ADAPTIVE ESTIMATION OF VECTOR AUTOREGRESSIVE MODELS WITH TIME-VARYING VARIANCE: APPLICATION TO TESTING LINEAR CAUSALITY IN MEAN

Valentin Patilea[a] and Hamdi Raïssi[b*]

[a] IRMAR-INSA & CREST-Ensai
[b] IRMAR-INSA

First version January 2010
This version July 2010

Abstract

Linear Vector AutoRegressive (VAR) models where the innovations could be unconditionally heteroscedastic and serially dependent are considered. The volatility structure is deterministic and quite general, including breaks or trending variances as special cases. In this framework we propose Ordinary Least Squares (OLS), Generalized Least Squares (GLS) and Adaptive Least Squares (ALS) procedures. The GLS estimator requires the knowledge of the time-varying variance structure while in the ALS approach the unknown variance is estimated by kernel smoothing with the outer product of the OLS residuals vectors. Different bandwidths for the different cells of the time-varying variance matrix are also allowed. We derive the asymptotic distribution of the proposed estimators for the VAR model coefficients and compare their properties. In particular we show that the ALS estimator is asymptotically equivalent to the infeasible GLS estimator. This asymptotic equivalence is obtained uniformly with respect to the bandwidth(s) in a given range and hence justifies data-driven bandwidth rules. Using these results we build Wald tests for the linear Granger causality in mean which are adapted to VAR processes driven by errors with a non stationary volatility. It is also shown that the commonly used standard Wald test for the linear Granger causality in mean is potentially unreliable in our framework. Monte Carlo experiments illustrate the use of the different estimation approaches for the analysis of VAR models with stable innovations.

Keywords: VAR model; Heteroscedatic errors; Adaptive least squares; Ordinary least squares; Kernel smoothing; Linear causality in mean.
JEL Classification: C01; C32

*20, avenue des buttes de Coësmes, CS 70839, F-35708 Rennes Cedex 7, France. Email: valentin.patilea@insa-rennes.fr and hamdi.raissi@insa-rennes.fr

1. Introduction

In the recent years the study of linear time series models in the context of unconditionally heteroscedastic innovations has become of increased interest. This interest may be explained by the fact that numerous applied works pointed out that unconditional volatility is a common feature in economic data. For instance Doyle and Faust (2005), Ramey and Vine (2006), McConnell and Perez-Quiros (2000), Blanchard and Simon (2001) among other references, pointed out a declining volatility for many economic data since the 1980s. Sensier and van Dijk (2004) found that 80% of 214 U.S. macroeconomic time series they considered exhibit a break in volatility.

In the univariate time series case Busetti and Taylor (2003), Cavaliere (2004), Cavaliere and Taylor (2007) and Kim, Leybourne and Newbold (2002) among other references, considered the test of unit roots with non stationary volatility, while Sanso, Arago and Carrion (2004) proposed tests to detect volatility breaks in the residuals. Robinson (1987) and Hansen (1995) studied univariate linear models with a non stationary volatility. Phillips and Xu (2005) investigated the Ordinary Least Squares (OLS) estimation of univariate stable autoregressive processes. Xu and Phillips (2008) considered the same model and proposed an Adaptive Least Squares (ALS) approach which are based on nonparametric estimation of the volatility of the innovations using OLS residuals. The main conclusion of Xu and Phillips (2008) is that the ALS estimating approach could be much more effective than the OLS estimation. They also found that the asymptotic behavior of the ALS estimator does not dependent on the volatility structure. Multivariate processes are often used in econometric applications because they allow to study cross-correlations between variables. In the multivariate framework Boswijk and Zu (2007) and Cavaliere, Rahbek and Taylor (2007) studied cointegrated systems in presence of non stationary volatility.

In this paper we study the inference in linear vector autoregressive (VAR) models with volatility changes and possibly serially dependent innovations. Three methods for estimating the VAR coefficients are investigated: OLS, infeasible Generalized Least Squares (GLS) based on the knowledge of the time-varying volatility structure, and ALS which is defined like the GLS but using a kernel estimate of the volatility structure. The kernel smoothing could be used with a single bandwidth for the whole volatility matrix or with different bandwidths for different cells. In some sense, we extend the approach of Phillips and Xu (2005) and Xu and Phillips (2008) to the VAR framework. In particular, we see that in the multivariate case the asymptotic distribution of the GLS and ALS estimators is no longer free from the time-varying volatility structure. Moreover, our asymptotic results are uniform with respect to the bandwidth in a given range. This opens the door to data-driven choices of the smoothing parameter, for instance by cross-validation. Such uniformity results seems new even for the univariate case.

As an application of the new estimation methodology, we also consider the problem of test linear causality in mean. The linear causality in mean, introduced by Granger (1969), is often used to investigate causal relations between subsets of variables. For instance Sims (1972), Feige and Pearce (1979) or Stock and Watson (1989) studied the money-income causality relation. Bataa *et al.* (2009) studied the links between the inflations of different countries by testing linear causality relations. This can be explained by the fact that linear causality in mean can be easily tested by considering tests of zero restrictions on the parameters of VAR models. However, the existing test procedures for checking the linear causality in mean are based on the iid innovation assumption, while several empirical analysis contradict this setting. For instance, Bataa *et al.* (2009) underlined the presence of volatility breaks in their data set. In this paper, we use our theoretical results on the OLS and ALS estimation to propose new Wald tests for linear causality in mean adapted to the framework of non-stationary volatility. The asymptotic chi-square distribution of the new Wald type statistic obtained from the ALS approach is derived uniformly with respect to the bandwidth(s).

The structure of the paper is as follows. Section 2 outlines the heteroscedastic VAR model, introduces the assumptions and the definitions of OLS and GLS estimators. Section 3 contains the results on the asymptotic behavior of the OLS and the infeasible Generalized Least Squares estimators. We also propose an estimator for the asymptotic variance of the OLS estimator. The ALS estimator based on kernel smoothing of OLS residuals is proposed in Section 4 as a feasible asymptotically equivalent version of GLS estimator. The asymptotic equivalence between ALS and GLS estimators is proved uniformly in the bandwidths involved in volatility estimation. To prove this equivalence we use, among other technical arguments, a recent version of a uniform CLT for martingale differences arrays obtained by Bae *et al.* (2010), Bae and Choi (1999). A procedure for estimating the asymptotic variance of the ALS estimator is also provided. The application of the new inference methodologies to the test of the linear Granger causality in mean in the presence of time-varying volatility is presented in Section 5. The benefit from using our new Wald type test statistics and the failure of the classical Wald test designed for iid innovations is illustrated through an example. In section 6 the finite sample properties of the different tests considered in this paper are studied by mean of Monte Carlo experiments. The better precision of the ALS estimator when compared to the OLS estimator is also highlighted. The proofs are relegated to the appendix.

The following notations will be used throughout in the paper. We denote by $A \otimes B$ the Kronecker product of two matrices A and B, and $A \otimes A$ by $A^{\otimes 2}$. The vector obtained by stacking the columns of A is denoted $\mathrm{vec}(A)$. The symbol $\Rightarrow$ denotes the convergence in distribution and we denote by $\xrightarrow{P}$ the convergence in probability. We denote by $[u]$ the integer part of a real number u. The determinant of a square matrix A is denoted by $\det A$.

2. The model and least squares estimation of the parameters

Let us consider the observations $X_{-p+1}, \dots, X_0, X_1, \dots, X_T$ generated by the following VAR model

$$\begin{aligned} X_t &= A_1 X_{t-1} + \cdots + A_p X_{t-p} + u_t \\ u_t &= H_t \epsilon_t, \end{aligned} \tag{2.1}$$

where the X_t's are d-dimensional vectors. The stability condition on the matrices A_i, $\det A(z) \neq 0$ for all $|z| \leq 1$ with $A(z) = I_d - \sum_{i=1}^{p} A_i z^i$ and I_d denotes the $d \times d$ identity matrix, is assumed to be hold. For a random variable x we define $\| x \|_r = (E \| x \|^r)^{1/r}$, where $\| x \|$ denotes the Euclidean norm. We also define $\mathcal{F}_t$ as the σ-field generated by $\{\epsilon_s : s \leq t\}$. The following assumption on the H_t's and the process (ϵ_t) gives the framework of our paper.

Assumption A1: (i) The $d \times d$ matrices H_t are invertible and satisfy $H_{[Tr]} = G(r)$, where the components of the matrix $G(r) := \{g_{kl}(r)\}$ are measurable deterministic functions on the interval $(0, 1]$, such that $\sup_{r \in (0,1]} |g_{kl}(r)| < \infty$, and each g_{kl} satisfies a Lipschitz condition piecewise on a finite number of some sub-intervals that partition $(0, 1]$. The matrix $\Sigma(r) = G(r)G(r)'$ is assumed positive definite for all r.
(ii) The process (ϵ_t) is α-mixing and such that $E(\epsilon_t \mid \mathcal{F}_{t-1}) = 0$, $E(\epsilon_t \epsilon_t' \mid \mathcal{F}_{t-1}) = I_d$ and the components ϵ_{kt} of the process (ϵ_t) satisfy $\sup_t \parallel \epsilon_{kt} \parallel_{4\mu} < \infty$ for some $\mu > 1$ and all $k \in \{1, \dots, d\}$.

The assumption **A1** generalizes the assumption of Xu and Phillips (2008) to the multivariate case. From the assumption $E(\epsilon_t \mid \mathcal{F}_{t-1}) = 0$, the innovations are possibly serially dependent. However since $G(r)$ is deterministic and $E(\epsilon_t \epsilon_t' \mid \mathcal{F}_{t-1}) = I_d$, we do not allow the error process to follow a multivariate GARCH model. Cavaliere, Rahbek and Taylor (2007) considered similar volatility structure to ours. Their assumption is slightly different from **A1** in the sense that they do not require a Lipschitz condition and allow for a countable number of jumps. Boswijk and Zu (2007) allow the matrix H_t to be possibly stochastic, but requires the volatility process to be continuous with other additional assumptions, which in particular excludes important cases like abrupt shifts. Hafner and Herwartz (2009) assumed no structure on the volatility of the error process (u_t) and allow for conditional heteroscedasticity. Nevertheless their framework excludes the use of information on the volatility structure and could result in a loss of efficiency in the statistical inference of the model. In addition Hafner and Herwartz (2009) also assumed

$$\lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \Sigma_t = \dot{\Sigma}, \quad \text{and} \quad \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E\{(\tilde{X}_{t-1} \tilde{X}_{t-1}') \otimes (u_t u_t')\} = W,$$

where $\Sigma_t = E(u_t u_t')$, $\tilde{X}_{t-1} = (X_{t-1}', \dots, X_{t-p}')' \in \mathbb{R}^{pd}$ and W, $\dot{\Sigma}$ are positive definite matrices, and this could be viewed as too restrictive. If we suppose that the volatility matrix H_t is constant, we retrieve the standard homoscedastic case. However the assumption of standard errors is often considered to be too restrictive for macroeconomic or financial applications. Indeed many applied studies pointed out that such data may display unconditional non-stationary volatility (see e.g. Kim and Nelson (1999), Warnock and Warnock (2000) or Batbekh *et al.* (2007)). Stărică and Granger (2005) found that when large samples of stock returns are considered, taking into account shifts for the unconditional volatility instead of assuming a stationary model as a GARCH(1,1) improve the volatility forecasts.

Let us denote by $\theta_0 = (\text{vec}\,(A_1)', \dots, \text{vec}\,(A_p)')' \in \mathbb{R}^{pd^2}$ the vector of the true parameters. The equation (2.1) becomes

$$\begin{aligned} X_t &= (\tilde{X}_{t-1}' \otimes I_d)\theta_0 + u_t \\ u_t &= H_t \epsilon_t, \end{aligned}$$

where we keep the notation $\tilde{X}_{t-1} = (X_{t-1}', \dots, X_{t-p}')'$. Using this expression we first define the OLS estimator

$$\hat{\theta}_{OLS} = \hat{\Sigma}_{\tilde{X}}^{-1} \text{vec}\left(\hat{\Sigma}_X\right),$$

where

$$\hat{\Sigma}_{\tilde{X}} = T^{-1} \sum_{t=1}^{T} \tilde{X}_{t-1}\tilde{X}'_{t-1} \otimes I_d \quad \text{and} \quad \hat{\Sigma}_{X} = T^{-1} \sum_{t=1}^{T} X_t \tilde{X}'_{t-1}.$$

Next, let us define the unconditional variance $\Sigma_t := H_t H_t'$ and the Generalized Least Squares (GLS) estimator that takes into account a time-varying Σ_t, that is

$$\hat{\theta}_{GLS} = \hat{\Sigma}_{\underline{\tilde{X}}}^{-1} \text{vec}\left(\hat{\Sigma}_{\underline{X}}\right), \tag{2.2}$$

with

$$\hat{\Sigma}_{\underline{\tilde{X}}} = T^{-1} \sum_{t=1}^{T} \tilde{X}_{t-1}\tilde{X}'_{t-1} \otimes \Sigma_t^{-1} \quad \text{and} \quad \hat{\Sigma}_{\underline{X}} = T^{-1} \sum_{t=1}^{T} \Sigma_t^{-1} X_t \tilde{X}'_{t-1}.$$

Note that since H_t is assumed invertible, Σ_t is positive definite for all t. If we suppose that the volatility matrix Σ_t is constant in time, it is easy to see that $\hat{\theta}_{GLS} = \hat{\theta}_{OLS}$. However the GLS estimator is in general infeasible since the true volatility matrix appears in the expression (2.2). In the next section we compare the efficiency of the OLS and GLS estimators.

3. Asymptotic behaviour of the estimators

In order to state the first result of the paper, we need to introduce the following notations. Since we assumed that $\det A(z) \neq 0$ for all $|z| \leq 1$, it is well known that

$$X_t = \sum_{i=0}^{\infty} \psi_i u_{t-i}, \tag{3.1}$$

where $\psi_0 = I_d$ and the components of the ψ_i's are absolutely summable (see e.g. Lütkepohl (2005, pp 14-16)). From the expression (3.1) we also write

$$\tilde{X}_t = \sum_{i=0}^{\infty} \tilde{\psi}_i u_{t-i}^p,$$

u_t^p is given by $u_t^p = \mathbf{1}_p \otimes u_t$, where $\mathbf{1}_p$ is the vector of ones of dimension p and

$$\tilde{\psi}_i = \begin{pmatrix} \psi_i & 0 & 0 & 0 \\ 0 & \psi_{i-1} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \psi_{i-p+1} \end{pmatrix},$$

taking $\psi_j = 0$ for $j < 0$. Let us define by $\mathbf{1}_{p\times p}$ the $p \times p$ matrix with components equal to one. The following proposition gives the asymptotic behavior of the OLS and GLS estimators. For the sake of brevity we only investigate the asymptotic normality, the consistency is in some sense an easier matter and is hence omitted.

Proposition 3.1. *If Assumption* **A1** *holds true, then:*

1.

$$T^{\frac{1}{2}}(\hat{\theta}_{GLS} - \theta_0) \Rightarrow \mathcal{N}(0, \Lambda_1^{-1}), \tag{3.2}$$

where

$$\Lambda_1 = \int_0^1 \sum_{i=0}^{\infty} \left\{ \tilde{\psi}_i (\mathbf{1}_{p\times p} \otimes \Sigma(r)) \tilde{\psi}_i' \right\} \otimes \Sigma(r)^{-1} dr$$

is positive definite;

2. $$T^{\frac{1}{2}}(\hat{\theta}_{OLS} - \theta_0) \Rightarrow \mathcal{N}(0, \Lambda_3^{-1}\Lambda_2\Lambda_3^{-1}), \tag{3.3}$$

where

$$\Lambda_2 = \int_0^1 \sum_{i=0}^{\infty} \left\{ \tilde{\psi}_i (\mathbf{1}_{p\times p} \otimes \Sigma(r)) \tilde{\psi}_i' \right\} \otimes \Sigma(r) dr$$

and

$$\Lambda_3 = \int_0^1 \sum_{i=0}^{\infty} \left\{ \tilde{\psi}_i (\mathbf{1}_{p\times p} \otimes \Sigma(r)) \tilde{\psi}_i' \right\} \otimes I_d \, dr$$

are positive definite;

3. *The asymptotic variance of $\hat{\theta}_{GLS}$ is smaller than the asymptotic variance of $\hat{\theta}_{OLS}$, that is the matrix $\Lambda_3^{-1}\Lambda_2\Lambda_3^{-1} - \Lambda_1^{-1}$ is positive semidefinite.*

If we suppose that the error process is homoscedastic, that is $\Sigma_t = \Sigma_u$ for all t, and since we assumed $E(\epsilon_t\epsilon_t' \mid \mathcal{F}_{t-1}) = I_d$, we obtain

$$\Lambda_1 = E\left[\tilde{X}_t\tilde{X}_t'\right] \otimes \Sigma_u^{-1}, \;\; \Lambda_2 = E\left[\tilde{X}_t\tilde{X}_t'\right] \otimes \Sigma_u \text{ and } \Lambda_3 = E\left[\tilde{X}_t\tilde{X}_t'\right] \otimes I_d,$$

so that we retrieve the standard result of the iid case (see e.g. Lütkepohl (2005, p 74))

$$\Lambda_1^{-1} = \Lambda_3^{-1}\Lambda_2\Lambda_3^{-1} = \{E[\tilde{X}_t\tilde{X}_t']\}^{-1} \otimes \Sigma_u,$$

although here the error process is assumed dependent. Note that in the homoscedastic case the OLS and ALS estimator have the same efficiency.

In the univariate case ($d = 1$), $\Sigma(r)$ belongs to the real line so that Λ_1 simplifies to

$$\Lambda_1 = \sum_{i=0}^{\infty} \tilde{\psi}_i \mathbf{1}_{p\times p} \tilde{\psi}_i, \tag{3.4}$$

where the $\tilde{\psi}_i$'s are $p \times p$ diagonal matrices. This expression corresponds to the asymptotic covariance matrix obtained in equation (10) of Xu and Phillips (2008). Moreover,

$$\Lambda_2 = \int_0^1 \Sigma(r)^2 dr \sum_{i=0}^{\infty} \left\{ \tilde{\psi}_i \mathbf{1}_{p\times p} \tilde{\psi}_i \right\}, \quad \Lambda_3 = \int_0^1 \Sigma(r) dr \sum_{i=0}^{\infty} \left\{ \tilde{\psi}_i \mathbf{1}_{p\times p} \tilde{\psi}_i \right\},$$

and then we retrieve equation (5) in Xu and Phillips (2008).

A nice feature of the GLS estimator in the univariate case is that the covariance matrix of the asymptotic distribution does not depend on the volatility function $\Sigma(r)$. In the multivariate case the simplification (3.4) is still possible if $\Sigma(r) = \sigma^2(r) I_d$, with $\sigma^2(r)$ a scalar function. Nevertheless, we show in Example 3.1 below that (3.4) does not hold in the general multivariate framework and the asymptotic covariance matrix in (3.2) depends on the volatility function $\Sigma(r)$. Moreover, our example shows that the covariance matrices in (3.2) and (3.3) can be equal in some particular cases of heteroscedasticity but in general they could be very different.

Example 3.1. *Consider the bivariate model (2.1) with* $p = 1$ *and*

$$A_1 = \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix}, \quad \Sigma(r) = \begin{pmatrix} \Sigma_1(r) & 0 \\ 0 & \Sigma_2(r) \end{pmatrix}.$$

In this simple case let us compare the asymptotic variances

$$Var_{as}\left(\hat{\theta}_{2,GLS}\right) = (1 - a_1^2) \times \left(\int_0^1 \Sigma_1(r)/\Sigma_2(r)dr\right)^{-1}$$

and

$$Var_{as}\left(\hat{\theta}_{2,OLS}\right) = (1 - a_1^2) \times \left\{ \frac{\int_0^1 \Sigma_1(r)\Sigma_2(r)dr}{\left(\int_0^1 \Sigma_1(r)dr\right)^2} \right\},$$

that is the asymptotic variances of the GLS and OLS estimators of the second component of the vector $\theta_0 = (a_1, 0, 0, a_2)'$ *(which corresponds to the element* $(2, 1)$ *of the matrix* A_1*).*

First we notice that $Var_{as}\left(\hat{\theta}_{2,GLS}\right)$ *depends on the volatility structure when* $\Sigma_1(r) \neq \Sigma_2(r)$*. In order to illustrate the difference between the variances of* $\hat{\theta}_{2,OLS}$ *and* $\hat{\theta}_{2,GLS}$*, we plot the ratio*

$$Var_{as}\left(\hat{\theta}_{2,OLS}\right) / Var_{as}\left(\hat{\theta}_{2,GLS}\right) \tag{3.5}$$

in Figure 7.1 taking

$$\Sigma_1(r) = \sigma_{10}^2 + (\sigma_{11}^2 - \sigma_{10}^2) \times \mathbf{1}_{[\tau_1,1]}(r) \quad \text{and} \quad \Sigma_2(r) = \sigma_{20}^2 + (\sigma_{21}^2 - \sigma_{20}^2) \times \mathbf{1}_{[\tau_2,1]}(r),$$

where and $\tau_i \in [0, 1]$ *with* $i \in \{1, 2\}$*. This specification of the volatility function is inspired by Example 1 of Xu and Phillips (2008) (see also Cavaliere (2004)). On the left graphic we take* $\tau_1 = \tau_2$ *and* $\sigma_{10}^2 = \sigma_{20}^2 = \sigma_{11}^2 = 1$ *but* $\sigma_{21}^2 \geq 1$*, so that only* (X_{2t}) *is heteroscedastic in general. When* $\sigma_{21}^2 = 1$ *or* $\tau_1 \in \{0, 1\}$*, the process* (X_t) *is homoscedastic. On the right graphic we take* $\sigma_{10}^2 = \sigma_{20}^2 = 1$ *and* $\sigma_{11}^2 = \sigma_{21}^2 = 3$ *but* $\tau_1 \neq \tau_2$ *in general. When* $\tau_1 = \tau_2$*, we have* $\Sigma_1(r) = \Sigma_2(r)$ *and hence we retrieve the case studied in Example 1 of Xu and Phillips (2008).*

As expected the ratio (3.5) is equal to one in the homoscedastic case in the left graphic. However, departure from this case clearly shows that the difference between the variances of the two estimators is increasing with σ_{21}^2*. In the right graphic we can see that when* $\tau_2 = 0$ *or* 1 *the ratio in (3.5) is equal to one although* (X_t) *is heteroscedastic. The variances* $Var_{as}(\hat{\theta}_{2,OLS})$ *and* $Var_{as}(\hat{\theta}_{2,GLS})$ *are different when* $\tau_2 \in (0, 1)$ *and the largest relative difference is attained when we set the volatility shifts in the middle of the sample.*

It appears that the GLS estimator is more efficient than the OLS estimator in general when the matrix Σ_t is time-varying. Nevertheless the assumption of known volatility structure needed to construct the GLS estimator could be unrealistic in practice. Moreover, the asymptotic distribution of the GLS estimator depends on the unknown volatility. In the OLS estimation approach only the asymptotic distribution of the coefficients estimator depends on the unknown volatility. In addition, we can provide simple consistent estimators of Λ_2 and Λ_3, which could be further used for instance to build confidence intervals for the OLS estimators. For the purpose of estimation of Λ_2 and Λ_3 let us consider the matrices $\Omega_2 := \int_0^1 \Sigma(r)^{\otimes 2} dr, \quad \Omega_3 := \int_0^1 \Sigma(r) dr$ and denote the OLS residuals by $\hat{u}_t$.

Proposition 3.2. *Under Assumption* **A1** *we have*

$$\hat{\Omega}_2 := T^{-1} \sum_{t=2}^{T} \hat{u}_{t-1}\hat{u}_{t-1}' \otimes \hat{u}_t \hat{u}_t' = \Omega_2 + o_p(1), \tag{3.6}$$

$$\hat{\Omega}_3 := T^{-1} \sum_{t=1}^{T} \hat{u}_t \hat{u}_t' = \Omega_3 + o_p(1), \tag{3.7}$$

$$\hat{\Lambda}_2 := T^{-1} \sum_{t=1}^{T} \tilde{X}_{t-1}\tilde{X}_{t-1}' \otimes \hat{u}_t \hat{u}_t' = \Lambda_2 + o_p(1). \tag{3.8}$$

$$\hat{\Lambda}_3 := \hat{\Sigma}_{\tilde{X}} = \Lambda_3 + o_p(1), \tag{3.9}$$

Using (3.6) and (3.7) and some additional algebra, we can define alternative consistent estimators of Λ_2 and Λ_3. Indeed, it is shown in the appendix that

$$\mathrm{vec}\,(\Lambda_2) = \left\{ I_{(pd^2)^2} - (\Delta \otimes I_d)^{\otimes 2} \right\}^{-1} \mathrm{vec} \begin{pmatrix} \Omega_2 & 0_{d^2 \times (p-1)d^2} \\ 0_{(p-1)d^2 \times d^2} & 0_{(p-1)d^2 \times (p-1)d^2} \end{pmatrix} \tag{3.10}$$

and

$$\mathrm{vec}\,(\Lambda_3) = \left\{ I_{(pd^2)^2} - (\Delta \otimes I_d)^{\otimes 2} \right\}^{-1} \mathrm{vec} \begin{pmatrix} \Omega_3 \otimes I_d & 0_{d^2 \times (p-1)d^2} \\ 0_{(p-1)d^2 \times d^2} & 0_{(p-1)d^2 \times (p-1)d^2} \end{pmatrix}, \tag{3.11}$$

where $0_{d^2 \times (p-1)d^2}$ is the null matrix of dimension $d^2 \times (p-1)d^2$ and

$$\Delta = \begin{pmatrix} A_1 & \dots & A_{p-1} & A_p \\ I_d & 0 & \dots & 0 \\ & \ddots & \ddots & \vdots \\ 0 & & I_d & 0 \end{pmatrix}$$

is a matrix of dimension $pd \times pd$. Therefore replacing Ω_2 and Ω_3 by respectively $\hat{\Omega}_2$ and $\hat{\Omega}_3$, and the $A_i's$ by their OLS estimates in the expression of Δ in (3.10) and (3.11), we obtain consistent estimators of Λ_2 and Λ_3. These estimators will be denoted by $\hat{\Lambda}_{2\delta}$ and $\hat{\Lambda}_{3\delta}$, where the subscript δ refer to the use of the OLS estimator of Δ.

4. Adaptive estimation

In the previous section we pointed out that the GLS estimator is generally infeasible in applications. Therefore we consider a feasible weighted estimator obtained using nonparametric estimation of the volatility function. Our approach generalizes the work of Xu and Phillips (2008) to the multivariate case. Let us denote by $A \odot B$ the

Hadamard (entrywise) product of two matrices of same dimension A and B. Define the symmetric matrix

$$\check{\Sigma}_t^0 = \sum_{i=1}^{T} w_{ti} \odot \hat{u}_i \hat{u}_i',$$

where, as before the $\hat{u}_i$'s are the OLS residuals and the $kl-$element, $k \leq l$, of the $d \times d$ matrix of weights w_{ti} is given by

$$w_{ti}(b_{kl}) = \left(\sum_{i=1}^{T} K_{ti}(b_{kl}) \right)^{-1} K_{ti}(b_{kl}),$$

with b_{kl} the bandwidth and

$$K_{ti}(b_{kl}) = \begin{cases} K(\frac{t-i}{Tb_{kl}}) & \text{if} \quad t \neq i, \\ 0 \quad \text{if} & t = i. \end{cases}$$

The kernel function $K(z)$ is bounded nonnegative and such that $\int_{-\infty}^{\infty} K(z)dz = 1$. For all $1 \leq k \leq l \leq d$ the bandwidth b_{kl} belongs to a range $\mathcal{B}_T = [c_{min} b_T, c_{max} b_T]$ with $c_{min}, c_{max} > 0$ some constants and $b_T \downarrow 0$ at a suitable rate that will be specified below.

When using the same bandwidth $b_{kl} \in \mathcal{B}_T$ for all the cells of $\check{\Sigma}_t^0$, since $\hat{u}_i$, $i = 1, ..., T$ are almost sure linear independent each other, $\check{\Sigma}_t^0$ is almost sure positive definite provided T is sufficiently large. A similar estimator is considered by Boswijk and Zu (2007). When using several bandwidths b_{kl} it is no longer clear that the symmetric matrix $\check{\Sigma}_t^0$ is positive definite. Then we propose to use a regularization of $\check{\Sigma}_t^0$, that is to replace it by the positive definite matrix

$$\check{\Sigma}_t = \left\{ \left(\check{\Sigma}_t^0 \right)^2 + \nu_T I_d \right\}^{1/2}$$

where $\nu_T > 0$, $T \geq 1$, is a sequence of real numbers decreasing to zero at a suitable rate that will be specified below. Our simulation experience indicates that in applications with moderate and large samples ν_T could be even set equal to 0.

In practice the bandwidths b_{kl} can be chosen by minimization of a cross-validation criterion like

$$\sum_{t=1}^{T} \parallel \check{\Sigma}_t - \hat{u}_t \hat{u}_t' \parallel^2,$$

with respect to all $b_{kl} \in \mathcal{B}_T$, $1 \leq k \leq l \leq d$, where $\parallel \cdot \parallel$ is some norm for a square matrix, for instance the Frobenius norm that is the square root of the sum of the squares of matrix elements. Our theoretical results below are obtained uniformly with respect to the bandwidths $b_{kl} \in \mathcal{B}_T$ and this brings a justification for the common cross-validation bandwidth selection approach in the framework we consider. To our best knowledge, this justification is new and hence completes previous procedures of Xu and Phillips (2008) and Boswijk and Zu (2007).

Let us now introduce the following adaptive least squares (ALS) estimator

$$\hat{\theta}_{ALS} = \check{\Sigma}_{\underline{\tilde{X}}}^{-1} \text{vec} \left(\check{\Sigma}_{\underline{X}} \right),$$

with

$$\check{\Sigma}_{\underline{\tilde{X}}} = T^{-1}\sum_{t=1}^{T}\tilde{X}_{t-1}\tilde{X}_{t-1}'\otimes\check{\Sigma}_t^{-1}, \quad \text{and} \quad \check{\Sigma}_{\underline{X}} = T^{-1}\sum_{t=1}^{T}\check{\Sigma}_t^{-1}X_t\tilde{X}_{t-1}'.$$

Assumption A1': Suppose that all the conditions in Assumption **A1**(i) hold true. In addition:

(i) $\inf_{r\in(0,1]}\lambda_{min}(\Sigma(r)) > 0$ where $\lambda_{min}(\Gamma)$ denotes the smallest eigenvalue of the symmetric matrix Γ.

(ii) $\sup_t \|\epsilon_{kt}\|_8 < \infty$ for all $k \in \{1, ..., d\}$.

Assumption A2: (i) The kernel $K(\cdot)$ is a bounded density function defined on the real line such that $K(\cdot)$ is nondecreasing on $(-\infty, 0]$ and decreasing on $[0, \infty)$ and $\int_{\mathbb{R}} v^2K(v)dv < \infty$. The function $K(\cdot)$ is differentiable except a finite number of points and the derivative $K'(\cdot)$ is an integrable function. Moreover, the Fourier Transform $\mathcal{F}[K](\cdot)$ of $K(\cdot)$ satisfies $\int_{\mathbb{R}} |s\mathcal{F}[K](s)|\, ds < \infty$.

(ii) The bandwidths b_{kl}, $1 \le k \le l \le d$, are taken in the range $\mathcal{B}_T = [c_{min}b_T, c_{max}b_T]$ with $0 < c_{min} < c_{max} < \infty$ and $b_T + 1/Tb_T^{2+\gamma} \to 0$ as $T \to \infty$, for some $\gamma > 0$.

Assumption **A1**' and **A2**(ii) are natural extensions to the multivariate framework of the assumptions used in Theorem 2 of Xu and Phillips (2008). The conditions on the kernel function are convenient assumptions satisfied by almost all commonly used kernels. These conditions allow us for simpler technical arguments when investigating the rates of convergence uniformly with respect to the bandwidths. The condition on the sequence b_T, $T \ge 1$, is slightly more restrictive than the one imposed by Xu and Phillips (2008) in the univariate case, that is $b_T + 1/Tb_T^2 \to 0$, and this is the price we pay for obtaining the results uniformly in the bandwidths in a range $\mathcal{B}_T$.

Let $\Omega_1 := \int_0^1 \Sigma(r)\otimes\Sigma(r)^{-1}dr$. In the sequel, we say that a sequence of random matrices A_T, $T \ge 1$ is $o_p(1)$ uniformly with respect to (w.r.t.) $b_{kl} \in \mathcal{B}_T$ as $T \to \infty$ if $\sup_{1\le k\le l\le d}\sup_{b_{kl}\in\mathcal{B}_T}\|\text{vec }(A_T)\| \xrightarrow{P} 0$. The following proposition gives the asymptotic behavior of the adaptive estimators uniformly w.r.t the bandwidths.

Proposition 4.1. *Under* **A1'** *and* **A2** *and provided* $T\nu_T^2 \to 0$*, uniformly w.r.t.* $b_{kl} \in \mathcal{B}_T$ *as* $T \to \infty$

$$\check{\Lambda}_1 := \check{\Sigma}_{\underline{\tilde{X}}} = \Lambda_1 + o_p(1),$$

$$\check{\Omega}_1 := T^{-1}\sum_{t=1}^{T}\check{\Sigma}_t\otimes\check{\Sigma}_t^{-1} = \Omega_1 + o_p(1)$$

and

$$\sqrt{T}(\hat{\theta}_{ALS} - \hat{\theta}_{GLS}) = o_p(1).$$

Proposition 4.1 shows that the ALS and GLS estimators have the same asymptotic behavior, that is the ALS estimator is consistent in probability and $\sqrt{T}-$asymptotically normal as soon as the GLS estimator has such properties. The results remains true even if the bandwidths $b_{kl} \in \mathcal{B}_T$ are data dependent.

On the other hand, similarly to (3.10) and (3.11),

$$\mathrm{vec}\,(\Lambda_1) = \left\{ I_{(pd^2)^2} - (\Delta \otimes I_d)^{\otimes 2} \right\}^{-1} \mathrm{vec} \begin{pmatrix} \Omega_1 & 0_{d^2 \times (p-1)d^2} \\ 0_{(p-1)d^2 \times d^2} & 0_{(p-1)d^2 \times (p-1)d^2} \end{pmatrix}. \quad (4.1)$$

Then we also obtain an alternative consistent estimator (uniformly w.r.t. $b_{kl} \in \mathcal{B}_T$) $\check{\Lambda}_{1\delta}$ of Λ_1 by replacing Ω_1 by $\check{\Omega}_1$, and the $A_i's$ by their ALS estimates in the expression of Δ in (4.1).

5. Application to the test of the linear Granger causality in mean

In this section we propose tests for linear causality in mean in our framework using the OLS and the adaptive approaches. Let us consider the subvectors X_{1t} and X_{2t} such that $X_t = (X_{1t}', X_{2t}')'$ where X_{1t} is of dimension $d_1 < d$, and $d_2 = d - d_1$. It is said that (X_{2t}) does not cause linearly (X_{1t}) in mean if we have

$$EL(X_{1t} \mid X_{1t-1}, \dots) = EL(X_{1t} \mid X_{1t-1}, X_{2t-1}, \dots),$$

where $EL(X_{1t} \mid \dots)$ is the linear conditional expectation. In our framework since we assumed that (ϵ_t) is a martingale difference, the linear predictor is optimal. Therefore we have $EL(X_{1t} \mid \dots) = E(X_{1t} \mid \dots)$, where $E(X_{1t} \mid \dots)$ is the conditional expectation, and we simply refer to the linear Granger causality in mean as Granger causality in mean in the sequel. We test the null hypothesis that (X_{2t}) does not Granger cause (X_{1t}) in mean. It is well known that this amounts to test the null hypothesis that $A_{i,12} = 0$ for all $1 \leq i \leq p$ versus the alternative that there exists $i \in \{1, \dots, p\}$ such that $A_{i,12} \neq 0$, where the $A_{i,12}$'s are the matrices given by the d_1 first rows and d_2 last columns of the A_i's (see e.g. Lütkepohl (2005)). Define the block diagonal matrix $R = diag(C, \dots, C)$ of dimension $pd_1d_2 \times pd^2$, where C is a $d_1d_2 \times d^2$-dimensional matrix given by

$$C = \begin{pmatrix} 0_{d_1 \times d_1 d} & I_{d_1} & 0_{d_1 \times d_2} & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & 0 & I_{d_1} & 0_{d_1 \times d_2} \end{pmatrix}.$$

The matrix R is such that we have $R\theta_0 = r$ with r is the null vector of dimension pd_1d_2 under the null hypothesis. Therefore the tested hypotheses can be written as

$$\mathcal{H}_0 : R\theta_0 = 0 \qquad \text{vs.} \qquad \mathcal{H}_1 : R\theta_0 \neq 0.$$

In this paper we focus on the Wald type tests, because they are the most commonly used tests by the practitioners. We first consider the ALS estimator to build tests for Granger causality in mean. Let us introduce the adaptive Wald test statistics

$$Q_{ALS} = T\hat{\theta}_{ALS}' R' (R\check{\Lambda}_1^{-1} R')^{-1} R\hat{\theta}_{ALS} \quad \text{and} \quad Q_{ALS}^{\delta} = T\hat{\theta}_{ALS}' R' (R\check{\Lambda}_{1\delta}^{-1} R')^{-1} R\hat{\theta}_{ALS}.$$

The following proposition gives the asymptotic distribution of the ALS test statistics as a simple consequence of Proposition 4.1. We say that a sequence of random variables A_T, $T \geq 1$, converges in law to a chi-square distribution χ_n^2 uniformly w.r.t. $b_{kl} \in \mathcal{B}_T$ as $T \to \infty$, if there exists a sequence of random variables $\tilde{A}_T$, $T \geq 1$, independent of $b_{kl} \in \mathcal{B}_T$ such that $\tilde{A}_T \Rightarrow \chi_n^2$ and $A_T - \tilde{A}_T = o_p(1)$ uniformly w.r.t. $b_{kl} \in \mathcal{B}_T$.

Proposition 5.1. *Under the assumptions of Proposition 4.1, uniformly w.r.t.* $b_{kl} \in \mathcal{B}_T$ *as* $T \to \infty$

$$Q_{ALS} \Rightarrow \chi^2_{pd_1d_2}, \tag{5.1}$$

$$Q^{\delta}_{ALS} \Rightarrow \chi^2_{pd_1d_2} \tag{5.2}$$

and

$$Q^{\max}_{ALS} = \max\{Q_{ALS}, Q^{\delta}_{ALS}\} \Rightarrow \chi^2_{pd_1d_2}. \tag{5.3}$$

Based on Proposition 5.1 we propose the following procedure for testing Granger causality in mean: for a fixed asymptotic level α, reject the null hypothesis $\mathcal{H}_0$ if $\chi^2_{pd_1d_2,1-\alpha} < Q^{\max}_{ALS}$, where $\chi^2_{pd_1d_2,1-\alpha}$ is the $(1-\alpha)th$ quantile of the $\chi^2_{pd_1d_2}$ law. Similar procedures could be defined using Q_{ALS} or Q^{δ}_{ALS} instead of $Q^{\max}_{ALS}$, but the latter statistic is expected to yield a more powerful test. The tests based on the ALS estimation will be denoted W_{ALS}, W^{δ}_{ALS} and $W^{\max}_{ALS}$ with obvious notations.

Let us now consider the following Wald test statistics based on the OLS estimation

$$Q_{OLS} = T\hat{\theta}'_{OLS}R'(R\hat{\Lambda}_3^{-1}\hat{\Lambda}_2\hat{\Lambda}_3^{-1}R')^{-1}R\hat{\theta}_{OLS},$$

$$Q^{\delta}_{OLS} = T\hat{\theta}'_{OLS}R'(R\hat{\Lambda}_{3\delta}^{-1}\hat{\Lambda}_{2\delta}\hat{\Lambda}_{3\delta}^{-1}R')^{-1}R\hat{\theta}_{OLS},$$

and the commonly used standard Wald test statistic

$$Q_S = T\hat{\theta}'_{OLS}R'(R\hat{J}^{-1}R')^{-1}R\hat{\theta}_{OLS}, \quad \text{with} \quad \hat{J} = \left\{ T^{-1}\sum_{t=1}^{T}\tilde{X}_{t-1}\tilde{X}'_{t-1} \right\} \otimes \hat{\Omega}_3^{-1}.$$

The following proposition gives the asymptotic behavior of the OLS and standard test statistics.

Proposition 5.2. *Under* **A1** *we have as* $T \to \infty$

$$Q_{OLS} \Rightarrow \chi^2_{pd_1d_2}, \tag{5.4}$$

$$Q^{\delta}_{OLS} \Rightarrow \chi^2_{pd_1d_2}, \tag{5.5}$$

$$Q^{\max}_{OLS} = \max\{Q_{OLS}, Q^{\delta}_{OLS}\} \Rightarrow \chi^2_{pd_1d_2}, \tag{5.6}$$

and

$$Q_S \Rightarrow Z(\delta) := \sum_{i=1}^{pd_1d_2} \kappa_i Z_i^2, \tag{5.7}$$

where the Z_i*'s are independent* $\mathcal{N}(0,1)$ *variables,* $\delta = (\kappa_1, \ldots, \kappa_{pd_1d_2})'$ *is the vector of the eigenvalues of the matrix*

$$\Psi = (RJ^{-1}R')^{-\frac{1}{2}}(R\Lambda_3^{-1}\Lambda_2\Lambda_3^{-1}R')(RJ^{-1}R')^{-\frac{1}{2}}, \tag{5.8}$$

with

$$J = \int_0^1 \sum_{i=0}^{\infty} \left\{ \tilde{\psi}_i(\mathbf{1}_{p\times p} \otimes \Sigma(r))\tilde{\psi}'_i \right\} dr \otimes \Omega_3^{-1}.$$

It is easy to see from (3.7) and (3.9) that $\hat{J}$ is a consistent estimator of J. The results (5.4), (5.5) and (5.7) are direct consequences of Proposition 3.1 and 3.2, so that the proof is omitted. In the Appendix we only give the proof of (5.6). Similarly to the tests built using the ALS approach, tests using the results (5.4), (5.5) and (5.6) can be proposed.

When the errors are homoscedastic ($\Sigma_t = \Sigma_u$ for all t), we obtain $J = E[\tilde{X}_t\tilde{X}_t'] \otimes \Sigma_u^{-1}$. Recall that in this case we also have $\Lambda_3^{-1}\Lambda_2\Lambda_3^{-1} = \{E[\tilde{X}_t\tilde{X}_t']\}^{-1} \otimes \Sigma_u$, so that we obtain $\Psi = I_{pd_1d_2}$ and hence we retrieve the standard result $Q_S \Rightarrow \chi^2_{pd_1d_2}$. However the κ_i's in (5.7) can be quite different from 1 if the volatility of the errors is not constant as illustrated in the following example.

Example 5.1. *Consider the bivariate VAR(1) process $X_t = AX_{t-1} + u_t$ with true parameter $A = 0$. Such a model may be used to test Granger causality in mean between the components of an uncorrelated process. Like in Example 3.1, let us take*

$$\Sigma(r) = \begin{pmatrix} \Sigma_1(r) & 0 \\ 0 & \Sigma_2(r) \end{pmatrix}.$$

Suppose that one is interested in testing if (X_{2t}) Granger causes (X_{1t}) in mean. Then $R = (0,0,1,0)$ and the matrix Ψ is a scalar such that

$$\Psi = \left(\int_0^1 \Sigma_1(r)dr\right)^{-1} \times \left(\int_0^1 \Sigma_2(r)dr\right)^{-1} \times \int_0^1 \Sigma_1(r)\Sigma_2(r)dr.$$

As a consequence the sum in (5.7) reduces to a single term corresponding to the coefficient $\kappa_1 = \Psi$. If we suppose that the error process is homoscedastic, we obtain $\kappa_1 = 1$. However in the general heteroscedastic case we have $\kappa_1 \neq 1$. To illustrate this let us take

$$\Sigma_1(r) = \sigma_{10}^2 + (\sigma_{11}^2 - \sigma_{10}^2)r^q$$

and

$$\Sigma_2(r) = \sigma_{20}^2 + (\sigma_{21}^2 - \sigma_{20}^2)r^q,$$

as in Example 2 of Xu and Phillips (2008). The values of κ_1 are plotted in Figure 7.2 for $q = 1$, $\sigma_{10} = \sigma_{20} = 1$ and $\sigma_{11}^2, \sigma_{21}^2 \in [0.25, 16]$. It can be seen that in the heteroscedastic case κ_1 can be quite different from 1 and therefore in this case using the standard Wald procedure based on Q_S for testing if (X_{2t}) Granger cause (X_{1t}) in mean could be quite a bad idea.

The tests based on the results (5.4), (5.5) and (5.6) will be denoted W_{OLS}, W_{OLS}^{δ}, $W_{OLS}^{\max}$, and the standard test based on the statistic Q_S and the $\chi^2_{pd_1d_2}$ distribution will be denoted W_S.

6. Monte Carlo experiments

We investigated the finite sample properties of the OLS, GLS and ALS estimating approaches for VAR analysis using simulations and several real data sets. For the sake of illustration we report here a small part of the results obtained using simulated data. More details are provided in Patilea and Raïssi (2010). Consider bivariate AR(1) processes $X_t = (X_{1t}, X_{2t})'$ simulated using the model $X_t = AX_{t-1} + u_t$ with

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad \text{and} \quad u_t = H_t\epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, I_2) \quad \text{iid}, \tag{6.1}$$

and taking $a_{21} = 0.1$ in all the experiments. In this framework (X_{2t}) Granger causes (X_{1t}) in mean iff $a_{12} \neq 0$. The volatility structure considered is given by

$$\Sigma(r) = \begin{pmatrix} (1+\gamma_1 r)(1+\rho^2) & \rho(1+\gamma_1 r)^{\frac{1}{2}}(1+\gamma_2 r)^{\frac{1}{2}} \\ \rho(1+\gamma_1 r)^{\frac{1}{2}}(1+\gamma_2 r)^{\frac{1}{2}} & 1+\gamma_2 r \end{pmatrix},$$

so that the variances of the error components have a linear trending behavior. Inspired by the various real data series that we investigated, we set $\rho = 0.6$ and $\gamma_1 = 20$, $\gamma_2 = 20/3$. For the ALS approach the bandwidth is chosen by cross-validation in a given range as described in Assumption **A2**, and we take $\nu_T = 0$ in all the experiments. In the sequel the results for the GLS estimation are given only for the sake of comparison, this method being infeasible in practice. In each experiment $N = 1000$ independent trajectories are simulated using (6.1).

First we examine the properties of the estimation methods presented in the previous sections. The Root Mean Squared Error (RMSE) of the OLS, ALS and GLS methods of the autoregressive parameters is considered in Figure 7.3. In this experiment only $a_{11} = a_{22}$ vary and we set $a_{21} = 0.1$, $a_{12} = 0$. The length of the simulated series is $T = 100$. As expected the infeasible GLS estimation outperforms the other methods. However the ALS procedure clearly better estimate the autoregressive parameters when compared to the OLS estimation.

Next we study the empirical size of the Wald tests under comparison. Therefore we take $a_{12} = 0$, so that (X_{2t}) does not Granger causes (X_{1t}) in mean. We set $a_{11} = a_{22} = 0.2$. The simulated processes are of lengths $T = 50$, $T = 100$, $T = 200$ and $T = 400$. We test the null hypothesis $a_{12} = 0$ at the asymptotic nominal level 5% in Table 1.[†] Since $N = 1000$ replications are performed and assuming that the size of the tests is 5%, the relative rejection frequencies should be between 3.65% and 6.35% with probability close to 0.95. In Table 1 the relative rejection frequencies are displayed in bold type when they are outside these size bounds. We first compare the W_{OLS}, W_S, W_{ALS} and W_{GLS}. In accordance with our theoretical results, the W_S test could be far from the nominal level even for large samples. The relative rejection frequencies of the W_{OLS}, W_{ALS} and W_{GLS} tests converge to the asymptotic nominal level as the samples increase. However we note that the W_{ALS} test have better results than the W_{OLS} test for $T = 50$. It also appears that the infeasible W_{GLS} test have a better control of the error of first kind than the other tests for $T = 50$. We remark that the tests W_{OLS}^{δ}, W_{ALS}^{δ}, W_{GLS}^{δ} and $W_{OLS}^{\max}$, $W_{ALS}^{\max}$, $W_{GLS}^{\max}$ are more liberal than the other tests for small samples.

A further set of Monte Carlo experiments has been conducted to analyze the empirical power of the studied tests. For each value $a_{12} \in \{\pm 0.8, \pm 0.6, \cdots, \pm 0.2, 0\}$, a number of $N = 1000$ independent trajectories of the VAR(1) defined in (6.1) are simulated with $a_{11} = a_{22} = 0.2$. The null hypothesis $a_{12} = 0$ is tested at the asymptotic nominal level 5%. We only consider samples of length $T = 100$. The results are given in Table 2. The W_{GLS} tests appear more powerful than the other tests. It also emerges that the ALS tests are more powerful than the OLS tests in presence of unconditional heteroscedasticity. This can be explained by the fact that the ALS tests are slightly more sophisticated than the OLS tests. We note a substantial gain of power for the

[†] Several other values of the autoregressive parameters and specifications of the heteroscedasticity were experimented, the conclusions were similar but are not reported here.

$W_{OLS}^{\max}$, $W_{ALS}^{\max}$, and $W_{GLS}^{\max}$ when compared to the other tests.

The conclusion of this small simulation experiment is that when the process is non stationary but stable, the ALS estimation procedure give significant improvements in the estimation of VAR models when compared to the standard OLS estimation method. We also noted significant improvements of the ALS based tests in the analysis of the Granger causality in mean when compared to the OLS based tests in the unconditionally heteroscedastic case. Indeed from our simulation results the ALS tests have a better control of the error of first kind and a greater ability to detect the causality in mean than the OLS based tests in our framework. Finally, as expected we found that the standard Wald test is not reliable for the test of autoregressive parameter restrictions when the process is stable but not stationary.

7. Appendix: Proofs

We first state some intermediate results. Define the linear processes

$$\vartheta_t = \sum_{i=0}^{\infty} C_i u_{t-i}^k \quad \text{and} \quad \zeta_t = \sum_{i=0}^{\infty} D_i u_{t-i}^q,$$

where the components of the C_i's and D_i's are absolutely summable. The vector u_t^k is given by $u_t^k = \mathbf{1}_k \otimes u_t$, where $\mathbf{1}_k$ is the vector of ones of dimension k. Let us introduce $v_t = \mathrm{vec}\,(\vartheta_t \zeta_t')$. The following lemmas are straight extensions of the results obtained in Xu and Phillips (2008) and Phillips and Xu (2005) to the multivariate case. The proof is omitted, but could be found in Patilea and Raïssi (2010).

Lemma 7.1. *(a) If $sup_{1\leq t\leq T}(\| \epsilon_{it} \|_{2\mu}) < \infty$, $1 \leq \mu \leq \infty$, for all $i \in \{1,\dots,d\}$, then we have $sup_{1\leq t\leq T}(\| v_{nt} \|_{\mu}) < \infty$.*

(b) If $sup_{1\leq t\leq T}(\| \epsilon_{it} \|_{4\mu}) < \infty$, $1 \leq \mu \leq \infty$, for all $i \in \{1,\dots,d\}$, then we have $sup_{1\leq t\leq T}(\| \vartheta_{jt} \|_{4\mu}) < \infty$ for all $j \in \{1,\dots,kd\}$.

(c) If $sup_{1\leq t\leq T}(\| \epsilon_{it} \|_{4\mu}) < \infty$, $1 \leq \mu \leq \infty$, for all $i \in \{1,\dots,d\}$, then we have $sup_{1\leq t\leq T}(\| \vartheta_{jt-1}\vartheta_{lt-1}u_{j't}u_{l't} \|_{\mu}) < \infty$ for all $j,j',l,l' \in \{1,\dots,kd\}$.

Lemma 7.2. *Under* **A1** *we have*

$$\lim_{T\to\infty} E\left[\vartheta_{[Tr]-1}\zeta_{[Tr]-1}'\right] = \sum_{i=0}^{\infty} C_i \left\{\mathbf{1}_{k\times q} \otimes \Sigma(r)\right\} D_i', \tag{7.1}$$

for values $r \in (0,1]$ at which the functions $g_{ij}(r)$ are continuous, and where $\mathbf{1}_{k\times q}$ is the matrix of ones of dimension $k \times q$.

We introduce $y_t = \mathrm{vec}\,(\vartheta_{t-1}\vartheta_{t-1}' \otimes \Sigma_t^{-1}u_t u_t'\Sigma_t^{-1})$ and $z_t = \mathrm{vec}\,(\vartheta_{t-1}\vartheta_{t-1}' \otimes u_t u_t')$.

Lemma 7.3. *Under* **A1** *we have*

$$T^{-1}\sum_{t=1}^{T} v_t \xrightarrow{P} \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E(v_t). \tag{7.2}$$

$$T^{-1}\sum_{t=1}^{T} y_t \xrightarrow{P} \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E(y_t) = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} vec\left\{E(\vartheta_{t-1}\vartheta_{t-1}') \otimes \Sigma_t^{-1}\right\}. \quad (7.3)$$

$$T^{-1}\sum_{t=1}^{T} z_t \xrightarrow{P} \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E(z_t) = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} vec\left\{E(\vartheta_{t-1}\vartheta_{t-1}') \otimes \Sigma_t\right\}. \quad (7.4)$$

Lemma 7.4. *Under* **A1** *we have*

$$T^{-1}\sum_{t=1}^{T} \vartheta_{t-1}\vartheta_{t-1}' \otimes \Sigma_t^{-1} \xrightarrow{P} \int_0^1 \sum_{i=0}^{\infty} \{C_i(\mathbf{1}_{k\times k} \otimes \Sigma(r))C_i'\} \otimes \Sigma(r)^{-1} dr, \quad (7.5)$$

$$T^{-1}\sum_{t=1}^{T} \vartheta_{t-1}\vartheta_{t-1}' \otimes I_d \xrightarrow{P} \int_0^1 \sum_{i=0}^{\infty} \{C_i(\mathbf{1}_{k\times k} \otimes \Sigma(r))C_i'\}\, dr \otimes I_d. \quad (7.6)$$

In addition we also have

$$T^{-\frac{1}{2}}\sum_{t=1}^{T} vec\,(\Sigma_t^{-1}u_t\vartheta_{t-1}') \Rightarrow \mathcal{N}(0,\Xi_1) \quad (7.7)$$

$$T^{-\frac{1}{2}}\sum_{t=1}^{T} vec\,(u_t\vartheta_{t-1}') \Rightarrow \mathcal{N}(0,\Xi_2), \quad (7.8)$$

where

$$\Xi_1 = \int_0^1 \sum_{i=0}^{\infty} \{C_i(\mathbf{1}_{k\times k} \otimes \Sigma(r))C_i'\} \otimes \Sigma(r)^{-1} dr,$$

and

$$\Xi_2 = \int_0^1 \sum_{i=0}^{\infty} \{C_i(\mathbf{1}_{k\times k} \otimes \Sigma(r))C_i'\} \otimes \Sigma(r) dr.$$

Now we have the ingredients for proving our results.

Proof of Proposition 3.1 1) and 2). For the proof of (3.2) we write using (2.1) and (2.2)

$$T^{\frac{1}{2}}(\hat{\theta}_{GLS} - \theta_0) = \hat{\Sigma}_{\underline{\tilde{X}}}^{-1}\mathrm{vec}\,(\hat{\Sigma}_{\underline{Xu}}), \quad (7.9)$$

with

$$\hat{\Sigma}_{\underline{Xu}} = T^{-\frac{1}{2}}\sum_{t=1}^{T} \Sigma_t^{-1}u_t\tilde{X}_{t-1}'.$$

Since we have $\tilde{X}_t = \sum_{i=0}^{\infty}\tilde{\psi}_i u_{t-i}^p$, it follow from Lemma 7.4 that

$$\hat{\Sigma}_{\underline{\tilde{X}}} = \int_0^1 \sum_{i=0}^{\infty} \left\{\tilde{\psi}_i(\mathbf{1}_{p\times p} \otimes \Sigma(r))\tilde{\psi}_i'\right\} \otimes \Sigma(r)^{-1} dr + o_p(1) = \Lambda_1 + o_p(1).$$

Using (7.7) we obviously have

$$\hat{\Sigma}_{\underline{Xu}} \Rightarrow \mathcal{N}(0, \Lambda_1),$$

so that we obtain the result (3.2).

For the proof of (3.3) we write similarly to (7.9)

$$T^{\frac{1}{2}}(\hat{\theta}_{OLS} - \theta_0) = \hat{\Sigma}_{\tilde{X}}^{-1} \text{vec}\,(\hat{\Sigma}_{Xu}),$$

with

$$\hat{\Sigma}_{Xu} = T^{-\frac{1}{2}} \sum_{t=1}^{T} u_t \tilde{X}'_{t-1}.$$

From (7.6) and (7.8) we write

$$\hat{\Sigma}_{\tilde{X}} = \int_0^1 \sum_{i=0}^{\infty} \left\{ \tilde{\psi}_i (\mathbf{1}_{p\times p} \otimes \Sigma(r)) \tilde{\psi}'_i \right\} dr \otimes I_d + o_p(1) = \Lambda_3 + o_p(1), \tag{7.10}$$

and

$$\text{vec}\,(\hat{\Sigma}_{Xu}) \Rightarrow \mathcal{N}(0, \Lambda_2),$$

with

$$\Lambda_2 = \int_0^1 \sum_{i=0}^{\infty} \left\{ \tilde{\psi}_i (\mathbf{1}_{p\times p} \otimes \Sigma(r)) \tilde{\psi}'_i \right\} \otimes \Sigma(r) dr,$$

so that we obtain the result (3.3).

In this part we show that Λ_3 is positive definite. To this aim it suffices to show that the matrix $\tilde{\Lambda}_3 = \sum_{i=0}^{\infty} \tilde{\psi}_i (\mathbf{1}_{p\times p} \otimes \Sigma(r)) \tilde{\psi}'_i$ is positive definite for all r. Let us consider a pd-dimensional vector $\lambda \neq 0$. If $\tilde{\Lambda}_3$ is not positive definite we have

$$\sum_{i=0}^{\infty} \lambda' \tilde{\psi}_i (\mathbf{1}_{p\times p} \otimes \Sigma(r)) \tilde{\psi}'_i \lambda = \sum_{i=0}^{\infty} \tilde{\lambda}'_i \tilde{\lambda}_i = \sum_{i=0}^{\infty} \tilde{\lambda}_i^2 = 0,$$

where $\tilde{\lambda}'_i = (\lambda'_1 \psi_i G(r), \dots, \lambda'_p \psi_{i-p+1} G(r))$ with obvious notations. Therefore we have $\tilde{\lambda}_i = 0$ for all $i \in \mathbb{N}$. First consider $\tilde{\lambda}_0$. In this case we have $\psi_0 = I_d$ and $\psi_{-1} = \psi_{-2} = \dots = \psi_{i-p+1} = 0$. Since we assumed that $\Sigma(r)$ is positive definite we can deduce that $\lambda_1 = 0$. Similarly $\tilde{\lambda}_1$ implies that $\lambda_2 = 0$, $\tilde{\lambda}_2$ implies that $\lambda_3 = 0$ and so on. Thus $\lambda = 0$, which shows that Λ_3 is positive definite. Using similar arguments and since the Kronecker product of two positive definite matrices is positive definite, it can be shown that the matrices Λ_1 and Λ_2 are positive definite.

3) Using the Cholesky decomposition for positive semidefinite matrix we can write

$$\sum_{i=0}^{\infty} \left\{ \tilde{\psi}_i (\mathbf{1}_{p\times p} \otimes \Sigma(r)) \tilde{\psi}'_i \right\} = Z(r) Z'(r),$$

and let

$$B_k(r) = \{Z(r) \otimes \Sigma^{k-3/2}(r)\}', \quad k = 1, 2 \quad r \in (0, 1].$$

Then, by the properties of the Kronecker product we have

$$\Lambda_k = \int_0^1 B_k'(r)B_k(r)dr, \quad k = 1, 2,$$

and

$$\Lambda_3 = \int_0^1 B_2'(r)B_1(r)dr = \int_0^1 B_1'(r)B_2(r)dr.$$

Define

$$\Lambda = \left\{ \int_0^1 B_2'(r)B_2(r)dr \right\}^{-1} \int_0^1 B_2'(r)B_1(r)dr = \Lambda_2^{-1}\Lambda_3.$$

Following the idea of Lavergne (2008) we can write

$$\begin{aligned} 0 &\ll \int_0^1 \{B_1(r) - B_2(r)\Lambda\}'\{B_1(r) - B_2(r)\Lambda\}dr \\ &= \Lambda_1 - \Lambda' \int_0^1 B_2'(r)B_1(r)dr - \int_0^1 B_1'(r)B_2(r)dr\Lambda + \Lambda'\Lambda_2\Lambda \\ &= \Lambda_1 - \Lambda_3\Lambda_2^{-1}\Lambda_3 \end{aligned}$$

and this prove the stated result. Notice that the equality between the two asymptotic variance holds if and only if $B_1(r) = B_2(r)\Lambda$ for almost all $r \in (0, 1]$. □

Proof of Proposition 3.2 For the proof of (3.7) we write

$$\begin{aligned} T^{-1}\sum_{t=1}^{T} \hat{u}_t\hat{u}_t' &= T^{-1}\sum_{t=1}^{T} u_t u_t' - \left[\sum_{i=1}^{p} \left\{ T^{-1}\sum_{t=1}^{T} u_t X_{t-i}' \right\} (\hat{A}_i^{OLS} - A_i)' \right] \\ &- \left[\sum_{i=1}^{p} (\hat{A}_i^{OLS} - A_i) \left\{ T^{-1}\sum_{t=1}^{T} X_{t-i}u_t' \right\} \right] \\ &- \left[\sum_{i=1}^{p} (\hat{A}_i^{OLS} - A_i) \left\{ T^{-1}\sum_{t=1}^{T} X_{t-i}X_{t-i}' \right\} (\hat{A}_i^{OLS} - A_i)' \right], \\ &= c_1 + c_2 + c_3 + c_4, \end{aligned} \tag{7.11}$$

with obvious notations. Using similar arguments to that of the proof of (7.8), it is easy to see that we have

$$T^{-1}\sum_{t=1}^{T} u_t X_{t-i}' = O_p(T^{-\frac{1}{2}}) \quad \text{and} \quad T^{-1}\sum_{t=1}^{T} X_{t-i}' u_t = O_p(T^{-\frac{1}{2}}).$$

Then since $\hat{A}_i^{OLS} - A_i = O_p(T^{-\frac{1}{2}})$, we write $c_2 = o_p(1)$ and $c_3 = o_p(1)$. From relation (7.6), it is also easy to see that we have $c_4 = o_p(1)$. Let us define $w_t = \text{vec}(u_t u_t' - vec(G(t/T)G(t/T)')$. Since $\{w_t, \mathcal{F}_{t-1}\}$ is α-mixing by Theorem 14.1 in Davidson (1994), and $E \parallel w_t \parallel^2 < \infty$ by **A1**, we have by the law of large numbers for L^1-mixingales (Andrews (2008))

$$
\begin{aligned}
T^{-1}\sum_{t=1}^{T}\mathrm{vec}(u_tu_t') &= \lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}E\{\mathrm{vec}(u_tu_t')\}+o_p(1)\\
&= \lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}\mathrm{vec}(G(t/T)G(t/T)')+o_p(1)\\
&= \mathrm{vec}\int_0^1\Sigma(r)dr+o_p(1),
\end{aligned}
$$

and we obtain (3.7). For the proof of (3.6) we have similarly to (7.11)

$$
T^{-1}\sum_{t=2}^{T}\hat{u}_{t-1}\hat{u}_{t-1}'\otimes\hat{u}_t\hat{u}_t' = T^{-1}\sum_{t=2}^{T}u_{t-1}u_{t-1}'\otimes u_tu_t'+o_p(1).
$$

From the Cauchy-Schwartz inequality and by Assumption **A1** we have

$$
E\mid u_{it-1}u_{jt-1}u_{kt}u_{lt}\mid^{\mu}<\{E(u_{it-1})^{4\mu}E(u_{jt-1})^{4\mu}E(u_{kt})^{4\mu}E(u_{lt})^{4\mu}\}^{\frac{1}{4}}<\infty.
$$

Then using again the law of large numbers for L^1-mixingales and since

$$
E(u_{t-1}u_{t-1}'\otimes u_tu_t') = E(u_{t-1}u_{t-1}'\otimes E(u_tu_t'\mid\mathcal{F}_{t-1})) = \Sigma_{t-1}\otimes\Sigma_t,
$$

we write

$$
\begin{aligned}
T^{-1}\sum_{t=2}^{T}u_{t-1}u_{t-1}'\otimes u_tu_t' &= \lim_{T\to\infty}T^{-1}\sum_{t=2}^{T}\Sigma_{t-1}\otimes\Sigma_t+o_p(1)\\
&= \lim_{T\to\infty}T^{-1}\sum_{t=2}^{T}\Sigma(t-1/T)\otimes\Sigma(t/T)+o_p(1).
\end{aligned}
$$

Finally noting that

$$
\lim_{T\to\infty}T^{-1}\sum_{t=2}^{T}\Sigma(t-1/T)\otimes\Sigma(t/T) = \int_0^1\Sigma(r)^{\otimes 2}dr+o_p(1).
$$

we obtain (3.6). The proof of (3.9) follows from (7.10). For the proof of (3.8), we write as above

$$
T^{-1}\sum_{t=1}^{T}\tilde{X}_{t-1}\tilde{X}_{t-1}'\otimes\hat{u}_t\hat{u}_t' = T^{-1}\sum_{t=1}^{T}\tilde{X}_{t-1}\tilde{X}_{t-1}'\otimes u_tu_t'+o_p(1),
$$

so that we obtain the desired result from similar arguments used for the proof of (7.7). □

Proof of (3.10), (3.11) and (4.1) We only prove (3.11). The proofs of (3.10) and (4.1) are similar. From the proof of Theorem 3.1 we have

$$
\hat{\Sigma}_{\tilde{X}} = \Lambda_3+o_p(1). \tag{7.12}
$$

Using Lemma 7.3 we also obtain

$$\mathrm{vec}\,\{\hat{\Sigma}_{\tilde{X}}\} \xrightarrow{P} \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\mathrm{vec}\,\{E(\tilde{X}_{t-1}\tilde{X}'_{t-1})\otimes I_d\}.$$

Straightforward computations show that

$$\mathrm{vec}\,\{E(\tilde{X}_{t-1}\tilde{X}'_{t-1})\otimes I_d\} = \sum_{i=0}^{\infty}\{(\Delta\otimes I_d)^{\otimes 2}\}^i\mathrm{vec}\,\left[\begin{pmatrix}\Sigma(\frac{t-i-1}{T}) & 0\\ 0 & 0\end{pmatrix}\otimes I_d\right].$$

Then considering similar arguments used in the proof of Lemma 7.2 we write

$$\begin{aligned}&\lim_{T\to\infty}\mathrm{vec}\,\{E(\tilde{X}_{[Tr]-1}\tilde{X}'_{[Tr]-1})\otimes I_d\}\\ =\ &\left\{I_{(pd^2)^2}-(\Delta\otimes I_d)^{\otimes 2}\right\}^{-1}\mathrm{vec}\,\left[\begin{pmatrix}\Sigma(r)\otimes I_d & 0\\ 0 & 0\end{pmatrix}\right],\end{aligned}$$

so that we obtain

$$\mathrm{vec}\,\{\hat{\Sigma}_{\tilde{X}}\} = \left\{I_{(pd^2)^2}-(\Delta\otimes I_d)^{\otimes 2}\right\}^{-1}\mathrm{vec}\,\left[\begin{pmatrix}\int_0^1\Sigma(r)dr\otimes I_d & 0\\ 0 & 0\end{pmatrix}\right]+o_p(1).$$

by using similar arguments of the proof of (7.5). Therefore the result (3.11) follow from (7.12). □

Proof of Proposition 4.1 In the following, c, C, ... denote constants with possibly different values from line to line. First, let us focus on the asymptotic equivalence between $\hat{\theta}_{ALS}$ and $\hat{\theta}_{GLS}$ uniformly w.r.t. the bandwidths $b_{kl}\in\mathcal{B}_T$. We extend the arguments of Theorem 2 in Xu and Phillips (2008). Consider the notation

$$A(\Gamma)=T^{-1}\sum_{t=1}^{T}\tilde{X}_{t-1}\tilde{X}'_{t-1}\otimes\Gamma_t^{-1},\quad\text{and}\quad a(\Gamma)=T^{-1/2}\sum_{t=1}^{T}\Gamma_t^{-1}u_t\tilde{X}'_{t-1}.$$

Then

$$\begin{aligned}\sqrt{T}(\hat{\theta}_{ALS}-\hat{\theta}_{GLS}) &= A(\check{\Sigma})^{-1}\mathrm{vec}\,\left(a(\check{\Sigma})\right)-A(\Sigma)^{-1}\mathrm{vec}\,\left(a(\Sigma)\right)\\ &= A(\check{\Sigma})^{-1}\left\{\mathrm{vec}\,\left(a(\check{\Sigma})\right)-\mathrm{vec}\,\left(a(\Sigma)\right)\right\}\\ &\quad -A(\Sigma)^{-1}\left\{A(\check{\Sigma})-A(\Sigma)\right\}A(\check{\Sigma})^{-1}\mathrm{vec}\,\left(a(\Sigma)\right).\end{aligned}$$

By our result (7.5), $A(\Sigma)\xrightarrow{P}\Lambda_1$ which is positive definite. Moreover, $a(\Sigma)$ is bounded in probability by Markov's inequality, Lemma 7.1-a) considered with $\mu\geq 2$ and the linear processes $\vartheta_t=u_t$ and $\zeta_t=\tilde{X}_{t-1}$, and the fact that Σ_t^{-1} is bounded. Hence, like in the proof of Theorem 2 of Xu and Phillips (2008), to prove that $\sqrt{T}(\hat{\theta}_{ALS}-\hat{\theta}_{GLS})=o_p(1)$, uniformly w.r.t. $b_{kl}\in\mathcal{B}_T$, it suffices to check

$$A(\check{\Sigma})-A(\Sigma)=o_p(1)\quad\text{and}\quad a(\check{\Sigma})-a(\Sigma)=o_p(1),\tag{7.13}$$

uniformly w.r.t. $b_{kl}\in\mathcal{B}_T$. As a direct by-product we also obtain $\check{\Lambda}_1-\Lambda_1=o_p(1)$ uniformly w.r.t. the bandwidths b_{kl}. Let us define

$$\mathring{\Sigma}_t=\sum_{i=1}^{T}w_{ti}\odot u_iu_i'\quad\text{and}\quad\bar{\Sigma}_t=\sum_{i=1}^{T}w_{ti}\odot\Sigma_i,\tag{7.14}$$

and, following Xu and Phillips (see also Robinson, 1987), notice that the results in (7.13) are consequences of the following eight rates obtained uniformly w.r.t. $b_{kl} \in \mathcal{B}_T$: (a) $a(\check{\Sigma}^0) - a(\overset{\circ}{\Sigma}) = o_p(1)$; (a') $a(\check{\Sigma}) - a(\check{\Sigma}^0) = o_p(1)$; (b) $a(\overset{\circ}{\Sigma}) - a(\bar{\Sigma}) = o_p(1)$; (c) $a(\bar{\Sigma}) - a(\Sigma) = o_p(1)$; (d) $A(\check{\Sigma}^0) - A(\overset{\circ}{\Sigma}) = o_p(1)$; (d') $A(\check{\Sigma}) - A(\check{\Sigma}^0) = o_p(1)$; (e) $A\left(\overset{\circ}{\Sigma}\right) - A(\bar{\Sigma}) = o_p(1)$; (f) $A(\bar{\Sigma}) - A(\Sigma) = o_p(1)$. In this proof the norm $\|\cdot\|$ is the Frobenius norm which in particular is a sub-multiplicative norm, that is $\|AB\| \leq \|A\|\|B\|$, and for a positive definite matrix A, $\|A\| \leq C[\lambda_{min}(A)]^{-1}$ with C a constant depending only on the dimension of A. Moreover, $\|A \otimes B\| = \|A\|\|B\|$. To simplify notation, let b denote the $d(d+1)$ vector of bandwidths b_{kl}, $1 \leq k \leq l \leq d$. Below we will simply write *uniformly w.r.t.* b instead of *uniformly w.r.t.* b_{kl}, $1 \leq k \leq l \leq d$, and $\sup_b$ instead of $\sup_{b_{kl} \in \mathcal{B}_T, 1 \leq k \leq l \leq d}$.

(a) Using the identity $A^{-1} - B^{-1} = A^{-1}(B-A)B^{-1}$ we can write

$$a(\check{\Sigma}^0) - a(\overset{\circ}{\Sigma}) = T^{-1/2} \sum_{t=1}^{T} \left(\check{\Sigma}_t^0\right)^{-1} \left\{ \overset{\circ}{\Sigma}_t - \check{\Sigma}_t^0 \right\} \overset{\circ}{\Sigma}_t^{-1} u_t \tilde{X}'_{t-1}.$$

Take the norm on the right-hand side and apply Lemma 7.6(f,h,i), Cauchy-Schwarz inequality and the fact that $T^{-1} \sum_{t=1}^{T} \|u_t \tilde{X}'_{t-1}\|^2 = O_p(1)$ by Lemma 7.1-a).

(a') Use the same decomposition to write

$$a(\check{\Sigma}^0) - a(\check{\Sigma}) = T^{-1/2} \sum_{t=1}^{T} \left(\check{\Sigma}_t^0\right)^{-1} \left\{ \check{\Sigma}_t - \check{\Sigma}_t^0 \right\} \check{\Sigma}_t^{-1} u_t \tilde{X}'_{t-1}. \tag{7.15}$$

Now, if $\|\cdot\|_2$ denotes the spectral norm, use the inequality

$$\|B^{1/2} - A^{1/2}\|_2 \leq \frac{1}{2} \left[\max\{\|A^{-1}\|_2, \|B^{-1}\|_2\}\right]^{1/2} \|B - A\|_2$$

(see for instance Horn and Johnson (1994), page 557), and deduce that

$$\|\check{\Sigma}_t^0 - \check{\Sigma}_t\|_2 \leq \frac{\nu_T}{2} \left[\max \left\{ \left\| \left[\left(\check{\Sigma}_t^0\right)^2 + \nu_T I_d \right]^{-1} \right\|_2, \left\| \left[\left(\check{\Sigma}_t^0\right)^2 \right]^{-1} \right\|_2 \right\} \right]^{1/2}.$$

Now, if $r \in (0,1]$ and $A_{[Tr]} - B = o_p(1)$ with B positive definite, it is easy to check that $\|A_{[Tr]}^{-1}\|_2 \leq \{1 + o_p(1)\}\|B^{-1}\|_2$. Use Lemma 7.5 and Assumption **A1'**(i) to deduce that the spectral norms of $[\left(\check{\Sigma}_t^0\right)^2 + \nu_T I_d]^{-1}$ and $[\left(\check{\Sigma}_t^0\right)^2]^{-1}$ are bounded in probability. Finally, take spectral norm on the right-hand side of (7.15), use the fact that $\nu_T = o(T^{-1/2})$ and deduce (a').

(b) Consider the identity

$$A^{-1} - B^{-1} = B^{-1}(B-A)B^{-1} + B^{-1}(B-A)A^{-1}(B-A)B^{-1}$$

and write

$$\begin{aligned}
a(\overset{\circ}{\Sigma}) - a(\bar{\Sigma}) &= T^{-1/2}\sum_{t=1}^{T}[\overset{\circ}{\Sigma}_t^{-1} - \bar{\Sigma}_t^{-1}]u_t\tilde{X}_{t-1}' \\
&= T^{-1/2}\sum_{t=1}^{T}\bar{\Sigma}_t^{-1}[\bar{\Sigma}_t - \overset{\circ}{\Sigma}_t]\bar{\Sigma}_t^{-1}u_t\tilde{X}_{t-1}' \\
&\quad + T^{-1/2}\sum_{t=1}^{T}\bar{\Sigma}_t^{-1}[\bar{\Sigma}_t - \overset{\circ}{\Sigma}_t]\overset{\circ}{\Sigma}_t^{-1}[\bar{\Sigma}_t - \overset{\circ}{\Sigma}_t]\bar{\Sigma}_t^{-1}u_t\tilde{X}_{t-1}' \\
&=: T^{-1/2}\sum_{t=1}^{T}\Delta_{1t}(b) + T^{-1/2}\sum_{t=1}^{T}\Delta_{2t}(b) \\
&=: \Delta_1(b) + \Delta_2(b).
\end{aligned}$$

Note that by equation (22) in Xu and Phillips (2008),

$$\{\Delta_{1t}(b), \mathcal{F}_t\} = \{\bar{\Sigma}_t^{-1}[\bar{\Sigma}_t - \overset{\circ}{\Sigma}_t]\bar{\Sigma}_t^{-1}u_t\tilde{X}_{t-1}', \mathcal{F}_t\}$$

is a martingale difference (m.d.) sequence indexed by the bandwidths b.[‡] To prove that $\Delta_1(b) = o_p(1)$ uniformly w.r.t. b we show that this uniform rate holds cellwise. For this purpose it easy to see that it suffices to prove that

$$S_T(h) = \frac{1}{\sqrt{T}}\sum_{i,t=1}^{T}\varepsilon_t\omega_i w_{ti}(h) = o_p(1)$$

uniformly w.r.t. $h \in \mathcal{B}_T$ where $\{\varepsilon_t, \mathcal{F}_t\}$ and $\{\omega_t, \mathcal{F}_t\}$ are univariate m.d. sequence satisfying suitable moment conditions and

$$\sup_{t\geq 1} E\{\varepsilon_t^2 + \omega_t^2 \mid \mathcal{F}_{t-1}\} < \infty. \tag{7.16}$$

More precisely, $\varepsilon_t\omega_i$ could be any cell of

$$\bar{\Sigma}_t^{-1}[\Sigma_i - u_i u_i']\bar{\Sigma}_t^{-1}u_t\tilde{X}_{t-1}'.$$

[‡]It is important to notice that for a fixed bandwidth the sequence $(\Delta_{1t}(b))$ is not adapted to the filtration $(\mathcal{F}_t)$. As a consequence, the expectation $E\{\Delta_{1t}(b)'\Delta_{1s}(b)\}$ is not necessarily zero and therefore the equality $E\{\|\Delta_1(b)\|^2\} = T^{-1}\sum_{t=1}^{T} E\{\|\Delta_{1t}(b)\|^2\}$ does not necessarily holds.

Using the Inverse Fourier Transform and a change of variables, we rewrite

$$\begin{aligned}
S_T(h) &= \frac{1}{T\sqrt{T}\,h}\sum_{t,i=1}^{T}\varepsilon_t\omega_i\widehat{f}_T^{-1}(t/T;h)K((t-i)/Th)-\Delta_T(h)\\
&= \frac{1}{T\sqrt{T}\,h}\int_{\mathbb{R}}\sum_{t,i=1}^{T}\varepsilon_t\omega_i\widehat{f}_T^{-1}(t/T;h)\exp\!\Big(2\pi\sqrt{-1}\,\frac{t-i}{Th}\,u\Big)\mathcal{F}[K](u)du-\Delta_T(h)\\
&\overset{u=sh}{=} \frac{1}{\sqrt{T}}\int_{\mathbb{R}}\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\varepsilon_t(1+|s|)^{-\tau}}{\widehat{f}_T(t/T;h)}\exp\left(2\pi\sqrt{-1}\,\frac{t}{T}\,s\right)\right\}\\
&\quad\times\left\{\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\omega_i(1+|s|)^{-\tau}\exp\left(-2\pi\sqrt{-1}\,\frac{i}{T}\,s\right)\right\}\frac{\mathcal{F}[K](sh)}{(1+|s|)^{-2\tau}}ds-\Delta_T(h)\\
&=: \frac{1}{\sqrt{T}}\int_{\mathbb{R}}\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widehat{S}_{1t}(h,s)\right\}\left\{\frac{1}{\sqrt{T}}\sum_{i=1}^{T}S_{2i}(s)\right\}\frac{\mathcal{F}[K](sh)}{(1+|s|)^{-2\tau}}\,ds-\Delta_T(h)
\end{aligned}$$

where

$$\Delta_T(h)=\frac{K(0)}{T\sqrt{T}\,h}\sum_{t=1}^{T}\varepsilon_t\omega_t\widehat{f}_T^{-1}(t/T;h),$$

$\widehat{f}_T(t/T;h) = T^{-1}h^{-1}\sum_{j=1}^{T}K((t-j)/Th)$ and τ is any (arbitrary small) positive constant. It is easy to see that Assumption **A1**'(i) and inequality (7.22) imply

$$\sup_{h\in\mathcal{B}_T}|\Delta_T(h)|=o_p(T^{-1/2}b_T^{-1})=o_p(1).$$

Since

$$\frac{1}{\sqrt{T}}\int_{\mathbb{R}}\frac{|\mathcal{F}[K](sh)|}{(1+|s|)^{-2\tau}}\,ds\le C\frac{1}{\sqrt{T}b_T^{1+2\tau}}\int_{\mathbb{R}}|s\mathcal{F}[K](s)|\,ds=O\left(\left(Tb_T^{2+\gamma}\right)^{-1/2}\right)$$

provided τ is sufficiently small, it suffices to prove

$$\sup_{h\in\mathcal{B}_T}\sup_{s\in\mathbb{R}}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widehat{S}_{1t}(h,s)\right|=o_p(1) \tag{7.17}$$

and

$$\sup_{s\in\mathbb{R}}\left|\frac{1}{\sqrt{T}}\sum_{i=1}^{T}S_{2i}(s)\right|=o_p(1) \tag{7.18}$$

Let us notice that

$$\widehat{f}_T(t/T;h)=h^{-1}\sum_{j=1}^{T}\int_{\frac{t-j}{T}}^{\frac{t-j+1}{T}}K\left(\frac{[Tr]}{Th}\right)dr\overset{z=r/h}{=}\int_{\frac{t-T}{Th}}^{\frac{t}{Th}}K\left(\frac{[Tzh]}{Th}\right)dz.$$

For $0<c_{max}^{-1}\le\vartheta\le c_{min}^{-1}$ define

$$f_T(t/T;b_T/\vartheta)=\int_{\frac{(t-T)\vartheta}{Tb_T}}^{\frac{t\vartheta}{Tb_T}}K(z)\,dz.$$

Then, if $K(\cdot)$ is differentiable, for any $1 \leq t \leq T$ and $h \in \mathcal{B}_T$

$$\begin{aligned}
\left|\widehat{f}_T(t/T;h) - f_T(t/T;h)\right| &\leq \int_{-\infty}^{\infty} \left| K(z) - K\left(\frac{[Tzh]}{Th}\right)\right| dz \\
&\leq \int_{-\infty}^{\infty} \int_{\frac{[Tzh]}{Th}}^{z} |K'(v)|\, dv dz \\
&\leq \int_{-\infty}^{\infty} \int_{z-\frac{1}{Th}}^{z} |K'(v)|\, dv dz \\
&= \int_{-\infty}^{\infty} |K'(v)| \int_{v}^{v+\frac{1}{Th}} dz dv \leq \frac{C}{Tb_T}.
\end{aligned}$$

When the $K(\cdot)$ is differentiable except a finite number of points, the same type of upper bound can be derived after minor and obvious changes. Hence, with the notation

$$S_{1t}(\vartheta, s) = \frac{\varepsilon_t (1+|s|)^{-\tau}}{f_T(t/T; b_T/\vartheta)} \exp\left(2\pi\sqrt{-1}\,\frac{t}{T}\, s\right), \quad 0 < c_{max}^{-1} \leq \vartheta \leq c_{min}^{-1},\ s \in \mathbb{R},$$

since for any real numbers a and b, $a^{-1} = b^{-1} + (b-a)/ab$ and knowing that $\widehat{f}_T(\cdot; b_T/\vartheta)$ and $f_T(\cdot; b_T/\vartheta)$ are uniformly bounded away from zero (see inequality (7.22) below), we obtain

$$\sup_{\vartheta} \sup_{s} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left\{ \widehat{S}_{1t}(\vartheta, s) - S_{1t}(b_T/\vartheta, s) \right\} \right| \leq \frac{C}{\sqrt{T}\, b_T} \frac{1}{T} \sum_{t=1}^{T} |\varepsilon_t| = O_p(T^{-1/2} b_T^{-1}) = o_p(1).$$

Therefore is suffices to prove

$$\sup_{c_{max}^{-1} \leq \vartheta \leq c_{min}^{-1}} \sup_{s \in \mathbb{R}} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} S_{1t}(\vartheta, s) \right| = o_p(1) \tag{7.19}$$

in place of (7.17). For proving (7.18) and (7.19) we use a uniform CLT for m.d. indexed by a class of functions, see Bae *et al.* (2010), Bae and Choi (1999). Here our indexing classes functions depend on $c_{max}^{-1} \leq \vartheta \leq c_{min}^{-1}$ and $s \in \mathbb{R}$, respectively on $s \in \mathbb{R}$, and we prove that their covering numbers are of polynomial order independent of T. Now we can explain the unique role of the $(1+|s|)^{\tau}$ function: it cuts the high frequencies of the complex exponential function and allows one to obtain polynomial covering numbers. Consider the family of functions $\mathcal{F}_{11} = \{\varphi_{11}(\cdot;\vartheta) : [0,1] \to [0,1] : c_{max}^{-1} \leq \vartheta \leq c_{min}^{-1}\}$ and $\mathcal{F}_{12} = \{\varphi_{12}(\cdot; s) : [0,1] \to \mathbb{C} : s \in \mathbb{R}\}$ where

$$\varphi_{11}(r;\vartheta) = f_T(r; b_T/\vartheta) = F_K(\vartheta r/b_T) - F_K(\vartheta(r-1)/b_T), \quad \varphi_{12}(r;s) = \frac{\exp\left(2\pi\sqrt{-1}\, rs\right)}{(1+|s|)^{\tau}},$$

where $F_K(\cdot)$ is the cumulative distribution function associated to the density $K(\cdot)$. By Lemma 22-ii) and Lemma 16 of Nolan and Pollard (1987), the class $\mathcal{F}_{11}$ is a $VC-$class (also called Euclidean) for a constant envelope. The $VC-$property for $\mathcal{F}_{12}$ is proved for instance in Lopez and Patilea (2010).

Now we check the conditions of Theorem 1 of Bae *et al.* (2010) in order to derive (7.18) and (7.19). For simplicity, we only provide the details for (7.19). With the notation of Bae *et al.* (2010), $j = t$, $n = j(n) = T$, $\mathcal{E}_{Tt} = \mathcal{F}_t$, $\mathbf{X} = \mathbb{R} \times [0,1]$, $V_{Tt}(f) = T^{-1/2}\varepsilon_t\varphi_{12}(t/T;s)\varphi_{11}(t/T;\vartheta)^{-1}$ with $(\vartheta, s) \in \mathcal{T} = [c_{max}^{-1}, c_{min}^{-1}] \times \mathbb{R}$, the family $\mathcal{F}$ being composed of the functions $f(\varepsilon, r) = \varepsilon\varphi_{12}(r;s)\varphi_{11}(r;\vartheta)^{-1}$, $(\vartheta, s) \in \mathcal{T}$, with envelope $F(\varepsilon, r) = C\varepsilon$ for some sufficiently large constant C. Moreover, define

$$\{d_{\mu_n}^{(2)}(f,g)\}^2 = d^2((\vartheta_1, s_1), (\vartheta_2, s_2)) = \int_0^1 E\left\{\varepsilon_{[Tr]}^2\left(\frac{\varphi_{12}(r;s_1)}{\varphi_{11}(r;\vartheta_1)} - \frac{\varphi_{12}(r;s_2)}{\varphi_{11}(r;\vartheta_2)}\right)^2\right\}dr$$

Notice that $\sup_{0<r\leq 1} E(\varepsilon_{[Tr]}^2) < \tilde{C}$ for some constant $\tilde{C}$, $\varphi_{11}(\cdot;\vartheta)$ is uniformly bounded and bounded away from zero (see (7.22)), $\varphi_{12}(\cdot;s)$ is uniformly bounded, and

$$\begin{aligned}\int_0^1 &\left(\frac{\varphi_{12}(r;s_1)}{\varphi_{11}(r;\vartheta_1)} - \frac{\varphi_{12}(r;s_2)}{\varphi_{11}(r;\vartheta_2)}\right)^2 dr \\ &\leq C_1(\vartheta_1 - \vartheta_2)^2 \int_0^1 \{K(r/c_{max}b_T)r/b_T\}^2 + \{K((r-1)/c_{max}b_T)(r-1)/b_T\}^2 dr \\ &\quad + C_2 \int_0^1 \{\varphi_{12}(r;s_1) - \varphi_{12}(r;s_2)\}^2 dr \\ &\leq C_3\{b_T(\vartheta_1 - \vartheta_2)^2 + (s_1 - s_2)^2\},\end{aligned}$$

$\forall (\vartheta_1, s_1), (\vartheta_2, s_2) \in \mathcal{T}$, for some constants $C_1, \cdots, C_3 > 0$. Moreover, for any $\rho > 0$ there exists $c_\rho > 0$ such that $\int_0^1 \varphi_{12}^2(r;s)dr \leq \rho$, $\forall s \geq c_\rho$. Using these properties, on one hand we check that the pseudometric space $(\mathcal{T}, d(\cdot,\cdot))$ is totally bounded and, on the other hand, we check that condition (2) of Bae *et al.* (2010) holds for some sufficiently large L given that the conditional variance of ε_t is deterministic and bounded. The convergence to zero for $L_n(\delta)$ in Bae *et al.* (2010) is a direct consequence of our unconditional moment conditions on ε_t. The uniformly integrable entropy condition is ensured by the $VC-$property satisfied by the classes $\mathcal{F}_{11}$ and $\mathcal{F}_{12}$ and the finite second order moment for ε_t. Now, all the required ingredients are gathered to apply Theorem 1 of Bae *et al.* (2010) and to deduce our property (7.19).

For the uniform $o_p(1)$ rate of Δ_2 take the norms and apply Lemma 7.6(c,e,f) and the moment assumptions.

The results (c) to (f) are obtained by obvious adaptation of the corresponding proofs in Xu and Phillips (2008), hence the details are omitted.

Finally, to derive the result for $\check{\Omega}_1$, use again the identity $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$, the inequality $\|A \otimes B\| = \|A\|\|B\|$, the triangle inequality and apply Lemma 7.6(e,g,h,j). □

Lemma 7.5. *Let* $g_{kl}(r-) = \lim_{\tilde{r}\uparrow r} g_{kl}(\tilde{r})$ *and* $g_{kl}(r+) = \lim_{\tilde{r}\downarrow r} g_{kl}(\tilde{r})$, *for* $r \in (0,1]$ *and* $1 \leq k, l, \leq d$. *Define the* $d \times d-$*matrices* $G(r-) = \{g_{kl}(r-)\}$ *and* $G(r+) = \{g_{kl}(r+)\}$ *and* $\Sigma(r-) = G(r-)G(r-)'$, $\Sigma(r+) = G(r+)G(r+)'$. *Set* $\Sigma(1+) = 0$. *Under the assumptions of Proposition 4.1,*

$$\check{\Sigma}_{[Tr]}^0 \xrightarrow{P} \Sigma(r-)\int_{-\infty}^0 K(z)dz + \Sigma(r+)\int_0^\infty K(z)dz,$$

uniformly with respect to $r \in (0,1]$.

Proof of Lemma 7.5 It suffices to notice that equation (19) of Xu and Phillips (2008) can be obtained uniformly w.r.t. $r \in (0,1]$ and $b_{kl} \in \mathcal{B}_T$, $1 \le k \le l \le d$ and to prove

$$\sup_{b_{kl} \in \mathcal{B}_T, 1 \le k \le l \le d} \sup_{r \in (0,1]} \{\|A_{[Tr]}\| + \|B_{[Tr]}\|\} = o_p(1),$$

for $A_{[Tr]} = \check{\Sigma}^0_{[Tr]} - \overset{\circ}{\Sigma}_{[Tr]}$ and $B_{[Tr]} = \overset{\circ}{\Sigma}_{[Tr]} - \bar{\Sigma}_{[Tr]}$ where $\overset{\circ}{\Sigma}_t$ and $\bar{\Sigma}_t$ are defined in equation (7.14). The uniform convergence of $A_{[Tr]}$ and $B_{[Tr]}$ is easily obtained from Lemma 7.6(d,g). □

In the following lemma, which is an extension of the statements (d) to (l) in Lemma A of Xu and Phillips (2008), we gather some results used in the proof of Proposition 4.1. Let $w_{ti,kl} = w_{ti}(b_{kl})$ denote the element kl of the $d \times d$ matrix w_{ti} that is a function of the bandwidth b_{kl}.

Lemma 7.6. *Let $\|\cdot\|$ denote the Frobenius norm. Under the assumptions of Proposition 4.1:*

(a) For all $T \ge 1$ and $1 \le k \le l \le d$,

$$\max_{1 \le t \le T} \sum_{i=1}^{T} \sup_{b_{kl} \in \mathcal{B}_T} w_{ti,kl} \le C < \infty,$$

for some constant C.

(b) For all $T \ge 1$ and $1 \le k \le l \le d$, $\max_{1 \le t,i \le T} \sup_{b_{kl} \in \mathcal{B}_T} w_{ti,kl} \le C/Tb_T$ for some constant $C > 0$.

(c) For all T, $\inf_{b \in \mathcal{B}_T} \min_{1 \le t \le T} \lambda_{min}(\bar{\Sigma}_t) \ge C > 0$ for some constant C.

(d) As $T \to \infty$,

$$\max_{1 \le t \le T} E\left(\sup_{b \in \mathcal{B}_T} \|\overset{\circ}{\Sigma}_t - \bar{\Sigma}_t\|^4\right) = O_p(\left(1/(Tb_T)^2\right). \tag{7.20}$$

(e) For $\delta = 1, 2$, as $T \to \infty$,

$$\max_{1 \le t \le T} \sup_{b \in \mathcal{B}_T} \|\overset{\circ}{\Sigma}_t - \bar{\Sigma}_t\|^\delta = O_p(T^{-\delta/4} b_T^{-\delta/2}).$$

(f) As $T \to \infty$,

$$\left(\inf_{b \in \mathcal{B}_T} \min_{1 \le t \le T} \lambda_{min}(\overset{\circ}{\Sigma}_t)\right)^{-1} = O_p(1).$$

(g) As $T \to \infty$,

$$\max_{1 \le t \le T} \sup_{b \in \mathcal{B}_T} \|\check{\Sigma}^0_t - \overset{\circ}{\Sigma}_t\| = O_p(T^{-1/2} b_T^{-1/2}).$$

(h) As $T \to \infty$,

$$\left(\inf_{b \in \mathcal{B}_T} \min_{1 \le t \le T} \lambda_{min}(\check{\Sigma}^0_t)\right)^{-1} = O_p(1).$$

(i) As $T \to \infty$, $\sup_{b \in \mathcal{B}_T} \sum_{i=1}^{T} \|\check{\Sigma}^0_t - \overset{\circ}{\Sigma}_t\|^2 = O_p(T^{-2} b_T^{-2})$

(j) As $T \to \infty$, $\sup_{b \in \mathcal{B}_T} T^{-1} \sum_{t=1}^{T} \|\bar{\Sigma}_t - \Sigma_t\| = o(1)$.

Proof of Lemma 7.6 (a) Using the monotonicity of $K(\cdot)$ we can write

$$\max_{1\leq t\leq T}\sum_{i=1}^{T}\sup_{b_{kl}\in\mathcal{B}_T} w_{ti,kl} \leq \max_{1\leq t\leq T}\frac{\sum_{i=1}^{T}K((t-i)/b_T c_{max})}{\sum_{i=1}^{T}K((t-i)/b_T c_{min})-K(0)}. \tag{7.21}$$

Now, using again the monotonicity of $K(\cdot)$ and adapting the lines of Lemma A(c) in Xu and Phillips (2008), for any $h\in\mathcal{B}_T$ and any $1\leq t\leq T$,

$$\frac{1}{Th}\sum_{i=1}^{T}K\Big(\frac{t-i}{Th}\Big)=\sum_{i=1}^{T}\int_{\frac{t-i}{Th}}^{\frac{t-i+1}{Th}}K\left(\frac{[Thz]}{Th}\right)dz\leq\int_{-\infty}^{\infty}\max\Big[K\left(z\right),K\Big(z-\frac{1}{Th}\Big)\Big]\,dz\leq 2.$$

This allows to control the numerator on the right-hand side of (7.21). On the other hand, using similar arguments and the fact that $K(0)>0$, for any $h\in\mathcal{B}_T$, any $1\leq t\leq T$ and any $0<\gamma_1<\gamma_2<\infty$,

$$\begin{aligned}\frac{1}{Th}\sum_{i=1}^{T}K\left(\frac{t-i}{Th}\right) &\geq \int_{\frac{t-T}{Th}}^{\frac{t}{Th}}\min\Big[K\left(z\right),K\Big(z-\frac{1}{Th}\Big)\Big]\,dz\\ &\geq \min\left[\int_{-\gamma_2}^{-\gamma_1}K(z)dz,\int_{\gamma_1}^{\gamma_2}K(z)dz\right],\end{aligned}\tag{7.22}$$

provided that T is sufficiently large. The last two integrals in the minimum are strictly positive for a suitable choice of γ_1,γ_2. This fixed lower bound considered for $h=c_{min}b_T$ allows to control the denominator on the right-hand side of (7.21) and thus to prove (a) with a constant depending on γ_1,γ_2 and c_{max}/c_{min}.

(b) For all $1\leq k\leq l\leq d$,

$$w_{ti,kl}\leq\frac{\frac{1}{Tb_Tc_{min}}K\left(\frac{t-i}{Tb_Tc_{max}}\right)}{\frac{1}{Tb_Tc_{max}}\sum_{j=1}^{T}K\left(\frac{t-j}{Tb_Tc_{min}}\right)}.$$

Now, use the fact that K is bounded, $c_{max}/c_{min}<\infty$ and Lemma A(c) of Xu and Phillips (2008) to derive the upper bound.

(c) This is an easy consequence of Assumption **A1'**(i) and the proof of Lemma 7.5, equation (19), that holds uniformly w.r.t. $r\in(0,1]$ and $b_{kl}\in\mathcal{B}_T$, $1\leq k\leq l\leq d$.

(d) Let $a_i(k,l)$ denote a generic element of the $d\times d-$matrix $u_iu_i'-\Sigma_i$. Then we can write

$$\begin{aligned}E\left(\sup_{b\in\mathcal{B}_T}\left\|\sum_{i=1}^{T}w_{ti}(u_iu_i'-\Sigma_i)\right\|^4\right) &\leq E\left(\sup_{b\in\mathcal{B}_T}\left[\sum_{k,l=1}^{d}\left|\sum_{i=1}^{T}w_{ti}a_i(k,l)\right|\right]^4\right)\\ &\leq c\sum_{k,l=1}^{d}E\left(\sup_{b\in\mathcal{B}_T}\left|\sum_{i=1}^{T}w_{ti,kl}a_i(k,l)\right|^4\right)\\ &\leq c\sum_{k,l=1}^{d}E\left(\left|\sum_{i=1}^{T}\sup_{b\in\mathcal{B}_T}w_{ti,kl}|a_i(k,l)|\right|^4\right)\end{aligned}$$

where c depends only on d. Now, by Lemma A(f) of Xu and Phillips (2008) and (a)-(b) above, for $1 \leq k \leq l \leq d$

$$E\left(\sum_{i=1}^{T} \sup_{b\in\mathcal{B}_T} w_{ti,kl}|a_i(k,l)|\right)^4 \leq \left[\max_{1\leq t\leq T} \sum_{i=1}^{T} \sup_{b\in\mathcal{B}_T} w_{ti,kl}\right]^2 \sum_{i=1}^{T} \sup_{b\in\mathcal{B}_T} w_{ti,kl} E|a_i(k,l)|^4 \leq \frac{c}{(Tb_T)^2},$$

for c a constant depending only on K, c_{max}/c_{min} and the upper bounds of the 4th order moments of the components of $u_i u_i' - \Sigma_i$. Now, (7.20) follows.

(e) By Markov's inequality and obvious algebra we can write

$$\begin{aligned}
&P\left(\max_{1\leq t\leq T} \sup_{b\in\mathcal{B}_T} \| \overset{\circ}{\Sigma}_t - \bar{\Sigma}_t\|^\delta > C T^{-\delta/4} b_T^{-\delta/2}\right) \\
&= P\left(\max_{1\leq t\leq T} \sup_{b\in\mathcal{B}_T} \| \overset{\circ}{\Sigma}_t - \bar{\Sigma}_t\|^4 > C^{4/\delta} T^{-1} b_T^{-2}\right) \\
&\leq C^{-4/\delta} T b_T^2 E\left(\max_{1\leq t\leq T} \sup_{b\in\mathcal{B}_T} \| \overset{\circ}{\Sigma}_t - \bar{\Sigma}_t\|^4\right) \\
&\leq C^{-4/\delta} T b_T^2 \sum_{t=1}^{T} E\left(\sup_{b\in\mathcal{B}_T} \| \overset{\circ}{\Sigma}_t - \bar{\Sigma}_t\|^4\right) \\
&= C^{-4/\delta} T b_T^2 T \max_{1\leq t\leq T} E\left(\sup_{b\in\mathcal{B}_T} \| \overset{\circ}{\Sigma}_t - \bar{\Sigma}_t\|^4\right) \\
&= C^{-4/\delta} O(1)
\end{aligned}$$

where for the last equality we use (d).

(f)+(h) Using equation (3.5.33) in Horn and Johnson (1994) and (e) above we have

$$\begin{aligned}
\min_{1\leq t\leq T} \lambda_{min}(\overset{\circ}{\Sigma}_t) &\geq \min_{1\leq t\leq T} \lambda_{min}(\bar{\Sigma}_t) - \max_{1\leq t\leq T} \left|\lambda_{min}(\overset{\circ}{\Sigma}_t) - \lambda_{min}(\bar{\Sigma}_t)\right| \\
&\geq \min_{1\leq t\leq T} \lambda_{min}(\bar{\Sigma}_t) - \sup_{b\in\mathcal{B}_T} \max_{1\leq t\leq T} \| \overset{\circ}{\Sigma}_t - \bar{\Sigma}_t\| \\
&= \min_{1\leq t\leq T} \lambda_{min}(\bar{\Sigma}_t) + o_p(1),
\end{aligned}$$

and hence (f) follows from (c). Similar algebra applies for (h) which will follow as a consequence of (g).

(g)+(i) Adapt the proof of Lemma A(i) and A(k) of Xu and Phillips (2008) using a decomposition like in our equation (7.11).

(j) Apply Lemma A(l) of Xu and Phillips (2008) componentwise, that is d^2 times. □

Proof of (5.3) and (5.6). Let us denote

$$\hat{\Xi} = R\hat{\Lambda}_3^{-1}\hat{\Lambda}_2\hat{\Lambda}_3^{-1}R' \quad \text{and} \quad \hat{\Xi}_\delta = R\hat{\Lambda}_{3\delta}^{-1}\hat{\Lambda}_{2\delta}\hat{\Lambda}_{3\delta}^{-1}R'.$$

From the expressions of Q_{OLS} and Q_{OLS}^δ we write

$$| Q_{OLS} - Q_{OLS}^\delta | \leq T \| R\hat{\theta}_{OLS} \|^2 \| \hat{\Xi}^{-1} - \hat{\Xi}_\delta^{-1} \| = o_p(1),$$

since $T \parallel R\hat{\theta}_{OLS} \parallel^2 = O_p(1)$ and $\parallel \hat{\Xi}^{-1} - \hat{\Xi}_{\delta}^{-1} \parallel = o_p(1)$ from the consistency of the estimators of Λ_2 and Λ_3. In addition we write

$$\frac{Q_{OLS} + Q_{OLS}^{\delta}}{2} = \frac{T}{2}\hat{\theta}'_{OLS}R'\{\hat{\Xi}^{-1} + \hat{\Xi}_{\delta}^{-1}\}R\hat{\theta}_{OLS}.$$

Noting that $\{\hat{\Xi}^{-1} + \hat{\Xi}_{\delta}^{-1}\}/2 = R\Lambda_3^{-1}\Lambda_2\Lambda_3^{-1}R' + o_p(1)$, we have $\{Q_{OLS} + Q_{OLS}^{\delta}\}/2 \Rightarrow \chi^2_{pd_1d_2}$. Since $\max(a,b) = \{a + b + \mid a - b \mid\}/2$, the result (5.6) follows from (5.4) and (5.5). The result (5.3) can be obtain in a similar way. $\square$

References

Andrews, D.W.K. (2008). Laws of large numbers for dependent nonidentically distributed random variables. *Econometric Theory* 4, 458-467.

Bae, J., and Choi, M.J. (1999). The uniform CLT for martingale difference of function-indexed process under uniform integrable entropy. *Communications Korean Mathematical Society* 14, 581-595.

Bae, J., Jun, D., and Levental, S. (2010). The uniform CLT for martingale differences arrays under the uniformly integrable entropy. *Bulletin of the Korean Mathematical Society* 47, 39-51.

Bataa, E., Osborn, D.R., Sensier, M., and van Dijk, D. (2009). Structural breaks in the international transmission of inflation. Discussion Paper, Centre for Growth and Business Cycle Research Economic Studies, University of Manchester.

Batbekh, S., Osborn, D.R., Sensier, M., and van Dick, D. (2007). Is there causality in the mean and volatility of inflation between the US and other G7 countries? Centre for Growth and Business Cycle Research Economic Studies, University of Manchester.

Blanchard, O., and Simon, J. (2001). The long and large decline in U.S. output volatility. Brookings Papers on Economic Activity 1, 135-164.

Boswijk, H.P., and Zu, Y. (2007). Testing for cointegration with nonstationary volatility. Working Paper, University of Amsterdam.

Busetti, F., and Taylor, A.M.R. (2003). Variance shifts, structural breaks, and stationarity tests. *Journal of Business and Economic Statistics* 21, 510-531.

Cavaliere, G. (2004). Unit root tests under time-varying variance shifts. *Econometric Reviews* 23, 259-292.

Cavaliere, G., Rahbek, A., and Taylor, A.M.R. (2010). Testing for co-integration in vector autoregressions with non-stationary volatility. *Journal of Econometrics*, forthcoming.

Cavaliere, G., and Taylor, A.M.R. (2007). Testing for unit roots in time series models with non-stationary volatility. *Journal of Econometrics* 140, 919-947.

Davidson, J. (1994). *Stochastic Limit Theory.* Oxford University Press.

Doyle, B.M., and Faust, J. (2005). Breaks in the variability and comovement of G-7 economic growth. *The Review of Economics and Statistics* 87, 721-740.

Feige, E.L., and Pierce, D.K. (1979). The casual causal relationship between money and income: some caveats for time series analysis. *Review of Economics and Statistics* 61, 21-33.

Granger, C.W.J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 12, 424-438.

Hafner, C.M., and Herwartz, H. (2009). Testing for linear vector autoregressive dynamics under multivariate generalized autoregressive heteroskedasticity. *Statistica Neerlandica* 63, 294-323.

Hansen, B.E. (1995). Regression with nonstationary volatility. *Econometrica* 63, 1113-1132.

Horn, R.A., and Johnson, C.R. (1994). *Topics in Matrix Analysis.* Cambridge University Press, Cambridge.

Kim, T-H., Leybourne, S., and Newbold, P. (2002). Unit root tests with a break in innovation variance. *Journal of Econometrics* 103, 365-387.

Kim, C.J., and Nelson, C.R. (1999). Has the U.S. economy become more stable? A bayesian approach based on a Markov-switching model of the business cycle. *The Review of Economics and Statistics* 81, 608-616.

Lavergne, P. (2008). A Cauchy-Schwartz inequality for expectation of matrices. Working Paper, Simon Fraser University.

Levental, S. (1989). A uniform CLT for uniformly bounded families of martingale differences. *Journal of Theoretical Probability* 2(3), 271-287.

Lopez, O., and Patilea, V. (2010). Testing conditional moment restrictions in the presence of right-censoring depending on the covariates. Discussion Paper, CREST-Ensai.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis.* Springer, Berlin.

McConnell, M.M., and Perez-Quiros, G. (2000). Output fluctuations in the United States: what has changed since the early 1980's? *The American Economic Review* 90, 1464-1476.

Nolan, D., and Pollard, D. (1987). *U*-Processes: Rates of Convergence. *The Annals of Statistics* 15(2), 780-799.

Patilea, V., and Raïssi, H. (2010). Adaptive estimation of vector autoregressive models with time-varying variance: application to testing linear causality in mean, arXiv:1007.1193v2.

Phillips, P.C.B., and Xu, K.L. (2005). Inference in autoregression under heteroskedasticity. *Journal of Time Series Analysis* 27, 289-308.

Ramey, V.A., and Vine, D.J. (2006). Declining volatility in the U.S. automobile industry. *The American Economic Review* 96, 1876-1889.

Robinson, P.M. (1987). Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form. *Econometrica* 55, 875-891.

Sanso, A., Aragó, V., and Carrion, J.L. (2004). Testing for changes in the unconditional variance of financial time series. DEA Working Paper, Universitat de les Illes Balears.

Sensier, M., and van Dijk, D. (2004). Testing for volatility changes in U.S. macroeconomic time series. *Review of Economics and Statistics* 86, 833-839.

Sims, C.A. (1972). Money income and causality. *American Economic Review* 62, 540-542.

Stărică, C., and Granger C.W.J. (2005). Nonstationarities in stock returns. *Reviews of Economics and Statistics* 87, 503-522.

Stock, J.H., and Watson, M.W. (1989). Interpreting the evidence on money-income causality. *Journal of Econometrics* 40, 161-181.

Warnock, M.V.C., and Warnock, F.E. (2000). The declining volatility of U.S. employment: was Arthur Burns right? Board of Governors of the Federal Reserve System, International finance discussion papers, 677.

Xu, K.L., and Phillips, P.C.B. (2008). Adaptive estimation of autoregressive models with time-varying variances. *Journal of Econometrics* 142, 265-280.

Table 1: Empirical size (in %) of the different Wald tests. The innovations are heteroscedastic with $\gamma_1 = 20$, $\rho = 0.6$, and we take $a_{11} = a_{22} = 0.2$, $a_{21} = 0.1$, $a_{12} = 0$.

T	50	100	200	400
W_{OLS}	**8.8**	5.8	4.8	5.2
W_{OLS}^{δ}	**9.7**	**6.5**	5.0	5.4
$W_{OLS}^{\max}$	**10.2**	**6.8**	5.0	5.5
W_S	**9.3**	**8.1**	**6.6**	**8.0**
W_{ALS}	**7.1**	5.5	4.9	4.8
W_{ALS}^{δ}	**8.3**	6.2	5.6	5.4
$W_{ALS}^{\max}$	**8.3**	6.3	5.6	5.4
W_{GLS}	5.2	4.1	5.2	4.2
W_{GLS}^{δ}	5.7	4.0	4.2	**3.4**
$W_{GLS}^{\max}$	**6.3**	4.4	5.4	4.2

Table 2: Empirical power (in %) of the different Wald tests. The innovations are heteroscedastic with $\gamma_1 = 20$ and $\rho = 0.6$. We take $a_{11} = a_{22} = 0.2$ and $a_{21} = 0.1$.

a_{12}	-0.8	-0.6	-0.4	-0.2	0.2	0.4	0.6	0.8
W_{OLS}	96.9	81.4	48.1	17.3	14.2	40.3	70.0	90.6
W_{OLS}^{δ}	97.5	82.4	49.9	18.2	15.5	41.5	70.8	90.9
$W_{OLS}^{\max}$	97.7	83.4	51.5	18.7	15.7	42.2	71.4	91.3
W_S	98.4	85.4	53.9	20.1	17.1	46.3	75.5	92.8
W_{ALS}	98.8	86.7	50.8	17.7	13.5	45.2	75.4	93.0
W_{ALS}^{δ}	98.9	87.8	54.2	19.4	15.6	48.4	77.4	94.5
$W_{ALS}^{\max}$	98.9	87.8	54.2	19.4	15.6	48.4	77.4	94.5
W_{GLS}	99.1	88.7	52.6	17.7	14.2	48.0	79.5	96.1
W_{GLS}^{δ}	99.1	89.9	52.9	18.1	12.4	46.9	78.3	95.1
$W_{GLS}^{\max}$	99.2	90.4	55.5	19.0	14.5	49.6	80.5	96.1

τ_1 σ_{21} τ_2

τ_1

FIGURE 7.1: The ratio $\mathrm{Var}_{as}\left(\hat{\theta}_{2,OLS}\right)/\mathrm{Var}_{as}\left(\hat{\theta}_{2,GLS}\right)$ of Example 3.1.

τ_1 σ_{21}

FIGURE 7.2: The coefficient κ_1 of Example 5.1.

RMSE$\times 10^2$

a_{11}

FIGURE 7.3: The RMSE of the estimators of the parameter a_{11} over $N = 1000$ replications with varying $a_{11} = a_{22}$, and $a_{12} = 0$, $a_{21} = 0.1$. We take $\gamma_1 = 20$, $\rho = 0.6$ and $T = 100$. The RMSE are displayed in blue for the ALS estimators, in green for the OLS estimators and in red for the GLS estimators.