilon_t|^\alpha, g_{t-j}^{\alpha/2} |\varepsilon_{t-j}|^\alpha)}{\sqrt{\sigma_t^{2\alpha} \sigma_{t-j}^{2\alpha} \text{var}(g_t^{\alpha/2} |\varepsilon_t|^\alpha) \text{var}(g_{t-j}^{\alpha/2} |\varepsilon_{t-j}|^\alpha)}} \\ &= \frac{\text{cov}(g_t^{\alpha/2} |\varepsilon_t|^\alpha, g_{t-j}^{\alpha/2} |\varepsilon_{t-j}|^\alpha)}{\text{var}(g_t^{\alpha/2} |\varepsilon_t|^\alpha)} \\ &= \rho_{|u|^\alpha}(t, j) = \rho_{|u|^\alpha}(j), \end{aligned}$$

where  $|u_t|^\alpha = g_t^{\alpha/2} |\varepsilon_t|^\alpha$ , and  $\rho_{|u|^\alpha}(t, j)$  is time invariant because of the stationarity of  $u_t = g_t^{1/2} \varepsilon_t$ .

<sup>1</sup>Note that when  $\Sigma(t/T) = \Sigma$ , this model collapses to the standard BEKK model.

$$\Omega_t = E(y_t y_t^\top | \mathcal{F}_{t-1}) = \tilde{C} \tilde{C}^\top + \tilde{A} y_{t-1} y_{t-1}^\top \tilde{A}^\top + \tilde{B} \Omega_{t-1} \tilde{B}^\top$$

for matrices  $\tilde{C}$ ,  $\tilde{A}$ , and  $\tilde{B}$ , where  $\tilde{A} = \Sigma^{1/2} A \Sigma^{-1/2}$  etc.

<sup>2</sup>A key property of the BEKK model, which makes the analysis easy is that it is possible to impose the restriction  $EG_t(\phi) = I_N$  by restricting the  $\phi$ . In other models (like the DCC model)  $EG_t(\phi)$  may in general depend on the distribution of the errors  $\varepsilon_t$ , and one cannot impose this assumption without specifying an error distribution parametrically. In that case one should consider different identification conditions, perhaps involving  $\Sigma(\cdot)$ . This affects the interpretation of the components of the model and makes estimation perhaps more complicated, but is still doable. In summary, any parametric model can be conceived for  $G_t$  but some are more convenient than others.

<sup>3</sup>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  $w_0, \dots, w_k, \delta$  are unknown parameters. In order to make sense of the asymptotic properties of such a (nonparametric) procedure one should rescale time as we have done.

The dynamics of the model only enter through the stationary process  $u_t$ . This means for example that any long memory properties in  $|y_t|^\alpha$  cannot be explained by structural breaks in the deterministic part of volatility  $\sigma^2(\cdot)$ , it has to arise from properties of  $u_t$ . Likewise,  $E[|y_t|^\alpha]/(E[y_t^2])^{\alpha/2} = E[|u_t|^\alpha]/(E[u_t^2])^{\alpha/2}$ , so that the local cumulants for example are time invariant and depend only on the corresponding cumulants of the stationary process  $u_t$ . In fact both these properties hold for whatever process  $g_t$  just so long as it is measurable only with respect to the past and possesses the required moments. This suggests a simple specification test of the model by looking at the local (in time) correlogram of powers of absolute returns and local cumulants and testing whether these quantities are constant over time. We will look at this in our application below.

In the multivariate case this time invariance does not hold, 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, consider  $\eta_t = \text{vech}(y_t y_t^\top)$ ,  $\bar{\eta}_t = E[\eta_t]$ , and define:

$$z_t = \text{vech}(G_t^{1/2} \varepsilon_t \varepsilon_t^\top G_t^{1/2} - I_N), \quad (4)$$

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

$$\Gamma(t, 0) = E[(\eta_t - \bar{\eta}_t)(\eta_t - \bar{\eta}_t)^\top] = D_N^+ W(t, 0) M_0 W(t, 0) D_N^{+\top}$$

$$\Gamma(t, j) = E[(\eta_t - \bar{\eta}_t)(\eta_{t-j} - \bar{\eta}_{t-j})^\top] = D_N^+ W(t, 0) M_j W(t, j) D_N^{+\top}.$$

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  $j$  (this corresponds to the correlogram of the locally stationary approximation). Consider the bivariate case  $y_{1t} = \sigma_{11}(t/T)u_{1t} + \sigma_{12}(t/T)u_{2t}$ , where  $u_{1t}, u_{2t}$  are stationary processes. Therefore, for example

$$\begin{aligned} \rho_{|y|^\alpha}(t, j) &= \frac{\text{cov}(|y_{1t}|^\alpha, |y_{1,t-j}|^\alpha)}{\sqrt{\text{var}(|y_{1t}|^\alpha) \text{var}(|y_{1,t-j}|^\alpha)}} \\ &= \frac{\text{cov}(|\sigma_{11}(t/T)u_{1t} + \sigma_{12}(t/T)u_{2t}|^\alpha, |\sigma_{11}(t-j/T)u_{1,t-j} + \sigma_{12}(t-j/T)u_{2,t-j}|^\alpha)}{\sqrt{\text{var}(|\sigma_{11}(t/T)u_{1t} + \sigma_{12}(t/T)u_{2t}|^\alpha) \text{var}(|\sigma_{11}(t-j/T)u_{1,t-j} + \sigma_{12}(t-j/T)u_{2,t-j}|^\alpha)}} \\ &\simeq \frac{\text{cov}(|\sigma_{11}(u)u_{1t} + \sigma_{12}(u)u_{2t}|^\alpha, |\sigma_{11}(u)u_{1,t-j} + \sigma_{12}(u)u_{2,t-j}|^\alpha)}{\sqrt{\text{var}(|\sigma_{11}(u)u_{1t} + \sigma_{12}(u)u_{2t}|^\alpha) \text{var}(|\sigma_{11}(u)u_{1,t-j} + \sigma_{12}(u)u_{2,t-j}|^\alpha)}}. \end{aligned}$$

This depends on  $t$  or, in the local stationary approximation, on  $u$ .<sup>4</sup>

---

<sup>4</sup>If  $\Sigma(u)$  is diagonal, then there is an invariance property that can be exploited [specifically the invariance of autocorrelation to transformations  $x_t \mapsto b + Ax_t$ , where  $A$  is diagonal] to show that  $\Psi(t, j) \simeq \Psi(j)$  for any fixed  $j$ . In the bivariate example this means that  $\sigma_{12}(u) = 0$  and so  $\rho_{|y|^\alpha}(t, j) = \rho_{|u|^\alpha}(t, j) = \rho_{|u|^\alpha}(j)$ .

Starica (2003) argued that the simple deterministic model  $y_t = \sigma_t \varepsilon_t = \sigma(t/T) \varepsilon_t$ , where  $\sigma(\cdot)$  is an unknown function of (rescaled) time and  $\varepsilon_t$  i.i.d. can perform as well as the GARCH(1,1) over some datasets. Rodriguez-Poo and Linton (2001) considered the multivariate generalization of this model in another context in which  $y_t = \Sigma(t/T)^{1/2} \varepsilon_t$ . In this model the conditional and unconditional variance of  $y_t$  are both equal to  $\Sigma(t/T)$ . If one finds time varying local autocorrelation, this would be inconsistent with this model too.

We remark that there are of course many reasons why we might empirically reject constant LACF.<sup>5</sup> For example, the presence of jumps in the return process can lead to an addition of an orthogonal stochastic term that would eliminate the cancellations of the low frequency terms in the LACF. Structural breaks can also affect such cancellations. Another possibility is the presence of an underlying factor structure that can be addressed in different ways within the multiplicative error framework (see for example Rangel and Engle (2009)).

### 3 Efficient Estimation

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

#### 3.1 The Parameters $\phi$

Here we consider the question of semiparametric efficiency, and our treatment follows a little the style of Bickel, Klaassen, Ritov, and Wellner (1993). See also Severini and Wong (1992). Consider the univariate model where

$$\sigma_t^2(\phi, h) = h(t/T)g_t(\phi) \tag{5}$$

for some unknown function  $h(\cdot)$  and  $\varepsilon_t \sim N(0, 1)$ . 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  $\psi_0(u) \equiv 1$ , and

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

where  $\delta_{jk} = 1$  if  $j = k$  and 0 if  $j \neq k$ . Thus  $\int_0^1 \psi_j(u) du = 0$  for all  $j \geq 1$  (which identifies  $\theta_0$  as the average value of  $\log h$ ). Then consider some finite order approximating model where

---

<sup>5</sup>As a trivial example, suppose that  $y_t = \rho y_{t-1} + u_t$  with  $u_t = \sigma(t/T)g_t^{1/2} \varepsilon_t$ , then the LACF of  $y_t$  will be time varying. However, for many financial time series it is believed that  $\rho$  is small and so we don't expect mean dynamics to drive the autocorrelation structure of high moments of returns.

$\sigma_t^2(\phi, \theta) = h_\theta(t/T)g_t(\phi)$  with  $\log h_\theta(t/T) = \sum_{j=0}^J \theta_j \psi_j(t/T)$ , and define the normalized likelihood function  $\ell_T(\phi, \theta) = -T^{-1} \sum_{t=1}^T \{\log \sigma_t^2(\phi, \theta) + y_t^2 / \sigma_t^2(\phi, \theta)\}$ . We have:

$$\begin{aligned} \frac{\partial \log h_\theta(t/T)}{\partial \theta_j} &= \psi_j(t/T) \\ I_{\theta\theta} &= \lim_{T \rightarrow \infty} TE \left[ \frac{\partial \ell_T(\phi_0, \theta_0)}{\partial \theta} \frac{\partial \ell_T(\phi_0, \theta_0)}{\partial \theta^\top} \right] = 2 \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T [\psi_j(t/T) \psi_k(t/T)]_{j,k} = 2I_J \\ I_{\phi\theta} &= 2E \left[ \frac{\partial \log g_t}{\partial \phi} \right] \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T [\psi_j(t/T)]_j = 2E \left[ \frac{\partial \log g_t}{\partial \phi} \right] (1, 0, \dots, 0) \\ \sqrt{T} \frac{\partial \ell_T(\phi_0, \theta_0)}{\partial \theta_j} &= \frac{1}{\sqrt{T}} \sum_{t=1}^T (\varepsilon_t^2 - 1) \psi_j(t/T), \end{aligned}$$

because  $E[(\varepsilon_t^2 - 1)^2] = 2$  for the normal distribution. From this we can obtain the efficient score function for  $\phi$  within the parametric model

$$\begin{aligned} \frac{\partial \ell_T^*(\phi_0, \theta_0)}{\partial \phi} &\equiv \frac{\partial \ell_T(\phi_0, \theta_0)}{\partial \phi} - I_{\phi\theta} I_{\theta\theta}^{-1} \frac{\partial \ell_T(\phi_0, \theta_0)}{\partial \theta} \\ &= \frac{1}{T} \sum_{t=1}^T (\varepsilon_t^2 - 1) \frac{\partial \log g_t}{\partial \phi} - E \left[ \frac{\partial \log g_t}{\partial \phi} \right] (1, 0, \dots, 0) \frac{1}{\sqrt{T}} \sum_{t=1}^T (\varepsilon_t^2 - 1) [\psi_j(t/T)]_j \\ &= \frac{1}{T} \sum_{t=1}^T (\varepsilon_t^2 - 1) \left[ \frac{\partial \log g_t}{\partial \phi} - E \left[ \frac{\partial \log g_t}{\partial \phi} \right] \right]. \end{aligned}$$

The simple structure allows us to pass  $J$  to infinity and obtain the semiparametric efficient score function and efficient information:

$$\frac{\partial \ell_T^*(\phi_0)}{\partial \phi} = \frac{1}{T} \sum_{t=1}^T (\varepsilon_t^2 - 1) \left[ \frac{\partial \log g_t}{\partial \phi} - E \left( \frac{\partial \log g_t}{\partial \phi} \right) \right] \quad (6)$$

$$\mathcal{I}_{\phi\phi}^* = E \left[ \left( \frac{\partial \log g_t}{\partial \phi} - E \left( \frac{\partial \log g_t}{\partial \phi} \right) \right) \left( \frac{\partial \log g_t}{\partial \phi} - E \left( \frac{\partial \log g_t}{\partial \phi} \right) \right)^\top \right], \quad (7)$$

see Bickel, Klaassen, Ritov, and Wellner (1993) for discussion of these concepts. This says that an efficient estimator in the semiparametric model would have asymptotic variance given by  $2\mathcal{I}_{\phi\phi}^{*-1}$ . If we compare this with the parametric estimator that knows  $h(\cdot)$ , we see that the information loss is

$$\begin{aligned} &E \left[ \frac{\partial \log g_t}{\partial \phi} \frac{\partial \log g_t}{\partial \phi^\top} \right] - E \left[ \left( \frac{\partial \log g_t}{\partial \phi} - E \left( \frac{\partial \log g_t}{\partial \phi} \right) \right) \left( \frac{\partial \log g_t}{\partial \phi} - E \left( \frac{\partial \log g_t}{\partial \phi} \right) \right)^\top \right] \\ &= E \left( \frac{\partial \log g_t}{\partial \phi} \right) E \left( \frac{\partial \log g_t}{\partial \phi^\top} \right) > 0. \end{aligned}$$

Therefore, the best one can do in the semiparametric model is strictly worse than the best one can do in the parametric model.

PROPOSITION 1. In the multivariate case, the semiparametric efficient score function and efficient information are given by:

$$\begin{aligned}\frac{\partial \ell_T^*(\phi_0)}{\partial \phi_i} &= \frac{1}{\sqrt{T}} \sum_{t=1}^T \text{vec} \left( G_t^{-1/2} \frac{\partial G_t}{\partial \phi_i} G_t^{-1/2} - E \left[ G_t^{-1/2} \frac{\partial G_t}{\partial \phi_i} G_t^{-1/2} \right] \right)^\top \text{vec} (\varepsilon_t \varepsilon_t^\top - I_N) \\ \mathcal{I}_{\phi_i \phi_j}^* &= 2E \left[ \text{vec} \left( G_t^{-1/2} \frac{\partial G_t}{\partial \phi_i} G_t^{-1/2} - E \left[ G_t^{-1/2} \frac{\partial G_t}{\partial \phi_i} G_t^{-1/2} \right] \right)^\top D_N D_N^+ \right. \\ &\quad \left. \times \text{vec} \left( G_t^{-1/2} \frac{\partial G_t}{\partial \phi_j} G_t^{-1/2} - E \left[ G_t^{-1/2} \frac{\partial G_t}{\partial \phi_j} G_t^{-1/2} \right] \right) \right]\end{aligned}$$

PROOF: See Appendix A.2.

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

$$\hat{\phi} = \tilde{\phi} - \tilde{\mathcal{I}}_{\phi\phi}^{*-1}(\tilde{\phi}) \frac{\partial \tilde{\ell}_T^*}{\partial \phi}(\tilde{\phi}), \quad (8)$$

where  $\partial \tilde{\ell}_T^* / \partial \phi$  and  $\tilde{\mathcal{I}}_{\phi\phi}^*$  are estimates of  $\partial \ell_T^* / \partial \phi$  and  $\mathcal{I}_{\phi\phi}^*$ , and  $\tilde{\phi}$  is an initial root-T consistent estimator of  $\phi$ . To construct estimates of  $\partial \ell_T^* / \partial \phi$  and  $\mathcal{I}_{\phi\phi}^*$  we also need preliminary estimates of  $\Sigma(\cdot)$ .

### 3.2 The Function $\Sigma(\cdot)$

We now discuss efficient estimation of  $\Sigma(\cdot)$ . 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:

$$\tilde{y}_t^2 = \frac{y_t^2}{g_t} = \sigma^2(t/T) + \sigma^2(t/T)(\varepsilon_t^2 - 1) \quad (9)$$

$$y_t^2 = \sigma^2(t/T) + \sigma^2(t/T)(g_t \varepsilon_t^2 - 1). \quad (10)$$

In both cases the error term is mean zero so that smoothing on time will yield consistent estimates of  $\sigma^2(t/T)$ . 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,  $\text{var}(g_t \varepsilon_t^2 - 1) = E[(g_t \varepsilon_t^2 - 1)^2] = E(g_t^2 \varepsilon_t^4) - 1 = E(\varepsilon_t^4) E g_t^2 - 1$ . Since  $\text{var}(g_t) = E g_t^2 - 1$ , we have  $E g_t^2 \geq 1$  and so  $\text{var}(g_t \varepsilon_t^2 - 1) \geq \text{var}(\varepsilon_t^2 - 1)$ . The process  $\varepsilon_t^2 - 1$  is uncorrelated but  $g_t \varepsilon_t^2 - 1$  is autocorrelated, and in particular  $\text{cov}(g_t \varepsilon_t^2 - 1, g_{t-j} \varepsilon_{t-j}^2 - 1) = E(g_t g_{t-j} \varepsilon_{t-j}^2) - 1 \neq 0$ . In fact,  $g_t \varepsilon_t^2 - 1$  is positively autocorrelated so that the long run variance exceeds the short run variance. This is one intuition why the improved estimator of  $\Sigma(\cdot)$  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

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

which corresponds to a standard regression smoother from (9). The local likelihood method with uniform kernel is efficient in the sense that it has the least variance amongst all estimators with the same bias function, ref.

In the multivariate case suppose that one knew the random variable  $G_t$ , how would you proceed to improve the estimate of  $\Sigma(t/T)$  and hence of  $\phi$ ? 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  $y_t^2/g_t$ . 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  $\varepsilon_t$  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

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

In the sequel we treat  $\Sigma(t/T)^{1/2}$  as an unknown parameter and replace it by  $\Theta$ . Let  $\theta = \text{vech}(\Theta) \in \mathbb{R}^{N(N+1)/2}$  be the unique elements of  $\Theta$ . 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  $\theta \in \Theta$ . The resulting estimator is denoted  $\hat{\theta}(u)$  and hence  $\hat{\Sigma}(u) = \hat{\Theta}^2(u)$ . In the case where  $\varepsilon_t$  is i.i.d. normal,  $\hat{\Sigma}(u)$  has smaller variance than the local sample variance defined below in (11).

## 4 Estimation

In the sequel we propose an estimation method for the parameters  $\phi$  along with the function  $\Sigma(\cdot)$ . The estimation method is designed to be efficient under the assumption that  $\varepsilon_t$  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.

#### 4.1 Step 1 Initial Estimation of $\Sigma$

Under the model assumptions,

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

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

$$\tilde{\Sigma}(u) = \frac{\sum_{t=1}^T K_h(u - t/T) y_t y_t^\top}{\sum_{t=1}^T K_h(u - t/T)}, \quad (11)$$

where  $K$  is a kernel function,  $h$  is a bandwidth, and  $K_h(\cdot) = K(\cdot/h)/h$ . Rodriguez-Poo and Linton (2001) established the consistency and asymptotic normality of  $\tilde{\Sigma}(u)$  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  $\omega = \text{vech}(\Omega)$ , we have

$$\frac{\partial l}{\partial \omega} = -D_N^\top \text{vec}(\Omega^{-1} - \Omega^{-1} y_t y_t^\top \Omega^{-1}) = -D_N^\top (\Omega^{-1} \otimes \Omega^{-1}) D_N \text{vech}(\Omega - y_t y_t^\top),$$

which, solving for  $\Omega$ , yields (11) exactly.

#### 4.2 Step 2 Initial Estimation of $\phi$

First, one computes the profiled  $G$  process, i.e., for each  $\phi$ , let

$$\tilde{G}_t(\phi) = \Gamma(\tilde{u}_{t-1}, \tilde{u}_{t-2}, \dots, \tilde{u}_{p+1}, 0, \dots; \phi), \quad (12)$$

where  $\tilde{u}_s = \tilde{\Sigma}(s/T)^{-1/2} y_s$ . For example in the BEKK case we might compute the simple recursion

$$\tilde{G}_t(\phi) = I_N - AA^\top - BB^\top + A\tilde{u}_{t-1}\tilde{u}_{t-1}^\top A^\top + B\tilde{G}_{t-1}(\phi)B^\top \quad (13)$$

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

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

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

$$\tilde{\ell}_T(\phi) = \sum_{t=1}^T l(\tilde{G}_t(\phi); \tilde{u}_t). \quad (14)$$

This estimator is expected to be consistent and asymptotically normal but inefficient.

### 4.3 Step 3 Improved Estimation

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

$$\tilde{L}_T(\theta; u) = \sum_{t=1}^T K_h(u - t/T) l(\tilde{\Omega}_t(\theta, \tilde{\phi}); y_t), \quad (15)$$

where  $\tilde{\Omega}_t(\theta, \tilde{\phi}) = \Theta \tilde{G}_t(\tilde{\phi}) \Theta$  and for any  $\phi$ ,  $\tilde{G}_t(\phi)$  is given in (12). Then let

$$\hat{\theta}(u) = \tilde{\theta}(u) - \left[ \frac{\partial^2 \tilde{L}_T(\tilde{\theta}; u)}{\partial \theta \partial \theta^\top} \right]^{-1} \frac{\partial \tilde{L}_T(\tilde{\theta}; u)}{\partial \theta}, \quad (16)$$

where  $\tilde{\theta}(u) = \text{vech}(\tilde{\Sigma}(u)^{1/2})$ , and let  $\hat{\Sigma}(u) = \hat{\Theta}^2(u)$ .

Next one computes a new profiled  $G$ , i.e., for each  $\phi$ , let

$$\hat{G}_t(\phi) = \Gamma(\hat{u}_{t-1}, \hat{u}_{t-2}, \dots, \hat{u}_{p+1}, 0, \dots; \phi), \quad (17)$$

for example

$$\hat{G}_t(\phi) = I_N - AA^\top - BB^\top + A\hat{u}_{t-1}\hat{u}_{t-1}^\top A^\top + B\hat{G}_{t-1}(\phi)B^\top, \quad (18)$$

where  $\hat{u}_s = \hat{\Sigma}(s/T)^{-1/2} y_s$  and some initialization  $\hat{G}_1(\phi)$  is chosen. Then compute the two-step estimator

$$\hat{\phi} = \tilde{\phi} - \left[ \frac{\partial^2 \hat{\ell}_T^*(\tilde{\phi})}{\partial \phi \partial \phi^\top} \right]^{-1} \frac{\partial \hat{\ell}_T^*(\tilde{\phi})}{\partial \phi}, \quad (19)$$

where the estimated efficient score function is

$$\begin{aligned}\frac{\partial \hat{\ell}_T^*(\tilde{\phi})}{\partial \phi_i} &= \frac{1}{T} \sum_{t=1}^T \text{vec} \left( \hat{G}_t^{-1/2}(\tilde{\phi}) \frac{\partial \hat{G}_t}{\partial \phi_i}(\tilde{\phi}) \hat{G}_t^{-1/2}(\tilde{\phi}) - \frac{1}{T} \sum_{t=1}^T \hat{G}_t^{-1/2}(\tilde{\phi}) \frac{\partial \hat{G}_t}{\partial \phi_i}(\tilde{\phi}) \hat{G}_t^{-1/2}(\tilde{\phi}) \right)^\top \\ &\quad \times \text{vec} \left( \hat{\varepsilon}_t(\tilde{\phi}) \hat{\varepsilon}_t^\top(\tilde{\phi}) - I_N \right),\end{aligned}$$

and  $\widehat{\varepsilon}_t(\widetilde{\phi}) = \widehat{G}_t(\widetilde{\phi})^{-1/2} \widehat{\Sigma}(t/T)^{-1/2} y_t$ .

One can iterate this procedure by updating the local likelihood using the new estimator of  $\phi$  and so on, but asymptotically this will not affect the variances of the procedures.

## 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 (4), which is a stationary mixing process with unconditional mean zero, and let

$$\Gamma_j = E[z_t z_{t-j}^\top], \quad j = 0, 1, \dots \quad (20)$$

Then let:

$$V_\sigma(u) = \|K\|_2^2 D_N [\Sigma(u)^{1/2} \otimes \Sigma(u)^{1/2}] D_N M_\infty D_N [\Sigma(u)^{1/2} \otimes \Sigma(u)^{1/2}] D_N \quad (21)$$

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

$$V_\phi = J^{-1} Q J^{-1} \quad (22)$$

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

$$\rho_t = \frac{\partial \text{vec}(G_t)^\top}{\partial \phi} (G_t^{-1/2} \otimes G_t^{-1/2}), \quad (23)$$

and where the matrix  $Q$  is defined in (45) in the appendix, while  $\|K\|_2^2 = \int K(s)^2 ds$ . Let  $\sigma(u) = \text{vech}(\Sigma(u))$  and  $\widetilde{\sigma}(u) = \text{vech}(\widetilde{\Sigma}(u))$ . The assumptions for the following theorems are given in Appendix A.1.

**THEOREM 1.** *Under Assumptions A1 to A10, there exist bounded continuous functions  $b_\sigma(u)$  such that*

$$\sqrt{T} h (\widetilde{\sigma}(u) - \sigma(u) - h^2 b_\sigma(u)) \implies N(0, V_\sigma(u)) \quad (24)$$

$$\sqrt{T} (\widetilde{\phi} - \phi) \implies N(0, V_\phi). \quad (25)$$

**PROOF:** See Appendix A.3.

The result in (24) 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  $\sigma''(u)$ . Clearly, the asymptotic distributions of these estimators are very complicated and practically unusable. The estimators are also inefficient in the leading Gaussian case. The result in Theorem 1 (25) is also specific only to the BEKK special case (3), whereas the Theorem 3 below holds in principle more generally along the lines of Bollerslev and Wooldridge (1992).

Define

$$\zeta_t = \text{vec}(\varepsilon_t \varepsilon_t^\top - I_N), \quad (26)$$

which is a vector martingale difference sequence, and let

$$\Xi_t = E[\zeta_t \zeta_t^\top | \mathcal{F}_{t-1}]. \quad (27)$$

Then let

$$\begin{aligned} V_\sigma^e(u) &= \|K\|_2^2 \Delta(u) \Lambda(u)^{-1} \Psi(u) \Lambda(u)^{-1} \Delta(u)^\top \\ \Delta(u) &= D_N^+ [(I_N \otimes \Sigma^{1/2}(u)) + (\Sigma^{1/2}(u) \otimes I_N)] D_N \end{aligned} \quad (28)$$

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

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

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

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

**THEOREM 2.** *Let  $\hat{\sigma}(u) = \text{vech}(\hat{\Sigma}(u))$ . Under Assumptions A1 to A10, there exist bounded continuous functions  $b_\sigma^e(u)$  such that*

$$\sqrt{T}h(\hat{\sigma}(u) - \sigma(u) - h^2 b_\sigma^e(u)) \implies N(0, V_\sigma^e(u)). \quad (29)$$

**PROOF:** See Appendix A.4.

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

Let

$$V_\phi^e = J^{e-1} Q^e J^{e-1} \quad (30)$$

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

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

**THEOREM 3.** *Under Assumptions A1 to A10*

$$\sqrt{T}(\hat{\phi} - \phi) \implies N(0, V_\phi^e). \quad (31)$$

**PROOF:** See Appendix A.5.

When  $\varepsilon_t$  are i.i.d. standard normal,  $\Xi_t = 2D_N D_N^+$  a.s. 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. The standard errors can be obtained from the estimated matrices:

$$\hat{V}_\sigma^e(u) = \|K\|_2^2 \hat{\Delta}(u) \hat{\Lambda}(u)^{-1} \hat{\Psi}(u) \hat{\Lambda}(u)^{-1} \hat{\Delta}(u)^\top$$

$$\hat{V}_\phi^e = \hat{J}^{e-1} \hat{Q}^e \hat{J}^{e-1}$$

$$\begin{aligned}
\widehat{\Delta}(u) &= D_N^+ \left[ \left( I_N \otimes \widehat{\Sigma}^{1/2}(u) \right) + \left( \widehat{\Sigma}^{1/2}(u) \otimes I_N \right) \right] D_N \\
\widehat{\Lambda}(u) &= 4D_N^\top \left( \widehat{\Sigma}^{-1/2}(u) \otimes I_N \right) \frac{1}{T} \sum_{t=1}^T (\widehat{G}_t^{-1} \otimes \widehat{G}_t) \left( \widehat{\Sigma}^{-1/2}(u) \otimes I_N \right) D_N \\
\widehat{\Psi}(u) &= 4D_N^\top \left( \widehat{\Sigma}^{-1/2}(u) \otimes I_N \right) \frac{1}{T} \sum_{t=1}^T (\widehat{G}_t^{-1/2} \otimes \widehat{G}_t^{1/2}) \widehat{\zeta}_t \widehat{\zeta}_t^\top (\widehat{G}_t^{-1/2} \otimes \widehat{G}_t^{1/2}) \left( \widehat{\Sigma}^{-1/2}(u) \otimes I_N \right) D_N \\
\widehat{J}^e &= \frac{1}{T} \sum_{t=1}^T \left( \widehat{\rho}_t - \frac{1}{T} \sum_{t=1}^T \widehat{\rho}_t \right) \left( \widehat{\rho}_t - \frac{1}{T} \sum_{t=1}^T \widehat{\rho}_t \right)^\top \\
\widehat{Q}^e &= \frac{1}{T} \sum_{t=1}^T \left( \widehat{\rho}_t - \frac{1}{T} \sum_{t=1}^T \widehat{\rho}_t \right) \widehat{\zeta}_t \widehat{\zeta}_t^\top \left( \widehat{\rho}_t - \frac{1}{T} \sum_{t=1}^T \widehat{\rho}_t \right)^\top,
\end{aligned}$$

where the hatted quantities have  $\widehat{\phi}$  and  $\widehat{\sigma}(\cdot)$  replacing the unknown quantities. Under our conditions  $\widehat{V}_\phi^e$  and  $\widehat{V}_\sigma^e(u)$  are consistent estimators of  $V_\phi^e$  and  $V_\sigma^e(u)$  respectively.

Regarding bandwidth choice, let  $a(\sigma)$  be a scalar function of  $\Sigma$  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

$$h_{opt}(T) = \left[ \frac{\int a_0^\top(u) V_\sigma^e(u) a_0(u) w(u) du}{4 \int a_0^\top(u) b_\sigma^e(u) b_\sigma^e(u)^\top a_0(u) w(u) du} \right] T^{-1/5} \quad (32)$$

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

## 6 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). Table 2 provides summary statistics of the two series, and Figure 1 shows the price and return series. The return distributions slightly negatively skewed and highly leptokurtic. For both series, the Jarque-Bera test clearly rejects normality at the 1% significance level. Figure 2 presents the autocorrelation function for the return series and for the absolute return series along with the so-called Bartlett confidence bands (at 95%) computed under the assumption of independence. The ACF for absolute returns shows very strong evidence of dependence and quite long memory effects.

We next present our estimates of the sample first order local autocorrelation function of the absolute value of returns. We estimated the LACF of order one by the sample equivalent

$$\hat{\rho}_{|y|}(u, 1) = \frac{\sum_{t=1}^T w_{ut} y_t^c y_{t-1}^c}{\sum_{t=1}^T w_{ut} y_t^{c2}},$$

where  $y_t^c = |y_t| - \sum_{t=1}^T w_{ut} |y_t|$  and  $w_{ut}$  are kernel weights. In Figures 3 and 4 we present these estimates for the NASDAQ and Dow Jones over the full sample period along with the local Bartlett bands. There is strong evidence of time variation, which is also supported by higher lags and other powers of absolute returns. This evidence remains strong when we removed also a local mean effect from the returns, i.e., replacing  $y_t$  by  $y_t - \sum_{t=1}^T w_{ut} y_t$ . We also computed (not shown here) the time varying first four cumulants of the data and they show substantial time variation consistent with our discussions above. As discussed earlier there may be many reasons for the violation.

Stock index autocorrelation is a well-documented empirical feature, see e.g. Table 2.4 of Campbell et al. (1997), where for CRSP data typical first order autocorrelations of daily index returns range from 10 to 40 percent and can be explained, for example, by time varying risk premia or illiquid trading. The latter is confirmed by the observation that equally weighted indices, putting more weight on illiquid assets, show higher autocorrelation than value-weighted indices. The Dow Jones IA is a price-weighted index, the NASDAQ a market valued index. As reported in Table 2, the autocorrelations are closer to zero, but many coefficients are significant. For example, the first order autocorrelation of the Dow Jones index using the full sample is -0.039, significant at the 5% level.

In order to whiten the series for the subsequent modelling of volatility, we have considered vector autoregressions (VAR) with a maximum order 10, and selected the optimal one using the Schwarz information criterion (SIC). According to the SIC, a VAR(1) was chosen. The estimated VAR(1) model is given by

$$\begin{aligned} DJ_t &= \nu_1 + \Phi_{11} DJ_{t-1} + \Phi_{12} NQ_{t-1} + y_{1t} \\ NQ_t &= \nu_2 + \Phi_{21} DJ_{t-1} + \Phi_{22} NQ_{t-1} + y_{2t}, \end{aligned}$$

and parameter estimates are reported in Table 3 together with heteroskedasticity consistent standard errors. At the five percent significant level, the coefficients  $\Phi_{21}$  is significant for the full sample, which indicates a possible causality or spillover from the Dow Jones to the NASDAQ index. In the following, we model the volatility of the residuals of the VAR(1) model,  $y_t$ .

For the nonparametric estimation of  $\Sigma(u)$  we use the 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 (32), 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. The estimated unconditional volatilities are depicted as the solid lines in Figure 5, and the estimated unconditional correlation is the solid line in Figure 6.

To test for constancy of the estimated  $\Sigma(u)$ , we use the test statistic

$$Th \sum_{j=1}^5 (\hat{\sigma}(u_j) - \bar{\sigma})^\top (\hat{V}_\sigma^e(u_j))^{-1} (\hat{\sigma}(u_j) - \bar{\sigma})$$

where  $u_j = 0.1, 0.3, 0.5, 0.7, 0.9$ , which under the null converges to a  $\chi_5^2$  distribution. For our data, the statistic takes the value 30.34, which is significant at 1%, so that the null hypothesis of a constant  $\Sigma$  function is clearly rejected.

We now turn to the parametric part of the model which describes the conditional volatilities and correlation. To take into account the asymmetry of positive and negative news, we specify the parametric part of the volatility model as

$$G_t = I_N - AA^\top - BB^\top - CC^\top/2 + Au_{t-1}u_{t-1}^\top A^\top + BG_{t-1}B^\top + Cu_{t-1}^*u_{t-1}^{*\top}C^\top,$$

where  $u_t = \Sigma(t/T)^{-1/2}y_t$  and  $u_t^* = u_t \odot I(y_t < 0)$  and where  $C = \text{diag}(\gamma_1, \gamma_2)$ , see Cappiello, Engle, and Sheppard (2006). Parameter estimates are given in Table 5 for the full sample and for two subsamples.

As a specification test, we estimate the model for two subsamples, obtain parameter estimates  $\hat{\phi}_1$  and  $\hat{\phi}_2$ , and then test  $H_0 : \phi_1 = \phi_2$  using the Wald statistic

$$W = (T/2)(\hat{\phi}_1 - \hat{\phi}_2)^\top \hat{V}^{-1}(\hat{\phi}_1 - \hat{\phi}_2) \rightarrow \chi_p^2$$

under  $H_0$ , where  $\hat{V} = \hat{V}_1 + \hat{V}_2$ ,  $\hat{V}_1$  and  $\hat{V}_2$  being the asymptotic variances of  $\hat{\phi}_1$  and  $\hat{\phi}_2$ , respectively, and where  $p = \dim(\phi)$ . We compare our locally stationary model with a stationary asymmetric BEKK model given by

$$G_t = S - ASA^\top - BSB^\top - CSC^\top/2 + Au_{t-1}u_{t-1}^\top A^\top + BG_{t-1}B^\top + Cu_{t-1}^*u_{t-1}^{*\top}C^\top,$$

where  $S$  is the sample covariance matrix of the residuals of the VAR(1) model, see the parameter estimates in Table 4. For the stationary BEKK model,  $W = 62.30$  which clearly rejects  $H_0$  for all reasonable significance levels. For the locally stationary model,  $W = 32.05$ , which is much closer to the 1% critical value of 23.2 of a  $\chi_{10}^2$  distribution. Hence, there is less statistical evidence against stationarity of  $G_t$  in the locally stationary model than in the stationary model, which corroborates our model specification.

To understand the impact of the nonparametric part on the parameter estimates, consider a measure of persistence of volatilities and correlations, which in the BEKK model is often defined as the largest eigenvalue of the matrix  $A \otimes A + B \otimes B + C \otimes C/2$ , given in Tables 4 and 5. For the model that ignores the presence of long term trends (Table 4), these measures are very close to one, indicating an extremely high persistence of shocks in volatilities and correlations. The half-life of a shock in this stationary model is about 157 days. On the other hand, for the locally stationary model in Table 5, persistence is measured much smaller, about 93 % for the full sample, and the half-life is

only about ten days. This shows the strong impact of the presence of long term trends in the model on the parameter estimates of the conditional volatilities.

The estimated conditional and unconditional standard deviation and correlation plots are shown in Figures 5 and 6. 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  $\Sigma(u)$  were computed. 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 7.4. The largest eigenvalue explodes towards the end of 2008 reflecting the big increase in volatility.

As a diagnostic test, we use the multivariate Portmanteau statistic given by (see e.g. Lütkepohl, 2005)

$$P_r = T^2 \sum_{i=1}^r (T-i)^{-1} \text{Tr}(\hat{C}_i^\top \hat{C}_0^{-1} \hat{C}_i \hat{C}_0^{-1}), \quad \hat{C}_i = \frac{1}{T} \sum_{t=i+1}^T \hat{\varepsilon}_t \hat{\varepsilon}_{t-i}^\top. \quad (33)$$

If  $\varepsilon_t$  were the original data, the statistic  $P_r$  would have an asymptotic  $\chi^2$  distribution with  $rN^2$  degrees of freedom. We use  $P_r$  as a measure for residual autocorrelation rather than as a formal test statistic, as to our knowledge the asymptotic theory for the present model framework has not been worked out, and we choose the order  $r = 10$ . The values for the  $P_{10}$  statistics are also reported in Tables 4 and 5. Except for the second half of the sample, they would reject correct specification at 5% if the standard  $\chi_{rN^2}^2$  critical values were used. However, note that the statistics have improved for the locally stationary model compared with the stationary one.

We conclude with some further diagnostics on our model. In Figure 7 we present the ACF of the residual series and of the absolute value of the residuals. Evidently the model has greatly reduced the amount of dependence in both series. In Figure 8 and 9 we present the first order LACF along with corresponding Bartlett bands. Although there are some violations of the bands in the case of the NASDAQ, this might be explained by sampling error, since the confidence bands do not allow for estimated parameters and functions and so would likely be quite a lot larger. In the case of the Dow Jones, there is only one violation of the bands over the whole sample. Finally, we present the bivariate density plots of the data and the residuals showing how the model has almost sphered the data.

## 7 Extensions

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

### 7.1 Forecasting

We next discuss how to do forecasting for the future values of volatility. First, let us accept that the model precludes long run forecasting because it does not prescribe the future trajectory of  $\Sigma(\cdot)$ . However, it does permit short run forecasting where the definition of short run is somewhat flexible. In particular, by a Taylor series expansion we have

$$\Sigma(1 + \alpha h) \simeq \Sigma(1) + \alpha h \Sigma'(1) + \frac{\alpha^2 h^2}{2} \Sigma''(1)$$

for  $\alpha \in [0, 1]$ . Therefore, the simplest forecast of  $\Sigma(1 + \alpha h)$  is given by  $\Sigma(1)$ , the next is by  $\Sigma(1) + \alpha h \Sigma'(1)$ , etc. Therefore, to forecast  $\Omega(1 + \alpha h) = E[y_{T+\alpha Th} y_{T+\alpha Th}^\top | \mathcal{F}_T]$ , we use

$$\Omega_{1+\alpha h | \mathcal{F}_T} = \Sigma(1)^{1/2} E[G_{T+\alpha Th} | \mathcal{F}_T] \Sigma(1)^{1/2},$$

where  $E[G_{T+\alpha Th} | \mathcal{F}_T]$  is a standard forecasting problem for parametric models (we assume without loss of generality that  $T + \alpha Th$  is an integer). In practice one replaces unknown quantities by their estimates. Specifically, one can only use one-sided (or boundary) kernels for this purpose. Li and Heckman (1997) have established some results in the purely nonparametric case.

### 7.2 Discontinuities or Regime shifts

One can allow  $\Sigma$  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  $\Sigma_d$  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  $\Sigma_d$ . Let

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

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

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

for some weighting sequence  $\{w_{\ell-}, w_{\ell+}\}_{\ell=1}^L$  with  $\sum_{\ell=1}^L w_{\ell-} + w_{\ell+} = 1$ . See Pelletier (2006) for an alternative approach.

### 7.3 Exogenous covariates

One could suppose also that  $\Sigma$  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  $\eta$ . This is like in the multiplicative model of Rangel and Engle (2009). It is straightforward to modify the estimation algorithms to accommodate this case.

### 7.4 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  $\Lambda(t/T) = \text{diag}\{\lambda_1(t/T), \dots, \lambda_N(t/T)\}$ . Now suppose that  $\lambda_j(\cdot) \equiv 0$  for  $j = K+1, \dots, N$ , 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).

## 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. The application demonstrates the value of our multivariate model.

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. Many practical issues arise. For instance, how can we distinguish breaks in the low frequency covariance component from breaks in the parameters of the high frequency one? What is a good way to choose bandwidth in the presence of possible breaks? We have provided a conceptual framework in which such issues could be analysed.

## A Appendix

### A.1 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  $\Phi$  is compact.

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

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

(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  $\phi_0$  is an interior point of  $\Phi$ .

(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)  $h(T) \rightarrow 0$  as  $T \rightarrow \infty$  such that  $Th^2 \rightarrow \infty$  and  $Th^4 \rightarrow 0$ .

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

The set of assumptions (A2-A8) is the weakest set of assumptions currently available to prove consistency and asymptotic normality of the BEKK model, see Hafner and Preminger (2009). See also Comte and Lieberman (2003) who need finite eight moments instead of finite six moments in A5. For the standard BEKK model with Gaussian innovations, primitive conditions on the parameters for A5 to hold could be worked out. The assumptions A10(a) are used to derive the properties of the estimators of  $\phi$ , while assumptions A10(b) are used to derive the properties of the estimators of  $\sigma(u)$ .

Define

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

Assumption A7 implies that  $\phi_0$  is the unique minimizer of  $\ell(\phi)$ .

## A.2 Proof of Proposition 1

Consider the following likelihood function

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

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

$$\begin{aligned}\frac{\partial l_t}{\partial \phi} &= \rho_t \text{vec}(\varepsilon_t \varepsilon_t^\top - I_N) \\ \rho_t &= \frac{\partial \text{vec}(G_t)^\top}{\partial \phi} (G_t^{-1/2} \otimes G_t^{-1/2})\end{aligned}$$

The matrix function  $\Sigma(t/T)$  is nonparametric. Consider a parametric submodel,  $\Sigma_\theta(t/T)$ , where  $\theta \in \Theta \subset \mathbb{R}^J$  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  $\theta$  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

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

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

$$P_t = \mathbb{E}[\rho_t] \text{vec}(\varepsilon_t \varepsilon_t^\top - I_N)$$

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

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

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

### A.3 Proof of Theorem 1

Let  $V_{tT} = V_t(t/T)$  with  $V_t(u) = \Sigma(u)^{1/2}[G_t^{1/2}\varepsilon_t\varepsilon_t^\top G_t^{1/2} - I_N]\Sigma(u)^{1/2}$  and  $v_t(u) = \text{vech}(V_t(u)) = D_N [\Sigma(u)^{1/2} \otimes \Sigma(u)^{1/2}] D_N^+ z_t$  where  $z_t = \text{vech}(G_t^{1/2}\varepsilon_t\varepsilon_t^\top G_t^{1/2} - I_N)$  is stationary and geometrically mixing. To establish (24) we use the following lemma.

LEMMA 1. *For some bounded continuous function  $b(u)$ ,*

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

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

It then follows by a CLT for mixing processes that

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

The bias function  $b(u)$  is  $\mu_2(K)\sigma''(u)/2$ .

We first establish consistency of  $\tilde{\phi}$ . By the triangle inequality

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

It follows from standard results that

$$\sup_{\phi \in \Phi} \left| T^{-1} \ell_T(\phi) - \ell(\phi) \right| = o_p(1). \quad (34)$$

We shall show that

$$\sup_{\phi \in \Phi} \left| T^{-1} \tilde{\ell}_T(\phi) - T^{-1} \ell_T(\phi) \right| = o_p(1). \quad (35)$$

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

Using the expression given in (14) for  $\tilde{\ell}_T(\phi)$ , we have

$$\tilde{\ell}_T(\phi) - \ell_T(\phi) = - \sum_{t=1}^T \log \det G_t(\phi) \tilde{G}_t^{-1}(\phi) - \sum_{t=1}^T y_t^\top \left( \tilde{G}_t^{-1}(\phi) - G_t^{-1}(\phi) \right) y_t,$$

so we obtain a representation for  $G_t(\phi)\tilde{G}_t^{-1}(\phi)$  and  $\tilde{G}_t^{-1}(\phi) - G_t^{-1}(\phi)$ . Note that  $\tilde{G}_t^{-1} = G_t^{-1} - G_t^{-1}(\tilde{G}_t - G_t)G_t^{-1} + o_p(\delta_T)$ , where we can show that  $\delta_T = T^{-1/2}$ . We have

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

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

Then, since

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

for any  $u \in [0, 1]$ , we have

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

Therefore,

$$\begin{aligned} \Delta_s &= \left[ \tilde{\Sigma}(s/T)^{-1/2} - \Sigma(s/T)^{-1/2} \right] y_s y_s^\top \Sigma(s/T)^{-1/2} \\ &\quad + \Sigma(s/T)^{-1/2} y_s y_s^\top \left[ \tilde{\Sigma}(s/T)^{-1/2} - \Sigma(s/T)^{-1/2} \right] \\ &\quad + \left[ \tilde{\Sigma}(s/T)^{-1/2} - \Sigma(s/T)^{-1/2} \right] y_s y_s^\top \left[ \tilde{\Sigma}(s/T)^{-1/2} - \Sigma(s/T)^{-1/2} \right] \\ &= -\frac{1}{2} \Sigma^{-3/4}(s/T) (\tilde{\Sigma}(s/T) - \Sigma(s/T)) \Sigma^{-3/4}(s/T) y_s y_s^\top \Sigma(s/T)^{-1/2} \\ &\quad - \frac{1}{2} \Sigma(s/T)^{-1/2} y_s y_s^\top \Sigma^{-3/4}(s/T) (\tilde{\Sigma}(s/T) - \Sigma(s/T)) \Sigma^{-3/4}(s/T) + o_p(\delta_T), \end{aligned}$$

so that

$$\begin{aligned} \tilde{G}_t(\phi) - G_t(\phi) &= -\frac{1}{2} \sum_{j=1}^{t-1} B^{j-1} A \Sigma^{-3/4}((t-j)/T) \left[ \tilde{\Sigma}((t-j)/T) - \Sigma((t-j)/T) \right] \\ &\quad \times \Sigma^{-3/4}((t-j)/T) y_{t-j} y_{t-j}^\top \Sigma(t-j/T)^{-1/2} A^\top (B^\top)^{j-1} \\ &\quad + \sum_{j=1}^{t-1} B^{j-1} A \Sigma(t-j/T)^{-1/2} y_{t-j} y_{t-j}^\top \Sigma^{-3/4}((t-j)/T) \\ &\quad \times \left[ \tilde{\Sigma}((t-j)/T) - \Sigma((t-j)/T) \right] \Sigma^{-3/4}((t-j)/T) A^\top (B^\top)^{j-1} \\ &\quad + o_p(\delta_T). \end{aligned} \tag{36}$$

We can bound this by the uniform convergence rate of  $\tilde{\Sigma}(\cdot)$  times a factor that is  $O_p(1)$ .

Note that

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T y_t^\top \tilde{G}_t^{-1}(\phi) y_t - \frac{1}{T} \sum_{t=1}^T y_t^\top G_t^{-1}(\phi) y_t &= -\frac{1}{T} \sum_{t=1}^T y_t^\top G_t^{-1}(\phi) \left[ \tilde{G}_t(\phi) - G_t(\phi) \right] \tilde{G}_t^{-1}(\phi) y_t \\ \frac{1}{T} \sum_{t=1}^T \log \det G_t(\phi) \tilde{G}_t^{-1}(\phi) &\simeq \frac{1}{T} \sum_{t=1}^T \text{Tr} \left[ (\tilde{G}_t(\phi) - G_t(\phi)) G_t^{-1}(\phi) \right]. \end{aligned}$$

Therefore, it suffices to show that

$$\max_{1 \leq t \leq T} \sup_{\phi \in \Phi} \left\| \tilde{G}_t(\phi) - G_t(\phi) \right\| = o_p(1) \quad (37)$$

$$\min_{1 \leq t \leq T} \inf_{\phi \in \Phi} \lambda_{\min}(G_t(\phi)) > 0. \quad (38)$$

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

We now turn to asymptotic normality of  $\tilde{\phi}$ . The general strategy is to show the following:

$$\sqrt{T}(\tilde{\phi} - \phi) = - \left[ \frac{\partial^2 \tilde{\ell}_T(\phi_0)}{\partial \phi \partial \phi^\top} \right]^{-1} \sqrt{T} \frac{\partial \tilde{\ell}_T(\phi_0)}{\partial \phi} + o_p(1), \quad (39)$$

$$\sqrt{T} \frac{\partial \tilde{\ell}_T(\phi_0)}{\partial \phi} \Rightarrow N(0, Q) \quad (40)$$

$$\frac{\partial^2 \tilde{\ell}_T(\phi_0)}{\partial \phi \partial \phi^\top} \xrightarrow{P} J. \quad (41)$$

We will consider the quantities:

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

We have

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

because  $I_N - u_t u_t^\top G_t^{-1}$  is a martingale difference sequence ( $\partial \tilde{G}_t / \partial \phi_i - \partial G_t / \partial \phi_i$  like  $\tilde{G}_t - G_t$  depends only on past). Then note that

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

to obtain that

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

Consider

$$\begin{aligned}
& \frac{1}{\sqrt{T}} \sum_{t=1}^T \text{Tr} \left[ \Sigma^{-3/4} (\tilde{\Sigma} - \Sigma) \Sigma^{-1/4} (t/T) \frac{\partial G_t}{\partial \phi_i} G_t^{-1} \right] \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^T \text{Tr} \left[ \Sigma^{-3/4} \frac{1}{T} \sum_{s=1}^T K_h((t-s)/T) V_{sT} \Sigma^{-1/4} (t/T) \frac{\partial G_t}{\partial \phi_i} G_t^{-1} \right] + o_p(1) \\
&= \frac{1}{\sqrt{T}} \sum_{s=1}^T \text{Tr} \left[ \frac{1}{T} \sum_{t=1}^T K_h((t-s)/T) \Sigma^{-3/4} (t/T) V_{sT} \Sigma^{-1/4} (t/T) \frac{\partial G_t}{\partial \phi_i} G_t^{-1} \right] + o_p(1) \\
&= \frac{1}{\sqrt{T}} \sum_{s=1}^T \frac{1}{T} \sum_{t=1}^T K_h((t-s)/T) \text{vec} \left( G_t^{-1} \frac{\partial G_t}{\partial \phi_i} \right)^\top [\Sigma^{-1/4} (t/T) \otimes \Sigma^{-3/4} (t/T)] \text{vec} (V_{sT}) + o_p(1) \\
&= \frac{1}{\sqrt{T}} \sum_{s=1}^T \text{vec} \left( E \left[ G_t^{-1} \frac{\partial G_t}{\partial \phi_i} \right] \right)^\top [\Sigma^{-1/4} (s/T) \otimes \Sigma^{-3/4} (s/T)] \text{vec} (V_{sT}) + o_p(1) \\
&= \text{vec} \left( E \left[ G_t^{-1} \frac{\partial G_t}{\partial \phi_i} \right] \right)^\top \frac{1}{\sqrt{T}} \sum_{s=1}^T [\Sigma^{1/4} (s/T) \otimes \Sigma^{-1/4} (s/T)] \text{vec} [(G_s^{1/2} \varepsilon_s \varepsilon_s^\top G_s^{1/2} - I_N)] + o_p(1) \tag{42}
\end{aligned}$$

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

$$\begin{aligned}
& \frac{1}{T} \sum_{t=1}^T K_h((t-s)/T) \text{vec} \left( G_t^{-1} \frac{\partial G_t}{\partial \phi_i} \right)^\top [\Sigma^{-1/4} (t/T) \otimes \Sigma^{-3/4} (t/T)] \\
&= \frac{1}{T} \sum_{t=1}^T K_h((t-s)/T) \text{vec} \left( E \left[ G_t^{-1} \frac{\partial G_t}{\partial \phi_i} \right] \right)^\top [\Sigma^{-1/4} (t/T) \otimes \Sigma^{-3/4} (t/T)] + o_p(1) \tag{43}
\end{aligned}$$

$$= \text{vec} \left( E \left[ G_t^{-1} \frac{\partial G_t}{\partial \phi_i} \right] \right)^\top [\Sigma^{-1/4} (s/T) \otimes \Sigma^{-3/4} (s/T)] + o_p(1). \tag{44}$$

The term (42) is asymptotically normal. Using (36), we obtain

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

Then

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

So we need the probability limit of

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

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  $\text{vec}(V_{tT}) = [\Sigma(t/T)^{1/2} \otimes \Sigma(t/T)^{1/2}] \text{vec}(G_t^{1/2} \varepsilon_t \varepsilon_t^\top G_t^{1/2} - I_N)$ , we have

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

where

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

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

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

and

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

where  $\tilde{\zeta}_t = \rho_t \text{vec} (\varepsilon_t \varepsilon_t^\top - I_N)$  is a martingale difference sequence. Furthermore,  $\rho_t$  is defined in (23),  $z_t$  in (4),  $C(t/T) = (C_1(t/T)^\top, \dots, C_p(t/T)^\top)^\top$ , and

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

We have

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

Letting  $\Gamma_u = E[z_s z_{s+u}^\top]$  as defined in (20),

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

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

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

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

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

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

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

Finally, we obtain the asymptotic covariance matrix,

$$Q = \lim_{T \rightarrow \infty} \text{var} \frac{1}{\sqrt{T}} \sum_{t=1}^T \aleph_t = \Phi + H + H^\top + E[\rho_t \Xi_t \rho_t^\top] \quad (45)$$

where  $\Xi_t$  is defined in (27).

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

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

with

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

after cancellation.

Q.E.D.

#### A.4 Proof of Theorem 2

For some  $\bar{\theta}(u)$  between  $\tilde{\theta}(u)$  and  $\theta(u)$  we have by Taylor expansion

$$\begin{aligned} \widehat{\theta}(u) - \theta(u) &= \tilde{\theta}(u) - \theta(u) - \left[ \frac{\partial^2 \tilde{L}_T(\tilde{\theta}; u)}{\partial \theta \partial \theta^\top} \right]^{-1} \left[ \frac{\partial \tilde{L}_T(\theta; u)}{\partial \theta} + \frac{\partial^2 \tilde{L}_T(\bar{\theta}; u)}{\partial \theta \partial \theta^\top} (\tilde{\theta}(u) - \theta(u)) \right] \\ &= - \left[ \frac{\partial^2 \tilde{L}_T(\theta; u)}{\partial \theta \partial \theta^\top} \right]^{-1} \frac{\partial \tilde{L}_T(\theta; u)}{\partial \theta} + o_p \left( \|\tilde{\theta}(u) - \theta(u)\| \right), \end{aligned}$$

using  $\tilde{\phi} = \phi_0 + O_p(T^{-1/2})$  and  $\sup_{u \in [0,1]} |\tilde{\sigma}(u) - \sigma(u)| = O_p(h^2) + O_p(\sqrt{\log T / Th})$ .

The local score function is

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

This score function is a kernel weighted sum of martingale difference sequence errors with finite variance and higher moments. The conditional variance process is mixing and satisfies a law of

large numbers. Therefore, the score function satisfies a CLT, Hall and Heyde (1980, p11). As  $G_t$  and  $\Sigma(t/T)$  are symmetric, we have (Lütkepohl (1996), p.190, (5))  $\partial \text{vec}(\Omega_t)^\top / \partial \theta = 2D_N^\top (I_N \otimes \Sigma(t/T)^{1/2} G_t) D_N D_N^+$ . 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

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

We calculate the variance matrix

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

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

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

In conclusion,

$$\sqrt{T}h \left( \hat{\theta}(u) - \theta(u) \right) \implies N(0, \|K\|_2^2 \Lambda(u)^{-1} \Psi(u) \Lambda(u)^{-1}).$$

Then note that

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

Therefore,

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

and so (29) follows.

Q.E.D.

### A.5 Proof of Theorem 3

For some  $\bar{\phi}$  between  $\phi$  and  $\phi_0$  we have by Taylor expansion

$$\begin{aligned}\widehat{\phi} - \phi_0 &= \widetilde{\phi} - \phi_0 - \left[ \frac{\partial^2 \widehat{\ell}_T^*(\widetilde{\phi})}{\partial \phi \partial \phi^\top} \right]^{-1} \left[ \frac{\partial \widehat{\ell}_T^*(\phi_0)}{\partial \phi} + \frac{\partial^2 \widehat{\ell}_T^*(\bar{\phi})}{\partial \phi \partial \phi^\top} (\widetilde{\phi} - \phi_0) \right] \\ &= - \left[ \frac{\partial^2 \widehat{\ell}_T^*(\phi_0)}{\partial \phi \partial \phi^\top} \right]^{-1} \frac{\partial \widehat{\ell}_T^*(\phi_0)}{\partial \phi} + o_p(T^{-1/2}) \\ &= - \left[ \frac{\partial^2 \ell_T^*(\phi_0)}{\partial \phi \partial \phi^\top} \right]^{-1} \frac{\partial \ell_T^*(\phi_0)}{\partial \phi} + o_p(T^{-1/2})\end{aligned}$$

because for any sequence  $\phi_T \rightarrow \phi_0$

$$\frac{\partial^2 \widehat{\ell}_T^*(\phi_T)}{\partial \phi \partial \phi^\top} - \frac{\partial^2 \ell_T^*(\phi_0)}{\partial \phi \partial \phi^\top} \xrightarrow{P} 0.$$

For asymptotic normality of  $\widehat{\phi}$  we show the following:

$$\sqrt{T} \frac{\partial \ell_T^*(\phi_0)}{\partial \phi} \Longrightarrow N(0, Q^e) \quad (46)$$

$$\frac{\partial^2 \ell_T^*(\phi_0)}{\partial \phi \partial \phi^\top} \xrightarrow{P} J^e. \quad (47)$$

We apply a CLT for martingale differences, Hall and Heyde (1980, p11), and an LLN for mixing processes. Q.E.D.

### A.6 Proof of Lemmas

Here we give proofs of Lemmas 1-2.

PROOF OF LEMMA 1. We have

$$\begin{aligned}\widetilde{\Sigma}(u) - \Sigma(u) &= \frac{\sum_{t=1}^T K_h(u - t/T) [y_t y_t^\top - \Sigma(t/T)]}{\sum_{t=1}^T K_h(u - t/T)} + \frac{\sum_{t=1}^T K_h(u - t/T) [\Sigma(t/T) - \Sigma(u)]}{\sum_{t=1}^T K_h(u - t/T)} \\ &= \frac{1}{Th} \sum_{t=1}^T K_h(u - t/T) V_{tT} + \frac{h^2}{2} \Sigma''(u) \int s^2 K(s) ds + o(h^2) + o_p(T^{-1/2} h^{-1/2}) \\ &= \frac{1}{Th} \sum_{t=1}^T K_h(u - t/T) V_t(u) + \frac{h^2}{2} \Sigma''(u) \int s^2 K(s) ds + o(h^2) + o_p(T^{-1/2} h^{-1/2})\end{aligned} \quad (48)$$

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

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

The error term in (48) is uniform in  $u$ , applying results from, for example, Masry (1996).<sup>6</sup>

Q.E.D.

PROOF OF LEMMA 2. The proof of Lemma 2 follows from the arguments given in Theorem 2 and Lemma 1. Q.E.D.

## B Appendix

### B.1 Derivatives w.r.t. $\phi$ and $\sigma$

$$\begin{aligned} \ell_t(\phi) &= \log |G_t(\phi)| + u_t^\top G_t^{-1}(\phi) u_t, \quad \phi = (\text{vec}(A)^\top, \text{vec}(B)^\top)^\top \\ u_t &= \Sigma(t/T)^{-1/2} y_t \\ G_t(\phi) &= \sum_{j=0}^{t-1} B^j (I_N - AA^\top - BB^\top + Au_{t-j} u_{t-j}^\top A^\top) (B^j)^\top \\ g_t = \text{vech}(G_t) &= \sum_{j=0}^{t-1} D_N^\top (B \otimes B)^j D_N \text{vech}(I_N - AA^\top - BB^\top + Au_{t-j} u_{t-j}^\top A^\top) \end{aligned} \tag{49}$$

#### B.1.1 First derivatives

Notation:  $\phi_i$  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:

---

<sup>6</sup>Actually, he considered a stochastic covariate process, but the arguments are even simpler in our case of deterministic covariate  $t/T$ .

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

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

### B.1.2 Second derivatives

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

$$\begin{aligned}
\frac{\partial^2 l_t}{\partial \phi_i \partial \phi_j} &= \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] \\
\frac{\partial^2 l_t}{\partial \phi_i \partial \sigma_j} &= \text{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. \\
&\quad \left. + 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} \right. \\
&\quad \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 \{ 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} \} (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}} \{ 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} \} (B^\top)^m \\
&\quad + \sum_{m=0}^{t-1} B^m \{ 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} \} \frac{\partial (B^\top)^m}{\partial B_{ij}} \\
\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}} (I_N - AA^\top - BB^\top + Au_{t-m} u_{t-m}^\top A^\top) (B^\top)^m \\
I_2 &= \sum_{m=0}^{t-1} \frac{\partial B^m}{\partial B_{ij}} (I_N - AA^\top - BB^\top + Au_{t-m} u_{t-m}^\top A^\top) \frac{\partial (B^\top)^m}{\partial B_{kl}} \\
I_3 &= \sum_{m=0}^{t-1} \frac{\partial B^m}{\partial B_{ij}} (-J_{kl} B^\top - B J_{lk}) (B^\top)^m \\
I_4 &= \sum_{m=0}^{t-1} \frac{\partial B^m}{\partial B_{kl}} (-J_{ij} B^\top - B J_{ji}) (B^\top)^m \\
I_5 &= -B^m J_{ij} J_{lk} (B^\top)^m
\end{aligned}$$

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

where:

$$\begin{aligned}
\frac{\partial^2 B^m}{\partial B_{ij} \partial B_{kl}} &= \sum_{n=0}^{m-1} \frac{\partial B^n}{\partial B_{kl}} J_{ij} B^{m-1-n} + B^n J_{ij} \frac{\partial B^{m-1-n}}{\partial B_{ij}}, \\
\frac{\partial \text{vech} \Sigma^{-1/2}(t/T)}{\partial \sigma_k} &= -D_N^+ \{ \Sigma^{-1/2}(t/T) \otimes \Sigma^{-1/2}(t/T) \} D_N \frac{\partial \text{vech} \Sigma^{1/2}(t/T)}{\partial \sigma_k},
\end{aligned}$$

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

## References

- [1] ALEXANDER, C. (2001) Orthogonal GARCH, *Mastering Risk*, Volume 2.
- [2] BAUWENS, L., S. LAURENT, AND J.V.K. ROMBOUTS (2006), Multivariate GARCH models: A survey. *Journal of Applied Econometrics* 21, 79-109.
- [3] BICKEL, P.J., KLAASSEN, C. A. J., RITOV, Y. AND J. A. WELLNER (1993). *Efficient and adaptive estimation for semiparametric models*. The Johns Hopkins University Press, Baltimore and London.
- [4] 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.
- [5] BOSWIJK, P. AND VAN DER WEIDE, R. (2006). Wake me up before you GO-GARCH, discussion paper, UvA Amsterdam.
- [6] CAMPBELL, J.Y., LO, A.W. AND MACKINLAY, A.C. (1997). *The Econometrics of Financial Markets*, Princeton University Press, Princeton.
- [7] CAPPIELLO, L., R.F.ENGLE AND K.SHEPPARD (2006) Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns, *Journal of Financial Econometrics*, Vol. 4, No. 4,537-572
- [8] COLACITO R., R. ENGLE AND E. GHYSELS (2009), “A Component Model for Dynamic Correlations” Working Paper, NYU Stern.
- [9] COMTE F. AND O. LIEBERMAN (2003). Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis*, 84, 61-84.
- [10] DAHLHAUS, R. (1997). Fitting time series models to nonstationary processes. *Annals of Statistics* 25, 1-37.
- [11] DAHLHAUS, R. AND S. SUBBA RAO (2006). Statistical inference for time-varying ARCH processes. *Annals of Statistics* 34, 1075-1114.
- [12] DREES, H. AND C. STARICA (2003). A simple non-stationary model for stock returns. Preprint.
- [13] DROST, F.C., C.A.J. KLAASSEN, AND B. WERKER (1997). Adaptive estimation in time-series models. *Annals of Statistics* 35, 786-817.
- [14] 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.

- [15] ENGLE, R.F. AND RANGEL (2008), The spline GARCH model for unconditional volatility and its global macroeconomic causes, *Review of Financial Studies* 21, 1187-1222.
- [16] ENGLE, R.F. AND K. SHEPPARD (2001). Theoretical and Empirical properties of Dynamic Conditional Correlation Multivariate GARCH, UCSD working paper.
- [17] HAFNER, C.M. AND A. PREMINGER (2009). On asymptotic theory of multivariate GARCH models, *Journal of Multivariate Analysis* 100 (2009), 2044-2054.
- [18] HALL, P. AND C. C. HEYDE (1980). *Martingale Limit Theory and its Application*. Academic Press, New York.
- [19] JONDEAU, E. AND M. ROCKINGER (2006). The Copula-GARCH model of conditional dependencies : An international stock market application, *Journal of International Money and Finance* 25 827-853
- [20] KAWAKATSU, H. (2006). Matrix exponential GARCH, *Journal of Econometrics* 134, 95-128
- [21] LANNE, M., AND P.SAIKKONEN (2005) A multivariate Generalized Orthogonal Factor GARCH model, *Helsinki Center of Economic Research*, Discussion Paper N0. 63.
- [22] LI, X., AND N.E. HECKMAN (1997). Local Linear Forecasting. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.57.4629>
- [23] LÜTKEPOHL, H. (1996). *Handbook of Matrices*. Wiley, New York.
- [24] LÜTKEPOHL, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer, Berlin.
- [25] MASRY, E. (1996), "Multivariate local polynomial regression for time series: Uniform strong consistency and rates," *J. Time Ser. Anal.* 17, 571-599.
- [26] MERCURIO, D. AND V. SPOKOINY (2004) Statistical inference for time-inhomogeneous volatility models. *The Annals of Statistics* 32, 577-602
- [27] PATTON, A (2006). Estimation of Multivariate Models for Time Series of Possibly Different Lengths, *Journal of Applied Econometrics*, 21, 147-174
- [28] PATTON, A. (2006). Modelling Asymmetric Exchange Rate Dependence, *International Economic Review*, Vol. 47, No. 2
- [29] PELLETIER, D. (2006). Regime Switching for Dynamic Correlations, *Journal of Econometrics* 131 445-473
- [30] PESARAN, H. AND A. TIMMERMAN (2004). How costly is it to ignore breaks when forecasting the direction of a time series. *International Journal of Forecasting* 20, 411-425.

- [31] POHLZEHL, J. AND V. SPOKOINY (2004). Varying coefficient GARCH versus local constant volatility modelling: comparison of the predictive power. WIAS Discussion Paper 977.
- [32] RANGEL, J.G. AND R.F. ENGLE (2009). The Factor-spline GARCH model for High and Low frequency correlations. Banco de Mexico working paper no 2009-03.
- [33] ROBINSON, P.M. (1989). Nonparametric estimation of time-varying parameters. In *Statistical Analysis and Forecasting of Economic Structural Change* (P. Hackl, ed) 253-264. Springer, Berlin.
- [34] RODRIGUEZ-POO, J.R. AND LINTON, O.B. (2001), Nonparametric Factor Analysis for Residual Time Series. *TEST* 10, 161-182.
- [35] SEVERINI, T.A. AND W.H. WONG (1992). Profile likelihood and conditionally parametric models. *The Annals of Statistics* 20, 1768-1802.
- [36] SILVENNOINEN, A. AND T. TERASVIRTA (2009a). Modelling multivariate autoregressive conditional heteroskedasticity with the smooth transition conditional correlation GARCH model. *Journal of Financial Econometrics* 7, 373-411.
- [37] SILVENNOINEN, A. AND T. TERASVIRTA (2009b). Multivariate GARCH Models. in *Handbook of Financial Time Series*, Springer. eds. T.G. Anderson, R.A. Davis, J-P. Kreiss, ad T. Mikosch
- [38] SPOKOINY, V. (1998). Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *The Annals of Statistics* 26, 1356-1378.
- [39] STARICA, C. (2003): "Is GARCH(1,1) as good a model as the Nobel prize accolades would imply?" Unpublished manuscript
- [40] TSE, Y.K AND TSUI, A.K.C (2002) A Multivariate Generalized Autoregressive Conditional Heteroscedasticity Model with Time-Varying Correlations, *Journal of Business & Economic Statistics*, Vol.20, No. 3 (Jul. 2002), pp. 351-362I.
- [41] VAN DER WEIDE, R. (2002) GO-GARCH: A Multivariate Generalized Orthogonal GARCH Model, *Journal of Applied Econometrics*, Vol. 17, No. 5, pp. 549-564
- [42] VRONTOS, D., P.DELLAPORTAS AND D.N.POLITIS (2003) A full-factor multivariate GARCH model, *Econometrics Journal*, Volume 6, pp311-333

## C Tables

	50-60	60-70	70-80	80-90	90-00	00-09
$c$	0.0213	0.0019	-0.0106	0.0111	0.0043	-0.0234
$\rho_1$	0.1584	0.2229	0.2429	0.0635	0.0603	-0.0814
$\rho_2$	-0.0977	-0.0288	-0.0563	-0.0033	0.0120	-0.0445
$\omega$	0.0425	0.0166	0.0039	0.0605	0.0132	0.0121
$\beta$	0.8330	0.8086	0.9543	0.8620	0.9230	0.9416
$\gamma$	0.0584	0.0574	0.0073	0.0362	0.0016	-0.0179
$\delta$	0.0692	0.2031	0.0691	0.0980	0.1264	0.1278
$R^2$	0.0165	0.0320	0.0515	0.0025	0.0000	0.0171
$mper$	0.0607	0.1941	0.1866	0.0602	0.0723	-0.1259
$vper$	0.9260	0.9676	0.9962	0.9472	0.9878	0.9876
$\mu_{year}$	0.1199	0.0714	0.0346	0.0939	0.0768	0.0161
$\sigma_{year}$	0.1144	0.1080	0.1520	0.1616	0.1571	0.1492

Table 1. Daily S&amp;P500 total returns

$$y_t = c + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \varepsilon_t \sigma_t$$

$$\sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \gamma u_{t-1}^2 + \delta u_{t-1}^2 1(u_{t-1} < 0)$$

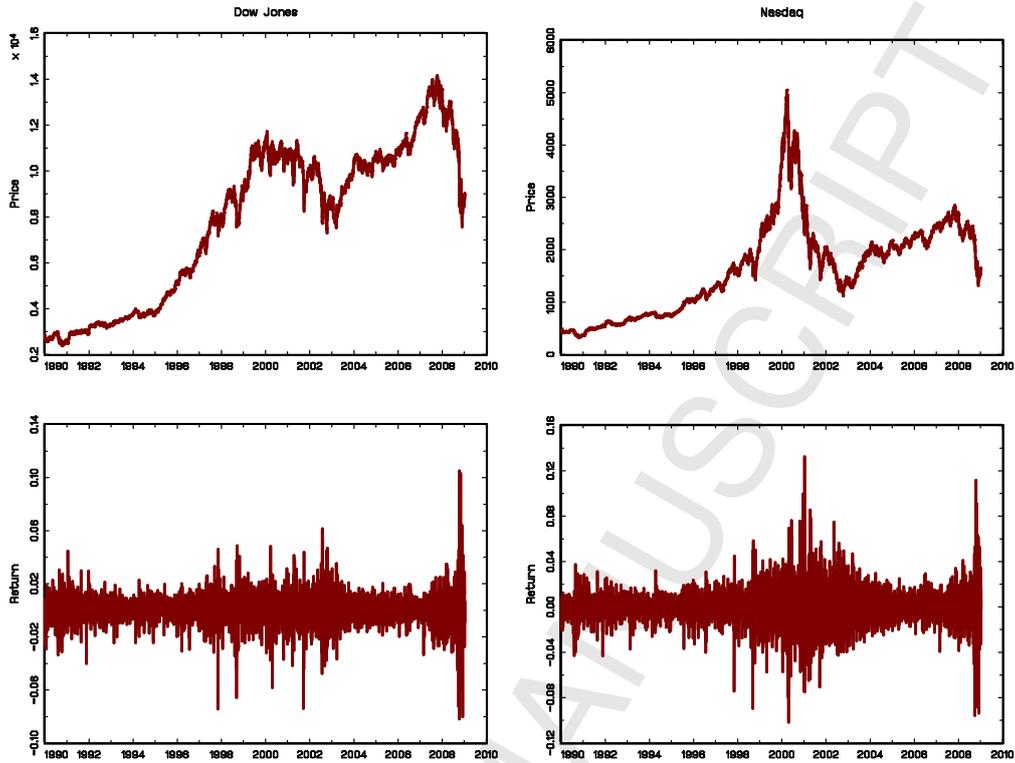


Figure 1. Stock price and total return

	full sample		first half		second half	
	DJ	NQ	DJ	NQ	DJ	NQ
mean	2.37E-04	2.60E-04	5.56E-04	7.24E-04	-8.12E-05	-1.95E-04
std.dev.	0.0110	0.0156	0.0088	0.0108	0.0127	0.0192
skew	-0.1483	-0.0562	-0.4354	-0.6099	0.0127	0.0979
kurt	12.20	9.24	8.57	8.2901	11.45	7.2277
$\rho(1)$	-0.039	0.009	0.029	0.115	-0.074	-0.026
$\rho(2)$	-0.063	-0.055	-0.030	-0.005	-0.081	-0.073
$\rho(3)$	0.033	0.022	-0.036	0.015	0.066	0.023
$\rho(4)$	-0.019	0.000	-0.013	-0.011	-0.023	0.002
$\rho(5)$	-0.033	-0.025	-0.003	-0.007	-0.049	-0.032

Table 2. Summary statistics for Dow Jones and NASDAQ returns

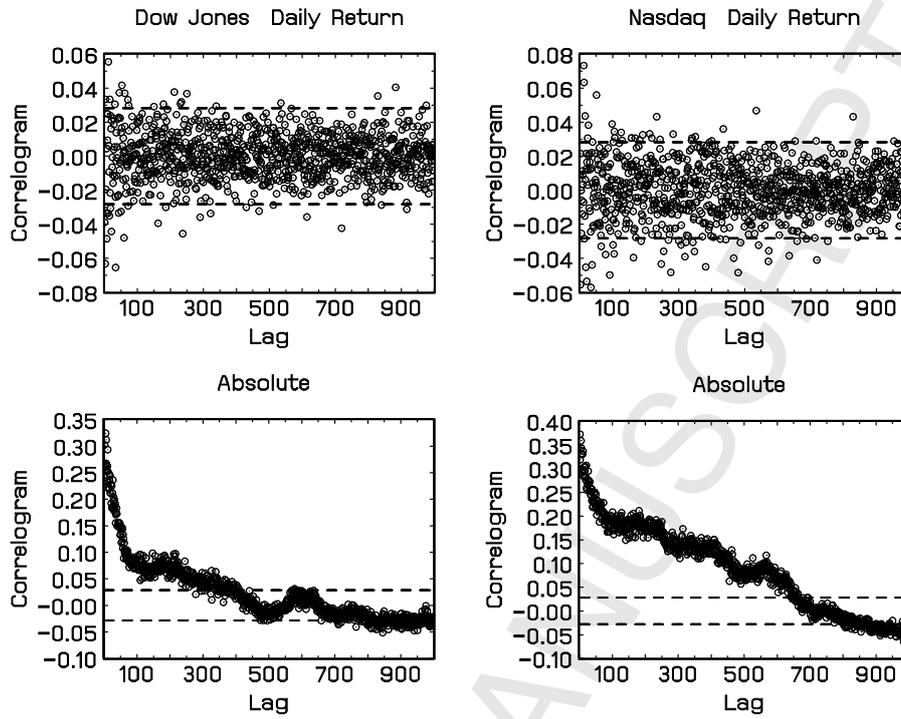


Figure 2. Autocorrelations of returns and absolute returns.

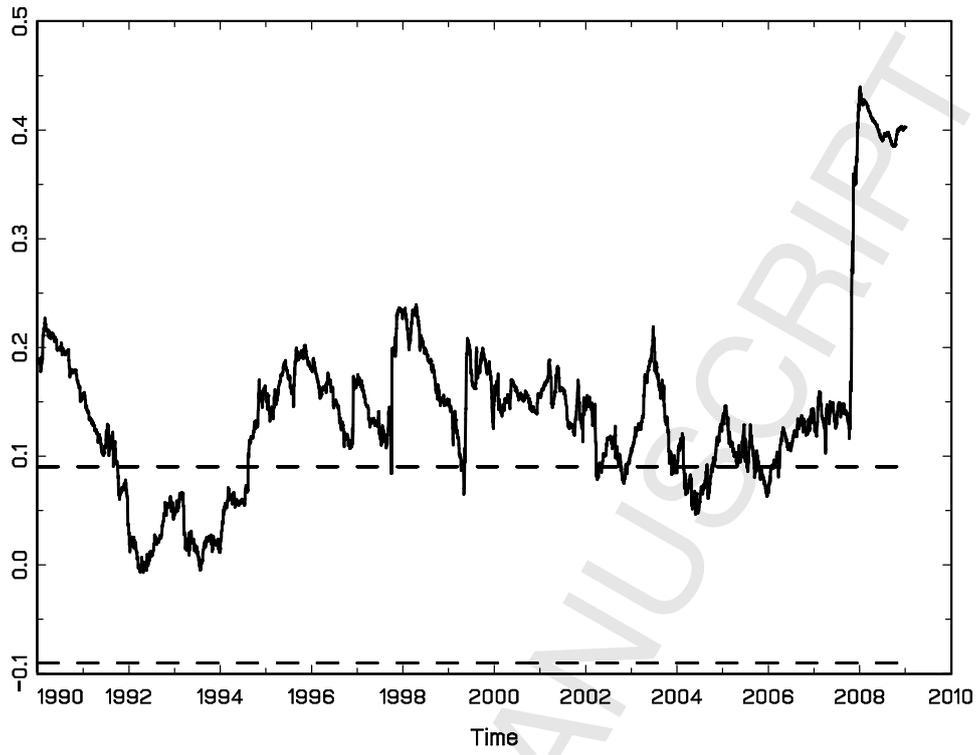


Figure 3. NASDAQ Local first order autocorrelation function

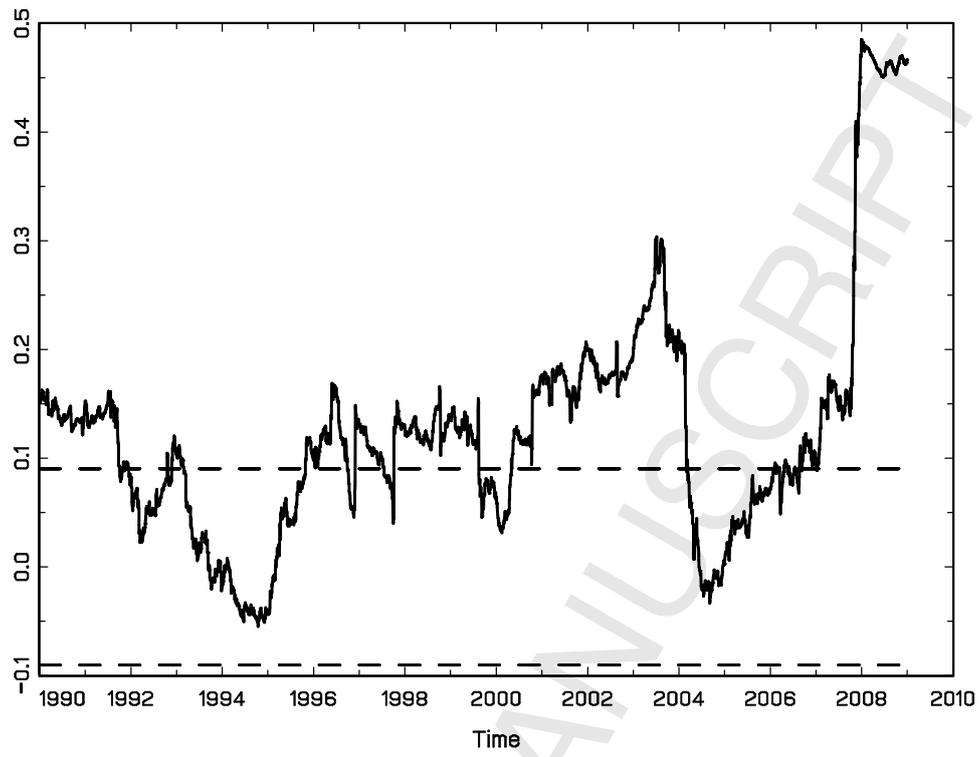


Figure 4. DOW local first order autocorrelation function

	full sample		first half		second half	
	est.	s.e.	est.	s.e.	est.	s.e.
$\nu_1$	0.000247	0.000160	0.000530	0.000187	-8.58E-05	0.000259
$\nu_2$	0.000268	0.000227	0.000644	0.000229	-0.000204	0.000391
$\Phi_{11}$	-0.047210	0.033287	-0.012809	0.036635	-0.076121	0.047746
$\Phi_{12}$	0.007684	0.019976	0.046989	0.028675	0.002380	0.025823
$\Phi_{21}$	-0.102697	0.050832	-0.034100	0.049310	-0.158907	0.073014
$\Phi_{22}$	0.063595	0.038714	0.136019	0.044622	0.054652	0.050555

Table 3. Estimation results of VAR(1) with heteroskedasticity consistent standard errors.

	full sample		first half		second half	
	est.	s.e.	est.	s.e.	est.	s.e.
$A_{11}$	0.1870	0.0428	0.0053	0.0386	0.1957	0.1027
$A_{12}$	-0.0266	0.0483	0.1337	0.0293	-0.0725	0.0491
$A_{21}$	0.0137	0.0431	-0.1673	0.0647	0.0288	0.0982
$A_{22}$	0.1812	0.0468	0.3694	0.0459	0.1270	0.0651
$B_{11}$	0.9627	0.0118	1.0071	0.0081	0.9568	0.0165
$B_{12}$	0.0063	0.0105	-0.0404	0.0123	0.0114	0.0079
$B_{21}$	-0.0028	0.0097	0.0460	0.0157	-0.0032	0.0196
$B_{22}$	0.9704	0.0086	0.9083	0.0251	0.9763	0.0116
$\gamma_1$	0.2389	0.0236	0.2019	0.0386	0.2814	0.0271
$\gamma_2$	0.2062	0.0261	0.2088	0.0608	0.2299	0.0244
<i>pers</i>	0.9956		0.9901		0.9967	
<i>port</i>	180.88		372.29		27.79	

Table 4. Results for stationary BEKK. The value *pers* is the maximum eigenvalue of the matrix  $A \otimes A + B \otimes B + C \otimes C/2$ , and *port* is the value of the multivariate portmanteau statistic of order

10.

	full sample		first half		second half	
	est.	s.e.	est.	s.e.	est.	s.e.
$A_{11}$	0.0586	0.0430	0.0301	0.0478	0.1567	0.0632
$A_{12}$	0.0497	0.0613	-0.1312	0.1332	-0.0785	0.0485
$A_{21}$	-0.1935	0.0273	-0.2007	0.0454	-0.1397	0.0901
$A_{22}$	0.0524	0.0544	-0.0918	0.0610	-0.0565	0.0608
$B_{11}$	0.9537	0.0146	0.8481	0.1073	0.9068	0.0484
$B_{12}$	-0.0226	0.0175	0.0522	0.0626	0.0324	0.0702
$B_{21}$	0.0694	0.0180	-0.1582	0.1812	0.1039	0.0503
$B_{22}$	0.8565	0.0330	0.7799	0.1032	0.8229	0.0904
$\gamma_1$	0.2757	0.0333	0.1731	0.0997	0.3137	0.0466
$\gamma_2$	0.3661	0.0378	0.4468	0.0615	0.2383	0.0465
<i>pers</i>	0.9317		0.7602		0.9253	
<i>port</i>	134.01		303.69		19.13	

Table 5. Results for local stationary model. The value *pers* is the maximum eigenvalue of the matrix  $A \otimes A + B \otimes B + C \otimes C/2$ , and *port* is the value of the multivariate portmanteau statistic of order 10.

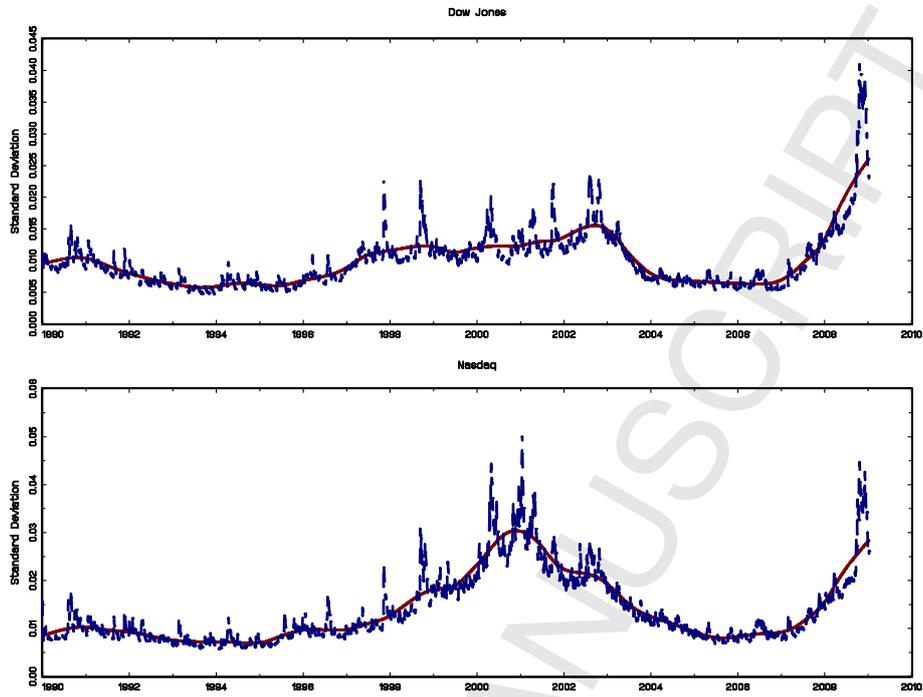


Figure 5. Estimates of conditional and unconditional standard deviations

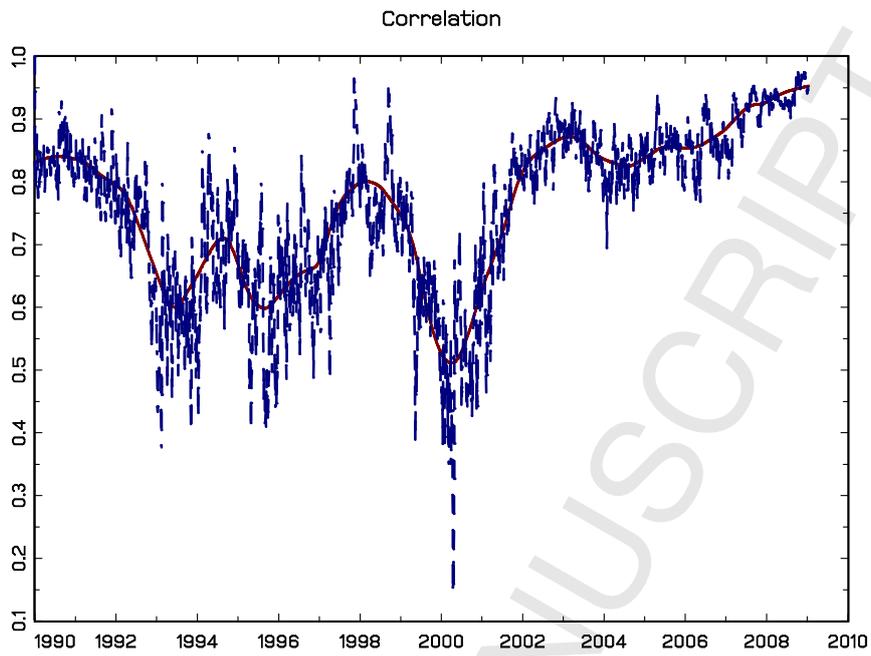


Figure 6. Estimates of conditional and unconditional correlation

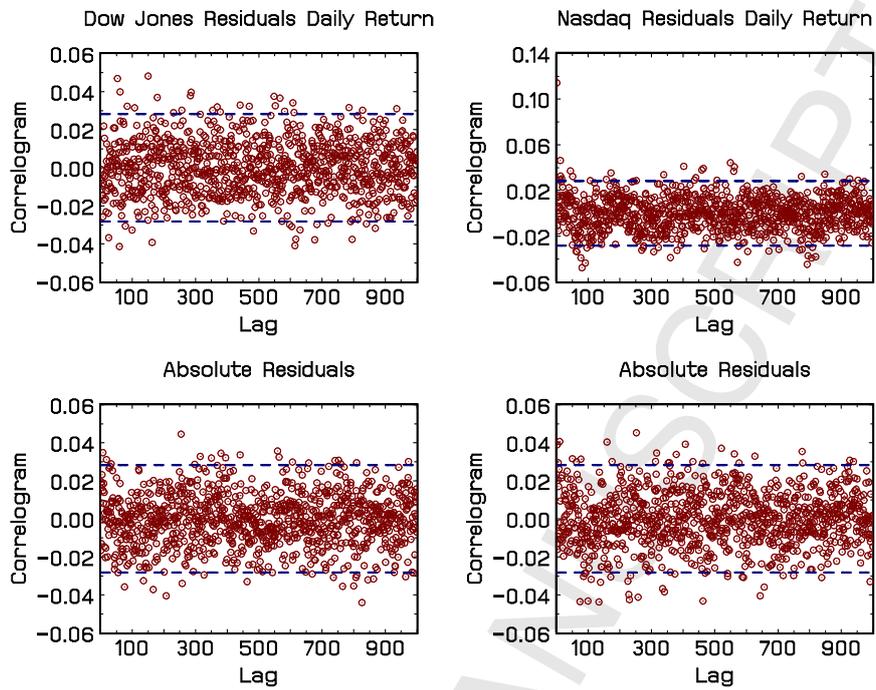


Figure 7. Autocorrelation function of residuals and absolute residuals

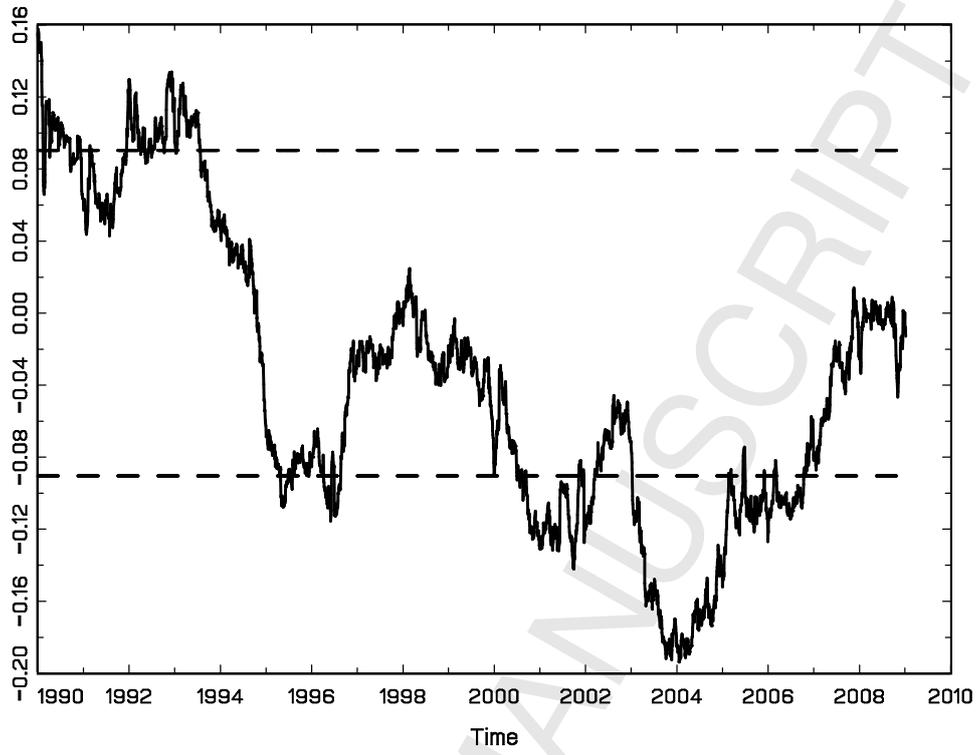


Figure 8. NASDAQ Residual first order local autocorrelation function

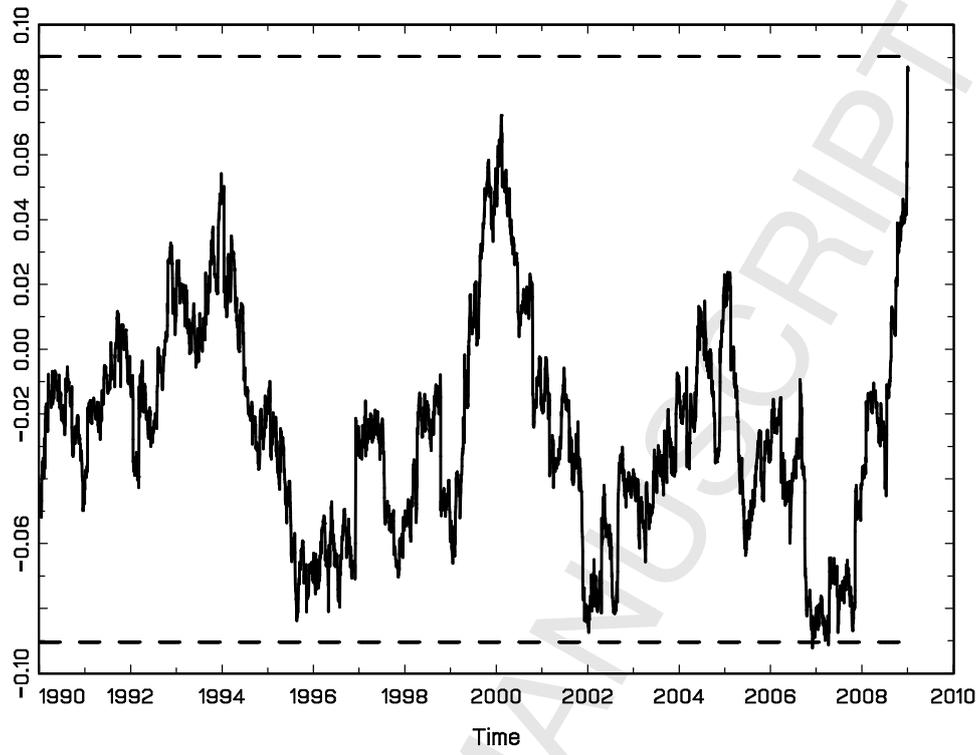


Figure 9. DOW Residual first order local autocorrelation function

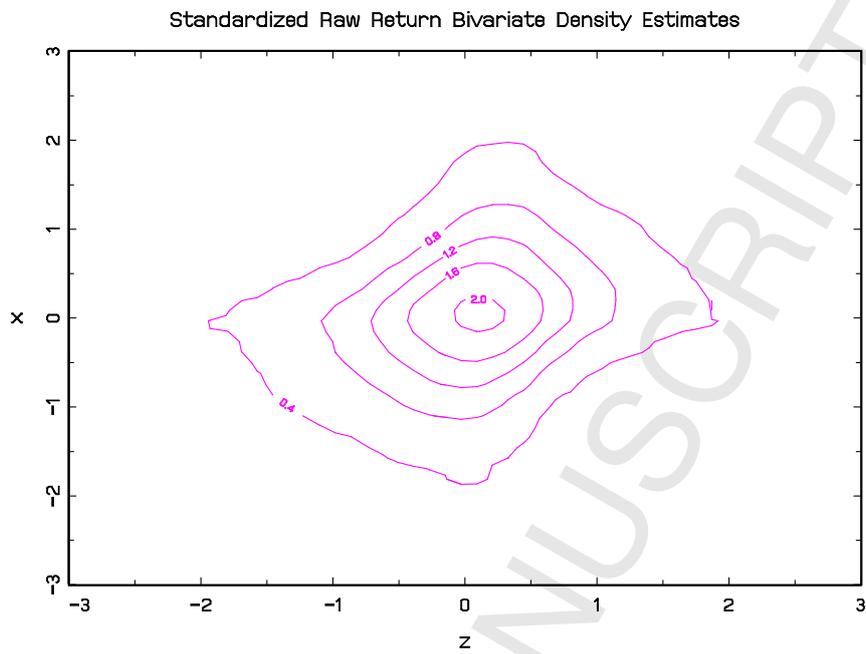


Figure 10. Bivariate density plot of data

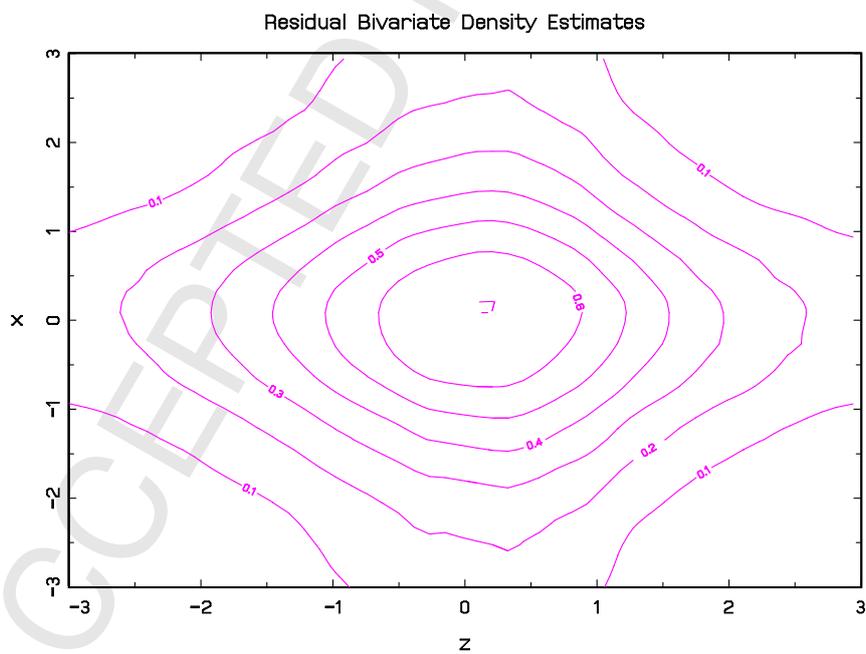


Figure 11. Bivariate density plot of residuals