Adaptive Estimation of VAR models with Time-Varying Variance: Application to Testing the VAR Order

Valentin Patilea(*), Hamdi Raïssi(**)
(*)CREST-Ensai & IRMAR
(**) IRMAR-INSA de Rennes

Résumé

Linear Vector AutoRegressive (VAR) models where the innovations could be unconditionally heteroscedastic are considered. In this framework we propose Ordinary Least Squares (OLS), Generalized Least Squares (GLS) and Adaptive Least Squares (ALS) procedures. The GLS estimator requires the knowledge of the time-varying variance structure while in the ALS approach the unknown variance is estimated by kernel smoothing with the outer product of the OLS residuals vectors. Different bandwidths for the different cells of the time-varying variance matrix are allowed. We derive the asymptotic distribution of the proposed estimators for the VAR coefficients and compare their properties. The ALS estimator is shown to be asymptotically equivalent to the infeasible GLS estimator. This asymptotic equivalence is obtained uniformly with respect to the bandwidth(s) and hence justifies data-driven bandwidth rules. Using these results we investigate the portmanteau tests when the innovations have time-varying variance and propose new corrected versions. The theoretical results are illustrated using a U.S. macro-economic data set.

Keywords : Vector Autoregressive Model, Kernel smoothing, Portmanteau tests

Introduction

In the recent years the study of linear time series models in the context of unconditionally heteroscedastic innovations has become of increased interest. This interest may be explained by the strong empirical evidence of non-constant unconditional volatility in macro-economic and financial data. For instance Sensier and Van Dijk (2004) found that approximately 80% among 214 US macro-economic data exhibit a volatility break. Stărică (2003) hypothesized that the returns of the Standard and Poors 500 stock market index have a non constant unconditional volatility. These findings stimulated an interest on the effects of non-stationary volatility in time series analysis. Reference can be made to Cavaliere, Rahbek and Taylor (2010) and Kim and Park (2010) who investigated the statistical analysis of cointegrated systems with non constant volatility.

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

The structure of the paper is as follows. Section 1 outlines the heteroscedastic VAR model, introduces the assumptions and the definitions of the OLS and GLS estimators. Section 2 contains the results on the asymptotic behavior of the OLS and the infeasible Generalized Least Squares estimators. The ALS estimator based on kernel smoothing of OLS residuals is proposed in Section 3 as a feasible asymptotically equivalent version of GLS estimator. The asymptotic equivalence between ALS and GLS estimators is proved uniformly in the bandwidths involved in volatility estimation. The asymptotic normality of the OLS, GLS and ALS residual autocovariances and autocorrelations is established 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. 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 they are asymptotically equivalent. This result is used to introduce portmanteau tests based on the ALS residuals. The mathematical proofs of the results stated in the paper are available in Patilea and Raïssi (2010, 2011). As an application of our theoretical findings we specify the autoregressive dynamics of a bivariate system of U.S. economic variables in section 7 : the U.S. balance on services and balance on merchandise trade data.

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

1 The model and least squares estimation of the parameters

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

$$X_t = A_1 X_{t-1} + \cdots + A_p X_{t-p} + u_t \qquad (1)$$
$$u_t = H_t \epsilon_t,$$

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

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

The assumption **A1** generalizes the assumption of Xu and Phillips (2008) to the multivariate case. From the assumptions $E(\epsilon_t \mid \mathcal{F}_{t-1}) = 0$ and $E(\epsilon_t \epsilon_t' \mid \mathcal{F}_{t-1}) = I_d$, the innovations are possibly serially dependent at the level of the third or higher order moments. However since $G(r)$ is deterministic, we do not allow the error process to follow a multivariate GARCH model. Cavaliere, Rahbek and Taylor (2010) considered similar volatility structure to ours. Their assumption is slightly different from **A1** in the sense that they do not require a Lipschitz condition and allow for a countable number of jumps. Boswijk and Zu (2007) allow the matrix H_t to be possibly stochastic, but requires the volatility process to be continuous with other additional assumptions, which in particular excludes important cases like abrupt shifts.

If we suppose that the volatility matrix H_t is constant, we retrieve the standard homoscedastic case. However the standard assumption on the errors is often considered to be too restrictive for macroeconomic or financial applications. Indeed many applied studies pointed out that such data may display unconditional non-stationary volatility (see e.g. Ramey and Vine (2006)). Stărică and Granger (2005) found that when large samples of stock returns are considered, taking into account shifts for the unconditional volatility instead of assuming a stationary model as a GARCH(1,1) improve the volatility forecasts.

Let us denote by $\theta_0 = (\mathrm{vec}\,(A_1)', \dots, \mathrm{vec}\,(A_p)')' \in \mathbf{R}^{pd^2}$ the vector of the true parameters. The equation (1) becomes

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

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

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

where

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

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

$$\hat{\theta}_{GLS} = \hat{\Sigma}_{\underline{\tilde{X}}}^{-1} \text{vec}\left(\hat{\Sigma}_{\underline{X}}\right), \tag{2}$$

with

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

Note that since H_t is assumed invertible, Σ_t is positive definite for all t. If we suppose that the volatility matrix Σ_t is constant in time, it is easy to see that $\hat{\theta}_{GLS} = \hat{\theta}_{OLS}$. However the GLS estimator is in general infeasible since the true volatility matrix appears in the expression (2). In the next section we compare the efficiency of the OLS and GLS estimators.

2 Asymptotic behaviour of the estimators

In order to state the first result of the paper, let us introduce the matrix

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

of dimension $pd \times pd$ and $e_p(1)$ the vector of dimension p such that the first component is equal to one and zero elsewhere. Note that if $\tilde{u}_t = (u'_t, 0 \dots, 0)'$, $\tilde{X}_t = \Delta \tilde{X}_{t-1} + \tilde{u}_t$. The following proposition gives the asymptotic behavior of the OLS and GLS estimators. For the sake of brevity we only investigate the asymptotic normality, the consistency is in some sense an easier matter and is hence omitted.

Proposition 1. *If Assumption* **A1** *holds true, then :*

1.

$$T^{\frac{1}{2}}(\hat{\theta}_{GLS} - \theta_0) \Rightarrow \mathcal{N}(0, \Lambda_1^{-1}), \tag{3}$$

where

$$\Lambda_1 = \int_0^1 \sum_{i=0}^{\infty} \left\{ \Delta^i (e_p(1) e_p(1)' \otimes \Sigma(r)) \Delta^{i'} \right\} \otimes \Sigma(r)^{-1} dr$$

is positive definite ;

2.

$$T^{\frac{1}{2}}(\hat{\theta}_{OLS} - \theta_0) \Rightarrow \mathcal{N}(0, \Lambda_3^{-1}\Lambda_2\Lambda_3^{-1}), \tag{4}$$

where

$$\Lambda_2 = \int_0^1 \sum_{i=0}^{\infty} \left\{ \Delta^i (e_p(1)e_p(1)' \otimes \Sigma(r))\Delta^{i'} \right\} \otimes \Sigma(r) dr$$

and

$$\Lambda_3 = \int_0^1 \sum_{i=0}^{\infty} \left\{ \Delta^i (e_p(1)e_p(1)' \otimes \Sigma(r))\Delta^{i'} \right\} \otimes I_d \, dr$$

are positive definite ;

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

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

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

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

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

although here the error process could present some dependence at the level of third or higher order moments. Note that in the homoscedastic case the OLS and ALS estimator have the same efficiency.

In the case where $\Sigma(r) = \sigma^2(r)I_d$ with $\sigma^2(\cdot)$ a real-valued function one obtains the following simplified formulae for Λ_k, $k = 1, 2, 3$:

$$\Lambda_1 = \sum_{i=0}^{\infty} \left\{ \tilde{\psi}_i(\mathbf{1}_{p\times p} \otimes I_d)\tilde{\psi}_i \right\} \otimes I_d, \tag{5}$$

and

$$\Lambda_2 = \int_0^1 \Sigma(r)^2 dr \Lambda_1, \qquad \Lambda_3 = \int_0^1 \Sigma(r) dr \Lambda_1.$$

In the case $d = 1$ these simplified formulae coincide with the equations (10) and (5) in Xu and Phillips (2008). Thereby a nice feature of the GLS estimator in the univariate case and in the multivariate case where $\Sigma(r) = \sigma^2(r)I_d$ is that the covariance matrix of the asymptotic distribution does not depend on the volatility function $\Sigma(r)$. Nevertheless an example provided in the extended version of the paper shows that (5) does not hold in the general multivariate framework and the asymptotic covariance matrix in (3) depends on the volatility function $\Sigma(r)$.

It appears that the GLS estimator is more efficient than the OLS estimator in general when the matrix Σ_t is time-varying. Nevertheless the assumption of known volatility structure needed to construct the GLS estimator could be unrealistic in practice. Moreover, the asymptotic distribution of the GLS estimator depends on the unknown volatility. In the OLS estimation approach only the asymptotic distribution of the coefficients estimator depends on the unknown volatility. In addition, we can provide simple consistent estimators of Λ_2 and Λ_3, which could be further used for instance to build confidence intervals for the OLS estimators. For the purpose of estimation of Λ_2 and Λ_3 let $\hat{u}_t := X_t - (\tilde{X}_{t-1}' \otimes I_d)\hat{\theta}_{OLS}$ denote the OLS residuals.

Proposition 2. *Under Assumption* **A1** *we have*

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

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

3 Adaptive estimation

In the previous section we pointed out that the GLS estimator is generally infeasible in applications. Therefore we consider a feasible weighted estimator obtained using non-parametric estimation of the volatility function. Our approach generalizes the work of Xu and Phillips (2008) to the multivariate case. Let us denote by $A\odot B$ the Hadamard (entrywise) product of two matrices of same dimension A and B. Define the symmetric matrix

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

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

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

with b_{kl} the bandwidth and

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

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

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

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

where $\nu_T > 0$, $T\ge 1$, is a sequence of real numbers decreasing to zero at a suitable rate that will be specified below. Our simulation experience reported in Patilea and Raïssi (2010, 2011) indicates that in applications with moderate and large samples ν_T could be even set equal to 0.

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

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

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

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

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

with

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

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

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

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

Assumption A2 : (i) The kernel $K(\cdot)$ is a bounded density function defined on the real line such that $K(\cdot)$ is nondecreasing on $(-\infty, 0]$ and decreasing on $[0, \infty)$ and $\int_{\mathbf{R}} v^2 K(v) dv < \infty$. 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 $\int_{\mathbf{R}} |s\mathcal{F}[K](s)| \, ds < \infty$.

(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 $b_T + 1/T b_T^{2+\gamma} \to 0$ as $T \to \infty$, for some $\gamma > 0$.

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

In the sequel, 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 $T \to \infty$ 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 the adaptive estimators uniformly w.r.t the bandwidths.

Proposition 3. *Under* **A1'** *and* **A2** *and provided $T\nu_T^2 \to 0$, uniformly w.r.t. $b_{kl} \in \mathcal{B}_T$ as $T \to \infty$*

$$\check{\Lambda}_1 := \check{\Sigma}_{\underline{\tilde{X}}} = \Lambda_1 + o_p(1), \tag{6}$$

and

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

Proposition 3 shows that the ALS and GLS estimators have the same asymptotic behavior, that is the ALS estimator is consistent in probability and $\sqrt{T}-$asymptotically normal as soon as the GLS estimator has such properties. The results remains true even if the bandwidths $b_{kl} \in \mathcal{B}_T$ are data dependent. It is clear from our results that using the adaptive estimators of the autoregressive parameters instead of the OLS estimators lead to a gain of efficiency. Consequently it is shown in Patilea and Raïssi (2010) that the test for zero restrictions in the autoregressive parameters based on the ALS estimators are more powerful than the test based on the OLS estimators. Simulation results in Patilea and Raïssi (2010) show that the gain of efficiency of the ALS estimator over the OLS estimator can be substantial.

4 Asymptotic behavior of the residual autocovariances

First note that the adaptive approach 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 the estimate of the matrix H_t. In the sequel we assume that the H_t's are positive definite matrices, so that $G(r)$ is identified as the square root of $\Sigma(r)$ and this is a convenient choice for the 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. Let us define the GLS-based estimates of ϵ_t

$$\hat{\epsilon}_t = H_t^{-1} X_t - H_t^{-1}(\tilde{X}'_{t-1} \otimes I_d)\hat{\theta}_{GLS},$$

and the residual autocovariances

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

where we recall that the $\hat{u}_t$'s are the OLS based residuals. In general the estimated residuals $\hat{\epsilon}_t$ as well as the autocovariances $\hat{\Gamma}_{GLS}(h)$ 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

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

Now, let $\Sigma_G = \int_0^1 \Sigma(r)dr$, $\Sigma_{G^{\otimes 2}} = \int_0^1 \Sigma(r)^{\otimes 2} dr$ and

$$\Phi_m^u = \sum_{i=0}^{m-1} \{e_m(i+1)e_p(1)' \otimes \Sigma_G \otimes I_d\} \left\{\Delta^{i\prime} \otimes I_d\right\}, \tag{7}$$

$$\Lambda_m^{u,\theta} = \sum_{i=0}^{m-1} \{e_m(i+1)e_p(1)' \otimes \Sigma_{G^{\otimes 2}}\} \left\{\Delta^{i\prime} \otimes I_d\right\}, \tag{8}$$

$$\Lambda_m^{\epsilon,\theta} = \sum_{i=0}^{m-1} \left\{e_m(i+1)e_p(1)' \otimes \int_0^1 G(r)' \otimes G(r)^{-1} dr\right\} \left\{\Delta^{i\prime} \otimes I_d\right\}, \tag{9}$$

and $\Lambda_m^{u,u} = I_m \otimes \Sigma_{G^{\otimes 2}}$, where $e_m(j)$ is the vector of dimension m such that the jth component is equal to one and zero elsewhere.

Proposition 4. *If model (1) is correct and Assumption* **A1** *holds true, we have*

$$T^{\frac{1}{2}}\hat{\gamma}_m^{OLS} \Rightarrow \mathcal{N}(0, \Sigma^{OLS}), \tag{10}$$

where

$$\Sigma^{OLS} = \Lambda_m^{u,u} - \Lambda_m^{u,\theta}\Lambda_3^{-1}\Phi_m^{u\,\prime} - \Phi_m^{u}\Lambda_3^{-1}\Lambda_m^{u,\theta\,\prime} + \Phi_m^{u}\Lambda_3^{-1}\Lambda_2\Lambda_3^{-1}\Phi_m^{u\,\prime},$$

$$T^{\frac{1}{2}}\hat{\gamma}_m^{GLS} \Rightarrow \mathcal{N}(0, \Sigma^{GLS}), \tag{11}$$

where

$$\Sigma^{GLS} = I_{d^2m} - \Lambda_m^{\epsilon,\theta}\Lambda_1^{-1}\Lambda_m^{\epsilon,\theta\,\prime}. \tag{12}$$

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

Let us discuss the conclusions of Proposition 4 in some particular situations. In the case where $\Sigma(\cdot) = \sigma^2(\cdot)I_d$ (that includes the univariate AR(p) models) for some positive scalar function $\sigma(\cdot)$, we have

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

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

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

where $\Sigma_u = G_uG_u'$, and

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

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

$$\Sigma^{OLS} = I_m \otimes \Sigma_u^{\otimes 2} - E\left[u_t^m\tilde{X}_t'\right]E\left[\tilde{X}_t\tilde{X}_t'\right]^{-1}E\left[u_t^m\tilde{X}_t'\right]' \otimes \Sigma_u, \tag{13}$$

$$\Sigma^{GLS} = I_{d^2m} - E\left[\epsilon_t^m\tilde{X}_t'\right]E\left[\tilde{X}_t\tilde{X}_t'\right]^{-1}E\left[\epsilon_t^m\tilde{X}_t'\right]' \otimes I_d. \tag{14}$$

Formula (13) (resp. (14)) 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 (14) indicates that the homoscedastic (time-constant variance) case is another situation where Σ^{GLS} does not depend on error process variance Σ_u. However Proposition 4 shows that in general VAR models with time-varying variance the covariance matrix Σ^{GLS} depends on $\Sigma(\cdot)$.

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

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

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

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

Proposition 5. *If model (1) is correct and Assumption* **A1** *holds true, we have*

$$T^{\frac{1}{2}}\hat{\rho}_m^{OLS} \Rightarrow \mathcal{N}(0, \Psi^{OLS}), \tag{15}$$

where

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

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

$$T^{\frac{1}{2}}\hat{\rho}_m^{GLS} \Rightarrow \mathcal{N}(0, \Sigma^{GLS}), \tag{16}$$

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

Using Proposition 5, $\hat{S}_u$ and a consistent estimator of Σ^{OLS} (that can build in a similar way to that of Δ_m^{OLS}, see Section 5), one can easily build a consistent estimate of Ψ^{OLS} and confidence intervals for the OLS residual autocorrelations.

5 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)

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

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

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

The following result, a direct consequence of Proposition 4 equation (10), provides the asymptotic distribution of Q_m^{OLS} and $\tilde{Q}_m^{OLS}$.

Theorem 1. *If model (1) is correct and Assumption* **A1** *holds true, the statistics* Q_m^{OLS} *and* $\tilde{Q}_m^{OLS}$ *converge in law to*

$$U(\delta_m^{OLS}) = \sum_{i=1}^{d^2 m} \delta_i^{ols} U_i^2, \tag{17}$$

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

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

$\Sigma_G = \int_0^1 \Sigma(r)dr$ *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 (17) can be very different from a chi-square law. An example which illustrate this point is provided in Patilea and Raïssi (2011).

Estimates of the weights which appear in (17) can be obtained as follows. First, let us recall the following results proved by Patilea and Raïssi (2010) :

$$\hat{\Sigma}_{G^{\otimes 2}} := T^{-1} \sum_{t=2}^{T} \hat{u}_{t-1}\hat{u}_{t-1}' \otimes \hat{u}_t\hat{u}_t' = \Sigma_{G^{\otimes 2}} + o_p(1), \tag{18}$$

$$\hat{\Sigma}_G := T^{-1} \sum_{t=1}^{T} \hat{u}_t\hat{u}_t' = \Sigma_G + o_p(1), \tag{19}$$

Using the results of Proposition 2 a consistent estimator of Φ_m^u and $\Lambda_m^{u,\theta}$ given in (7) and (8) is easily obtained by replacing $A_1, \dots, A_p$ with their OLS estimators in Δ and considering (18) and (19). Thus from this and the equations (18) and (19), one can easily define a consistent estimator of Δ_m^{OLS}. Denote the estimated eigenvalues of Δ_m^{OLS} by $\hat{\delta}_m^{OLS} = (\hat{\delta}_1^{ols}, \dots, \hat{\delta}_{d^2m}^{ols})'$.

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). 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 of uncorrelated error process (u_t) 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, for instance, using the Imhof algorithm (Imhof, 1961).

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

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

and clearly, $c_\sigma \geq 1$. If in addition $p = 0$, by Proposition 4 we have $\Sigma^{GLS} = I_{d^2m}$ and hence $\delta_m^{OLS} = c_\sigma(1, \cdots, 1)'$.

6 Adaptive portmanteau tests

First note that since we assumed a deterministic volatility structure $\mathrm{Cov}(u_t, u_{t-h}) = 0$ is equivalent to $\mathrm{Cov}(\epsilon_t, \epsilon_{t-h}) = 0$. Therefore an alternative way to build portmanteau tests for VAR(p) models with time-varying variance is to use approximations of the innovation

ϵ_t. A nonparametric estimate of the volatility function is needed for building such approximations. The ALS residuals, proxies of the infeasible GLS residuals, are defined as $\check{\epsilon}_t = \check{H}_t^{-1} X_t - \check{H}_t^{-1}(\tilde{X}'_{t-1} \otimes I_d)\hat{\theta}_{ALS}$, and the adaptive autocovariances and autocorrelations

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

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

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

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

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. 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 6. *If model (1) is correct and under the assumptions of Proposition 3, uniformly w.r.t. $b \in \mathcal{B}_T$*

$$T^{-1} \sum_{t=1}^{T} \check{H}'_t \otimes \check{H}_t^{-1} = \int_0^1 G(r)' \otimes G(r)^{-1} dr + o_p(1). \tag{20}$$

Moreover, given any $m \geq 1$,

$$T^{\frac{1}{2}} \left\{ \hat{\gamma}_m^{ALS} - \hat{\gamma}_m^{GLS} \right\} = o_p(1) \quad \text{and} \quad T^{\frac{1}{2}} \left\{ \hat{\rho}_m^{ALS} - \hat{\rho}_m^{GLS} \right\} = o_p(1), \tag{21}$$

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

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

and

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

Consider also the Ljung-Box type statistics

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

and

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

The following theorem is a direct consequence of (11) and Proposition 3.

Theorem 2. *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*

$$U(\delta_m^{ALS}) = \sum_{i=1}^{d^2m} \delta_i^{als} U_i^2, \tag{22}$$

as $T \to \infty$, where $\delta_m^{ALS} = (\delta_1^{als}, \dots, \delta_{d^2m}^{als})'$ 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 $\Lambda_m^{\epsilon,\theta}$ given in (9) by replacing $A_{01}, \dots, A_{0p}$ by their ALS estimators in K and using (20). Next we consider the estimate of Λ_1 given in (6). Plugging these estimates into the formula (12), we obtain a consistent estimator of Σ^{GLS} with eigenvalues $\hat{\delta}_m^{ALS} = (\hat{\delta}_1^{als}, \dots, \hat{\delta}_{d^2m}^{als})'$ 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 4) and hence the asymptotic distribution of the four test statistics in Theorem 2 would be $\chi^2_{d^2m}$, that means independent of the variance structure given by $\Sigma(\cdot)$. In the general case $p \geq 1$ where the autoregressive coefficients A_{0i}, $i = 1, ..., p$ have to be estimated, the matrix $I_{d^2m} - \Sigma^{GLS}$ being positive semi-definite, the eigenvalues $\delta_1^{als}, ..., \delta_{d^2m}^{als}$ are smaller or equal to 1. Since, in some sense, the unconditional heteroscedasticity is removed in the ALS residuals, one could expect that the $\chi^2_{d^2(m-p)}$ asymptotic approximation is reasonably accurate for the ALS tests. A theoretical example provided in Patilea and Raïssi (2011) indicates that this is may not the case. The asymptotic distribution we obtain for the ALS portmanteau statistics can be very different from the $\chi^2_{d^2(m-p)}$ approximation when the errors are heteroscedastic. Finally, recall that from Sections 2 and 3 using the adaptive estimators of the 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. In this case, at the validation stage, the ALS tests will not be more complicated than the OLS tests to implement.

Let us also point out that the eigenvalues $\delta_1^{als}, ..., \delta_{d^2m}^{als}$ 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) one can easily show that for large values of m, the law of $U(\delta_m^{ALS})$ is accurately approximated by a $\chi^2_{d^2(m-p)}$ distribution. However, in the general the multivariate setup the asymptotic distribution in (22) depend on the variance function $\Sigma(\cdot)$.

7 Illustrative example

For real data illustration we consider the first differences of the quarterly U.S. international finance data from January 1, 1970 to October 1, 2009 : the balance on services and the balance on merchandise trade in billions of USD. The length of the series is $T = 159$. From Figure 1 it seems that the series are stable but have a non constant volatility. The series are available seasonally adjusted from the website of the research division of the Federal Reserve Bank of Saint Louis.

In our VAR system the first component corresponds to the balance on merchandise trade and the second corresponds to the balance on services trade. We adjusted a VAR(1) model to capture the linear dynamics of the series. The ALS and OLS estimators are given in Table 1. The standard deviations into brackets are computed using the results (3) and (4). In accordance with the results in Sections 2 and 3, 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.

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 4 we plotted the ALS residual autocorrelations in Figure 2, and the OLS residual autocorrelations in Figure 3, where we denote

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

The ALS 95% confidence bounds obtained using (16) and (21) are displayed in Figure 2, while in Figure 3 we plotted the standard 95% confidence bounds obtained using (13) and the OLS 95 % confidence bounds obtained using (15). We can remark that the ALS residual autocorrelations are inside the confidence bands 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 possible presence of unconditional volatility in the analyzed series. In particular we note that the $\hat{R}^{21}_{OLS}(5)$ 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 the corresponding p-values are displayed in Table 2. It appears that the p-values of the standard tests are very small, that means 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 (17) is very different from the $\chi^2_{d^2(m-p)}$ standard distribution. For instance we obtained $\sup_{i\in\{1,\dots,d^2m\}}\left\{\hat{\delta}_i^{ols}\right\} = 11.18$ 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. We also noticed that the weights (not reported here but available upon request) of the sums in (17) and in (22) are quite different for our example.

BIBLIOGRAPHY

1. Boswijk, H.P., and Zu, Y. (2007), "Testing for cointegration with nonstationary volatility," Working Paper, University of Amsterdam.
2. 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.
3. Cavaliere, G., Rahbek, A., and Taylor, A.M.R. (2010), "Testing for co-integration in vector autoregressions with non-stationary volatility," *Journal of Econometrics*, 158, 7-24.
4. Chitturi, R. V. (1974), "Distribution of residual autocorrelations in multiple autoregressive schemes," *Journal of the American Statistical Association* 69, 928-934.
5. Hosking, J. R. M. (1980), "The multivariate portmanteau statistic," *Journal of the American Statistical Association* 75, 343-386.
6. Imhof, J. P. (1961), "Computing the distribution of quadratic forms in normal variables," *Biometrika* 48, 419-426.
7. Kim C., and Park, J. (2010), "Cointegrating Regressions with Time Heterogeneity," *Econometric Reviews*, 29, 397-438.
8. Ljung, G.M. and Box, G.E.P. (1978), "On measure of lack of fit in time series models," *Biometrika* 65, 297-303.
9. Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*. Springer, Berlin.
10. Patilea, V., and Raïssi, H. (2010), "Adaptive estimation of vector autoregressive models with time-varying variance : application to testing linear causality in mean," Working paper IRMAR-INSA ; arXiv :1007.1193v2
11. Patilea, V., and Raïssi, H. (2011), "Corrected portmanteau tests for VAR models with time-varying variance," Working paper IRMAR-INSA ; arXiv :1105.3638v2
12. Phillips, P.C.B., and Xu, K.L. (2005), "Inference in autoregression under heteroskedasticity," *Journal of Time Series Analysis* 27, 289-308.
13. Ramey, V.A., and Vine, D.J. (2006), "Declining volatility in the U.S. automobile industry," *The American Economic Review* 96, 1876-1889.
14. 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.
15. Stărică, C. (2003), "Is GARCH(1,1) as good a model as the Nobel prize accolades would imply ?" Working paper.
16. Stărică, C., and Granger C.W.J. (2005), "Nonstationarities in stock returns.," *Reviews of Economics and Statistics* 87, 503-522.
17. Xu, K.L., and Phillips, P.C.B. (2008), "Adaptive estimation of autoregressive models with time-varying variances," *Journal of Econometrics* 142, 265-280.

TABLE 1 – The estimators of the autoregressive parameters of the VAR(1) model for the balance data for the U.S. The standard deviations are into brackets.

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 2 – The values of the Ljung-Box portmanteau test statistics (standard, modified OLS-based and ALS-based) and the associated p-values for checking of the adequacy of the VAR(1) model for the U.S. trade balance data.

m	5		15	
	Test statistic	$p-$value	Test statistic	$p-$value
LB_m^S	6.84	0.000	106.34	0.000
LB_m^{OLS}	6.84	0.508	106.34	0.999
LB_m^{ALS}	25.73	0.064	66.83	0.159

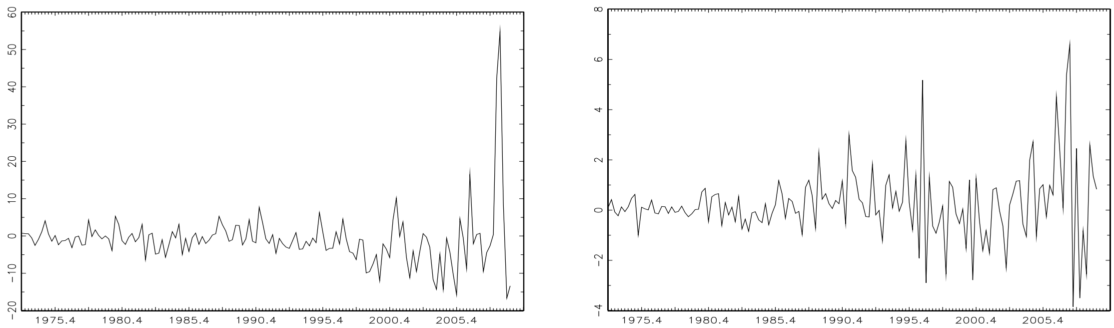

FIGURE 1 – The differences of the balance on merchandise trade for the U.S. on the left panel and the differences of the balance on services for the U.S. on the right panel in billions of dollars from 1/1/1970 to 10/1/2009, T=159.

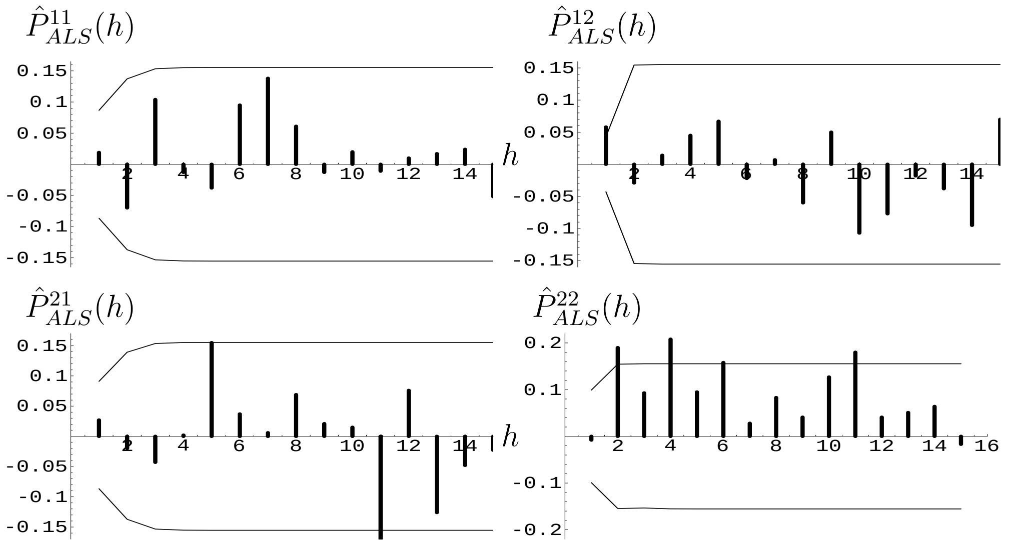

FIGURE 2 – The balance data for the U.S. : the ALS residual autocorrelations. The 95% confidence bounds are obtained using (11) and (21).

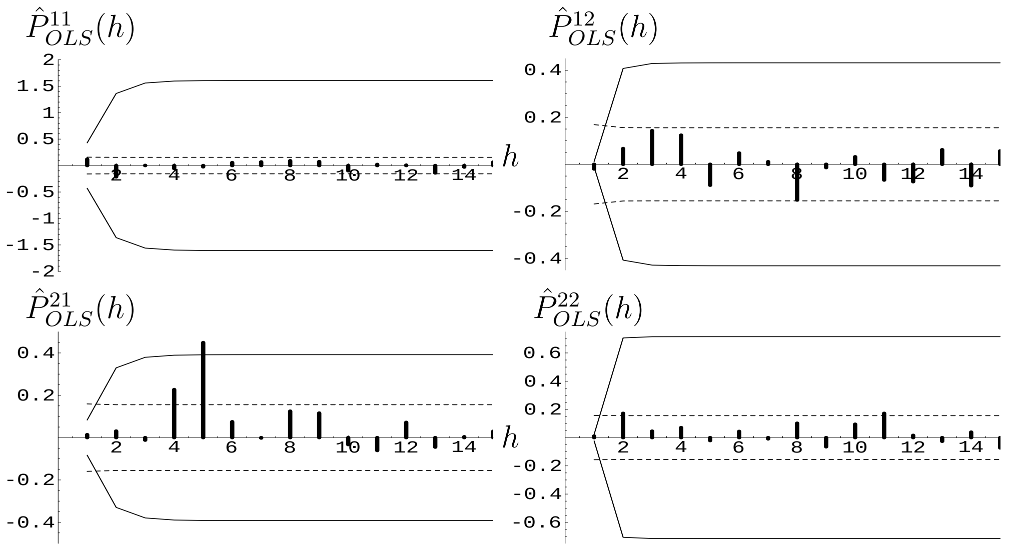

FIGURE 3 – The balance data for the U.S. : the OLS residual autocorrelations. The full lines 95% confidence bounds are obtained using (15). The dotted lines 95% confidence bounds are obtained using the standard result (13).