

Corrected portmanteau tests for VAR models with time-varying variance

VALENTIN PATILEA*, AND HAMDI RAÏSSI†

June 1, 2011

Abstract: The problem of test of fit for Vector AutoRegressive (VAR) processes with unconditionally heteroscedastic errors is studied. The volatility structure is deterministic but time-varying and allows for changes that are commonly observed in economic or financial multivariate series such as breaks or smooth transitions. Our analysis is based on the residual autocovariances and autocorrelations obtained from Ordinary Least Squares (OLS), Generalized Least Squares (GLS) and Adaptive Least Squares (ALS) estimation of the autoregressive parameters. The OLS residuals are the standard estimates of the VAR model errors. To build the GLS residuals we use the GLS estimate of the VAR coefficients to estimate the model errors that we further standardize by the time-varying volatility. Hence, the GLS estimates require the knowledge of the variance structure. The ALS approach is the GLS approach adapted to the *unknown* time-varying volatility that is then estimated

*CREST (Ensaï) & IRMAR (UEB), ENSAI - Campus de Ker-Lann, Rue Blaise Pascal - BP 37203, 35172 BRUZ cedex, France. E-mail: valentin.patilea@insa-rennes.fr

†IRMAR-INSA, 20 avenue des buttes de Coësmes, CS 70839, F-35708 Rennes Cedex 7, France.

E-mail: hamdi.raïssi@insa-rennes.fr

by kernel smoothing. The properties of the three types of residual autocovariances and autocorrelations are derived. In particular it is shown that the ALS and GLS residual autocorrelations are asymptotically equivalent. It is also found that the asymptotic distribution of the OLS residual autocorrelations can be quite different from the standard chi-square asymptotic distribution obtained in a correctly specified VAR model with iid innovations. As a consequence the standard portmanteau tests are unreliable in our framework. The correct critical values of the standard portmanteau tests based on the OLS residuals are derived. Moreover, modified portmanteau statistics based on ALS residual autocorrelations are introduced and their asymptotic critical values are obtained. The finite sample properties of the goodness-of-fit tests we consider are investigated by Monte Carlo experiments. The theoretical results are also illustrated using a U.S. economic data set.

Keywords: VAR model; Unconditionally heteroscedastic errors; Residual autocorrelations; Portmanteau tests.

1 Introduction

In the econometric analysis numerous tools are routinely used in the framework of VAR (Vector AutoRegressive) modeling of time series variables (see Lütkepohl (2005) and references therein). Nevertheless it is well known that these tools are in general noticeably affected by the adjusted autoregressive order. For instance Thornton and Batten (1985), Stock and Watson (1989) or Jones (1989) discussed the importance of well specified VAR model for the test of linear Granger causality in mean. Therefore the checking of goodness-of-fit of the autoregressive order is commonly performed in applied works before proceeding to the analysis of the dynamics of time series. The dominant tests for the adequacy of the autoregressive order are the portmanteau tests introduced in the VAR framework by Chitturi (1974) and Hosking (1980). The properties of the tests based on the residual autocorrelations are well explored in the case of stationary processes (see e.g. Francq, Roy and Zakoïan (2005) in the univariate case, Francq and Raïssi (2007) or Boubacar Mainasara (2010) in the multivariate case). Duchesne (2005), Brüggemann, Lütkepohl and Saikkonen (2006) and Raïssi (2010) developed tests for residual autocorrelation in a cointegrated framework with stationary innovations.

However many applied studies pointed out the presence of non stationary volatility in economic time series. For instance Ramey and Vine (2006) found a declining volatility in the U.S. automobile industry. Watson (1999) noted a declining volatility of short-term U.S. interest rates and increasing volatility for long-term U.S. interest rates. Sensier and van Dijk (2004) considered 214 U.S. macroeconomic variables and found that approximately 80% of these variables have a volatility that changes in time. These findings stimulated an interest on the effects of non-stationary volatility in time series analysis amongst econometricians (see e.g. Kim, Leybourne and

Newbold (2002) or Cavaliere and Taylor (2007)).

The present paper is motivated by the need of reliable tools for testing the adequacy of the autoregressive order of VAR models with non stationary volatility. On one hand, we show that in such cases the use of standard procedures for testing the adequacy of the autoregressive order can be quite misleading. On the other hand, valid portmanteau tests based on Ordinary Least Squares (OLS) and Adaptive Least Squares (ALS) residual autocovariances are proposed for testing the goodness-of-fit tests of non-stationary but stable VAR processes. More precisely we consider the VAR model of order $p \geq 0$ and dimension $d \geq 1$

$$\begin{aligned} X_t &= A_{01}X_{t-1} + \cdots + A_{0p}X_{t-p} + u_t, \\ u_t &= H_t\epsilon_t, \quad t = 1, 2, \dots \end{aligned} \tag{1.1}$$

where X_t are random vectors of dimension d and the $d \times d$ -matrices A_{0i} , $i \in \{1, \dots, p\}$, are such that the process (X_t) is stable, that means $\det(A(z)) \neq 0$ for all $|z| \leq 1$, with $A(z) = I_d - \sum_{i=1}^p A_{0i}z^i$. Here, H_t is an *unknown* $d \times d$ matrix-valued deterministic function of time and (ϵ_t) is an innovation process of unit variance that could be serially dependent. Phillips and Xu (2005), Xu and Phillips (2008) already studied the problem of estimation of such univariate stable autoregressive processes. Patilea and Raïssi (2010) investigated the estimation and the test of parameter restrictions of multivariate stable autoregressive processes like in (1.1).

The usual way to check the adequacy of a stable VAR(p) model, implemented in any specialized software, is to assume that the error term u_t is second order stationary, to fix an integer $m > 0$ and to test

$$\mathcal{H}_0 : \text{Cov}(u_t, u_{t-h}) = 0, \text{ for all } 0 < h \leq m, \tag{1.2}$$

using a classical (Box-Pierce or Ljung-Box) portmanteau test statistic and chi-square

type critical values. The errors u_t are approximated using the OLS type estimates of the coefficients A_{0i} . With the volatility structure we assumed in model (1.1), the variance of u_t depends on t and the usual chi-square type critical values are in general inaccurate, that is the asymptotic distribution of the classical portmanteau statistics under \mathcal{H}_0 is no longer of chi-square type.

Here we propose two ways to correct this problem. First, we derive the correct asymptotic distribution of the classical portmanteau test statistics under \mathcal{H}_0 and the conditions of model (1.1). This asymptotic distribution is a weighted sum of d^2m independent chi-square distributions. Next, we indicate how the correct critical values can be approximated.

To explain our second approach, let us notice that $\text{Cov}(u_t, u_{t-h}) = 0$ is equivalent to $\text{Cov}(\epsilon_t, \epsilon_{t-h}) = 0$ and the variance of ϵ_t does not depend on the time t . Thus an alternative idea for checking the adequacy of a model like (1.1) is to test

$$\mathcal{H}'_0 : \text{Cov}(\epsilon_t, \epsilon_{t-h}) = 0, \text{ for all } 0 < h \leq m. \quad (1.3)$$

The values ϵ_t are approximated by residuals built using a nonparametric estimate of the deterministic function H_t and Adaptive Least Squares (ALS) type estimates of the coefficients A_{0i} that take into account the volatility structure. More precisely, to build the ALS residual vector at time t we use the ALS estimate of the VAR coefficients to estimate the VAR model error vector at time t that we further standardize by the nonparametric estimate of time-varying volatility H_t . Next, we build classical portmanteau test statistics using the estimates of ϵ_t and we derive the asymptotic distribution under \mathcal{H}_0 . The asymptotic distribution is again a weighted sum of d^2m independent chi-square distributions and the weights can be easily estimated from the data. In some important particular cases, including the univariate (i.e. $d = 1$) autoregressive models, we retrieve the standard chi-squared asymptotic distribution.

The remainder of the paper is organized as follows. In section 2 we specify the framework of our study and state the asymptotic behavior of the OLS and the Generalized Least Squares (GLS) estimators of the VAR coefficients. The asymptotic normality of the OLS and the infeasible GLS residual autocovariances and autocorrelations is established in section 3. The GLS residuals are defined as the standardized (by the true volatility H_t) estimates of the model error vector obtained with the GLS estimates of the VAR coefficients. In section 4 we highlight the unreliability of the chi-square type critical values for standard portmanteau statistics and we derive their correct critical values in our framework. In section 5 the ALS residual autocovariances and autocorrelations are introduced. Since the GLS residual autocovariances and autocorrelations are infeasible, we investigate the relationship between the GLS and ALS residual autocovariances and autocorrelations and we show that, in some sense, they are asymptotically equivalent. This result is used to introduce portmanteau tests based on the ALS residuals that have the same critical values like those based on the infeasible GLS residuals. In section 6 we propose suitably modified quadratic forms of OLS and ALS residual autocovariances in order to obtain alternative test statistics with chi-square asymptotic distributions under the null hypothesis. Such modified statistics are nothing but Wald type test statistics for testing the nullity of a vector of autocovariances. In section 7 some theoretical comparisons of the asymptotic power, in the Bahadur sense, are carried out: classical Box-Pierce portmanteau test vs. modified quadratic forms of OLS residual autocorrelations based test; and ALS vs. OLS residual autocorrelations based portmanteau tests. A possible extension of our findings on testing the order of a VAR model to the case of heteroscedastic co-integrated variables is briefly described in section 8. The finite sample properties of the different tests considered in this

paper are studied by mean of Monte Carlo experiments in section 9. In section 10 applications to U.S. economic real data sets are used to illustrate the theoretical results: the U.S. balance on services and balance on merchandise trade data, and the U.S. energy-transport consumer price indexes. The summary of our finding and some concluding remarks are given in section 11. The proofs and the tables and figures are relegated in the appendices.

2 Parameters estimation

In the following weak convergence is denoted by \Rightarrow while \xrightarrow{P} stands for convergence in probability. The symbol \otimes denotes the usual Kronecker product for matrices and $A^{\otimes 2}$ stands for $A \otimes A$. The symbol $\text{vec}(\cdot)$ is used for the column vectorization operator. We denote by $[a]$ the integer part of a real number a . For a squared matrix A , $\text{tr}(A)$ denotes the trace. For a random variable x we define $\|x\|_r = (E \|x\|^r)^{1/r}$, where $\|x\|$ denotes the Euclidean norm. We also define the σ -field $\mathcal{F}_t = \sigma(\epsilon_s : s \leq t)$. The following conditions on the innovations process (u_t) are assumed to hold.

Assumption A1: (i) The $d \times d$ matrices H_t are positive definite and the components $\{g_{ij}(r) : 1 \leq k, l \leq d\}$ of the matrix $G(r)$ are measurable deterministic functions on the interval $(0, 1]$, such that $H_t = G(t/T)$ and, $\forall 1 \leq k, l \leq d$, $\sup_{r \in (0, 1]} |g_{k,l}(r)| < \infty$ and $g_{k,l}(\cdot)$ satisfies a Lipschitz condition piecewise on a finite number of some sub-intervals that partition $(0, 1]$ (the partition may depend on k, l). 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 $E(\epsilon_t | \mathcal{F}_{t-1}) = 0$, $E(\epsilon_t \epsilon_t' | \mathcal{F}_{t-1}) = I_d$ and $\sup_t \|\epsilon_{it}\|_{4\mu} < \infty$ for some $\mu > 1$ and all $i \in \{1, \dots, d\}$.

The second approach we propose for checking the adequacy of a VAR(p) model requires the estimation of the innovations ϵ_t , and hence we will need an identification condition for $G(r)$ and an estimate of the matrix H_t . The condition H_t is positive definite matrix identifies $G(r)$ as the square root of $\Sigma(r)$ and this is a convenient choice for the mathematical proofs. Nevertheless one can notice from the following that our results could be stated using alternative conditions, like for instance H_t is a lower triangular matrix with diagonal components restricted to be positive. The conditions on the unknown volatility function $G(r)$ are general and allow for a large set of dynamics for the innovation variance as for instance abrupt shifts or piecewise affine functions. This assumption generalizes to a multivariate framework the specification of the innovation variance considered in Xu and Phillips (2008). The conditional homoscedasticity of (ϵ_t) imposed in (ii) ensures the identifiability of $\Sigma(r)$. We call a model like in (1.1) with the innovation process (u_t) satisfying Assumption **A1** a stable VAR(p) model with time-varying variance.

To introduce the OLS and GLS estimators of the autoregressive parameters, set the observations X_{-p+1}, \dots, X_0 equal to the null vector of \mathbb{R}^d (or any other initial values) and denote by $\theta_0 = (\text{vec}(A_{01})' \dots \text{vec}(A_{0p})')' \in \mathbb{R}^{pd^2}$ the vector of true parameters. The equation (1.1) becomes

$$X_t = (\tilde{X}'_{t-1} \otimes I_d)\theta_0 + u_t, \quad t = 1, 2, \dots \quad (2.1)$$

$$u_t = H_t \epsilon_t,$$

with $\tilde{X}'_{t-1} = (X'_{t-1}, \dots, X'_{t-p})'$. Then the OLS estimator is

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

where

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

Multiplying by H_t^{-1} on the left in equation (2.1) we obtain

$$H_t^{-1} X_t = H_t^{-1} (\tilde{X}'_{t-1} \otimes I_d) \theta_0 + \epsilon_t,$$

and then the GLS estimator is

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

with

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

In general, the GLS estimator is infeasible since it involves the true volatility matrix.

Due to the stability condition, we can write $X_t = \sum_{i=0}^{\infty} \psi_i u_{t-i}$, where $\psi_0 = I_d$ and the components of the ψ_i 's are absolutely summable $d \times d$ -matrices. Then

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

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

$$\tilde{\psi}_i = \text{diag}\{\psi_i, \psi_{i-1}, \dots, \psi_{i-p+1}\},$$

taking $\psi_j = 0$ for $j < 0$.

Let $\mathbf{1}_{p \times p}$ stand for the $p \times p$ -matrix with all components equal to one. Patilea and Raïssi (2010) proved that under **A1**

$$T^{\frac{1}{2}}(\hat{\theta}_{GLS} - \theta_0) \Rightarrow \mathcal{N}(0, \Lambda_1^{-1}), \quad (2.3)$$

where

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

and

$$T^{\frac{1}{2}}(\hat{\theta}_{OLS} - \theta_0) \Rightarrow \mathcal{N}(0, \Lambda_3^{-1} \Lambda_2 \Lambda_3^{-1}), \quad (2.4)$$

with

$$\begin{aligned} \Lambda_2 &= \int_0^1 \sum_{i=0}^{\infty} \left\{ \tilde{\psi}_i(\mathbf{1}_{p \times p} \otimes \Sigma(r)) \tilde{\psi}_i' \right\} \otimes \Sigma(r) dr, \\ \Lambda_3 &= \int_0^1 \sum_{i=0}^{\infty} \left\{ \tilde{\psi}_i(\mathbf{1}_{p \times p} \otimes \Sigma(r)) \tilde{\psi}_i' \right\} dr \otimes I_d. \end{aligned}$$

Moreover, they showed that $\Lambda_3^{-1} \Lambda_2 \Lambda_3^{-1} - \Lambda_1^{-1}$ is positive semi-definite.

3 Asymptotic behavior of the residual autocovariances

Let us define the OLS-based estimates of u_t and the GLS-based estimates of ϵ_t

$$\hat{u}_t = X_t - (\tilde{X}'_{t-1} \otimes I_d) \hat{\theta}_{OLS} \quad \text{and} \quad \hat{\epsilon}_t = H_t^{-1} X_t - H_t^{-1} (\tilde{X}'_{t-1} \otimes I_d) \hat{\theta}_{GLS}.$$

The corresponding residual autocovariances are defined as

$$\hat{\Gamma}_{OLS}^u(h) = T^{-1} \sum_{t=h+1}^T \hat{u}_t \hat{u}'_{t-h} \quad \text{and} \quad \hat{\Gamma}_{GLS}^\epsilon(h) = T^{-1} \sum_{t=h+1}^T \hat{\epsilon}_t \hat{\epsilon}'_{t-h}.$$

In general the estimated residuals $\hat{\epsilon}_t$ as well as the autocovariances $\hat{\Gamma}_{GLS}^\epsilon(h)$ are not computable since they depend on the unknown matrices H_t and the infeasible estimator $\hat{\theta}_{GLS}$.

For any fixed integer $m \geq 1$, the estimates of the first m residual autocovariances are defined by

$$\hat{\gamma}_m^{u,OLS} = \text{vec} \left\{ \left(\hat{\Gamma}_{OLS}^u(1), \dots, \hat{\Gamma}_{OLS}^u(m) \right) \right\}, \quad \hat{\gamma}_m^{\epsilon,GLS} = \text{vec} \left\{ \left(\hat{\Gamma}_{GLS}^\epsilon(1), \dots, \hat{\Gamma}_{GLS}^\epsilon(m) \right) \right\}.$$

To state the asymptotic behavior of $\hat{\gamma}_m^{u,OLS}$ and $\hat{\gamma}_m^{\epsilon,GLS}$ let us define

$$K = \begin{pmatrix} A_{01} & \dots & A_{0p-1} & A_{0p} \\ I_d & 0 & \dots & 0 \\ & \ddots & \ddots & \vdots \\ 0 & & I_d & 0 \end{pmatrix}.$$

Note that if $\tilde{u}_t = (u'_t, 0 \dots, 0)'$, $\tilde{X}_t = K\tilde{X}_{t-1} + \tilde{u}_t$. Now, let $\Sigma_G = \int_0^1 \Sigma(r)dr$, $\Sigma_{G^{\otimes 2}} = \int_0^1 \Sigma(r)^{\otimes 2}dr$ and

$$\Phi_m^u = \sum_{i=0}^{m-1} \{e_m(i+1)e_p(1)' \otimes \Sigma_G \otimes I_d\} \{K^{i'} \otimes I_d\}, \quad (3.1)$$

$$\Lambda_m^{u,\theta} = \sum_{i=0}^{m-1} \{e_m(i+1)e_p(1)' \otimes \Sigma_{G^{\otimes 2}}\} \{K^{i'} \otimes I_d\}, \quad (3.2)$$

$$\Lambda_m^{\epsilon,\theta} = \sum_{i=0}^{m-1} \left\{ e_m(i+1)e_p(1)' \otimes \int_0^1 G(r)' \otimes G(r)^{-1} dr \right\} \{K^{i'} \otimes I_d\}, \quad (3.3)$$

$$\Lambda_m^{u,u} = I_m \otimes \Sigma_{G^{\otimes 2}}, \quad (3.4)$$

where $e_m(j)$ is the vector of dimension m such that the j th component is equal to one and zero elsewhere.¹

Proposition 1 *If model (1.1) is correct and Assumption A1 holds true, we have*

$$T^{\frac{1}{2}}\hat{\gamma}_m^{u,OLS} \Rightarrow \mathcal{N}(0, \Sigma^{u,OLS}), \quad (3.5)$$

where

$$\Sigma^{u,OLS} = \Lambda_m^{u,u} - \Lambda_m^{u,\theta} \Lambda_3^{-1} \Phi_m^{u'} - \Phi_m^u \Lambda_3^{-1} \Lambda_m^{u,\theta'} + \Phi_m^u \Lambda_3^{-1} \Lambda_2 \Lambda_3^{-1} \Phi_m^{u'}, \quad (3.6)$$

$$T^{\frac{1}{2}}\hat{\gamma}_m^{\epsilon,GLS} \Rightarrow \mathcal{N}(0, \Sigma^{\epsilon,GLS}), \quad (3.7)$$

where

$$\Sigma^{\epsilon,GLS} = I_{d^2m} - \Lambda_m^{\epsilon,\theta} \Lambda_1^{-1} \Lambda_m^{\epsilon,\theta'}. \quad (3.8)$$

In the particular case $p = 0$, $\Sigma^{u,OLS} = \Lambda_m^{u,u}$ and $\Sigma^{\epsilon,GLS} = I_{d^2m}$.

¹Recall that our identification condition for H_t implies $G(r) = \Sigma(r)^{1/2}$.

Let us discuss the conclusions of Proposition 1 in some particular situations. In the case where $\Sigma(\cdot) = \sigma^2(\cdot)I_d$ for some positive scalar function $\sigma(\cdot)$, we have

$$\Lambda_m^{\epsilon, \theta} = \sum_{i=0}^{m-1} \{e_m(i+1)e_p(1)' \otimes I_d \otimes I_d\} \{K^{i'} \otimes I_d\}, \quad \Lambda_1 = \sum_{i=0}^{\infty} \left\{ \tilde{\psi}_i(\mathbf{1}_{p \times p} \otimes I_d) \tilde{\psi}_i' \right\} \otimes I_d,$$

so that in this case the asymptotic distribution of the ϵ_t autocovariances estimates $\hat{\gamma}_m^{\epsilon, GLS}$ do not depend on the volatility function $\Sigma(\cdot)$. Meanwhile, the (asymptotic) covariance matrix $\Sigma^{u, OLS}$ still depends on the volatility function.

If we suppose that (u_t) have a time-constant variance $\Sigma(r) \equiv \Sigma_u$, we obtain

$$\Lambda_1 = E \left[\tilde{X}_t \tilde{X}_t' \right] \otimes \Sigma_u^{-1}, \quad \Lambda_m^{\epsilon, \theta} = E \left[\epsilon_t^m \tilde{X}_t' \right] \otimes G_u^{-1}, \quad \Lambda_m^{u, u} = I_m \otimes \Sigma_u^{\otimes 2}, \quad \Lambda_3 = E \left[\tilde{X}_t \tilde{X}_t' \right] \otimes I_d,$$

where $\Sigma_u = G_u G_u'$, and

$$\Lambda_m^{u, \theta} = E \left[u_t^m \tilde{X}_t' \right] \otimes \Sigma_u, \quad \Lambda_2 = E \left[\tilde{X}_t \tilde{X}_t' \right] \otimes \Sigma_u, \quad \Phi_m^u = E \left[u_t^m \tilde{X}_t' \right] \otimes I_d,$$

where $u_t^m = (u_t', \dots, u_{t-m}')'$ and $\epsilon_t^m = (\epsilon_t', \dots, \epsilon_{t-m}')'$. By straightforward computations

$$\Sigma^{u, OLS} = I_m \otimes \Sigma_u^{\otimes 2} - E \left[u_t^m \tilde{X}_t' \right] E \left[\tilde{X}_t \tilde{X}_t' \right]^{-1} E \left[u_t^m \tilde{X}_t' \right]' \otimes \Sigma_u, \quad (3.9)$$

$$\Sigma^{\epsilon, GLS} = I_{d^2 m} - E \left[\epsilon_t^m \tilde{X}_t' \right] E \left[\tilde{X}_t \tilde{X}_t' \right]^{-1} E \left[\epsilon_t^m \tilde{X}_t' \right]' \otimes I_d. \quad (3.10)$$

Formula (3.9) (resp. (3.10)) corresponds to the (asymptotic) covariance matrix obtained in the standard case with an i.i.d. error process of variance Σ_u (resp. I_d), see Lütkepohl (2005), Proposition 4.5. Herein, some dependence of the error process is allowed. In particular, equation (3.10) indicates that the homoscedastic (time-constant variance) case is another situation where $\Sigma^{\epsilon, GLS}$ does not depend on error process variance Σ_u .

Proposition 1 shows that in general VAR models with time-varying variance the covariance matrix $\Sigma^{\epsilon, GLS}$ depends on $\Sigma(\cdot)$. For the sake of simpler notation,

hereafter we write $\hat{\Gamma}_{OLS}(h)$ (resp. $\hat{\gamma}_m^{OLS}$) (resp. Σ^{OLS}) instead of $\hat{\Gamma}_{OLS}^u(h)$ (resp. $\hat{\gamma}_m^{u,OLS}$) (resp. $\Sigma^{u,OLS}$). Similar notation simplification will be applied for $\hat{\Gamma}_{\epsilon,GLS}(h)$, $\hat{\gamma}_m^{\epsilon,GLS}$ and $\Sigma^{\epsilon,GLS}$.

The following example shows that when the error process is heteroscedastic, the covariance matrices Σ^{OLS} and Σ^{GLS} can be quite different and far from the covariance matrices obtained in the stationary case.

Example 3.1 Consider a bivariate $AR(1)$ model $X_t = A_0 X_{t-1} + u_t$ with true parameter A_0 equal to the zero 2×2 -matrix. One can use such a model to study linear Granger causality in mean between uncorrelated variables. However in practice one has first to check that the error process is a white noise. If we assume for simplicity that

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

we obtain diagonal covariance matrices $\Sigma^{OLS} = \text{diag}\{0_{4 \times 4}, I_{m-1} \otimes \check{\Sigma}^{OLS}\}$ and $\Sigma^{GLS} = \text{diag}\{\check{\Sigma}^{GLS}, I_{4(m-1)}\}$, with

$$\check{\Sigma}^{OLS} = \begin{pmatrix} \int_0^1 \Sigma_1(r)^2 dr & 0 & 0 & 0 \\ 0 & \int_0^1 \Sigma_1(r) \Sigma_2(r) dr & 0 & 0 \\ 0 & 0 & \int_0^1 \Sigma_1(r) \Sigma_2(r) dr & 0 \\ 0 & 0 & 0 & \int_0^1 \Sigma_2(r)^2 dr \end{pmatrix}$$

and

$$\check{\Sigma}^{GLS} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 - \frac{(\int_0^1 \Sigma_1(r)^{\frac{1}{2}} \Sigma_2(r)^{-\frac{1}{2}} dr)^2}{\int_0^1 \Sigma_1(r) \Sigma_2(r)^{-1} dr} & 0 & 0 \\ 0 & 0 & 1 - \frac{(\int_0^1 \Sigma_1(r)^{-\frac{1}{2}} \Sigma_2(r)^{\frac{1}{2}} dr)^2}{\int_0^1 \Sigma_2(r) \Sigma_1(r)^{-1} dr} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

We denote by $0_{q \times q}$ the null matrix of dimension $q \times q$. Note that the matrix $I_{4(m-1)}$ which appears in the expression of Σ^{GLS} is a consequence of the assumption $A_0 = 0_{2 \times 2}$. If we suppose that the errors are homoscedastic, that is $\Sigma(r)$ is constant, equal to some Σ_u , we obtain $\Sigma^{OLS} = \text{diag}\{0_{4 \times 4}, I_{m-1} \otimes \Sigma_u^{\otimes 2}\}$ and $\Sigma^{GLS} = \text{diag}\{0_{4 \times 4}, I_{4(m-1)}\}$. Therefore in the OLS approach and if the innovations variance is spuriously assumed constant, the asymptotic spurious covariance matrix $\Sigma_S^{OLS} = \text{diag}\{0_{4 \times 4}, I_{m-1} \otimes \Sigma_{u,S}^{\otimes 2}\}$ is used with

$$\Sigma_{u,S} = \begin{pmatrix} \int_0^1 \Sigma_1(r) dr & 0 \\ 0 & \int_0^1 \Sigma_2(r) dr \end{pmatrix}.$$

Now we illustrate the difference between the covariance matrices obtained if we take into account the unconditional heteroscedasticity of the process and the case where the process is spuriously supposed homoscedastic. We take

$$\Sigma_1(r) = \sigma_{10}^2 + (\sigma_{11}^2 - \sigma_{10}^2) \times \mathbf{1}_{\{r \geq \tau_1\}}(r) \quad (3.11)$$

and

$$\Sigma_2(r) = \sigma_{20}^2 + (\sigma_{21}^2 - \sigma_{20}^2) \times \mathbf{1}_{\{r \geq \tau_2\}}(r), \quad (3.12)$$

where $\tau_i \in [0, 1]$ 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)). In Figure 1, we take $\tau_1 = \tau_2$, $\sigma_{10}^2 = \sigma_{20}^2 = 1$ and $\sigma_{11}^2 = 0.5$, so that only the break dates and σ_{21}^2 vary freely. In figure 2 only the break dates vary with $\tau_1 \neq \tau_2$ in general, and $\sigma_{10}^2 = \sigma_{20}^2 = 1$, $\sigma_{11}^2 = \sigma_{21}^2 = 4$. In Figure 1 and 2 we plot in the left graphics the second component $\Sigma^{GLS}(2, 2)$ on the diagonal of Σ^{GLS} and in the right graphics the ratio of $\Sigma^{OLS}(6, 6)/\Sigma_S^{OLS}(6, 6)$.

From the left graphic of Figure 1 it turns out that $\Sigma^{GLS}(2, 2)$ could be far from zero for larger values of σ_{21} and when the breaking point τ_1 is located early

in the sample. From the right graphic of Figure 1 we can see that the ratio of $\Sigma^{OLS}(6,6)/\Sigma_S^{OLS}(6,6)$ can be far from 1, however the relation between this ratio and the variations of τ_1, σ_{21} is not clear. From the left graphic of Figure 2, it appears that $\Sigma^{GLS}(2,2)$ can be far from zero. According to the right graphic of Figure 2, the relative difference between $\Sigma^{OLS}(6,6)$ and $\Sigma_S^{OLS}(6,6)$ is significantly larger when the breaking points τ_1 and τ_2 are located at the end of the sample. This example shows that the standard results for the analysis of the autocovariances can be quite misleading when the unconditional homoscedasticity assumption on the innovations process does not hold.

We also consider the vector of residual autocorrelations: for a given integer $m \geq 1$, define

$$\hat{\rho}_m^{OLS} = \text{vec} \left\{ \left(\hat{R}_{OLS}(1), \dots, \hat{R}_{OLS}(m) \right) \right\} \quad \text{where} \quad \hat{R}_{OLS}(h) = \hat{S}_u^{-1} \hat{\Gamma}_{OLS}(h) \hat{S}_u^{-1}$$

with $\hat{S}_u^2 = \text{Diag}\{\hat{\sigma}_u^2(1), \dots, \hat{\sigma}_u^2(d)\}$, $\hat{\sigma}_u^2(i) = T^{-1} \sum_{t=1}^T \hat{u}_{it}^2$, and

$$\hat{\rho}_{a,m}^{GLS} = \text{vec} \left\{ \left(\hat{R}_{GLS}(1), \dots, \hat{R}_{GLS}(m) \right) \right\} \quad \text{where} \quad \hat{R}_{GLS}(h) = \hat{S}_\epsilon^{-1} \hat{\Gamma}_{GLS}(h) \hat{S}_\epsilon^{-1},$$

with $\hat{S}_\epsilon^2 = \text{Diag}\{\hat{\sigma}_\epsilon^2(1), \dots, \hat{\sigma}_\epsilon^2(d)\}$, $\hat{\sigma}_\epsilon^2(i) = T^{-1} \sum_{t=1}^T \hat{\epsilon}_{it}^2$. Since ϵ_t has identity variance matrix, we can also define

$$\hat{\rho}_{b,m}^{GLS} = \hat{\gamma}_m^{GLS}.$$

Proposition 2 *If model (1.1) is correct and Assumption A1 holds true, we have*

$$T^{\frac{1}{2}} \hat{\rho}_m^{OLS} \Rightarrow \mathcal{N}(0, \Psi^{OLS}), \quad (3.13)$$

where

$$\Psi^{OLS} = \{I_m \otimes (S_u \otimes S_u)^{-1}\} \Sigma^{OLS} \{I_m \otimes (S_u \otimes S_u)^{-1}\},$$

where $S_u^2 = \text{Diag}\{\Sigma_{G,11}, \dots, \Sigma_{G,dd}\}$. Moreover,

$$T^{\frac{1}{2}} \hat{\rho}_m^{GLS} \Rightarrow \mathcal{N}(0, \Sigma^{GLS}), \quad (3.14)$$

where $\hat{\rho}_m^{GLS}$ stands for any of $\hat{\rho}_{a,m}^{GLS}$ or $\hat{\rho}_{b,m}^{GLS}$.

Using Proposition 2, \hat{S}_u and a consistent estimator of Σ^{OLS} (that can build in a similar way to that of Δ_m^{OLS} , see Section 4, p. 18), one can easily build a consistent estimate of Ψ^{OLS} and confidence intervals for the OLS residual autocorrelations.

4 Modified portmanteau tests based on OLS estimation

Corrected portmanteau tests based on the OLS residual autocorrelations are proposed below. We use the standard Box-Pierce statistic, Box and Pierce (1970), introduced in the VAR framework by Chitturi (1974)

$$\begin{aligned} Q_m^{OLS} &= T \sum_{h=1}^m \text{tr} \left(\hat{\Gamma}'_{OLS}(h) \hat{\Gamma}_{OLS}^{-1}(0) \hat{\Gamma}_{OLS}(h) \hat{\Gamma}_{OLS}^{-1}(0) \right) \\ &= T \hat{\gamma}_m^{OLS'} \left(I_m \otimes \hat{\Gamma}_{OLS}^{-1}(0) \otimes \hat{\Gamma}_{OLS}^{-1}(0) \right) \hat{\gamma}_m^{OLS}. \end{aligned} \quad (4.1)$$

We also consider the Ljung-Box statistic (Ljung and Box (1978)) introduced in the VAR framework by Hosking (1980)

$$\tilde{Q}_m^{OLS} = T^2 \sum_{h=1}^m (T-h)^{-1} \text{tr} \left(\hat{\Gamma}'_{OLS}(h) \hat{\Gamma}_{OLS}^{-1}(0) \hat{\Gamma}_{OLS}(h) \hat{\Gamma}_{OLS}^{-1}(0) \right).$$

The following result, a direct consequence of Proposition 1 equation (3.13), provides the asymptotic distribution of Q_m^{OLS} and \tilde{Q}_m^{OLS} .

Theorem 4.1 *If model (1.1) is correct and Assumption A1 holds true, the statistics*

Q_m^{OLS} and \tilde{Q}_m^{OLS} converge in law to

$$U(\delta_m^{OLS}) = \sum_{i=1}^{d^2 m} \delta_i^{ols} U_i^2, \quad (4.2)$$

as $T \rightarrow \infty$, where $\delta_m^{OLS} = (\delta_1^{ols}, \dots, \delta_{d^2 m}^{ols})'$ is the vector of the eigenvalues of the matrix

$$\Delta_m^{OLS} = (I_m \otimes \Sigma_G^{-1/2} \otimes \Sigma_G^{-1/2}) \Sigma^{OLS} (I_m \otimes \Sigma_G^{-1/2} \otimes \Sigma_G^{-1/2}),$$

$\Sigma_G = \int_0^1 \Sigma(r) dr$ and the U_i 's are independent $\mathcal{N}(0, 1)$ variables.

When the error process is homoscedastic i.i.d. and m is large, it is well known that the asymptotic distribution of the statistics Q_m^{OLS} and \tilde{Q}_m^{OLS} under the null hypothesis \mathcal{H}_0 can be approximated by a chi-square law with $d^2(m - p)$ degrees of freedom, see Box and Pierce (1970). In our framework, even for large m , the limit distribution in (4.2) can be very different from a chi-square law. The following example illustrate this point.

Example 4.1 Consider the bivariate process in Example 3.1. Then

$$\Delta_m^{OLS} = \text{diag}\{0_{4 \times 4}, I_{m-1} \otimes \check{\Delta}_{OLS}\}$$

with

$$\check{\Delta}_{OLS} = \begin{pmatrix} \frac{\int_0^1 \Sigma_1(r)^2 dr}{(\int_0^1 \Sigma_1(r) dr)^2} & 0 & 0 & 0 \\ 0 & \frac{\int_0^1 \Sigma_1(r) \Sigma_2(r) dr}{\int_0^1 \Sigma_1(r) dr \int_0^1 \Sigma_2(r) dr} & 0 & 0 \\ 0 & 0 & \frac{\int_0^1 \Sigma_1(r) \Sigma_2(r) dr}{\int_0^1 \Sigma_2(r) dr \int_0^1 \Sigma_1(r) dr} & 0 \\ 0 & 0 & 0 & \frac{\int_0^1 \Sigma_2(r)^2 dr}{(\int_0^1 \Sigma_2(r) dr)^2} \end{pmatrix}.$$

If we suppose that $\Sigma(r)$ is constant and $A_0 = 0_{2 \times 2}$, we obtain $\check{\Delta}_{OLS} = I_4$, so that the asymptotic distribution of Q_m^{OLS} and \tilde{Q}_m^{OLS} is $\chi^2(d^2(m - p))$ with $p = 1$ and

$d = 2$. However it is easy to see that the $d^2(m - p)$ non zero diagonal elements in Δ_m^{OLS} can be far from one if the error process is heteroscedastic. From the Jensen inequality the components $\check{\Delta}_{OLS}(1, 1)$ and $\check{\Delta}_{OLS}(4, 4)$ are greater or equal than one. For illustration, in the right graphics of Figures 1 and 2 we present the second diagonal element of $\check{\Delta}_{OLS}$ when the volatility function is like in (3.11)-(3.12).

Estimates of the weights which appear in (4.2) can be obtained as follows. First, let us recall the following results proved by Patilea and Raïssi (2010):

$$\hat{\Sigma}_{G^{\otimes 2}} := T^{-1} \sum_{t=2}^T \hat{u}_{t-1} \hat{u}'_{t-1} \otimes \hat{u}_t \hat{u}'_t = \Sigma_{G^{\otimes 2}} + o_p(1), \quad (4.3)$$

$$\hat{\Sigma}_G := T^{-1} \sum_{t=1}^T \hat{u}_t \hat{u}'_t = \Sigma_G + o_p(1), \quad (4.4)$$

$$\hat{\Lambda}_2 := T^{-1} \sum_{t=1}^T \tilde{X}_{t-1} \tilde{X}'_{t-1} \otimes \hat{u}_t \hat{u}'_t = \Lambda_2 + o_p(1), \quad (4.5)$$

and

$$\hat{\Lambda}_3 := \hat{\Sigma}_{\tilde{X}} = \Lambda_3 + o_p(1). \quad (4.6)$$

A consistent estimator of Φ_m^u and $\Lambda_m^{u,\theta}$ given in (3.1) and (3.2) is easily obtained by replacing A_{01}, \dots, A_{0p} with their OLS estimators in K and using (4.3) and (4.4). Thus from this and the equations (4.3) to (4.6), one can easily define a consistent estimator of Δ_m^{OLS} . Denote the estimated eigenvalues of Δ_m^{OLS} by $\hat{\delta}_m^{OLS} = (\hat{\delta}_1^{ols}, \dots, \hat{\delta}_{d^2 m}^{ols})'$.

We are now ready to introduce the OLS residuals-based corrected versions of the Box-Pierce (resp. Ljung-Box) portmanteau tests for testing the order of the VAR model (1.1). With at hand a vector $\hat{\delta}_m^{OLS}$, at the asymptotic level α , the Box-Pierce (resp. Ljung-Box) procedure consists in rejecting the null hypothesis (1.2) of uncorrelated innovations when

$$P(Q_m^{OLS} > U_{OLS}(\hat{\delta}_m^{OLS}) \mid X_1, \dots, X_T) < \alpha$$

(resp. $P(\tilde{Q}_m^{OLS} > U_{OLS}(\hat{\delta}_m^{OLS}) \mid X_1, \dots, X_T) < \alpha$). The p -values can be evaluated using the Imhof algorithm (Imhof, 1961) or the saddle point method, see e.g. Kuonen (1999).

Let us end this section with some remarks on the particular case $\Sigma(\cdot) = \sigma^2(\cdot)I_d$ (that includes the univariate AR(p) models with time-varying variance). In this case

$$\Delta_m^{OLS} = \left[\int_0^1 \sigma^2(r) dr \right]^{-2} \Sigma^{OLS} = \left[\int_0^1 \sigma^2(r) dr \right]^{-2} \left[\int_0^1 \sigma^4(r) dr \right] \Sigma^{GLS} =: c_\sigma \Sigma^{GLS}, \quad (4.7)$$

and clearly, $c_\sigma \geq 1$. If in addition $p = 0$, by Proposition 1 we have $\Sigma^{GLS} = I_{d^2m}$ and hence $\delta_m^{OLS} = c_\sigma(1, \dots, 1)'$.

5 Adaptive portmanteau tests

An alternative way to build portmanteau tests for VAR(p) models with time-varying variance we consider herein is to use approximations of the innovation ϵ_t . A nonparametric estimate of the volatility function is needed for building such approximations. For this purpose we generalize the approach of Xu and Phillips (2008) to the multivariate case, see also Patilea and Raïssi (2010). 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

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

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 , $i = 1, \dots, T$ are almost sure linear independent each other, $\check{\Sigma}_t^0$ is almost sure positive definite provided T is sufficiently large. When using several bandwidths b_{kl} a regularization of $\check{\Sigma}_t^0$ could be necessary in order to ensure positive definiteness. Let us consider

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

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 experiments indicate 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

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

with respect to all $b_{kl} \in \mathcal{B}_T$, $1 \leq k \leq l \leq d$, where $\|\cdot\|$ 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. Like in Patilea and Raïssi (2010), the theoretical results below are obtained uniformly with respect to the bandwidths $b_{kl} \in \mathcal{B}_T$ and this provides a justification for the common cross-validation bandwidth selection approach in the framework we consider.

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

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

with

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

The ALS residuals, proxies of the infeasible GLS residuals, are defined as $\check{\epsilon}_t = \check{H}_t^{-1} X_t - \check{H}_t^{-1} (\check{X}'_{t-1} \otimes I_d) \hat{\theta}_{ALS}$, and the adaptive autocovariances and autocorrelations

$$\hat{\Gamma}_{ALS}(h) = \hat{\Gamma}_{ALS}^\epsilon(h) = T^{-1} \sum_{t=h+1}^T \check{\epsilon}_t \check{\epsilon}'_{t-h}, \quad \hat{R}_{ALS}(h) = \check{S}_\epsilon^{-1} \hat{\Gamma}_{ALS}(h) \check{S}_\epsilon^{-1},$$

where $\check{S}_\epsilon = \text{Diag}\{\check{\sigma}_\epsilon(1), \dots, \check{\sigma}_\epsilon(d)\}$, $\check{\sigma}_\epsilon^2(i) = T^{-1} \sum_{t=1}^T \check{\epsilon}_{it}^2$, and \check{H}_t is the nonparametric estimator obtained from $\check{\Sigma}_t$ and the identification condition on H_t (see Assumption **A1**(i)), that is $\check{H}_t = \check{\Sigma}_t^{1/2}$.

Let $\hat{\gamma}_m^{ALS} = \text{vec}\{\hat{\Gamma}_{ALS}(1), \dots, \hat{\Gamma}_{ALS}(m)\}$. Following the notation of the previous section, for a given integer $m \geq 1$, define the residual autocorrelations

$$\hat{\rho}_{a,m}^{ALS} = \text{vec} \left\{ \left(\hat{R}_{ALS}(1), \dots, \hat{R}_{ALS}(m) \right) \right\} \quad \text{and} \quad \hat{\rho}_{b,m}^{ALS} = \hat{\gamma}_m^{ALS}.$$

The main result of this section shows that $\hat{\gamma}_m^{ALS}$ and $\hat{\rho}_{a,m}^{ALS}$ are asymptotic equivalent to $\hat{\gamma}_m^{GLS}$ and $\hat{\rho}_{a,m}^{GLS}$. This will allow us to define new portmanteau statistics based on the ALS residuals. For this purpose, we need the following assumptions.

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 for any symmetric matrix A the real value $\lambda_{\min}(A)$ denotes its smallest eigenvalue.
- (ii) $\sup_t \|\epsilon_{kt}\|_8 < \infty$ for all $k \in \{1, \dots, d\}$.

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 $\int_{\mathbb{R}} v^2 K(v) dv < \infty$. 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} , $1 \leq k \leq l \leq d$, are taken in the range $\mathcal{B}_T = [c_{min}b_T, c_{max}b_T]$ with $0 < c_{min} < c_{max} < \infty$ and $b_T + 1/Tb_T^{2+\gamma} \rightarrow 0$ as $T \rightarrow \infty$, for some $\gamma > 0$.

(iii) The sequence ν_T is such that $T\nu_T^2 \rightarrow 0$.

Below, we say that a sequence of random matrices A_T , $T \geq 1$ is $o_p(1)$ uniformly with respect to (w.r.t.) $b_{kl} \in \mathcal{B}_T$ as $T \rightarrow \infty$ if $\sup_{1 \leq k \leq l \leq d} \sup_{b_{kl} \in \mathcal{B}_T} \|\text{vec}(A_T)\| \xrightarrow{P} 0$. The following proposition gives the asymptotic behavior of variances, autocovariances and autocorrelations estimators based on the ALS estimator of θ_0 and the nonparametric estimate of the time-varying variance structure Σ_t . The results are uniformly w.r.t the bandwidths.

Proposition 3 *If model (1.1) is correct and Assumptions A1' and A2 hold, uniformly w.r.t. $b \in \mathcal{B}_T$*

$$T^{-1} \sum_{t=1}^T \check{H}'_t \otimes \check{H}_t^{-1} = \int_0^1 G(r)' \otimes G(r)^{-1} dr + o_p(1), \quad (5.1)$$

$$\check{\Sigma}_{\check{X}} = \Lambda_1 + o_p(1). \quad (5.2)$$

Moreover, given any $m \geq 1$,

$$T^{\frac{1}{2}} \{ \hat{\gamma}_m^{ALS} - \hat{\gamma}_m^{GLS} \} = o_p(1) \quad \text{and} \quad T^{\frac{1}{2}} \{ \hat{\rho}_m^{ALS} - \hat{\rho}_m^{GLS} \} = o_p(1), \quad (5.3)$$

where $\hat{\rho}_m^{ALS}$ (resp. $\hat{\rho}_m^{GLS}$) stands for any of $\hat{\rho}_{a,m}^{ALS}$ and $\hat{\rho}_{b,m}^{ALS}$ (resp. $\hat{\rho}_{a,m}^{GLS}$ and $\hat{\rho}_{b,m}^{GLS}$).

This asymptotic equivalence result allows us to propose portmanteau test statistics adapted to the case of time-varying variance. Consider the Box-Pierce type

statistic

$$\begin{aligned} Q_{a,m}^{ALS} &= T \sum_{h=1}^m \text{tr} \left(\hat{\Gamma}'_{ALS}(h) \hat{\Gamma}_{ALS}^{-1}(0) \hat{\Gamma}_{ALS}(h) \hat{\Gamma}_{ALS}^{-1}(0) \right) \\ &= T \hat{\gamma}_m^{ALS'} \left(I_m \otimes \hat{\Gamma}_{ALS}^{-1}(0) \otimes \hat{\Gamma}_{ALS}^{-1}(0) \right) \hat{\gamma}_m^{ALS}, \end{aligned}$$

and

$$Q_{b,m}^{ALS} = T \hat{\rho}_{b,m}^{ALS'} \hat{\rho}_{b,m}^{ALS}.$$

Consider also the Ljung-Box type statistics

$$\tilde{Q}_{a,m}^{ALS} = T^2 \sum_{h=1}^m (T-h)^{-1} \text{tr} \left(\hat{\Gamma}'_{ALS}(h) \hat{\Gamma}_{ALS}^{-1}(0) \hat{\Gamma}_{ALS}(h) \hat{\Gamma}_{ALS}^{-1}(0) \right)$$

and

$$\tilde{Q}_{b,m}^{ALS} = T^2 \sum_{h=1}^m (T-h)^{-1} \text{tr} \left(\hat{\Gamma}'_{ALS}(h) \hat{\Gamma}_{ALS}(h) \right).$$

The following theorem is a direct consequence of (3.7) and Proposition 3 and hence the proof is omitted.

Theorem 5.1 *Under the assumptions of Proposition 3, the statistics $Q_{a,m}^{ALS}$, $Q_{b,m}^{ALS}$ and $\tilde{Q}_{a,m}^{ALS}$, $\tilde{Q}_{b,m}^{ALS}$ converge in distribution to*

$$U(\delta_m^{ALS}) = \sum_{i=1}^{d^2 m} \delta_i^{als} U_i^2, \quad (5.4)$$

as $T \rightarrow \infty$, where $\delta_m^{ALS} = (\delta_1^{als}, \dots, \delta_{d^2 m}^{als})'$ is the vector of the eigenvalues of Σ^{GLS} , and the U_i 's are independent $\mathcal{N}(0,1)$ variables.

To compute the critical values of the adaptive portmanteau tests, we first obtain a consistent estimator of $\Lambda_m^{c,\theta}$ given in (3.3) by replacing A_{01}, \dots, A_{0p} by their ALS estimators in K and using (5.1). Next we consider the estimate of Λ_1 given in (5.2). Plugging these estimates into the formula (3.8), we obtain a consistent estimator of Σ^{GLS} with eigenvalues $\hat{\delta}_m^{ALS} = (\hat{\delta}_1^{als}, \dots, \hat{\delta}_{d^2 m}^{als})'$ that consistently estimate δ_m^{ALS} .

There are several important particular cases that could be mentioned. In the case of a VAR(0) model (i.e., the process (u_t) is observed), $\Sigma^{GLS} = I_{d^2m}$ (see Proposition 1) and hence the asymptotic distribution of the four test statistics in Theorem 5.1 would be $\chi_{d^2m}^2$, that means independent of the variance structure given by $\Sigma(\cdot)$. In the general case $p \geq 1$ where the autoregressive coefficients A_{0i} , $i = 1, \dots, p$ have to be estimated, the matrix $I_{d^2m} - \Sigma^{GLS}$ being positive semi-definite, the eigenvalues $\delta_1^{als}, \dots, \delta_{d^2m}^{als}$ are smaller or equal to 1. Since, in some sense, the unconditional heteroscedasticity is removed in the ALS residuals, one could expect that the $\chi_{d^2(m-p)}^2$ asymptotic approximation is reasonably accurate for the ALS tests. Example 3.1 indicates that this is may not the case, the asymptotic distribution we obtain for the ALS portmanteau statistics can be very different from the $\chi_{d^2(m-p)}^2$ approximation when the errors are heteroscedastic. Finally note that Patilea and Raïssi (2010) pointed out that using the adaptive estimators of autoregressive parameters instead of the OLS estimators lead to a gain of efficiency, so that it is advisable to compute the kernel smoothing estimator of the variance function $\Sigma(\cdot)$ at the estimation stage. Therefore since the kernel estimator of the variance Σ_t is available for the validation stage, the ALS tests are not more complicated than the OLS tests to implement.

Let us also point out that the eigenvalues $\delta_1^{als}, \dots, \delta_{d^2m}^{als}$ will not depend on the variance structure when $\Sigma(\cdot) = \sigma^2(\cdot)I_d$ (in particular in the univariate case), whatever the value of p is. Moreover, using the arguments of Box and Pierce (1970), see also Brockwell and Davis (1991, pp. 310–311), one can easily show that for large values of m , the law of $U(\delta_m^{ALS})$ is accurately approximated by a $\chi_{d^2(m-p)}^2$ distribution. However, in the general the multivariate setup the asymptotic distribution in (5.4) depend on the variance function $\Sigma(\cdot)$.

6 Modified portmanteau statistics with standard chi-square asymptotic distributions

In the previous sections we considered portmanteau tests for which the asymptotic critical values are given by weighted sums of chi-squares in the general VAR(p) case. Using a suitable change of our quadratic forms one can propose alternative portmanteau test statistics with chi-squared asymptotic distributions under the null hypothesis. This type of modification was already proposed in the recent time series literature but in different contexts.

First note that as remarked above when testing that the observed process is uncorrelated ($p = 0$) and using the standard portmanteau statistic (4.1) we obtain a non standard asymptotic distribution. Then following the approach of Lobato, Nankervis and Savin (2002) we consider the modified portmanteau test statistic

$$\underline{Q}_m^{OLS} = T \hat{\gamma}_m^{OLS'} \left(\hat{\Lambda}_m^{u,u} \right)^{-1} \hat{\gamma}_m^{OLS},$$

where $\hat{\Lambda}_m^{u,u} = I_m \otimes \hat{\Sigma}_{G^{\otimes 2}}$ with $\hat{\Sigma}_{G^{\otimes 2}}$ defined in equation (4.3). The invertibility of $\hat{\Lambda}_m^{u,u}$ is guaranteed asymptotically by our assumptions. In view of Proposition 1 it is clear that under the null hypothesis of uncorrelated observed process, the asymptotic distribution of the \underline{Q}_m^{OLS} statistic is $\chi_{d^2 m}^2$. Recall that this kind of statistic correction is not necessary to obtain a standard asymptotic distribution for the adaptive portmanteau tests when the non correlation of the observed process is tested.

This approach can be generalized to the case of VAR(p) models with possibly $p > 0$ using the approach of Katayama (2008) for building tests with standard asymptotic distributions. In this part we take $p < m < T$. Let us introduce

$$D_m^{OLS} = \Phi_m \left\{ \Phi_m' (\Lambda_m^{u,u})^{-1} \Phi_m \right\}^{-1} \Phi_m' (\Lambda_m^{u,u})^{-1}$$

$$D_m^{GLS} = \Lambda_m^{\epsilon, \theta} \left\{ \Lambda_m^{\epsilon, \theta'} \Lambda_m^{\epsilon, \theta} \right\}^{-1} \Lambda_m^{\epsilon, \theta'}$$

so that $(I_{d^2m} - D_m^{OLS})\Phi_m = 0$ and $(I_{d^2m} - D_m^{GLS})\Lambda_m^{\epsilon, \theta} = 0$. From the proof of Proposition 1 (equation (12.17)), it is easy to see that

$$(I_{d^2m} - D_m^{OLS})T^{\frac{1}{2}}\hat{\gamma}_m^{OLS} = (I_{d^2m} - D_m^{OLS})T^{\frac{1}{2}}c_m^u + o_p(1)$$

where $T^{1/2}c_m^u$ is asymptotically normal of mean 0 and variance $\Lambda_m^{u,u}$. Deduce that

$$(I_{d^2m} - D_m^{OLS})T^{\frac{1}{2}}\hat{\gamma}_m^{OLS} \Rightarrow \mathcal{N}(0, V),$$

where $V = (I_{d^2m} - D_m^{OLS})\Lambda_m^{u,u}(I_{d^2m} - D_m^{OLS})'$. Now, notice that

$$(I_{d^2m} - D_m^{OLS})\Lambda_m^{u,u} = \Lambda_m^{u,u} - \Phi_m \left\{ \Phi_m' (\Lambda_m^{u,u})^{-1} \Phi_m \right\}^{-1} \Phi_m' = \Lambda_m^{u,u} (I_{d^2m} - D_m^{OLS})'.$$

From this and the fact that $I_{d^2m} - D_m^{OLS}$ is a projector, deduce that the matrix AV is idempotent, where

$$A = (I_{d^2m} - D_m^{OLS})' (\Lambda_m^{u,u})^{-1} (I_{d^2m} - D_m^{OLS}).$$

Moreover, since Φ_m is of full column rank d^2p , it is easy to see that the rank of A is $d^2(m-p)$. A classical result in multivariate data analysis implies

$$T\hat{\gamma}_m^{OLS'} (I_{d^2m} - D_m^{OLS})' (\Lambda_m^{u,u})^{-1} (I_{d^2m} - D_m^{OLS})\hat{\gamma}_m^{OLS} \Rightarrow \chi_{d^2(m-p)}^2. \quad (6.1)$$

Similarly we obtain

$$(I_{d^2m} - D_m^{GLS})T^{\frac{1}{2}}\hat{\gamma}_m^{GLS} \Rightarrow \mathcal{N}(0, I_{d^2m} - D_m^{GLS}) \quad (6.2)$$

and we deduce

$$T\hat{\gamma}_m^{GLS'} (I_{d^2m} - D_m^{GLS})\hat{\gamma}_m^{GLS} \Rightarrow \chi_{d^2(m-p)}^2. \quad (6.3)$$

The matrices D_m^{OLS} and $\Lambda_m^{u,u}$ could be estimated as suggested in Section 4, see comments after equation (4.6), and hence a modified portmanteau test statistic

based on the OLS estimates $\hat{\gamma}_m^{OLS}$ and having standard chi-square critical values could be derived from equation (6.1). On the other hand, using a nonparametric estimate of H_t one could easily estimate D_m^{GLS} , see Proposition 3 and the comments after Theorem 5.1. Moreover, Proposition 3 allows us to replace $\hat{\gamma}_m^{GLS}$ with $\hat{\gamma}_m^{ALS}$ and thus to introduce an adaptive portmanteau test with a modified statistic and standard chi-square critical values based on equation (6.3). Clearly one can consider similar modification for Ljung-Box type statistics.

The chi-square critical values are certainly more convenient for portmanteau tests. Moreover, in section 7 we provide evidence that the test based on the statistic (6.1) could be more powerful, in the Bahadur slope sense, than the OLS estimates based test based on the Q_m^{OLS} statistic investigated in Theorem 1.1. However, it is not necessarily true that the modified procedures presented in this section are preferable in applications. Indeed, the empirical evidence presented in Section 9 shows that test statistics like in (6.1) and (6.3) are unstable and induce bad levels even with series of few hundred observations.

7 Testing for autocorrelation in heteroscedastic series: some theoretical power comparisons

In this part we carry out some theoretical power comparisons for the tests we considered above in the important case where the non correlation of the observed process $X_t = u_t$ is tested. The case $p \geq 1$ will be considered elsewhere. On one hand we compare the classical Box-Pierce portmanteau test and modified quadratic forms of OLS residual autocorrelations based test introduced in section 6. On the other hand we compare ALS and OLS residual autocorrelations based portmanteau tests. For

this purpose we use the Bahadur slope approach that we briefly recall here. Let Q_A denote a test statistic and, for any $x > 0$, define $q_A(x) = -\log P_0(Q_A > x)$ where P_0 stands for the limit distribution of Q_A under the null hypothesis. Following Bahadur (1960) (see also van der Vaart 1998, chapter 14), consider the (*asymptotic*) slope $c_A(\varrho) = 2 \lim_{T \rightarrow \infty} T^{-1} q_A(Q_A)$ under a *fixed* alternative \mathcal{H}_1 such that the limit exists in probability. The asymptotic relative efficiency of the test based on Q_A with respect to a competing test based on a test statistic Q_B is then defined as the ratio $ARE_{A,B}(\varrho) = c_A(\varrho)/c_B(\varrho)$. A relative efficiency $ARE_{A,B}(\varrho) \geq 1$ suggests that the test given by Q_A is better suited to detect \mathcal{H}_1 because the associated p -values wanes faster or equally faster compared to the p -values of the test based on Q_B .

For the sake of simplicity we restrict our attention to the BP statistics and consider the case where one tests the non correlation of the observed process, while the underlying true process is the autoregressive process of order 1

$$u_t = Bu_{t-1} + \tilde{H}_t \epsilon_t, \quad (7.1)$$

where $\det(I_d - Bz) \neq 0$ for all $|z| \leq 1$ and $B \neq 0$. We keep the notation $E(X_t X_t') = E(u_t u_t') = \Sigma_t$ and we introduce $E(\tilde{H}_t \epsilon_t \epsilon_t' \tilde{H}_t') = \tilde{H}_t \tilde{H}_t' := \tilde{\Sigma}_t$. Under the alternative hypothesis (7.1) we have the relationship

$$\Sigma_t = \sum_{i=0}^{\infty} B^i \tilde{\Sigma}_{t-i} B^{i'}. \quad (7.2)$$

Using similar arguments to that of the proofs of Lemmas 12.1 to 12.3 in the Appendix, deduce that

$$\hat{\Gamma}_{OLS}(h) = T^{-1} \sum_{t=h+1}^T u_t u_{t-h}' = B^h \int_0^1 \Sigma(r) dr + o_p(1) \quad (7.3)$$

and

$$\hat{\Gamma}_{GLS}(h) = T^{-1} \sum_{t=h+1}^T H_t^{-1} u_t u_{t-h}' H_{t-h}^{-1'} = \int_0^1 G(r)^{-1} B^h G(r) dr + o_p(1). \quad (7.4)$$

Using basic properties of the $\text{vec}(\cdot)$ operator and the Kronecker product we obtain

$$T^{-1}Q_m^{OLS} = \mathcal{B}' \left\{ I_m \otimes \int_0^1 \Sigma(r) dr \otimes \left(\int_0^1 \Sigma(r) dr \right)^{-1} \right\} \mathcal{B} + o_p(1)$$

$$T^{-1}\underline{Q}_m^{OLS} = \mathcal{B}' \left\{ I_m \otimes \left(\int_0^1 \Sigma(r) dr \otimes I_d \right) \left(\int_0^1 \Sigma(r) \otimes \Sigma(r) dr \right)^{-1} \left(\int_0^1 \Sigma(r) dr \otimes I_d \right) \right\} \mathcal{B} + o_p(1)$$

and

$$T^{-1}Q_{i,m}^{ALS} = \mathcal{B}' \left\{ I_m \otimes \left(\int_0^1 G(r)' \otimes G(r)^{-1} dr \right)^2 \right\} \mathcal{B} + o_p(1),$$

with $i \in \{a, b\}$ and $\mathcal{B} = \text{vec} \{(B^1, \dots, B^m)\}$.

Proposition 4 (i) *If Assumption **A1** holds true and the observations follow the model (7.1), the asymptotic relative efficiency of the portmanteau test based on \underline{Q}_m^{OLS} with respect to the portmanteau tests based on Q_m^{OLS} is larger or equal to 1.*

(ii) *Suppose that $\Sigma(\cdot) = \sigma^2(\cdot)I_d$ where $\sigma(\cdot)$ is some positive scalar function. Suppose that Assumptions **A1'** and **A2** holds true and the observations follow the model (7.1). Then asymptotic relative efficiencies of the portmanteau test based on Q_m^{ALS} with respect to the portmanteau tests based on Q_m^{OLS} or \underline{Q}_m^{OLS} are larger or equal to 1.*

In the first part of Proposition 4 the result is obtained without additional restriction on $\Sigma(\cdot)$ while in the second part we impose $\Sigma(\cdot) = \sigma^2(\cdot)I_d$ which is for instance true in the univariate case ($d = 1$). In the general multivariate case the portmanteau test based on ALS residual autocorrelations does not necessarily outperforms, in the sense of the Bahadur slope, the tests based on the OLS residual autocorrelations considered above.

8 Extending the scope: testing the order of a heteroscedastic co-integration model

Consider the case of a unit root multivariate process (X_t) with time-varying volatility, see for instance Cavaliere, Rahbek and Taylor (2010) or Boswijk (2010). With $Z_{0t} := X_t - X_{t-1}$ the model (1.1) can be rewritten in its error correction form

$$Z_{0t} = \Pi_{01}X_{t-1} + \sum_{i=2}^p \Pi_{0i}Z_{0t-i+1} + u_t \quad (8.1)$$

$$u_t = H_t\epsilon_t.$$

The matrices Π_{0i} are functions of the A_i 's, and such that the assumptions of the Granger representation theorem hold (see for instance Assumption 1 of Cavaliere, Rahbek and Taylor (2010)), $\Pi_{01} = \alpha_0\beta_0'$ where the $d \times s$ -dimensional matrices α_0 and β_0 are identified in some appropriate way (see e.g. Johansen 1995, p. 72, for the identification problem). If $p = 1$ the sum in (8.1) vanishes. In this section we follow Cavaliere, Rahbek and Taylor (2010) and we slightly strengthen **A1** assuming that (ϵ_t) is iid. Then it follows from their Lemma 1 that (X_t) have a random walk behavior and also that $(\beta_0'X_t)$ is stable. By analogy with the homoscedastic case, the number s of independent linear stable combinations in $(\beta_0'X_t)$ is the cointegrating rank (see section 2.3 of Cavaliere, Rahbek and Taylor 2010, for a detailed discussion on the concept of cointegration in our framework). If $s = 0$ the process (X_t) is not cointegrated and the procedures described in the previous sections apply directly to the process (Z_{0t}) . Many contributions in the literature that considered the standard homoscedastic framework pointed out that the choice of the lag length is important for the cointegrating rank analysis, see e.g. Boswijk and Franses (1992, section 4). It seems reasonable to imagine that a similar remark remains true with a time-varying variance.

To describe the estimation procedure of (8.1), let us define $Z_{1t}(\beta) = (X'_{t-1}\beta, Z'_{0t-1}, \dots, Z'_{0t-p+1})'$ for any $d \times s$ -matrix β and rewrite model (8.1) under the form

$$Z_{0t} = (Z_{1t}(\beta_0)' \otimes I_d)\vartheta_0 + u_t,$$

where $\vartheta_0 = \text{vec}\{(\alpha_0, \Pi_{02}, \dots, \Pi_{0p})\}$. The estimator of the long run relationships $\hat{\beta}$ can be obtained using the reduced rank regression method introduced by Anderson (1951). Cavaliere *et al* (2010) showed that in our framework

$$T(\hat{\beta} - \beta) = O_p(1). \quad (8.2)$$

Now, let us define

$$\hat{\vartheta}_{OLS}(\beta) = \hat{\Sigma}_{Z_1}^{-1}(\beta) \text{vec} \left(\hat{\Sigma}_{Z_0}(\beta) \right),$$

where

$$\hat{\Sigma}_{Z_1}(\beta) = T^{-1} \sum_{t=1}^T Z_{1t}(\beta) Z_{1t}(\beta)' \otimes I_d \quad \text{and} \quad \hat{\Sigma}_{Z_0}(\beta) = T^{-1} \sum_{t=1}^T Z_{0t} Z_{1t}(\beta)',$$

and similarly to (2.2) let us introduce

$$\hat{\vartheta}_{GLS}(\beta) = \hat{\Sigma}_{Z_1}(\beta)^{-1} \text{vec} \left(\hat{\Sigma}_{Z_0}(\beta) \right),$$

with

$$\hat{\Sigma}_{Z_1} = T^{-1} \sum_{t=1}^T Z_{1t}(\beta) Z_{1t}(\beta)' \otimes \Sigma_t^{-1}, \quad \hat{\Sigma}_{Z_0} = T^{-1} \sum_{t=1}^T \Sigma_t^{-1} Z_{0t} Z_{1t}(\beta)',$$

where the volatility Σ_t is assumed known. Next, for any fixed β , let us define the estimated residuals

$$\hat{u}_t(\beta) = Z_{0t} - (Z_{1t}(\beta)' \otimes I_d) \hat{\vartheta}_{OLS}(\beta),$$

$$\hat{\epsilon}_t(\beta) = H_t^{-1} Z_{0t} - H_t^{-1} (Z_{1t}(\beta)' \otimes I_d) \hat{\vartheta}_{GLS}(\beta)$$

and the corresponding estimated autocovariance matrices

$$\hat{\Gamma}_{OLS}(h, \beta) = T^{-1} \sum_{t=h+1}^T \hat{u}_t(\beta) \hat{u}_{t-h}(\beta)' \quad \text{and} \quad \hat{\Gamma}_{GLS}(h, \beta) = T^{-1} \sum_{t=h+1}^T \hat{\varepsilon}_t(\beta) \hat{\varepsilon}_{t-h}(\beta)'.$$

From (8.2) we obviously have

$$\hat{\Gamma}_{OLS}(h, \hat{\beta}) = \hat{\Gamma}_{OLS}^u(h, \beta_0) + o_p(T^{-\frac{1}{2}}) \quad \text{and} \quad \hat{\Gamma}_{GLS}(h, \hat{\beta}) = \hat{\Gamma}_{GLS}^u(h, \beta_0) + o_p(T^{-\frac{1}{2}}).$$

Defining

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

it is also obvious that

$$\check{\Sigma}_t^0(\hat{\beta}) = \check{\Sigma}_t^0(\beta_0) + o_p(T^{-\frac{1}{2}}).$$

It is clear now that one can treat β_0 as known and, following the lines of the previous sections, one can use $\hat{\Gamma}_{OLS}(h, \hat{\beta})$ and the ALS version of $\hat{\Gamma}_{GLS}(h, \hat{\beta})$ to build portmanteau tests for checking the order p of the model (8.1).

9 Monte Carlo experiments

In the sequel LB_m^{OLS} and LB_m^{ALS} will denote the Ljung-Box type portmanteau tests based on the adaptive approach with non standard distributions (4.2) and (5.4). For the sake of brevity, only the results obtained with the test statistic $\tilde{Q}_{a,m}^{ALS}$ will be reported. Moreover, since we found similar results for the BP and LB tests, we only report on LB tests. The LB tests based on modified statistics which are built using the results in Section 6 are denoted by \widetilde{LB}_m^{OLS} and \widetilde{LB}_m^{ALS} . If we assume that the volatility function is known, one can also build portmanteau tests using the result in (3.14) in a similar way to the adaptive portmanteau tests. These infeasible tests denoted by LB_m^{GLS} and \widetilde{LB}_m^{GLS} will serve as a benchmark for comparison with the ALS tests LB_m^{ALS} and \widetilde{LB}_m^{ALS} . It is clear that the asymptotic critical values of

the ALS tests are the same as the critical values of the GLS tests. In this section we investigate by simulations the finite sample properties of the ALS and GLS portmanteau tests and we compare them with the OLS estimation-based tests. In the next section we study the model adequacy of two real data sets: the US energy and transportation price indexes for all urban consumers on one hand and the US balances on merchandise trade and on services on the other hand.

9.1 Empirical size

Our Monte Carlo experiments are based on the following Data Generating Process (DGP) specification

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} 0.3 & -0.3 \\ 0 & -0.1 \end{pmatrix} \begin{pmatrix} X_{1t-1} \\ X_{2t-1} \end{pmatrix} + \begin{pmatrix} \mathbf{a} & 0 \\ 0 & \mathbf{a} \end{pmatrix} \begin{pmatrix} X_{1t-2} \\ X_{2t-2} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}, \quad (9.1)$$

where $\mathbf{a} = 0$ in the empirical size part of the study and $\mathbf{a} = -0.3$ in the empirical power part. The autoregressive parameters are such that the stability condition hold and are inspired from the ALS estimation obtained for the U.S. balance on services and merchandise trade data (see Table 5 below). In the case of smooth variance structure we consider

$$\Sigma(r) = \begin{pmatrix} (1 + \pi_1 r)(1 + \varpi^2) & \varpi(1 + \pi_1 r)^{\frac{1}{2}}(0.1 + \pi_2 r)^{\frac{1}{2}} \\ \varpi(1 + \pi_1 r)^{\frac{1}{2}}(0.1 + \pi_2 r)^{\frac{1}{2}} & (0.1 + \pi_2 r) \end{pmatrix}, \quad (9.2)$$

where we take $\pi_1 = 250$ and $\pi_2 = 5$. In order to investigate the properties of the tests when a volatility break is present we also consider the following specification

$$\Sigma(r) = \begin{pmatrix} (6 + f_1(r))(1 + \varpi^2) & \varpi(6 + f_1(r))^{\frac{1}{2}}(0.5 + f_2(r))^{\frac{1}{2}} \\ \varpi(0.5 + f_2(r))^{\frac{1}{2}}(6 + f_1(r))^{\frac{1}{2}} & (0.5 + f_2(r))(1 + \rho^2) \end{pmatrix}, \quad (9.3)$$

with $f_1(r) = 54 \times \mathbf{1}_{(r \geq 1/2)}(r)$ and $f_2(r) = 3 \times \mathbf{1}_{(r \geq 1/2)}(r)$. In this case we have a common volatility break at the date $t = T/2$. In all experiments we fix $\varpi = 0.2$. These volatility specifications are inspired by the real data studies we consider in the next section. For instance to fix the $(1, 1)$ -component of the specification (9.2) we noticed that the last estimated variances with the balance services and merchandise trade data are all greater than 200. In the energy-transport price indexes data some of the last estimated variances of the first component are even greater than 300. The amplitudes of the functions $f_1(\cdot)$ and $f_2(\cdot)$ in the volatility specification with an abrupt break defined in (9.3) were calibrated close to the means of the first $T/2$ estimated volatilities and of the $T/2$ last volatilities for the balance services and merchandise trade data. To assess the finite sample properties of the tests under comparison when the errors are stationary, we also considered standard i.i.d. Gaussian error processes. For each experiment $N = 1000$ independent trajectories are simulated using DGP (9.1). Samples of length $T = 50$, $T = 100$ and $T = 200$ are simulated.

We first study the empirical size of the tests taking $\mathbf{a} = 0$, and adjusting a VAR(1) model to the simulated processes. The portmanteau tests for the non correlation of the error terms are applied using $m = 5$ and $m = 15$ at the asymptotic nominal level 5%. The results are given in Tables 1, 2 and 3. Since $N = 1000$ replications are performed and assuming that the finite sample size of the tests is 5%, the relative rejection frequencies should be between the significant limits 3.65% and 6.35% with probability 0.95. Then the relative rejection frequencies are displayed in bold type when they are outside these significant limits. Note that the distributions (4.2) and (5.4) are given for fixed m , while the $\chi_{d^2(m-p)}^2$ approximation (see discussion after Theorem 4.1 above) should be accurate only for large m . Therefore we only

comment the results for small and moderate samples ($T = 50$ and $T = 100$) for the standard portmanteau tests with $m = 5$, while the results for large samples ($T = 200$) are considered when $m = 15$ is taken.

In Table 1 i.i.d. standard Gaussian errors are used, that means $\Sigma(\cdot) \equiv I_2$. In this simple case the relative rejection frequencies of the tests converge to the asymptotic nominal level. In general we do not remark a major loss of efficiency of the LB_m^{ALS} test when compared to the standard test. Therefore one can use the LB_m^{ALS} test in case of doubt of the presence of unconditional heteroscedasticity. The same remark can be made for the LB_m^{OLS} test when m is small. However we note that the LB_m^{OLS} is oversized for small samples and when m is large. It also appears that the tests based on modified statistics are oversized when m is large. This can be explained by the fact that the matrices $\Lambda_m^{\epsilon, \theta'} \Lambda_m^{\epsilon, \theta}$ and $\Phi'_m (\Lambda_m^{u, u})^{-1} \Phi_m$ are difficult to invert in such situations.

In Table 2 heteroscedastic errors with an abrupt volatility break are considered, while the trending specification (9.2) is used for the volatility in Table 3. In line with the theoretical the relative rejection frequencies of the ALS, GLS and OLS tests converge to the asymptotic nominal level. As expected the standard portmanteau test is not valid. In general it emerges from this part that the LB_m^{ALS} and LB_m^{GLS} tests have similar results. Then it seems that estimating the volatility entails no major loss of efficiency when building this kind of test. We did not found clear advantage for the LB_m^{ALS} when compared to the LB_m^{OLS} in the presented experiments. However note that the asymptotic distribution (4.2) of the standard portmanteau statistic seems estimated with less precision than the asymptotic distribution of the ALS portmanteau statistic. For instance we see in Table 4 that for $m = 5$ the standard deviations of the ALS weights are lower or equal to the standard deviations of the

OLS weights. We found that this difference is more marked for $m = 15$ which may lead to problems for the control of the error of first kind for the LB_m^{OLS} as already noted in the homoscedastic case (see Table 1 for $m = 15$). Other simulation results not reported here show that the LB_m^{OLS} test can be oversized when m is large as in the homoscedastic case with $m = 15$ and we noted that the estimation of the weights requires a relatively large number of observations for the OLS approach. A possible explanation is also that $\delta_i^{als} \in [0, 1]$ while $\delta_i^{ols} \in [0, \infty)$ for $i \in \{1, \dots, d^2m\}$, which may proceed more instable estimation for δ_i^{ols} in many cases. In addition the energy-transportation price indexes example below show that the estimation of the weights may be problematic in the OLS case. Therefore we recommend to choose small m when the samples are small and use large m only when the samples are large despite the asymptotic results hold true for fixed m when the LB_m^{OLS} is considered. We again note that the tests with modified statistics are in general clearly oversized. A possible explanation is that White (1980) type correction matrices are inverted in the modified portmanteau statistics which may lead to oversized tests as pointed out in Vilasuso (2001) in the stationary case. We can conclude that with a data generating process close to our simulation design the LB_m^{ALS} test controls reasonably well the error of the first kind in all the situations we considered.

9.2 Empirical power

In the empirical power part of this section, we examine the ability of the different tests to detect underspecified autoregressive order. The power investigation is realized in the Bahadur sense, that is the sample size is increased while the alternative is kept fixed. More precisely we set $\mathbf{a} = -0.3$, and we adjusted a VAR(1) model to the simulated VAR(2) processes with $T = 50, 100, 200, 300$. We simulated $N = 1000$

independent trajectories using DGP (9.1) with standard Gaussian innovations and heteroscedastic volatility specifications (9.2) and (9.3). The non correlation of the error process is again tested at the asymptotic nominal level 5%, but taking $m = 10$ in all the experiments. In Figure 3 we consider the homoscedastic case, errors where an abrupt volatility shift is observed and errors with smoothly varying variance.

When the variance is constant it appears that the tests with non standard distribution have similar power to the standard test when the errors are homoscedastic. Therefore we again note that there is no major loss of efficiency when the tests with modified distribution are used while the variance is constant. The tests with standard distribution may seem more powerful when the samples are small, but this mainly comes from the fact that these tests are oversized. When the errors are heteroscedastic the standard test seems more powerful than the other tests. However the standard test is oversized and in this case and the comparison is again not fair. A similar comment can be made when the tests with modified distribution are compared to the tests with standard distribution. It emerges that the LB_m^{ALS} test is more powerful than the LB_m^{OLS} . The relation between the powers of \widetilde{LB}_m^{OLS} and \widetilde{LB}_m^{ALS} is not clear. Finally we can remark that in the presented experiments the GLS type tests are not necessarily more powerful than the ALS tests.

The results of section 7 are also illustrated. To this aim $N = 1000$ independent trajectories are simulated using a VAR(1) DGP with $A_{01} = -0.3I_2$. We consider heteroscedastic errors with trending behavior volatility specification:

$$\Sigma(r) = \begin{pmatrix} (1 + \pi_1 r) & 0 \\ 0 & (1 + \pi_1 r) \end{pmatrix}$$

where we take $\pi_1 = 150$, and a volatility with an abrupt shift at $T/2$:

$$\Sigma(r) = \begin{pmatrix} (1 + f_1(r)) & 0 \\ 0 & (1 + f_1(r)) \end{pmatrix},$$

where $f_1(r) = 10 \times \mathbf{1}_{(r \geq 1/2)}(r)$. The lengths of the simulated series are $T = 50, 100, 200$. The non correlation of the observed process is tested at the asymptotic nominal level 5% taking $m = 10$. From Figure 4 the \widetilde{LB}_{OLS} test may appear more powerful than the LB_{ALS} test when the samples are small. However this comes from the fact that the \widetilde{LB}_{OLS} test is strongly oversized in small samples. In accordance with the theoretical we see that the LB_{ALS} test clearly outperform the \widetilde{LB}_{OLS} and LB_{OLS} as the samples become large.

10 Illustrative examples

For our real data illustrations we use two U.S. economic data sets. First we consider the quarterly U.S. international finance data for the period from January 1, 1970 to October 1, 2009: the balance on services and the balance on merchandise trade in billions of Dollars. The length of the balance data series is $T = 159$. We also consider monthly data on the U.S. consumer price indexes of energy and transportation for all urban consumers for the period from January 1, 1957 to February 1, 2011. The length of the energy-transportation series is $T = 648$. The series are available seasonally adjusted from the website of the research division of the Federal Reserve Bank of Saint Louis.

10.1 VAR modeling of the U.S. balance trade data

In our VAR system the first component corresponds to the balance on merchandise trade and the second corresponds to the balance on services trade. From Figure 5 it

seems that the series have a random walk behavior. We applied the approach of unit root testing proposed by Beare (2008) in presence of non constant volatility using the Augmented Dickey-Fuller (ADF) test for each series. The ADF statistic is 0.72 for the merchandise trade balance data and is -0.15 for the services balance data. These statistics are greater than the 5% critical value -1.94 of the ADF test, so that the stability hypothesis have to be rejected for the studied series. Furthermore we also applied the Kolmogorov-Smirnov (KS) type test for homoscedasticity considered by Cavaliere and Taylor (2008, Theorem 3) for each series. The KS statistic is 3.05 for the merchandise trade balance data and is 7.62 for the services balance data. These statistics are greater than the 5% critical value 1.36, so that the homoscedasticity hypothesis has to be rejected for our series. Therefore we consider the first differences of the series to get stable processes, so that the evolutions of the U.S. balance data are studied in the sequel. From Figure 6 we see that the first differences of the series are stable but have a non constant volatility. We adjusted a VAR(1) model to capture the linear dynamics of the series. The ALS and OLS estimators are given in Table 5. The standard deviations into brackets are computed using the results (2.3) and (2.4). In accordance with Patilea and Raïssi (2010), we find that the ALS estimation method seems better estimate the autoregressive parameters than the OLS estimation method, in the sense that the standard deviations of the ALS estimators are smaller than those of the OLS estimators. The bandwidth we use for the ALS estimation, $b = 7.67 \times 10^{-2}$, is selected by cross-validation in a given range and using 200 grid points (Figure 7).

A quick inspection of Figures 8 and 9 suggests that the OLS residuals have a varying volatility, while the stationarity of the ALS residuals is plausible. Thus it seems that the volatility of the error process (u_t) is satisfactorily estimated by the

adaptive method. Now we examine the possible presence conditional heteroscedasticity in the ALS residuals. From Figure 10 we see that the autocorrelations of the squared ALS residuals components are not significant. In addition we also considered the ARCH-LM test of Engle (1982) with different lags in Table 6 for testing the presence of ARCH effects in the ALS residuals. It appears that the null hypothesis of conditional homoscedasticity cannot be rejected at the level 5%. These diagnostics give some evidence that the conditional homoscedasticity assumption on (ϵ_t) is plausible in our case. To analyze the changes of the variance of the OLS error terms, we plotted the estimated variances and cross correlation of the components of the error process in Figures 11 and 12. It appears from Figure 11 that the variance of the first component of the residuals does not vary much until the end of the 90's and then increase. Similarly the volatility of the second component of the residuals does not vary much until the end of the 80's and then increase. From Figure 12 we also note that the correlation between the components of the innovations seems to be positive until the beginning of the 90's and then become negative.

Now we turn to the check of the goodness-of-fit of the VAR(1) model adjusted to the first differences of the series. To illustrate the results of Proposition 1 we plotted the ALS residual autocorrelations in Figures 13 and 14, and the OLS residual autocorrelations in Figures 15 and 16, where we denote

$$\hat{R}_{OLS}^{ij}(h) = \frac{T^{-1} \sum_{t=h+1}^T \hat{u}_{it} \hat{u}_{jt-h}}{\hat{\sigma}_u(i) \hat{\sigma}_u(j)} \quad \text{and} \quad \check{R}_{ALS}^{ij}(h) = \frac{T^{-1} \sum_{t=h+1}^T \check{\epsilon}_{it} \check{\epsilon}_{jt-h}}{\check{\sigma}_\epsilon(i) \check{\sigma}_\epsilon(j)}.$$

The ALS 95% confidence bounds obtained using (3.14) and (5.3) are displayed in Figures 13 and 14. In Figures 15 and 16 the standard 95% confidence bounds obtained using (3.9) and the OLS 95 % confidence bounds obtained using (3.13) are plotted. We can remark that the ALS residual autocorrelations are inside or not

much larger than the ALS significance limits. A similar comment can be made for the OLS residual autocorrelations when compared to the OLS significance limits. However we found that the OLS significance limits can be quite different from the standard significance limits. This can be explained by the presence of unconditional volatility in the analyzed series. In particular we note that the $\hat{R}_{OLS}^{21}(5)$ is far from the standard confidence bounds. We also apply the different portmanteau tests considered in this paper for testing if the errors are uncorrelated. The test statistics and p -values of the tests are displayed in Tables 7 and 8. It appears that the p -values of the standard tests are very small. Therefore the standard tests clearly reject the null hypothesis. We also remark that the p -values of the modified tests based on the OLS estimation and of the adaptive tests are far from zero. Thus in view of the tests introduced in this paper the null hypothesis is not rejected. These contradictory results can be explained by the fact that we found that the distribution in (4.2) is very different from the $\chi_{d^2(m-p)}^2$ standard distribution. For instance we obtained $\sup_{i \in \{1, \dots, d^2 m\}} \left\{ \hat{\delta}_i^{ols} \right\} = 11.18$ for $m = 15$ in our case. Our findings may be viewed as a consequence of the presence of unconditional heteroscedasticity in the data. Since the theoretical basis of the standard tests do not include the case of stable processes with non constant volatility, we can suspect that the results of the standard tests are not reliable. Therefore we can draw the conclusion that the practitioner is likely to select a too large autoregressive order in our case when using the standard tools for checking the adequacy of the VAR model. From Table 8 we see that the OLS and ALS statistics are quite different. We also noted that the weights (not reported here) of the sums in (4.2) and in (5.4) are quite different for our example.

10.2 VAR modeling of the U.S. energy-transportation data

In this example the first component of the VAR system corresponds to the transportation price index and the second corresponds to the energy price index. We first briefly describe some features of the energy-transportation price indexes. In Figure 17 we again see that the studied series seems to have a random walk behavior and then we again consider the first differences of the series. The KS type statistic is 7.05 for the energy price index and is 6.81 for the transportation price index, so that the homoscedasticity hypothesis has to be rejected for these series. From Figure 18 we see that the first differences of the series are stable but have a non constant volatility. We adjusted a VAR(4) model to capture the linear dynamics of the series. The results in Table 9 indicate that the ALS approach is more precise than the OLS approach for the estimation of the autoregressive parameters. The bandwidth obtained by the cross-validation for the ALS estimation is $b = 9.53 \times 10^{-2}$ (Figure 19). From Figure 21 we see that the OLS residuals seem to have a varying volatility. The stationarity of the ALS residuals is plausible (Figure 20) and ARCH-LM tests (not reported here) show that the conditional homoscedasticity of the ALS residuals cannot be rejected. From Figure 22 we see that the shape of the variance structure of the components of the OLS residuals are similar. More precisely it can be noted that the variance of the components of the OLS residuals is relatively low and seems constant until the beginning of the 80's. The OLS residual variance seems to switch to an other regime where the variance is increased but constant from the beginning of the 80's to the end of the 90's. The volatility of the OLS residual variance seems to increase from the end of the 90's. From Figure 12 we also note that the components of the OLS residuals seem highly correlated.

Now we check the adequacy of the VAR(4) model. The ALS and OLS residual

autocorrelations are given in Figures 24, 25 and 26, 27. The OLS residual autocorrelations with the standard bounds are given in Figures 28 and 29. From Figures 24 and 25 it can be noted that the ALS residual autocorrelations are inside or not much larger than the ALS significance limits. The OLS confidence bounds seems not reliable since we remark that some confidence bounds (corresponding to the $\hat{R}_{OLS}^{ij}(2)$'s) are unexpectedly much larger than the OLS residual autocorrelations. From Figures 28 and 29 it can also be noted that some of the OLS residual autocorrelations are far from the standard confidence bounds. However considering the standard confidence bounds in presence of non constant variance can be misleading in view of our theoretical findings. In addition we also remark that some standard confidence bounds (corresponding to the $\hat{R}_{OLS}^{11}(2)$, $\hat{R}_{OLS}^{21}(2)$ and $\hat{R}_{OLS}^{22}(8)$, $\hat{R}_{OLS}^{12}(8)$) are unexpectedly far from the residual autocorrelations. The non correlation of the residuals is tested using the different portmanteau tests considered in this paper. The test statistics and p -values of the tests are displayed in Tables 10 and 11. The standard tests again lead to select complicated models since the adequacy of the VAR(4) model is rejected. We also remark that the p -values of the OLS tests are very large. In fact we found that the OLS tests do not reject the hypothesis of the adequacy of a VAR(1) model for the studied series. From Table 12 we see that some of the estimated weights for the asymptotic distribution of the standard statistic take very large values. It appears that the OLS method seems not able to estimate correctly the asymptotic distribution of the standard statistics and may be suspected to have a low power in this example. Then the OLS portmanteau test seem not reliable in this example on the contrary to the ALS portmanteau test. Indeed it can be noted that the asymptotic distribution seems well estimated. Finally note that we found that the estimators of the $\Lambda_m^{\epsilon,\theta'} \Lambda_m^{\epsilon,\theta}$ and $\Phi_m' (\Lambda_m^{u,u})^{-1} \Phi_m$ are not

invertible, so that the \widetilde{LB}_m^{ALS} and \widetilde{LB}_m^{OLS} are not feasible in this example. In view of the different outputs we presented, it seems that the LB_m^{ALS} test is the only test which give reliable conclusions in this example. This may be explained by the fact that the variance strongly change in this second example when compared to the first example and hence make the tests based on inverted matrices or the tests which do not exploit the variance structure difficult to implement.

11 Conclusion

In this paper the problem of specification of the linear autoregressive dynamics of multivariate series with deterministic but non constant volatility is studied. Considering such situations is important since it is well known that economic or financial series commonly exhibit non stationary volatility. The unreliability of the standard portmanteau tests for testing the adequacy of the autoregressive order of VAR processes is highlighted through theoretical results and empirical illustrations. From the statistical methodology point of view, the main contribution of the paper is two-fold. In the setup of a stable VAR with time-varying variance, (a) we show how to compute corrected critical values for the standard portmanteau statistics implemented in all specialized software; and (b) we propose new portmanteau statistics based on the model residuals filtered for the non constant variance. Moreover, we provide some theoretical and empirical power comparisons of the two approaches and we show that they are well-suited for replacing the usual test procedures even when the volatility is constant. The new portmanteau statistics require the estimation of the time-varying variance that is done by classical kernel smoothing of the outer product of the OLS residuals vectors. Then another contribution of the paper

is represented by the fact that our asymptotic results are derived uniformly in the bandwidth used in the kernel. This makes the theory compatible with the practice where people usually use the data to determine the bandwidth. Our theoretical and empirical investigation could be extended to other setups, like for instance the co-integrated systems. We briefly mention this extension but a deeper investigation is left for future work.

12 Appendix A: Proofs

For the sake of a simpler presentation, hereafter we stick to our identification condition for H_t and hence we replace everywhere $G(r)$ by $\Sigma(r)^{1/2}$. Recall that

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

(see pages 9 and 11). Let us introduce

$$\Upsilon_{t-1}^u = (u'_{t-1}, \dots, u'_{t-m}, \tilde{X}'_{t-1})' = (u_{t-1}^{m'}, \tilde{X}'_{t-1})'$$

and

$$\Upsilon_{t-1}^\epsilon = (\epsilon'_{t-1}, \dots, \epsilon'_{t-m}, \tilde{X}'_{t-1})' = (\epsilon_{t-1}^{m'}, \tilde{X}'_{t-1})',$$

for a given $m > 0$. To prove Propositions 1 and 2 we need several preliminary results that are gathered in Lemmas 12.1 to 12.3 below.

Lemma 12.1 *Under Assumption A1 we have*

$$\lim_{T \rightarrow \infty} E \left[\tilde{X}_{[Tr]-1} \tilde{X}'_{[Tr]-1} \right] = \sum_{i=0}^{\infty} \tilde{\psi}_i \{ \mathbf{1}_{p \times p} \otimes \Sigma(r) \} \tilde{\psi}'_i := \Omega(r), \quad (12.1)$$

$$\lim_{T \rightarrow \infty} E \left[\Upsilon_{[Tr]-1}^u \Upsilon'_{[Tr]-1} \right] = \Omega^u(r), \quad (12.2)$$

$$\lim_{T \rightarrow \infty} E \left[\Upsilon_{[Tr]-1}^\epsilon \Upsilon'_{[Tr]-1} \right] = \Omega^\epsilon(r), \quad (12.3)$$

for values $r \in (0, 1]$ where the functions $g_{ij}(\cdot)$ are continuous. The matrices in (12.2) and (12.3) are given by

$$\Omega^u(r) = \begin{pmatrix} I_m \otimes \Sigma(r) & \Theta_m^u(r) \\ \Theta_m^u(r)' & \Omega(r) \end{pmatrix}, \quad \Theta_m^u(r) = \sum_{i=0}^{m-1} \{e_m(i+1)e_p(1)' \otimes \Sigma(r)\} K^{i'}$$

and

$$\Omega^\epsilon(r) = \begin{pmatrix} I_{dm} & \Theta_m^\epsilon(r) \\ \Theta_m^\epsilon(r)' & \Omega(r) \end{pmatrix}, \quad \Theta_m^\epsilon(r) = \sum_{i=0}^{m-1} \{e_m(i+1)e_p(1)' \otimes \Sigma(r)^{1/2}\} K^{i'}.$$

Proof of Lemma 12.1 Statement 12.1 is a direct consequence of Lemma 7.2 in Patilea and Raïssi (2010). For the proof of (12.2) we write²

$$\begin{aligned} E(u_{t-1}^m \tilde{X}'_{t-1}) &= \sum_{i=0}^{\infty} E(u_{t-1}^m \tilde{u}'_{t-i-1} K^{i'}) \\ &= \sum_{i=0}^{m-1} E(u_{t-1}^m \tilde{u}'_{t-i-1} K^{i'}) \\ &= \sum_{i=0}^{m-1} \{e_m(i+1)e_p(1)' \otimes H_{t-i-1} H'_{t-i-1}\} K^{i'} \\ &= \sum_{i=0}^{m-1} \{e_m(i+1)e_p(1)' \otimes \Sigma((t-i-1)/T)\} K^{i'}. \end{aligned}$$

Therefore

$$\begin{aligned} \lim_{T \rightarrow \infty} E(u_{[Tr]-1}^m \tilde{X}'_{[Tr]-1}) &= \lim_{T \rightarrow \infty} \sum_{i=0}^{m-1} \{e_m(i+1)e_p(1)' \otimes \Sigma(([Tr]-i-1)/T)\} K^{i'} \\ &= \sum_{i=0}^{m-1} \{e_m(i+1)e_p(1)' \otimes \Sigma(r)\} K^{i'}. \end{aligned}$$

Similarly we have

$$\lim_{T \rightarrow \infty} E(u_{[Tr]-1}^m u_{[Tr]-1}^{m'}) = I_m \otimes \Sigma(r), \quad (12.4)$$

²Here we make a common abuse of notation because in Assumption **A1** the matrix-valued function $\Sigma(\cdot)$ is not defined for negative values. To remedy this problem it suffices to extend the function $\Sigma(\cdot)$ to the left of the origin, for instance by setting $\Sigma(r)$ equal to the identity matrix if $r \leq 0$.

so that using (12.1) we obtain the result (12.2). The proof of (12.3) is similar. \square

Let us define $v_t^u = \text{vec}(\Upsilon_{t-1}^u \Upsilon_{t-1}^{u'} \otimes u_t u_t')$ and $v_t^\epsilon = \text{vec}(\Upsilon_{t-1}^\epsilon \Upsilon_{t-1}^{\epsilon'} \otimes \epsilon_t \epsilon_t')$. The following lemma is similar to Lemma 7.3 of Patilea and Raïssi (2010), and hence the proof is omitted.

Lemma 12.2 *Under A1 we have*

$$T^{-1} \sum_{t=i+1}^T \text{vec}(u_t \tilde{X}'_{t-i}) \xrightarrow{P} 0, \quad (12.5)$$

$$T^{-1} \sum_{t=i+1}^T \text{vec}(\epsilon_t \tilde{X}'_{t-i}) \xrightarrow{P} 0, \quad (12.6)$$

for $i > 0$, and

$$T^{-1} \sum_{t=m+1}^T \text{vec}(u_{t-1}^m \tilde{X}'_{t-1}) \xrightarrow{P} \lim_{T \rightarrow \infty} T^{-1} \sum_{t=m+1}^T \text{vec} \left\{ E(u_{t-1}^m \tilde{X}'_{t-1}) \right\}, \quad (12.7)$$

$$T^{-1} \sum_{t=m+1}^T \text{vec}(\epsilon_{t-1}^m \tilde{X}'_{t-1}) \xrightarrow{P} \lim_{T \rightarrow \infty} T^{-1} \sum_{t=m+1}^T \text{vec} \left\{ E(\epsilon_{t-1}^m \tilde{X}'_{t-1}) \right\}. \quad (12.8)$$

In addition we have

$$T^{-1} \sum_{t=1}^T v_t^u \xrightarrow{P} \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T E(v_t^u) = \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \text{vec} \left\{ E(\Upsilon_{t-1}^u \Upsilon_{t-1}^{u'}) \otimes \Sigma_t \right\}, \quad (12.9)$$

$$T^{-1} \sum_{t=1}^T v_t^\epsilon \xrightarrow{P} \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T E(v_t^\epsilon) = \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \text{vec} \left\{ E(\Upsilon_{t-1}^\epsilon \Upsilon_{t-1}^{\epsilon'}) \otimes I_d \right\}. \quad (12.10)$$

Lemma 12.3 *Under A1 we have*

$$\hat{\Sigma}_{\tilde{X}} = \Lambda_1 + o_p(1), \quad (12.11)$$

$$\hat{\Sigma}_{\tilde{X}} = \Lambda_3 + o_p(1). \quad (12.12)$$

In addition we also have

$$T^{-\frac{1}{2}} \sum_{t=1}^T \Upsilon_{t-1}^u \otimes u_t \Rightarrow \mathcal{N}(0, \Xi_u), \quad (12.13)$$

$$T^{-\frac{1}{2}} \sum_{t=1}^T J_t^{-1}(\Upsilon_{t-1}^\epsilon \otimes \epsilon_t) \Rightarrow \mathcal{N}(0, \Xi_\epsilon), \quad (12.14)$$

where

$$J_t = \begin{pmatrix} I_{d^2m} & 0 \\ 0 & I_{dp} \otimes H_t \end{pmatrix}$$

and

$$\begin{aligned} \Xi_u &= \int_0^1 \begin{pmatrix} I_m \otimes \Sigma(r)^{\otimes 2} & \Theta_m^u(r) \otimes \Sigma(r) \\ \Theta_m^{u'}(r) \otimes \Sigma(r) & \Omega(r) \otimes \Sigma(r) \end{pmatrix} dr = \begin{pmatrix} \Lambda_m^{u,u} & \Lambda_m^{u,\theta} \\ \Lambda_m^{u,\theta'} & \Lambda_2 \end{pmatrix}, \\ \Xi_\epsilon &= \int_0^1 \begin{pmatrix} I_{d^2m} & \Theta_m^\epsilon(r) \otimes \Sigma(r)^{-\frac{1}{2}} \\ \Theta_m^{\epsilon'}(r) \otimes \Sigma(r)^{-\frac{1}{2}} & \Omega(r) \otimes \Sigma(r)^{-1} \end{pmatrix} dr = \begin{pmatrix} \Lambda_m^{u,u} & \Lambda_m^{\epsilon,\theta} \\ \Lambda_m^{\epsilon,\theta'} & \Lambda_1 \end{pmatrix}. \end{aligned}$$

Proof of Lemma 12.3 Statements (12.11) and (12.12) lemma are direct consequences of Lemma 7.4 of Patilea and Raïssi (2010). We only give the proof of (12.14) and (12.13). To prove (12.14), using the well known identity $(B \otimes C)(D \otimes F) = (BD) \otimes (CF)$ for matrices of appropriate dimensions, we obtain

$$J_t^{-1}(\Upsilon_{t-1}^\epsilon \otimes \epsilon_t)(\Upsilon_{t-1}^{\epsilon'} \otimes \epsilon'_t)J_t^{-1} = J_t^{-1}(\Upsilon_{t-1}^\epsilon \Upsilon_{t-1}^{\epsilon'} \otimes \epsilon_t \epsilon'_t)J_t^{-1}.$$

From (12.10) we write

$$T^{-1} \sum_{t=1}^T J_t^{-1}(\Upsilon_{t-1}^\epsilon \Upsilon_{t-1}^{\epsilon'} \otimes \epsilon_t \epsilon'_t)J_t^{-1} \xrightarrow{P} \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T J_t^{-1} \left[E \left\{ \Upsilon_{t-1}^\epsilon \Upsilon_{t-1}^{\epsilon'} \right\} \otimes I_d \right] J_t^{-1}.$$

Now let us denote the discontinuous points of the functions $g_{ij}(\cdot)$ by $\xi_1, \xi_2, \dots, \xi_q$ where q is a finite number independent of T . We write

$$\begin{aligned}
& \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T J_t^{-1} \left[E(\Upsilon_{t-1}^\epsilon \Upsilon_{t-1}^{\epsilon'} \otimes I_d) \right] J_t^{-1} \\
&= \lim_{T \rightarrow \infty} \sum_{t=1}^T \int_{t/T}^{(t+1)/T} J_{[Tr]}^{-1} \left[E(\Upsilon_{[Tr]-1}^\epsilon \Upsilon_{[Tr]-1}^{\epsilon'} \otimes I_d) \right] J_{[Tr]}^{-1} dr + o_p(1) \\
&= \lim_{T \rightarrow \infty} \int_{1/T}^{\xi_1} J_{[Tr]}^{-1} \left[E(\Upsilon_{[Tr]-1}^\epsilon \Upsilon_{[Tr]-1}^{\epsilon'} \otimes I_d) \right] J_{[Tr]}^{-1} dr + \dots \\
&\dots + \int_{\xi_q}^{(T+1)/T} J_{[Tr]}^{-1} \left[E(\Upsilon_{[Tr]-1}^\epsilon \Upsilon_{[Tr]-1}^{\epsilon'} \otimes I_d) \right] J_{[Tr]}^{-1} dr + o_p(1).
\end{aligned}$$

Then from (12.3) we obtain

$$\begin{aligned}
& T^{-1} \sum_{t=1}^T J_t^{-1} (\Upsilon_{t-1}^\epsilon \Upsilon_{t-1}^{\epsilon'} \otimes \epsilon_t \epsilon_t') J_t^{-1} \\
& \xrightarrow{P} \int_0^1 J(r)^{-1} \begin{pmatrix} I_{d^2m} & \Theta_m^\epsilon(r) \otimes I_d \\ \Theta_m^{\epsilon'}(r) \otimes I_d & \Omega(r) \otimes I_d \end{pmatrix} J(r)^{-1} dr,
\end{aligned}$$

where

$$J(r) = \begin{pmatrix} I_{d^2m} & 0 \\ 0 & I_{dp} \otimes \Sigma(r)^{-\frac{1}{2}} \end{pmatrix}$$

and $\Sigma(r)^{1/2} = G(r) = H_{[Tr]}$. Noting that $J_t^{-1}(\Upsilon_{t-1}^\epsilon \otimes \epsilon_t)$ are martingale differences, we obtain the result (12.14) using the Lindeberg central limit theorem. Using relations (12.2) and (12.9), the proof of (12.13) is similar. Finally, the equivalent compact expressions of Ξ_ϵ and Ξ_u can be easily derived using elementary properties of the Kronecker product. \square

Proof of Proposition 1 First we establish the result (3.5). Let us define

$$\Gamma_u(h) = T^{-1} \sum_{t=h+1}^T u_t u_{t-h}' \quad \text{and} \quad c_m^u = \text{vec} \{ (\Gamma_u(1), \dots, \Gamma_u(m)) \}.$$

Let us first show the asymptotic normality of $T^{1/2}(c_m^u, (\hat{\theta}_{OLS} - \theta_0)')'$. Note that

$$c_m^u = T^{-1} \sum_{t=1}^T \tilde{u}_{t-1}^m \otimes u_t \quad \text{and} \quad \hat{\theta}_{OLS} - \theta_0 = \hat{\Sigma}_{\tilde{X}}^{-1} \left\{ T^{-1} \sum_{t=1}^T (\tilde{X}_{t-1} \otimes u_t) \right\},$$

with $\tilde{u}_{t-1}^m = (\mathbf{1}_{(0,\infty)}(t-1) \times u'_{t-1}, \dots, \mathbf{1}_{(0,\infty)}(t-m) \times u'_{t-m})'$. From (12.12) we write

$$T^{\frac{1}{2}} \begin{pmatrix} c_m^u \\ \hat{\theta}_{OLS} - \theta_0 \end{pmatrix} = \dot{\Lambda}_3^{-1} \left\{ T^{-\frac{1}{2}} \sum_{t=1}^T \Upsilon_{t-1}^u \otimes u_t \right\} + o_p(1),$$

with

$$\dot{\Lambda}_3 = \begin{pmatrix} I_{d^2m} & 0 \\ 0 & \Lambda_3 \end{pmatrix}.$$

Then we can obtain from (12.13)

$$T^{\frac{1}{2}} \begin{pmatrix} c_m^u \\ \hat{\theta}_{OLS} - \theta_0 \end{pmatrix} \Rightarrow \mathcal{N}(0, \dot{\Lambda}_3^{-1} \Xi_u \dot{\Lambda}_3^{-1}), \quad (12.15)$$

with Ξ_u defined in Lemma 12.3. Now, define $u_t(\theta) = X_t - (\tilde{X}'_{t-1} \otimes I_d)\theta$ with $\theta \in \mathbb{R}^{d^2p}$.

Considering $\hat{\gamma}_m^{u,OLS}$ and c_m^u as values of the same function at the points θ_0 and $\hat{\theta}_{OLS}$,

by the Mean Value Theorem

$$\hat{\gamma}_m^{u,OLS} = c_m^u + T^{-1} \sum_{t=1}^T \left\{ \tilde{u}_{t-1}^m(\theta) \otimes \frac{\partial u_t(\theta)}{\partial \theta'} + \frac{\partial \tilde{u}_{t-1}^m(\theta)}{\partial \theta'} \otimes u_t(\theta) \right\}_{\theta=\theta^*} (\hat{\theta}_{OLS} - \theta_0),$$

with θ^* between $\hat{\theta}_{OLS}$ and θ_0 .³ Using $T^{1/2}(\hat{\theta}_{OLS} - \theta_0) = O_p(1)$ and since $\partial u_{t-i}(\theta)/\partial \theta' = -(\tilde{X}'_{t-i-1} \otimes I_d)$, it follows from (12.5) and (12.7) that

$$\begin{aligned} \hat{\gamma}_m^{u,OLS} &= c_m^u + \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T E \left\{ \tilde{u}_{t-1}^m \otimes \frac{\partial u_t}{\partial \theta'} \right\} (\hat{\theta}_{OLS} - \theta_0) + o_p(T^{-\frac{1}{2}}) \\ &= c_m^u + \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T -E \left\{ \tilde{u}_{t-1}^m \otimes \tilde{X}'_{t-1} \otimes I_d \right\} (\hat{\theta}_{OLS} - \theta_0) \\ &\quad + o_p(T^{-\frac{1}{2}}). \end{aligned} \quad (12.16)$$

From (12.2) and using arguments like in the proof of (12.14), it is easy to see that

$$\lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T -E \left\{ \tilde{u}_{t-1}^m \otimes \tilde{X}'_{t-1} \otimes I_d \right\} = - \int_0^1 \Theta_m^u(r) dr \otimes I_d + o_p(1) = -\Phi_m^u + o_p(1).$$

³The value θ^* between $\hat{\theta}_{OLS}$ and θ_0 may be different for different components of $\hat{\gamma}_m^{u,OLS}$ and c_m^u .

Finally from (12.16) we have

$$\hat{\gamma}_m^{u,OLS} = c_m^u - \Phi_m^u (\hat{\theta}_{OLS} - \theta_0) + o_p(T^{-\frac{1}{2}}), \quad (12.17)$$

so that it follows from (12.15) that $T^{1/2}\hat{\gamma}_m^{OLS}$ is asymptotically normal with covariance matrix

$$\Sigma^{u,OLS} = \Lambda_m^{u,u} + \Phi_m^u \Lambda_3^{-1} \Lambda_2 \Lambda_3^{-1} \Phi_m^{u'} - \Lambda_m^{u,\theta} \Lambda_3^{-1} \Phi_m^{u'} - \Phi_m^u \Lambda_3^{-1} \Lambda_m^{u,\theta'}. \quad (12.18)$$

The proof of (3.7) is very similar, here we only present a sketch. Let us define

$$\Gamma_\epsilon(h) = T^{-1} \sum_{t=h+1}^T \epsilon_t \epsilon'_{t-h} \quad \text{and} \quad c_m^\epsilon = \text{vec} \{(\Gamma_\epsilon(1), \dots, \Gamma_\epsilon(m))\}.$$

Using (12.11) and (12.14) it can be shown that

$$T^{\frac{1}{2}} \begin{pmatrix} c_m^\epsilon \\ \hat{\theta}_{GLS} - \theta_0 \end{pmatrix} \Rightarrow \mathcal{N}(0, \dot{\Lambda}_1^{-1} \Xi_\epsilon \dot{\Lambda}_1^{-1}), \quad (12.19)$$

with Ξ_ϵ defined in Lemma 12.3 and

$$\dot{\Lambda}_1 = \begin{pmatrix} I_{d^2 m} & 0 \\ 0 & \Lambda_1 \end{pmatrix}.$$

From (12.6), (12.8) and since $\partial \epsilon_{t-i}(\theta) / \partial \theta' = -(\tilde{X}'_{t-i-1} \otimes H_t^{-1})$ we write

$$\begin{aligned} \hat{\gamma}_m^{\epsilon,GLS} &= c_m^\epsilon + \lim_{T \rightarrow \infty} T^{-1} \sum_{t=m+1}^T -E \left\{ \epsilon_{t-1}^m \otimes \tilde{X}'_{t-1} \otimes H_t^{-1} \right\} (\hat{\theta}_{GLS} - \theta_0) + o_p(T^{-\frac{1}{2}}) \\ &= c_m^\epsilon - \Lambda_m^{\epsilon,\theta} (\hat{\theta}_{GLS} - \theta_0) + o_p(T^{-\frac{1}{2}}). \end{aligned}$$

By (12.19), $T^{1/2}\hat{\gamma}_m^{\epsilon,GLS}$ is asymptotically normal with covariance matrix $\Sigma^{\epsilon,GLS} = I_{d^2 m} - \Lambda_m^{\epsilon,\theta} \Lambda_1^{-1} \Lambda_m^{\epsilon,\theta'}$. The particular case where the order of the VAR model is $p = 0$ is an easy consequence of the arguments above in this proof. \square

Proof of Proposition 2 For the proof of (3.13), we write

$$u_{it} = \sum_{j=1}^d h_{ij,t} \epsilon_{jt} \quad \text{and} \quad E(u_{it}^2) = \sum_{j=1}^d h_{ij,t}^2 = \sigma_{ii,t}^2, \quad \text{say.}$$

It is clear from (12.4) that

$$\lim_{T \rightarrow \infty} E(u_{i[Tr]}^2) = \sigma_{ii}^2(r),$$

where $\sigma_{ii}^2(r)$ is the i th diagonal element of $\Sigma(r)$. Following similar arguments used in Phillips and Xu (2005 p 303) for the proof of Lemma 1 (iii), we write

$$T^{-1} \sum_{t=1}^T u_{it}^2 = \int_0^1 \sigma_{ii}^2(r) dr + o_p(1).$$

Let us define $\Sigma^u = T^{-1} \sum_{t=1}^T u_t u_t'$ and $\hat{\Sigma}^u = T^{-1} \sum_{t=1}^T \hat{u}_t \hat{u}_t'$. We have using again the Mean Value Theorem

$$\text{vec}(\hat{\Sigma}^u) = \text{vec}(\Sigma^u) + T^{-1} \sum_{t=1}^T \left\{ u_t(\theta) \otimes \frac{\partial u_t(\theta)}{\partial \theta'} + \frac{\partial u_t(\theta)}{\partial \theta'} \otimes u_t(\theta) \right\}_{\theta=\theta^*} (\hat{\theta}_{OLS} - \theta_0),$$

with θ^* is between $\hat{\theta}_{OLS}$ and θ_0 .⁴ Therefore using $\partial u_t(\theta)/\partial \theta' = -(\tilde{X}'_{t-1} \otimes I_d)$ and from the consistency of $\hat{\theta}_{OLS}$, we write

$$T^{\frac{1}{2}} \text{vec} \hat{\Sigma}^u = T^{\frac{1}{2}} \text{vec} \Sigma^u + o_p(1),$$

and

$$T^{-1} \sum_{t=1}^T \hat{u}_{it}^2 = \int_0^1 \sigma_{ii}^2(r) dr + o_p(1),$$

so that the result follows from the Slutsky lemma. We obtain the expression (3.13)

noting that

$$\hat{\rho}_m^{OLS} = \{I_m \otimes (\hat{S}_u \otimes \hat{S}_u)^{-1}\} \hat{\gamma}_m^{OLS}.$$

The proof of (3.14) is similar to that of (3.13) and hence is omitted. \square

⁴The value θ^* may be different for different components of $\text{vec}(\hat{\Sigma}^u)$ and $\text{vec}(\Sigma^u)$.

Proof of Proposition 3 In the following, c, C, \dots denote constants with possibly different values from line to line. To simplify notation, let b denote the $d(d+1)/2$ 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 $\sup_{b_{kl} \in \mathcal{B}_T, 1 \leq k \leq l \leq d}$. Here 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 obtain the asymptotic equivalences in equation (5.3) it suffices to notice that for all $1 \leq i \leq d$, $\hat{\sigma}_\epsilon^2(i) - 1 = o_p(1)$, and to prove

$$\sup_{1 \leq i \leq d} \sup_b |\check{\sigma}_\epsilon^2(i) - \hat{\sigma}_\epsilon^2(i)| = o_p(1) \quad (12.20)$$

and

$$\sup_b \left| T^{\frac{1}{2}} \{ \Gamma_{ALS}(h) - \Gamma_{GLS}(h) \} \right| = o_p(1), \quad (12.21)$$

for any fixed $h \geq 1$. Let us write

$$\begin{aligned} \check{\epsilon}_t - \hat{\epsilon}_t &= (\check{\Sigma}_t^{-\frac{1}{2}} - \Sigma_t^{-\frac{1}{2}})u_t + \check{\Sigma}_t^{-\frac{1}{2}}(\tilde{X}'_{t-1} \otimes I_d)(\hat{\theta}_{GLS} - \hat{\theta}_{ALS}) \\ &\quad + (\Sigma_t^{-\frac{1}{2}} - \check{\Sigma}_t^{-\frac{1}{2}})(\tilde{X}'_{t-1} \otimes I_d)(\hat{\theta}_{GLS} - \theta_0) \\ &=: (\check{\Sigma}_t^{-\frac{1}{2}} - \Sigma_t^{-\frac{1}{2}})u_t + \delta_t^\epsilon \end{aligned}$$

where $\|\delta_t^\epsilon\| \leq \|\tilde{X}_{t-1}\| \check{R}_T(b)$ with

$$\check{R}_T(b) = d \left\{ \|\hat{\theta}_{GLS} - \hat{\theta}_{ALS}\| \sup_{1 \leq t \leq T} \left\| \check{\Sigma}_t^{-\frac{1}{2}} \right\| + \|\hat{\theta}_{GLS} - \theta_0\| \sup_{1 \leq t \leq T} \left\| \check{\Sigma}_t^{-\frac{1}{2}} - \Sigma_t^{-\frac{1}{2}} \right\| \right\}$$

By Lemma 12.4-(a,b) and given that $\hat{\theta}_{GLS} - \theta_0 = O_p(T^{-1/2})$ and $\sup_b \|\hat{\theta}_{GLS} - \hat{\theta}_{ALS}\| = o_p(T^{-1/2})$, we obtain that

$$\sup_b \check{R}_T(b) = o_p(T^{-1/2}). \quad (12.22)$$

From this and the moment conditions on (X_t) induced by Assumption **A1**, deduce that (12.20) holds true. On the other hand,

$$\begin{aligned}
\Gamma_{ALS}(h) - \Gamma_{GLS}(h) &= \frac{1}{T} \sum_{t=h+1}^T \hat{\epsilon}_t (\check{\epsilon}_{t-h} - \hat{\epsilon}_{t-h})' + \frac{1}{T} \sum_{t=h+1}^T (\check{\epsilon}_t - \hat{\epsilon}_t) \check{\epsilon}'_{t-h} \\
&\quad + \frac{1}{T} \sum_{t=h+1}^T (\check{\epsilon}_t - \hat{\epsilon}_t) (\check{\epsilon}_{t-h} - \hat{\epsilon}_{t-h})' \\
&=: R_{1T}(h) + R_{2T}(h) + R_{3T}(h).
\end{aligned}$$

The terms $R_{1T}(h)$ and $R_{2T}(h)$ could be handled in a similar manner, hence we will only analyze $R_{2T}(h)$. Let us write

$$\begin{aligned}
R_{2T}(h) &= \frac{1}{T} \sum_{t=h+1}^T \left[(\check{\Sigma}_t^{-\frac{1}{2}} - \Sigma_t^{-\frac{1}{2}}) u_t + \delta_t^\epsilon \right] \left[\Sigma_{t-h}^{-\frac{1}{2}} u_{t-h} - \Sigma_{t-h}^{-\frac{1}{2}} (\tilde{X}'_{t-h-1} \otimes I_d) (\hat{\theta}_{GLS} - \theta_0) \right]' \\
&=: \frac{1}{T} \sum_{t=h+1}^T (\check{\Sigma}_t^{-\frac{1}{2}} - \Sigma_t^{-\frac{1}{2}}) u_t u'_{t-h} \Sigma_{t-h}^{-\frac{1}{2}'} + R_{22T}(h; b) =: R_{21T}(h; b) + R_{22T}(h; b).
\end{aligned}$$

By (12.22) and the moment conditions on the innovation process (ϵ_t) , and the rate of convergence of $\hat{\theta}_{GLS}$ and 12.4-(b), it is clear that $\sup_b \|R_{22T}(h; b)\| = o_p(T^{-1/2})$.

Next let us write

$$\begin{aligned}
R_{21T}(h) &= \frac{1}{T} \sum_{t=h+1}^T \left(\check{\Sigma}_t^{-\frac{1}{2}} - (\check{\Sigma}_t^0)^{-\frac{1}{2}} \right) u_t \epsilon'_{t-h} + \frac{1}{T} \sum_{t=h+1}^T \left((\check{\Sigma}_t^0)^{-\frac{1}{2}} - \overset{\circ}{\Sigma}_t^{-\frac{1}{2}} \right) u_t \epsilon'_{t-h} \\
&\quad + \frac{1}{T} \sum_{t=h+1}^T \left(\overset{\circ}{\Sigma}_t^{-\frac{1}{2}} - \bar{\Sigma}_t^{-\frac{1}{2}} \right) u_t \epsilon'_{t-h} + \frac{1}{T} \sum_{t=h+1}^T \left(\bar{\Sigma}_t^{-\frac{1}{2}} - \Sigma_t^{-\frac{1}{2}} \right) u_t \epsilon'_{t-h} \\
&=: R_{211T}(h; b) + R_{212T}(h; b) + R_{213T}(h; b) + R_{214T}(h; b),
\end{aligned}$$

where, like in Patilea and Raïssi (2010),

$$\overset{\circ}{\Sigma}_t = \overset{\circ}{\Sigma}_t(b) = \sum_{i=1}^T w_{ti} \odot u_i u_i' \quad \text{and} \quad \bar{\Sigma}_t = \bar{\Sigma}_t(b) = \sum_{i=1}^T w_{ti} \odot \Sigma_i.$$

From classical matrix norm inequalities (see for instance Horn and Johnson, 1994), we have that for any $d \times d$ -positive definite matrices A and B , for $a = 1$ or $a = -1$,

$$\|A^{-\frac{a}{2}} - B^{-\frac{a}{2}}\| \leq c_a (\max\{\|A^a\|, \|B^a\|\})^{\frac{1}{2}} \left\| A^{-\frac{1+a}{2}} \right\| \left\| B^{-\frac{1+a}{2}} \right\| \|A - B\|, \quad (12.23)$$

where c_a is a constant that depends only on d (by definition $A^0 = B^0 = I_d$). Applying this inequality twice we deduce

$$\begin{aligned} \left\| \check{\Sigma}_t^{-\frac{1}{2}} - (\check{\Sigma}_t^0)^{-\frac{1}{2}} \right\| &\leq \nu_T c_1 c_{-1} \left\| \check{\Sigma}_t^{-1} \right\| \left\| (\check{\Sigma}_t^0)^{-1} \right\| \\ &\quad \times \left(\max\{ \|\check{\Sigma}_t\|, \|\check{\Sigma}_t^0\| \} \right)^{\frac{1}{2}} \left(\max\{ \|[(\check{\Sigma}_t^0)^2 + \nu_T I_d]^{-1}\|, \|(\check{\Sigma}_t^0)^{-2}\| \} \right)^{\frac{1}{2}}. \end{aligned}$$

Take the norm of R_{211T} , use the inequality in the last display, Lemma 12.4-(a) below, the moment conditions on the innovation process and the condition $T\nu_T^2 \rightarrow \infty$ to deduce that $\sup_b \|R_{211T}(h; b)\| = o_p(T^{-1/2})$. Next, using similar matrix norm inequalities, Lemma 12.4-(a) and the Cauchy-Schwarz inequality,

$$\begin{aligned} \sup_b \|R_{212T}(h; b)\| &\leq O_p(1) \sup_b \left\{ \frac{1}{T} \sum_{t=h+1}^T \left\| (\check{\Sigma}_t^0) - \overset{\circ}{\Sigma}_t \right\| \|u_t \epsilon'_{t-h}\| \right\} \\ &\leq O_p(1) \left(\sup_b \left\{ \frac{1}{T} \sum_{t=h+1}^T \left\| (\check{\Sigma}_t^0) - \overset{\circ}{\Sigma}_t \right\|^2 \right\} \right)^{1/2} \left(\frac{1}{T} \sum_{t=h+1}^T \|u_t \epsilon'_{t-h}\|^2 \right)^{1/2} \\ &= O_p(1) O_p(T^{-1} b_T^{-1}) \left(\frac{1}{T} \sum_{t=h+1}^T \|u_t \epsilon'_{t-h}\|^2 \right)^{1/2}, \end{aligned}$$

where for the equality we used Lemma 7.6-(i) in Patilea and Raïssi (2010). Deduce that $\sup_b \|R_{212T}(h; b)\| = o_p(T^{-1/2})$. The uniform rate of convergence for $R_{213T}(h; b)$ is obtained after replacing $\overset{\circ}{\Sigma}_t^{-1/2} - \bar{\Sigma}_t^{-1/2}$ by a Taylor expansion of the power $-1/2$ function for positive definite matrix, a key and apparently new ingredient we provide in section 12.1 below. The reminder term of the Taylor expansion could be controlled taking expectation, using Cauchy-Schwarz inequality and Lemma 12.4-(d). The term under the integral that represents the first order term of this Taylor expansion could be treated similarly to the term $\bar{\Sigma}_t^{-1} [\Sigma_i - u_i u_i'] \bar{\Sigma}_t^{-1} u_t \tilde{X}'_{t-1}$ in the proof of the Proposition 4.1 of Patilea and Raïssi (2010). That means we use the CLT for m.d. sequences indexed by classes of functions, see Bae, Jun and Levental (2010), see also Bae and Choi (1999). Here the uniformity to be considered is also with respect to

the integration variable v , but this can be handled with very little additional effort, like in Patilea and Raïssi (2010). The details are omitted. Finally, to derive the uniform order $R_{214T}(h; b)$, let us write it as

$$R_{214T}(h; b) = R_{214T}(h; b) - R_{214T}(h; b_T) + R_{214T}(h; b_T) =: r_{214T}(b) + R_{214T}(h; b_T).$$

The term $R_{214T}(h; b_T)$ is centered and the variance of each element of this matrix decreases to zero at the rate $o(1/T)$ (use Lemma 12.4-(d) and Assumption **A1** to derive the rate of the variance). Deduce that $R_{214T}(h; b_T) = o_p(T^{-1/2})$. Next consider the d^2 stochastic processes corresponding to the elements of $r_{214T}(b)$ and indexed by $\vartheta \in [c_{min}, c_{max}]$ where $b = \vartheta b_T$. For each such process apply Theorem 1 of Bae, Jun and Levental (2010) to deduce that $\sup_b \|r_{214T}(b)\| = o_p(T^{-1/2})$. Finally, deduce that $\sup_b \|R_{214T}(h; b)\| = o_p(T^{-1/2})$

To handle the term $R_{3T}(h)$, let us write

$$\begin{aligned} R_{3T}(h) &= \frac{1}{T} \sum_{t=h+1}^T \left[(\check{\Sigma}_t^{-\frac{1}{2}} - \Sigma_t^{-\frac{1}{2}}) u_t + \delta_t^\epsilon \right] \left[(\check{\Sigma}_{t-h}^{-\frac{1}{2}} - \Sigma_{t-h}^{-\frac{1}{2}}) u_{t-h} + \delta_{t-h}^\epsilon \right]' \\ &=: \frac{1}{T} \sum_{t=h+1}^T (\check{\Sigma}_t^{-\frac{1}{2}} - \Sigma_t^{-\frac{1}{2}}) u_t u_{t-h}' (\check{\Sigma}_{t-h}^{-\frac{1}{2}} - \Sigma_{t-h}^{-\frac{1}{2}})' + R_{32T}(h) \\ &=: R_{31T}(h) + R_{32T}(h). \end{aligned}$$

The term $R_{32T}(h)$ could be easily handled taking the norm, using the bound on δ_t^ϵ and Lemma 12.4-(b) below. For $R_{31T}(h)$, we could decompose $\check{\Sigma}_t^{-1/2} - \Sigma_t^{-1/2}$ in four term exactly as we did for $R_{21T}(h)$ and apply the same techniques. The details are omitted and are available from the authors upon request. \square

Lemma 12.4 *Let $\|\cdot\|$ denote the Frobenius norm. Under the Assumptions of Proposition 3 we have:*

(a) *As $T \rightarrow \infty$, for $a = 1$ or $a = -1$, we have*

$$\sup_{1 \leq t \leq T} \sup_{b \in \mathcal{B}_T} \left\{ \left\| \check{\Sigma}_t^{-\frac{1}{2}} \right\| + \left\| \check{\Sigma}_t^a \right\| + \left\| (\check{\Sigma}_t^0)^a \right\| + \left\| \check{\Sigma}_t^{\circ a} \right\| + \left\| \bar{\Sigma}_t^a \right\| \right\} = O_p(1).$$

(b) As $T \rightarrow \infty$,

$$\sup_{1 \leq t \leq T} \sup_{b \in \mathcal{B}_T} \left\| \check{\Sigma}_t^{-\frac{1}{2}} - \Sigma_t^{-\frac{1}{2}} \right\| = o_p(1).$$

(c) As $T \rightarrow \infty$,

$$\sup_{b \in \mathcal{B}_T} \frac{1}{T} \sum_{t=1}^T \left\| \bar{\Sigma}_t - \Sigma_t \right\|^2 = o(1).$$

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

$$\max_{1 \leq t \leq T} E \left(\sup_{b \in \mathcal{B}_T} \left\| \overset{\circ}{\Sigma}_t - \bar{\Sigma}_t \right\|^4 \right) = O((1/(Tb_T))^2). \quad (12.24)$$

The proof of Lemma 12.4 is a direct consequence of Lemmas 7.5 and 7.6 of Patilea and Raïssi (2010) and Lemma A of Xu and Phillips (2008) applied elementwise, and hence will be omitted.

Proof of Proposition 4 The notation in this proof are those of section 7. First let us notice that $q_A(x) = x/2\{1 + o(1)\}$ for large values of x , provided the asymptotic law of a test statistic Q_A under the null hypothesis is χ_m^2 with some $m \geq 1$. In the case where a test statistic Q_A has the asymptotic distribution of $U(\delta_m^{OLS})$ defined in equation (4.2),

$$\begin{aligned} q_A(x) &= -\log P(U(\delta_m^{OLS}) > x) \leq -\log P\left(\max_i \{\delta_i^{OLS}\} U^2 > x\right) \\ &= \frac{x}{2 \max_i \{\delta_i^{OLS}\}} \{1 + o(1)\}, \end{aligned}$$

and

$$\begin{aligned} q_A(x) &= -\log P(U(\delta_m^{OLS}) > x) \geq -\log P\left(\max_i \{\delta_i^{OLS}\} \sum_{j=1}^{d^2 m} U_j^2 > x\right) \\ &= \frac{x}{2 \max_i \{\delta_i^{OLS}\}} \{1 + o(1)\}, \end{aligned}$$

with U and U_j independent $\mathcal{N}(0, 1)$ variables. Thus to prove (i) it suffices to show that

$$\int_0^1 \Sigma dr \otimes \left(\int_0^1 \Sigma dr \right)^{-1} \ll \max_i \{\delta_i^{OLS}\} \left(\int_0^1 \Sigma dr \otimes I_d \right) \Sigma_{G^{\otimes 2}}^{-1} \left(\int_0^1 \Sigma dr \otimes I_d \right). \quad (12.25)$$

Herein, for any A and B symmetric matrices, $A \ll B$ means that $B - A$ is positive semidefinite. Now, in the last display, multiply both sides of the order relationship on the left and on the right by $\left(\int_0^1 \Sigma dr\right)^{-1/2} \otimes \left(\int_0^1 \Sigma dr\right)^{1/2}$ and deduce that it suffices to prove

$$\begin{aligned} I_d \otimes I_d &\ll \max_i \{\delta_i^{OLS}\} \left[\left(\int_0^1 \Sigma dr\right)^{1/2} \otimes \left(\int_0^1 \Sigma dr\right)^{1/2} \right] \Sigma_{G^{\otimes 2}}^{-1} \left[\left(\int_0^1 \Sigma dr\right)^{1/2} \otimes \left(\int_0^1 \Sigma dr\right)^{1/2} \right] \\ &=: \max_i \{\delta_i^{OLS}\} \tilde{\Delta}_m^{-1}. \end{aligned}$$

To obtain (12.25) it remains to notice that $\Delta_m^{OLS} = I_m \otimes \tilde{\Delta}_m$ and that δ_i^{OLS} , $1 \leq i \leq d^2 m$ are the eigenvalues of Δ_m^{OLS} .

In (ii) we suppose $\Sigma(\cdot) = \sigma^2(\cdot)I_d$ and in this case it suffices to notice that

$$\left(\int_0^1 \Sigma dr \otimes I_d\right) \Sigma_{G^{\otimes 2}}^{-1} \left(\int_0^1 \Sigma dr \otimes I_d\right) = \frac{\left(\int_0^1 \sigma^2(r) dr\right)^2}{\int_0^1 \sigma^4(r) dr} I_d \otimes I_d \ll I_d \otimes I_d,$$

where for the order relationship we use Cauchy-Schwarz inequality, while

$$\left(\int_0^1 G(r)' \otimes G(r)^{-1} dr\right)^2 = I_d \otimes I_d.$$

□

12.1 A Taylor expansion of the matrix function $f(A) = A^{-1/2}$

Recall that the differential of a function F that maps a $r \times r$ matrix X into a $r \times r$ matrix $F(X)$ is defined by the equation

$$\text{vec}(dF) = d\mathbf{f}$$

where \mathbf{f} is a $r^2 \times 1$ vector function such that $\mathbf{f}(\text{vec}(X)) = \text{vec}(F(X))$. In other words, the (first-order) differential of F at X in the $r \times r$ matrix obtained by unstacking the differential of $d\mathbf{f}$ at $\text{vec}(X)$. See also Schott (2005), section 9.3. Basic properties

of vector differentials implies

$$0 = d(X^{-1/2}XX^{-1/2}) = d(X^{-1/2})X^{1/2} + X^{-1/2}d(X)X^{-1/2} + X^{1/2}d(X^{-1/2}).$$

Now, recall that for any A positive definite matrix, the Lyapunov equation $AY + YA = B$ has a unique solution that can be represented as

$$Y = \int_0^\infty \exp(-vA)B \exp(-vA)dv.$$

See Horn and Johnson (1991), section 6.5. All these facts brings us to the following technical result.

Lemma 12.5 *Let A and \hat{A} be two positive definite $r \times r$ -matrices such that $0 < c_1 \leq \lambda_{\min}(B) < \infty$ for some constant c_1 and $B = A$ and $B = \hat{A}$, where $\lambda_{\min}(B)$ is the smallest eigenvalue of the symmetric matrix B . Moreover, suppose that $\|\hat{A} - A\| \leq c_2$ for some small constant c_2 . Then*

$$\hat{A}^{-1/2} - A^{-1/2} = - \int_0^\infty \exp(-vA)A^{-1/2}\{\hat{A} - A\}A^{-1/2} \exp(-vA)dv + R_n$$

where R_n is a $r \times r$ -symmetric matrix with $\|R_n\| \leq C\|\hat{A} - A\|^2$ and C is a constant depending only on c_1, c_2 .

Proof of Lemma 12.5 Let Δ be some arbitrary matrix. By Taylor expansion, for sufficiently small values of ε and for some matrices $G_i, i = 1, 2, \dots$

$$(A + \varepsilon\Delta)^{-1/2} = A^{-1/2} + \varepsilon G_1 + \varepsilon^2 G_2 + \dots = A^{-1/2} + \varepsilon G_1 + R_1 \quad (12.26)$$

with $\|R_1\| \leq C_1\varepsilon^2$ and C_1 a constant depending on c_1 and the norm of Δ . This kind of representation could be derived from a Taylor formula for the vector function \mathbf{f} defined by the equation $\mathbf{f}(\text{vec}(X)) = \text{vec}(X^{-1/2})$ considered in a neighborhood of the $\text{vec}(A)$ for a positive definite matrix A . See, for instance, Schott (2005),

section 9.6. On the other hand, recall that for B a square matrix with $\|B\| < 1$, $(I - B)^{-1} = I + B + B^2 + \dots = I + B + R_2$ where R_2 is the remainder of the expansion with $\|R_2\| \leq \|B\|^2(1 - \|B\|)^{-1}$. Thus for sufficiently small values of ε we can write

$$\begin{aligned} (A + \varepsilon\Delta)^{-1} &= A^{-1/2}(I + \varepsilon A^{-1/2}\Delta A^{-1/2})^{-1}A^{-1/2} \\ &= A^{-1/2}(I - \varepsilon A^{-1/2}\Delta A^{-1/2} + \varepsilon^2 A^{-1/2}\Delta A^{-1}\Delta A^{-1/2} + \dots)A^{-1/2} \end{aligned} \quad (12.27)$$

Taking the square on both sides of the first equality in (12.26) and identifying the coefficients of the power of ε in equation (12.27) deduce that G_1 is the solution of the Lyapunov equation

$$A^{-1/2}Y + YA^{-1/2} = -A^{-1}\Delta A^{-1}.$$

Finally, the result follows by taking $\Delta = (\widehat{A} - A)/\|\widehat{A} - A\|$ and $\varepsilon = \|\widehat{A} - A\|$. \square

References

- AHN, S.K. (1988) Distribution for residual autocovariances in multivariate autoregressive models with structured parameterization. *Biometrika* 75, 590-593.
- ANDERSON, T.W. (1951) Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics* 22, 327-351.
- 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.
- BAEK, E., AND BROCK, W. (1992) A general test for nonlinear Granger causality: bivariate model. Working paper, University of Wisconsin-Madison.
- BAHADUR, R.R. (1960) Stochastic comparison of tests. *Annals of Mathematical Statistics* 31, 276-295.
- BEARE B.K. (2008) Unit root testing with unstable volatility. Working paper, Nuffield College, University of Oxford.
- BOSWIJK, H.P. (2010) Nuisance parameter free inference on cointegration parameters in the presence of a variance shift. *Economics Letters* 107, 190-193.
- BOSWIJK, H.P., AND FRANCES, H.F. (1992) Dynamic specification and cointegration. *Oxford Bulletin of Economics and Statistics* 54, 369-381.
- BOUBACAR MAINASSARA, Y. (2010) Multivariate portmanteau test for structural VARMA models with uncorrelated but non-independent error terms. Working paper. EQUIPPE Université Lille 3.
- BOX, G.E.P., AND PIERCE, D.A. (1970) Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association* 65, 1509-1526.
- BROCKWELL, P.J., AND DAVIS, R.A. (1991) *Time Series: Theory and methods*. Springer, New York.

- BRÜGGEMANN R., LÜTKEPOHL H., AND SAIKKONEN P. (2006) Residual autocorrelation testing for vector error correction models. *Journal of Econometrics* 134, 579-604.
- 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 cointegration in vector autoregressions with non-stationary volatility. *Journal of Econometrics* 158, 7-24.
- 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.
- CAVALIERE, G., AND TAYLOR, A.M.R. (2008) Time-transformed unit root tests for models with non-stationary volatility. *Journal of Time Series Analysis* 29, 300-330.
- CHITTURI, R. V. (1974) Distribution of residual autocorrelations in multiple autoregressive schemes. *Journal of the American Statistical Association* 69, 928-934.
- DUCHESNE, P. (2005) Testing for serial correlation of unknown form in cointegrated time series models. *Annals of the Institute of Statistical Mathematics* 57, 575-595.
- EDGERTON, D., AND SHUKUR, G. (1999) Testing autocorrelation in a system perspective. *Econometric Reviews* 18, 343-386.
- ENGLE, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica* 50, 987-1008.

- FRANCQ, C., AND RAÏSSI, H. (2007) Multivariate portmanteau test for autoregressive models with uncorrelated but nonindependent errors. *Journal of Time Series Analysis* 28, 454-470.
- FRANCQ, C., ROY, R., AND ZAKOÏAN, J.-M. (2005) Diagnostic checking in ARMA models with uncorrelated errors. *Journal of the American Statistical Association* 13, 532-544.
- HORN, R.A., AND JOHNSON, C.R. (1994) *Topics in Matrix Analysis*. Cambridge University Press, Cambridge.
- HOSKING, J. R. M. (1980) The multivariate portmanteau statistic. *Journal of the American Statistical Association* 75, 343-386.
- IMHOF, J. P. (1961) Computing the distribution of quadratic forms in normal variables. *Biometrika* 48, 419-426.
- JOHANSEN, S. (1995) *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, New York.
- JONES, J.D. (1989) A comparison of lag-length selection techniques in tests of Granger causality between money growth and inflation: evidence for the US, 1959-86. *Applied Economics* 21, 809-822.
- KATAYAMA, N. (2008) An improvement of the portmanteau statistic. *Journal of Time Series Analysis* 29, 359-370.
- KIM, T.-H., LEYBOURNE, S., AND NEWBOLD, P. (2002) Unit root tests with a break in innovation variance. *Journal of Econometrics* 103, 365-387.

- KUONEN, D. (1999) Saddlepoint approximations for distributions of quadratic forms in normal variables. *Biometrika* 86, 929-935.
- LJUNG, G.M. AND BOX, G.E.P. (1978) On measure of lack of fit in time series models. *Biometrika* 65, 297-303.
- LOBATO, I., NANKERVIS, J.C., AND SAVIN, N.E. (2002) Testing for zero autocorrelation in the presence of statistical dependence. *Econometric Theory* 18, 730-743.
- LÜTKEPOHL, H. (2005) *New Introduction to Multiple Time Series Analysis*. Springer, Berlin.
- PATILEA, V., AND RAÏSSI, H. (2010) Adaptive estimation of vector autoregressive models with time-varying variance: application to testing linear causality in mean. Working document IRMAR-INSA; 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.
- RAÏSSI, H. (2010) Autocorrelation based tests for vector error correction models with uncorrelated but nonindependent errors. *Test* 19, 304-324.
- 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.
- SCHOTT, J.R. (2005) *Matrix analysis for statistics* (2nd ed.). Wiley series in probability and statistics, Wiley, Hoboken, N.J.

- STOCK, J.H., AND WATSON, M.W. (1989) Interpreting the evidence on money-income causality. *Journal of Econometrics* 40, 161-181.
- VAN DER VAART, A.W. (1998) *Asymptotic Statistics*. Cambridge University Press, Cambridge.
- THORNTON, D.L. AND BATTEN, D.S. (1985) Lag-length selection and tests of Granger causality between money and income. *Journal of Money, Credit, and Banking* 17, 164-178.
- VILASUSO, J. (2001) Causality tests and conditional heteroskedasticity: Monte Carlo evidence. *Journal of Econometrics* 101, 25-35.
- WATSON, M.W. (1999) Explaining the increased variability in long-term interest rates. *Federal Reserve Bank of Richmond Economic Quarterly* 85/4, 71-96.
- WHITE, H. (1980) A heteroskedasticity consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, 817-838.
- XU, K.L., AND PHILLIPS, P.C.B. (2008) Adaptive estimation of autoregressive models with time-varying variances. *Journal of Econometrics* 142, 265-280.

13 Appendix B: Tables and Figures

Table 1: Empirical size (in %) of the portmanteau tests with iid standard Gaussian errors.

Case	$m = 5$			$m = 15$		
T	50	100	200	50	100	200
LB_m^S	2.6	4.6	5.5	4.4	4.1	4.6
LB_m^{OLS}	4.2	4.9	5.2	11.5	8.1	6.9
LB_m^{ALS}	2.2	4.1	5.1	4.1	3.7	4.4
LB_m^{GLS}	2.0	3.9	5.1	3.7	3.8	4.3
\widetilde{LB}_m^{OLS}	14.2	9.0	6.5	30.9	15.4	10.5
\widetilde{LB}_m^{ALS}	6.3	5.9	5.6	15.8	7.4	6.8
\widetilde{LB}_m^{GLS}	4.6	4.7	4.8	8.6	8.1	8.1

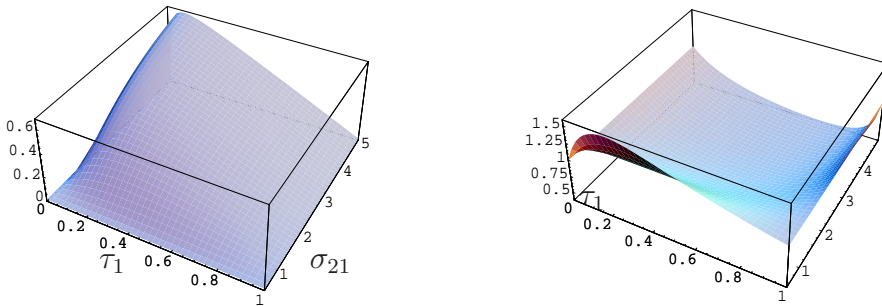


Figure 1: The asymptotic variance $\Sigma_{GLS}(2,2)$ on the left and the ratio $\Sigma_{OLS}(6,6)/\Sigma_{OLS}^S(6,6)$ on the right for $\tau_1 = \tau_2$ in Example 3.1.

Table 2: Empirical size (in %) of the portmanteau tests. The innovations are heteroscedastic with an abrupt break at $T/2$.

Case	$m = 5$			$m = 15$		
T	50	100	200	50	100	200
LB_m^S	27.9	35.3	40.1	35.7	63.0	76.7
LB_m^{OLS}	4.5	3.3	4.8	5.4	6.1	6.0
LB_m^{ALS}	3.2	3.7	5.0	3.8	3.8	3.9
LB_m^{GLS}	2.6	4.2	5.7	3.7	4.2	4.7
\widetilde{LB}_m^{OLS}	28.9	13.8	9.7	30.9	21.9	15.2
\widetilde{LB}_m^{ALS}	18.6	8.7	7.1	35.0	15.9	9.0
\widetilde{LB}_m^{GLS}	6.4	5.3	6.3	12.5	10.0	9.7

Table 3: Empirical size (in %) of the portmanteau tests. The innovations are heteroscedastic with trending behaviour.

Case	$m = 5$			$m = 15$		
T	50	100	200	50	100	200
LB_m^S	12.8	15.1	19.2	18.4	27.5	36.8
LB_m^{OLS}	4.9	4.5	4.8	9.2	7.3	6.3
LB_m^{ALS}	4.4	5.0	5.0	8.0	6.0	6.0
LB_m^{GLS}	2.3	3.7	5.2	2.8	4.0	4.0
\widetilde{LB}_m^{OLS}	32.4	22.6	15.3	38.3	25.1	16.5
\widetilde{LB}_m^{ALS}	10.0	4.2	3.3	22.7	6.7	5.2
\widetilde{LB}_m^{GLS}	5.7	5.3	6.3	10.8	9.7	9.2

Table 4: The empirical means and standard deviations of the weights in the sums (4.2), (5.4) and their GLS counterparts over the $N = 1000$ iterations. The innovations are heteroscedastic with trending behaviour.

i	1	2	3	4	5
$\hat{\delta}_i^{ols}$	0.02 _[0.02]	0.06 _[0.04]	0.11 _[0.06]	0.2 _[0.09]	0.89 _[0.11]
$\hat{\delta}_i^{als}$	0.03 _[0.02]	0.05 _[0.02]	0.09 _[0.03]	0.11 _[0.03]	1.00 _[0.00]
$\hat{\delta}_i^{gls}$	0.05 _[0.04]	0.06 _[0.04]	0.15 _[0.08]	0.17 _[0.08]	1.00 _[0.00]
i	6	7	8	9	10
$\hat{\delta}_i^{ols}$	0.89 _[0.11]	0.89 _[0.11]	0.91 _[0.11]	1.11 _[0.13]	1.11 _[0.13]
$\hat{\delta}_i^{als}$	1.00 _[0.00]	1.00 _[0.00]	1.00 _[0.00]	1.00 _[0.00]	1.00 _[0.00]
$\hat{\delta}_i^{gls}$	1.00 _[0.00]	1.00 _[0.00]	1.00 _[0.00]	1.00 _[0.00]	1.00 _[0.00]
i	11	12	13	14	15
$\hat{\delta}_i^{ols}$	1.11 _[0.13]	1.12 _[0.13]	1.36 _[0.17]	1.36 _[0.17]	1.36 _[0.17]
$\hat{\delta}_i^{als}$	1.00 _[0.00]	1.00 _[0.00]	1.00 _[0.00]	1.00 _[0.00]	1.00 _[0.00]
$\hat{\delta}_i^{gls}$	1.00 _[0.00]	1.00 _[0.00]	1.00 _[0.00]	1.00 _[0.00]	1.00 _[0.00]
i	16	17	18	19	20
$\hat{\delta}_i^{ols}$	1.36 _[0.17]	1.69 _[0.27]	1.71 _[0.27]	1.71 _[0.27]	1.71 _[0.27]
$\hat{\delta}_i^{als}$	1.00 _[0.00]	1.00 _[0.00]	1.00 _[0.00]	1.00 _[0.00]	1.00 _[0.00]
$\hat{\delta}_i^{gls}$	1.00 _[0.00]	1.00 _[0.00]	1.00 _[0.00]	1.00 _[0.00]	1.00 _[0.00]

Table 5: The estimators of the autoregressive parameters of the VAR(1) model for the balance data for the U.S..

Parameter	θ_1	θ_2	θ_3	θ_4
ALS estimate	0.33 _[0.08]	0.02 _[0.02]	-0.35 _[0.30]	-0.07 _[0.08]
OLS estimate	0.45 _[0.23]	0.00 _[0.02]	-1.02 _[0.60]	0.1 _[0.17]

Table 6: The balance data for the U.S.: the p -values of the ARCH-LM tests (in %) for the components of the ALS-residuals of a VAR(1).

lags	2	5	10
$\check{\epsilon}_{1t}$	22.26	45.05	36.44
$\check{\epsilon}_{2t}$	25.32	73.32	77.18

Table 7: The p -values of the portmanteau tests (in %) for the checking of the adequacy of the VAR(1) model for the U.S. trade balance data.

m	5	15
LB_m^S	0.00	0.01
LB_m^{OLS}	50.80	99.94
LB_m^{ALS}	6.36	15.95
\widetilde{LB}_m^{OLS}	0.00	7.87
\widetilde{LB}_m^{ALS}	5.61	15.50

Table 8: The balance data for the U.S.: the test statistics of the portmanteau tests used for checking the adequacy of the VAR(1) model. The $\underline{\tilde{Q}}_m^{OLS}$ and $\underline{\tilde{Q}}_m^{ALS}$ correspond to the statistics of the LB version of the Katayama portmanteau tests with standard asymptotic distribution.

m	5	15
\tilde{Q}_m^{OLS}	6.84	106.34
\tilde{Q}_m^{ALS}	25.73	66.83
$\underline{\tilde{Q}}_m^{OLS}$	6.84	106.34
$\underline{\tilde{Q}}_m^{ALS}$	48.73	66.83

Table 9: The estimators of the autoregressive parameters of the VAR(4) model for the U.S. energy-transportation price indexes.

Parameter	θ_1	θ_2	θ_3	θ_4
ALS estimate	0.36 _[0.08]	0.37 _[0.14]	0.10 _[0.04]	0.43 _[0.08]
OLS estimate	0.74 _[0.32]	1.08 _[0.67]	-0.08 _[0.13]	0.10 _[0.28]
Parameter	θ_5	θ_6	θ_7	θ_8
ALS estimate	0.06 _[0.08]	-0.02 _[0.14]	-0.09 _[0.04]	-0.13 _[0.08]
OLS estimate	-0.53 _[0.35]	-1.26 _[0.73]	0.10 _[0.14]	0.27 _[0.30]
Parameter	θ_9	θ_{10}	θ_{11}	θ_{12}
ALS estimate	0.18 _[0.08]	0.13 _[0.14]	-0.05 _[0.04]	0.01 _[0.08]
OLS estimate	0.21 _[0.24]	0.13 _[0.52]	-0.05 _[0.14]	0.03 _[0.31]
Parameter	θ_{13}	θ_{14}	θ_{15}	θ_{16}
ALS estimate	0.17 _[0.08]	0.15 _[0.14]	-0.07 _[0.04]	0.03 _[0.08]
OLS estimate	0.32 _[0.26]	0.64 _[0.57]	-0.17 _[0.15]	0.31 _[0.32]

Table 10: The p -values of the portmanteau tests (in %) for the checking of the adequacy of the VAR(4) model for the U.S. energy-transportation price indexes (n.a.: not available).

m	3	6	12
LB_m^S	n.a.	2.08	0.00
LB_m^{OLS}	100.00	100.00	100.00
LB_m^{ALS}	85.44	98.72	10.07
\widetilde{LB}_m^{OLS}	n.a.	n.a.	n.a.
\widetilde{LB}_m^{ALS}	n.a.	n.a.	n.a.

Table 11: VAR modeling of the energy-transportation price indexes: the test statistics of the portmanteau tests used for checking the adequacy of the VAR(4) model.

m	3	6	12
\tilde{Q}_m^{OLS}	1.87	18.06	117.51
\tilde{Q}_m^{ALS}	5.03	9.50	57.69
$\underline{\tilde{Q}}_m^{OLS}$	n.a.	n.a.	n.a.
$\underline{\tilde{Q}}_m^{ALS}$	n.a.	n.a.	n.a.

Table 12: VAR modeling of the energy-transportation price indexes: the weights of the non standard distributions of the portmanteau tests used for checking the adequacy of the VAR(4) model with $m = 3$.

i	1	2	3	4	5	6	7	8	9	10	11	12
$\hat{\delta}_i^{ols}$	0.05	0.48	1.94	2.05	2.66	2.85	4.97	6.58	10.51	16.06	64.14	312.18
$\hat{\delta}_i^{als}$	0.01	0.09	0.24	0.89	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

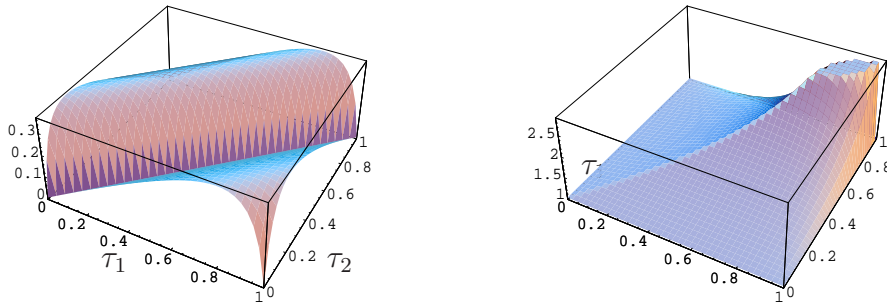


Figure 2: The same as in Figure 1 but for $\tau_1 \neq \tau_2$ in general.

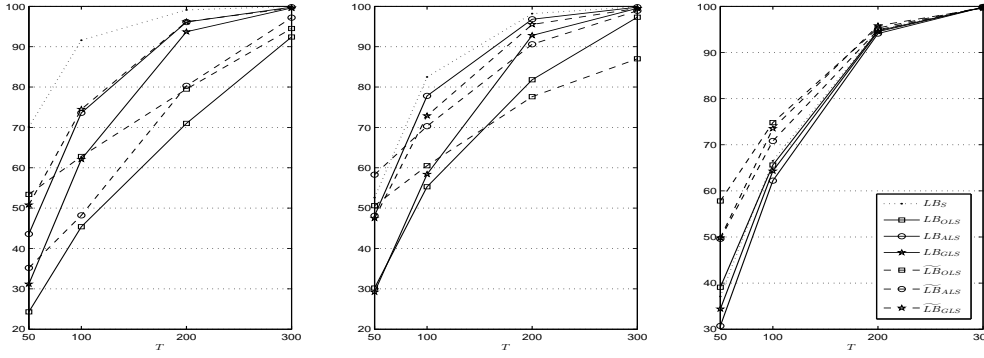


Figure 3: Empirical power (in %) of the portmanteau tests with $m = 10$. The adequacy of a VAR(1) model to VAR(2) processes is tested. The innovations are homoscedastic on the right. The variance exhibits a break at $T/2$ on the left and have a trending behavior in the middle.

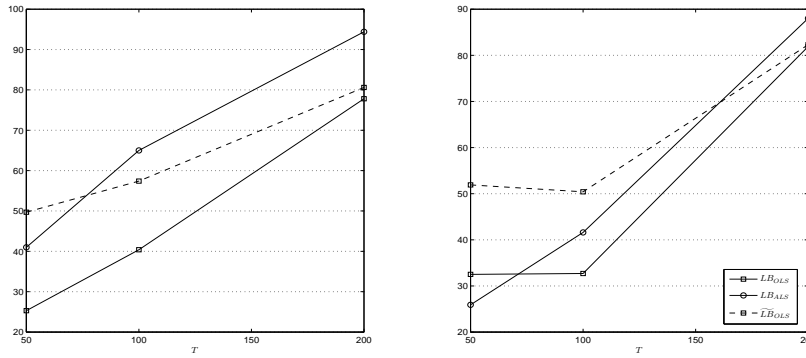


Figure 4: Empirical power (in %) of the portmanteau tests with $m = 10$. The non correlation of VAR(1) processes is tested. The variance have a trending behavior on the left and exhibits an abrupt shift on the right.

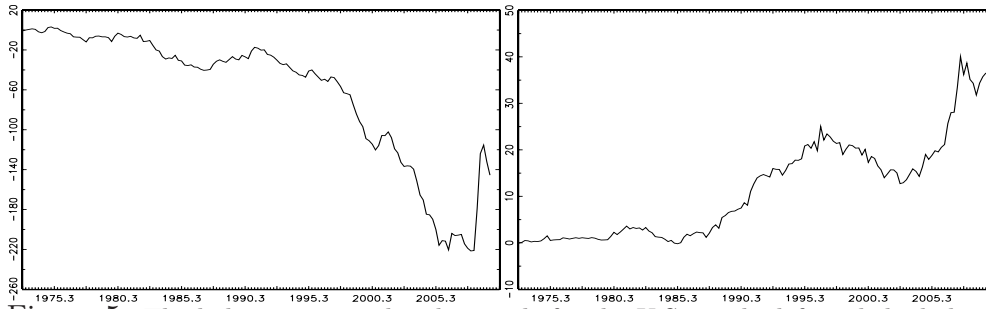


Figure 5: The balance on merchandise trade for the U.S. on the left and the balance on services for the U.S. on the right in billions of dollars from 1/1/1970 to 10/1/2009, $T=160$. Data source: The research division of the federal reserve bank of Saint Louis, www.research.stlouis.org.

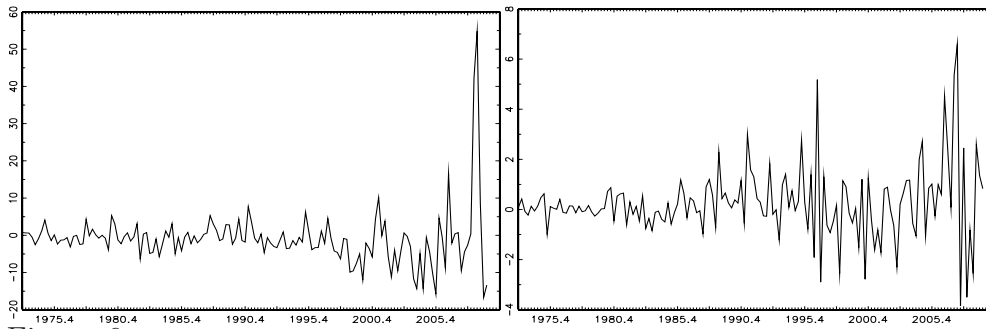


Figure 6: The differences of the balance on merchandise trade (on the left) and of the balance on services for the U.S. (on the right).

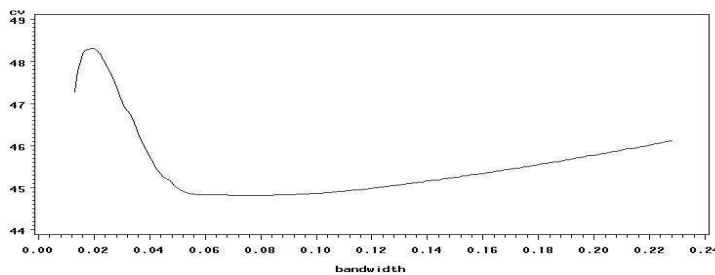


Figure 7: The cross validation score (CV) for the ALS estimation of the VAR(1) model for the differences of the balance on merchandise trade and on services in the U.S..

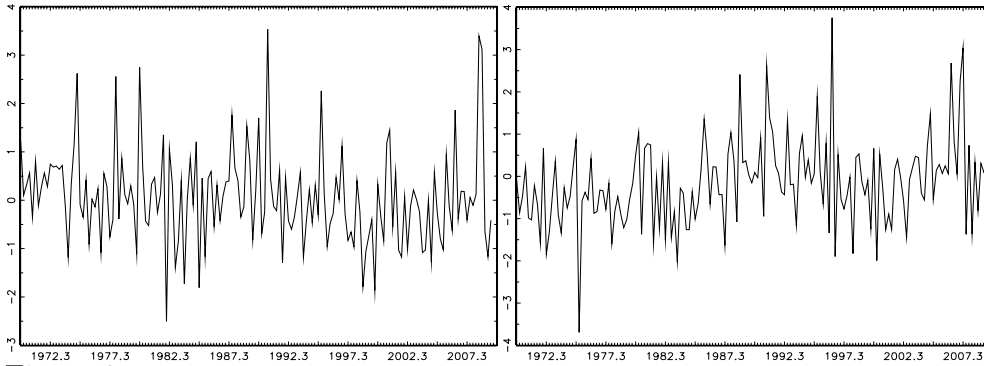


Figure 8: The ALS residuals of a VAR(1) for the differences of the balance on merchandise trade

and on services in the U.S.. The first component of the ALS residuals is on the left and the second is on the right.

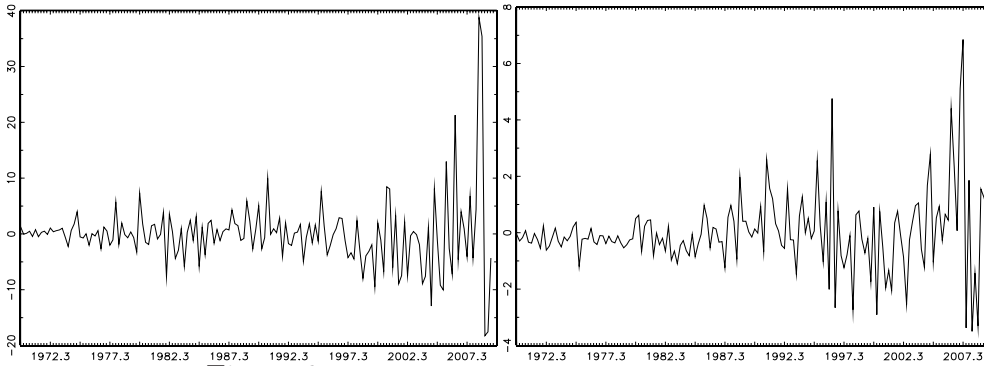


Figure 9: The same as in Figure 8 but for the OLS residuals.

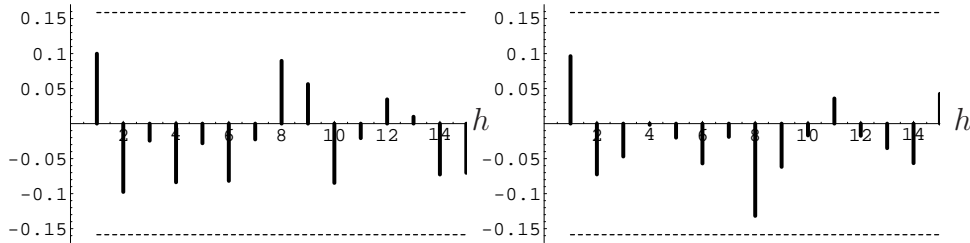


Figure 10: The balance data for the U.S.: the autocorrelations of the squares of the first component of the ALS residuals (on the left) and of the second component of the ALS residual (on the right).

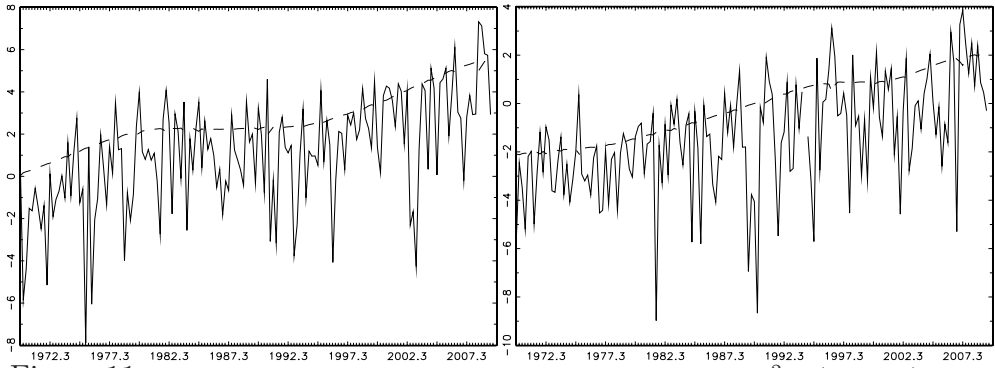


Figure 11: The balance data for the U.S.: the logarithms of the \hat{u}_{1t}^2 's (full line) and the logarithms of the non parametric estimates of $\text{Var}(u_{1t})$ (dotted line) on the left and the same for the \hat{u}_{2t}^2 's and $\text{Var}(u_{2t})$ on the right.

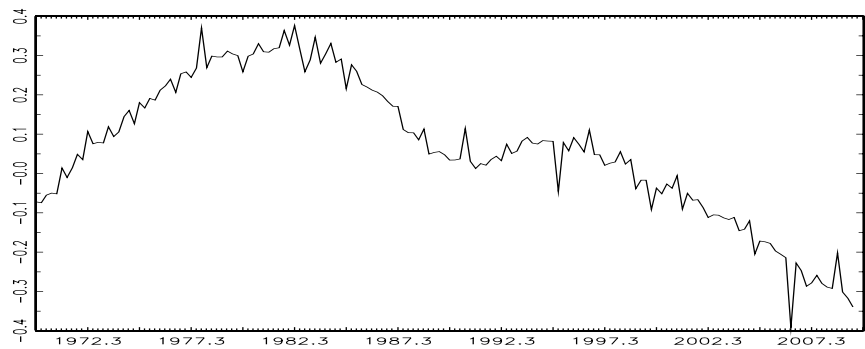


Figure 12: The balance data for the U.S.: estimation of the correlation between of the components of the error process.

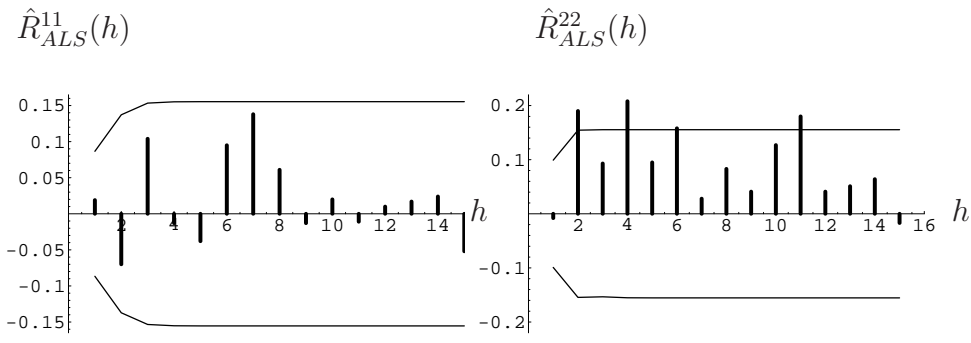


Figure 13: The balance data for the U.S.: the ALS residual autocorrelations $\hat{R}_{ALS}^{11}(h)$ (on the left) and $\hat{R}_{ALS}^{22}(h)$ (on the right), with obvious notations. The 95% confidence bounds are obtained using (3.7) and (5.3).

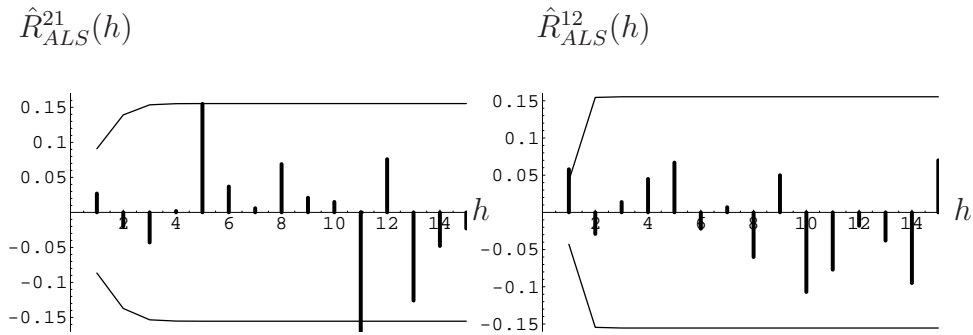


Figure 14: The same as in Figure 13 but for $\hat{R}_{ALS}^{21}(h)$ (on the left) and $\hat{R}_{ALS}^{12}(h)$ (on the right).

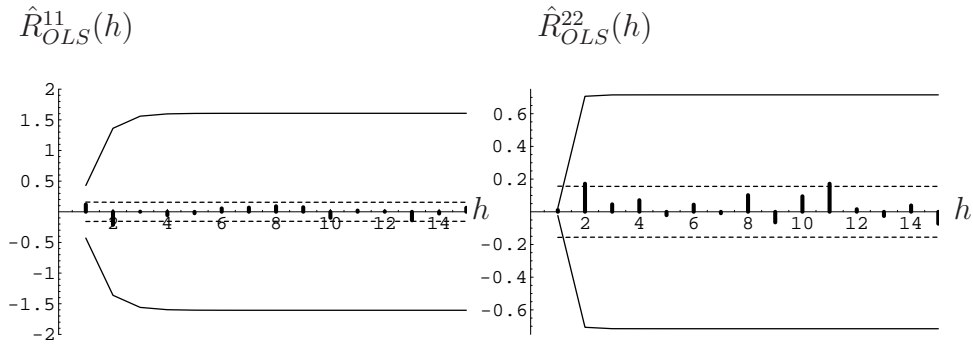


Figure 15: The balance data for the U.S.: the OLS residual autocorrelations $\hat{R}_{OLS}^{11}(h)$ (on the left) and $\hat{R}_{OLS}^{22}(h)$ (on the right). The full lines 95% confidence bounds are obtained using (3.13). The dotted lines 95% confidence bounds are obtained using the standard result (3.9).

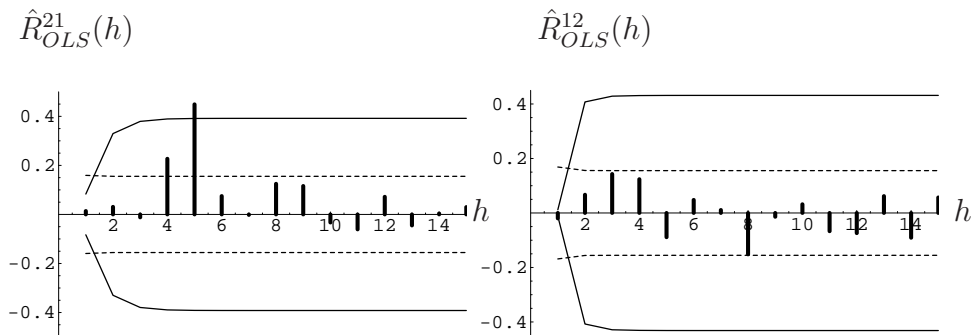


Figure 16: The same as in Figure 13 but for $\hat{R}_{OLS}^{21}(h)$ (on the left) and $\hat{R}_{OLS}^{12}(h)$ (on the right).

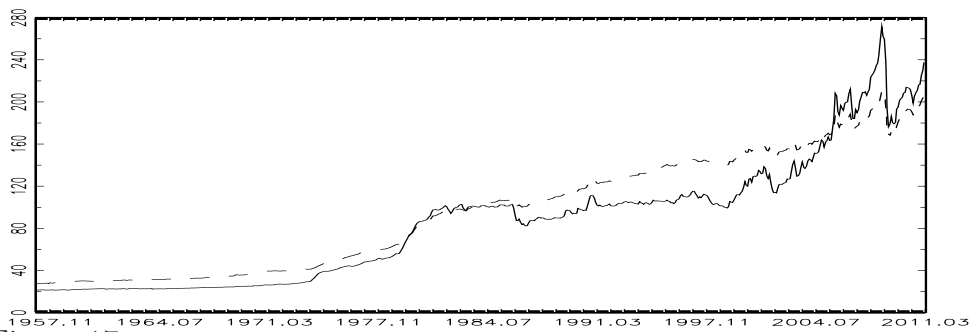


Figure 17: The energy price index (full line) and the transportation price index (dotted line) in the U.S. from 1/1/1957 to 2/1/2011, $T=648$. Data source: The research division of the federal reserve bank of Saint Louis, www.research.stlouis.org.

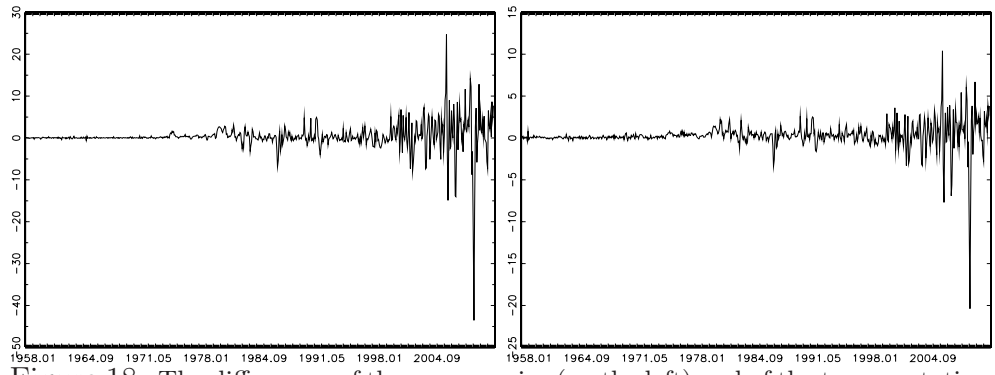


Figure 18: The differences of the energy price (on the left) and of the transportation price indexes for the U.S. (on the right).

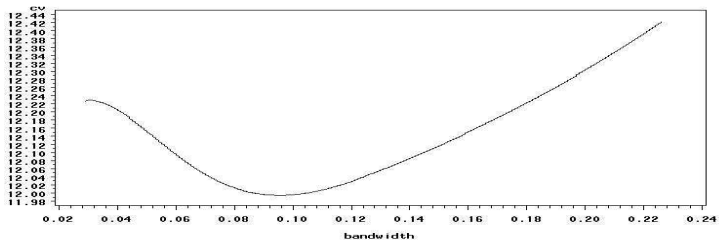


Figure 19: The cross validation score (CV) for the ALS estimation of the VAR(4) model for the differences of the energy-transportation price indexes in the U.S..

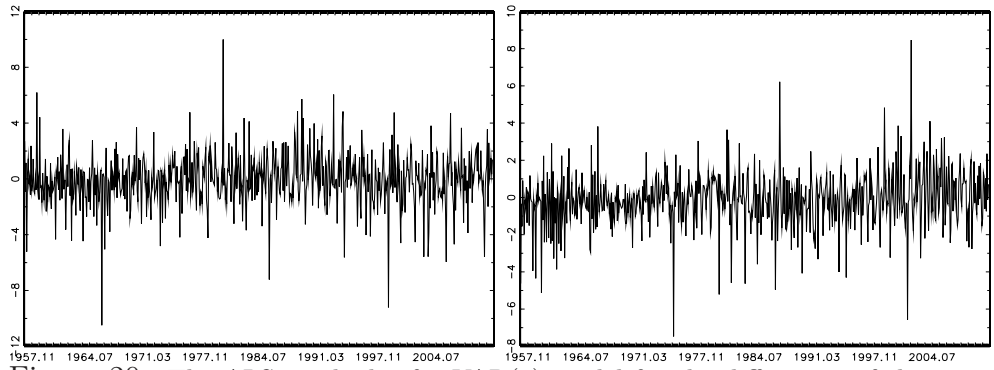


Figure 20: The ALS residuals of a VAR(4) model for the differences of the energy and transportation price indexes for the U.S.. The first component of the ALS residuals is on the left and the second is on the right.

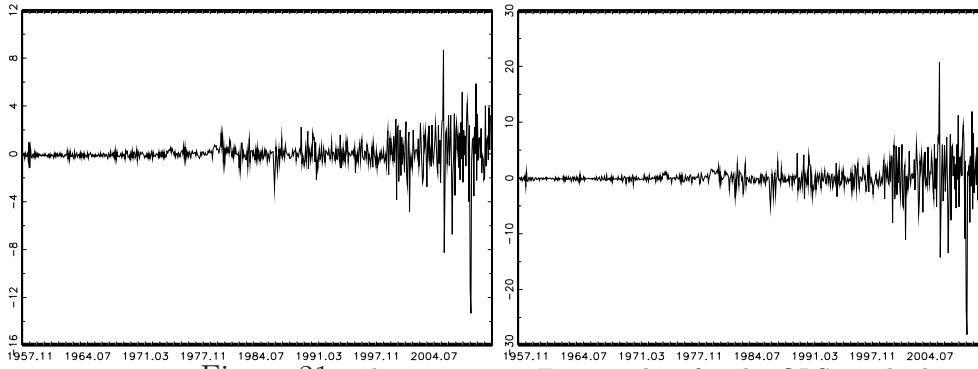


Figure 21: The same as in Figure 8 but for the OLS residuals.

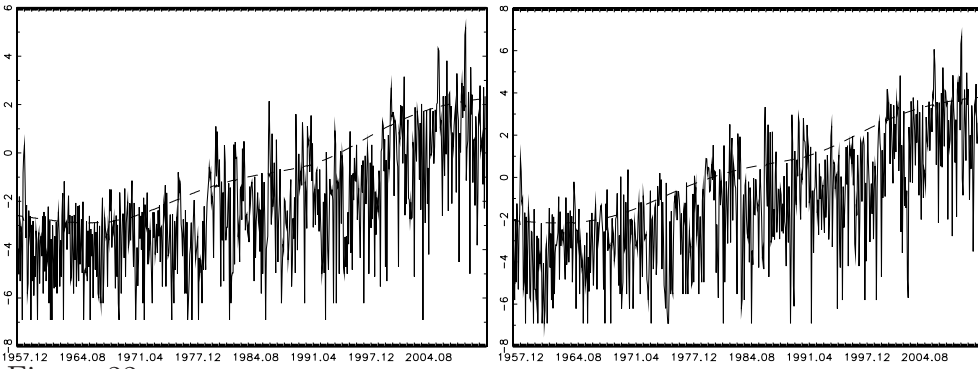


Figure 22: The energy-transportation data for the U.S.: the logarithms of the \hat{u}_{1t}^2 's (full line) and the logarithms of the non parametric estimation of $\text{Var}(u_{1t})$ (dotted line) on the left and the same for the \hat{u}_{2t}^2 's and $\text{Var}(u_{2t})$ on the right.

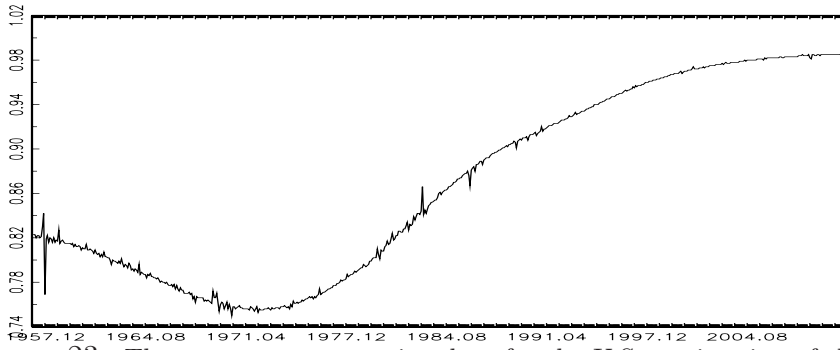


Figure 23: The energy-transportation data for the U.S.: estimation of the correlation between the components of the error process.

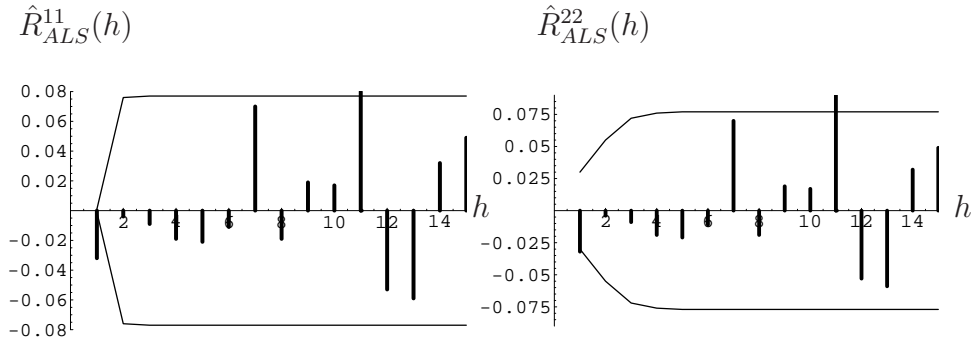


Figure 24: The energy-transportation data for the U.S.: the ALS residual autocorrelations $\hat{R}_{ALS}^{11}(h)$ (on the left) and $\hat{R}_{ALS}^{22}(h)$ (on the right), with obvious notations. The 95% confidence bounds are obtained using (3.7) and (5.3).

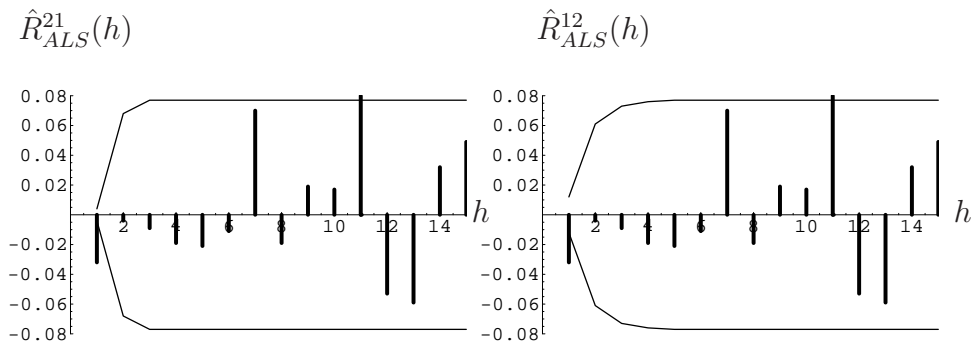


Figure 25: The same as in Figure 24 but for $\hat{R}_{ALS}^{21}(h)$ (on the left) and $\hat{R}_{ALS}^{12}(h)$ (on the right).

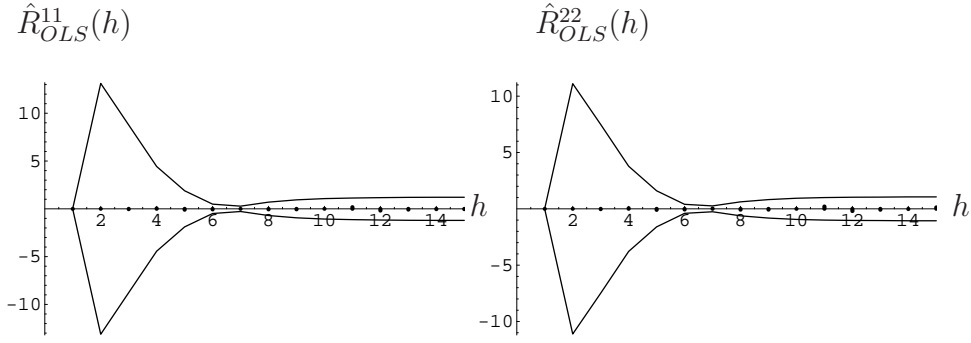


Figure 26: The energy-transportation data for the U.S.: the OLS residual autocorrelations $\hat{R}_{OLS}^{11}(h)$ (on the left) and $\hat{R}_{OLS}^{22}(h)$ (on the right). The full line 95% confidence bounds are obtained using (3.13). The dotted lines 95% confidence bounds are obtained using the standard result (3.9).

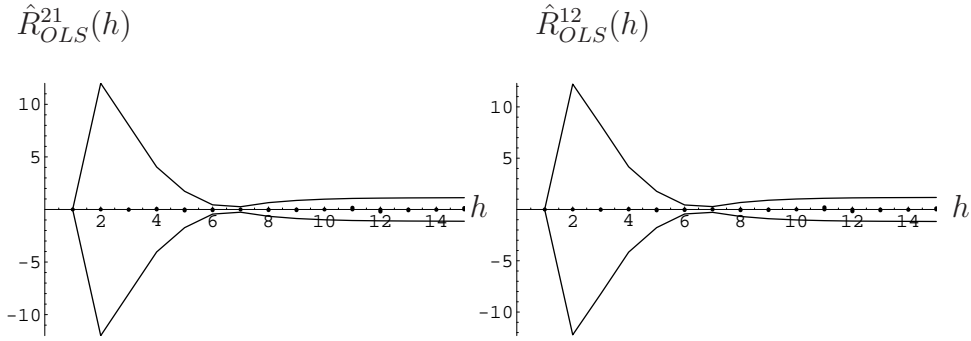


Figure 27: The same as in Figure 26 but for $\hat{R}_{OLS}^{21}(h)$ (on the left) and $\hat{R}_{OLS}^{12}(h)$ (on the right).

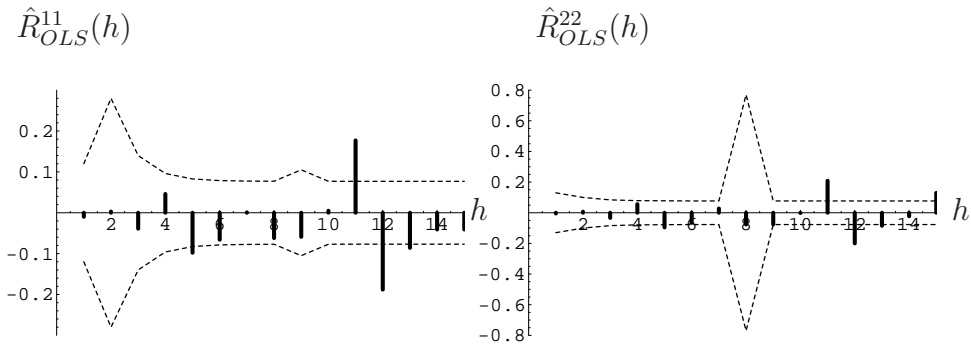


Figure 28: The energy-transportation data for the U.S.: the OLS residual autocorrelations $\hat{R}_{OLS}^{11}(h)$ (on the left) and $\hat{R}_{OLS}^{22}(h)$ (on the right). The dotted lines 95% confidence bounds are obtained using the standard result (3.9).

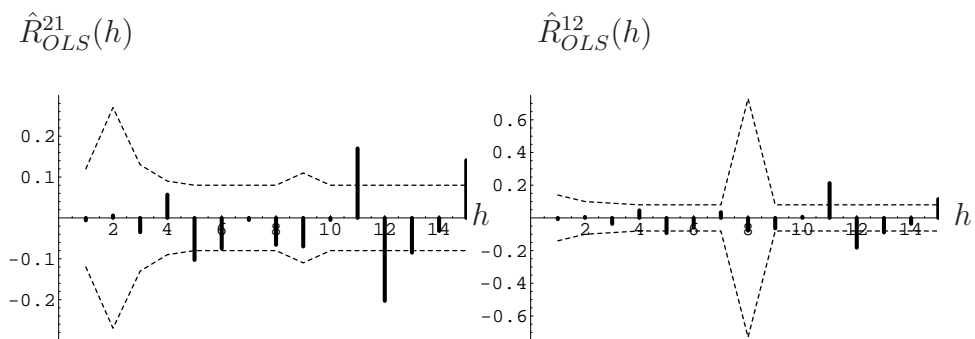


Figure 29: The same as in Figure 28 but for $\hat{R}_{OLS}^{21}(h)$ (on the left) and $\hat{R}_{OLS}^{12}(h)$ (on the right).

This figure "cvbalanc.jpg" is available in "jpg" format from:

<http://arxiv.org/ps/1105.3638v2>

This figure "cvenergy.jpg" is available in "jpg" format from:

<http://arxiv.org/ps/1105.3638v2>