

## ESTIMATION AND INFERENCE OF THE VECTOR AUTOREGRESSIVE PROCESS UNDER HETEROSCEDASTICITY

UDC 519.21

T. BODNAR AND T. ZABOLOTSKY

**ABSTRACT.** In this paper we derive the asymptotic distribution of the estimator for the parameters of the vector autoregressive process of order  $p$  with an unconditionally heteroscedastic error process. The covariance matrix of the error process is modeled as a deterministic matrix function and it is estimated nonparametrically at each time point. This estimator is used for deriving inference procedures for the parameters of the vector autoregressive process.

### 1. INTRODUCTION

Autoregressive processes are widely spread in the statistical and economical literature. The idea of modeling the conditional mean of the process by using the autoregressive structure is also applied for other purposes. For instance, Engle (1982) [8] suggested a model for conditional volatility, the so-called ARCH (autoregressive conditionally heteroscedastic) process, which implements the design of the autoregressive process for modeling the conditional variance. Moreover, it is the case that the squared values of the ARCH process follow an autoregressive process. Hansen (1994) [14] applied the autoregressive method for modeling the density function by suggesting a process that is based on the  $t$ -distribution with time-varying degrees of freedom. Rockinger and Jondeau (2002) [21] considered the process with time-varying higher moments; i.e., the conditional time-varying skewness and kurtosis were modeled. Finally, Darolles et al. (2006) [5] considered a general class of autoregressive processes. As a partial case, the autoregressive Wishart process was introduced by Gouriéroux et al. (2009) [12].

A common assumption, which is imposed when an autoregressive process is fitted to real data, is that its parameters are time-invariant. This assumption significantly simplifies the estimation of the parameters and the derivation of the distributional properties of the estimator (see, e.g., Brockwell and Davis (1991) [3]). However, it can be unrealistic in a practical situation. Usually, the variance of the error process used in the equation of the autoregressive process is heteroscedastic. This problem has been extensively treated in the case of the linear regression model (see, e.g., Greene (2008) [13]), where several tests for the homoscedasticity versus the heteroscedasticity were derived by White (1980) [24], Godfrey (1978) [11], Breusch and Pagan (1979) [2].

Although the ARCH-type models, which are developed to model the conditionally time-varying variance, can capture a lot of properties of financial and economical data, Xu and Phillips (2008) [26] argued that economical time series, like exchange rates, interest rates, GDP, usually follow a heteroscedastic process. An improvement can be obtained

---

2010 *Mathematics Subject Classification.* Primary 62H12, 62M15; Secondary 62H10.

*Key words and phrases.* Heteroscedasticity, inference procedure, parameter estimation, vector autoregressive process.

by modeling the unconditional heteroscedasticity by complicated ARCH-type processes (see, e.g., Engle and Rangel (2004) [9], Polzehl and Spokoiny (2006) [20]). Alternatively, the time-varying variance is modeled as a nonparametric function by using a smooth deterministic framework (see, e.g., Hsu et al. (1974) [16], Merton (1980) [18], Drees and Stărică (2002) [7], Xu and Phillips (2008) [26]). Drees and Stărică (2002) [7] and Stărică (2003) [22] compared the deterministic framework with the ARCH-type models by using the S&P 500 return series and concluded that the deterministic framework has a better performance in the sense of both fitting data and forecasting the next day volatility.

Although several series of economic variables are, usually, modeled simultaneously, the impact of the heteroscedasticity in the error process has not been treated in the literature up to now. This is a goal of the present paper. Namely, we contribute to the existing literature by extending the findings of Phillips and Xu (2006) [19] to the vector autoregressive process (VAR), a multivariate generalization of the autoregressive process. The obtained results can be widely applied in economic theory. For instance, the cointegration analysis, which is based on the VAR process, is, usually, performed for studying the relationship between the macroeconomic variables and stock market returns (see, e.g., Fama (1981) [10], Kwon and Shin (1999) [17]). Since the univariate time series of the macroeconomic variables are heteroscedastic, it is obvious to assume that the joint model possesses a heteroscedastic covariance matrix.

The rest of the paper is structured as follows. In the next section, some preliminary results are given. Here, we provide the definition of the VAR process with heteroscedastic errors and discuss the imposed conditions. The main results are given in Section 3 where the asymptotic distribution of the estimated parameters is derived. In Section 3.2, the inference procedures are presented.

## 2. PRELIMINARY RESULTS

Let  $\{\mathbf{u}_\tau\}$ ,  $\tau \in (-\infty, T]$ ,  $\tau \in \mathbb{Z}$ , be a  $k$ -dimensional martingale difference process with respect to the increasing filtration  $\mathcal{F}_\tau$ , where  $\mathcal{F}_\tau = \sigma(\mathbf{u}_s, s \leq \tau)$  is the  $\sigma$ -field generated by the events  $\{\mathbf{u}_s, s \leq \tau\}$  with  $\mathbb{E}(\mathbf{u}_\tau \mathbf{u}_\tau' | \mathcal{F}_{\tau-1}) = \mathbf{I}_k$ , a.s., for all  $\tau$ . The symbol  $\mathbf{I}_k$  stands for the  $k$ -dimensional identity matrix. We assume that  $\{\mathbf{Y}_t\}$  follows a  $k$ -dimensional vector autoregressive process of known and finite order  $p$  expressed as

$$(1) \quad \begin{aligned} \mathbf{Y}_t &= \boldsymbol{\theta}_0 + \boldsymbol{\theta}_1 \mathbf{Y}_{t-1} + \cdots + \boldsymbol{\theta}_p \mathbf{Y}_{t-p} + \boldsymbol{\varepsilon}_t, \\ \boldsymbol{\varepsilon}_t &= \boldsymbol{\Sigma}_t \mathbf{u}_t \end{aligned}$$

with  $\boldsymbol{\Sigma}_t = \mathbf{G}(t/T)$  for  $t \in [0, T]$ ,  $t, T \in \mathbb{N}$ . The  $k$ -dimensional vector  $\boldsymbol{\theta}_0$  and the  $k \times k$  matrices  $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p$  determine the parameters of the process. We also assume that  $\mathbf{G}(\cdot)$  is a deterministic matrix function such that  $\text{rank}(\mathbf{G}(t/T)) = k$  for all  $t$ , where  $\text{rank}(\mathbf{A})$  denotes the rank of the matrix  $\mathbf{A}$ . This assumption ensures that the covariance matrix  $\text{Var}(\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1}) = \boldsymbol{\Sigma}_t \boldsymbol{\Sigma}_t'$  is a positive definite deterministic matrix function for all  $t$ . Because the infinite moving average representation of the process  $\{\mathbf{Y}_t\}$  is used in the derivation of the results we assume that both the process  $\{\boldsymbol{\varepsilon}_t\}$  and the matrix function  $\mathbf{G}(\cdot)$  are also defined for negative values of  $t$ .

Throughout the paper we assume that the following four assumptions hold:

**Assumption 1.** All eigenvalues of the matrix  $\sum_{i=1}^{p_1} \boldsymbol{\theta}_i$  have modulus smaller than one.

When  $\{\mathbf{u}_t\}$  is a weakly stationary process and  $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}$ , i.e. the matrix  $\boldsymbol{\Sigma}_t$  is time-invariant, Assumption 1 ensures that the process  $\{\mathbf{Y}_t\}$  is also weakly stationary with the mean vector  $\mathbb{E}(\mathbf{Y}_t) = (\mathbf{I} - \sum_{i=1}^p \boldsymbol{\theta}_i)^{-1} \boldsymbol{\theta}_0$  and the cross-covariance matrix  $\boldsymbol{\Gamma}(h)$  of lag  $h$ ,

which is given by

$$(2) \quad \mathbf{\Gamma}(h) = \sum_{i=0}^{+\infty} \mathbf{A}_{i+h} \mathbf{\Sigma} \mathbf{A}'_i,$$

where the matrices  $\mathbf{A}_i$  are uniquely defined by the matrices  $\boldsymbol{\theta}_i$ , i.e.

$$(3) \quad \mathbf{A}_0 = \mathbf{I}_k \quad \text{and} \quad \mathbf{A}_j = \sum_{i=1}^j \boldsymbol{\theta}_i \mathbf{A}_{j-i}, \quad j = 1, 2, \dots,$$

with  $\boldsymbol{\theta}_i = \mathbf{0}$  for  $i > p$ .

**Assumption 2.** The matrix function  $\mathbf{G}(\cdot)$  is nonstochastic, measurable and uniformly bounded on the interval  $(-\infty; 1]$  with a finite numbers of points of discontinuity and it satisfies a Lipschitz condition except at points of discontinuity.

This assumption is similar to Assumption (ii) of Phillips and Xu (2006) [19]. It implies that each component of the matrix  $\mathbf{G}(\cdot)$  is integrable on  $[0, 1]$  up to any finite order, i.e.  $\int_0^1 |g_{ij}(r)|^m dr < \infty$  for all  $m > 0$ , where  $\mathbf{G}(r) = (g_{ij}(r))_{i,j=1,\dots,k}$ . Note, that the matrix function  $\mathbf{G}(\cdot)$  is defined for  $r \in (-\infty, 1]$  since the initial conditions are in the infinite past and we make use of the infinite moving average representation of the process  $\{\mathbf{Y}_t\}$ .

In the following we use the notions of strong mixing ( $\alpha$ -mixing) and of near-epoch dependence which are defined by

**Definition 1.** A sequence of random variables  $\{\xi_t(w)\}_{-\infty}^{+\infty}$  with

$$\mathcal{F}_{-\infty}^t = \sigma(\dots, \xi_{t-2}, \xi_{t-1}, \xi_t)$$

and

$$\mathcal{F}_{t+m}^{+\infty} = \sigma(\xi_{t+m}, \xi_{t+m+1}, \xi_{t+m+2}, \dots)$$

is strong mixing (or  $\alpha$ -mixing) if  $\lim_{m \rightarrow +\infty} \alpha_m = 0$ , where  $\alpha_m = \sup_t \alpha(\mathcal{F}_{-\infty}^t, \mathcal{F}_{t+m}^{+\infty})$  with

$$\alpha(\mathcal{F}_{-\infty}^t, \mathcal{F}_{t+m}^{+\infty}) = \sup_{F_1 \in \mathcal{F}_{-\infty}^t, F_2 \in \mathcal{F}_{t+m}^{+\infty}} |\mathbf{P}(F_1 \cap F_2) - \mathbf{P}(F_1)\mathbf{P}(F_2)|.$$

**Definition 2.** For a sequence  $\{\xi_t(w)\}_{-\infty}^{+\infty}$ , let  $\mathcal{F}_{t-m}^{t+m} = \sigma(\xi_{t-m}, \dots, \xi_{t+m})$ , such that  $\{\mathcal{F}_{t-m}^{t+m}\}_{m=0}^{+\infty}$  is an increasing sequence of  $\sigma$ -fields. If, for  $p > 0$ , a sequence of integrable random variables  $\{\zeta_t(w)\}_{-\infty}^{+\infty}$  satisfies

$$\|\zeta_t - \mathbf{E}(\zeta_t | \mathcal{F}_{t-m}^{t+m})\|_p \leq d_t \nu_m,$$

where  $\nu_m \rightarrow 0$  and  $\{d_t\}_{-\infty}^{+\infty}$  is a sequence of positive constants, then  $\zeta_t$  is near-epoch dependent in the  $L_p$ -norm ( $L_p$ -NED) on  $\{\xi_t(w)\}_{-\infty}^{+\infty}$ .

**Assumption 3.**  $\{\mathbf{u}_t\}$  is a strong mixing ( $\alpha$ -mixing) martingale difference process with  $\mathbf{E}(\mathbf{u}_t | \mathcal{F}_{t-1}) = \mathbf{0}$  and  $\mathbf{E}(\mathbf{u}_t \mathbf{u}'_t | \mathcal{F}_{t-1}) = \mathbf{I}_k$ , a.s., for all  $t$ , with the natural filtration  $\mathcal{F}_t = \sigma(\mathbf{u}_s, s \leq t)$ . There exist  $\delta > 1$  and  $C > 0$ , such that  $\sup_t \mathbf{E} \|\mathbf{u}_t\|^{4\delta} < C < \infty$ .

This assumption implies that  $\mathbf{E}(\mathbf{u}_t \mathbf{u}'_s) = \mathbf{0}_k$  for  $s \neq t$ , where  $\mathbf{0}_k$  is a  $k \times k$  zero matrix. We also have that  $\mathbf{E}(|u_{t,i}|^{4\delta}) < C$  with  $\mathbf{u}_t = (u_{t,1}, \dots, u_{t,k})'$  and, consequently, by the Lyapunov's inequality it follows that  $\mathbf{E}(|u_{t,i}|^\eta) < C$  for all  $\eta \leq 4\delta$ . The same results are true for any expression of  $u_{t,i}$  with the overall power less than or equal to  $4\delta$ . From  $\text{Var}(\mathbf{u}_t | \mathcal{F}_{t-1}) = \mathbf{I}_k$  we obtain  $\text{Var}(\boldsymbol{\varepsilon}_t) = \text{Var}(\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1}) = \mathbf{\Sigma}_t \mathbf{\Sigma}'_t$ . The last identity specifies the unconditional heteroscedasticity in the process  $\{\mathbf{Y}_t\}$  modeled in (1). In the univariate case, Cavaliere (2004) [4] discussed the allowable variance patterns in detail.

For the rest of the paper it is assumed that a multivariate sample of  $T+p$  observations is available given by  $\{\mathbf{Y}_{-p+1}, \mathbf{Y}_{-p+2}, \dots, \mathbf{Y}_{-1}, \mathbf{Y}_0, \mathbf{Y}_1, \dots, \mathbf{Y}_T\}$ . Let

$$\mathbf{X}_{t-1} = (\mathbf{1}', \mathbf{Y}'_{t-1}, \dots, \mathbf{Y}'_{t-p})'.$$

The parameters of the process (1) we denote by  $\boldsymbol{\theta} = (\boldsymbol{\theta}_0, \dots, \boldsymbol{\theta}_p)'$ . The expression (1) is rewritten as

$$(4) \quad \mathbf{Y}_t = \boldsymbol{\theta}' \mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t.$$

Under Assumption 1,  $\mathbf{Y}_t$  has the infinite moving average representation, the so-called Wold representation, given by

$$(5) \quad \mathbf{Y}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \mathbf{A}_i \boldsymbol{\varepsilon}_{t-i},$$

where the matrices  $\mathbf{A}_i = (A_{i,jl})_{j,l=1,\dots,k}$  are defined in (3).

**Assumption 4.** Let  $\sum_{i=0}^{\infty} \|\mathbf{A}_i\|^2 < \infty$ , where  $\|\mathbf{A}_i\|^2 = \sum_{j,l=1}^k A_{i,jl}^2$ .

The estimator of the parameter matrix  $\boldsymbol{\theta}$  is, usually, derived by considering (4) as a multivariate regression given by

$$\begin{pmatrix} \mathbf{Y}'_1 \\ \mathbf{Y}'_2 \\ \vdots \\ \mathbf{Y}'_T \end{pmatrix} = \begin{pmatrix} \mathbf{X}'_0 \\ \mathbf{X}'_1 \\ \vdots \\ \mathbf{X}'_{T-1} \end{pmatrix} \boldsymbol{\theta} + \begin{pmatrix} \boldsymbol{\varepsilon}'_1 \\ \boldsymbol{\varepsilon}'_2 \\ \vdots \\ \boldsymbol{\varepsilon}'_T \end{pmatrix}$$

and estimating  $\boldsymbol{\theta}$  of this regression by the OLS (ordinary least square) estimator. This leads to

$$(6) \quad \hat{\boldsymbol{\theta}} = \left( \sum_{t=1}^T \mathbf{X}_{t-1} \mathbf{X}'_{t-1} \right)^{-1} \sum_{t=1}^T \mathbf{X}_{t-1} \mathbf{Y}'_t.$$

### 3. MAIN RESULTS

In Section 3.1, we derive the asymptotic distribution of  $\hat{\boldsymbol{\theta}}$ , while the inference procedures are given in Section 3.2.

**3.1. Asymptotic distribution of  $\hat{\boldsymbol{\theta}}$ .** Let  $\|\xi\|_k = (\mathbb{E} \|\xi\|^k)^{1/k}$ . In Theorem 1, we present the asymptotic distribution of  $\hat{\boldsymbol{\theta}}$  and show that  $\hat{\boldsymbol{\theta}}$  is consistent. For the proof of Theorem 1 we need the following two lemmas.

**Lemma 1.** *Under Assumptions 1–4, we obtain*

- (a)  $\sup_{1 \leq t \leq T} \mathbb{E} \|\mathbf{Y}_t - \boldsymbol{\mu}\|^{4\delta} < \infty$ , for  $\delta > 1$ .
- (b) Let  $1 \leq h \leq p$ ,  $0 \leq j \leq p-h$ . Then

$$\lim_{T \rightarrow \infty} \mathbb{E} (\mathbf{Y}_{[Tr]-h} - \boldsymbol{\mu}) (\mathbf{Y}_{[Tr]-h-j} - \boldsymbol{\mu})' = \sum_{i=0}^{\infty} \mathbf{A}_{i+j} \mathbf{G}(r) \mathbf{G}'(r) \mathbf{A}'_i,$$

for values  $r \in (0, 1]$  at which the function  $\mathbf{G}(\cdot)$  is continuous.

The proof of this lemma is given in the Appendix.

By  $\mathbf{1}_k = (1, 1, \dots, 1)'$  we denote the  $k$ -dimensional vector of ones. Let

$$(7) \quad \boldsymbol{\Omega}_1 = \begin{pmatrix} \mathbf{1}\mathbf{1}' & \mathbf{1}\boldsymbol{\mu}' & \dots & \mathbf{1}\boldsymbol{\mu}' \\ \boldsymbol{\mu}\mathbf{1}' & \boldsymbol{\mu}\boldsymbol{\mu}' + \tilde{\boldsymbol{\Omega}}^0 & \dots & \boldsymbol{\mu}\boldsymbol{\mu}' + \tilde{\boldsymbol{\Omega}}^{(p-1)} \\ \dots & \dots & \dots & \dots \\ \boldsymbol{\mu}\mathbf{1}' & \boldsymbol{\mu}\boldsymbol{\mu}' + \tilde{\boldsymbol{\Omega}}^{(p-1)'} & \dots & \boldsymbol{\mu}\boldsymbol{\mu}' + \tilde{\boldsymbol{\Omega}}^0 \end{pmatrix}$$

with  $\tilde{\Omega}^j = \sum_{i=0}^{\infty} \mathbf{A}_{i+j} (\int_0^1 \mathbf{G}(r) \mathbf{G}'(r) dr) \mathbf{A}'_i$ . Let  $\mathbf{K}(i, j, t/T) = \sum_{l=1}^k g_{il}(t/T) g_{jl}(t/T)$  and  $\Omega_2 = \{\Omega^{i,j}\}_{i,j=1,\dots,k}$  with

$$(8) \quad \Omega^{i,j} = \begin{pmatrix} \nu_{i,j} \mathbf{1} \mathbf{1}' & \nu_{i,j} \mathbf{1} \boldsymbol{\mu}' & \dots & \nu_{i,j} \mathbf{1} \boldsymbol{\mu}' \\ \nu_{i,j} \boldsymbol{\mu} \mathbf{1}' & \nu_{i,j} \boldsymbol{\mu} \boldsymbol{\mu}' + \Omega_{ij}^0 & \dots & \nu_{i,j} \boldsymbol{\mu} \boldsymbol{\mu}' + \Omega_{ij}^{(p-1)} \\ \dots & \nu_{i,j} \boldsymbol{\mu} \boldsymbol{\mu}' + \Omega_{ij}^1 & \dots & \nu_{i,j} \boldsymbol{\mu} \boldsymbol{\mu}' + \Omega_{ij}^{(p-2)} \\ \dots & \dots & \dots & \dots \\ \boldsymbol{\mu} \mathbf{1}' & \nu_{i,j} \boldsymbol{\mu} \boldsymbol{\mu}' + \Omega_{ij}^{(p-1)'} & \dots & \nu_{i,j} \boldsymbol{\mu} \boldsymbol{\mu}' + \Omega_{ij}^0 \end{pmatrix},$$

where  $\nu_{i,j} = \int_0^1 \mathbf{K}(i, j, r) dr$  and  $\Omega_{ij}^l = \sum_{m=0}^{\infty} \mathbf{A}_{m+l} (\int_0^1 \mathbf{K}(i, j, r) \mathbf{G}(r) \mathbf{G}'(r) dr) \mathbf{A}'_m$ .

**Lemma 2.** Under Assumptions 1–4 as  $T \rightarrow \infty$ ,

- (a)  $T^{-1} \sum_{t=1}^T \mathbf{X}_{t-1} \boldsymbol{\varepsilon}'_t \xrightarrow{P} \mathbf{0}$ ;
- (b)  $T^{-1} \sum_{t=1}^T \mathbf{X}_{t-1} \mathbf{X}'_{t-1} \xrightarrow{P} \Omega_1$ ;
- (c)  $T^{-1} \sum_{t=1}^T \mathbf{vec}(\mathbf{X}_{t-1} \boldsymbol{\varepsilon}'_t) \mathbf{vec}(\mathbf{X}_{t-1} \boldsymbol{\varepsilon}'_t)' \xrightarrow{P} \Omega_2$ ;
- (d)  $T^{-1/2} \sum_{t=1}^T \mathbf{vec}(\mathbf{X}_{t-1} \boldsymbol{\varepsilon}'_t) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Omega_2)$ .

The proof of this lemma is given in the Appendix.

Applying the OLS estimator (6) of  $\boldsymbol{\theta}$  we obtain the scaled error expressed as

$$(9) \quad \sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \left( \frac{1}{T} \sum_{t=1}^T \mathbf{X}_{t-1} \mathbf{X}'_{t-1} \right)^{-1} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{X}_{t-1} \boldsymbol{\varepsilon}'_t \right).$$

**Theorem 1.** Under Assumptions 1–4,  $\hat{\boldsymbol{\theta}}$  is consistent and asymptotically normally distributed and

$$\mathbf{vec} \left( \sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}),$$

with  $T \rightarrow \infty$  and  $\boldsymbol{\Lambda} = \tilde{\Omega}_1^{-1} \Omega_2 \tilde{\Omega}_1^{-1}$ , where  $\tilde{\Omega}_1 = \mathbf{I}_p \otimes \Omega_1$  and the symbol  $\otimes$  denotes the Kronecker product.

*Proof.* The results of Theorem 1 follow from Lemma 2.  $\square$

**3.2. Inference procedure.** The results of Theorem 1 can be applied to construct tests for some hypotheses on  $\boldsymbol{\theta}$  by using a consistent estimator of  $\boldsymbol{\Lambda}$ . If the matrix function  $\mathbf{G}(\cdot)$  is known, the consistent estimator of  $\boldsymbol{\Lambda}$  is obtained by replacing  $\boldsymbol{\mu}$  by

$$\hat{\boldsymbol{\mu}} = \left( \mathbf{I} - \sum_{i=1}^p \hat{\boldsymbol{\theta}}_i \right)^{-1} \hat{\boldsymbol{\theta}}_0,$$

which is consistent since  $\hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\theta}}_1, \dots, \hat{\boldsymbol{\theta}}_p$  are consistent.

However, the function  $\mathbf{G}(\cdot)$  is, usually, unknown and has to be estimated. In the univariate case, Phillips and Xu (2006) [19] suggested three possible consistent estimators of  $\boldsymbol{\Lambda}$  when  $\mathbf{G}(\cdot)$  is an unknown deterministic function. The first two approaches are based on the Eicker–White correction for heteroscedasticity and on the modification of the Eicker–White correction. In the third approach, an application of the nonparametric estimator of  $\mathbf{G}(\cdot)$  is suggested.

Next, we generalize this approach to the multivariate case. We estimate the matrix function  $\mathbf{G}(\cdot) \mathbf{G}'(\cdot)$  nonparametrically by a weighted sum of squared OLS residuals using kernel smoothing. Let

$$\hat{\mathbf{G}}(r) \hat{\mathbf{G}}'(r) = \sum_{t=1}^T \omega_{rt} \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}'_t,$$

where  $\hat{\boldsymbol{\varepsilon}}_t = \mathbf{Y}_t - \hat{\boldsymbol{\theta}}\mathbf{X}'_{t-1}$  are the OLS residuals,  $r \in [0, 1]$  and the weights  $\omega_{rt}$  ( $t = 1, \dots, T$ ) are given by

$$\omega_{rt} = \frac{K((\lceil Tr \rceil - t)/(Tb))}{\sum_{t=1}^T K((\lceil Tr \rceil - t)/(Tb))},$$

where the kernel function  $K: \mathbb{R} \rightarrow [0, \infty)$  satisfies the following two inequalities:

$$0 \leq K(x) < C_1 < \infty \quad \text{and} \quad \int_{-\infty}^{\infty} K(x) dx < \infty$$

with  $C_1$  being a positive constant. The bandwidth parameter  $b$  depends on  $T$ , such that  $b + (1/Tb) \rightarrow 0$ , as  $T \rightarrow \infty$ . In the univariate case, Wong (1983) [25] proposed to use cross-validation on the average squared error for choosing the bandwidth parameter  $b$ .

Let  $\mathbf{W} = \{\mathbf{W}(i, j)\}_{i,j=1,\dots,k}$ , where  $\mathbf{W}(i, j) = \mathbf{E}(u_{t,i}u_{t,j}(\mathbf{u}_t\mathbf{u}'_t))$ .

**Lemma 3.** *Under Assumptions 1–4, as  $T \rightarrow \infty$ ,*

- (a)  $T^{-1} \sum_{t=1}^T \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}'_t \xrightarrow{P} \int_0^1 \mathbf{G}(r)\mathbf{G}'(r) dr$ ;
- (b)  $T^{-1} \sum_{t=1}^T \mathbf{vec}(\mathbf{vec}(\hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}'_t) \mathbf{vec}(\hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}'_t)') \xrightarrow{P} \int_0^1 (\mathbf{G}(r) \otimes \mathbf{G}(r) \otimes \mathbf{G}(r) \otimes \mathbf{G}(r)) dr \mathbf{vec} \mathbf{W}$ ;
- (c)  $T^{-1} \sum_{t=1}^T \mathbf{vec}(\mathbf{X}_{t-1} \hat{\boldsymbol{\varepsilon}}'_t) \mathbf{vec}(\mathbf{X}_{t-1} \hat{\boldsymbol{\varepsilon}}'_t)' \xrightarrow{P} \boldsymbol{\Omega}_2$ ;
- (d)  $\hat{\mathbf{G}}(r)\hat{\mathbf{G}}'(r) \xrightarrow{P} \mathbf{G}(r)\mathbf{G}'(r)$ , for all  $r \in [0, 1]$  for which the function  $\mathbf{G}$  is continuous.

The proof of this lemma is given in the Appendix. The results of Lemma 3 imply three consistent estimators of  $\boldsymbol{\Lambda}$ . Let  $\mathbf{Q} = \sum_{t=1}^T \mathbf{X}_{t-1}\mathbf{X}'_{t-1}$ . The first one is given by

$$\hat{\boldsymbol{\Lambda}}_1 = T(\mathbf{I}_p \otimes \mathbf{Q})^{-1} \left( \sum_{t=1}^T \mathbf{vec}(\mathbf{X}_{t-1} \hat{\boldsymbol{\varepsilon}}'_t) (\mathbf{vec}(\mathbf{X}_{t-1} \hat{\boldsymbol{\varepsilon}}'_t))' \right) (\mathbf{I}_p \otimes \mathbf{Q})^{-1},$$

which is based on the Eicker–White correction for heteroscedasticity. The second estimator is

$$\hat{\boldsymbol{\Lambda}}_2 = \hat{\boldsymbol{\Omega}}_1^{-1} \left( \frac{1}{T} \sum_{t=1}^T \mathbf{vec}(\mathbf{X}_{t-1} \hat{\boldsymbol{\varepsilon}}'_t) (\mathbf{vec}(\mathbf{X}_{t-1} \hat{\boldsymbol{\varepsilon}}'_t))' \right) \hat{\boldsymbol{\Omega}}_1^{-1},$$

where  $\hat{\boldsymbol{\Omega}}_1 = \mathbf{I}_p \otimes \hat{\boldsymbol{\Omega}}_1$  with  $\hat{\boldsymbol{\Omega}}_1$  obtained from (7) by substituting

$$\hat{\boldsymbol{\Omega}}^j = \sum_{i=0}^{\infty} \hat{\mathbf{A}}_{i+j} \left( \frac{1}{T} \sum_{t=1}^T \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}'_t \right) \hat{\mathbf{A}}'_i$$

and  $\hat{\boldsymbol{\mu}}$  instead of  $\tilde{\boldsymbol{\Omega}}^j$  and  $\boldsymbol{\mu}$ .

The third estimator is given by

$$\hat{\boldsymbol{\Lambda}}_3 = \hat{\boldsymbol{\Omega}}_1^{-1} \hat{\boldsymbol{\Omega}}_2 \hat{\boldsymbol{\Omega}}_1^{-1},$$

where  $\hat{\boldsymbol{\Omega}}_2$  is obtained by substituting

$$\hat{K} \left( i, j, \frac{t}{T} \right) = \sum_{l=1}^k \hat{g}_{il} \left( \frac{t}{T} \right) \hat{g}_{jl} \left( \frac{t}{T} \right),$$

$\hat{\nu}_{i,j} = \int_0^1 \hat{K}(i, j, r) dr$ , and  $\hat{\boldsymbol{\Omega}}_{ij}^l = \sum_{m=0}^{\infty} \hat{\mathbf{A}}_{m+l} (\int_0^1 \hat{K}(i, j, r) \hat{\mathbf{G}}(r)\hat{\mathbf{G}}'(r) dr) \hat{\mathbf{A}}'_m$  in (8).

**Theorem 2.** *Under Assumptions 1–4, as  $T \rightarrow \infty$  it follows that:*

- (a)  $\hat{\boldsymbol{\Lambda}}_j \xrightarrow{P} \boldsymbol{\Lambda}$ , for  $j = 1, 2, 3$ .

(b) Under  $H_0: \boldsymbol{\theta}_0 = \tilde{\boldsymbol{\theta}}_0$  ( $H_0: \boldsymbol{\theta}_l = \tilde{\boldsymbol{\theta}}_l$  for  $l = 1, \dots, p$ ),

$$T_{0;j} = \sqrt{T}(\hat{\boldsymbol{\theta}}_0 - \tilde{\boldsymbol{\theta}}_0)'(\hat{\boldsymbol{\Lambda}}_j)_{00}^{-1}(\hat{\boldsymbol{\theta}}_0 - \tilde{\boldsymbol{\theta}}_0) \xrightarrow{d} \chi_k^2$$

$$\left( T_{l;j} = \sqrt{T}(\mathbf{vec}(\hat{\boldsymbol{\theta}}_l) - \mathbf{vec}(\tilde{\boldsymbol{\theta}}_l))'(\hat{\boldsymbol{\Lambda}}_j)_{ll}^{-1}(\mathbf{vec}(\hat{\boldsymbol{\theta}}_l) - \mathbf{vec}(\tilde{\boldsymbol{\theta}}_l)) \xrightarrow{d} \chi_{k^2}^2 \right),$$

where  $(\hat{\boldsymbol{\Lambda}}_j)_{ll}$  is the  $(l, l)$ -th block of  $\hat{\boldsymbol{\Lambda}}_j$  for  $l = 0, 1, \dots, p$ .

(c) Under  $H_0: \mathbf{a}(\boldsymbol{\theta}) = \mathbf{0}$ , where  $\mathbf{a}(\boldsymbol{\theta})$  is an  $s$ -vector of continuously differentiable restrictions for which the  $s \times (p+1)k$  matrix  $\mathbf{A}(\boldsymbol{\theta})$  of first derivatives is of full row rank,

$$W_j = T\mathbf{a}(\hat{\boldsymbol{\theta}})'(\mathbf{A}(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{\Lambda}}_j\mathbf{A}(\hat{\boldsymbol{\theta}})')^{-1}\mathbf{a}(\hat{\boldsymbol{\theta}}) \xrightarrow{d} \chi_s^2.$$

*Proof.* The results follow from Theorem 1 and Lemma 3.  $\square$

A test for general linear restrictions is obtained from Theorem 2 by setting

$$\mathbf{a}(\boldsymbol{\theta}) = \mathbf{R} \mathbf{vec}(\boldsymbol{\theta}),$$

where  $\mathbf{R}$  is a  $q \times (pk^2 + k)$  matrix of constants and  $q$  defines the number of linear restrictions to be tested. The testing problem is given by

$$H_0: \mathbf{R} \mathbf{vec}(\boldsymbol{\theta}) = \mathbf{0} \quad \text{against} \quad H_1: \mathbf{R} \mathbf{vec}(\boldsymbol{\theta}) \neq \mathbf{0},$$

with the test statistic expressed as

$$L_j = (\mathbf{vec}(\boldsymbol{\theta}))' \mathbf{R}' \hat{\boldsymbol{\Lambda}}_j^{-1} \mathbf{R} \mathbf{vec}(\boldsymbol{\theta}),$$

which is asymptotically  $\chi_q^2$ -distributed under  $H_0$ .

#### 4. APPENDIX

In this section we present the proofs of Lemmas 1–3.

*Proof of Lemma 1.* (a) The following holds:

$$\mathbb{E} \left( \|\mathbf{Y}_t - \boldsymbol{\mu}\|^{4\delta} \right) = \|\mathbf{Y}_t - \boldsymbol{\mu}\|_{4\delta}^{4\delta} = \left\| \sum_{i=0}^{\infty} \mathbf{A}_i \boldsymbol{\varepsilon}_{t-i} \right\|_{4\delta}^{4\delta}.$$

Let  $\mathbf{A}_{i,j}$  be the  $j$ -th column of the matrix  $\mathbf{A}_i$ . The application of the Minkowski inequality leads to

$$\mathbb{E} \left( \|\mathbf{Y}_t - \boldsymbol{\mu}\|^{4\delta} \right) = \left\| \sum_{i=0}^{\infty} \sum_{j=1}^k \mathbf{A}_{i,j} \boldsymbol{\varepsilon}_{t-i,j} \right\|_{4\delta}^{4\delta} \leq \left( \sum_{i=0}^{\infty} \sum_{j=1}^k \|\mathbf{A}_{i,j} \boldsymbol{\varepsilon}_{t-i,j}\|_{4\delta} \right)^{4\delta},$$

where  $\boldsymbol{\varepsilon}_t = (\varepsilon_{t,1}, \dots, \varepsilon_{t,k})'$ . Using the fact that  $\|\mathbf{A}_{i,j}\| \leq \|\mathbf{A}_i\|$  we get

$$\begin{aligned} \mathbb{E} \left( \|\mathbf{Y}_t - \boldsymbol{\mu}\|^{4\delta} \right) &\leq \left( \sum_{i=0}^{\infty} \|\mathbf{A}_i\| \sum_{j=1}^k \|\boldsymbol{\varepsilon}_{t-i,j}\|_{4\delta} \right)^{4\delta} \\ &= \left( \sum_{i=0}^{\infty} \|\mathbf{A}_i\| \sum_{j=1}^k \left\| \mathbf{G}_j \left( \frac{t-i}{T} \right) \mathbf{u}_{t-i} \right\|_{4\delta} \right)^{4\delta}, \end{aligned}$$

where  $\mathbf{G}_j((t-i)/T)$  is the  $j$ -th row of the matrix  $\mathbf{G}((t-i)/T)$ . Finally, from Assumption 3 we get

$$\begin{aligned} \mathbb{E} \left( \|\mathbf{Y}_t - \boldsymbol{\mu}\|^{4\delta} \right) &\leq \left( \sum_{i=0}^{\infty} \|\mathbf{A}_i\| \left\| \mathbf{G} \left( \frac{t-i}{T} \right) \right\| \sum_{j=1}^k \|\mathbf{u}_{t-i}\|_{4\delta} \right)^{4\delta} \\ &\leq Ck \left( \sum_{i=0}^{\infty} \|\mathbf{A}_i\| \left\| \mathbf{G} \left( \frac{t-i}{T} \right) \right\| \right)^{4\delta} < \infty. \end{aligned}$$

(b) Let  $t = [Tr]$ . Then

$$\begin{aligned} \mathbb{E}((\mathbf{Y}_{t-h} - \boldsymbol{\mu})(\mathbf{Y}_{t-h-j} - \boldsymbol{\mu})') &= \mathbb{E} \left( \sum_{i=0}^{\infty} \mathbf{A}_i \boldsymbol{\varepsilon}_{t-h-i} \sum_{i=0}^{\infty} \boldsymbol{\varepsilon}'_{t-h-j-i} \mathbf{A}_i' \right) \\ &= \sum_{i=0}^{\infty} \mathbf{A}_{i+j} \mathbb{E}(\boldsymbol{\varepsilon}_{t-h-j-i} \boldsymbol{\varepsilon}'_{t-h-j-i}) \mathbf{A}_i' \\ &= \sum_{i=0}^{\infty} \mathbf{A}_{i+j} \mathbf{G} \left( \frac{t-h-j-i}{T} \right) \mathbf{G}' \left( \frac{t-h-j-i}{T} \right) \mathbf{A}_i'. \end{aligned}$$

Let  $L_T > 0$  such that

$$\frac{L_T}{T} + \frac{1}{L_T} \rightarrow 0.$$

We obtain

$$\begin{aligned} &\mathbb{E}((\mathbf{Y}_{t-h} - \boldsymbol{\mu})(\mathbf{Y}_{t-h-j} - \boldsymbol{\mu})') \\ &= \sum_{i=0}^{L_T} \mathbf{A}_{i+j} \mathbf{G} \left( \frac{t-h-j-i}{T} \right) \mathbf{G}' \left( \frac{t-h-j-i}{T} \right) \mathbf{A}_i' \\ &\quad + \sum_{i=L_T+1}^{\infty} \mathbf{A}_{i+j} \mathbf{G} \left( \frac{t-h-j-i}{T} \right) \mathbf{G}' \left( \frac{t-h-j-i}{T} \right) \mathbf{A}_i'. \end{aligned}$$

Since  $Tr - 1 \leq [Tr] \leq Tr$  and  $(Tr - 1 - h - j - i)/T \rightarrow r$  as well as

$$\frac{Tr - h - j - i}{T} \rightarrow r,$$

we get that

$$\frac{[Tr] - h - j - i}{T} \rightarrow r.$$

This limit implies that

$$\sum_{i=0}^{L_T} \mathbf{A}_{i+j} \mathbf{G} \left( \frac{t-h-j-i}{T} \right) \mathbf{G}' \left( \frac{t-h-j-i}{T} \right) \mathbf{A}_i' \rightarrow \sum_{i=0}^{\infty} \mathbf{A}_{i+j} \mathbf{G}(r) \mathbf{G}'(r) \mathbf{A}_i',$$

as  $T \rightarrow \infty$ .



Using the fact that the function  $\mathbf{G}(\cdot)$  is uniformly bounded on  $(-\infty, 1]$  we get

$$\begin{aligned} & \left\| \sum_{i=L_T+1}^{\infty} \mathbf{A}_{i+j} \mathbf{G} \left( \frac{t-h-j-i}{T} \right) \mathbf{G}' \left( \frac{t-h-j-i}{T} \right) \mathbf{A}'_i \right\| \\ & \leq \sum_{i=L_T+1}^{\infty} \|\mathbf{A}_{i+j}\| \left\| \mathbf{G} \left( \frac{t-h-j-i}{T} \right) \right\| \left\| \mathbf{G}' \left( \frac{t-h-j-i}{T} \right) \right\| \|\mathbf{A}'_i\| \\ & \leq \|C\|^2 \sum_{i=L_T+1}^{\infty} \|\mathbf{A}_{j+k}\| \cdot \|\mathbf{A}'_i\| \rightarrow 0, \end{aligned}$$

as  $T \rightarrow \infty$ . Hence,

$$\lim_{T \rightarrow \infty} \mathbf{E}((\mathbf{Y}_{t-h} - \boldsymbol{\mu})(\mathbf{Y}_{t-h-j} - \boldsymbol{\mu})') = \sum_{i=0}^{\infty} \mathbf{A}_{i+j} \mathbf{G}(r) \mathbf{G}'(r) \mathbf{A}'_i. \quad \square$$

*Proof of Lemma 2.* (a) We have

$$\frac{1}{T} \sum_{t=1}^T \mathbf{X}_{t-1} \boldsymbol{\varepsilon}'_t = \frac{1}{T} \sum_{t=1}^T (\boldsymbol{\varepsilon}_t \mathbf{1}', \boldsymbol{\varepsilon}_t \mathbf{Y}'_{t-1}, \dots, \boldsymbol{\varepsilon}_t \mathbf{Y}'_{t-p})',$$

where

$$\mathbf{1}\boldsymbol{\varepsilon}'_t = \begin{pmatrix} \varepsilon_{t,1} & \dots & \varepsilon_{t,k} \\ & \dots & \\ \varepsilon_{t,1} & \dots & \varepsilon_{t,k} \end{pmatrix} \quad \text{and} \quad \mathbf{Y}_{t-h} \boldsymbol{\varepsilon}'_t = \begin{pmatrix} Y_{t-h,1} \varepsilon_{t,1} & \dots & Y_{t-h,1} \varepsilon_{t,k} \\ & \dots & \\ Y_{t-h,k} \varepsilon_{t,1} & \dots & Y_{t-h,k} \varepsilon_{t,k} \end{pmatrix}.$$

First, consider  $\mathbf{1}\boldsymbol{\varepsilon}'_t$ . The application of Assumption 3 leads to  $\mathbf{E}(\varepsilon_{t,i} | \mathcal{F}_{t-1}) = 0$ . From Assumption 2 we obtain

$$\mathbf{E}(\varepsilon_{t,i}^2) = \mathbf{E}(|\varepsilon_{t,i}|^2) \leq \mathbf{E} \|\varepsilon_t\|^2 = \mathbf{E} \left\| \mathbf{G} \left( \frac{t}{T} \right) \mathbf{u}_t \right\|^2 = \left\| \mathbf{G} \left( \frac{t}{T} \right) \right\|^2 \mathbf{E} \|\mathbf{u}_t\|^2 < \infty.$$

Next, we consider the product  $\mathbf{Y}_{t-h} \boldsymbol{\varepsilon}'_t$ ,  $1 \leq h \leq p$ . Assumption 3 implies that

$$\mathbf{E}(Y_{t-h,i} \varepsilon_{t,j} | \mathcal{F}_{t-1}) = Y_{t-h,i} \mathbf{E}(\varepsilon_{t,j} | \mathcal{F}_{t-1}) = 0 \quad \text{for } i, j \in \{1, \dots, k\}.$$

The application of Assumptions 2, 3 and Lemma 1 leads to

$$\begin{aligned} \mathbf{E}(Y_{t-h,j}^2 \varepsilon_{t,i}^2) & \leq \sqrt{\mathbf{E} Y_{t-h,j}^4 \mathbf{E} \varepsilon_{t,i}^4} \leq \sqrt{\mathbf{E} \|\mathbf{Y}_{t-h}\|^4 \mathbf{E} \|\varepsilon_t\|^4} \\ & \leq \sqrt{\mathbf{E} \sup_t \|\mathbf{Y}_{t-h}\|^4 C \left\| \mathbf{G} \left( \frac{t}{T} \right) \right\|^4 \sup_t \mathbf{E} \|\mathbf{u}_t\|^4} < \infty. \end{aligned}$$

Applying the law of large numbers for martingale differences we get that

$$\frac{1}{T} \sum_{t=1}^T \varepsilon_{t,i} \xrightarrow{P} 0 \quad \text{and} \quad \frac{1}{T} \sum_{t=1}^T Y_{t-h,j} \varepsilon_{t,i} \xrightarrow{P} 0.$$

(b) Let

$$\mathbf{X}_{t-1} \mathbf{X}'_{t-1} = \begin{pmatrix} \mathbf{1}\mathbf{1}' & \mathbf{1}\mathbf{Y}'_{t-1} & \dots & \mathbf{1}\mathbf{Y}'_{t-p} \\ \mathbf{Y}_{t-1} \mathbf{1}' & \mathbf{Y}_{t-1} \mathbf{Y}'_{t-1} & \dots & \mathbf{Y}_{t-1} \mathbf{Y}'_{t-p} \\ \dots & \dots & \dots & \dots \\ \mathbf{Y}_{t-p} \mathbf{1}' & \mathbf{Y}_{t-p} \mathbf{Y}'_{t-1} & \dots & \mathbf{Y}_{t-p} \mathbf{Y}'_{t-p} \end{pmatrix}.$$

In order to prove the statement (b) of Lemma 2, we have to calculate the following three limits in probability:

- (i)  $\lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \mathbf{1}\mathbf{1}'$ ;
- (ii)  $\lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \mathbf{1}\mathbf{Y}'_{t-h}$  and  $\lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \mathbf{Y}_{t-h} \mathbf{1}'$ , where  $1 \leq h \leq p$ ;

(iii)  $\lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \mathbf{Y}_{t-h} \mathbf{Y}'_{t-h-j}$ , where  $1 \leq h \leq p$  and  $1 \leq j \leq p-h$ .

Because  $\mathbf{1}$  is a constant vector, we get that  $\lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \mathbf{1} \mathbf{1}' = \mathbf{1} \mathbf{1}'$ .

For calculating the limits in probability of (ii), we have to check all the conditions of the law of large numbers for martingale differences. We choose  $m \geq p$  such that  $\|\mathbf{A}_{m-p+1}\| \neq 1$ . Then,

$$\mathbf{Y}_{t-p} - \boldsymbol{\mu} = \sum_{i=0}^{m-p} \mathbf{A}_i \boldsymbol{\varepsilon}_{t-p-i} + \mathbf{A}_{m-p+1} (\mathbf{Y}_{t-m-1} + \mathbf{B}_2 \mathbf{Y}_{t-m-2} + \cdots + \mathbf{B}_p \mathbf{Y}_{t-m-p}),$$

where  $\mathbf{B}_2, \dots, \mathbf{B}_p$  are some constant matrices. Let  $\mathcal{F}_{t-m}^{t+m}$  be a sigma field generated by  $\{\mathbf{u}_{t-m}, \dots, \mathbf{u}_{t+m}\}$ . We obtain

$$\begin{aligned} & \mathbb{E}(\mathbf{Y}_{t-p} - \boldsymbol{\mu} \mid \mathcal{F}_{t-m}^{t+m}) \\ &= \sum_{i=0}^{m-p} \mathbf{A}_i \boldsymbol{\varepsilon}_{t-p-i} \\ & \quad + \mathbf{A}_{m-p+1} \mathbb{E}(\mathbf{Y}_{t-m-1} + \mathbf{B}_2 \mathbf{Y}_{t-m-2} + \cdots + \mathbf{B}_p \mathbf{Y}_{t-m-p} \mid \mathcal{F}_{t-m}^{t+m}). \end{aligned}$$

Using the previous equality we get

$$\begin{aligned} & \|\mathbf{Y}_{t-p} - \boldsymbol{\mu} - \mathbb{E}(\mathbf{Y}_{t-p} - \boldsymbol{\mu} \mid \mathcal{F}_{t-m}^{t+m})\|_2 \\ &= \|\mathbf{A}_{m-p+1}\| \\ & \quad \times \left\| \mathbf{Y}_{t-m-1} + \mathbf{B}_2 \mathbf{Y}_{t-m-2} + \cdots + \mathbf{B}_p \mathbf{Y}_{t-m-p} \right. \\ & \quad \left. - \mathbb{E}(\mathbf{Y}_{t-m-1} + \mathbf{B}_2 \mathbf{Y}_{t-m-2} + \cdots + \mathbf{B}_p \mathbf{Y}_{t-m-p} \mid \mathcal{F}_{t-m}^{t+m}) \right\|_2. \end{aligned}$$

Using the Minkowski inequality, the conditional Jensen inequality and the law of iterated expectation we conclude that

$$\begin{aligned} & \|\mathbf{Y}_{t-p} - \boldsymbol{\mu} - \mathbb{E}(\mathbf{Y}_{t-p} - \boldsymbol{\mu} \mid \mathcal{F}_{t-m}^{t+m})\|_2 \\ & \leq \|\mathbf{A}_{m-p+1}\| \\ & \quad \times \left( \|\mathbf{Y}_{t-m-1} + \mathbf{B}_2 \mathbf{Y}_{t-m-2} + \cdots + \mathbf{B}_p \mathbf{Y}_{t-m-p}\|_2 \right. \\ & \quad \left. + \|\mathbb{E}(\mathbf{Y}_{t-m-1} + \mathbf{B}_2 \mathbf{Y}_{t-m-2} + \cdots + \mathbf{B}_p \mathbf{Y}_{t-m-p} \mid \mathcal{F}_{t-m}^{t+m})\|_2 \right) \\ & \leq 2 \|\mathbf{A}_{m-p+1}\| \|\mathbf{Y}_{t-m-1} + \mathbf{B}_2 \mathbf{Y}_{t-m-2} + \cdots + \mathbf{B}_p \mathbf{Y}_{t-m-p}\|_2 \\ & \leq C \|\mathbf{A}_{m-p+1}\| \sup_t \|\mathbf{Y}_t\|_2. \end{aligned}$$

Using the fact that  $\|\mathbf{A}_{m-p+1}\| \rightarrow 0$  as  $m \rightarrow \infty$  and the fact that  $\sup_t \|\mathbf{Y}_t\|_2 < \infty$  we get that  $\{\mathbf{Y}_{t-p} - \boldsymbol{\mu}\}$  is mean-zero near-epoch dependent in the  $L^2$ -norm on the  $\alpha$ -mixing sequence  $\{\mathbf{u}_t\}$ . From Theorem 17.10 of Davidson (1994) [6],  $\{\mathbf{Y}_{t-h} - \boldsymbol{\mu}\}$ ,  $1 \leq h \leq p$ , is also mean-zero near-epoch dependent in the  $L^2$ -norm on  $\{\mathbf{u}_t\}$ , and necessarily an  $L^1$ -mixingale. Lemma 1 gives us that it is uniformly integrable, and the application of the law of large numbers (Andrews (1988) [1, Theorem 1]) leads to

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbf{Y}_{t-h} \mathbf{1}' = \boldsymbol{\mu} \mathbf{1}' \quad \text{and} \quad \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbf{1} \mathbf{Y}'_{t-h} = \mathbf{1} \boldsymbol{\mu}'.$$

To prove (iii) we note that  $\{\mathbf{Y}_{t-h} - \boldsymbol{\mu}\}$  and  $\{\mathbf{Y}_{t-h-j} - \boldsymbol{\mu}\}$  are mean-zero near-epoch dependent in the  $L^2$ -norm on  $\{\mathbf{u}_t\}$ . From Theorem 17.9 of Davidson (1994) [6], it follows that  $\{(\mathbf{Y}_{t-h} - \boldsymbol{\mu})(\mathbf{Y}_{t-h-k} - \boldsymbol{\mu})' - \mathbb{E}(\mathbf{Y}_{t-h} - \boldsymbol{\mu})(\mathbf{Y}_{t-h-k} - \boldsymbol{\mu})'\}$  is mean-zero near-epoch dependent in the  $L^1$ -norm on  $\{\mathbf{u}_t\}$  and necessarily an  $L^1$ -mixingale. Lemma 1 gives

us that this sequence is uniformly integrable. From the law of large numbers for  $L^1$ -mixingales we obtain

$$\frac{1}{T} \sum_{t=1}^T ((\mathbf{Y}_{t-h} - \boldsymbol{\mu})(\mathbf{Y}_{t-h-j} - \boldsymbol{\mu})' - \mathbb{E}(\mathbf{Y}_{t-h} - \boldsymbol{\mu})(\mathbf{Y}_{t-h-j} - \boldsymbol{\mu})') \xrightarrow{P} \mathbf{0}.$$

Hence,

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T ((\mathbf{Y}_{t-h} - \boldsymbol{\mu})(\mathbf{Y}_{t-h-j} - \boldsymbol{\mu})') &= \frac{1}{T} \sum_{t=1}^T (\mathbb{E}(\mathbf{Y}_{t-h} - \boldsymbol{\mu})(\mathbf{Y}_{t-h-j} - \boldsymbol{\mu})') + o_P(1)\mathbf{1}\mathbf{1}' \\ &= \sum_{t=1}^T \int_{t/T}^{(t+1)/T} \mathbb{E}(\mathbf{Y}_{[Tr]-h} - \boldsymbol{\mu})(\mathbf{Y}_{[Tr]-h-j} - \boldsymbol{\mu})' dr + o_P(1)\mathbf{1}\mathbf{1}' \\ &= \int_{1/T}^{(T+1)/T} \mathbb{E}(\mathbf{Y}_{[Tr]-h} - \boldsymbol{\mu})(\mathbf{Y}_{[Tr]-h-j} - \boldsymbol{\mu})' dr + o_P(1)\mathbf{1}\mathbf{1}' \\ &\xrightarrow{P} \sum_{i=0}^{\infty} \mathbf{A}_{i+j} \left( \int_0^1 \mathbf{G}(r)\mathbf{G}'(r) dr \right) \mathbf{A}'_i. \end{aligned}$$

Finally, taking (ii) into account we get

$$\frac{1}{T} \sum_{t=1}^T (\mathbf{Y}_{t-h}\mathbf{Y}'_{t-h-j}) \xrightarrow{P} \boldsymbol{\mu}\boldsymbol{\mu}' + \sum_{i=0}^{\infty} \mathbf{A}_{i+j} \left( \int_0^1 \mathbf{G}(r)\mathbf{G}'(r) dr \right) \mathbf{A}'_i.$$

(c) Using the properties of the **vec** operator and the Kronecker product (see, e.g., Harville (1997) [15]) we get

$$\mathbf{vec}(\mathbf{X}_{t-1}\boldsymbol{\varepsilon}'_t) = \mathbf{vec}(\boldsymbol{\varepsilon}'_t \otimes \mathbf{X}_{t-1}) = \boldsymbol{\varepsilon}_t \otimes \mathbf{X}_{t-1}.$$

Hence,

$$\begin{aligned} \mathbf{vec}(\mathbf{X}_{t-1}\boldsymbol{\varepsilon}'_t) (\mathbf{vec}(\mathbf{X}_{t-1}\boldsymbol{\varepsilon}'_t))' &= (\boldsymbol{\varepsilon}_t \otimes \mathbf{X}_{t-1}) (\boldsymbol{\varepsilon}'_t \otimes \mathbf{X}'_{t-1}) \\ &= (\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t) \otimes (\mathbf{X}_{t-1}\mathbf{X}'_{t-1}) = \begin{pmatrix} \varepsilon_{t,1}^2 \mathbf{X}_{t-1}\mathbf{X}'_{t-1} & \dots & \varepsilon_{t,1}\varepsilon_{t,k} \mathbf{X}_{t-1}\mathbf{X}'_{t-1} \\ \varepsilon_{t,1}\varepsilon_{t,k} \mathbf{X}_{t-1}\mathbf{X}'_{t-1} & \dots & \varepsilon_{t,k}^2 \mathbf{X}_{t-1}\mathbf{X}'_{t-1} \end{pmatrix}. \end{aligned}$$

Without loss of generality we calculate  $\lim_{T \rightarrow \infty} \varepsilon_{t,1}^2 \mathbf{X}_{t-1}\mathbf{X}'_{t-1}$ . The other limits in probability can be computed in the same way. It follows that

$$\varepsilon_{t,1}^2 \mathbf{X}_{t-1}\mathbf{X}'_{t-1} = \begin{pmatrix} \varepsilon_{t,1}^2 \mathbf{1}\mathbf{1}' & \varepsilon_{t,1}^2 \mathbf{1}\mathbf{Y}'_{t-1} & \dots & \varepsilon_{t,1}^2 \mathbf{1}\mathbf{Y}'_{t-p} \\ \varepsilon_{t,1}^2 \mathbf{Y}_{t-1}\mathbf{1}' & \varepsilon_{t,1}^2 \mathbf{Y}_{t-1}\mathbf{Y}'_{t-1} & \dots & \varepsilon_{t,1}^2 \mathbf{Y}_{t-1}\mathbf{Y}'_{t-p} \\ \dots & \dots & \dots & \dots \\ \varepsilon_{t,1}^2 \mathbf{Y}_{t-p}\mathbf{1}' & \varepsilon_{t,1}^2 \mathbf{Y}_{t-p}\mathbf{Y}'_{t-1} & \dots & \varepsilon_{t,1}^2 \mathbf{Y}_{t-p}\mathbf{Y}'_{t-p} \end{pmatrix}.$$

It is necessary to find the next three limits in probability:

- (i)  $\lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \varepsilon_{t,1}^2 \mathbf{1}\mathbf{1}'$ ;
- (ii)  $\lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \varepsilon_{t,1}^2 \mathbf{Y}_{t-h}\mathbf{1}'$  and  $\lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \varepsilon_{t,1}^2 \mathbf{1}\mathbf{Y}'_{t-h}$ , where  $1 \leq h \leq p$ ;
- (iii)  $\lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \varepsilon_{t,1}^2 \mathbf{Y}_{t-h}\mathbf{Y}'_{t-h-j}$ , where  $1 \leq h \leq p$  and  $1 \leq j \leq p-h$ .

First, (i) is proved. The identity

$$\varepsilon_{t,1}^2 = \left( \sum_{j=1}^k g_{1j}(t/T)u_{t,j} \right)^2$$

and the application of Assumption 3 lead to  $\mathbb{E}(\varepsilon_{t,1}^2 \mid \mathcal{F}_{t-1}) = \sum_{j=1}^k g_{1j}^2(t/T)$ . Because  $\{\varepsilon_{t,1}^2 - \sum_{j=1}^k g_{1j}^2(t/T)\}$  is  $\alpha$ -mixing (see Davidson (1994) [6, Theorem 14.1]), and

$$\mathbb{E}\left(\varepsilon_{t,1}^2 - \sum_{j=1}^k g_{1j}^2(t/T)\right)^2 < \infty$$

(Assumptions 2 and 3), the law of large numbers for  $L^1$ -mixingales implies that

$$\frac{1}{T} \sum_{t=1}^T \varepsilon_{t,1}^2 = \frac{1}{T} \sum_{t=1}^T \mathbb{E}(\varepsilon_{t,1}^2) + o_P(1) = \frac{1}{T} \sum_{t=1}^T \sum_{j=1}^k g_{1j}^2\left(\frac{t}{T}\right) + o_P(1) \xrightarrow{P} \sum_{j=1}^k \int_0^1 g_{1j}^2(r) dr.$$

Hence,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \varepsilon_{t,1}^2 \mathbf{1}\mathbf{1}' = \left( \sum_{j=1}^k \int_0^1 g_{1j}^2(r) dr \right) \mathbf{1}\mathbf{1}'.$$

(ii) Using the fact that  $\{\varepsilon_{t,1}^2\}$  is  $\alpha$ -mixing and therefore near-epoch dependent in the  $L^2$ -norm on  $\{\mathbf{u}_t\}$  as well as the fact that  $\{\mathbf{Y}_{t-h} - \boldsymbol{\mu}\}$  is also near-epoch dependent in the  $L^2$ -norm on  $\{\mathbf{u}_t\}$ , we obtain that  $\{\varepsilon_{t,1}^2(\mathbf{Y}_{t-h} - \boldsymbol{\mu})\}$  is near-epoch dependent in the  $L^1$ -norm on  $\{\mathbf{u}_t\}$  (see Davidson (1994) [6, Theorem 17.9]). Moreover, from Assumption 3 and Lemma 1 it follows that

$$\mathbb{E} \|\varepsilon_{t,1}^2(\mathbf{Y}_{t-h} - \boldsymbol{\mu})\|^\delta \leq \mathbb{E} \varepsilon_{t,1}^{4\delta} \mathbb{E} \|(\mathbf{Y}_{t-h} - \boldsymbol{\mu})\|^{2\delta} < \infty.$$

The application of the law of large numbers for  $L^1$ -mixingales yields

$$T^{-1} \sum_{t=1}^T \varepsilon_{t,1}^2(\mathbf{Y}_{t-h} - \boldsymbol{\mu}) \xrightarrow{P} \mathbf{0}$$

and, consequently,

$$T^{-1} \sum_{t=1}^T \varepsilon_{t,1}^2 \mathbf{Y}_{t-h} \xrightarrow{P} \left( \sum_{j=1}^k \int_0^1 g_{1j}^2(r) dr \right) \boldsymbol{\mu}.$$

Hence,

$$\frac{1}{T} \sum_{t=1}^T \varepsilon_{t,1}^2 \mathbf{Y}_{t-h} \mathbf{1}' \xrightarrow{P} \left( \sum_{j=1}^k \int_0^1 g_{1j}^2(r) dr \right) \boldsymbol{\mu} \mathbf{1}'$$

and

$$\frac{1}{T} \sum_{t=1}^T \varepsilon_{t,1}^2 \mathbf{1} \mathbf{Y}'_{t-h} \xrightarrow{P} \left( \sum_{j=1}^k \int_0^1 g_{1j}^2(r) dr \right) \mathbf{1} \boldsymbol{\mu}'.$$

(iii) Consider

$$\left\{ \varepsilon_{t,1}^2(\mathbf{Y}_{t-h} - \boldsymbol{\mu})(\mathbf{Y}_{t-h-j} - \boldsymbol{\mu})' - \left( \sum_{j=i}^k g_{1i}^2\left(\frac{t}{T}\right) \right) (\mathbf{Y}_{t-h} - \boldsymbol{\mu})(\mathbf{Y}_{t-h-j} - \boldsymbol{\mu})', \mathcal{F}_t \right\}.$$

It follows that

$$\begin{aligned} & \mathbb{E} \left( \varepsilon_{t,1}^2(\mathbf{Y}_{t-h} - \boldsymbol{\mu})(\mathbf{Y}_{t-h-j} - \boldsymbol{\mu})' - \left( \sum_{i=1}^k g_{1i}^2\left(\frac{t}{T}\right) \right) (\mathbf{Y}_{t-h} - \boldsymbol{\mu})(\mathbf{Y}_{t-h-j} - \boldsymbol{\mu})' \mid \mathcal{F}_{t-1} \right) \\ &= (\mathbf{Y}_{t-h} - \boldsymbol{\mu})(\mathbf{Y}_{t-h-j} - \boldsymbol{\mu})' \mathbb{E} \left( \varepsilon_{t,1}^2 - \sum_{i=1}^k g_{1i}^2\left(\frac{t}{T}\right) \mid \mathcal{F}_{t-1} \right) = 0. \end{aligned}$$

From the Minkowski and the Cauchy–Schwarz inequalities and the application of Lemma 1, we obtain

$$\begin{aligned}
& \left\| \varepsilon_{t,1}^2 (\mathbf{Y}_{t-h} - \boldsymbol{\mu})(\mathbf{Y}_{t-h-j} - \boldsymbol{\mu})' - \left( \sum_{i=1}^k g_{1i}^2 \left( \frac{t}{T} \right) \right) (\mathbf{Y}_{t-h} - \boldsymbol{\mu})(\mathbf{Y}_{t-h-j} - \boldsymbol{\mu})' \right\|_{\delta} \\
& \leq \left\| \varepsilon_{t,1}^2 (\mathbf{Y}_{t-h} - \boldsymbol{\mu})(\mathbf{Y}_{t-h-j} - \boldsymbol{\mu})' \right\|_{\delta} \\
& \quad + \left\| \left( \sum_{i=1}^k g_{1i}^2 \left( \frac{t}{T} \right) \right) (\mathbf{Y}_{t-h} - \boldsymbol{\mu})(\mathbf{Y}_{t-h-j} - \boldsymbol{\mu})' \right\|_{\delta} \\
& = \left( \mathbb{E} \varepsilon_{t,1}^{2\delta} \|\mathbf{Y}_{t-h} - \boldsymbol{\mu}\|^{\delta} \|\mathbf{Y}_{t-h-j} - \boldsymbol{\mu}\|^{\delta} \right)^{1/\delta} \\
& \quad + \left( \sum_{i=1}^k g_{1i}^2 \left( \frac{t}{T} \right) \right) \left( \mathbb{E} \|\mathbf{Y}_{t-h} - \boldsymbol{\mu}\|^{\delta} \|\mathbf{Y}_{t-h-j} - \boldsymbol{\mu}\|^{\delta} \right)^{1/\delta} \\
& \leq \left( \mathbb{E} \varepsilon_{t,1}^{4\delta} \left( \mathbb{E} \|\mathbf{Y}_{t-h} - \boldsymbol{\mu}\|^{4\delta} \mathbb{E} \|\mathbf{Y}_{t-h-j} - \boldsymbol{\mu}\|^{4\delta} \right)^{1/2} \right)^{1/(2\delta)} \\
& \quad + \left( \sum_{i=1}^k g_{1i}^2 \left( \frac{t}{T} \right) \right) \left( \mathbb{E} \|\mathbf{Y}_{t-h} - \boldsymbol{\mu}\|^{2\delta} \mathbb{E} \|\mathbf{Y}_{t-h-j} - \boldsymbol{\mu}\|^{2\delta} \right)^{1/(2\delta)} \\
& < \infty.
\end{aligned}$$

Hence,

$$\left\{ \varepsilon_{t,1}^2 (\mathbf{Y}_{t-h} - \boldsymbol{\mu})(\mathbf{Y}_{t-h-j} - \boldsymbol{\mu})' - \left( \sum_{i=1}^k g_{1i}^2 \left( \frac{t}{T} \right) \right) (\mathbf{Y}_{t-h} - \boldsymbol{\mu})(\mathbf{Y}_{t-h-j} - \boldsymbol{\mu})', \mathcal{F}_t \right\}$$

is a martingale difference sequence. From the law of large numbers, we get

$$\begin{aligned}
& \frac{1}{T} \sum_{t=1}^T \varepsilon_{t,1}^2 (\mathbf{Y}_{t-h} - \boldsymbol{\mu})(\mathbf{Y}_{t-h-j} - \boldsymbol{\mu})' \\
& \quad - \frac{1}{T} \sum_{t=1}^T \left( \sum_{i=1}^k g_{1i}^2 \left( \frac{t}{T} \right) \right) (\mathbf{Y}_{t-h} - \boldsymbol{\mu})(\mathbf{Y}_{t-h-j} - \boldsymbol{\mu})' \\
& \xrightarrow{P} \mathbf{0}.
\end{aligned}$$

Next, we calculate

$$\begin{aligned}
& \frac{1}{T} \sum_{t=1}^T \left( \sum_{i=1}^k g_{1i}^2 \left( \frac{t}{T} \right) \right) (\mathbf{Y}_{t-h} - \boldsymbol{\mu})(\mathbf{Y}_{t-h-j} - \boldsymbol{\mu})' \\
& = \frac{1}{T} \sum_{t=1}^T \left( \sum_{i=1}^k g_{1i}^2 \left( \frac{t}{T} \right) \right) \mathbb{E} (\mathbf{Y}_{t-h} - \boldsymbol{\mu})(\mathbf{Y}_{t-h-j} - \boldsymbol{\mu})' + o_P(1) \mathbf{1} \mathbf{1}' \\
& = \frac{1}{T} \sum_{t=1}^T \int_{t/T}^{(t+1)/T} \left( \sum_{i=1}^k g_{1i}^2 \left( \frac{[Tr]}{T} \right) \right) \mathbb{E} (\mathbf{Y}_{[Tr]-h} - \boldsymbol{\mu})(\mathbf{Y}_{[Tr]-h-j} - \boldsymbol{\mu})' dr \\
& \quad + o_P(1) \mathbf{1} \mathbf{1}' \\
& = \int_{1/T}^{(T+1)/T} \left( \sum_{i=1}^k g_{1i}^2 \left( \frac{[Tr]}{T} \right) \right) \mathbb{E} (\mathbf{Y}_{[Tr]-h} - \boldsymbol{\mu})(\mathbf{Y}_{[Tr]-h-j} - \boldsymbol{\mu})' dr + o_P(1) \mathbf{1} \mathbf{1}' \\
& \xrightarrow{P} \sum_{l=0}^{\infty} \mathbf{A}_{l+j} \left( \int_0^1 \left( \sum_{i=1}^k g_{1i}^2(r) \right) \mathbf{G}(r) \mathbf{G}'(r) dr \right) \mathbf{A}_l'.
\end{aligned}$$

Thus,

$$\begin{aligned} & \frac{1}{T} \sum_{t=1}^T \varepsilon_{t,1}^2 (\mathbf{Y}_{t-h} - \boldsymbol{\mu})(\mathbf{Y}_{t-h-j} - \boldsymbol{\mu})' \frac{1}{T} \sum_{t=1}^T \left( \sum_{i=1}^k g_{1i}^2 \left( \frac{t}{T} \right) \right) (\mathbf{Y}_{t-h} - \boldsymbol{\mu})(\mathbf{Y}_{t-h-j} - \boldsymbol{\mu})' \\ & \quad + o_P(1) \mathbf{1} \mathbf{1}' \\ & \xrightarrow{P} \sum_{l=0}^{\infty} \mathbf{A}_{l+j} \left( \int_0^1 \left( \sum_{i=1}^k g_{1i}^2(r) \right) \mathbf{G}(r) \mathbf{G}'(r) dr \right) \mathbf{A}_l'. \end{aligned}$$

Finally we get

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \varepsilon_{t,1}^2 \mathbf{Y}_{t-h} \mathbf{Y}'_{t-h-j} &= \left( \sum_{i=1}^k \int_0^1 g_{1i}^2(r) dr \right) \boldsymbol{\mu} \boldsymbol{\mu}' \\ & \quad + \sum_{l=0}^{\infty} \mathbf{A}_{l+j} \left( \int_0^1 \left( \sum_{i=1}^k g_{1i}^2(r) \right) \mathbf{G}(r) \mathbf{G}'(r) dr \right) \mathbf{A}_l'. \end{aligned}$$

(d) By the Cramer–Wold device it is sufficient to show that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \boldsymbol{\lambda}' \mathbf{vec}(\mathbf{X}_{t-1} \boldsymbol{\varepsilon}'_t) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\lambda}' \boldsymbol{\Omega}_2 \boldsymbol{\lambda})$$

for every fixed  $k^2(p+1)$ -vector  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_{k^2(p+1)})' \neq \mathbf{0}$ . Note that

$$\{\boldsymbol{\lambda}' \mathbf{vec}(\mathbf{X}_{t-1} \boldsymbol{\varepsilon}'_t), \mathcal{F}_t\}$$

is a martingale difference sequence and by (c),

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \boldsymbol{\lambda}' \mathbf{vec}(\mathbf{X}_{t-1} \boldsymbol{\varepsilon}'_t) (\mathbf{vec}(\mathbf{X}_{t-1} \boldsymbol{\varepsilon}'_t))' \boldsymbol{\lambda} \xrightarrow{P} \boldsymbol{\lambda}' \boldsymbol{\Omega}_2 \boldsymbol{\lambda}.$$

Moreover,

$$\begin{aligned} & \mathbf{E} \left( \boldsymbol{\lambda}' \mathbf{vec}(\mathbf{X}_{t-1} \boldsymbol{\varepsilon}'_t) \right)^{2\delta} \\ &= \left\| (\lambda_1 + \dots + \lambda_k) \varepsilon_{t,1} + (\lambda_{k+1} Y_{t-1,1} + \dots + \lambda_{2k} Y_{t-1,k}) \varepsilon_{t,1} + \dots \right. \\ & \quad \left. + (\lambda_{pk+1} Y_{t-p,1} + \dots + \lambda_{p(k+1)} Y_{t-p,k}) \varepsilon_{t,1} + \dots \right\|_{2\delta}^{2\delta} \\ &\leq \left( \left\| (\lambda_1 + \dots + \lambda_k) \varepsilon_{t,1} \right\|_{2\delta} + \left\| \lambda_{k+1} Y_{t-1,1} \varepsilon_{t,1} \right\|_{2\delta} + \dots \right. \\ & \quad \left. + \left\| \lambda_{2k} Y_{t-1,k} \varepsilon_{t,1} \right\|_{2\delta} + \dots + \left\| \lambda_{pk+1} Y_{t-p,1} \varepsilon_{t,1} \right\|_{2\delta} + \dots \right. \\ & \quad \left. + \left\| \lambda_{p(k+1)} Y_{t-p,k} \varepsilon_{t,1} \right\|_{2\delta} + \dots \right)^{2\delta} \\ &< \infty, \end{aligned}$$

since for  $1 \leq h \leq p$  and  $1 \leq i, j \leq k$ ,

$$\|Y_{t-h,i} \varepsilon_{t,j}\|_{2\delta} = (\mathbf{E}(Y_{t-h,i} \varepsilon_{t,j}))^{1/(2\delta)} \leq (\mathbf{E}(Y_{t-h,i})^{4\delta} \mathbf{E}(\varepsilon_{t,j})^{4\delta})^{1/(4\delta)} < \infty$$

by Lemma 1. Thus from the central limit theorem for martingale differences (e.g., White (1984) [24, Corollary 5.25]) we get that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \boldsymbol{\lambda}' \mathbf{vec}(\mathbf{X}_{t-1} \boldsymbol{\varepsilon}'_t) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\lambda}' \boldsymbol{\Omega}_2 \boldsymbol{\lambda})$$

for every fixed  $k^2(p+1)$ -vector  $\boldsymbol{\lambda} \neq \mathbf{0}$ . Consequently,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \text{vec}(\mathbf{X}_{t-1} \boldsymbol{\varepsilon}'_t) \rightarrow^d \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_2). \quad \square$$

*Proof of Lemma 3.* (a) From the definition of  $\hat{\boldsymbol{\varepsilon}}_t$  we get

$$\hat{\boldsymbol{\varepsilon}}_t = \mathbf{Y}_t - \hat{\boldsymbol{\theta}} \mathbf{X}'_{t-1} = \boldsymbol{\varepsilon}_t - (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \mathbf{X}'_{t-1} = \hat{\boldsymbol{\varepsilon}}_t - \boldsymbol{\zeta}_t,$$

where

$$\boldsymbol{\zeta}_t = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \mathbf{X}'_{t-1}.$$

From Theorem 1, Lemma 1 and Markov's inequality, it follows that

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = O_P(T^{-1/2}) \mathbf{1} \mathbf{1}', \quad \mathbf{X}_{t-1} = O_P(1) \mathbf{1}, \quad \boldsymbol{\zeta}_t = O_P(T^{-1/2}) \mathbf{1}.$$

Then

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}'_t &= \frac{1}{T} \sum_{t=1}^T \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t - \frac{1}{T} \sum_{t=1}^T \boldsymbol{\varepsilon}_t \boldsymbol{\zeta}'_t - \frac{1}{T} \sum_{t=1}^T \boldsymbol{\zeta}_t \boldsymbol{\varepsilon}'_t + \frac{1}{T} \sum_{t=1}^T \boldsymbol{\zeta}_t \boldsymbol{\zeta}'_t \\ &= \frac{1}{T} \sum_{t=1}^T \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t - \frac{2}{T} O_P(\sqrt{T}) \mathbf{1} \mathbf{1}' + \frac{1}{T} O_P(1) \mathbf{1} \mathbf{1}' = \frac{1}{T} \sum_{t=1}^T \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t + o_P(1) \mathbf{1} \mathbf{1}' \\ &\xrightarrow{P} \int_0^1 \mathbf{G}(r) \mathbf{G}'(r) dr. \end{aligned}$$

(b) Using the definition of the matrix  $\mathbf{W}$  and the properties of the Kronecker product and the  $\text{vec}$  operator we obtain

$$\begin{aligned} &\mathbf{E}(\text{vec}(\text{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t) (\text{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t))')) \\ &= \mathbf{E} \left( \text{vec} \left( \text{vec} \left( \mathbf{G} \left( \frac{t}{T} \right) \mathbf{u}_t \mathbf{u}'_t \mathbf{G}' \left( \frac{t}{T} \right) \right) \left( \text{vec} \left( \mathbf{G} \left( \frac{t}{T} \right) \mathbf{u}_t \mathbf{u}'_t \mathbf{G}' \left( \frac{t}{T} \right) \right) \right)' \right) \right) \\ &= \mathbf{E} \left( \text{vec} \left( \left( \mathbf{G} \left( \frac{t}{T} \right) \otimes \mathbf{G} \left( \frac{t}{T} \right) \right) \text{vec}(\mathbf{u}_t \mathbf{u}'_t) (\text{vec}(\mathbf{u}_t \mathbf{u}'_t))' \right. \right. \\ &\quad \left. \left. \times \left( \mathbf{G}' \left( \frac{t}{T} \right) \otimes \mathbf{G}' \left( \frac{t}{T} \right) \right) \right) \right) \\ &= \mathbf{E} \left( \text{vec} \left( \left( \mathbf{G} \left( \frac{t}{T} \right) \otimes \mathbf{G} \left( \frac{t}{T} \right) \right) ((\mathbf{u}_t \mathbf{u}'_t) \otimes (\mathbf{u}_t \mathbf{u}'_t)) \left( \mathbf{G}' \left( \frac{t}{T} \right) \otimes \mathbf{G}' \left( \frac{t}{T} \right) \right) \right) \right) \\ &= \left( \mathbf{G} \left( \frac{t}{T} \right) \otimes \mathbf{G} \left( \frac{t}{T} \right) \otimes \mathbf{G} \left( \frac{t}{T} \right) \otimes \mathbf{G} \left( \frac{t}{T} \right) \right) \text{vec}(\mathbf{E}((\mathbf{u}_t \mathbf{u}'_t) \otimes (\mathbf{u}_t \mathbf{u}'_t))) \\ &= \left( \mathbf{G} \left( \frac{t}{T} \right) \otimes \mathbf{G} \left( \frac{t}{T} \right) \otimes \mathbf{G} \left( \frac{t}{T} \right) \otimes \mathbf{G} \left( \frac{t}{T} \right) \right) \text{vec} \mathbf{W}. \end{aligned}$$

Since

$$\left\{ \text{vec}(\text{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t) (\text{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t))') - \left( \mathbf{G}(t/T) \otimes \mathbf{G}(t/T) \otimes \mathbf{G}(t/T) \otimes \mathbf{G}(t/T) \right) \text{vec} \mathbf{W} \right\}$$

is  $\alpha$ -mixing and

$$\mathbf{E}(\text{vec}(\text{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t) (\text{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t))')^{4\delta}) < \infty$$

for some  $\delta > 1$ , it is an  $L^1$ -mixingale. The application of the law of large numbers leads to

$$\frac{1}{T} \sum_{t=1}^T \left( \text{vec}(\text{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t) (\text{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t))') - \mathbf{G}^{(4)} \left( \frac{t}{T} \right) \text{vec} \mathbf{W} \right) \xrightarrow{P} \mathbf{0},$$

where

$$\mathbf{G}^{(4)}(t/T) = \mathbf{G}(t/T) \otimes \mathbf{G}(t/T) \otimes \mathbf{G}(t/T) \otimes \mathbf{G}(t/T).$$

From the arguments similar to (a) we conclude that

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T \mathbf{vec}(\mathbf{vec}(\hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_t') (\mathbf{vec}(\hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_t'))') &= \frac{1}{T} \sum_{t=1}^T \mathbf{vec}(\mathbf{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') (\mathbf{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'))') + o_P(1) \mathbf{1} \\ &= \frac{1}{T} \sum_{t=1}^T \left( \mathbf{G}^{(4)} \left( \frac{t}{T} \right) \mathbf{vec} \mathbf{W} \right) + o_P(1) \mathbf{1} \xrightarrow{P} \int_0^1 \mathbf{G}^{(4)}(r) dr \mathbf{vec} \mathbf{W}. \end{aligned}$$

(c) Similarly to (a), we obtain

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T (\mathbf{vec}(\mathbf{X}_{t-1} \hat{\boldsymbol{\varepsilon}}_t') \mathbf{vec}(\mathbf{X}_{t-1} \hat{\boldsymbol{\varepsilon}}_t')) &= \frac{1}{T} \sum_{t=1}^T ((\hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_t') \otimes (\mathbf{X}_{t-1} \mathbf{X}_{t-1}')) \\ &= \frac{1}{T} \sum_{t=1}^T (((\boldsymbol{\varepsilon}_t - \boldsymbol{\zeta}_t)(\boldsymbol{\varepsilon}_t' - \boldsymbol{\zeta}_t') \otimes (\mathbf{X}_{t-1} \mathbf{X}_{t-1}')) \\ &= \frac{1}{T} \sum_{t=1}^T ((\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') \otimes (\mathbf{X}_{t-1} \mathbf{X}_{t-1}')) + \frac{1}{T} \sum_{t=1}^T ((-\boldsymbol{\varepsilon}_t \boldsymbol{\zeta}_t' - \boldsymbol{\zeta}_t \boldsymbol{\varepsilon}_t' + \boldsymbol{\zeta}_t \boldsymbol{\zeta}_t') \otimes (\mathbf{X}_{t-1} \mathbf{X}_{t-1}')) \\ &= \frac{1}{T} \sum_{t=1}^T ((\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') \otimes (\mathbf{X}_{t-1} \mathbf{X}_{t-1}')) + o_P(1) \mathbf{1} \xrightarrow{P} \boldsymbol{\Omega}_2, \end{aligned}$$

where the last equality holds since

$$\begin{aligned} &\left\| \frac{1}{T} \sum_{t=1}^T ((-\boldsymbol{\varepsilon}_t \boldsymbol{\zeta}_t' - \boldsymbol{\zeta}_t \boldsymbol{\varepsilon}_t' + \boldsymbol{\zeta}_t \boldsymbol{\zeta}_t') \otimes (\mathbf{X}_{t-1} \mathbf{X}_{t-1}')) \right\| \\ &\leq \frac{2}{T} \sum_{t=1}^T \|\boldsymbol{\varepsilon}_t\| \|\boldsymbol{\zeta}_t'\| \|\mathbf{X}_{t-1} \mathbf{X}_{t-1}'\| + \frac{1}{T} \sum_{t=1}^T \|\boldsymbol{\zeta}_t\|^2 \|\mathbf{X}_{t-1} \mathbf{X}_{t-1}'\| \\ &= \frac{2}{T} \sum_{t=1}^T O_P(1/\sqrt{T}) + \frac{1}{T} \sum_{t=1}^T O_P(1/T) = o_P(1). \end{aligned}$$

(d) Let

$$K_{rt} = K \left( \frac{[Tr] - t}{Tb} \right)$$

and for  $1 \leq t \leq T$  let

$$[Tr] - t = [Tx],$$

where  $-1 \leq x < 1$ . Then

$$\begin{aligned} \frac{1}{Tb} \sum_{t=1}^T K_{rt} &= \sum_{t=1}^T \int_{([Tr]-t)/T}^{([Tr]-t+1)/T} K \left( \frac{[Tx]}{Tb} \right) d \left( \frac{x}{b} \right) = \sum_{t=1}^T \int_{([Tr]-t)/Tb}^{([Tr]-t+1)/Tb} K \left( \frac{[Tbx]}{Tb} \right) dx \\ &= \int_{([Tr]-T)/Tb}^{[Tr]/T} K \left( \frac{[Tx]}{Tb} \right) d \left( \frac{x}{b} \right) \rightarrow \int_{-\infty}^{+\infty} K(x) dx < \infty. \end{aligned}$$



The application of  $\hat{\varepsilon}_t = \varepsilon_t - \zeta_t$  yields

$$\begin{aligned}
& \left\| \left( \frac{1}{Tb} \sum_{t=1}^T K_{rt} \right) \left( \hat{\mathbf{G}}(r) \hat{\mathbf{G}}(r)' - \mathbf{G}(r) \mathbf{G}'(r) \right) \right\| = \left\| \frac{1}{Tb} \sum_{t=1}^T K_{rt} \left( \hat{\varepsilon}_t \hat{\varepsilon}_t' - \mathbf{G}(r) \mathbf{G}'(r) \right) \right\| \\
& = \left\| \frac{1}{Tb} \sum_{t=1}^T K_{rt} \left( \varepsilon_t \varepsilon_t' - \zeta_t \zeta_t' - \varepsilon_t \zeta_t' + \zeta_t \zeta_t' + \mathbf{G} \left( \frac{t}{T} \right) \mathbf{G}' \left( \frac{t}{T} \right) \right. \right. \\
& \qquad \qquad \qquad \left. \left. - \mathbf{G} \left( \frac{t}{T} \right) \mathbf{G}' \left( \frac{t}{T} \right) - \mathbf{G}(r) \mathbf{G}'(r) \right) \right\| \\
(10) \quad & \leq \left\| \frac{1}{Tb} \sum_{t=1}^T K_{rt} \left( \varepsilon_t \varepsilon_t' - \mathbf{G} \left( \frac{t}{T} \right) \mathbf{G}' \left( \frac{t}{T} \right) \right) \right\| \\
& \quad + \left\| \frac{1}{Tb} \sum_{t=1}^T K_{rt} \left( \mathbf{G} \left( \frac{t}{T} \right) \mathbf{G}' \left( \frac{t}{T} \right) - \mathbf{G}(r) \mathbf{G}'(r) \right) \right\| \\
& \quad + 2 \left\| \frac{1}{Tb} \sum_{t=1}^T K_{rt} \zeta_t \varepsilon_t' \right\| + \left\| \frac{1}{Tb} \sum_{t=1}^T K_{rt} \zeta_t \zeta_t' \right\|.
\end{aligned}$$

Taking into account the fact that  $\{\varepsilon_t \varepsilon_t' - \mathbf{G}(t/T) \mathbf{G}'(t/T)\}$  is a martingale difference sequence we obtain

$$\begin{aligned}
& \mathbb{E} \left\| \frac{1}{Tb} \sum_{t=1}^T K_{rt} \left( \varepsilon_t \varepsilon_t' - \mathbf{G} \left( \frac{t}{T} \right) \mathbf{G}' \left( \frac{t}{T} \right) \right) \right\|^2 \\
& = \frac{1}{T^2 b^2} \sum_{t=1}^T K_{rt}^2 \mathbb{E} \left\| \varepsilon_t \varepsilon_t' - \mathbf{G} \left( \frac{t}{T} \right) \mathbf{G}' \left( \frac{t}{T} \right) \right\|^2 \\
& \leq \frac{1}{Tb} \left( \sup_t K_{rt} \right) \left( \sup_t \mathbb{E} \left\| \varepsilon_t \varepsilon_t' - \mathbf{G} \left( \frac{t}{T} \right) \mathbf{G}' \left( \frac{t}{T} \right) \right\|^2 \right) \left( \frac{1}{Tb} \sum_{t=1}^T K_{rt} \right) \rightarrow 0.
\end{aligned}$$

Hence, the first term in (10) vanishes in the  $L^2$ -norm as  $T \rightarrow \infty$ .

For the second term in (10) we get

$$\begin{aligned}
& \left\| \frac{1}{Tb} \sum_{t=1}^T K_{rt} \left( \mathbf{G} \left( \frac{t}{T} \right) \mathbf{G}' \left( \frac{t}{T} \right) - \mathbf{G}(r) \mathbf{G}'(r) \right) \right\| \\
& \leq \left\| \frac{1}{Tb} \sum_{|t-Tr| \leq TMb} K_{rt} \left( \mathbf{G} \left( \frac{t}{T} \right) \mathbf{G}' \left( \frac{t}{T} \right) - \mathbf{G}(r) \mathbf{G}'(r) \right) \right\| \\
& \quad + \left\| \frac{1}{Tb} \sum_{|t-Tr| > TMb} K_{rt} \left( \mathbf{G} \left( \frac{t}{T} \right) \mathbf{G}' \left( \frac{t}{T} \right) - \mathbf{G}(r) \mathbf{G}'(r) \right) \right\|.
\end{aligned}$$

Because the function  $\mathbf{G}(\cdot)$  is uniformly bounded and satisfies the Lipschitz condition (Assumption 2), we obtain

$$\begin{aligned}
& \left\| \frac{1}{Tb} \sum_{|t-Tr| \leq TMb} K_{rt} \left( \mathbf{G} \left( \frac{t}{T} \right) \mathbf{G}' \left( \frac{t}{T} \right) - \mathbf{G}(r) \mathbf{G}'(r) \right) \right\| \\
& \leq \frac{C}{Tb} \sum_{|t-Tr| \leq TMb} K_{rt} \left( \frac{t}{T} - r \right) \leq \frac{CMb}{Tb} \sum_{|t-Tr| \leq TMb} K_{rt} \leq 2CM^2 b^2
\end{aligned}$$

and

$$\begin{aligned} & \left\| \frac{1}{Tb} \sum_{|t-Tr|>TMb} K_{rt} \left( \mathbf{G} \left( \frac{t}{T} \right) \mathbf{G}' \left( \frac{t}{T} \right) - \mathbf{G}(r) \mathbf{G}'(r) \right) \right\| \\ & \leq \frac{C}{Tb} \sum_{|t-Tr|>TMb} K_{rt} \rightarrow C \int_{|x| \geq M} K(x) dx. \end{aligned}$$

Letting  $T \rightarrow \infty$  and taking  $M$  such that  $M \rightarrow \infty$  and  $Mb \rightarrow 0$ , we get that the second term in (10) converges to zero.

Since

$$\frac{1}{Tb} \sum_{t=1}^T K_{rt} \zeta_t \zeta_t' = O_P \left( \frac{1}{T} \right) \mathbf{11}',$$

we conclude that the third and the fourth terms of (10) converge in probability to zero as  $T \rightarrow \infty$ . Therefore,  $\hat{\mathbf{G}}(r) \hat{\mathbf{G}}'(r) \xrightarrow{P} \mathbf{G}(r) \mathbf{G}'(r)$ , for all  $r \in [0, 1]$  for which the function  $\mathbf{G}(\cdot)$  is continuous.  $\square$

#### ACKNOWLEDGMENTS

The authors are thankful to the referees and the editor for their suggestions which have significantly improved the presentation in the paper.

#### BIBLIOGRAPHY

1. D. W. K. Andrews, *Laws of large numbers for dependent nonidentically distributed random variables*, *Econometric Theory* **4** (1988), 458–467. MR985156 (90c:60013)
2. T. Breusch and A. Pagan, *A simple test of heteroscedasticity and random coefficient variation*, *Econometrica* **47** (1979), 1287–1294. MR545960 (81b:62081)
3. P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, Springer, New York, 1991. MR1093459 (92d:62001)
4. G. Cavaliere, *Unit root tests under time-varying variance shifts*, *Econometric Reviews* **23** (2004), 259–292. MR2089641 (2005d:62138)
5. S. Darolles, C. Gourieroux, and J. Jasiak, *Structural Laplace transform and compound autoregressive models*, *Journal of Time Series Analysis* **27** (2006), 477–503. MR2245710
6. J. Davidson, *Stochastic Limit Theory: An Introduction for Econometricians*, Oxford, Oxford University Press, 1994. MR1430804 (97k:60002)
7. H. Drees and C. Stărică, *A Simple non-Stationary Model for Stock Returns*, Working paper, Chalmers University of Technology, 2002.
8. R. F. Engle, *Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation*, *Econometrica* **50**(4) (1982), 987–1007. MR666121 (83j:62158)
9. R. F. Engle and J. G. Rangel, *The Spline GARCH Model for Unconditional Volatility and its Global Macroeconomic Causes*, Working paper, New York University and University of California, San Diego, 2004.
10. E. F. Fama, *Stock return, real activity, inflation and money*, *American Economic Review* **71** (1981), 545–565.
11. L. G. Godfrey, *Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables*, *Econometrica*, *Econometric Society* **46**(6) (1978), 1293–1301.
12. C. Gourieroux, J. Jasiak, and R. Sufana, *The Wishart autoregressive process of multivariate stochastic volatility*, *Journal of Econometrics* **150** (2009), 167–181. MR2535514 (2011a:62289)
13. W. H. Greene, *Econometric Analysis*, Pearson/Prentice Hall, New Jersey, 2008.
14. B. E. Hansen, *Autoregressive conditional density estimation*, *International Economic Review* **35**(3) (1994), 705–730.
15. D. A. Harville, *Matrix Algebra: Exercises and Solutions*, Springer, Berlin, 1997. MR1874239
16. D. A. Hsu, R. Miller, and D. Wichern, *On the stable Paretian behavior of stock-market prices*, *Journal of American Statistical Association* **69** (1974), 108–113.
17. C. S. Kwon and T. S. Shin, *Cointegration and causality between macroeconomic variables and stock market returns*, *Global Finance Journal* **10**(1) (1999), 71–81.

18. R. Merton, *On estimating the expected return on the market: an exploratory investigation*, Journal of Financial Economics **8** (1980), 323–361.
19. P. C. B. Phillips and K.-L. Xu, *Inference in autoregression under heteroscedasticity*, Journal of Time Series Analysis **27(2)** (2006), 289–308. MR2235847 (2007g:62092)
20. J. Polzehl and V. Spokoiny, *Varying Coefficient GARCH Versus Local Constant Volatility Modeling: Comparison of Predictive Power*, Working paper, Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany, 2006.
21. M. Rockinger and E. Jondeau, *Entropy densities with an application to autoregressive conditional skewness and kurtosis*, Journal of Econometrics **106** (2002), 119–142. MR1875530 (2003a:62024)
22. C. Stărică, *Is GARCH (1,1) as Good a Model as the Nobel Prize Accolades would Imply?*, Working paper, Chalmers University of Technology, 2003.
23. H. White, *A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity*, Econometrica **48(4)** (1980), 817–838. MR575027 (81k:62097)
24. H. White, *Nonlinear regression with dependent observations*, Econometrica **52(1)** (1984), 143–161. MR729213 (86b:62172)
25. W. H. Wong, *On the consistency of cross validation in kernel nonparametric regression*, Annals of Statistics **11** (1983), 1136–1141. MR720259 (85k:62094)
26. K.-L. Xu and P. C. B. Phillips, *Adaptive estimation of autoregressive models with time-varying variances*, Journal of Econometrics **142** (2008), 265–280. MR2408736

DEPARTMENT OF STATISTICS, EUROPEAN UNIVERSITY VIADRINA, PO Box 1786, 15207 FRANKFURT (ODER), GERMANY

*E-mail address:* bodnar@euv-frankfurt-o.de

DEPARTMENT OF STATISTICS, EUROPEAN UNIVERSITY VIADRINA, PO Box 1786, 15207 FRANKFURT (ODER), GERMANY

*E-mail address:* zabolotsky@euv-frankfurt-o.de

Received 5/OCT/2009  
Originally published in English