Goodness-of-fit tests for ARMA models with uncorrelated errors*

Christian Francq† Roch Roy‡
Jean-Michel Zakoïan§

CRM-2925

July 2003 (This version: April 2004)

*This work was supported by grants to the second named author from the Natural Science and Engineering Research Council of Canada, the Network of Centres of Excellence on The Mathematics of Information Technology and Complex Systems (MITACS) and the Fonds FQRNT (Government of Quebec).

†Université Lille III, GREMARS, BP 149, 59653 Villeneuve d'Ascq cedex, France; francq@univ-lille3.fr

‡Université de Montréal, Département de mathématiques et de statistique et Centre de recherches mathématiques, C.P. 6128, succ. Centre-ville, Montréal, QC H3C 3J7, Canada; royr@dms.umontreal.ca

§GREMARS and CREST, 3 Avenue Pierre Larousse, 92245 Malakoff Cedex, France; zakoian@ensae.fr

Abstract

We consider tests for lack of fit in ARMA models with non independent innovations. In this framework, the standard Box-Pierce and Ljung-Box portmanteau tests can perform poorly. Specifically, the usual text book formulas for asymptotic distributions are based on strong assumptions and should not be applied without careful consideration. In this paper, we derive the asymptotic covariance matrix $\Sigma_{\hat{\rho}_m}$ of a vector of autocorrelations for residuals of ARMA models under weak assumptions on the noise. The asymptotic distribution of the portmanteau statistics follows. A consistent estimator of $\Sigma_{\hat{\rho}_m}$, and a modification of the portmanteau tests are proposed. This allows to construct valid asymptotic significance limits for the residual autocorrelations, and (asymptotically) valid goodness-of-fit tests, when the underlying noise process is assumed to be non correlated rather than independent or a martingale difference. A set of Monte-Carlo experiments, and an application to the Standard & Poor's 500 returns, illustrate the practical relevance of our theoretical results.

Keywords: Residual Autocorrelations, Approximate Significance Limits, Portmanteau Tests, Weak ARMA Models, GARCH.

1 Introduction

Since the papers by Box and Pierce (1970) and Ljung and Box (1978), portmanteau tests have been popular diagnostic checking tools in the ARMA modelling of time series. Based on the residual empirical autocorrelations $\hat{\rho}(h)$, the Box-Pierce and Ljung-Box statistics (BP and LB hereafter) are defined by

$$Q_m = n\sum_{h=1}^{m}\hat{\rho}^2(h) \qquad \text{and} \qquad \tilde{Q}_m = n(n+2)\sum_{h=1}^{m}\frac{\hat{\rho}^2(h)}{n-h} \tag{1.1}$$

where n is the length of the series and m is a fixed integer.

The standard test procedure consists in rejecting the null hypothesis of an ARMA (p,q) model if $Q_m > \chi^2_{m-(p+q)}(1-\alpha)$ (or $\tilde{Q}_m > \chi^2_{m-(p+q)}(1-\alpha)$), where $m > p+q$ and $\chi^2_\ell(1-\alpha)$ denotes the $(1-\alpha)$-quantile of a χ^2 distribution with ℓ degrees of freedom. Box and Pierce (1970) noted that if the noise sequence is independent and identically distributed (iid), the level of the Q_m-test is approximately α when both m and n are large. The reader is referred to their paper, and to McLeod (1978), for a mathematical basis to this statement. The statistics Q_m and $\tilde{Q}_m$ have the same asymptotic distribution, but the LB statistic has the reputation of doing a better work for small or medium sample sizes (see Ljung-Box (1978) or Davies, Triggs and Newbold (1977)).

In the last decade, the time series literature has been characterized by a growing interest in nonlinear models. On the one hand these models provide useful characterizations of nonlinearities that are present in the mean or the variance of economic and financial time series. But on the other hand, nonlinear models appear difficult to handle. Many practitioners still use linear (ARMA) models, even when there is evidence of nonlinearities. Actually, this practice is not meaningless because many important classes of nonlinear processes admit ARMA representations. The representations are called weak because the innovations are dependent, though uncorrelated, and therefore constitute a weak white noise. However, some caution should be exercised in fitting ARMA models to nonlinear time series. The time series packages currently available for linear ARMA building (e.g. GAUSS, RATS, SAS, SPSS) rely on strong assumptions on the noise process such as independence or martingale difference that are typically not satisfied for linear representations of nonlinear processes. In recent years, a large part of the time-series and econometric literatures have been devoted to weaken the noise dependence assumptions; see for instance the papers by Francq and Zakoïan (1998) or Romano and Thombs (1996). In this paper we focus on the validation step of the standard time series procedures. This validation stage is not only based on portmanteau tests, but also on the examination of the autocorrelation function of the residuals.

It is a common practice to draw the sample autocorrelations of the observed time series and their 5%-significance limits, computed on the basis of asymptotic results obtained for an iid process. Romano and Thombs (1996) show that these significance limits can be quite misleading if the underlying process is uncorrelated rather than independent. They also show that the moving block bootstrap does a very good job in estimating the asymptotic variance of the lag 1 sample autocorrelation.

The above-mentioned significance limits are also used extensively as a diagnostic check on the residuals of a fitted ARMA model. In this article we show that these limits, as well as the BP and LB test procedures, are not (asymptotically) valid when the noise sequence is only uncorrelated and we propose valid procedures. To this aim we study the behaviour of the residuals autocorrelations, and of the portmanteau tests, in the framework of ARMA models with non independent error terms. By establishing the asymptotic distribution of a vector of m residual autocorrelations, we are able to provide the exact asymptotic distribution of the portmanteau statistics Q_m and $\tilde{Q}_m$ without requiring that the noise is independent or a martingale difference. The aim of this work is therefore to considerably widen the application area of these adequacy tests.

Several papers in the recent time-series literature consider goodness-of-fit tests. We briefly review the most significant contributions. Chen and Deo (2003) propose a generalized portmanteau test based on the discrete spectral average estimator, in the framework of linear processes (allowing for long memory). The asymptotic distribution is derived under an iid assumption on the noise process. Extending the work of

Durlauf (1991), Deo (2000) considers testing the martingale difference hypothesis in presence of conditional heteroskedasticity. Hong and Lee (2003) consider nonlinear time series models of a general form (including ARMA models) and propose a test based on a spectral density estimator. They derive the asymptotic distribution of their test under the assumption that the noise is a martingale difference. Lobato, Nankervis and Savin (2001, 2002) address the problem of testing the null hypothesis that a time series is uncorrelated up to some fixed order K and propose an extension of the Box-Pierce statistic. For the same problem, Lobato (2001) proposes an alternative test statistic that is asymptotically distribution-free under the null. None of these tests can be applied to the residuals of weak ARMA models because the underlying noise process is (i) neither iid, nor a martingale difference, (ii) not observable.

As noted by Romano and Thombs (1996), although the distinction between uncorrelated and iid noise may appear superficial, the implications in terms of statistical inference are broad. Many examples of ARMA processes in which the noise is not a martingale difference are given in Section 2. The rest of the paper is organized as follows. Section 3 introduces notations and provides the asymptotic distribution of a vector of sample autocorrelations of ARMA residuals. Then we obtain the limiting distribution of the portmanteau statistics. The noise is not required to be a martingale difference. Important particular cases are considered in Section 4, showing huge discrepancies between the true asymptotic distribution and the commonly used χ^2 approximation. These examples justify the need for adequate modifications of the standard tests. To this aim, the asymptotic covariance matrix of the residual autocorrelations vector is estimated in Section 5. The method consists in estimating the spectral density of a multivariate process by means of auxiliary autoregressive models. The performances of the corrected portmanteau tests are evaluated in Section 6 through Monte-Carlo experiments. We concentrate on the LB test since it is more used than the BP test. Section 7 is devoted to an application to the Standard & Poor's 500 returns. Section 8 concludes. Technical lemmas and proofs are collected in an appendix.

2 Examples of processes with weak ARMA representations

A second order stationary process $(X_t)_{t\in\mathbb{Z}}$ is said to satisfy an ARMA(p,q) representation if, for all $t \in \mathbb{Z}$,

$$X_t = \sum_{i=1}^{p} a_i X_{t-i} + \epsilon_t - \sum_{i=1}^{q} b_i \epsilon_{t-i} \tag{2.1}$$

where the ϵ_t are error terms, with zero mean and common variance $\sigma^2 > 0$, and where the polynomials $\phi(z) = 1 - a_1 z - \cdots - a_p z^p$ and $\psi(z) = 1 - b_1 z - \cdots - b_q z^q$ have all their zeros outside the unit disk and have no zero in common. It is said that (2.1) is a weak ARMA model when the ϵ_t are only supposed to be uncorrelated. We will say that the representation (2.1) is semi-strong when (ϵ_t) is a martingale difference, and that (2.1) is a strong ARMA representation when (ϵ_t) is an iid sequence.

These noise assumptions are of crucial importance for the interpretation of Model (2.1) in terms of predictions. The linear predictor of X_t given $X_{t-1}, X_{t-2}, \ldots$, is

$$P(X_t|X_{t-1},\ldots) := \sum_{i=1}^{p} a_i X_{t-i} - \sum_{i=1}^{q} b_i \frac{\phi(B)}{\psi(B)} X_{t-i},$$

where B denotes the backshift operator. Therefore, $\epsilon_t = X_t - P(X_t|X_{t-1},\ldots)$ is the linear innovation of X_t. By elementary properties of projection mappings, the linear innovation process of a stationary process is always a weak white noise (*i.e.* a stationary sequence of centered and uncorrelated random variables), but is not always a martingale difference. Illustrations will be provided below. In fact, in (2.1), (ϵ_t) is a martingale difference if and only if the best predictor of X_t is linear, that is $E(X_t|X_{t-1},\ldots) = P(X_t|X_{t-1},\ldots)$, where $E(X_t|X_{t-1},\ldots)$ denotes the conditional expectation of X_t given the σ-field generated by $\{X_u,\ u < t\}$. Since the assumption that the best predictor of X_t is a linear function of its past values is questionable, weak ARMA models seem more realistic (at least as approximations) than semi-strong ones.

Note that (2.1) is just a model, not a data generating process (DGP). Many DGP can be compatible with a weak ARMA representation. DGP admitting weak ARMA representations can be obtained as i) certain transformations of strong ARMA processes, ii) causal representations of non causal ARMA processes, iii) linear representations of non linear processes, or iv) approximations of the so-called Wold decomposition. In the following examples, (η_t) is a sequence of iid random variables with $E(\eta_t) = 0, E\eta_t^2 = 1$.

2.1 Transformations of strong ARMA processes

Consider a process (X_t) satisfying an ARMA model, and an aggregated process (Y_t) of the form $Y_t = c_1 X_{mt} + c_2 X_{mt+1} + \cdots + c_m X_{mt+m-1}$. There exist concrete situations in which only (Y_t) is observed. This is the case, for instance, when only low frequency data, $(X_{mt})_{1\leq t\leq n}$, from a high frequency strong ARMA are available. The aggregation properties of ARMA models are well-known (see e.g. Amemiya and Wu, 1972, Harvey and Pierse, 1984, Palm and Nijman, 1984). The aggregated process (Y_t) also satisfies an ARMA model. It is perhaps less well known that, in general, this ARMA representation is neither strong nor semi-strong, even when (X_t) satisfies a strong ARMA representation.

As a simple example, consider the strong ARMA(1,1) process

$$X_t - aX_{t-1} = \eta_t - b\eta_{t-1}, \quad a \neq b \in (-1, 1).$$

Consider the process $(Y_t) = (X_{2t})$. We have $Y_t - a^2 Y_{t-1} = u_t := \eta_{2t} + (a-b)\eta_{2t-1} - ab\eta_{2t-2}$. Since $Eu_t^2 = 1 + (a-b)^2 + a^2b^2$, $Eu_t u_{t-1} = -ab$, and $Eu_t u_{t-h} = 0$ for all $h > 1$, it is seen that (u_t) is a MA(1) of the form $u_t = \epsilon_t - \theta\epsilon_{t-1}$ where (ϵ_t) is a weak white noise with variance $E\epsilon_t^2 = Eu_t^2/(1+\theta^2)$ and $\theta \in (-1, 1)$ is such that $\theta/(1+\theta^2) = -Eu_t u_{t-1}/Eu_t^2$. We have $\epsilon_t = -ab\eta_{2t-2} + \theta u_{t-1} + \theta^2\eta_{2t-4} + R_t$ where $R_t = \eta_{2t} + (a-b)\eta_{2t-1} + \theta^2[(a-b)\eta_{2t-5} - ab\eta_{2t-6}] + \sum_{i\geq 3}\theta^i u_{t-i}$ is centered and independent of u_{t-1}, which is a function of $(\eta_{2t-2}, \eta_{2t-3}, \eta_{2t-4})$. Therefore, provided $\mu_3 := E\eta_t^3$ exists, $E\epsilon_t u_{t-1}^2 = -abE\eta_{2t-2}u_{t-1}^2 + \theta Eu_{t-1}^3 + \theta^2 E\eta_{2t-4}u_{t-1}^2 = \mu_3\left[-ab + \theta\left\{1 + (a-b)^3 - a^3b^3\right\} + a^2b^2\theta^2\right]$, which is generally not equal to 0 when $\mu_3 \neq 0$. In this case, since u_{t-1} belongs to the σ-field generated by $\{\epsilon_u,\ u < t\}$, we have $E\epsilon_t u_{t-1}^2 = E\left\{u_{t-1}^2 E(\epsilon_t|\epsilon_{t-1}, \dots)\right\} \neq 0$. This shows that (ϵ_t) is not a martingale difference. Hence, (Y_t) satisfies a weak ARMA(1,1) representation, which is generally not semi-strong.

Other weak ARMA representations are obtained by considering components of strong multivariate ARMA processes. More generally, processes of the form $Y_t = c'X_t$, where $c \in \mathbb{R}^d$ and (X_t) is a d-variate ARMA process, admit ARMA representations (see Lütkepohl, Chapter 6, 1991; Nsiri and Roy, 1993), and these representations are generally weak.

2.2 Causal representations of non causal ARMA processes

Let $X_t = \eta_t - \phi\eta_{t-1}$ where $|\phi| > 1$. The X_t process is a non causal MA(1). Now let $\epsilon_t = \sum_{i\geq 0}\phi^{-i}X_{t-i}$. The ϵ_t process is centered and uncorrelated, and we have $X_t = \epsilon_t - \phi^{-1}\epsilon_{t-1}$, which is the causal MA(1) representation of (X_t). Obviously $E(\epsilon_t X_{t-1}) = 0$ because ϵ_t is the linear innovation of X_t. However, straightforward computations show that $E(\epsilon_t X_{t-1}^2) = E\eta_t^3(1-\phi^2)(1+\phi^{-1})$ and $E(\epsilon_t X_{t-1}^3) = (E\eta_t^4 - 3)(1-\phi^2)^2/\phi$. Thus the ϵ_t process is not a martingale difference in general (for instance if $E\eta_t^3 \neq 0$ or $E\eta_t^4 \neq 3$).

2.3 Nonlinear processes

Nonlinear processes are often opposed to ARMA models, though these classes can be compatible. There exist numerous examples of nonlinear processes admitting ARMA representations: bilinear processes (see Pham, 1985), GARCH processes and their powers (Francq and Zakoïan, 2003), Hidden Markov models, Markov-switching ARMA and GARCH processes (Francq and Zakoïan, 2001), autoregressive conditional duration processes (Engle, 1998), logarithm of stochastic volatility processes (Ruiz, 1994). Other examples of ARMA models with innovations that are only uncorrelated are given by Romano and Thombs (1996). For a nonlinear process (X_t), the best predictor is generally not linear. In other words, $E(X_t|X_{t-1}, \dots) \neq$

$P(X_t|X_{t-1},\dots)$. If (X_t) admits an ARMA representation of the form (2.1), then the σ-fields generated by $\{X_u,\ u<t\}$ and $\{\epsilon_u,\ u<t\}$ coincide. Therefore, $E(\epsilon_t|\epsilon_{t-1},\dots) = E\left\{X_t - P(X_t|X_{t-1},\dots)|\epsilon_{t-1},\dots\right\} = E(X_t|X_{t-1},\dots) - P(X_t|X_{t-1},\dots) \neq 0$, which shows that, when it exists, the ARMA representation of a nonlinear process is only weak.

For the sake of concreteness, let us consider the simple bilinear model $X_t = \phi X_{t-1} + \eta_t + bX_{t-1}\eta_{t-2}$. If $\phi^2 + b^2 < 1$ it can be shown that this model has a unique nonanticipative (*i.e.* X_t is function of the $\eta_{t-i}, i \geq 0$) strictly stationary solution with a finite second-order moment. For this solution, with $u_t = \phi + b\eta_t$, we have the following expansion

$$X_t = \eta_t + \eta_{t-1}u_{t-2} + \sum_{k=2}^{\infty} u_{t-2}u_{t-3}\dots u_{t-k}\eta_{t-k}u_{t-k-1},$$

from which the second-order structure of (X_t) can be derived. Tedious calculations show that $E(X_t) = b\phi/(1-\phi)$, and that for $h>3$, $\gamma(h) = \phi\gamma(h-1)$, where $\gamma(h) = \mathrm{Cov}(X_t, X_{t-h})$. It follows that X_t admits an ARMA(1,3) representation of the form

$$X_t - \phi X_{t-1} = b\phi + \epsilon_t - \alpha_1\epsilon_{t-1} - \alpha_2\epsilon_{t-2} - \alpha_3\epsilon_{t-3}.$$

The values of the coefficients α_i can be obtained from the first four autocovariances of X_t. In particular, in the pure bilinear case when $\phi = 0$ it can be shown that $\alpha_1 = \alpha_2 = 0$. The coefficient α_3 is obtained by solving $\alpha_3/(1+\alpha_3^2) = b^2(1-b^2)/(1-b^4+b^4E\eta_t^4)$, $|\alpha_3| < 1$. It is clear that for this class of bilinear processes, the ARMA representations are only weak ones. This can be seen by comparing the optimal predictions derived from the DGP and those derived from the ARMA equations. For instance in the case $\phi = 0$, the optimal *linear* prediction of X_t given its past takes the form $\alpha_3 X_{t-3} - \alpha_3^2 X_{t-6} + \alpha_3^3 X_{t-9} + \cdots$ whereas the optimal prediction is $bX_{t-1}X_{t-2} - b^2X_{t-1}X_{t-3}X_{t-4} + b^3X_{t-1}X_{t-3}X_{t-5}X_{t-6} + \cdots$, provided this expansion exists.

2.4 Approximations of the Wold decomposition

Wold (1938) has shown that any purely non deterministic, second-order stationary process (X_t) admits an infinite MA representation of the form

$$X_t = \epsilon_t + \sum_{i=1}^{\infty} c_i\epsilon_{t-i}, \tag{2.2}$$

where (ϵ_t) is the linear innovation process of X and $\sum_i c_i^2 < \infty$. Defining the MA(q) process $X_t(q) = \epsilon_t + \sum_{i=1}^{q} c_i\epsilon_{t-i}$, it is straightforward that

$$\|X_t(q) - X_t\|_2^2 = E\epsilon_t^2 \sum_{i>q} c_i^2 \to 0, \quad \text{when } q \to \infty.$$

Therefore any purely non deterministic, second-order stationary process is the limit, in the mean-square sense, of weak finite-order ARMA processes.

The examples discussed above show that weak ARMA models can arise from various situations. Making strong or semi-strong assumptions on the noise process precludes most of these DGP, as well as many others. The linear model (2.2), which consists of the ARMA models and their limits, is arguably very general under the noise uncorrelatedness, but can be restrictive if stronger assumptions are made.

3 Limiting distributions

In this section we derive the limiting distribution of the residual autocorrelations and that of the port-manteau statistics, in the framework of weak ARMA. We will first recall asymptotic results concerning the estimation of weak ARMA models.

3.1 ARMA estimation under non-iid innovations

Let $(X_t)_{t\in\mathbb{Z}}$ be an ARMA(p,q) process, *i.e.* a second order stationary solution of Model (2.1). Without loss of generality, assume that a_p and b_q are not both equal to zero (by convention $a_0 = b_0 = 1$). When p and q are not both equal to 0, let $\theta_0 = (a_1, \ldots, a_p, b_1, \ldots, b_q)'$, $\theta = (\theta_1, \ldots, \theta_p, \theta_{p+1}, \ldots, \theta_{p+q})'$ and denote by Θ the parameter space, $\Theta = \{\theta \in \mathbb{R}^{p+q} : \ \phi_\theta(z) = 1 - \theta_1 z - \cdots - \theta_p z^p$ and $\psi_\theta(z) = 1 - \theta_{p+1} z - \cdots - \theta_{p+q} z^q$ have all their zeros outside the unit disk$\}$. For all $\theta \in \Theta$, let $\epsilon_t(\theta) = \psi_\theta^{-1}(B)\phi_\theta(B)X_t$. Note that $\epsilon_t = \epsilon_t(\theta_0)$. For simplicity, we will omit the notation θ in all quantities taken at the true value, θ_0. Given a realization $X_1, X_2, \ldots, X_n$, the variable $\epsilon_t(\theta)$ can be approximated, for $0 < t \leq n$, by $e_t(\theta)$ defined recursively by

$$e_t(\theta) = X_t + \sum_{i=1}^{p} \theta_i X_{t-i} - \sum_{i=1}^{q} \theta_{p+i} e_{t-i}(\theta)$$

where the unknown starting values are set to zero: $e_0(\theta) = e_{-1}(\theta) = \cdots = e_{-q+1}(\theta) = X_0 = X_{-1} = \cdots = X_{-p+1} = 0$.

Let Θ^* be a compact subset of Θ such that θ_0 is in the interior of Θ^*. The random variable $\hat{\theta}$ is called least squares estimator of θ if it satisfies, almost surely,

$$O_n(\hat{\theta}) = \min_{\theta\in\Theta^*} O_n(\theta) \quad \text{where} \quad O_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} e_t^2(\theta).$$

We denote by $(\alpha_X(k))_{k\in N^*}$ the sequence of strong mixing coefficients of any process $(X_t)_{t\in\mathbb{Z}}$. We consider the following assumption.

Assumption 1 *(X_t) is strictly stationary, satisfies the ARMA(p,q) model (2.1), $E|X_t|^{4+2\nu} < \infty$ and $\sum_{k=0}^{\infty}\{\alpha_X(k)\}^{\nu/(2+\nu)} < \infty$ for some $\nu > 0$.*

Francq and Zakoïan (1998) showed that under this assumption, $\hat{\theta}$ is strongly consistent and asymptotically normal:

$$\hat{\theta} \longrightarrow \theta_0 \quad a.s., \qquad \text{and} \qquad \sqrt{n}(\hat{\theta} - \theta_0) \overset{d}{\rightsquigarrow} \mathcal{N}(0, J^{-1}IJ^{-1}), \qquad \text{as } n \to \infty, \tag{3.1}$$

where $I = I(\theta_0)$ and $J = J(\theta_0)$ with

$$I(\theta) = \lim_{n\to\infty} \text{var}\left\{\sqrt{n}\frac{\partial}{\partial\theta}O_n(\theta)\right\} \qquad \text{and} \qquad J(\theta) = \lim_{n\to\infty} \frac{\partial^2}{\partial\theta\partial\theta'}O_n(\theta) \quad a.s.$$

Notice that Assumption 1 does not require independence of the noise, nor the fact that it is a martingale difference. The mixing condition is valid for large classes of processes (see Pham (1986), Carrasco and Chen (2002)) for which the (stronger) condition of β-mixing with exponential decay can be shown. However, this condition could be replaced by any weak dependence condition allowing to apply a central limit theorem. The moment condition is mild since $EX_t^4 < \infty$ is required for the existence of I and J.

Before turning to the residuals autocovariances, it will be useful to consider the joint asymptotic behaviour of the estimator $\hat{\theta}$ and sample autocovariances of the (non-observed) noise $\epsilon_t(\theta_0)$.

3.2 Joint distribution of $\hat{\theta}$ and the noise empirical autocovariances

Let, for $\ell \geq 0$,

$$\gamma(\ell) = \frac{1}{n}\sum_{t=1}^{n-\ell} \epsilon_t\epsilon_{t+\ell} \quad \text{and} \quad \rho(\ell) = \frac{\gamma(\ell)}{\gamma(0)}$$

denote the white noise "empirical" autocovariances and autocorrelations. Notice that these quantities are not statistics (unless if $p = q = 0$) since they depend on the unknown parameter θ_0. They are introduced as a device to facilitate forthcoming derivations.

For any fixed $m \geq 1$, let $\gamma_m = (\gamma(1), \ldots, \gamma(m))'$ and $\rho_m = (\rho(1), \ldots, \rho(m))'$ and let

$$\Gamma(\ell, \ell') = \sum_{h=-\infty}^{\infty} E(\epsilon_t \epsilon_{t+\ell} \epsilon_{t+h} \epsilon_{t+h+\ell'}),$$

for $(\ell, \ell') \neq (0, 0)$. Remark that $\Gamma(\ell, \ell') = \Gamma(\ell', \ell) = \Gamma(\ell, -\ell')$. The existence of $\Gamma(\ell, \ell')$ is justified in the appendix (see Lemma A.1).

A few additional notations are required. We denote by ϕ_i^* and ψ_i^* the coefficients defined by

$$\phi^{-1}(z) = \sum_{i=0}^{\infty} \phi_i^* z^i, \quad \psi^{-1}(z) = \sum_{i=0}^{\infty} \psi_i^* z^i, \qquad |z| \leq 1$$

for $i \geq 0$. Take $\phi_i^* = \psi_i^* = 0$ when $i < 0$. For $m, m' = 1, \ldots, \infty$, let $\Gamma_{m,m'} = (\Gamma(\ell, \ell'))_{1 \leq \ell \leq m, 1 \leq \ell' \leq m'}$. Let $\lambda_i = (-\phi_{i-1}^*, \ldots, -\phi_{i-p}^*, \psi_{i-1}^*, \ldots, \psi_{i-q}^*)' \in \mathbb{R}^{p+q}$ and let the $(p+q) \times m$ matrix

$$\Lambda_m = (\lambda_1 \quad \lambda_2 \quad \cdots \quad \lambda_m) = \begin{pmatrix} -1 & -\phi_1^* & \cdots & & & & -\phi_{m-1}^* \\ 0 & -1 & \ddots & & & & \vdots \\ \vdots & & \ddots & & & & \\ 0 & \cdots & & -1 & -\phi_1^* & \cdots & -\phi_{m-p}^* \\ 1 & \psi_1^* & \cdots & & \psi_p^* & \cdots & \psi_{m-1}^* \\ 0 & 1 & \ddots & & & & \vdots \\ \vdots & & \ddots & & & & \\ 0 & \cdots & & 1 & \psi_1^* & \cdots & \psi_{m-q}^* \end{pmatrix}. \tag{3.2}$$

It will be convenient to denote the matrix $\sum_{i=1}^{+\infty} \lambda_i \lambda_i'$ by $\Lambda_\infty \Lambda_\infty'$. This matrix is well defined because the components of the λ_i, defined through the series expansions of ϕ^{-1} and ψ^{-1}, decrease exponentially fast to zero as i goes to infinity. Similarly we define $\Lambda_\infty \Gamma_{\infty,\infty} \Lambda_\infty' = \sum_{\ell,\ell'=1}^{+\infty} \lambda_\ell \Gamma(\ell, \ell') \lambda_{\ell'}'$ and $\Gamma_{m,\infty} \Lambda_\infty' = \sum_{\ell'=1}^{+\infty} \sum_{\ell=1}^{m} \Gamma(\ell, \ell') \lambda_{\ell'}'$. We will show in the appendix (Lemma A.1) that $|\Gamma(\ell, \ell')| \leq K \max(\ell, \ell')$ for some constant K, which is sufficient to ensure the existence of these matrices.

Theorem 3.1 *Assume $p > 0$ or $q > 0$. Under Assumption 1, $\sqrt{n}(\hat{\theta} - \theta_0, \gamma_m)' \overset{d}{\rightsquigarrow} \mathcal{N}(0, \Sigma_{\hat{\theta},\gamma_m})$, where*

$$\Sigma_{\hat{\theta},\gamma_m} = \begin{pmatrix} \{\sigma^2 \Lambda_\infty \Lambda_\infty'\}^{-1} \Lambda_\infty \Gamma_{\infty,\infty} \Lambda_\infty' \{\sigma^2 \Lambda_\infty \Lambda_\infty'\}^{-1} & -\{\sigma^2 \Lambda_\infty \Lambda_\infty'\}^{-1} \Lambda_\infty \Gamma_{\infty,m} \\ -\Gamma_{m,\infty} \Lambda_\infty' \{\sigma^2 \Lambda_\infty \Lambda_\infty'\}^{-1} & \Gamma_{m,m} \end{pmatrix}.$$

Obviously, the top-left term of this covariance matrix is the matrix $J^{-1} I J^{-1}$ of (3.1). In this theorem it is explicitly given in terms of the AR and MA polynomials (through the matrix Λ_∞), and in terms of the noise variance and fourth-order structure (through the matrix $\Gamma_{\infty,\infty}$).

3.3 Limiting distribution of residual autocorrelations

We now turn to the residuals. Let $\hat{\epsilon}_t = e_t(\hat{\theta})$ when $p > 0$ or $q > 0$, and let $\hat{\epsilon}_t = \epsilon_t = X_t$ when $p = q = 0$. The residuals autocovariances and autocorrelations are defined by

$$\hat{\gamma}(\ell) = \frac{1}{n} \sum_{t=1}^{n-\ell} \hat{\epsilon}_t \hat{\epsilon}_{t+\ell} \quad \text{and} \quad \hat{\rho}(\ell) = \frac{\hat{\gamma}(\ell)}{\hat{\gamma}(0)}. \tag{3.3}$$

Let $\hat{\rho}_m = (\hat{\rho}(1), \ldots, \hat{\rho}(m))'$. Let $R(\ell, \ell') = \Gamma(\ell, \ell')/\sigma^4$ and $R_{i,j} = (R(\ell, \ell'))_{1 \leq \ell \leq i, 1 \leq \ell' \leq j}$ for $i, j = 1, \ldots, \infty$.

Theorem 3.2 *Under Assumption 1,* $\sqrt{n}\hat{\rho}_m \overset{d}{\rightsquigarrow} \mathcal{N}(0, \Sigma_{\hat{\rho}_m})$, *where*

$$\Sigma_{\hat{\rho}_m} = R_{m,m} \quad \text{when } p = q = 0,$$

and, when $p > 0$ *or* $q > 0$,

$$\begin{aligned}\Sigma_{\hat{\rho}_m} &= R_{m,m} + \Lambda'_m\{\Lambda_\infty\Lambda'_\infty\}^{-1}\Lambda_\infty R_{\infty,\infty}\Lambda'_\infty\{\Lambda_\infty\Lambda'_\infty\}^{-1}\Lambda_m \\ &\quad -\Lambda'_m\{\Lambda_\infty\Lambda'_\infty\}^{-1}\Lambda_\infty R_{\infty,m} - R_{m,\infty}\Lambda'_\infty\{\Lambda_\infty\Lambda'_\infty\}^{-1}\Lambda_m.\end{aligned}$$

To validate an ARMA(p, q) model, the most basic technique is to examine the autocorrelation function of the residuals. Theorem 3.2 can be used to obtain asymptotic significance limits for the residual autocorrelations.

Example 3.1 *Romano and Thombs (1996) derived the asymptotic distribution of a vector of empirical autocorrelations for processes satisfying Assumption 1. As an example they considered the weak white noise* $\epsilon_t = \eta_t\eta_{t-1}$, *where* η_t *is iid* $\mathcal{N}(0, 1)$. *Applying the first part of Theorem 3.2 to this weak white noise, we find that* $Var_{as}\{\sqrt{n}\hat{\rho}(1)\} = R_{1,1} = \Gamma_{1,1} = 3$, *which is the result obtained by Romano and Thombs. For a strong white noise, recall that* $Var_{as}\{\sqrt{n}\hat{\rho}(1)\} = 1$.

Example 3.2 *We now consider the MA(1) model* $X_t = \epsilon_t - b\epsilon_{t-1}$, *where* (ϵ_t) *is the weak white noise defined in Example 3.1. We have* $\lambda_i = \psi^*_{i-1} = b^{i-1}$ *and* $R(i + 1, i + 1) = 1$ *for all* $i \geq 1$, $R(1, 1) = 3$, *and* $R(i, j) = 0$ *when* $i \neq j$. *Thus,* $\Lambda_\infty\Lambda'_\infty = 1/(1-b^2)$, $\Lambda_\infty R_{\infty,\infty}\Lambda'_\infty = 3+b^2/(1-b^2)$, $\Lambda_\infty R_{\infty,1} = 3$. *Applying the second part of Theorem 3.2 to the weak MA(1) model, we find that* $Var_{as}\{\sqrt{n}\hat{\rho}(1)\} = b^2(1 + 2b^2)$. *For a strong MA(1) model* $X_t = \eta_t - b\eta_{t-1}$, *with* η_t *iid* $(0, 1)$, *we find that* $Var_{as}\{\sqrt{n}\hat{\rho}(1)\} = b^2$, *which is the result obtained by Box and Pierce (1970).*

The previous example shows that the significance limits of the residual autocorrelations can be very different for ARMA models with iid noise and ARMA models with a noise that is only uncorrelated (see also Figure 1 and 2 below). Of course, in practice $\Sigma_{\hat{\rho}_m}$ has to be estimated (see Section 5).

Remark 3.1 *It is clear from Theorem 3.2 that, for a given ARMA model, the asymptotic distribution of the residual autocorrelations depends on the noise distribution through the quantities* $\Gamma(\ell, \ell')$ *only. It is also worth noting that this asymptotic distribution depends on the ARMA parameters through the vectors* $\lambda_i\lambda'_j$ *only. So the asymptotic distribution of* $\sqrt{n}\hat{\rho}_m$ *will be the same for the AR(p) model:* $\phi(B)X_t = \epsilon_t$, *and for the MA(p) model:* $X_t = \phi(B)\epsilon_t$. *In the former case we have* $\lambda_i = (-\phi^*_{i-1}, \ldots, -\phi^*_{i-p})$ *and in the latter case* $\lambda_i = (\phi^*_{i-1}, \ldots, \phi^*_{i-p})$.

3.4 Limiting distribution of the portmanteau statistics

Portmanteau statistics, as defined in (1.1) and (3.3), are generally used to test the null hypothesis

$$H_0\text{: } (X_t) \text{ satisfies an ARMA}(p, q) \text{ representation } [5]$$

against the alternative

$$H_1\text{: } (X_t) \text{ does not admit an ARMA representation, or admits an ARMA } (p', q') \text{ representation with } p' > p \text{ or } q' > q.$$

[5] A more formal notation and formulation of the null hypothesis is as follows:

$H_0(p, q)$: there exist polynomials ϕ and ψ of orders p and q, with roots outside the unit disk and no common root, such that $\psi^{-1}(L)\phi(L)X_t$ is a centered and uncorrelated stationary sequence,

where L denotes the standard lag operator.

Under H_0, the residuals should resemble to a realization of a weak white noise. In other words, if the fitted ARMA(p,q) model is appropriate, the residuals autocorrelations should be close to zero, and the portmanteau statistics should not take too large values. Consequently, the adequacy of the ARMA model with (canonical) orders (p,q) is usually rejected when the portmanteau statistics take too large values. Of course, portmanteau statistics, as any statistic based on a finite number m of autocorrelations, are unable to detect all departures from H_0. Practitioners generally use several values of m (in the ARIMA procedure of SAS, by default the LB portmanteau tests are provided for $m = 6, 12, 18$ and 24).

It is not possible to derive the asymptotic distribution of Q_m and $\tilde{Q}_m$ under H_0 only. The *standard* limit distribution of those statistics, and hence the standard tests, is obtained under the stronger assumption that (ϵ_t) is iid. The main result of this paper is a direct consequence of Theorem 3.2. It gives the exact asymptotic distribution of the standard portmanteau statistics under weak dependence assumptions.

Theorem 3.3 *Let Assumption 1, in particular H_0, hold. Then the statistics Q_m and $\tilde{Q}_m$ defined in (1.1) converge in distribution, as $n \to \infty$, to*

$$\mathrm{Z}_m(\xi_m) := \sum_{i=1}^{m} \xi_{i,m} Z_i^2$$

where $\xi_m = (\xi_{1,m}, \ldots, \xi_{m,m})'$ is the eigenvalues vector of the matrix $\Sigma_{\hat{\rho}_m}$ defined in Theorem 3.2, and $Z_1, \ldots, Z_m$ are independent $\mathcal{N}(0,1)$ variables.

Consider the strong ARMA case, *i.e.* the case when (ϵ_t) in (2.1) is iid. In this case, for $\ell, \ell' > 0$, $\Gamma(\ell, \ell') = 0$ when $\ell \neq \ell'$ and $\Gamma(\ell, \ell) = \sigma^4$. Thus $R_{m,m} = I_m$, where I_m denotes the $m \times m$ identity matrix, and $\Lambda_\infty R_{\infty,m} = \Lambda_m$. We deduce that

$$\Sigma_{\hat{\rho}_m} \quad = \quad I_m - \Lambda_m' \{\Lambda_\infty \Lambda_\infty'\}^{-1} \Lambda_m, \tag{3.4}$$

which is the result obtained by McLeod (1978). When m is large, $\Sigma_{\hat{\rho}_m} \simeq I_m - \Lambda_m' \{\Lambda_m \Lambda_m'\}^{-1} \Lambda_m$ is close to a projection matrix with $m - (p+q)$ eigenvalues equal to 1, and $(p+q)$ eigenvalues equal to 0, and we retrieve the result given in Box and Pierce (1970). Therefore, in the strong ARMA case, the asymptotic distribution of the BP and LB statistics can be approximated by a $\chi^2_{m-(p+q)}$ distribution. In the next section, we will see that the approximation in no longer valid when the innovations (ϵ_t) are not iid.

Remark 3.2 *The distribution of the quadratic form $\mathrm{Z}_m(\xi_m)$ can be computed using the algorithm by Imhof (1961). An alternative approach would consist in modifying the standard statistics. This approach was followed by Lobato (2001) for testing that a dependent process is uncorrelated. It seems natural to consider the statistic $n\hat{\rho}_m' \hat{\Sigma}_{\hat{\rho}_m}^{-1} \hat{\rho}_m$ for some consistent estimate $\hat{\Sigma}_{\hat{\rho}_m}$ of $\Sigma_{\hat{\rho}_m}$. Indeed it is clear that Theorem 3.2 implies*

$$n\hat{\rho}_m' \hat{\Sigma}_{\hat{\rho}_m}^{-1} \hat{\rho}_m \overset{n\to\infty}{\rightsquigarrow} \chi^2_m$$

provided $\Sigma_{\hat{\rho}_m}$ is non singular. Unfortunately, when $p+q \neq 0$, the matrix $\Sigma_{\hat{\rho}_m}$ may be singular: for example when, an AR(1) model is fitted to a strong white noise (see Corollary 4.1 below with $a = 0$). In other cases $\Sigma_{\hat{\rho}_m}$ may be close to be singular: for strong ARMA(p,q) models we have already seen that, when m is large, $p+q$ eigenvalues of $\Sigma_{\hat{\rho}_m}$ are very close to 0. For these reasons we do not follow this approach.

Remark 3.3 *It can be shown that the previous results remain valid when the moment and mixing assumptions are made on the noise rather than on the observed process. Namely, Assumption* **1** *can be replaced by*

Assumption 1' $(X_t)_{t\in\mathbb{Z}}$ is strictly stationary, satisfies the ARMA(p,q) model (2.1), $E|\epsilon_t|^{4+2\nu} < \infty$ and $\sum_{k=0}^{\infty} \{\alpha_\epsilon(k)\}^{\nu/(2+\nu)} < \infty$ for some $\nu > 0$.

Note however that Assumptions **1** *and* **1'** *are not equivalent.*

4 Examples

In the first example, the results of the previous section are particularized in the AR(1) and MA(1) cases. The second example concerns an ARMA model with martingale difference innovations. The third example is an ARMA model with innovations that are only uncorrelated.

4.1 AR(1) or MA(1)

In the AR(1) and MA(1) cases Theorem 3.2 takes a simpler form. For instance in the AR(1) case we have, with $a = a_1$,

$$\lambda_i = -a^{i-1}, \quad \Lambda_i = -(1, a, \ldots, a^{i-1}), \quad \Lambda_\infty \Lambda_\infty' = \frac{1}{1-a^2}. \tag{4.1}$$

The next result follows.

Corollary 4.1 *Let $p = 1, q = 0$. Under Assumption 1, $\sqrt{n}\hat{\rho}_m \overset{d}{\rightsquigarrow} \mathcal{N}(0, \Sigma_{\hat{\rho}_m})$, where, for $i, j = 1, \ldots, m$*

$$\begin{aligned}
\Sigma_{\hat{\rho}_m}(i,j) &= \lim_{n\to\infty} n Cov\{\hat{\rho}(i), \hat{\rho}(j)\} \\
&= R(i,j) + (1-a^2)^2 a^{i+j-2} \sum_{\ell,\ell'=1}^{\infty} a^{\ell+\ell'-2} R(\ell,\ell') \\
&\quad -(1-a^2) \sum_{\ell=1}^{\infty} \left\{ a^{\ell+i-2} R(\ell,j) + a^{\ell+j-2} R(\ell,i) \right\}.
\end{aligned}$$

When $p = 0, q = 1$ the same result holds with a replaced by $b = b_1$.

4.2 ARMA models with GARCH errors

Assume that in the ARMA(p, q) equation (2.1), the innovation process (ϵ_t) is a GARCH(P, Q) process, solution of the model

$$\left\{ \begin{array}{ll} \epsilon_t & = \sigma_t \eta_t \\ \sigma_t^2 & = \omega + \displaystyle\sum_{i=1}^{Q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{P} \beta_j \sigma_{t-j}^2 \end{array} \right. \tag{4.2}$$

where (η_t) is a sequence of iid centered variables with unit variance, the α_i and β_j are nonnegative constants and ω is a positive constant. Under the assumption that $\sum_{i=1}^{Q} \alpha_i + \sum_{j=1}^{P} \beta_j < 1$, there exists a unique strictly stationary and nonanticipative solution process (ϵ_t) and this process has a finite variance. Under more restrictive conditions on the coefficients and if $E(\eta_t^4) < \infty$, then $E(\epsilon_t^4) < \infty$ (see e.g. Ling and McAleer 2002). If we further assume that the variables η_t have a symmetric distribution, then it can be shown that for $\ell, \ell' > 0$,

$$E(\epsilon_t \epsilon_{t+\ell} \epsilon_{t+h} \epsilon_{t+h+\ell'}) = \left\{ \begin{array}{ll} E(\epsilon_t^2 \epsilon_{t+\ell}^2) & \text{if } h = 0 \text{ and } \ell = \ell' \\ 0 & \text{otherwise.} \end{array} \right.$$

Hence $\Gamma(\ell, \ell') = 0$ when $\ell \neq \ell'$ and $\Gamma(\ell, \ell) = E(\epsilon_t^2 \epsilon_{t+\ell}^2)$.

To compute the autocovariance structure of (ϵ_t^2) we can use the fact that (ϵ_t^2) is solution of the ARMA$(P \wedge Q, P)$ model

$$\epsilon_t^2 = \omega + \sum_{i=1}^{P \wedge Q} (\alpha_i + \beta_i) \epsilon_{t-i}^2 + \nu_t - \sum_{j=1}^{P} \beta_j \nu_{t-j}^2 \tag{4.3}$$

where $P \wedge Q = \max(P, Q)$ and $\nu_t = \epsilon_t^2 - \sigma_t^2$ is the innovation of ϵ_t^2.

For example consider the case of an ARCH(1), *i.e.* $P = 0, Q = 1$. Let $\kappa = E(\eta_t^4)$ and assume $\kappa > 1$ and $0 < \kappa\alpha_1^2 < 1$. We have,

$$\begin{aligned} E(\epsilon_t^2) &= \sigma^2 = \frac{\omega}{1-\alpha_1} \\ \text{Var } (\epsilon_t^2) &= \gamma_{\epsilon^2}(0) = \omega^2 \left\{ \frac{(\alpha_1+1)\kappa}{(1-\kappa\alpha_1^2)(1-\alpha_1)} - \frac{1}{(1-\alpha_1)^2} \right\} = \sigma^4 \frac{\kappa-1}{1-\kappa\alpha_1^2} \\ \text{Cov } (\epsilon_t^2, \epsilon_{t+\ell}^2) &= \gamma_{\epsilon^2}(\ell) = \alpha_1^\ell \gamma_{\epsilon^2}(0), \quad \ell \geq 0 \\ \Gamma(\ell,\ell) &= \sigma^4 \left\{ 1 + \frac{(\kappa-1)\alpha_1^\ell}{1-\kappa\alpha_1^2} \right\}. \end{aligned}$$

It follows that the matrix $R_{m,m}$ of Theorem 3.2 is given by

$$R_{m,m} = I_m + \frac{(\kappa-1)}{1-\kappa\alpha_1^2} \text{ diag}(\alpha_1, \alpha_1^2, \dots \alpha_1^m) := I_m + \tilde{R}_{m,m}.$$

For simplicity, let us consider the case of an AR(1) process (X_t), endowed with a sequence of ARCH(1) errors (ϵ_t). Denote by $\Sigma^0_{\hat{\rho}_m}$ the asymptotic variance obtained in (3.4) for the strong AR(1), *i.e.* when $\alpha_1 = 0$. We have

$$\Sigma^0_{\hat{\rho}_m} = I_m - (1-a^2) \begin{pmatrix} 1 \\ a \\ \vdots \\ a^{m-1} \end{pmatrix} \begin{pmatrix} 1 & a & \dots & a^{m-1} \end{pmatrix}. \tag{4.4}$$

In view of (4.1) and the above expression for $R_{m,m}$

$$\Lambda_\infty \tilde{R}_{\infty,\infty} \Lambda'_\infty = \frac{\alpha_1(\kappa-1)}{(1-a^2\alpha_1)(1-\kappa\alpha_1^2)}, \quad \Lambda'_m \Lambda_\infty \tilde{R}_{\infty,m} = \left(\alpha_1^j a^{i+j-2} \frac{\kappa-1}{1-\kappa\alpha_1^2} \right),$$

from which we deduce

$$\Sigma_{\hat{\rho}_m} = \Sigma^0_{\hat{\rho}_m} + \frac{(\kappa-1)}{1-\kappa\alpha_1^2} \left(\alpha_1^i \mathbb{1}_{i=j} + (1-a^2)a^{i+j-2} \left\{ \frac{\alpha_1(1-a^2)}{1-\alpha_1 a^2} - \alpha_1^i - \alpha_1^j \right\} \right)_{i,j}. \tag{4.5}$$

The square roots of the diagonal terms of $\Sigma_{\hat{\rho}_m}$ correspond to the asymptotic standard deviations of the residual autocorrelations (multiplied by $\sqrt{n}$). Figure 1 displays these asymptotic standard deviations for AR(1) models with ARCH(1) innovations. These quantities can be very different from those of strong AR models (which are close to 1 when the lag is large enough).

For instance when $m = 2$, $\kappa = 3$ and $a = 0.5$ we have:

	$\Sigma_{\hat{\rho}_2}$	eigenvalues $\xi_{2,1}, \xi_{2,2}$	$Z_2(\xi_2)$
$\alpha_1 = 0$	$\begin{pmatrix} 0.25 & -0.375 \\ -0.375 & 0.8125 \end{pmatrix}$	(0.0625, 1)	$\chi_1^2 + 0.0625\chi_1^2$
$\alpha_1 = 0.5$	$\begin{pmatrix} 1.6786 & -0.9107 \\ -0.9107 & 2.9196 \end{pmatrix}$	(0.1260, 3.4008)	$0.13\chi_1^2 + 3.40\chi_1^2$

It is clear that for $\alpha_1 = 0.5$, the BP approximation by a χ_1^2 distribution will be disastrous. More precisely, with the LB portmanteau test, applied to residuals, the adequacy of the AR(1) model is rejected at level $\alpha = 5\%$ if

$$\tilde{Q}_m > \chi^2_{m-1}(0.95). \tag{4.6}$$

Table 1 gives the true asymptotic level of the test defined by the rejection region (4.6), for several values of m and a.[6] As expected, the true asymptotic size of the test is far from the 5% nominal level when the ARCH coefficient $\alpha_1 \neq 0$. We also note that the difference between the true level and the nominal level

Table 1: Exact asymptotic level (in %), corresponding to the LB 5% nominal level test (4.6), for the AR(1) model: $X_t = aX_{t-1} + \epsilon_t$, with ARCH(1) innovations $\epsilon_t = \sqrt{\omega + \alpha_1 \epsilon_{t-1}^2}\eta_t$, (η_t) iid $\mathcal{N}(0,1)$. The nominal level is obtained from the χ^2_{m-1} approximation, where m is the number of autocorrelations used in the statistics.

m	(a, α_1)				
	(0.0,0.0)	(0.0,0.2)	(0.5,0.2)	(0.9,0.2)	(0.9,0.4)
2	5.00	6.06	7.17	15.12	29.17
3	5.00	5.84	6.62	10.97	23.83
4	5.00	5.68	6.29	9.35	20.97
12	5.00	5.32	5.60	6.53	13.04
24	5.00	5.20	5.39	5.91	10.14
36	5.00	5.16	5.30	5.71	8.96

increases when $a \to 1$, *i.e.* when the DGP is close to a unit root model. The case $a = 0$ (a pure ARCH series) was studied by Diebold (1986).

4.3 An AR(1) model with weak innovations

Markov switching models have been found to be appropriate for series displaying sudden breaks, modelled through the change of state of a finite state space Markov chain (Δ_t). Following Hamilton (1989) and Hamilton and Susmel (1994) several authors have considered ARMA models in which the noise is subject to Markov-Switching. For instance, consider the AR(1) model with Markov-switching innovations

$$X_t = aX_{t-1} + \epsilon_t, \quad \epsilon_t = \begin{cases} \eta_t + c\eta_{t-1} & \text{if } \Delta_t = 0, \\ \eta_t - c\eta_{t-1} & \text{if } \Delta_t = 1, \end{cases} \tag{4.7}$$

where $|a| < 1$, $c \in \mathbb{R}$, (η_t) iid $\mathcal{N}(0,1)$, (Δ_t) is a stationary Markov chain independent of (η_t), with state space $\{0,1\}$ and transition probabilities $0 < \mathrm{p} := P(\Delta_t = 1|\Delta_{t-1} = 0) = P(\Delta_t = 0|\Delta_{t-1} = 1) < 1$. It is not difficult to see that (ϵ_t) is a white noise, of variance $\sigma^2 = 1 + c^2$, but its fourth-order cross-moments structure is more complicated than in the previous example. From Broze, Francq and Zakoïan (2002), we have

$$\begin{aligned} \Gamma(1,1) &= 1 + 4c^2 + c^4 + 4c^2(1-2\mathrm{p}) + \frac{c^2(1-2\mathrm{p})^2}{\mathrm{p}}, \\ \Gamma(\ell,\ell) &= (1+c^2)^2 + 2c^2(1-2\mathrm{p})^\ell, \quad \forall \ell \geq 2, \\ \Gamma(\ell,\ell+1) &= 0, \quad \forall \ell \geq 1, \\ \Gamma(\ell,\ell+2) &= c^2(1-2\mathrm{p})^{\ell+1}, \quad \forall \ell \geq 1, \\ \Gamma(\ell,\ell') &= 0, \quad \forall \ell \geq 1, \quad \ell' > \ell + 2. \end{aligned}$$

The numerical computation of $\Sigma_{\hat{\rho}_m}$ follows.

Table 2 displays the exact asymptotic level of the LB test defined by (4.6), for $c = 1$ and several values of a, p and m. This table shows that, for this weak AR(1) model, the standard χ^2_{m-1} approximation is not a valid approximation of the asymptotic distribution of $\tilde{Q}_m$, even when m is large or/and a is close to zero. The most dramatic differences between the true asymptotic level and the nominal 5% level are observed when p is close to 0. In this case, the noise is likely to stay in a same regime, the first MA(1) regime $\epsilon_t = \eta_t + \eta_{t-1}$ or the second MA(1) regime $\epsilon_t = \eta_t - \eta_{t-1}$, for a very long time, before switching to the other regime. Consequently, the AR(1) model (4.7) is likely to be confused with an ARMA(1,1)

[6]The level was obtained from Theorem 3.3, computing the eigenvalues of the matrix given by (4.5) and using the algorithm by Imhof (1961).

model, which could explain the bad performances of any adequacy test in this situation. Concerning the LB portmanteau test, the high probability of incorrectly rejecting the AR(1) model is explained by the fact that, when p is small, most of the eigenvalues of $\Sigma_{\hat{\rho}_m}$ are much larger than 1. Amazingly, the exact asymptotic level does not decrease with m, and is not even monotonous. Note also that for certain values of the parameters, the true asymptotic level can be less than the nominal 5% level. The Monte-Carlo experiments made in Section 6.1 below are in agreement with the values of this array.

Table 2: Exact asymptotic level (in %), corresponding to the LB 5% nominal level test (4.6), for the AR(1) model (4.7) with $c = 1$. The nominal level is obtained from the χ^2_{m-1} approximation, where m is the number of autocorrelations used in the statistics.

m	(a, p)				
	(0.0,0.95)	(0.6,0.95)	(0.6,0.05)	(0.9,0.5)	(0.99,0.95)
2	9.82	8.45	25.29	14.61	16.92
3	5.84	4.50	24.83	10.50	9.74
4	7.85	6.78	24.60	8.96	11.23
12	6.47	5.91	23.43	6.36	7.72
24	5.84	5.47	20.24	5.80	6.41
36	5.60	5.31	17.51	5.62	5.89

In ARMA modelling, the validation stage not only consists in using portmanteau tests. Another basic technique is to examine the residual autocorrelations. Figure 2 is the same as Figure 1, but for weak AR(1) models of the form (4.7). The right-hand graph shows an oscillatory behaviour of the significance bands which is impossible to obtain with a strong AR(1). It is also seen that the autocorrelations for odd h have a smaller asymptotic variance than the same autocorrelations for the strong AR(1) model.

5 Modifying the standard tests

To be able to use the statistics Q_m and $\tilde{Q}_m$ under more general assumptions on the noise process than those of independence or martingale difference, it is clear from the previous sections that the standard portmanteau tests deserve correction.

5.1 Modified portmanteau tests

First assume that a given consistent estimator $\hat{\Sigma}_{\hat{\rho}_m}$ of $\Sigma_{\hat{\rho}_m}$ is available. Denote by $\hat{\xi}_m = (\hat{\xi}_{m,1}, \ldots, \hat{\xi}_{m,m})'$ the eigenvalues vector of $\hat{\Sigma}_{\hat{\rho}_m}$. The next result, which is a straightforward consequence of Theorem 3.3, provides a critical region of asymptotic level $\alpha \in]0, 1[$.

Theorem 5.1 *Let Assumption 1, in particular H_0, hold. Then*

$$\lim_{n\to\infty} P\left\{\tilde{Q}_m > \mathrm{z}_m(1-\alpha)\right\} = \lim_{n\to\infty} P\left\{Q_m > \mathrm{z}_m(1-\alpha)\right\} = \alpha, \tag{5.1}$$

where $\mathrm{z}_m(1-\alpha)$ is such that: $P\left\{\mathrm{Z}_m(\hat{\xi}_m) > \mathrm{z}_m(1-\alpha)\right\} = \alpha$.

This test is not consistent against any form of the alternative H_1, just as standard portmanteau tests are not consistent against all violations of the strong ARMA(p, q) assumption. We illustrate the influence of m on the consistency of the portmanteau tests, by means of a simple example.

Example 5.1 *Let us consider the MA(1) DGP $X_t = \epsilon_t - b\epsilon_{t-1}$, where $b \neq 0$ and (ϵ_t) is an ergodic weak white noise with variance σ^2. Assume that the fitted model is an AR(1). The least squares estimator of the*

AR coefficient is the first-order empirical autocorrelation of X, namely $\hat{a} = \hat{\rho}_X(1)$. By the ergodic theorem, $\hat{a}$ converges a.s. to the first-order autocorrelation $\rho_X(1)$ of X, and it is easily seen that $e_t(\hat{a}) = X_t - \hat{a}X_{t-1}$ converges a.s. to

$$Z_t = X_t - \rho_X(1)X_{t-1} = X_t - \frac{b}{1+b^2}X_{t-1}.$$

Consequently, for $k > 0$, a.s.

$$\begin{aligned}\hat{\rho}(k) &\rightarrow Corr(Z_t, Z_{t-k}) = \left\{1 + \rho_X^2(1)\right\}\rho_X(k) - \rho_X(1)\left\{\rho_X(k-1) + \rho_X(k+1)\right\} \\ &= \begin{cases} 0 & when \quad k > 2, \\ -\rho_X^2(1) & when \quad k = 2, \\ \rho_X^3(1) & when \quad k = 1. \end{cases}\end{aligned}$$

Therefore for any $m \geq 1$, $\lim_{n\to\infty} \tilde{Q}_m = \lim_{n\to\infty} Q_m \geq \lim_{n\to\infty} n\rho_X^6(1) = \infty$, a.s., which proves that the tests are consistent:

$$\lim_{n\to\infty} P\left\{\tilde{Q}_m > \mathrm{z}_m(1-\alpha)\right\} = 1 \text{ and } \lim_{n\to\infty} P\left\{Q_m > \mathrm{z}_m(1-\alpha)\right\} = 1,$$

for any $m \geq 1$. If instead of a MA(1), the DGP is MA(12), similar computations show that the tests are consistent for $m \geq 12$ only.

This example might lead to think that large values of m are recommended in portmanteau tests. This is of course not the case since it is well known that choosing m too large would alter the power in finite samples. This point is not particular to our approach, and it also holds for the standard tests.

Evaluation of the critical value $\mathrm{z}_m(1-\alpha)$ can be done by using the Imhof algorithm. To implement the modified test, we have to estimate the eigenvalues of $\Sigma_{\hat{\rho}_m}$. There are two main approaches for estimating this matrix. A model-dependent method assumes a model for the noise (e.g. iid, GARCH, Markov-switching) such that $\Sigma_{\hat{\rho}_m}$ can be explicitly computed (as in (4.5) for instance). Then this model is fitted to the residuals, and a plug-in approach leads, under appropriate assumptions, to a consistent estimator $\Sigma_{\hat{\rho}_m}$. The second approach, which we will use, is a fully nonparametric one in which no specific time series model is assumed for the noise.

Nonparametric estimators of $\Gamma_{m,m} = \sigma^4 R_{m,m}$ rely on the fact that this matrix can be interpreted as the spectral density at frequency zero of a multivariate process (Υ_t). Kernel-based procedures, as proposed by e.g. Newey and West (1987), Andrews (1991), Andrews and Monahan (1992), use a weighted sum of autocovariances to estimate $\Gamma_{m,m}$, where the weights are determined by the kernel and the bandwidth parameter. Another estimator of $\Gamma_{m,m}$, which we will consider in the next section, is obtained by fitting to (Υ_t) an auxiliary AR model of large order, and by taking an empirical estimate of the spectral density of the auxiliary model.

5.2 Estimation of $\Gamma_{m,m}$

Note that $\Gamma_{m,m} = \sum_{h=-\infty}^{+\infty} \mathrm{Cov}\left(\Upsilon_t, \Upsilon_{t+h}\right)$, where

$$\Upsilon_t = \epsilon_t(\epsilon_{t+1}, \cdots, \epsilon_{t+m})'. \tag{5.2}$$

The stationary process (Υ_t) admits the Wold decomposition

$$\Upsilon_t = u_t + \sum_{i=1}^{\infty} B_i u_{t-i},$$

where (u_t) is a m-variate weak white noise. Assume that the covariance matrix $\Sigma_{u,u} := \mathrm{Var}(u_t)$ is non-singular, that $\sum_{i=1}^{\infty} \|B_i\| < \infty$, where $\|\cdot\|$ denotes any norm on the space of the real $m \times m$ matrices, and that $\det\left(I_m + \sum_{i=1}^{\infty} B_i z^i\right) \neq 0$ if $|z| \leq 1$. Then (Υ_t) admits an AR(∞) representation of the form

$$\mathcal{A}(B)\Upsilon_t := \Upsilon_t - \sum_{i=1}^{\infty} A_i \Upsilon_{t-i} = u_t, \tag{5.3}$$

such that $\sum_{i=1}^{\infty} \|A_i\| < \infty$ and $\det(\mathcal{A}(z)) \neq 0$ if $|z| \leq 1$. Interpreting $(2\pi)^{-1}\Gamma_{m,m}$ as the spectral density of (Υ_t) evaluated at the frequency zero (see Brockwell and Davis (1991) p. 459), we obtain

$$\Gamma_{m,m} = \mathcal{A}(1)^{-1}\Sigma_{u,u}\mathcal{A}'(1)^{-1}. \tag{5.4}$$

In the framework of univariate linear processes with independent innovations, Berk (1974) showed that the spectral density can be consistently estimated by fitting autoregressive models of order $r = r(n)$, whenever $r \to \infty$ and $r^3/n \to 0$ as $n \to \infty$. We will show that this result remains valid for the multivariate linear process (Υ_t), though its innovation (u_t) is not an independent process. Another difference with Berk (1974), is that the construction of the estimator requires replacing the non observable vector Υ_t by the statistic $\hat{\Upsilon}_t = \hat{\epsilon}_t(\hat{\epsilon}_{t+1}, \cdots, \hat{\epsilon}_{t+m})'$.

We make the following additional assumptions.

Assumption 2 *The innovation process (ϵ_t) of the ARMA(p, q) model (2.1) is such that the process (Υ_t) defined by (5.2) admits the AR(∞) representation (5.3), where $\|A_i\| = o\left(i^{-2}\right)$ as $i \to \infty$, the roots of $\det(\mathcal{A}(z)) = 0$ are outside the unit disk, and $\Sigma_{u,u} = Var(u_t)$ is non-singular. Moreover, $E|\epsilon_t|^{8+4\nu} < \infty$ and $\sum_{k=0}^{\infty}\{\alpha_\epsilon(k)\}^{\nu/(2+\nu)} < \infty$ for some $\nu > 0$.*

Note that the moment condition required by Berk (1974), for the univariate linear AR(∞) model (5.3) with (u_t) iid, was $E\Upsilon_t^4$. By comparison, our moment condition on (ϵ_t) entails that $E\|\Upsilon_t\|^{4+2\nu} < \infty$. If the mixing coefficients tend to zero at an exponential rate, which is the case for a wide class of processes, ν can be chosen arbitrarily small. Therefore our moment assumption is only slightly stronger than the one made by Berk.

Let $\hat{\mathcal{A}}_r(z) = I_m - \sum_{i=1}^{r} \hat{A}_{r,i}z^i$, where $\hat{A}_{r,1} \cdots \hat{A}_{r,r}$ are the coefficients of the multivariate regression of $\hat{\Upsilon}_t$ on $\hat{\Upsilon}_{t-1}, \ldots, \hat{\Upsilon}_{t-r}$ for $t = 1, \ldots, n-m$ (by convention $\hat{\Upsilon}_t = 0$ when $t \leq 0$). We denote by $\hat{\Sigma}_{\hat{u}_r,\hat{u}_r}$ the empirical estimator of the variance of the error term of this regression. The matrices $\hat{A}_{r,1}, \ldots, \hat{A}_{r,r}$ and $\hat{\Sigma}_{\hat{u}_r,\hat{u}_r}$ may be computed recursively using the algorithm given by Whittle (1963).

We are now in a position to state the main result of this section.

Theorem 5.2 *Under Assumptions 1 and 2,*

$$\hat{\Gamma}_{m,m} = \hat{\mathcal{A}}_r(1)^{-1}\hat{\Sigma}_{\hat{u}_r,\hat{u}_r}\hat{\mathcal{A}}_r'(1)^{-1} \to \Gamma_{m,m} = \mathcal{A}(1)^{-1}\Sigma_{u,u}\mathcal{A}'(1)^{-1}$$

in probability when $r = r(n) \to \infty$ and $r^3/n \to 0$ as $n \to \infty$.

The proof of Theorem 5.2 is given in the appendix.

5.3 Estimation of $\Sigma_{\hat{\rho}_m}$

In the asymptotic covariance matrix involved in Theorem 3.2, the infinite-dimensional matrices $R_{\infty,\infty}$ and $R_{\infty,m}$ have to be approximated by finite-dimensional ones. We therefore consider the matrix

$$\begin{aligned}\Sigma_{\hat{\rho}_m}^M &= R_{m,m} + \Lambda_m'\{\Lambda_\infty\Lambda_\infty'\}^{-1}\Lambda_M R_{M,M}\Lambda_M'\{\Lambda_\infty\Lambda_\infty'\}^{-1}\Lambda_m \\ &\quad -\Lambda_m'\{\Lambda_\infty\Lambda_\infty'\}^{-1}\Lambda_M R_{M,m} - R_{m,M}\Lambda_M'\{\Lambda_\infty\Lambda_\infty'\}^{-1}\Lambda_m\end{aligned} \tag{5.5}$$

for some $M \geq m$. Write

$$\Lambda_\infty = (\Lambda_M, \Lambda_{M+}), \quad R_{\infty,m} = \begin{pmatrix} R_{M,m} \\ R_{M+,m} \end{pmatrix}, \quad R_{\infty,\infty} = \begin{pmatrix} R_{M,M} & R_{M,M+} \\ R_{M+,M} & R_{M+,M+} \end{pmatrix}.$$

We have

$$\begin{aligned}\Sigma_{\hat{\rho}_m} - \Sigma_{\hat{\rho}_m}^M &= \Lambda_m'\{\Lambda_\infty\Lambda_\infty'\}^{-1}\{\Lambda_\infty R_{\infty,\infty}\Lambda_\infty' - \Lambda_M R_{M,M}\Lambda_M'\}\{\Lambda_\infty\Lambda_\infty'\}^{-1}\Lambda_m \\ &\quad -\Lambda_m'\{\Lambda_\infty\Lambda_\infty'\}^{-1}\{\Lambda_\infty R_{\infty,m} - \Lambda_M R_{M,m}\} \\ &\quad -\{R_{m,\infty}\Lambda_\infty' - R_{m,M}\Lambda_M'\}\{\Lambda_\infty\Lambda_\infty'\}^{-1}\Lambda_m\end{aligned} \tag{5.6}$$

For any multiplicative norm, the norm of the first term in the right-hand side is bounded by

$$\begin{aligned} & \|\Lambda'_m\{\Lambda_\infty\Lambda'_\infty\}^{-1}\|\|\Lambda_\infty R_{\infty,\infty}\Lambda'_\infty - \Lambda_M R_{M,M}\Lambda'_M\|\|\{\Lambda_\infty\Lambda'_\infty\}^{-1}\Lambda_m\| \\ \leq \; & K\left(\|\Lambda_{M+}R_{M+,M}\Lambda'_M\| + \|\Lambda_M R_{M,M+}\Lambda'_{M+}\| + \|\Lambda_{M+}R_{M+,M+}\Lambda'_{M+}\|\right). \end{aligned} \tag{5.7}$$

We have $\phi_i^* = O(\rho^i)$ and $\psi_i^* = O(\rho^i)$ for some $\rho \in [0,1[$. Consequently, any element of the j-th column of Λ_M is of order $O(\rho^j)$, and any element of the j-th column of Λ_{M+} is of order $O(\rho^{M+j})$. Arguing as in the proof of Lemmas A.1 and A.3, it can be shown that $\sup_{\ell,\ell'>0}|R(\ell,\ell')| < \infty$. It follows that any element of $R_{M+,M}\Lambda'_M$ is bounded uniformly in M, and that any element of $\Lambda_{M+}R_{M+,M}\Lambda'_M$ is bounded by $K\rho^M$, where K is independent of M. Thus the first norm in the right-hand side of (5.7) is of order $O(\rho^M)$. The same bound is obtained for the two other norms. The other terms in the right-hand side of (5.6) can be handled in the same way. Thus

$$\|\Sigma_{\hat{\rho}_m} - \Sigma^M_{\hat{\rho}_m}\| = O(\rho^M).$$

Now note that $\Lambda_m = \Lambda_m(\theta)$ and $\Lambda_\infty = \Lambda_\infty(\theta)$ are continuous functions of θ . Therefore $\hat{\Lambda}_m = \Lambda_m(\hat{\theta})$ and $\hat{\Lambda}_\infty = \Lambda_\infty(\hat{\theta})$ are consistent estimates of Λ_m and Λ_∞ defined in (3.2). The following algorithm was used in the numerical experiments presented below.

Step 1: Fit an ARMA(p,q) model (2.1) to the time series $X_1,\ldots,X_n$. Let $\hat{\theta}$ be the least squares estimate of the θ, let $\hat{\epsilon}_t,\ t=1,\ldots,n$ be the residuals and let $\hat{\sigma}^2 = n^{-1}\sum_{t=1}^n \hat{\epsilon}_t^2$ be an estimator of σ^2. Let $\hat{\Lambda}_k = \Lambda_k(\hat{\theta})$ for $k \in \{m, M, \infty\}$.

Step 2: Given a small number $\tau > 0$, define an integer $M \geq m$ such that $\|\Lambda_{M+}(\hat{\theta})\| \leq \tau$. If for instance $(p,q) = (1,0)$, then, by (4.1), $\|\Lambda_{M+}(\hat{\theta})\| = \|(\hat{a}_1^M, \hat{a}_1^{M+1}, \ldots)\| = \left\{\hat{a}_1^{2M}/(1-\hat{a}_1^2)\right\}^{1/2}$, and we can therefore choose $M = \max\{m, [\{\ln\tau^2(1-\hat{a}_1^2)\}/\ln\hat{a}_1^2] + 1\}$, where $[\cdot]$ denotes the integer part.

Step 3: For $r = 0,\ldots,r_{max}$, fit a M-variate AR(r) model to $\hat{\Upsilon}_1,\ldots,\hat{\Upsilon}_{n-M}$, where $\hat{\Upsilon}_t = (\hat{\epsilon}_t\hat{\epsilon}_{t+1},\ldots,\hat{\epsilon}_t\hat{\epsilon}_{t+M})'$. This can be done very rapidly using the multivariate version of the Durbin-Levinson algorithm (see Brockwell and Davis (1991), p. 422). An order $r_0 \in \{0,\ldots,r_{max}\}$ can be selected by minimizing an information criterion. In view of Theorem 5.2, we define

$$\hat{\Gamma}_{M,M} = \hat{\mathcal{A}}_{r_0}(1)^{-1}\hat{\Sigma}_{\hat{u}_{r_0},\hat{u}_{r_0}}\hat{\mathcal{A}}'_{r_0}(1)^{-1}$$

and we denote by $\hat{\Gamma}_{m,m}$ and $\hat{\Gamma}_{m,M}$ the corresponding sub-matrices of $\hat{\Gamma}_{M,M}$.

Step 4: An estimate $\hat{\Sigma}_{\hat{\rho}_m}$ of $\Sigma_{\hat{\rho}_m}$ is obtained by replacing, in (5.5), $R_{i,j}$ and Λ_k by $\hat{\sigma}^{-4}\hat{\Gamma}_{i,j}$ and $\hat{\Lambda}_k$, for $i \in \{m,M\}$, $j \in \{m,M\}$ and $k \in \{m,M,\infty\}$.

For the numerical illustrations presented in this paper, we used $\tau^2 = 0.1$ in Step 2 to select M. To save computation time, we also imposed $M \leq m + 20$. For the selection of r_0, in Step 3, we used the BIC criterion with $r_{max} = 5$. In the numerical experiments we made, the AIC criterion led to very similar results. The parameters τ^2 and r_{max} have influence on the time performance. Experiments made on a small number of replications, lead us to think that the results, presented in Section 6 below, can be slightly improved with a smaller τ^2 and a higher r_{max}.

In practice r_{max} should be chosen as large as possible, and τ^2 as small as possible, in function of the computer hardware performance.

6 Monte Carlo results

In this section, we investigate the finite-sample properties of the test introduced in this paper. For illustrative purpose, we also consider the standard LB portmanteau test, and the recent test proposed by Hong and Lee (2003) (HL hereafter). Other tests, already mentioned, could have been considered as well, but our goal is not to make a comparative study. Although the three tests considered were developed for checking the hypothesis of serial correlation of the noise in time series models, it is important to point out that their asymptotic properties were derived under very different hypotheses. The null hypotheses tested are, for our test, H_0 (as defined in Section 3.4), for the HL test,

$$H_0'\text{: } (X_t) \text{ satisfies a semi-strong ARMA}(p,q) \text{ model,}$$

and for the standard LB test,

$$H_0''\text{: } (X_t) \text{ satisfies a strong ARMA}(p,q) \text{ model.}$$

These hypotheses are the main assumptions under which the asymptotic levels of these tests are valid but, of course, other regularity conditions (such as our Assumption 1') are required. To our knowledge none of the tests of the literature are really designed for H_0.

Monte Carlo experiments have been conducted for the AR(1)-ARCH(1) and the Markov-switching models considered in Section 4.

6.1 Empirical size of the modified LB test

First we generate $N = 1,000$ replications of size $n = 1,000$ of the AR(1)-ARCH(1) model

$$X_t = aX_{t-1} + \epsilon_t, \quad \epsilon_t = \sqrt{1 + \alpha_1 \epsilon_{t-1}^2}\eta_t, \quad (\eta_t)\ iid\ \mathcal{N}(0,1),$$

for different values of the parameters a and α_1. Recall that this model is a semi-strong AR(1).

Concerning the LB test, we compute, for each replication, and for different values of m, a and α_1, the statistic $\tilde{Q}_m$. The model is rejected when $\tilde{Q}_m$ is greater than $\chi^2_{m-1}(0.95)$. This corresponds to a nominal level $\alpha = 5\%$ in the strong case. The rejection frequencies are presented in the first array of Table 3 and are similar to those of Table 1. With $N = 1,000$ replications, the standard error of the relative rejection frequencies is at most 1.58%. Therefore, the finite-sample sizes of the tests should be at most at $\pm 3.1\%$ from the relative rejection frequencies with a probability greater than 0.95. Comparing Table 1 with the first array of Table 3, the asymptotic behaviour of the LB portmanteau test and its behaviour for the sample size $n = 1,000$ do not appear significantly different. Most of the relative rejection frequencies are considerably greater than the theoretical 5%, and the conclusion drawn from the asymptotic study remains valid for this Monte Carlo experiment: the standard test should not be used to test H_0.

The HL test statistic is of the form $\hat{M}(m) = \left\{ n \sum_{h=1}^{n-1} k^2(h/m)\hat{\rho}^2(h) - \hat{C}(m) \right\} / \sqrt{\hat{D}(m)}$ where $\hat{C}(m)$ and $\hat{D}(m)$ are centering and scaling factors, $m = m_n = cn^{1/5}$ and $k(\cdot)$ is the Parzen kernel (other choices of the parameters are possible). Under some regularity conditions, $\hat{M}(m)$ is asymptotically $\mathcal{N}(0,1)$ distributed when the $\hat{\rho}(h)$ are the residual autocorrelations of a time series model with martingale difference errors. In our Monte-Carlo experiment, the test defined by the rejection region $\{\hat{M}(m) > 1.645\}$ should therefore be approximately of size 5%. For $N = 1,000$ independent replications of a 5%-level test, the standard error of the relative rejection frequencies is 0.69% and the latter should vary between 3.65% and 6.35% with an approximate 0.95 probability. The second array of Table 3 indicates that, in this example, the HL test is slightly undersized when $a = 0.9$ and considerably more undersized when $a = 0.0$ or 0.5. Indeed in the latter case, the only relative rejection frequency between the significance limits is when $(a, \alpha_1) = (0,0)$ and $m = 39$. This is in agreement with Chen and Deo (2003) who found that the finite sample distributions of statistics of the form $\hat{M}(m)$ are generally heavily skewed and they proposed level corrections.

Now consider the test of the present paper. We reject the model when $\tilde{Q}_m$ is greater than $\mathrm{z}_m(0.95)$ (see (5.1)). The relative rejection frequencies, presented in the third array of Table 3, are almost always within the 0.05 significant limits 3.65% and 6.35%. We conclude that, for this set of experiments, the modified LB portmanteau test satisfactorily controls the type I error for $n = 1,000$. For $n = 100$ (the results are not reported here) our test still accurately controls the type I error when $m \leq 12$, but the results are less satisfactory for $m = 24$ and $m = 36$.

The same experiments have been repeated for the Markov-switching model (4.7). Recall that this model is only a weak AR(1). The results are presented in Table 4. The first array reports the relative rejection frequencies of the standard LB test, which are similar to the asymptotic probabilities of type I error displayed in Table 2. The second array reports the relative rejection frequencies of the HL test. The most significant output is that, as for the standard LB tests, numerous relative rejection frequencies are

very far from 5%. Since the length of the simulated series is $n = 10,000$, it is very likely that this poor performance under H_0 is not due to a small-sample effect. This experiment lead us to think that tests designed to test H_0' should not be used to test H_0. The third array concerns the modified version of the LB test. For $(a, \mathrm{p}) = (0.6, 0.05)$ the test has a slight tendency to over-reject. It is probably due to the complex nature of this model when p is close to zero (see Section 4.3). However, most of the relative rejection frequencies are satisfactorily close to the theoretical 5%.

6.2 Empirical power

In this section we examine the power of the tests for the null hypothesis of an AR(1) against the ARMA(1,1)-ARCH(1) alternative

$$X_t = aX_{t-1} + \epsilon_t + b\epsilon_{t-1}, \quad \epsilon_t = \sqrt{1 + \alpha_1 \epsilon_{t-1}^2}\eta_t, \quad (\eta_t) \text{ iid } \mathcal{N}(0,1).$$

We generate $N = 1,000$ replications of size $n = 1,000$ of this model and fit an AR(1) for each replication. Since the standard LB test suffers from a serious size distortion, the critical values of the standard LB test have been adjusted so that the relative rejection frequencies be 5% under the null (*i.e.* for each value of a and α, we take as critical value the 95%-quantile of the $\tilde{Q}_m$ distribution over $1,000$ replications of the model with $b = 0$). Note however that in practice, when the data-generating process (DGP) is unknown, it is not possible to do the adjustment for the type I error. The results are reported in Table 5. It is seen that the tests have high power, except in the case $a = 0$. The lack of power in the latter case may be explained by the fact that the MA(1) process is very closed to an AR(1), as can be seen from its causal representation: $X_t = \epsilon_t + 0.2X_{t-1} - 0.04X_{t-2} + 0.008X_{t-3} + \cdots$ (The causal representation when $a = 0.5$ is $X_t = \epsilon_t + 0.7X_{t-1} - 0.14X_{t-2} + 0.028X_{t-3} + \cdots$, and when $a = 0.9$ it is $X_t = \epsilon_t + 1.1X_{t-1} - 0.49X_{t-2} + 0.368X_{t-3} + \cdots$). The powers of the three tests is in general decreasing when m increases. This is not surprising because, for this DGP, it is mainly in the first autocorrelations that the lack of AR(1) structure can be detected. The decrease of the power, as m increases, is moderate for the HL test which gives decreasing weights to higher lag empirical autocorrelations. Of course for other DGPs, the high values of m may be helpful to detect departure from the null hypothesis. The modified and (adjusted) standard versions of the LB test have very similar powers. This is not surprising because the tests are based on the same statistic and reject the null when this statistic is high. In the case $a = 0$, the HL test is slightly outperformed by the LB tests, especially for small values of m.

7 Standard & Poor's index

We now consider an application to the daily returns of the Standard & Poor's 500 index, from January 3, 1979 to December 31, 2001. The length of the series is $n = 5804$. First, we apply portmanteau tests for checking the hypothesis that the S&P returns constitute a white noise (the standard economic theory asserts that such stock indices should be martingale differences, though they are not generally independent sequences). Table 6 displays the p-values of the standard and modified LB tests. Since the p-values of the standard test are very small, the strong white noise hypothesis is rejected. The weak white noise hypothesis is not rejected, since for the modified test, the p-values are far from zero. This is in accordance with other works devoted to the analysis of stock-market returns (see Lobato, Nankervis and Savin (2001)).

Next, we fit an ARMA(1,1) model to the squares of the S&P returns. Denoting by (X_t) the mean-corrected series of the squared returns, we obtain the model $X_t - 0.832X_{t-1} = \epsilon_t - 0.727\epsilon_{t-1}$, $\mathrm{Var}\epsilon_t = 0.526\times$

10^{-6}. Table 7 displays the p-values of the standard and modified LB tests. The strong ARMA(1,1) model is rejected, but a weak ARMA(1,1) model is not rejected. Figure 3 displays the residual autocorrelations and their 5% significance limits under the strong ARMA(1,1) assumption (left graph) and under the assumption of weak ARMA(1,1) model (right graph). This figure confirms the conclusions drawn from Table 7.

Note that, in view of (4.3), the first and second-order structures we found for the S&P returns, namely a weak white noise for the returns and a weak ARMA(1,1) model for the squares of the returns, are compatible with a GARCH(1,1) model.

8 Conclusion

At the model criticism stage of the Box-Jenkins methodology, residual autocorrelations are plotted and portmanteau tests are performed in order to detect departures from the hypothesis that the disturbances of the specified ARMA(p, q) model are uncorrelated. However, the standard significance limits for the residual autocorrelations are not valid and the portmanteau tests do not well control the Type I error, when the actual underlying process is an ARMA(p, q) with uncorrelated but non independent disturbances. Thus, when the standard BP and LB portmanteau tests reject the specified ARMA(p_0, q_0) model and/or many residual autocorrelations are outside the non-significance band, the practitioner does not know whether (i) the ARMA(p_0, q_0) model is rejected because the disturbances are not uncorrelated, or (ii) the ARMA(p_0, q_0) model is rejected because the disturbances are not independent, though they are uncorrelated. The same problem occurs with all recent extensions of the LB test because, as we have seen, these extensions are designed for testing the adequacy of models with independent or martingale difference errors. The issue of distinguishing between weak and (semi-)strong ARMA models is important because, in the case (i) one should search another ARMA(p, q) model with $(p, q) \neq (p_0, q_0)$, in the case (ii) the actual underlying process displays non linearities but the orders of the weak ARMA(p_0, q_0) representation are well selected. In this paper we have partially solved the problem. The modifications proposed for the significance limits and the portmanteau tests lead to reject a given ARMA(p_0, q_0) model in the case (i) but not in the case (ii).

A computer program implementing the method is available on request.

Acknowledgements

The authors gratefully acknowledge the quick and careful reading of the manuscript by the Editor, the Associate Editor and two referees. Their detailed comments led to a greatly improved presentation.

Appendix: Proofs and technical results

Proof of Theorem 3.1. A standard expansion of the derivative of O_n about θ_0, taken at $\hat{\theta}$, yields

$$0 = \frac{\partial}{\partial\theta}O_n(\theta_0) + \left\{\frac{\partial^2}{\partial\theta\partial\theta'}O_n(\theta_n^*)\right\}(\hat{\theta}-\theta_0)$$

where θ_n^* is between θ_0 and $\hat{\theta}$. Thus, by standard arguments,

$$\hat{\theta}-\theta_0 = J^{-1}Y_n + O_p(1/n)$$

where $Y_n = -\frac{2}{n}\sum_{t=1}^n \epsilon_t \frac{\partial\epsilon_t}{\partial\theta}$. It is easily seen (McLeod, 1978) that the noise derivatives can be represented as

$$\frac{\partial\epsilon_t(\theta)}{\partial\theta} = (v_{t-1}(\theta),\ldots,v_{t-p}(\theta),u_{t-1}(\theta),\ldots,u_{t-q}(\theta))'$$

where

$$v_t(\theta) = -\phi_\theta^{-1}(B)\epsilon_t(\theta), \qquad u_t(\theta) = \psi_\theta^{-1}(B)\epsilon_t(\theta)$$

and

$$\epsilon_t(\theta) = \psi_\theta^{-1}(B)\phi_\theta(B)X_t.$$

Hence, at θ_0

$$\frac{\partial\epsilon_t}{\partial\theta} = \sum_{i\geq 1}\epsilon_{t-i}\lambda_i.$$

The asymptotic normality of $\sqrt{n}(J^{-1}Y_n,\gamma_m)'$ can be established along the same lines as in Francq and Zakoïan (1998). The detailed proof is omitted. It is easily shown that for $\ell,\ell'\geq 1$,

$$\begin{aligned}
\mathrm{Cov}(\sqrt{n}\gamma(\ell),\sqrt{n}\gamma(\ell')) &= \frac{1}{n}\sum_{t=1}^{n-\ell}\sum_{t'=1}^{n-\ell'}E(\epsilon_t\epsilon_{t+\ell}\epsilon_{t'}\epsilon_{t'+\ell'}) \\
&\to \quad \Gamma(\ell,\ell') \quad \text{as} \quad n\to\infty, \\
\mathrm{Cov}(\sqrt{n}J^{-1}Y_n,\sqrt{n}\gamma(\ell)) &= -J^{-1}\frac{2}{n}\sum_{t=1}^{n}\sum_{t'=1}^{n-\ell}E(\epsilon_t\frac{\partial\epsilon_t}{\partial\theta}\epsilon_{t'}\epsilon_{t'+\ell}) \\
&\to -2J^{-1}\sum_{h=-\infty}^{\infty}E(\epsilon_t\frac{\partial\epsilon_t}{\partial\theta}\epsilon_{t+h}\epsilon_{t+h+\ell}) \\
&= -2J^{-1}\sum_{h=-\infty}^{\infty}\sum_{i\geq 1}E(\epsilon_t\epsilon_{t-i}\epsilon_{t+h}\epsilon_{t+h+\ell})\lambda_i \\
&= -2J^{-1}\sum_{i\geq 1}\Gamma(i,\ell)\lambda_i \quad \text{as} \quad n\to\infty, \\
\mathrm{Var}_{as}(\sqrt{n}J^{-1}Y_n) &= J^{-1}IJ^{-1} \\
&= 4J^{-1}\sum_{i,j\geq 1}\Gamma(i,j)\lambda_i\lambda_j'J^{-1}, \\
J &= 2\sigma^2\sum_{i\geq 1}\lambda_i\lambda_i'
\end{aligned}$$

from which the asymptotic covariance matrix of Theorem 3.1 can be deduced.

□

Proof of Theorem 3.2. We have, for $\ell = 1, \ldots, m$,

$$\begin{aligned}
\hat{\gamma}(\ell) &= \gamma(\ell) + \frac{1}{n}\sum_{t=1}^{n-\ell}\left(\epsilon_{t+\ell}\frac{\partial \epsilon_t}{\partial \theta'} + \epsilon_t \frac{\partial \epsilon_{t+\ell}}{\partial \theta'}\right)_{\theta=\theta_n^*}(\hat{\theta}-\theta_0) + O_p(1/n) \\
&= \gamma(\ell) + E\left(\epsilon_t \frac{\partial \epsilon_{t+\ell}}{\partial \theta'}\right)_{\theta=\theta_0}(\hat{\theta}-\theta_0) + O_p(1/n) \\
&= \gamma(\ell) + \sigma^2 \lambda_\ell'(\hat{\theta}-\theta_0) + O_p(1/n)
\end{aligned}$$

where θ_n^* is between θ_0 and $\hat{\theta}$. Then $\hat{\gamma}_m := (\hat{\gamma}(1), \ldots, \hat{\gamma}(m))' = \gamma_m + \sigma^2 \Lambda_m'(\hat{\theta}-\theta_0) + O_p(1/n)$. Hence, by Theorem 3.1, the asymptotic distribution of $\sqrt{n}\hat{\gamma}_m$ is normal, with mean zero and covariance matrix

$$\begin{aligned}
\mathrm{Var}_{as}(\sqrt{n}\hat{\gamma}_m) &= \mathrm{Var}_{as}(\sqrt{n}\gamma_m) + \sigma^4 \Lambda_m' \mathrm{Var}_{as}(\sqrt{n}\hat{\theta})\Lambda_m \\
&\quad + \sigma^2 \Lambda_m' \mathrm{Cov}_{as}(\sqrt{n}\hat{\theta}, \sqrt{n}\gamma_m) + \sigma^2 \mathrm{Cov}_{as}(\sqrt{n}\gamma_m, \sqrt{n}\hat{\theta})\Lambda_m.
\end{aligned}$$

Finally, we have

$$n\left\{\frac{\hat{\gamma}(\ell)}{\hat{\gamma}(0)} - \frac{\hat{\gamma}(\ell)}{\sigma^2}\right\} = \sqrt{n}\hat{\gamma}(\ell)\frac{\sqrt{n}\{\sigma^2 - \hat{\gamma}(0)\}}{\sigma^2 \hat{\gamma}(0)}$$

so

$$\hat{\rho}_m = \hat{\gamma}_m/\sigma^2 + O_p(1/n).$$

Theorem 3.2 now follows from standard arguments.

□

The proof of Theorem 5.2 is based on a series of lemmas. We first justify the existence of the $\Gamma(\ell, \ell')$ in the following result.

Lemma A.1 *Under Assumption 1,*

$$\sum_{h=-\infty}^{\infty} |E(\epsilon_t \epsilon_{t+\ell}\epsilon_{t+h}\epsilon_{t+h+\ell'})| < \infty, \quad \textit{for } (\ell, \ell') \neq (0,0).$$

Proof. Note that, for all $h \in \mathbb{Z}$ and all $(\ell, \ell') \neq (0,0)$, $E(\epsilon_t \epsilon_{t+\ell}\epsilon_{t+h}\epsilon_{t+h+\ell'}) = \mathrm{Cov}\,(\epsilon_t \epsilon_{t+\ell}, \epsilon_{t+h}\epsilon_{t+h+\ell'})$. Note also that, the assumption made on $\psi_{\theta_0}(\cdot)$ entails

$$\epsilon_t = \sum_{i=0}^{+\infty} \pi_i X_{t-i} \quad \text{with } \pi_0 = 0 \text{ and } |\pi_i| \leq K\rho^i, \quad 0 < K < \infty, \quad \rho \in]0,1[.$$

Using the Davydov inequality (1968), we deduce that

$$\begin{aligned}
&\sum_{h=-\infty}^{\infty} |E(\epsilon_t \epsilon_{t+\ell} \epsilon_{t+h} \epsilon_{t+h+\ell'})| \\
\leq & \sum_{h=-\infty}^{\infty} \sum_{i_1,i_2,i_3,i_4=0}^{\infty} |\pi_{i_1}\pi_{i_2}\pi_{i_3}\pi_{i_4}| \, |\mathrm{Cov}\left(X_{t-i_1}X_{t+\ell-i_2}, X_{t+h-i_3}X_{t+h+\ell'-i_4}\right)| \\
\leq & K_0 K^4 \sum_{k=1}^{4} \sum_{(i_1,i_2,i_3,i_4)\in\mathcal{S}_k} \sum_{h=-\infty}^{\infty} \rho^{i_1+i_2+i_3+i_4} \left\|X_{t-i_1}X_{t+\ell-i_2}\right\|_{2+\nu} \\
& \times \left\|X_{t+h-i_3}X_{t+h+\ell'-i_4}\right\|_{2+\nu} \left\{\alpha\left(X_{t-i_1}X_{t+\ell-i_2}, X_{t+h-i_3}X_{t+h+\ell'-i_4}\right)\right\}^{\frac{\nu}{2+\nu}},
\end{aligned}$$

where $\|X\|_p = (EX^p)^{1/p}$ denotes the usual L^p-norm, $\alpha(X,Y)$ denotes the strong mixing coefficient between the σ-field generated by the random variable X and that generated by Y, K_0 is a universal constant, and

$$\begin{aligned}
\mathcal{S}_1 &= \left\{(i_1,i_2,i_3,i_4)\in\{0,1,\ldots\}^4 :\ i_1 \geq i_2-\ell,\ i_4 \leq i_3+\ell'\right\}, \\
\mathcal{S}_2 &= \left\{(i_1,i_2,i_3,i_4)\in\{0,1,\ldots\}^4 :\ i_1 \geq i_2-\ell,\ i_4 \geq i_3+\ell'\right\}, \\
\mathcal{S}_3 &= \left\{(i_1,i_2,i_3,i_4)\in\{0,1,\ldots\}^4 :\ i_1 \leq i_2-\ell,\ i_4 \leq i_3+\ell'\right\}, \\
\mathcal{S}_4 &= \left\{(i_1,i_2,i_3,i_4)\in\{0,1,\ldots\}^4 :\ i_1 \leq i_2-\ell,\ i_4 \geq i_3+\ell'\right\}.
\end{aligned}$$

For $(i_1,i_2,i_3,i_4)\in\mathcal{S}_1$, we have $t-i_1 \leq t+\ell-i_2, \quad t+h-i_3 \leq t+h+\ell'-i_4$, and

$$\begin{aligned}
\left|\alpha\left(X_{t-i_1}X_{t+\ell-i_2}, X_{t+h-i_3}X_{t+h+\ell'-i_4}\right)\right| &\leq \alpha_X(h-i_3-\ell+i_2), \quad \forall h \geq i_3+\ell-i_2, \\
\left|\alpha\left(X_{t-i_1}X_{t+\ell-i_2}, X_{t+h-i_3}X_{t+h+\ell'-i_4}\right)\right| &\leq \alpha_X(-i_1-h-\ell'+i_4), \quad \forall h \leq i_4-\ell'-i_1, \\
\left|\alpha\left(X_{t-i_1}X_{t+\ell-i_2}, X_{t+h-i_3}X_{t+h+\ell'-i_4}\right)\right| &\leq \alpha_X(0) \leq 1/4, \\
& \qquad \forall h = i_4-\ell'-i_1, \ldots, i_3+\ell-i_2.
\end{aligned}$$

Note also that, by the Hölder inequality,

$$\left\|X_{t-i_1}X_{t+\ell-i_2}\right\|_{2+\nu} \leq \|X_t\|_{4+2\nu}^2 := K_1 < \infty.$$

Therefore

$$\begin{aligned}
& \sum_{(i_1,i_2,i_3,i_4)\in\mathcal{S}_1} \sum_{h=-\infty}^{\infty} \rho^{i_1+i_2+i_3+i_4} \left\|X_{t-i_1}X_{t+\ell-i_2}\right\|_{2+\nu} \left\|X_{t+h-i_3}X_{t+h+\ell'-i_4}\right\|_{2+\nu} \\
& \times \left\{\alpha\left(X_{t-i_1}X_{t+\ell-i_2}, X_{t+h-i_3}X_{t+h+\ell'-i_4}\right)\right\}^{\frac{\nu}{2+\nu}} \\
\leq & K_1^2 \sum_{(i_1,i_2,i_3,i_4)\in\mathcal{S}_1} \rho^{i_1+i_2+i_3+i_4} \left\{i_3+\ell-i_2-i_4+\ell'+i_1+1+\sum_{k=0}^{\infty}\{\alpha_X(k)\}^{\nu/(2+\nu)}\right\} \\
< & \infty.
\end{aligned}$$

Continuing in this way, the conclusion follows.

□

We now turn to the estimation of $\Gamma_{m,m}$ given by (5.4). We need to introduce additional notations. Consider the regression of Υ_t on $\Upsilon_{t-1}, \ldots, \Upsilon_{t-r}$ defined by

$$\Upsilon_t = \sum_{i=1}^{r} A_{r,i}\Upsilon_{t-i} + u_{r,t}, \qquad u_{r,t} \perp \{\Upsilon_{t-1} \cdots \Upsilon_{t-r}\}. \tag{A.1}$$

If $\Upsilon_1, \ldots, \Upsilon_n$ were observed, the least squares estimators of $\underline{A}_r = (A_{r,1} \cdots A_{r,r})$ and $\Sigma_{u_r,u_r} = \mathrm{Var}(u_{r,t})$ would be given by

$$\breve{\underline{A}}_r = \hat{\Sigma}_{\Upsilon,\underline{\Upsilon}_r}\hat{\Sigma}^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r} \qquad \text{and} \qquad \hat{\Sigma}_{\breve{u}_r,\breve{u}_r} = \frac{1}{n}\sum_{t=1}^{n}\left(\Upsilon_t - \breve{\underline{A}}_r\underline{\Upsilon}_{r,t}\right)\left(\Upsilon_t - \breve{\underline{A}}_r\underline{\Upsilon}_{r,t}\right)'$$

where $\underline{\Upsilon}_{r,t} = (\Upsilon'_{t-1} \cdots \Upsilon'_{t-r})'$,

$$\hat{\Sigma}_{\Upsilon,\underline{\Upsilon}_r} = \frac{1}{n-r}\sum_{t=1}^{n}\Upsilon_t\underline{\Upsilon}'_{r,t}, \qquad \hat{\Sigma}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r} = \frac{1}{n-r}\sum_{t=1}^{n}\underline{\Upsilon}_{r,t}\underline{\Upsilon}'_{r,t},$$

with by convention $\Upsilon_t = 0$ when $t \leq 0$, and assuming $\hat{\Sigma}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}$ is non singular.[7]

Actually, we just observe $X_1, \ldots, X_n$. The residuals $\hat{\epsilon}_t = e_t(\hat{\theta})$ are then available for $t = 1, \ldots, n$, and the vectors $\hat{\Upsilon}_t$ are available for $t = 1, \ldots, n-m$. We therefore define the estimators

$$\hat{\underline{A}}_r = \hat{\Sigma}_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r}\hat{\Sigma}^{-1}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r} \qquad \text{and} \qquad \hat{\Sigma}_{\hat{u}_r,\hat{u}_r} = \frac{1}{n}\sum_{t=1}^{n-m}\left(\hat{\Upsilon}_t - \hat{\underline{A}}_r\underline{\hat{\Upsilon}}_{r,t}\right)\left(\hat{\Upsilon}_t - \hat{\underline{A}}_r\underline{\hat{\Upsilon}}_{r,t}\right)'$$

where $\underline{\hat{\Upsilon}}_{r,t} = (\hat{\Upsilon}'_{t-1} \cdots \hat{\Upsilon}'_{t-r})'$,

$$\hat{\Sigma}_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r} = \frac{1}{n-r}\sum_{t=1}^{n-m}\hat{\Upsilon}_t\underline{\hat{\Upsilon}}'_{r,t}, \qquad \hat{\Sigma}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r} = \frac{1}{n-r}\sum_{t=1}^{n-m}\underline{\hat{\Upsilon}}_{r,t}\underline{\hat{\Upsilon}}'_{r,t}.$$

Thus we have $\hat{\mathcal{A}}_r(z) = I_m - \sum_{i=1}^{r}\hat{A}_{r,i}z^i$, where $\hat{\underline{A}}_r = \left(\hat{A}_{r,1} \cdots \hat{A}_{r,r}\right)$.

In the sequel we employ the multiplicative matrix norm defined by: $\|A\| = \sup_{\|x\|\leq 1}\|Ax\| = \rho^{1/2}(A'A)$, where A is a $d_1 \times d_2$ matrix, $\|x\|$ is the Euclidean norm of the vector $x \in \mathbb{R}^{d_2}$, and $\rho(\cdot)$ denotes the spectral radius. This norm satisfies

$$\|A\|^2 \leq \sum_{i,j} a_{i,j}^2 \tag{A.2}$$

with obvious notations. This choice of the norm is crucial for the following lemma to hold (with e.g. the Euclidean norm, this result is not valid). Let

$$\Sigma_{\Upsilon,\underline{\Upsilon}_r} = E\Upsilon_t\underline{\Upsilon}'_{r,t}, \quad \Sigma_{\Upsilon,\Upsilon} = E\Upsilon_t\Upsilon'_t, \quad \Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r} = E\underline{\Upsilon}_{r,t}\underline{\Upsilon}'_{r,t}, \quad \hat{\Sigma}_{\hat{\Upsilon},\hat{\Upsilon}} = \frac{1}{n-r}\sum_{t=1}^{n-m}\hat{\Upsilon}_t\hat{\Upsilon}'_t.$$

Lemma A.2 *Under Assumption 2,*

$$\sup_{r\geq 1}\max\left\{\left\|\Sigma_{\Upsilon,\underline{\Upsilon}_r}\right\|, \left\|\Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\|, \left\|\Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\|\right\} \leq \infty.$$

[7] This condition means that there exists no $\lambda \neq 0$ such that $\lambda'\Upsilon_t = 0$ for $t = 1, \ldots, n$. In view of the non-singularity of $\Sigma_{u,u}$, this condition holds true, at least asymptotically.

Proof. We readily have

$$\|\Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}x\| \leq \|\Sigma_{\underline{\Upsilon}_{r+1},\underline{\Upsilon}_{r+1}}(x',0_m')'\| \quad \text{and} \quad \|\Sigma_{\Upsilon,\underline{\Upsilon}_r}x\| \leq \|\Sigma_{\underline{\Upsilon}_{r+1},\underline{\Upsilon}_{r+1}}(0_m',x')'\|$$

for any $x \in \mathbb{R}^{mr}$ and $0_m = (0,\ldots,0)' \in \mathbb{R}^m$. Therefore

$$0 < \|\mathrm{Var}\,(\Upsilon_t)\| = \left\|\Sigma_{\underline{\Upsilon}_1,\underline{\Upsilon}_1}\right\| \leq \left\|\Sigma_{\underline{\Upsilon}_2,\underline{\Upsilon}_2}\right\| \leq \cdots$$

and

$$\left\|\Sigma_{\Upsilon,\underline{\Upsilon}_r}\right\| \leq \left\|\Sigma_{\underline{\Upsilon}_{r+1},\underline{\Upsilon}_{r+1}}\right\|.$$

Let $f(\lambda)$ be the spectral density of Υ_t. Because the autocovariance function of Υ_t is absolutely summable (clearly, Lemma A.1 holds when Assumption 1 is replaced by Assumption 2), $\|f(\lambda)\|$ is bounded by a finite constant M, say. Denoting by $\delta = (\delta_1',\ldots,\delta_r')'$ an eigenvector of $\Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}$ associated with its largest eigenvalue, such that $\|\delta\| = 1$ and $\delta_i \in \mathbb{R}^m$ for $i = 1,\ldots,r$, we have

$$\begin{aligned}\left\|\Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\| &= \rho^{1/2}(\Sigma^2_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}) = \rho(\Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}) = \delta'\Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\delta \\ &= \sum_{j,k=1}^{r}\delta_j'\int_{-\pi}^{\pi}e^{i(k-j)\lambda}f(\lambda)d(\lambda)\delta_k \leq 2\pi M.\end{aligned}$$

By similar arguments, the smallest eigenvalue of $\Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}$ is greater than a positive constant independent of r. Using the fact that $\|\Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\|$ is equal to the inverse of the smallest eigenvalue of $\Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}$, the proof is completed.

□

Lemma A.3 *Let $\epsilon = (\epsilon_t)$ be a sequence of centered and uncorrelated variables, with $E|\epsilon_t|^{8+4\nu} < \infty$ and $\sum_{k=0}^{\infty}\{\alpha_\epsilon(k)\}^{\nu/(2+\nu)} < \infty$ for some $\nu > 0$. Let $\epsilon^{(2)}_{t,\ell} = \epsilon_t\epsilon_{t+\ell}$. Then, for $m_1, m_2 \geq 1$*

$$\sup_{s\in\mathbb{Z}}\sum_{h=-\infty}^{\infty}\left|Cov\left(\epsilon^{(2)}_{1,m_1}\epsilon^{(2)}_{1+s,m_2},\epsilon^{(2)}_{1+h,m_1}\epsilon^{(2)}_{1+h+s,m_2}\right)\right| < \infty.$$

Proof. Without loss of generality, we can take the supremum over the integers $s > 0$, and consider the sum for positive h. Let $m_0 = m_1 \wedge m_2$. We have, using again the Davydov (1968) inequality

$$\sum_{h=s+m_0}^{\infty}\left|\mathrm{Cov}\left(\epsilon^{(2)}_{1,m_1}\epsilon^{(2)}_{1+s,m_2},\epsilon^{(2)}_{1+h,m_1}\epsilon^{(2)}_{1+h+s,m_2}\right)\right| \leq K_0\sum_{h=s+m_0}^{\infty}\|\epsilon_t\|^8_{8+4\nu}\left\{\alpha_\epsilon(h-s-m_0)\right\}^{\frac{\nu}{2+\nu}}$$

which is bounded by a constant independent of s. To deal with the terms obtained for $h < s + m_0$, we write

$$\begin{aligned}\mathrm{Cov}\left(\epsilon^{(2)}_{1,m_1}\epsilon^{(2)}_{1+s,m_2},\epsilon^{(2)}_{1+h,m_1}\epsilon^{(2)}_{1+h+s,m_2}\right) &= \mathrm{Cov}\left(\epsilon^{(2)}_{1,m_1}\epsilon^{(2)}_{1+h,m_1},\epsilon^{(2)}_{1+s,m_2}\epsilon^{(2)}_{1+h+s,m_2}\right) \\ &+E\left\{\epsilon^{(2)}_{1,m_1}\epsilon^{(2)}_{1+h,m_1}\right\}E\left\{\epsilon^{(2)}_{1+s,m_2}\epsilon^{(2)}_{1+h+s,m_2}\right\} \\ &-E\left\{\epsilon^{(2)}_{1,m_1}\epsilon^{(2)}_{1+s,m_2}\right\}E\left\{\epsilon^{(2)}_{1+h,m_1}\epsilon^{(2)}_{1+h+s,m_2}\right\} \qquad (A.3)\end{aligned}$$

With the convention $\alpha_\epsilon(k) = 1/4$ for $k \leq 0$, we have

$$\begin{aligned}
&\sum_{h=0}^{s+m_0-1} \left|\mathrm{Cov}\left(\epsilon^{(2)}_{1,m_1}\epsilon^{(2)}_{1+h,m_1}, \epsilon^{(2)}_{1+s,m_2}\epsilon^{(2)}_{1+h+s,m_2}\right)\right| \\
\leq\ & K_0 \sum_{i=0}^{s+m_0-1} \|\epsilon_t\|^8_{8+4\nu} \left\{\alpha_\epsilon(i+1-m_0-m_1)\right\}^{\frac{\nu}{2+\nu}},
\end{aligned}$$

$$\begin{aligned}
\sum_{h=0}^{s+m_0-1} \left|E\left\{\epsilon^{(2)}_{1,m_1}\epsilon^{(2)}_{1+h,m_1}\right\}\right| &\leq K_0 \sum_{h=0}^{s+m_0-1} \|\epsilon_t\|^4_{4+2\nu} \left\{\alpha_\epsilon(h-m_1)\right\}^{\frac{\nu}{2+\nu}}, \\
\sum_{h=0}^{s+m_0-1} \left|E\left\{\epsilon^{(2)}_{1,m_1}\epsilon^{(2)}_{1+s,m_2}\right\}\right| &\leq (s+m_0)K_0\|\epsilon_t\|^4_{4+2\nu} \left\{\alpha_\epsilon(s-m_1)\right\}^{\frac{\nu}{2+\nu}}.
\end{aligned}$$

The right-hand sides of the first two inequalities are clearly bounded by constants independent of s. The same is true for the right-hand side of the third inequality because it can be shown that $\sup_{h\geq 1} h\left\{\alpha_\epsilon(h)\right\}^{\frac{\nu}{2+\nu}} < \infty$. Finally, the expectations in (A.3) are bounded, in absolute value, by $E\epsilon_t^4$. Hence the lemma is proved.

□

Lemma A.4 *Under Assumptions 1 and 2,* $\sqrt{r}\|\hat{\Sigma}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r} - \Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\|$, $\sqrt{r}\|\hat{\Sigma}_{\hat{\Upsilon},\hat{\Upsilon}} - \Sigma_{\Upsilon,\Upsilon}\|$, *and* $\sqrt{r}\|\hat{\Sigma}_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r} - \Sigma_{\Upsilon,\underline{\Upsilon}_r}\|$ *tend to zero in probability as* $n \to \infty$ *when* $r = o(n^{1/3})$.

Proof. For $1 \leq m_1, m_2 \leq m$ and $1 \leq r_1, r_2 \leq r$, the element of the $\{(r_1-1)m+m_1\}$-th row and $\{(r_2-1)m+m_2\}$-th column of $\hat{\Sigma}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}$ is of the form $\frac{1}{n-r}\sum_{t=r_0+1}^n \epsilon_t^{(4)}$, where $\epsilon_t^{(4)} = \epsilon_t^{(4)}(m_1,m_2,r_1,r_2) = \epsilon_{t-r_1}\epsilon_{t-r_1+m_1}\epsilon_{t-r_2}\epsilon_{t-r_2+m_2}$ and $r_0 = r_1 \wedge r_2$. By stationarity of $\left(\epsilon_t^{(4)}\right)_t$, we have

$$\begin{aligned}
&E\left\{\left(\frac{1}{n-r}\sum_{t=r_0+1}^n \epsilon_t^{(4)}\right) - E\epsilon_t^{(4)}\right\}^2 \\
=\ & E\left\{\frac{1}{n-r}\sum_{t=r_0+1}^n (\epsilon_t^{(4)} - E\epsilon_t^{(4)})\right\}^2 + \left(\frac{r-r_0}{n-r}\right)^2\left(E\epsilon_t^{(4)}\right)^2 \\
=\ & \frac{1}{(n-r)^2}\sum_{h=-(n-r_0-1)}^{n-r_0-1} (n-r_0)\,\mathrm{Cov}\left(\epsilon_t^{(4)}, \epsilon_{t-h}^{(4)}\right) + \left(\frac{r-r_0}{n-r}\right)^2\left(E\epsilon_t^{(4)}\right)^2 \\
\leq\ & \frac{n}{(n-r)^2}\sum_{h=-\infty}^{\infty}\left|\mathrm{Cov}(\epsilon_t^{(4)}, \epsilon_{t-h}^{(4)})\right| + \left(\frac{r-r_0}{n-r}\right)^2\left(E\epsilon_t^{(4)}\right)^2, \\
\leq\ & K\left(\frac{n}{(n-r)^2} + \frac{r^2}{(n-r)^2}\right),
\end{aligned} \tag{A.4}$$

for some constant K independent of r_1, r_2, m_1, m_2 and r, n. The last inequality holds because, by Lemma A.3, $\sum_{h=-\infty}^{\infty}\left|\mathrm{Cov}(\epsilon_t^{(4)}, \epsilon_{t-h}^{(4)})\right|$ is uniformly bounded in r_1, r_2. In view of (A.2) and (A.4) we have

$$\begin{aligned}
&E\left\{r\|\hat{\Sigma}_{\Upsilon,\Upsilon} - \Sigma_{\Upsilon,\Upsilon}\|^2\right\} \leq E\left\{r\|\hat{\Sigma}_{\Upsilon,\underline{\Upsilon}_r} - \Sigma_{\Upsilon,\underline{\Upsilon}_r}\|^2\right\} \\
&\leq E\left\{r\|\hat{\Sigma}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r} - \Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\|^2\right\} \leq Km^2r^3\left(\frac{n}{(n-r)^2} + \frac{r^2}{(n-r)^2}\right) = o(1)
\end{aligned}$$

as $n \to \infty$ when $r = r(n) = o(n^{1/3})$. Hence, when $r = o(n^{1/3})$

$$\begin{aligned}\sqrt{r}\|\hat{\Sigma}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r} - \Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\| &= o_P(1),\\ \sqrt{r}\|\hat{\Sigma}_{\Upsilon,\Upsilon} - \Sigma_{\Upsilon,\Upsilon}\| &= o_P(1), \quad \sqrt{r}\|\hat{\Sigma}_{\Upsilon,\underline{\Upsilon}_r} - \Sigma_{\Upsilon,\underline{\Upsilon}_r}\| = o_P(1). \end{aligned} \tag{A.5}$$

Let us introduce additional notations. Recall that $\hat{\epsilon}_t = e_t(\hat{\theta})$. We will first show that replacing $\hat{\epsilon}_t$ by $\tilde{\epsilon}_t = \epsilon_t(\hat{\theta})$ does not modify the asymptotic behaviour of the estimators. Then we will show that $\tilde{\epsilon}_t$ can be replaced $\epsilon_t = \epsilon_t(\theta_0)$. Let $\hat{\Sigma}_{\underline{\tilde{\Upsilon}}_r,\underline{\tilde{\Upsilon}}_r}$ and $\hat{\Sigma}_{\tilde{\Upsilon},\underline{\tilde{\Upsilon}}_r}$ be the matrices obtained by replacing $\hat{\Upsilon}_t$ by $\tilde{\Upsilon}_t = \tilde{\epsilon}_t(\tilde{\epsilon}_{t+1},\ldots,\tilde{\epsilon}_{t+m})'$ in $\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r}$ and $\hat{\Sigma}_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r}$. Similarly, we define $\hat{\epsilon}_t^{(4)}$ and $\tilde{\epsilon}_t^{(4)}$ obtained by replacing the variables ϵ by $e(\hat{\theta})$ and $\epsilon(\hat{\theta})$, respectively, in $\epsilon_t^{(4)}$. It can be easily shown that, almost surely, there exist constants $K > 0$ and $\rho \in]0,1[$ such that, $\sup_{\theta\in\Theta^*} |\epsilon_t(\theta) - e_t(\theta)| \leq K\rho^t$. Hence, we have for $t \geq r_0$,

$$\begin{aligned}&\left|\hat{\epsilon}_t^{(4)} - \tilde{\epsilon}_t^{(4)}\right|\\ \leq\ & K\rho^{t-r_1}\left|\hat{\epsilon}_{t-r_1+m_1}\hat{\epsilon}_{t-r_2}\hat{\epsilon}_{t-r_2+m_2}\right| + K\rho^{t-r_1+m_1}\left|\tilde{\epsilon}_{t-r_1}\hat{\epsilon}_{t-r_2}\hat{\epsilon}_{t-r_2+m_2}\right|\\ &+K\rho^{t-r_2}\left|\tilde{\epsilon}_{t-r_1}\tilde{\epsilon}_{t-r_1+m_1}\hat{\epsilon}_{t-r_2+m_2}\right| + K\rho^{t-r_2+m_2}\left|\tilde{\epsilon}_{t-r_1}\tilde{\epsilon}_{t-r_1+m_1}\tilde{\epsilon}_{t-r_2}\right|\\ \leq\ & K\rho^{t-r_0}\left\{|\tilde{\epsilon}_{t-r_1+m_1}| + K\rho^{t-r_0}\right\}\left\{|\tilde{\epsilon}_{t-r_2}| + K\rho^{t-r_0}\right\}\left\{|\tilde{\epsilon}_{t-r_2+m_2}| + K\rho^{t-r_0}\right\}\\ &+K\rho^{t-r_0}\left|\tilde{\epsilon}_{t-r_1}\right|\left\{|\tilde{\epsilon}_{t-r_2}| + K\rho^{t-r_0}\right\}\left\{|\tilde{\epsilon}_{t-r_2+m_2}| + K\rho^{t-r_0}\right\}\\ &+K\rho^{t-r_0}\left|\tilde{\epsilon}_{t-r_1}\tilde{\epsilon}_{t-r_1+m_1}\right|\left\{|\tilde{\epsilon}_{t-r_2+m_2}| + K\rho^{t-r_0}\right\} + K\rho^{t-r_0}\left|\tilde{\epsilon}_{t-r_1}\tilde{\epsilon}_{t-r_1+m_1}\tilde{\epsilon}_{t-r_2}\right|\\ \leq\ & K\rho^{t-r_0}\left\{K^3\rho^{3(t-r_0)} + \sum_{i=1}^{3} K^{(3-i)}\rho^{(3-i)(t-r_0)} \sum_{t_1,\ldots,t_i\in\mathcal{T}_i}\prod_{k=1}^{i}|\tilde{\epsilon}_{t_k}|\right\},\end{aligned}$$

where $\mathcal{T}_i = \mathcal{T}_i(t, r_1, r_2, m_1, m_2)$ denotes a set of indices $t_1, \ldots, t_i$ such that

$$t_k \in \{t - r_1, t - r_1 + m_1, t - r_2, t - r_2 + m_2\} \qquad \text{for } 1 \leq k \leq i.$$

By the Hölder and Lyapunov inequalities,

$$E\prod_{k=1}^{i}|\tilde{\epsilon}_{t_k}|^2 \leq E\sup_{\theta\in\Theta^*}|\epsilon_t(\theta)|^6 < \infty.$$

Consequently, for some finite constant K^* independent of t, r_1, r_2, m_1 and m_2, we have

$$\left\|\hat{\epsilon}_t^{(4)} - \tilde{\epsilon}_t^{(4)}\right\|_2 \leq K^*\rho^{t-r_0}$$

and, when $r = o(n)$,

$$\left\|\frac{1}{n-r}\sum_{t=r_0+1}^{n-m}\left(\hat{\epsilon}_t^{(4)} - \tilde{\epsilon}_t^{(4)}\right)\right\|_2 \leq \frac{1}{n-r}K^*\sum_{k=1}^{\infty}\rho^k = O\left(n^{-1}\right). \tag{A.6}$$

From (A.6) we deduce that, when $r = o(n^{2/3})$

$$\begin{aligned}\lim_{n\to\infty}\sqrt{r}\|\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r} - \hat{\Sigma}_{\underline{\tilde{\Upsilon}}_r,\underline{\tilde{\Upsilon}}_r}\| &= \lim_{n\to\infty}\sqrt{r}\|\hat{\Sigma}_{\hat{\Upsilon},\hat{\Upsilon}} - \hat{\Sigma}_{\tilde{\Upsilon},\tilde{\Upsilon}}\|\\ &= \lim_{n\to\infty}\sqrt{r}\|\hat{\Sigma}_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r} - \hat{\Sigma}_{\tilde{\Upsilon},\underline{\tilde{\Upsilon}}_r}\| = 0\end{aligned} \tag{A.7}$$

tend to zero as $n \to \infty$.

A Taylor expansion about θ_0 yields $\left|\epsilon_t(\hat{\theta}) - \epsilon_t\right| \leq \epsilon_t^* \left\|\hat{\theta} - \theta_0\right\|$, $\epsilon_t^* = \left\|\frac{\partial}{\partial\theta'}\epsilon_t(\theta^*)\right\|$ where $\theta^* = \theta^*(t,n)$ is between $\hat{\theta}$ and θ_0.

Thus, for $t \geq r_0$,

$$\begin{aligned}\left|\epsilon_t^{(4)} - \tilde{\epsilon}_t^{(4)}\right| \quad \leq \quad & \left\|\hat{\theta} - \theta_0\right\| \left\{\epsilon^*_{t-r_1} \left|\epsilon_{t-r_1+m_1}\epsilon_{t-r_2}\epsilon_{t-r_2+m_2}\right|\right. \\ & +\epsilon^*_{t-r_1+m_1} \left|\tilde{\epsilon}_{t-r_1}\epsilon_{t-r_2}\epsilon_{t-r_2+m_2}\right| \\ & \left.+\epsilon^*_{t-r_2} \left|\tilde{\epsilon}_{t-r_1}\tilde{\epsilon}_{t-r_1+m_1}\epsilon_{t-r_2+m_2}\right| + \epsilon^*_{t-r_2+m_2} \left|\tilde{\epsilon}_{t-r_1}\tilde{\epsilon}_{t-r_1+m_1}\tilde{\epsilon}_{t-r_2}\right|\right\}.\end{aligned}$$

Note that, in the previous inequality, the L^2-norm of the terms into brackets is bounded, uniformly in t, n, r_1, r_2, m_1 and m_2 because

$$E\left|\epsilon_t^*\right|^8 \leq E \sup_{\theta\in\Theta^*} \left\|\frac{\partial\epsilon_t}{\partial\theta'}(\theta)\right\|^8 < \infty, \qquad E\left|\epsilon_t\right|^8 < \infty, \qquad \text{and} \qquad E\left|\tilde{\epsilon}_t\right|^8 < \infty.$$

Using the Jensen inequality, it follows

$$\left\{\frac{1}{n-r}\sum_{t=r_0+1}^{n}\left(\epsilon_t^{(4)} - \tilde{\epsilon}_t^{(4)}\right)\right\}^2 \leq \left\|\hat{\theta} - \theta_0\right\|^2 D_{n,r_1,m_1,r_2,m_2},$$

where $E\left|D_{n,r_1,m_1,r_2,m_2}\right| \leq K^*$ for some constant K^* independent of n, r_1, r_2, m_1 and m_2. Thus $r\|\hat{\Sigma}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r} - \hat{\Sigma}_{\underline{\tilde{\Upsilon}}_r,\underline{\tilde{\Upsilon}}_r}\|^2$, $r\|\hat{\Sigma}_{\Upsilon,\Upsilon} - \hat{\Sigma}_{\tilde{\Upsilon},\tilde{\Upsilon}}\|^2$, and $r\|\hat{\Sigma}_{\Upsilon,\underline{\Upsilon}_r} - \hat{\Sigma}_{\tilde{\Upsilon},\underline{\tilde{\Upsilon}}_r}\|$ are respectively bounded by $r^3\left\|\hat{\theta} - \theta_0\right\|^2 O_P(1)$, $r\left\|\hat{\theta} - \theta_0\right\|^2 O_P(1)$, and $r^2\left\|\hat{\theta} - \theta_0\right\|^2 O_P(1)$. Since $\left\|\hat{\theta} - \theta_0\right\| = O_P\left(n^{-1/2}\right)$, we obtain for $r = o(n^{1/3})$

$$\begin{aligned}\sqrt{r}\|\hat{\Sigma}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r} - \hat{\Sigma}_{\underline{\tilde{\Upsilon}}_r,\underline{\tilde{\Upsilon}}_r}\| &= o_P(1), \\ \sqrt{r}\|\hat{\Sigma}_{\Upsilon,\Upsilon} - \hat{\Sigma}_{\tilde{\Upsilon},\tilde{\Upsilon}}\|\| &= o_P(1), \\ \sqrt{r}\|\hat{\Sigma}_{\Upsilon,\underline{\Upsilon}_r} - \hat{\Sigma}_{\tilde{\Upsilon},\underline{\tilde{\Upsilon}}_r}\| &= o_P(1).\end{aligned} \tag{A.8}$$

The proof of the lemma follows from (A.5), (A.7) and (A.8).

□

Write $\underline{A}_r^* = (A_1 \cdots A_r)$ where the A_i's are defined by (5.3).

Lemma A.5 *Under Assumption 2,*

$$\sqrt{r}\,\|\underline{A}_r^* - \underline{A}_r\| \to 0,$$

as $r \to \infty$.

Proof. Recall that by (5.3) and (A.1)

$$\begin{aligned}\Upsilon_t &= \underline{A}_r\underline{\Upsilon}_{r,t} + u_{r,t} \\ &= \underline{A}_r^*\underline{\Upsilon}_{r,t} + \sum_{i=r+1}^{\infty} A_i\Upsilon_{t-i} + u_t := \underline{A}_r^*\underline{\Upsilon}_{r,t} + u_{r,t}^*.\end{aligned}$$

Hence, using the orthogonality conditions in (5.3) and (A.1)

$$\underline{A}_r^* - \underline{A}_r \quad = \quad -\Sigma_{u_r^*,\underline{\Upsilon}_r}\Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r} \tag{A.9}$$

where $\Sigma_{u_r^*,\underline{\Upsilon}_r} = Eu_{r,t}^*\underline{\Upsilon}_{r,t}'$. By the Cauchy-Schwarz inequality and (A.2),

$$\left\|\text{Cov}\left(\Upsilon_{t-r-h},\underline{\Upsilon}_{r,t}\right)\right\| \leq r^{1/2}m\|\epsilon_t\|_4^4.$$

Thus,

$$\begin{aligned}\|\Sigma_{u_r^*,\underline{\Upsilon}_r}\| &= \|\sum_{i=r+1}^{\infty} A_i E\Upsilon_{t-i}\underline{\Upsilon}_{r,t}'\| \leq \sum_{h=1}^{\infty}\|A_{r+h}\|\left\|\text{Cov}\left(\Upsilon_{t-r-h},\underline{\Upsilon}_{r,t}\right)\right\| \\ &= O(1)r^{1/2}\sum_{h=1}^{\infty}\|A_{r+h}\|. \end{aligned} \tag{A.10}$$

Note that the assumption $\|A_i\| = o\left(i^{-2}\right)$ entails $r\sum_{h=1}^{\infty}\|A_{r+h}\| = o(1)$ as $r\to\infty$. The Lemma therefore follows from (A.9), (A.10) and Lemma A.2.

□

The following lemma is similar to Lemma 3 in Berk (1974).

Lemma A.6 *Under Assumptions 1 and 2,*

$$\sqrt{r}\|\hat{\Sigma}^{-1}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r} - \Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\| \quad = \quad o_P(1)$$

as $n\to\infty$ when $r = o(n^{1/3})$ and $r\to\infty$.

Proof. We have

$$\begin{aligned}&\left\|\hat{\Sigma}^{-1}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r} - \Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\| \\ =\ &\left\|\left\{\hat{\Sigma}^{-1}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r} - \Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r} + \Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\}\left\{\Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r} - \hat{\Sigma}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r}\right\}\Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\| \\ \leq\ &\left\|\Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\|\sum_{i=1}^{\infty}\left\|\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r} - \Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\|^i\left\|\Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\|^i.\end{aligned}$$

Thus, for every $\varepsilon > 0$,

$$\begin{aligned}&P\left(\sqrt{r}\left\|\hat{\Sigma}^{-1}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r} - \Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\| > \varepsilon\right) \\ \leq\ &P\left(\sqrt{r}\frac{\left\|\Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\|^2\left\|\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r} - \Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\|}{1 - \left\|\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r} - \Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\|\left\|\Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\|} > \varepsilon \text{ and } \left\|\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r} - \Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\|\left\|\Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\| < 1\right) \\ &+P\left(\sqrt{r}\left\|\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r} - \Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\|\left\|\Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\| \geq 1\right) \\ \leq\ &P\left(\sqrt{r}\left\|\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r} - \Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\| > \frac{\varepsilon}{\left\|\Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\|^2 + \varepsilon r^{-1/2}\left\|\Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\|}\right) \\ &+P\left(\sqrt{r}\left\|\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r} - \Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\| \geq \left\|\Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\|^{-1}\right) = o(1)\end{aligned}$$

by Lemmas A.4 and A.2. This establishes Lemma A.6.

□

Lemma A.7 *Under Assumptions 1 and 2,*

$$\sqrt{r}\left\|\underline{\hat{A}}_r - \underline{A}_r\right\| = o_P(1)$$

as $r \to \infty$ *and* $r = o(n^{1/3})$.

Proof. By the triangle inequality and Lemmas A.2 and A.6, we have

$$\left\|\hat{\Sigma}^{-1}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r}\right\| \leq \left\|\hat{\Sigma}^{-1}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r} - \Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\| + \left\|\Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\| = O_p(1) \tag{A.11}$$

when $r = o(n^{1/3})$. Note that the orthogonality conditions in (A.1) entail that $\underline{A}_r = \Sigma_{\Upsilon,\underline{\Upsilon}_r}\Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}$. This, Lemmas A.2, A.4, A.6, and (A.11) give

$$\begin{aligned}
\sqrt{r}\left\|\underline{\hat{A}}_r - \underline{A}_r\right\| &= \sqrt{r}\left\|\hat{\Sigma}_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r}\hat{\Sigma}^{-1}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r} - \Sigma_{\Upsilon,\underline{\Upsilon}_r}\Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\| \\
&= \sqrt{r}\left\|\left(\hat{\Sigma}_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r} - \Sigma_{\Upsilon,\underline{\Upsilon}_r}\right)\hat{\Sigma}^{-1}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r} + \Sigma_{\Upsilon,\underline{\Upsilon}_r}\left(\hat{\Sigma}^{-1}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r} - \Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right)\right\| = o_P(1).
\end{aligned}$$

□

Proof of Theorem 5.2. It suffices to show that $\hat{\mathcal{A}}_r(1) \to \mathcal{A}(1)$ and $\hat{\Sigma}_{\hat{u}_r,\hat{u}_r} \to \Sigma_{u,u}$ in probability. Let the $rm \times m$ matrix $\mathrm{E}_r = (1,\ldots,1)'$. Using (A.2), and Lemmas A.5, A.7, we obtain

$$\begin{aligned}
\left\|\hat{\mathcal{A}}_r(1) - \mathcal{A}(1)\right\| &\leq \left\|\sum_{i=1}^{r}\hat{A}_{r,i} - A_{r,i}\right\| + \left\|\sum_{i=1}^{r}A_{r,i} - A_i\right\| + \left\|\sum_{i=r+1}^{\infty}A_i\right\| \\
&= \left\|\left(\underline{\hat{A}}_r - \underline{A}_r\right)\mathrm{E}_r'\right\| + \left\|\left(\underline{A}^*_r - \underline{A}_r\right)\mathrm{E}_r'\right\| + \left\|\sum_{i=r+1}^{\infty}A_i\right\| \\
&\leq \sqrt{m}\sqrt{r}\left\{\left\|\underline{\hat{A}}_r - \underline{A}_r\right\| + \left\|\underline{A}^*_r - \underline{A}_r\right\|\right\} + \left\|\sum_{i=r+1}^{\infty}A_i\right\| = o_P(1).
\end{aligned}$$

Now note that

$$\hat{\Sigma}_{\hat{u}_r,\hat{u}_r} = \hat{\Sigma}_{\hat{\Upsilon},\hat{\Upsilon}} - \underline{\hat{A}}_r\hat{\Sigma}'_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r}$$

and, by (5.3)

$$\begin{aligned}
\Sigma_{u,u} &= Eu_tu_t' = Eu_t\Upsilon_t' = E\left\{\left(\Upsilon_t - \sum_{i=1}^{\infty}A_i\Upsilon_{t-i}\right)\Upsilon_t'\right\} \\
&= \Sigma_{\Upsilon,\Upsilon} - \sum_{i=1}^{\infty}A_iE\Upsilon_{t-i}\Upsilon_t' = \Sigma_{\Upsilon,\Upsilon} - \underline{A}^*_r\Sigma'_{\Upsilon,\underline{\Upsilon}_r} - \sum_{i=r+1}^{\infty}A_iE\Upsilon_{t-i}\Upsilon_t'.
\end{aligned}$$

Thus,

$$\begin{aligned}\left\|\hat{\Sigma}_{\hat{u}_r,\hat{u}_r}-\Sigma_{u,u}\right\| &= \Big\|\hat{\Sigma}_{\hat{\Upsilon},\hat{\Upsilon}}-\Sigma_{\Upsilon,\Upsilon}-\left(\underline{\hat{A}}_r-\underline{A}_r^*\right)\hat{\Sigma}'_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r} \\ &\quad -\underline{A}_r^*\left(\hat{\Sigma}'_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r}-\Sigma'_{\Upsilon,\underline{\Upsilon}_r}\right)+\sum_{i=r+1}^{\infty}A_iE\Upsilon_{t-i}\Upsilon_t'\Big\| \\ &\leq \left\|\hat{\Sigma}_{\hat{\Upsilon},\hat{\Upsilon}}-\Sigma_{\Upsilon,\Upsilon}\right\|+\left\|\left(\underline{\hat{A}}_r-\underline{A}_r^*\right)\left(\hat{\Sigma}'_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r}-\Sigma'_{\Upsilon,\underline{\Upsilon}_r}\right)\right\| \\ &\quad +\left\|\left(\underline{\hat{A}}_r-\underline{A}_r^*\right)\Sigma'_{\Upsilon,\underline{\Upsilon}_r}\right\|+\left\|\underline{A}_r^*\left(\hat{\Sigma}'_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r}-\Sigma'_{\Upsilon,\underline{\Upsilon}_r}\right)\right\| \\ &\quad +\left\|\sum_{i=r+1}^{\infty}A_iE\Upsilon_{t-i}\Upsilon_t'\right\|. \end{aligned} \tag{A.12}$$

In the right-hand side of this inequality, the first norm is $o_P(1)$ by Lemma A.4. By Lemmas A.5 and A.7, we have $\|\underline{\hat{A}}_r-\underline{A}_r^*\|=o_p(r^{-1/2})=o_p(1)$, and by Lemma A.4, $\|\hat{\Sigma}'_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r}-\Sigma'_{\Upsilon,\underline{\Upsilon}_r}\|=o_p(r^{-1/2})=o_p(1)$. Therefore the second norm in the right-hand side of (A.12) tends to zero in probability. The third norm tends to zero in probability because $\|\underline{\hat{A}}_r-\underline{A}_r^*\|=o_p(1)$ and, by Lemma A.2, $\|\Sigma'_{\Upsilon,\underline{\Upsilon}_r}\|=O(1)$. The fourth norm tends to zero in probability because, in view of Lemma A.4, $\|\hat{\Sigma}'_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r}-\Sigma'_{\Upsilon,\underline{\Upsilon}_r}\|=o_p(1)$, and, in view of (A.2), $\|\underline{A}_r^*\|^2\leq\sum_{i=1}^{\infty}\mathrm{Tr}(A_iA_i')<\infty$. Clearly, the last norm tends to zero, which completes the proof.

References

Amemiya, T. and Wu, R. Y. (1972) The Effect of Aggregation on Prediction in the Autoregressive Model *J. Amer. Statist. Assoc.*, **67**, 628–632.

Andrews, D. W. K. (1991) Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica*, **59**, 817–858.

Andrews, D. W. K. and Monahan, J.C. (1992) An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator. *Econometrica*, **60**, 953–966.

Berk, K. N. (1974) Consistent Autoregressive Spectral Estimates. *Ann. Statist.*, **2**, 489–502.

Bollerslev, T. (1986) Generalized Autoregressive Conditional Heteroskedasticity. *J. Econometrics*, **31**, 307–327.

Box, G. E. P. and Pierce, D. A. (1970) Distribution of Residual Autocorrelations in Autoregressive-Integreted Moving Average Time Series Models. *J. Amer. Statist. Assoc.*, **65**, 1509–1526.

Brockwell, P. J. and Davis, R. A. (1991) *Time Series: Theory and Methods*, Springer-Verlag, New York, 2nd edition.

Broze, L., Francq, C. and Zakoïan, J-M. (2002) Efficient use of higher-lag autocorrelations for estimating autoregressive processes *J. Time Series Analysis*, **23**, 287–312.

Carrasco, M. and Chen, X. (2002) Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models. *Econometric Theory*, **18**, 17-39.

Chen, W. W. and Deo, R. S. (2003) A Generalized Portmanteau Goodness-of-fit Test for Time Series Models. *Econometric Theory*, forthcoming.

Chen, W. W. and Deo, R. S. (2003) Power Transformations to Induce Normality and Their Applications. *J. Roy. Statist. Soc. B*, forthcoming.

Davies, N., Triggs, C. M. and Newbold, P. (1977) Significance levels of the Box-Pierce Portmanteau Statistic in Finite Samples. *Biometrika*, **64**, 517–522.

Davydov, Y. A. (1968) Convergence of Distributions Generated by Stationary Stochastic Processes. *Theor. Probab. Appl.*, **13**, 691–696.

Deo, R. S. (2000) Spectral Tests of the Martingale Hypothesis under Conditional Heteroskedasticity. *J. Econometrics*, **99**, 291–315.

Diebold, F. X. (1986) Testing for Serial Correlation in the Presence of ARCH. 1986 *Proceedings of the Business and Economics Statistics Section*, American Statistical Association, 323–328.

Durlauf, S. N. (1991) Spectral based Testing of the Martingale Hypothesis. *J. Econometrics*, **50**, 355–376.

Engle, R. F. (1982) Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of the United Kingdom Inflation. *Econometrica*, **50**, 987–1007.

Engle, R. F. (1998) Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. *Econometrica*, **66**, 1127–1162.

Francq, C. and Zakoïan, J-M. (1998) Estimating Linear Representations of Nonlinear Processes. *J. Statist. Plan. Infer.*, **68**, 145–165.

Francq, C. and Zakoïan, J-M. (2001) Stationarity of Multivariate Markov-switching ARMA Models. *J. Econometrics*, **102**, 339-364.

Francq, C. and Zakoïan, J-M. (2003) The L^2-Structures of Standard and Switching-Regime GARCH Models. *Unpublished document*, CREST.

Hamilton, J. D. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* **57**, 357–384.

Hamilton, J. D. (1994) *Time Series Analysis.* Princeton University Press.

Hamilton, J. D., and Susmel, R. (1994) Autoregressive conditional heteroskedasticity and changes in regime. *J. Econometrics* 64, 307-333.

Harvey, A. C. and Pierse, R. G. (1984) Estimating Missing Observations in Economic Time Series. *J. Amer. Statist. Assoc.*, **79**, 125–131.

Hong, Y. (1996) Consistent Testing for Serial Correlation of Unknown Form. *Econometrica*, **64**, 837–864.

Hong, Y. and Lee Y. J. (2003) Consistent Testing for Serial Correlation of Unknown Form Under General Conditional Heteroskedasticity. *Preprint, Cornell University.*

Imhof, J. P. (1961) Computing the Distribution of Quadratic Forms in Normal Variables. *Biometrika*, **48**, 419–426.

Ling, S. and McAleer, M. (2002) Necessary and Sufficient Moment Conditions for the GARCH(r, s) and Asymmetric Power GARCH(r, s) Models. *Econometric Theory*, **18**, 722–729.

Ljung, G. M. and Box, G. E. P. (1978) On the Measure of Lack of Fit in Time Series Models *Biometrika*, **65**, 297–303.

Lobato, I. N. (2001) Testing that a Dependent Process is Uncorrelated. *J. Amer. Statist. Assoc.*, **96**, 1066–1076.

Lobato, I. N., Nankervis J. C. and Savin, N. E. (2001) Testing for Autocorrelation Using a Modified Box-Pierce Q Test. *Inter. Econ. Review*, **42**, 187–205.

Lobato, I. N., 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. (1991) *Introduction to Multiple Time Series Analysis,* Springer Verlag , Berlin.

McLeod, A. I. (1978) On the Distribution of Residual Autocorrelations in Box-Jenkins Method. *J. Roy. Statist. Soc. B*, **40**, 296–302.

Newey, W.K. and West, K.D. (1987) A Simple Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, **55**, 703–708.

Nsiri, S. and Roy, R. (1993) On the invertibility of Multivariate Linear Processes. *J. Time Ser. Anal.*, **14**, 305–316.

Palm, F. C. and Nijman, T. E. (1984) Missing Observations in the Dynamic Regression Model. *Econometrica*, **52**, 1415–1435.

Pham, D. T. (1986) The Mixing Property of Bilinear and Generalized Random Coefficient Autoregressive Models. *Stochastic Process. Appl.*, **23**, 291–300.

Romano, J. L. and Thombs, L. A. (1996) Inference for Autocorrelations under Weak Assumptions. *J. Amer. Statist. Assoc.*, **91**, 590–600.

Ruiz, E. (1994) Quasi-Maximum Likelihood Estimation of Stochastic Volatility Models. *J. Econometrics*, 63, 289–306.

Whittle, P. (1963) On the fitting of multivariate autoregressions and the approximate canonical factorization of a spectral density matrix. *Biometrika*, **40**, 129–134.

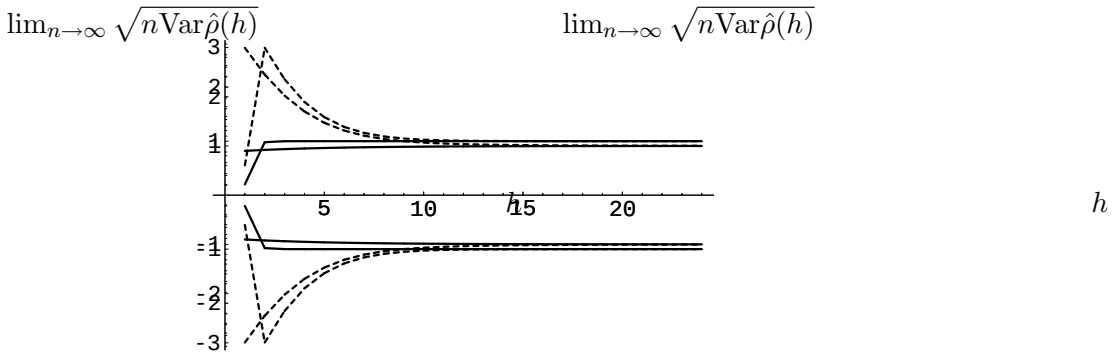

Figure 1: Asymptotic significance limits for the residual autocorrelations in the AR(1)-ARCH(1) model: $X_t = aX_{t-1} + \epsilon_t$, with $\epsilon_t = \sqrt{\omega + \alpha_1 \epsilon_t^2}\eta_t, \quad (\eta_t)$ iid $\mathcal{N}(0,1)$. Dotted lines correspond to $\pm\lim_{n\to\infty}\sqrt{n\mathrm{Var}\hat{\rho}(h)}$ (obtained from Theorem 3.2). Full lines correspond to $\pm\lim_{n\to\infty}\sqrt{n\mathrm{Var}\hat{\rho}(h)}$ for a strong AR(1) model, with $a = 0.9$ for the left graph, and $a = 0.2$ for the right graph.

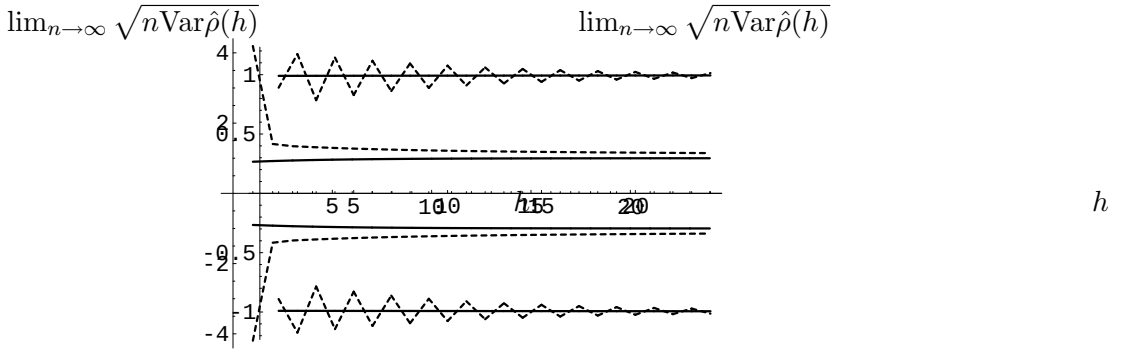

Figure 2: Asymptotic significance limits for the residual autocorrelations in the Markov-switching AR(1) model (4.7). Dotted lines correspond to $\pm\lim_{n\to\infty}\sqrt{n\mathrm{Var}\hat{\rho}(h)}$ (obtained from Theorem 3.2). Full lines correspond to $\pm\lim_{n\to\infty}\sqrt{n\mathrm{Var}\hat{\rho}(h)}$ for a strong AR(1) model, with $a = 0.9$ for the left graph, and $a = 0.99$ for the right graph.

Table 3: Relative rejection frequencies (in %) at the 5% nominal level of the standard LB, the modified LB and the HL tests applied to the AR(1) model with ARCH(1) innovations $\epsilon_t = \sqrt{1 + \alpha_1 \epsilon_{t-1}^2}\eta_t$, (η_t) iid $\mathcal{N}(0,1)$. The sample size is $n = 1,000$ and the number of replications is $N = 1,000$.

Standard LB test

m	(a, α_1)				
	(0.0,0.0)	(0.0,0.2)	(0.5,0.2)	(0.9,0.2)	(0.9,0.4)
2	5.30	7.20	7.70	15.40	26.30
3	5.80	5.60	6.40	11.60	21.40
4	5.60	4.20	5.90	9.20	20.00
12	4.60	5.90	5.10	6.40	12.20
24	5.60	4.50	5.10	7.40	10.70
36	5.40	5.60	3.60	4.70	8.20

HL test

m	(a, α_1)				
	(0.0,0.0)	(0.0,0.2)	(0.5,0.2)	(0.9,0.2)	(0.9,0.4)
7	1.20	0.20	0.20	3.80	3.80
15	2.10	1.10	1.30	4.50	3.70
23	2.90	1.30	2.30	3.10	3.80
31	3.10	1.50	2.60	3.80	4.20
39	3.70	1.60	3.00	4.00	4.20

Modified LB test

m	(a, α_1)				
	(0.0,0.0)	(0.0,0.2)	(0.5,0.2)	(0.9,0.2)	(0.9,0.4)
2	5.20	5.80	5.80	5.70	5.10
3	5.60	4.80	5.10	5.10	4.90
4	5.20	3.70	4.30	5.20	5.60
12	4.30	5.00	3.80	5.00	5.30
24	4.90	3.50	3.90	5.00	5.70
36	4.90	4.60	2.80	4.00	4.70

Table 4: Relative rejection frequencies (in %) at the 5% nominal level of the standard LB, the modified LB and the HL tests applied to the Markov switching AR(1) model (4.7). The sample size is $n = 10,000$ and the number of replications is $N = 1,000$.

Standard LB test

m	(a, p)				
	(0.0,0.95)	(0.6,0.95)	(0.6,0.05)	(0.9,0.5)	(0.99,0.95)
2	10.00	9.60	25.20	16.90	15.30
3	6.70	4.00	26.00	10.60	10.30
4	7.30	6.50	25.50	9.70	11.00
12	6.30	5.60	23.70	6.10	9.60
24	6.00	4.80	19.40	6.10	5.90
36	6.60	5.80	15.80	5.80	6.20

HL test

m	(a, p)				
	(0.0,0.95)	(0.6,0.95)	(0.6,0.05)	(0.9,0.5)	(0.99,0.95)
12	1.70	1.50	14.00	4.50	3.00
25	2.60	2.60	17.50	4.20	4.60
37	3.70	4.60	19.00	4.00	4.40
50	3.20	3.00	19.20	4.70	6.70
63	3.80	3.20	20.30	4.30	4.90

Modified LB test

m	(a, p)				
	(0.0,0.95)	(0.6,0.95)	(0.6,0.05)	(0.9,0.5)	(0.99,0.95)
2	5.50	5.10	8.40	6.00	3.60
3	5.90	4.60	7.50	5.00	4.80
4	5.00	5.30	9.50	4.50	3.70
12	4.60	4.40	7.70	5.40	5.80
24	4.70	3.80	7.30	5.70	3.60
36	5.30	4.70	6.70	5.20	3.80

Table 5: Empirical power (in %) at the 5% nominal level of the standard LB test, the modified LB test and the HL test for goodness of fit of an AR(1) model. The DGP is an ARMA(1,1) process: $X_t = aX_{t-1} + \epsilon_t + b\epsilon_{t-1}$, $\epsilon_t = \sqrt{1 + \alpha_1 \epsilon_{t-1}^2}\eta_t$, (η_t) iid $\mathcal{N}(0,1)$. The sample size is $n = 1,000$ and the number of replications is $N = 1,000$.

Standard LB test (adjusted for a correct type I error)

m	(a, b, α_1)				
	(0.0,0.2,0.0)	(0.0,0.2,0.2)	(0.5,0.2,0.2)	(0.9,0.2,0.2)	(0.9,0.2,0.4)
2	27.00	19.10	90.70	99.70	96.70
3	16.50	20.20	89.80	99.50	96.00
4	19.30	17.20	89.70	99.30	96.20
12	11.80	10.10	76.20	97.70	92.50
24	9.90	9.60	56.80	91.30	86.10
36	6.90	10.50	47.80	88.70	81.40

HL test

m	(a, b, α_1)				
	(0.0,0.2,0.0)	(0.0,0.2,0.2)	(0.5,0.2,0.2)	(0.9,0.2,0.2)	(0.9,0.2,0.4)
7	10.30	4.10	82.30	99.60	95.90
15	9.70	8.60	83.30	99.40	95.00
23	11.80	8.80	83.10	99.20	95.00
31	10.30	8.30	80.70	99.10	93.90
39	10.10	7.80	77.30	98.40	92.60

Modified LB test

m	(a, b, α_1)				
	(0.0,0.2,0.0)	(0.0,0.2,0.2)	(0.5,0.2,0.2)	(0.9,0.2,0.2)	(0.9,0.2,0.4)
2	27.80	23.10	91.70	99.60	95.90
3	16.30	20.00	90.50	99.50	95.90
4	18.00	15.70	89.50	99.20	95.00
12	11.00	9.10	73.50	97.50	91.80
24	8.20	8.10	55.90	91.60	85.00
36	7.50	7.50	47.30	89.00	80.60

Table 6: Standard and modified LB white noise tests on the S&P returns.

lag	1	2	3	4	5
LB statistic	8.04	16.11	22.35	24.55	24.57
p-value of the standard version	0.0046	0.0003	0.0001	0.0001	0.0002
p-value of the modified version	0.2688	0.3242	0.3104	0.3139	0.4432

6	7	8	9	12	18	24	30	36
24.59	25.51	26.59	26.61	35.34	38.43	43.93	60.46	78.11
0.0004	0.0006	0.0008	0.0016	0.0004	0.0034	0.0078	0.0008	0.0001
0.4872	0.5057	0.5351	0.5839	0.5042	0.6173	0.6862	0.5527	0.4376

Table 7: Standard and modified LB tests for goodness-of-fit of an ARMA(1,1) model for the squares of the S&P returns.

lag				1	2	3	4	5
LB statistic				2.25	15.30	16.60	37.26	84.12
p-value of the standard version				0.0000	0.0000	0.0000	0.0000	0.0000
p-value of the modified version				0.5827	0.5190	0.5652	0.4285	0.3796
6	7	8	9	12	18	24	30	36
88.31	94.21	95.84	95.94	97.39	98.32	100.34	106.60	111.01
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.3788	0.3735	0.3733	0.3755	0.3759	0.3765	0.3752	0.3702	0.3655

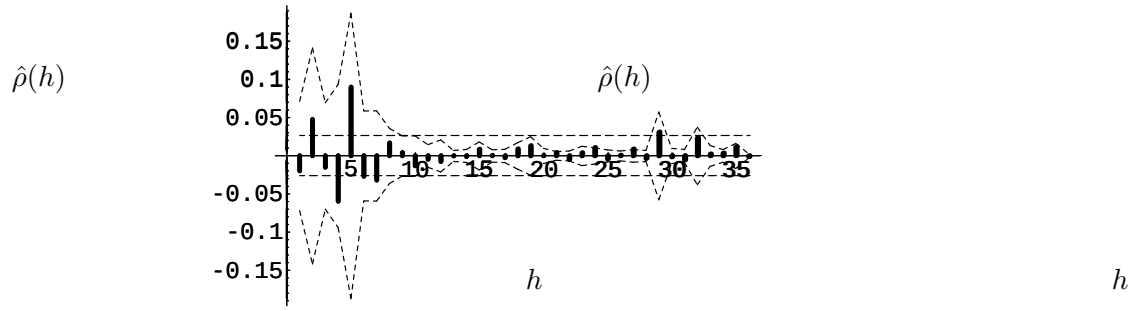

Figure 3: Autocorrelation of the ARMA(1,1) residuals for the squares of the S&P returns.