Journal of Time Series Analysis, 36, 39-60, 2015

DOI:http://dx.doi.org/10.1111/jtsa.12091

A Joint Portmanteau Test for Conditional Mean and Variance Time Series Models

Carlos Velasco*

Universidad Carlos III de Madrid

Xuexin Wang†

Xiamen University

July 23, 2014

Abstract

In this paper we propose a new joint Portmanteau test for checking the specification of parametric conditional mean and variance functions of linear and nonlinear time series models. The use of a joint test is motivated for complete control of the asymptotic size since marginal tests for the conditional variance may lead to misleading conclusions when the conditional mean is misspecified. The new test is based on an asymptotically distribution-free transformation on the sample autocorrelations of both normalized residuals and squared normalized residuals, extending Delgado and Velasco (2011). This makes unnecessary to full

*Department of Economics, Universidad Carlos III de Madrid, 28903 Getafe (Madrid), Spain. carlos.velasco@uc3m.es.

Research support from the Spanish Plan Nacional de I+D+I (ECO2012-31748) is gratefully acknowledged. We are grateful to the Co-editor, Peter M. Robinson, and two anonymous referees for their helpful comments and suggestions.

†The Wang Yanan Institute of Studies in Economics, Xiamen University, Fujian, China. xuexinwang@xmu.edu.cn.

Research support from National Natural Science Funds of China (Grant No.71301136) is gratefully acknowledged.

detail the asymptotic properties of the estimates used to obtain residuals, which could be inefficient two-step ones, avoiding also choices of maximum lag parameters increasing with sample length to control asymptotic size. The robust versions of the new test also properly account for higher order moment dependence at a reduced cost. The finite-sample performance of the new test is compared with those of well known tests through simulations.

JEL Classification: C12, C22

Keywords: Model Diagnostic Checking; Portmanteau Statistic; Estimation Effect; GARCH Model Specification Testing; Residual serial correlation.

1 Introduction

During the time series model building process, it is critical to check whether the residuals of a time series model are approximately uncorrelated, since a good model should be able to describe the dependence structure of the data adequately, and one important measurement of dependence is via the autocorrelation function of residuals. But the autocorrelation only attends to linear dependence, so when modeling other dynamic aspects, such as the conditional variance, or using general nonlinear specifications, further serial dependence measures of the residuals have to be considered.

It has been a long history of studying the distribution of residual autocorrelations in linear time series models. With the popularization of the Box-Jenkins modeling approach in 1970s, Box and Pierce (1970) and Ljung and Box (1978) propose portmanteau tests to check the adequacy of ARMA models. Box and Pierce (1970) and Durbin (1970) show that, although the sample autocorrelations of ARMA residuals under the true parameters are asymptotic independently normal distributed, this does not hold when genuine innovations are substituted by estimated residuals. More specifically, consider the ARMA (p, q) model

$$e_{\theta t} = \varphi_\theta (L) Y_t,\ t \in \mathbb{Z},$$

where $\varphi_\theta (z) = A_\theta (z) B_\theta^{-1} (z)$, $A_\theta (z) = 1 - \sum_{j=1}^{p} a_j z^j$, $B_\theta (z) = 1 - \sum_{j=1}^{q} b_j z^j$, in which $A_\theta (z)$ and $B_\theta (z)$ have no roots in common and all roots are outside the unit circle for $\theta \in \Theta$.

Consider the residuals $\{e_{\theta t}\}_{t\in\mathbb{Z}}$, and define the residual sample autocorrelation function

$$\rho_\theta (j) = \frac{\gamma_\theta (j)}{\gamma_\theta (0)},\ j \in 1, 2, \ldots,$$

where $\gamma_\theta (j) = Cov (e_{\theta t}, e_{\theta t-j})$, $j \in \mathbb{Z}$, is the corresponding autocorrelation function.

The null hypothesis of correct specification for such linear model is

$$H_0 : \rho_{\theta_0}(j) = 0 \text{ for all } j \in 1, 2, \ldots \text{ and some } \theta_0 \in \Theta.$$

Given observations $\{Y_t\}_{t=1}^T$, ρ_θ is estimated by the sample autocorrelation function

$$\hat{\rho}_{T\theta}(j) = \frac{\hat{\gamma}_{T\theta}(j)}{\hat{\gamma}_{T\theta}(0)}, \; j \in 1, \ldots, T-1,$$

where

$$\hat{\gamma}_{T\theta}(j) = \frac{1}{T} \sum_{t=j+1}^{T} (e_{\theta t} - \bar{e}_\theta)(e_{\theta t-j} - \bar{e}_\theta)$$

is the sample autocovariance function and $\bar{e}_\theta = T^{-1} \sum_{t=1}^T e_{\theta t}$ is the residual sample mean. When some further conditions are imposed on the model errors, such as $\{e_{\theta_0 t}\}_{t\in\mathbb{Z}}$ being an independently and identically distributed (i.i.d.), martingale difference, or mixing sequence with some restrictions on higher order moments, it is well known that $\left\{\sqrt{T}\hat{\rho}_{T\theta}(j)\right\}_{j=1}^{s}$ are asymptotically distributed as independent standard normals, so the Box-Pierce-Ljung portmanteau test

$$BPL(s) = T(T+2) \sum_{j=1}^{s} (T-j)^{-1} \hat{\rho}_{T\theta_0}^2(j)$$

follows an asymptotic χ^2 distribution with s degrees of freedom under H_0.

When θ_0 is unknown and we replace it by a pseudo-maximum likelihood estimator $\hat{\theta}_T$, then the null distribution of

$$\widehat{BPL}(s) = T(T+2) \sum_{j=1}^{s} (T-j)^{-1} \hat{\rho}_{T\hat{\theta}_T}^2(j)$$

is approximated by a χ^2 distribution with $s-(p+q)$ degrees of freedom. Note that the degrees of freedom of the Box-Pierce-Ljung test depend on the number of the estimated parameters due to the impact of the parameter estimation uncertainty. However, when it comes to nonlinear time series models, the previous asymptotic properties of portmanteau tests for ARMA models breaks down. Usually, the analysis of Lagrange multiplier and Portmanteau tests requires to derive the asymptotic theory for quadratic forms of the residual sample autocorrelations, which depends on the model and the estimator considered, see e.g. Francq, Roy, and Zakoïan (2005) for weak ARMA models.

Quite recently Delgado and Velasco (2011) develop an asymptotically distribution-free transform of the sample autocorrelations of residuals in general parametric linear time series models. This paper shows that the proposed Box-Pierce type test statistic based on the transformed autocorrelation is not affected by the estimation effect.

For financial time series, where dynamic conditional heteroskedasticity is the norm, the ARMA model with constant variance is inadequate to describe the data. A non-constant conditional variance of $e_{\theta t}$ can be modeled by

$$e_{\theta t} = h\left(I_{t-1}, \theta\right) \varepsilon_{\theta t},\ \varepsilon_{\theta t} \sim i.i.d.\left(0, \sigma^2\right),$$

where I_{t-1} denotes the information set at t. There are many possible specifications of the function $h\left(I_{t-1}, \theta\right)$. Engle (1982) proposes the autoregressive conditional heteroskedasticity (ARCH) model, $h^2\left(I_{t-1}, \theta\right) = \varpi_0 + \alpha_1 e_{\theta t-1}^2 + \cdots + \alpha_m e_{\theta t-m}^2$. Bollerslev (1986) proposes the GARCH models, which involve infinite lags of $e_{\theta t}^2$, as ARMA models involve all lags of $e_{\theta t}$. Since then, GARCH models have become more and more popular and successful in economics and finance. In this case the autocorrelations of squared normalized residuals derived from these models should be useful in checking the adequacy of $h\left(I_{t-1}, \theta\right)$. In this regard, a Portmanteau statistic on the first s autocorrelations of squared normalized residuals is proposed by Higgins and Bera (1992) for checking of the adequacy of the ARCH model specifications. However, the proposed approximation of the asymptotic distribution for the test statistic by a χ^2 distribution with s degrees of freedom turned out to be not appropriate. Li and Mak (1994) propose a portmanteau test based on the correct asymptotic distribution of the autocorrelations of squared normalized residuals, while Lundbergh and Teräsvirta (2002) establish the asymptotic equivalence between Li and Mak's statistic and the LM statistic.

Nowadays, dynamic econometric models that jointly parameterize conditional means and conditional variances are becoming increasingly popular in the analysis of economic and financial time series. This class of models appears in several dynamic contexts, such as asset pricing, portfolio choices, and market risk management. While there exist portmanteau tests for conditional mean models or for conditional variance models, the literature on joint model checking for the conditional mean and variance functions is rather scarce. The joint portmanteau test is motivated for complete control of the asymptotic size since marginal portmanteau tests for the conditional variance may lead to misleading conclusions when the conditional mean is misspecified. Wong and Ling (2005) consider simultaneously the Box-Pierce-Ljung and Li-Mak test statistics to jointly test the model adequacy of the conditional mean and variance models when pseudo-maximum likelihood estimates are used. Escanciano (2008) proposes a class of joint and marginal spectral diagnostic tests for parametric conditional mean and variance functions of linear and nonlinear time series models. Escanciano (2008) approach enjoys a consistency property by considering an increasing number of lags, but the asymptotic null distributions of these tests depend on the data generating process because of the parameter estimation uncertainty so that a bootstrap procedure has to be applied.

In this paper, instead, we propose an asymptotic simultaneous distribution-free transform of the sample autocorrelations of standardized residuals and their squares, extending Delgado and Velasco (2011) approach to the conditional mean and variance models diagnosis. We then consider portmanteau type tests based on these transformation. This makes unnecessary to full detail the asymptotic properties of the estimates used to obtain residuals, which could be inefficient two-step ones, avoiding also choices of maximum lag parameters increasing with sample length to control asymptotic size of tests. The robust versions of the new test can properly account for higher order moment dependence at a reduced computational cost. The outline of the rest of the paper is as following. In Section 2, we establish the transform. Section 3 studies its asymptotic properties, and propose the Box-Pierce test statistic. Section 4 is a Monte Carlo study of the joint portmanteau test. Proofs are contained in a technical Appendix.

2 Transformed Residual autocorrelations

We consider the following mean-scale model for observations Y_t,

$$Y_t = f\left(I_{t-1}, \theta\right) + h\left(I_{t-1}, \theta\right) \varepsilon_{\theta t}, \;\; t \in \mathbb{Z}, \tag{1}$$

where $f\left(I_{t-1}, \theta\right)$ and $h^2\left(I_{t-1}, \theta\right)$ are the parametric specifications for the first two conditional moments of Y_t, $f\left(I_{t-1}\right) = E\left(Y_t | I_{t-1}\right)$ and $h^2\left(I_{t-1}\right) = Var\left(Y_t | I_{t-1}\right)$, respectively; I_t is the information set generated by $\{Y_t, Y_{t-1}, \ldots\}$; θ is a finite-dimensional unknown parameter vector such that $\theta \in \Theta \subset \mathbb{R}^k$. If we assume $\varepsilon_{\theta_0 t}$ in such a way that for some $\theta_0 \in \Theta$

$$E(\varepsilon_{\theta_0 t} | I_{t-1}) = 0, \;\; E(\varepsilon^2_{\theta_0 t} | I_{t-1}) = 1 \tag{2}$$

hold, then we have $f\left(I_{t-1}, \theta_0\right) = E\left(Y_t | I_{t-1}\right)$, $h^2\left(I_{t-1}, \theta_0\right) = Var\left(Y_t | I_{t-1}\right)$. This assumption is weaker than assuming $\{\varepsilon_{\theta_0 t}\}_{t \in \mathbb{Z}}$ is i.i.d. with mean 0 and variance 1, which is usually assumed in much of the related literature. This provides additional generality, since there is a growing econometrics and finance literature documenting time-varying conditional skewness and kurtosis in economic and financial time series, see, e.g. Gallant et al. (1991), Hansen (1994), Harvey and Siddique (1999) and Jondeau and Rockinger (2003). This specification covers most commonly used linear and nonlinear dynamic time series models. Examples include the autoregressive conditional heteroskedasticity (ARCH), autoregressive moving average (ARMA), bilinear, nonlinear moving average, Markov regime-switching, smooth transition, exponential, and threshold autoregressive models.

Consider $\{\varepsilon_{\theta t}\}_{t \in \mathbb{Z}}$ and $\{\varepsilon^2_{\theta t}\}_{t \in \mathbb{Z}}$, and define the residual sample autocorrelation functions

$$\rho_\theta\left(j\right) = \tfrac{\gamma_\theta(j)}{\gamma_\theta(0)}, \;\; j \in 1, 2, \ldots,$$
$$\delta_\theta\left(j\right) = \tfrac{\eta_\theta(j)}{\eta_\theta(0)}, \;\; j \in 1, 2, \ldots,$$

where $\gamma_\theta(j) = Cov(\varepsilon_{\theta t}, \varepsilon_{\theta t-j})$ and $\eta_\theta(j) = Cov\left(\varepsilon^2_{\theta t}, \varepsilon^2_{\theta t-j}\right)$, $j \in 1, 2, \ldots$, are, respectively, the corresponding autocovariance functions of the standardized residuals and the square of the residuals.

Then, if the model (1) is correctly specified the null hypothesis

$$H_0^{(m)}: \begin{array}{c} \rho_{\theta_0}(1) = \rho_{\theta_0}(2) = \cdots = \rho_{\theta_0}(m) = 0 \\ \delta_{\theta_0}(1) = \delta_{\theta_0}(2) = \cdots = \delta_{\theta_0}(m) = 0 \end{array}$$

is satisfied for any $m = 1, 2, \ldots$, expressing the limited serial dependence of the errors in the first two (unconditional) moments.

Given observations $\{Y_t\}_{t=1}^T$, $\rho_\theta(j)$ and $\delta_\theta(j)$ are estimated by the sample autocorrelation functions

$$\begin{aligned} \hat{\rho}_{T\theta}(j) &= \frac{\hat{\gamma}_{T\theta}(j)}{\hat{\gamma}_{T\theta}(0)}, \; j \in 1, 2, \ldots, \\ \hat{\delta}_{T\theta}(j) &= \frac{\hat{\eta}_{T\theta}(j)}{\hat{\eta}_{T\theta}(0)}, \; j \in 1, 2, \ldots, \end{aligned}$$

where

$$\begin{aligned} \hat{\gamma}_{T\theta}(j) &= \frac{1}{T}\sum_{t=j+1}^{T}(\varepsilon_{\theta t} - \bar{\varepsilon}_\theta)(\varepsilon_{\theta t-j} - \bar{\varepsilon}_\theta), \; j \in \mathbb{Z} \\ \hat{\eta}_{T\theta}(j) &= \frac{1}{T}\sum_{t=j+1}^{T}\left(\varepsilon^2_{\theta t} - \bar{\varepsilon}^2_\theta\right)\left(\varepsilon^2_{\theta t-j} - \bar{\varepsilon}^2_\theta\right), \; j \in \mathbb{Z} \end{aligned}$$

are the corresponding sample autocovariance functions and $\bar{\varepsilon}_\theta = T^{-1}\sum_{t=1}^T \varepsilon_{\theta t}$, $\bar{\varepsilon}^2_\theta = T^{-1}\sum_{t=1}^T \varepsilon^2_{\theta t}$.

Define $\hat{\rho}^{(m)}_{T\theta_0} = (\hat{\rho}_{T\theta_0}(1), \ldots, \hat{\rho}_{T\theta_0}(m))'$, $\hat{\delta}^{(m)}_{T\theta_0} = \left(\hat{\delta}_{T\theta_0}(1), \ldots, \hat{\delta}_{T\theta_0}(m)\right)'$ for a fixed m. If $\{\varepsilon_{\theta_0 t}\}$ were a sequence of i.i.d. innovations, then

$$\sqrt{T}\begin{bmatrix} \hat{\rho}^{(m)}_{T\theta_0} \\ \hat{\delta}^{(m)}_{T\theta_0} \end{bmatrix} \to_d N\left(0, I^{(2m)}\right),$$

where $I^{(2m)}$ is the identity matrix of dimension $2m$.

In general, i.e., in absence of further limitations of the dependence structure of model errors apart from $H_0^{(m)}$, we can only expect that under weak dependence conditions

$$\sqrt{T}\left[\begin{array}{c}\hat{\rho}_{T\theta_0}^{(m)}\\ \hat{\delta}_{T\theta_0}^{(m)}\end{array}\right]\rightarrow_d N\left(0,\Omega_{\theta_0}^{(2m)}\right),$$

for an unrestricted positive definite covariance matrix $\Omega_{\theta_0}^{(2m)}$, while under errors with constant first two conditional models as in equation (2),

$$\Omega_{\theta_0}^{(2m)}=\left(\begin{array}{cc}I^{(m)} & \Pi_{\theta_0}^{(m)}\\ \Pi_{\theta_0}^{(m)\prime} & \Sigma_{\theta_0}^{(m)}\end{array}\right),$$

with

$$\Pi_{\theta_0}^{(m)}=\left[\frac{\upsilon_{\theta_0}^{(i,j)}}{\eta_{\theta_0}(0)}\right]_{i,j=1}^{m}\quad\text{and}\quad\Sigma_{\theta}^{(m)}=\left[\frac{\sigma_{\theta}^{(i,j)}}{\eta_{\theta}(0)^2}\right]_{i,j=1}^{m},$$

and $\upsilon_{\theta_0}^{(i,j)}=E\left[\varepsilon_{\theta_0 t}^3\varepsilon_{\theta_0 t-i}\left(\varepsilon_{\theta_0 t-j}^2-1\right)\right]$ and $\sigma_{\theta_0}^{(i,j)}=E\left[\left(\varepsilon_{\theta_0 t}^2-1\right)^2\left(\varepsilon_{\theta_0 t-i}^2-1\right)\left(\varepsilon_{\theta_0 t-j}^2-1\right)\right]$. If further $E\left(\varepsilon_{\theta_0 t}^3|I_{t-1}\right)=0$, then $\upsilon_{\theta_0}^{(i,j)}=0$, $i,j>0$, and we have

$$\Omega_{\theta_0}^{(2m)}=\left(\begin{array}{cc}I^{(m)} & 0\\ 0 & \Sigma_{\theta_0}^{(m)}\end{array}\right).$$

Therefore, when dealing with $\hat{\rho}_{T\theta_0}^{(m)}$ and $\hat{\delta}_{T\theta_0}^{(m)}$, we would need to account for proper standardization, but in practice, we do not know the true values of the parameters, so they have to be estimated in the first place. Assume that there exists an estimator $\hat{\theta}_T$ such that

$$\hat{\theta}_T=\theta_0+O_p\left(T^{-1/2}\right),\tag{3}$$

and we compute the sample serial correlation of residuals and their squares up to lag

m as

$$\begin{aligned}
\hat{\rho}^{(m)}_{T\hat{\theta}_T} &= \left(\hat{\rho}_{T\hat{\theta}_T}(1), \cdots, \hat{\rho}_{T\hat{\theta}_T}(m)\right)' \\
\hat{\delta}^{(m)}_{T\hat{\theta}_T} &= \left(\hat{\delta}_{T\hat{\theta}_T}(1), \cdots, \hat{\delta}_{T\hat{\theta}_T}(m)\right)'.
\end{aligned}$$

In the following proposition, we show that the residual autocorrelations suffer from estimation effects. Its proof and the proof of all our results are contained in the Appendix, together with the technical assumptions used.

Proposition 1. *Under $H^{(m)}_0$, (3) and Assumptions A1 to A4 in the Appendix,*

$$\begin{aligned}
\sqrt{T}\hat{\rho}^{(m)}_{T\hat{\theta}_T} &= \sqrt{T}\hat{\rho}^{(m)}_{T\theta_0} + \nabla\rho^{(m)}_{\theta_0}\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) + o_p(1) \\
\sqrt{T}\hat{\delta}^{(m)}_{T\hat{\theta}_T} &= \sqrt{T}\hat{\delta}^{(m)}_{T\theta_0} + \nabla\delta^{(m)}_{\theta_0}\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) + o_p(1),
\end{aligned}$$

where $\nabla\rho^{(m)}_{\theta_0} = p\lim \frac{\partial}{\partial\theta'}\hat{\rho}^{(m)}_{T\theta_0}$ and $\nabla\delta^{(m)}_{\theta_0} = p\lim \frac{\partial}{\partial\theta'}\hat{\delta}^{(m)}_{T\theta_0}$.

The assumptions of the proposition further restrict the dependence of the residual and observation processes so that the linear expansion holds under mixing conditions. In Section 3.3 we provide expressions for $\nabla\rho^{(m)}_{\theta_0}$ and $\nabla\delta^{(m)}_{\theta_0}$ for simple models, and see that in general those will be nonzero. Then, the asymptotic distribution of $\hat{\rho}^{(m)}_{T\hat{\theta}_T}$ and $\hat{\delta}^{(m)}_{T\hat{\theta}_T}$ depends both on that of $\sqrt{T}\left(\hat{\rho}^{(m)\prime}_{T\theta_0}, \hat{\delta}^{(m)\prime}_{T\theta_0}\right)'$ and $\sqrt{T}\left(\hat{\theta}_T - \theta_0\right)$. For instance, to derive the covariance matrix of $\sqrt{T}\left(\hat{\rho}^{(m)\prime}_{T\hat{\theta}_T}, \hat{\delta}^{(m)\prime}_{T\hat{\theta}_T}\right)'$ correctly, the asymptotic joint distribution of $\sqrt{T}\left(\hat{\theta}_T - \theta_0\right)$ and $\sqrt{T}\left(\hat{\rho}^{(m)\prime}_{T\theta_0}, \hat{\delta}^{(m)\prime}_{T\theta_0}\right)'$ has to be considered, which depends on the model characteristics, the method of estimating $\hat{\theta}_T$ and the unknown parameter value θ_0.

In this article, we propose an asymptotically distribution-free transform of the sample autocorrelations of residuals that accounts for both problems, standardization and estimation effects, simultaneously. Consider first a positive definite matrix of estimates $\hat{\Omega}^{(2m)}_{\theta}$ such that

$$\hat{\Omega}^{(2m)}_{T\hat{\theta}_T} = \Omega^{(2m)}_{\theta_0} + o_p(1) \tag{4}$$

under $H_0^{(m)}$. This problem has been pursued in related contexts by e.g. Lobato, Nankervis and Savin (2002) and Francq, Roy and Zakoïan (2005) using two different approaches. Both methods can be seen as extensions of usual heteroskedasticity and autocorrelation consistent (HAC) robust estimation of asymptotic covariance matrices of regression coefficients. The first reference proposes using nonparametric spectral estimates as originally developed by Jowett (1955), Hannan (1957) and Brillinger (1979) and later popularized by Newey and West (1987) in the econometrics literature, while the second one uses a VAR approximation to build the spectral estimates as suggested by, e.g., Den Haan and Levin (1997) following early ideas in Press and Tukey (1956), Blackman and Tukey (1958) and Grenander and Rosenblatt (1957).

We first normalize $\hat{\rho}_{T\hat{\theta}_T}^{(m)}$ and $\hat{\delta}_{T\hat{\theta}_T}^{(m)}$ into

$$\sqrt{T}\begin{bmatrix}\tilde{\rho}_{T\hat{\theta}_T}^{(m)}\\ \tilde{\delta}_{T\hat{\theta}_T}^{(m)}\end{bmatrix}=\sqrt{T}\hat{\Omega}_{T\hat{\theta}_T}^{(2m)-1/2}\begin{bmatrix}\hat{\rho}_{T\hat{\theta}_T}^{(m)}\\ \hat{\delta}_{T\hat{\theta}_T}^{(m)}\end{bmatrix}.$$

Based on Proposition 1, it is easy to normalize the linear expansions of residuals autocorrelations under (4) and obtain that

$$\sqrt{T}\begin{bmatrix}\tilde{\rho}_{T\hat{\theta}_T}^{(m)}\\ \tilde{\delta}_{T\hat{\theta}_T}^{(m)}\end{bmatrix}=\begin{bmatrix}\nabla\tilde{\rho}_{\theta_0}^{(m)}\\ \nabla\tilde{\delta}_{\theta_0}^{(m)}\end{bmatrix}\sqrt{T}\left(\hat{\theta}_T-\theta_0\right)+\sqrt{T}\begin{bmatrix}\tilde{\rho}_{T\theta_0}^{(m)}\\ \tilde{\delta}_{T\theta_0}^{(m)}\end{bmatrix}+o_p\left(1\right), \tag{5}$$

where

$$\begin{bmatrix}\nabla\tilde{\rho}_{\theta_0}^{(m)}\\ \nabla\tilde{\delta}_{\theta_0}^{(m)}\end{bmatrix}=\Omega_{\theta_0}^{(2m)-1/2}\begin{bmatrix}\nabla\rho_{\theta_0}^{(m)}\\ \nabla\delta_{\theta_0}^{(m)}\end{bmatrix}.$$

Now expression (5) can be recasted as an approximated linear regression model, where the errors $\sqrt{T}\left(\tilde{\rho}_{T\theta_0}^{(m)\prime},\tilde{\delta}_{T\theta_0}^{(m)\prime}\right)'$ have identity covariance matrix and the estimation effect $\sqrt{T}\left(\hat{\theta}_T-\theta_0\right)$ is the vector of unknown coefficients. Our transformation approach tries to project out this nuisance effect through a simple recursive least squares algorithm which transmits such orthogonality of true errors into the projection residuals, see Brown, Durbin and Evans (1975).

For that, first group $\tilde{\rho}^{(m)}_{T\hat{\theta}_T}$ and $\tilde{\delta}^{(m)}_{T\hat{\theta}_T}$ into $\tilde{\lambda}_{T\hat{\theta}_T}(1) = \left(\tilde{\rho}^{(m)}_{T\hat{\theta}_T}(1), \tilde{\delta}^{(m)}_{T\hat{\theta}_T}(1)\right)', \ldots, \tilde{\lambda}_{T\hat{\theta}_T}(m) = \left(\tilde{\rho}^{(m)}_{T\hat{\theta}_T}(m), \tilde{\delta}^{(m)}_{T\hat{\theta}_T}(m)\right)'$, and define

$$\tilde{\Lambda}_{T\hat{\theta}_T}(i) = \begin{pmatrix} \nabla\tilde{\rho}^{(m)}_{T\hat{\theta}_T}(i) \\ \nabla\tilde{\delta}^{(m)}_{T\hat{\theta}_T}(i) \end{pmatrix}, \quad \text{with} \begin{bmatrix} \nabla\tilde{\rho}^{(m)}_{T\hat{\theta}_T} \\ \nabla\tilde{\delta}^{(m)}_{T\hat{\theta}_T} \end{bmatrix} = \hat{\Omega}^{(2m)-1/2}_{T\hat{\theta}_T} \begin{bmatrix} \nabla\rho^{(m)}_{T\hat{\theta}_T} \\ \nabla\delta^{(m)}_{T\hat{\theta}_T} \end{bmatrix},$$

and

$$\tilde{\Lambda}_{\theta_0}(i) = \begin{pmatrix} \nabla\tilde{\rho}^{(m)}_{\theta_0}(i) \\ \nabla\tilde{\delta}^{(m)}_{\theta_0}(i) \end{pmatrix},$$

for $i = 1, \ldots, m$, where for instance $\nabla\tilde{\rho}^{(m)}_{T\hat{\theta}_T}$ is the actual derivative of the sample autocorrelation $\tilde{\rho}^{(m)}_{T\hat{\theta}_T}$, with limit $\nabla\tilde{\rho}^{(m)}_{\theta_0}$. These (empirical) derivatives terms can be computed using simplified explicit formulae which only accounts for the derivative of the autocovariances in the numerator (since the estimation effect in the sample variance of residuals plays no role asymptotically) but can also be approximated with numerical methods perturbing the residuals autocorrelations around the parameter estimate $\hat{\theta}_T$.

For $i = 1, \ldots, m-k$, the transformation of the pairs of residual autocorrelations is

$$\bar{\lambda}_{T\hat{\theta}_T}(i) = \tilde{\Xi}_{T\hat{\theta}_T}(i) \left[\tilde{\lambda}_{T\hat{\theta}_T}(i) - \tilde{\Lambda}_{T\hat{\theta}_T}(i) \left(\sum_{j=i+1}^{m} \tilde{\Lambda}_{T\hat{\theta}_T}(j)' \tilde{\Lambda}_{T\hat{\theta}_T}(j) \right)^{-1} \sum_{j=i+1}^{m} \tilde{\Lambda}_{T\hat{\theta}_T}(j)' \tilde{\lambda}_{T\hat{\theta}_T}(j) \right],$$

where we make a recursive projection of $\tilde{\lambda}_{T\hat{\theta}_T}(i)$ on $\tilde{\Lambda}_{T\hat{\theta}_T}(j)$, employing only forward observations $j = i+1, \ldots, m$, and

$$\tilde{\Xi}_{T\hat{\theta}_T}(i) = \left[I^{(2)} + \tilde{\Lambda}_{T\hat{\theta}_T}(i) \left(\sum_{j=i+1}^{m} \tilde{\Lambda}_{T\hat{\theta}_T}(j)' \tilde{\Lambda}_{T\hat{\theta}_T}(j) \right)^{-1} \tilde{\Lambda}_{T\hat{\theta}_T}(i)' \right]^{-1/2}$$

accounts for the projection residuals standardization as can be shown by simple algebra. Only up to $m-k$ projected coefficients can be computed due to the restriction on the minimum number of correlations to perform the projection.

3 Main Results

In this section we show that our transformed (squared) residual autocorrelations are asymptotically distribution-free and propose new specification tests based on them.

We prove in the following theorem that, under $H_0^{(m)}$, the vector of projected residual autocorrelations,

$$\bar{\lambda}_{T\hat{\theta}_T}^{(m-k)} = \left(\bar{\lambda}_{T\hat{\theta}_T}(1)', \ldots, \bar{\lambda}_{T\hat{\theta}_T}(m-k)'\right)',$$

and $\bar{\lambda}_{T\theta_0}^{(m-k)}$, the vector of projected true error autocorrelations, are asymptotically equivalent, and therefore $\sqrt{T}\bar{\lambda}_{T\theta_0}^{(m-k)}$ is asymptotically distributed as a vector of independent standard normals, which renders asymptotic inference feasible for any fixed m.

Theorem 1. *Under $H_0^{(m)}$, $m > k$, Assumptions A1 to A4 in the Appendix and with $\hat{\theta}_T$ and $\hat{\Omega}_{T\hat{\theta}_T}^{(2m)}$ satisfying (3) and (4), respectively,*

$$\bar{\lambda}_{T\hat{\theta}_T}^{(m-k)} = \bar{\lambda}_{T\theta_0}^{(m-k)} + o_p\left(T^{-1/2}\right)$$

and

$$\sqrt{T}\bar{\lambda}_{T\theta_0}^{(m-k)} \to_d N\left(0, I^{(2(m-k))}\right).$$

The regularity conditions of the theorem guarantee that a central limit theorem holds for the error autocorrelations without further restrictions on the asymptotic covariance matrix and also impose an identification condition so that the recursive projection has enough degrees of freedom.

3.1 Local Alternatives

We consider the following local alternative sequence

$$H_{1T}: \rho_{\theta_0}(j) = \frac{r_{\theta_0\rho}(j)}{\sqrt{T}}, \ \delta_{\theta_0}(j) = \frac{r_{\theta_0\delta}(j)}{\sqrt{T}}, \text{ for all } j = 1, 2, ...,$$

where $r_{\theta_0\rho}$ and $r_{\theta_0\delta}$ are square summable so that ρ_{θ_0} and δ_{θ_0} are positive semidefinite sequences for all T. In order to describe the asymptotic distribution of $\bar{\lambda}_{T\hat{\theta}_T}^{(m-k)}$ under

H_{1T}, define first the vector $\bar{\tau}_{\theta}^{(m-k)} = \left(\bar{\tau}_{\theta}(1)', \ldots, \bar{\tau}_{\theta}(m-k)'\right)'$ as the projected and standardized drift of the residual autocovariances, where

$$\bar{\tau}_{\theta}(i)' = \tilde{\Xi}_{\theta}(i)\left[\tilde{\tau}_{\theta}(i) - \tilde{\Lambda}_{\theta}(i)\left(\sum_{j=i+1}^{m}\tilde{\Lambda}_{\theta}(j)'\tilde{\Lambda}_{\theta}(j)\right)^{-1}\sum_{j=i+1}^{m}\tilde{\Lambda}_{\theta}(j)'\tilde{\tau}_{\theta}(j)\right]$$

for $i = 1, 2, \ldots, m-k$, and

$$\tilde{\tau}_{\theta}^{m} = \Omega_{\theta}^{(2m)-1/2}\tau_{\theta}^{m},$$

where $\tau_{\theta}^{m} = \left(\tau_{\theta}(1)', \ldots, \tau_{\theta}(m)'\right)'$ with $\tau_{\theta}(i) = \left(r_{\theta\rho}(i), r_{\theta\delta}(i)\right)'$.

Theorem 2. *Under H_{1T}, $m > k$, Assumptions A1 to A4 in the Appendix and with $\hat{\theta}_T$ and $\hat{\Omega}_{\hat{\theta}_T}^{(2m)}$ satisfying (3) and (4), respectively,*

$$\bar{\lambda}_{T\hat{\theta}_T}^{(m-k)} = \bar{\lambda}_{T\theta_0}^{(m-k)} + o_p\left(T^{-1/2}\right)$$

and

$$\sqrt{T}\bar{\lambda}_{T\theta_0}^{(m-k)} \rightarrow_d N\left(\bar{\tau}_{\theta_0}^{(m-k)}, I^{2(m-k)}\right).$$

Theorem 3 shows that the projected autocorrelations have nonzero mean if the drifts of ρ_{θ_0} and δ_{θ_0} are not fully explained by the autocorrelations scores.

3.2 Box-Pierce Type Tests

Based on Theorems 1 and 2, we can establish the asymptotic properties of a portmanteau type test statistic for both autocorrelation of residuals and squared residuals,

$$\bar{B}_{T\hat{\theta}_T}^{(2m)}(s) = T\sum_{j=1}^{s}\bar{\lambda}_{T\hat{\theta}_T}(j)'\,\bar{\lambda}_{T\hat{\theta}_T}(j),$$

where s, $1 \leq s \leq m-k$, is a user chosen lag parameter, fixed in the asymptotics.

Theorem 3. *Under the regularity conditions of Theorem 1 and $H_0^{(m)}$, $\bar{B}_{T\hat{\theta}_T}^{(2m)}(s) \rightarrow_d \chi_{2s}^2$,* *while under H_{1T},*

$$\bar{B}_{T\hat{\theta}_T}^{(2m)}(s) \to_d \chi_{2s}^2\left(\sum_{j=1}^{s} \bar{\tau}_{\theta_0}^{(m-k)}(j)' \, \bar{\tau}_{\theta_0}^{(m-k)}(j)\right).$$

Here s can be chosen as small as 1 without problems to control asymptotic size in contrast with tests that need to smooth out the estimation effect in the residuals by means of choosing s growing with T, which typically leads to a reduction of power. Theorem 3 confirms this by showing that our joint portmanteau test statistic has power against nonparametric local alternatives converging to the null at the rate $T^{-1/2}$ as long as the projected drift is non-zero.

In the same way, individual or marginal tests for correlation in the residuals, such as

$$\bar{B}_{T\hat{\theta}_T,res}^{(2m)}(s) = T\sum_{j=1}^{s} \bar{\lambda}_{T\hat{\theta}_T}^{(1)}(j)^2,$$

or in their squares,

$$\bar{B}_{T\hat{\theta}_T,rsq}^{(2m)}(s) = T\sum_{j=1}^{s} \bar{\lambda}_{T\hat{\theta}_T}^{(2)}(j)^2,$$

can be developed and compared against χ_s^2 critical values to gain information on the potential source of the failure of specification when the joint test $\bar{B}_{T\hat{\theta}_T}^{(2m)}(s)$ rejects. Here, alternative projections and standardizations could be envisaged for each particular set of autocorrelation coefficients to isolate information from each particular moment, but, in principle, simultaneously accounting for estimation effect for both mean and variance parameterizations is simpler and could provide some advantage is terms of defining the transformations with a reduced loss of degrees of freedom, cf. Assumption A4 in the Appendix.

3.3 ARMA-GARCH Model

We consider in this subsection the ARMA(P,Q)-GARCH(p,q) model

$$
\begin{aligned}
Y_t &= \sum_{j=1}^{P} a_{0j} Y_{t-j} + e_t - \sum_{j=1}^{Q} b_{0j} e_{t-j}, \\
e_t &= h_t \varepsilon_t, \quad h_t^2 = \varpi_0 + \sum_{j=1}^{q} \alpha_{0j} e_{t-j}^2 + \sum_{j=1}^{p} \beta_{0j} h_{t-j}^2,
\end{aligned}
$$

with $E(\varepsilon_t|I_{t-1}) = 0$, $E(\varepsilon_t^2|I_{t-1}) = 1$, and show how to compute our new test statistics.

The parameter vector is denoted by $\vartheta = (a_1, \ldots, a_P, b_1, \ldots, b_Q)'$ for the conditional mean part of the model, $\nu = (\varpi, \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p)'$ for the conditional variance part of the model, $\theta = (\vartheta', \nu')'$. The true parameter vector is $\theta_0 = (\vartheta_0', \nu_0')'$. The usual identification conditions are assumed. In the following, for any generic function $g = g_\theta$ indexed by parameters $\theta \in \Theta_0$,

$$\dot{g}_\theta = \frac{\partial g_\theta}{\partial \theta'}.$$

We can write the model residuals as

$$\varepsilon_t(\vartheta, \nu) = \frac{e_t(\vartheta)}{h_{\theta t}} = \frac{\varphi_\vartheta(L) Y_t}{h_{\theta t}} = \frac{Y_t - \{1 - \varphi_\vartheta(L)\} Y_t}{h_{\theta t}},$$

where $\varphi_\vartheta(z) = A_\vartheta(z) B_\vartheta^{-1}(z)$, $A_\vartheta(z) = 1 - \sum_{j=1}^{P} a_j z^j$, $B_\vartheta(z) = 1 - \sum_{j=1}^{Q} b_j z^j$ and $h_{\theta t}^2 = \varpi + \sum_{j=1}^{q} \alpha_j e_{t-j}^2(\vartheta) + \sum_{j=1}^{p} \beta_j h_{\theta t-j}^2$. Then

$$
\begin{aligned}
\dot{\varepsilon}_{\vartheta t}(\vartheta, \nu) &= \frac{\dot{\varphi}_\vartheta(L) Y_t}{h_{\theta t}} - \frac{\dot{h}_{\vartheta t}}{h_{\theta t}} \frac{\varphi_\vartheta(L) Y_t}{h_{\theta t}} \\
&= \left\{ \frac{\dot{\varphi}_\vartheta(L)}{\varphi_\vartheta(L)} - \frac{\dot{h}_{\vartheta t}}{h_{\theta t}} \right\} \frac{e_t(\vartheta)}{h_{\theta t}} \\
\dot{\varepsilon}_{\nu t}(\vartheta, \nu) &= -\frac{\dot{h}_{\nu t}}{h_{\theta t}} \frac{\varphi_\theta(L) Y_t}{h_{\theta t}} \\
&= -\frac{\dot{h}_{\nu t}}{h_{\theta t}} \frac{e_t(\vartheta)}{h_{\theta t}}.
\end{aligned}
$$

If a semistrong GARCH model is assumed, it is easy to obtain that under H_0

$$\hat{\rho}^{(m)}_{T\hat{\theta}_T}(j) = \hat{\rho}^{(m)}_{T\theta_0}(j) - \left(\hat{\vartheta}_T - \vartheta_0\right)' E\left[\frac{\dot{\varphi}_{\vartheta_0}(L)}{\varphi_{\vartheta_0}(L)}\varepsilon_{\theta_0 t}\varepsilon_{\theta_0 t-j}\right] + o_p\left(T^{-1/2}\right),$$

$$\hat{\delta}^{(m)}_{T\hat{\theta}_T}(j) = \hat{\delta}^{(m)}_{T\theta_0}(j) - 2\left(\hat{\theta}_T - \theta_0\right)' E\left[\frac{\dot{h}_{\theta_0 t}}{h_{\theta_0 t}}\varepsilon^2_{\theta_0 t}\left(\varepsilon^2_{\theta_0 t-j} - 1\right)\right] + o_p\left(T^{-1/2}\right).$$

Example: ARMA$(1,1)$-GARCH$(1,1)$ model $\theta_0 = (\vartheta_0', \nu_0')'$, where $\vartheta_0 = (a_{01}, b_{01})'$, $\nu_0 = (\varpi_0, \alpha_{01}, \beta_{01})'$

$$\begin{aligned}
\frac{\dot{h}_{\vartheta t}}{h_{\theta t}} &= \frac{1}{2}h_{\theta t}^{-2}\left(2\alpha\sum_{j=0}^{\infty}\beta_1^j e_{\vartheta t-1-j}\dot{e}_{\vartheta t-1-j}\right) \\
&= \alpha_1\sum_{j=0}^{\infty}\beta_1^j\frac{e_{\vartheta t-1-j}}{h_{\theta t}}\frac{\dot{e}_{\vartheta t-1-j}}{h_{\theta t}} \\
&= \alpha_1\frac{e_{\vartheta t-1}}{h_{\theta t}}\frac{\dot{e}_{\vartheta t-1}}{h_{\theta t}} \quad \text{if ARCH(1).}
\end{aligned}$$

and

$$\begin{aligned}
\frac{\dot{h}_{\nu t}}{h_{\theta t}} &= \frac{1}{2}\begin{pmatrix} h_{\theta t}^{-2}/(1-\beta_1) \\ \sum_{j=0}^{\infty}\beta_1^j\frac{e^2_{\vartheta t-1-j}}{h^2_{\theta t}} \\ -h_{\theta t}^{-2}/(1-\beta_1)^2 + \alpha\sum_{j=1}^{\infty}j\beta_1^{j-1}\frac{e^2_{\vartheta t-1-j}}{h^2_{\theta t}}\end{pmatrix} \\
&= \frac{1}{2}\begin{pmatrix} h_{\theta t}^{-2} \\ \frac{e^2_{\vartheta t-1}}{h^2_{\theta t}}\end{pmatrix} \quad \text{if ARCH(1).}
\end{aligned}$$

4 Monte Carlo Simulations

We carry out some Monte Carlo simulations to compare the finite-sample performance of the new joint test statistics with those of the Wong-Ling test and different versions of the marginal Portmanteau tests. The null model is the AR(1)-ARCH(1) model,

$$\begin{aligned} Y_t &= 0.5Y_{t-1} + e_t, \\ e_t &= h_t \varepsilon_t, \quad h_t^2 = 0.1 + 0.4e_{t-1}^2. \end{aligned}$$

Parameters are estimated by Quasi Maximum Likelihood. Nominal size of all tests is 5%. The sample sizes are $T = 200$ and $T = 500$. We consider $\varepsilon_t \sim$i.i.d. $\mathrm{N}(0,1)$ or i.i.d. standardized student t distribution with 10 degrees of freedom, and we also consider the semistrong version of the AR(1)-ARCH(1) model,[1] which only satisfies (2). Replications are $10,000$ in each experiment. Given the fact that the ε_t considered here follow a symmetric distribution, we use different estimates of $\Omega_{\theta_0}^{(2m)}$ that exploit that is block diagonal and has first block equal to identity for this specification,

$$\hat{\Omega}_{T\hat{\theta}_T}^{(2m)} = \begin{pmatrix} I^{(m)} & 0 \\ 0 & \hat{\Sigma}_{T\hat{\theta}_T}^{(m)} \end{pmatrix}.$$

We use four different estimators of $\Sigma_{\theta_0}^{(m)}$. The first one sets $\hat{\Sigma}_{T\hat{\theta}_T}^{(m)} = I^{(m)}$, which exploits the asymptotic i.i.d. property of the sample autocorrelations of the centered square errors $\varepsilon_{\theta_0 t}^2 - 1$ for the strong version of the ARCH model. The second one sets

$$\hat{\Sigma}_{T\hat{\theta}_T}^{(m)} = \left[\frac{\hat{\sigma}_{T\hat{\theta}_T}^{(i,j)}}{\hat{\eta}_{T\hat{\theta}_T}(0)^2} \right]_{i,j=1}^{m}$$

with $\hat{\sigma}_{\theta}^{(i,j)} = T^{-1} \sum_{t=1+\max(i,j)}^{T} \left[\left(\varepsilon_{\theta t}^2 - 1\right)^2 \left(\varepsilon_{\theta t-i}^2 - 1\right) \left(\varepsilon_{\theta t-j}^2 - 1\right) \right]$, which only exploits that the centered squared errors $\varepsilon_{\theta_0 t}^2 - 1$ are a martingale difference sequence for both ARCH versions. The last two estimators of $\Sigma_{\theta_0}^{(m)}$ consider the HAC robust estimation of the asymptotic covariance matrix of $\hat{\delta}_{T\hat{\theta}_T}^{(m)}$ by the spectral approach and by the VAR approximation, so do not exploit the ARCH structure. For the spectral approach, we

[1] To obtain the semistrong ARCH model, first generate the ARCH (1) model $h_t = 0.1/(1+\sqrt{0.4}) + \sqrt{0.4}e_{t-1}^2$ with sample size $2T$, then choose the even-number observations. It could be shown that these observations follow a semistrong ARCH (1) model $h_t = 0.1 + 0.4e_t^2$, see Franq and Zakoian (2010) Chapter 4.1.1 for more details.

use Bartlett's kernel as in Newey and West (1987). We set the lag by $b = floor(1.2T^{1/3})$ in both estimators and choose $m = m_s = s + 3$ for both sample sizes, where 3 is the minimum number of the degrees of freedom needed to conduct the residuals autocorrelation projections given the dimension of $\theta = (a_1, \varpi_0, \alpha_1)'$. As in Delgado and Velasco (2011), we find that using a larger m than needed when $\Omega_{\theta_0}^{(2m)}$ is estimated can affect the finite sample properties of tests.

We compare our new recursive Box-Pierce type statistic with Wong-Ling statistic. In Wong and Ling (2005), the properties of the quasi-maximun likelihood estimator are used to handle the estimation effect of the joint Portmanteau test. To be specific, Wong and Ling (2005) test statistic has the form

$$\hat{Q}_{2m} = n \begin{pmatrix} \hat{\rho}_{T\hat{\theta}_T}^{(m)} \\ \hat{\delta}_{T\hat{\theta}_T}^{(m)} \end{pmatrix}' \hat{\Psi}^{-1} \begin{pmatrix} \hat{\rho}_{T\hat{\theta}_T}^{(m)} \\ \hat{\delta}_{T\hat{\theta}_T}^{(m)} \end{pmatrix},$$

where $\hat{\Psi}$ is the estimate of the asymptotic variance of $\sqrt{T}\left(\hat{\rho}_{T\hat{\theta}_T}^{(m)\prime}, \hat{\delta}_{T\hat{\theta}_T}^{(m)\prime}\right)'$ derived from that of $\sqrt{T}\left(\hat{\rho}_{T\theta_0}^{(m)\prime}, \hat{\delta}_{T\theta_0}^{(m)\prime}\right)'$ and $\sqrt{T}\left(\hat{\theta}_T - \theta_0\right)$, so that asymptotically $\hat{Q}_{2m}$ follows a χ_{2m}^2 distribution. When $E\left[\varepsilon_t^3\right] = 0, E\left[\varepsilon_t^4\right] = 3$, and $\varepsilon_t \sim i.i.d.\,(0, \sigma^2)$, their test statistic returns to the mixed statistic of Box–Pierce and Li–Mak test statistics.

Figures 1 and 2 report the simulated size. We can observe that when the innovations follow an i.i.d. N(0,1) distribution, Wong-Ling statistic has good size properties. Our new recursive Box-Pierce statistics with $\hat{\Omega}_{T\hat{\theta}_T}^{(2m)} = I^{(2m)}$ underreject for $T = 200$, but have good size levels for $T = 500$. On the other hand, the new recursive Box-Pierce statistics with $\hat{\Omega}_{T\hat{\theta}_T}^{(2m)} = diag\left(I^{(m)}, \hat{\Sigma}_{T\hat{\theta}_T}^{(m)}\right)$ overreject for $T = 200$, but have good size levels very close to nominal size for $T = 500$. When the innovations follow an i.i.d. standardized Student's t distribution with 10 degrees of freedom, Wong-Ling statistics heavily overreject for both $T = 200$ and $T = 500$. Similar results are obtained for Wong-Ling statistics when the innovations follow a semistrong ARCH model with conditional normal distribution. However the new recursive Box-Pierce statistics have good size properties in both cases. When it comes to the case that the innovations follow the semistrong ARCH model with standardized Student's t distribution, Wong-Ling statistics overreject for $T = 200$, but become closer to the nominal size for $T = 500$. When $T = 200$, the VAR approximation and the spectral approach have size distortions for large s, but when $T = 500$, test statistics based on both the VAR approximation and the spectral approach have nice size properties overall. In the case of i.i.d. errors, the unnecessary robust estimation of asymptotic covariance matrices makes the simulated size worse with the VAR approximation providing better finite sample size than the spectral approach. When it comes to semistrong ARCH cases, both the spectral approach and the VAR approximation work fine.

To study the power properties of the new tests, we keep fitting the same AR(1)-ARCH(1) model but consider the following alternative models to generate data,

$$\begin{aligned}
\mathbf{M1}: \quad Y_t &= 0.5Y_{t-1} + 0.2Y_{t-2} + e_t,\ e_t = h_t\varepsilon_t,\ h_t^2 = 0.1 + 0.4e_{t-1}^2, \\
\mathbf{M2}: \quad Y_t &= 0.5Y_{t-1} + e_t,\ e_t = h_t\varepsilon_t,\ h_t^2 = 0.1 + 0.4e_{t-1}^2 + 0.2e_{t-2}^2, \\
\mathbf{M3}: \quad Y_t &= 0.5Y_{t-1} + 0.2Y_{t-2} + e_t,\ e_t = h_t\varepsilon_t,\ h_t^2 = 0.1 + 0.4e_{t-1}^2 + 0.2e_{t-2}^2, \\
\mathbf{M4}: \quad Y_t &= 0.5Y_{t-1} + 0.2Y_{t-2} + e_t,\ e_t = h_t\varepsilon_t,\ h_t^2 = 0.1 + 0.4e_{t-1}^2 + 0.5h_{t-1}^2,
\end{aligned}$$

where $\varepsilon_t \sim$i.i.d. $N(0,1)$. The first alternative is misspecified in conditional mean part; the second in the conditional variance part, and the third and fourth alternatives are misspecified in both the conditional mean and conditional variance functions. We plot the percentage of rejections under these four alternative hypotheses in Figures 3-6 respectively. It is confirmed that the joint portmanteau tests can detect the misspecifications in conditional mean and/or conditional variance. The power of the new transformed Box-Pierce statistics is comparable to Wong-Ling test, especially when $\hat{\Omega}^{(2m)}_{T\hat{\theta}_T} = I^{(2m)}$. However robust estimation with $\hat{\Omega}^{(2m)}_{T\hat{\theta}_T} = diag\left(I^{(m)}, \hat{\Sigma}^{(m)}_{T\hat{\theta}_T}\right)$ introduces some costs in terms of power.

In Tables 1-8 we present the simulation results for the power of marginal Portmanteau tests against the four previous alternatives and both sample sizes, and report also the result for the recursive Portmanteau joint test with $\hat{\Sigma}^{(m)}_{T\hat{\theta}_T} = I^{(m)}$. The aim is to

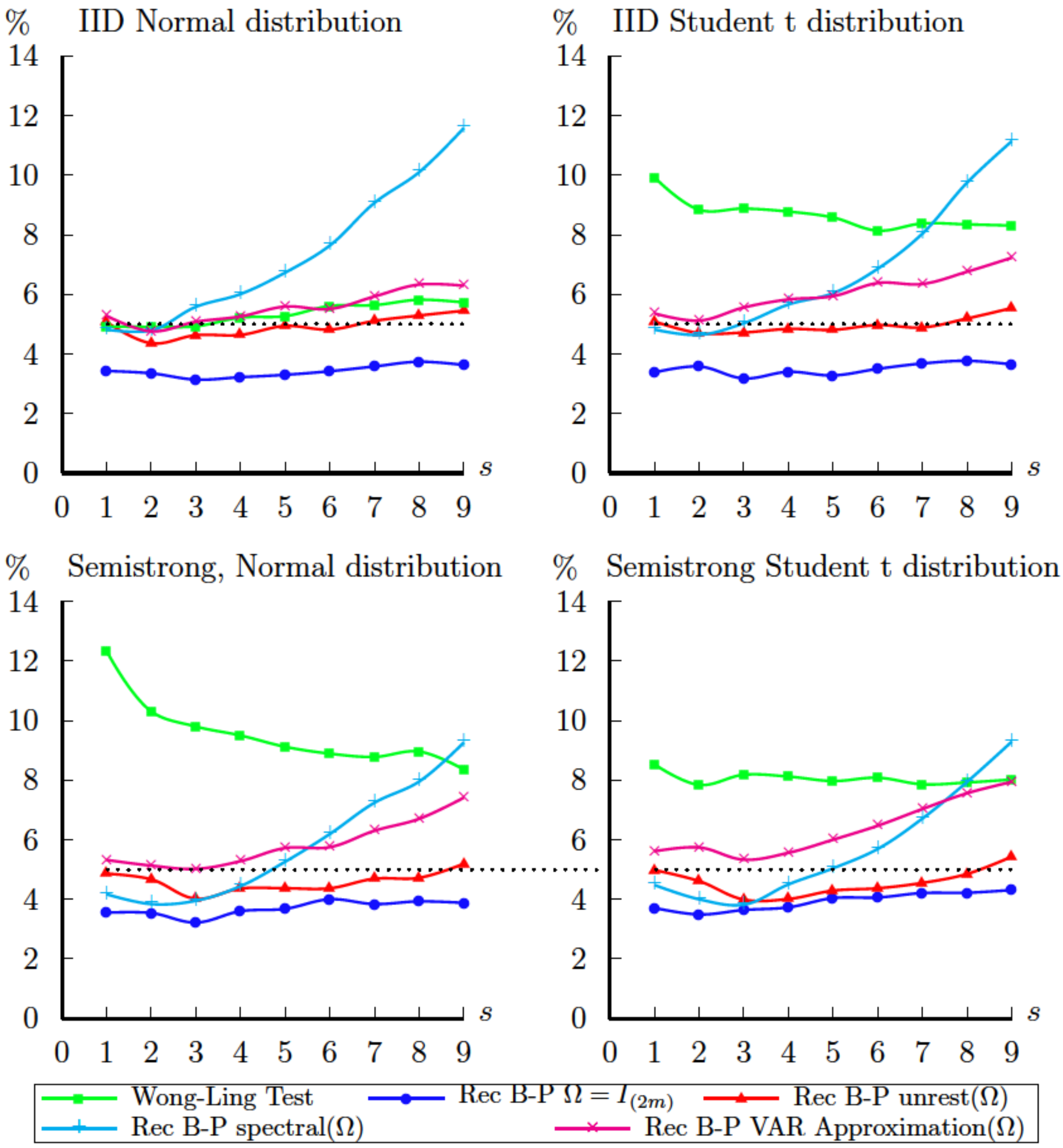


Figure 1: Size simulation $T = 200$. Percentage of rejections of Portmanteau tests in terms of the lag s. Nominal level is 5%. Wong-Ling tests compare with a χ^2_{2s} critical value, Rec B-P are tests $\bar{B}^{(2m)}_{T\hat{\theta}_T}(s)$ based on recursive projected autocorrelations compared to χ^2_{2s}, $m = s + 3$. Rec B-P spectral(Ω) applies Newey-West (1987) approach, Rec B-P VAR Approximation(Ω) applies VAR approximation of Den Haan and Levin (1997). Models are AR(1)-ARCH(1) with i.i.d. normal distribution, i.i.d. Student's t distribution with 10 degrees of freedom, semistrong ARCH with normal distribution and semistrong ARCH with Student's t distribution with 10 degrees of freedom.

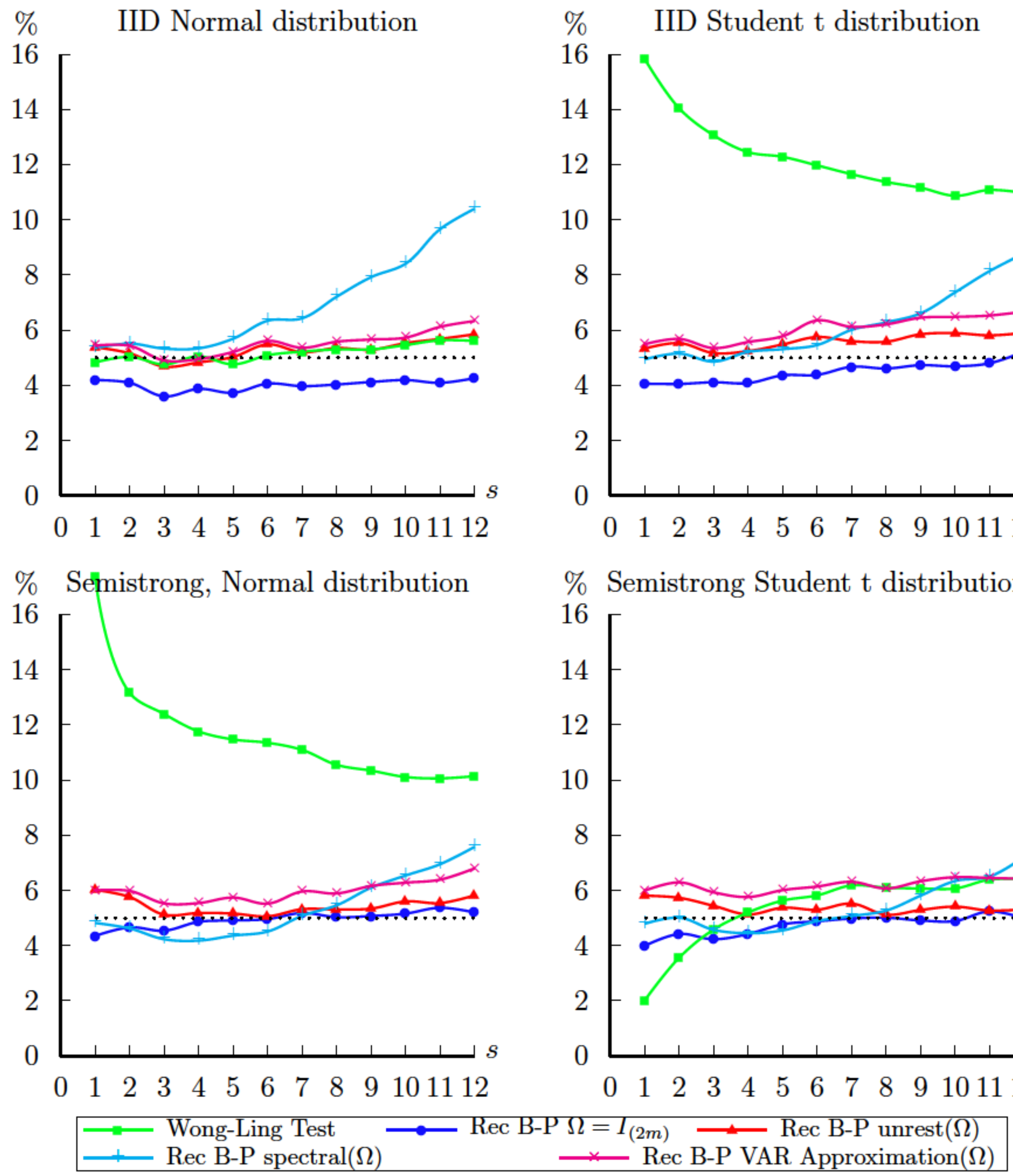


Figure 2: Size Simulation $T = 500$. Percentage of rejections of Portmanteau tests in terms of the lag s. Test statistics as described in Figure 1.

s	1	2	3	4	5	6	7	8	9
$\bar{B}^{(2m)}_{T\hat{\theta}_T}(s)$	49.1	41.4	32.9	28.7	25.8	23.9	22.8	20.1	18.5
$\bar{B}^{(2m)}_{T\hat{\theta}_T,res}(s)$	61.2	52.5	46.6	43.0	38.5	37.1	34.1	33.1	31.8
BPL (s)	57.8	54.4	48.0	42.8	40.0	38.0	35.1	33.6	32.7
$\bar{B}^{(2m)}_{T\hat{\theta}_T,rsq}(s)$	4.2	3.4	3.9	3.7	3.3	3.5	3.9	3.5	3.1
Li-Mak (s)	5.9	5.4	5.7	5.9	5.2	5.5	5.9	6.1	6.1

Table 1: Power simulation Model M1, $T = 200$. Percentage of rejections of marginal Portmanteau tests for lag s against alternative $M1$, AR(2)-ARCH(1) model, $T = 200$. The null is AR(1)-ARCH(1) model. Nominal level is 5%. $\bar{B}^{(2m)}_{T\hat{\theta}_T}(s)$ is the recursive Portmanteau joint test with $\hat{\Sigma}^{(m)}_{T\hat{\theta}_T} = I^{(m)}$. $\bar{B}^{(2m)}_{T\hat{\theta}_T,res}(s)$ is the recursive Portmanteau test of conditional mean, $\bar{B}^{(2m)}_{T\hat{\theta}_T,rsq}(s)$ is the recursive Portmanteau test of conditional variance, with $\hat{\Sigma}^{(m)}_{T\hat{\theta}_T} = I^{(m)}$. BPL (s) represents LM type Portmanteau test of conditional mean, Li-Mak (s) represents LM type Portmanteau test of conditional variance.

investigate whether joint testing can provide good performance compared to specific tests directed to the true alternative in one of the first two conditional moments. We consider the recursive Portmanteau test of conditional mean, $\bar{B}^{(2m)}_{T\hat{\theta}_T,res}(s)$ and the recursive Portmanteau test of conditional variance, $\bar{B}^{(2m)}_{T\hat{\theta}_T,rsq}(s)$, with $\hat{\Sigma}^{(m)}_{T\hat{\theta}_T} = I^{(m)}$. For comparison we also present the results for the LM type Portmanteau test of conditional mean, denoted as BPL, and for the LM type Portmanteau test of conditional variance, denoted as Li-Mak.

We can observe in Tables 1-4 that the joint test is competitive against alternatives that misspecify only one the two conditional moments compared to the appropriate marginal tests. Marginal tests for the conditional mean are not very robust against misspecification in the variance as they use restricted estimates of the asymptotic variance, cf. Tables 3 and 4. However, when testing against simultaneous misspecification of the conditional mean and variance in Tables 5-8, joint tests exploiting both residuals and squared residuals autocorrelations, appear noticeably more powerful that any of the marginal tests.

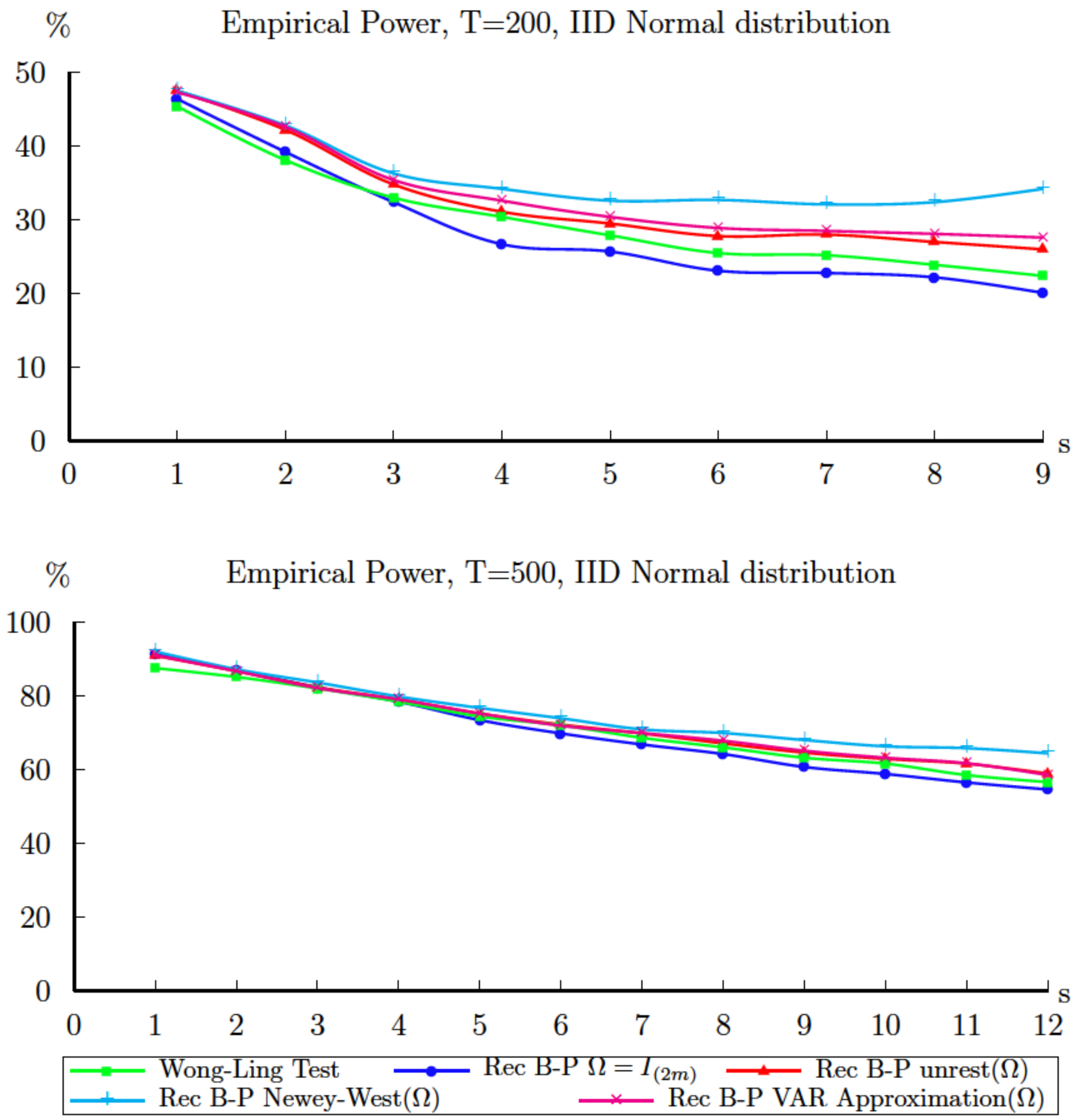


Figure 3: Power Simulation Model M1. Percentage of rejections of joint Portmanteau tests in terms of the lag s. The null is AR(1)-ARCH(1) model. The alternative is AR(2)-ARCH(1), $M1$. $T = 200$ and $T = 500$. Nominal level is 5%. Wong-Ling tests compare with a χ^2_{2s} critical value, Rec B-P are tests $\bar{B}^{(2m)}_{T\hat{\theta}_T}(s)$ based on recursive projected autocorrelations compared to χ^2_{2s}, $m = s + 3$.

s	1	2	3	4	5	6	7	8	9	10	11	12
$\bar{B}^{(2m)}_{T\hat{\theta}_T}(s)$	93.7	88.9	83.5	78.4	74.1	68.4	65.8	62.8	60.1	57.9	56.1	55.6
$\bar{B}^{(2m)}_{T\hat{\theta}_T,res}(s)$	96.6	94.7	91.1	89.8	86.3	83.8	81.2	78.5	77.1	76.6	74.5	73.3
BPL (s)	94.1	93.0	90.1	89.7	86.4	84.1	81.9	79.0	77.5	76.5	75.6	73.3
$\bar{B}^{(2m)}_{T\hat{\theta}_T,rsq}(s)$	5.1	5.2	5.5	5.3	5.0	5.6	5.0	4.7	4.3	5.5	5.0	4.7
Li-Mak (s)	4.9	6.9	7.1	7.4	7.0	7.2	7.4	7.8	8.0	7.7	6.6	6.4

Table 2: Power simulation Model M1, $T = 500$. Percentage of rejections of marginal Portmanteau tests against alternative $M1$, AR(2)-ARCH(1) model. Test statistics as described in Table 1.

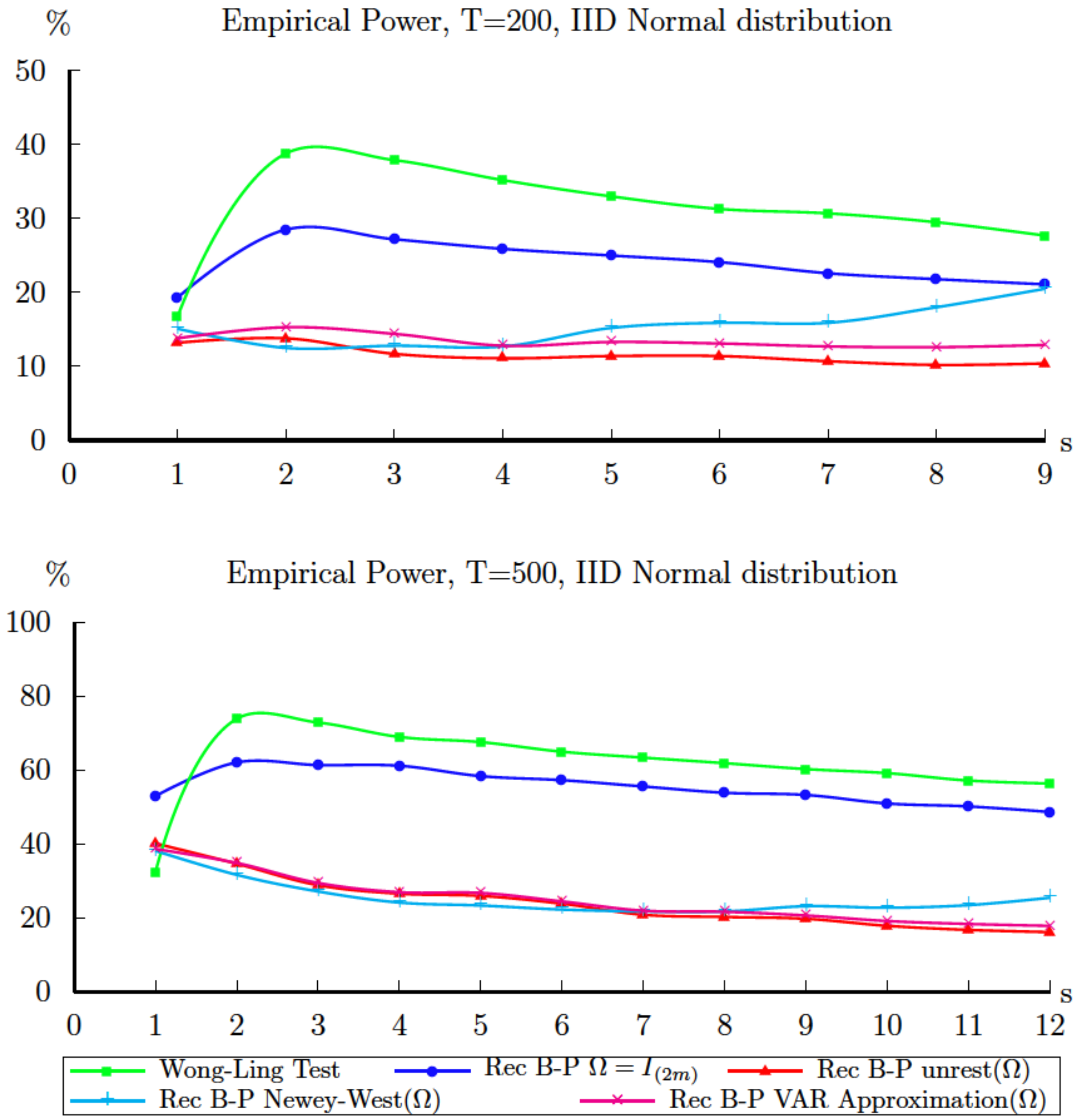
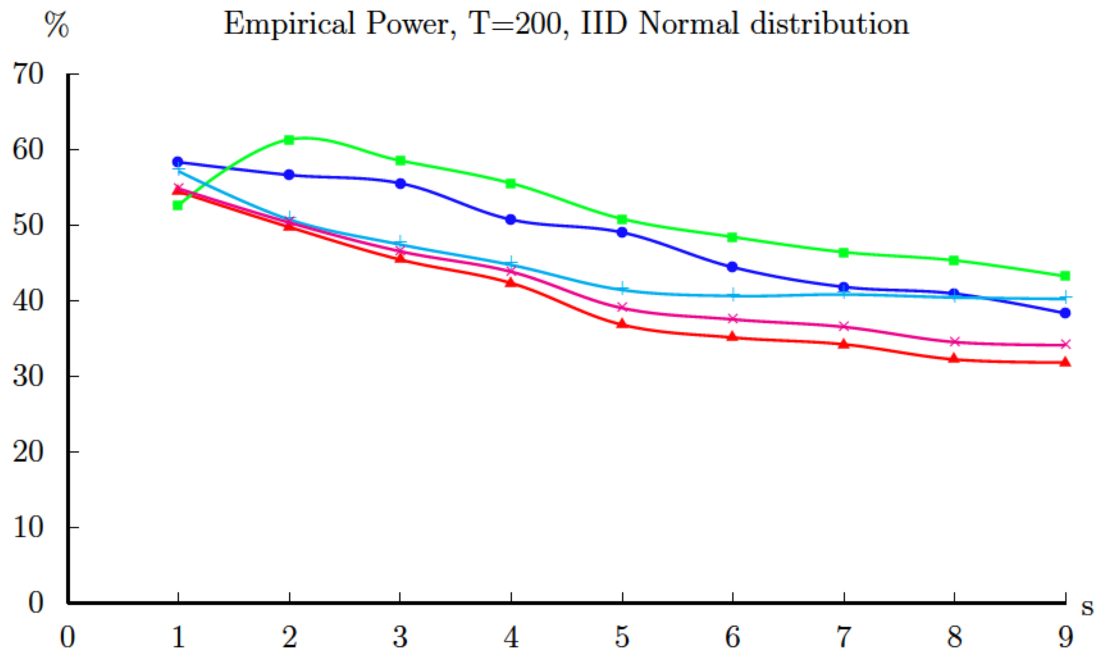


Figure 4: Power Simulation Model M2. Percentage of rejections of joint Portmanteau tests in terms of the lag s. The null is AR(1)-ARCH(1) model. The alternative is AR(1)-ARCH(2), $M2$. $T = 200$ and $T = 500$. Tests statistics as described in Figure 3.

s	1	2	3	4	5	6	7	8	9
$\bar{B}^{(2m)}_{T\hat{\theta}_T}(s)$	19.5	28.3	25.3	25.3	24.6	22.6	22.0	19.8	18.9
$\bar{B}^{(2m)}_{T\hat{\theta}_T,res}(s)$	7.0	7.0	7.3	7.4	7.7	7.9	7.7	6.9	7.2
BPL (s)	8.9	10.8	11.3	10.3	11.0	10.0	10.1	10.3	11.2
$\bar{B}^{(2m)}_{T\hat{\theta}_T,rsq}(s)$	21.5	32.1	30.4	29.6	28.5	24.9	24.5	22.9	22.4
Li-Mak (s)	18.1	39.9	38.8	36.6	34.1	32.4	30.8	29.5	28.5

Table 3: Power simulation Model M2, $T = 200$. Percentage of rejections of marginal Portmanteau tests against alternative $M2$, AR(1)-ARCH(2) model. Test statistics as described in Table 1.



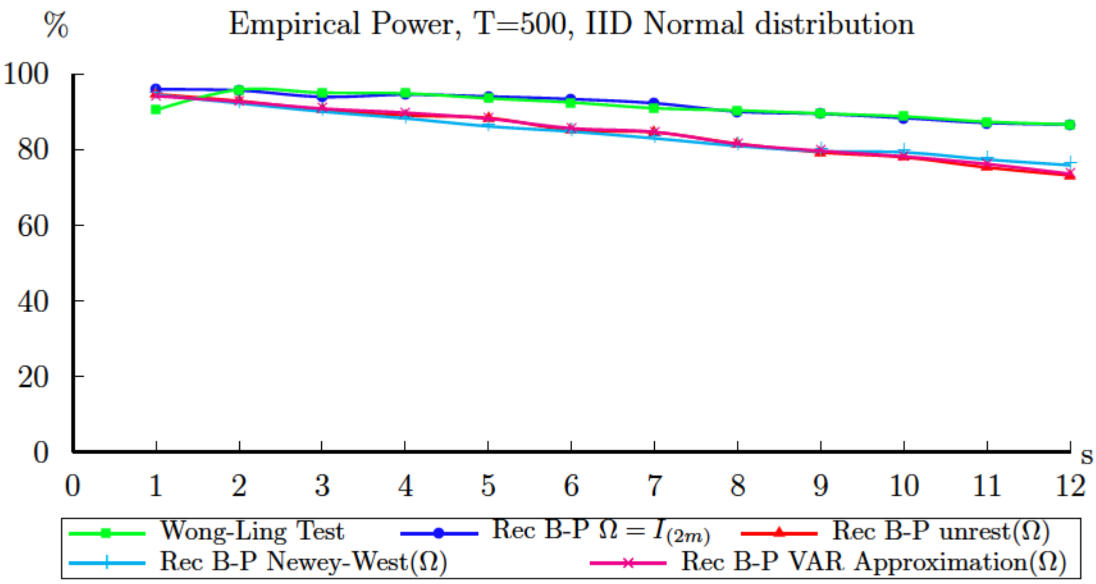


Figure 5: Power simulation Model M3. Percentage of rejections of joint Portmanteau tests in terms of the lag s. The null is AR(1)-ARCH(1) model. The alternative is AR(2)-ARCH(2), $M3$. $T = 200$ and $T = 500$. Tests statistics as described in Figure 3.

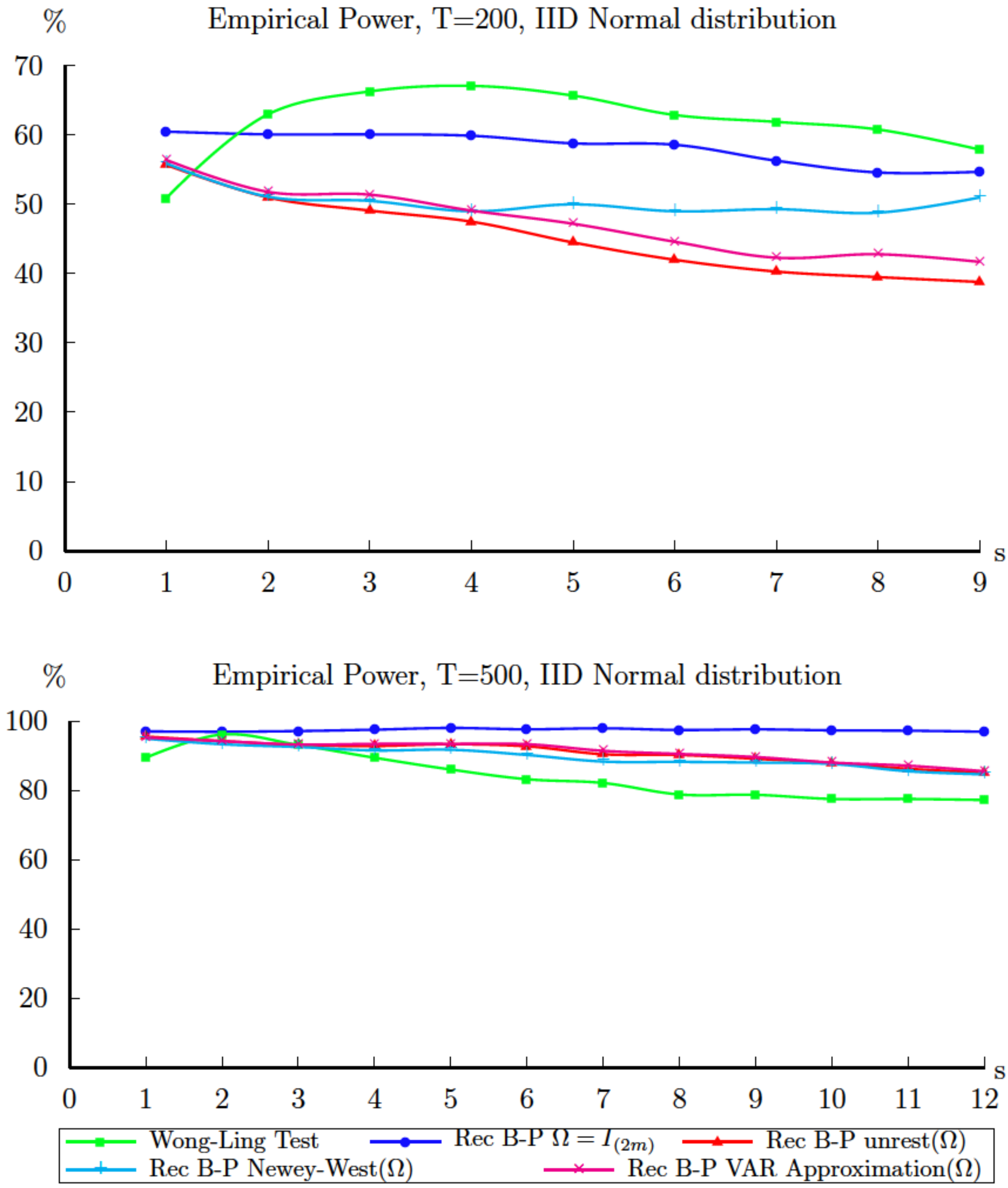


Figure 6: Power simulation Model M4. Percentage of rejections of joint Portmanteau tests in terms of the lag s. The null is AR(1)-ARCH(1) model. The alternative is AR(2)-GARCH(1,1), $M4$. $T = 200$ and $T = 500$. Tests statistics as described in Figure 3.

s	1	2	3	4	5	6	7	8	9	10	11	12
$\bar{B}^{(2m)}_{T\hat{\theta}_T}(s)$	53.3	61.0	62.3	61.5	60.0	58.2	56.6	53.7	51.3	50.1	48.3	46.7
$\bar{B}^{(2m)}_{T\hat{\theta}_T,res}(s)$	8.4	7.5	8.9	8.5	8.9	8.1	7.4	9.1	8.5	8.7	7.7	7.0
BPL (s)	12.1	14.2	15.1	14.9	14.7	14.1	14.2	13.5	13.3	12.6	12.5	12.0
$\bar{B}^{(2m)}_{T\hat{\theta}_T,rsq}(s)$	59.6	69.9	70.4	69.8	67.9	65.4	64.9	63.0	62.4	60.6	58.2	57.7
Li-Mak (s)	37.3	80.4	80.5	77.6	76.0	73.9	71.8	70.1	68.9	67.6	65.3	63.6

Table 4: Power simulation Model M2, $T = 500$. Percentage of rejections of marginal Portmanteau tests against alternative $M2$, AR(1)-ARCH(2) model. Test statistics as described in Table 1.

s	1	2	3	4	5	6	7	8	9
$\bar{B}^{(2m)}_{T\hat{\theta}_T}(s)$	57.1	57.2	53.7	50.4	47.4	44.8	42.5	39.9	38.2
$\bar{B}^{(2m)}_{T\hat{\theta}_T,res}(s)$	57.7	51.3	46.3	42.1	39.2	37.2	35.8	35.8	33.6
BPL (s)	53.2	52.6	48.0	45.0	42.1	40.1	38.5	37.3	35.8
$\bar{B}^{(2m)}_{T\hat{\theta}_T,rsq}(s)$	20.8	29.2	29.0	27.9	26.6	25.4	24.8	23.7	22.6
Li-Mak (s)	16.4	37.7	37.4	35.5	33.7	31.8	30.3	29.2	28.5

Table 5: Power simulation Model M3, $T = 200$. Percentage of rejections of marginal Portmanteau tests against alternative $M3$, AR(2)-ARCH(2) model. Test statistics as described in Table 1.

s	1	2	3	4	5	6	7	8	9	10	11	12
$\bar{B}^{(2m)}_{T\hat{\theta}_T}(s)$	97.4	96.7	95.8	94.9	93.7	91.7	90.2	88.5	88.2	86.2	84.4	83.9
$\bar{B}^{(2m)}_{T\hat{\theta}_T,res}(s)$	95.0	93.3	89.8	88.0	86.1	83.9	80.9	78.8	77.4	74.9	73.6	71.6
BPL (s)	89.9	92.8	89.7	88.7	87.4	86.0	83.7	80.7	80.1	78.0	76.3	75.0
$\bar{B}^{(2m)}_{T\hat{\theta}_T,rsq}(s)$	55.6	62.3	63.3	64.2	62.9	60.7	59.9	57.6	57.0	56.3	54.0	53.5
Li-Mak (s)	32.9	75.0	75.5	72.9	70.7	69.3	66.5	64.6	62.6	61.5	59.9	59.0

Table 6: Power simulation Model M3, $T = 500$. Percentage of rejections of marginal Portmanteau tests against alternative $M3$, AR(2)-ARCH(2) model. Test statistics as described in Table 1.

s	1	2	3	4	5	6	7	8	9
$\bar{B}^{(2m)}_{T\hat{\theta}_T}(s)$	60.3	61.2	62.6	62.6	62.7	61.5	60.9	59.5	58.3
$\bar{B}^{(2m)}_{T\hat{\theta}_T,res}(s)$	54.8	50.2	46.5	44.1	42.2	40.3	39.2	37.9	36.3
BPL (s)	50.9	54.0	51.3	49.0	47.7	45.9	44.4	43.2	41.9
$\bar{B}^{(2m)}_{T\hat{\theta}_T,rsq}(s)$	30.0	35.6	41.0	43.5	44.8	45.4	45.8	44.9	44.6
Li-Mak (s)	21.7	37.3	48.2	52.0	52.8	51.6	48.9	48.1	46.3

Table 7: Power simulation Model M4, $T = 200$. Percentage of rejections of marginal Portmanteau tests against alternative $M4$, AR(2)-GARCH(1,1) model. Test statistics as described in Table 1.

s	1	2	3	4	5	6	7	8	9	10	11	12
$\bar{B}^{(2m)}_{T\hat{\theta}_T}(s)$	95.6	96.2	96.6	96.8	97.2	97.2	97.2	97.3	97.1	96.9	96.7	96.5
$\bar{B}^{(2m)}_{T\hat{\theta}_T,res}(s)$	92.4	89.8	87.5	85.3	83.7	82.3	80.8	79.5	77.8	76.5	75.4	74.1
BPL (s)	81.6	89.2	88.8	88.0	86.9	86.0	84.9	83.3	81.8	80.7	79.3	78.4
$\bar{B}^{(2m)}_{T\hat{\theta}_T,rsq}(s)$	60.9	70.5	77.6	81.7	84.1	85.8	86.9	87.6	87.6	87.6	87.6	87.4
Li-Mak (s)	44.6	72.0	83.6	82.2	79.1	75.6	73.1	71.0	69.5	68.6	68.0	67.6

Table 8: Power simulation Model M4, $T = 500$. Percentage of rejections of marginal Portmanteau tests against alternative $M4$, AR(2)-GARCH(1,1) model. Test statistics as described in Table 1.

Appendix

In this appendix we present the sufficient assumptions for the proofs of our results. First we introduce some notations. For any generic function g_θ indexed by parameters $\theta \in \Theta_0$,

$$\ddot{g}_\theta = \frac{\partial^2 g_\theta}{\partial\theta\partial\theta'}$$

Assumption A1. $(Y_t, \varepsilon_{\theta_0 t})'$ is strictly stationary, $E\left[\varepsilon_{\theta_0 t}\right] = 0$, $E\left[\varepsilon^2_{\theta_0 t}\right] = 1$, $E\left(\varepsilon^{8+4\kappa}_{\theta_0 t}\right) < \infty$ for some $\kappa > 0$. $(Y_t, \varepsilon_{\theta_0 t})'$ is strong mixing with coefficient α_j satisfying $\sum_{j=1}^{\infty} \alpha_j^{\kappa/(2+\kappa)} < \infty$, where

$$\alpha_j = \sup_{A,B} |\Pr(AB) - \Pr(A)\Pr(B)|,$$

where A and B vary over events in the σ fields generated by $\left\{(Y_t, \varepsilon_{\theta_0 t})', t \leq 0\right\}$, and $\left\{(Y_{t+j}, \varepsilon_{\theta_0 t+j})', t \geq j\right\}$.

Assumption A2. The functions $f_t(\cdot) = f(I_{t-1}, \cdot)$ and $h_t(\cdot) = h(I_{t-1}, \cdot)$ are twice continuously differentiable with respect to $\theta \in \Theta_0$ a.s., with $E\left\|\varepsilon_{\theta_0 t}\frac{\dot{h}_t(\theta_0)}{h_t(\theta_0)}\right\|^{8+4\kappa} + E\left\|\frac{\dot{f}_t(\theta_0)}{h_t(\theta_0)}\right\|^{8+4\kappa} < \infty$ and $\varepsilon_{\theta_0 t}\frac{\dot{h}_t(\theta_0)}{h_t(\theta_0)}$ and $\frac{\dot{f}_t(\theta_0)}{h_t(\theta_0)}$ are mixing for the same $\kappa > 0$ as in Assumption A1.

Assumption A3. Let Θ_0 be a small convex neighborhood of θ_0, then

$$E\sup_{\theta\in\Theta_0}\left\|\varepsilon_{\theta t}\frac{\dot{h}_t(\theta)'\dot{h}_t(\theta)}{h_t^2(\theta)}\right\| + E\sup_{\theta\in\Theta_0}\left\|\varepsilon_{\theta t}\frac{\ddot{h}_t(\theta)}{h_t(\theta)}\right\| + E\sup_{\theta\in\Theta_0}\left\|\frac{\dot{f}_t(\theta)'\dot{h}_t(\theta)}{h_t^2(\theta)}\right\|$$
$$+ E\sup_{\theta\in\Theta_0}\left\|\frac{\ddot{f}_t(\theta)}{h_t(\theta)}\right\| + E\sup_{\theta\in\Theta_0}\left\|\frac{f_t(\theta)}{h_t(\theta)}\frac{\dot{h}_t(\theta)'\dot{h}_t(\theta)}{h_t^2(\theta)}\right\| + E\sup_{\theta\in\Theta_0}\left\|\frac{f_t(\theta)}{h_t(\theta)}\frac{\ddot{h}_t(\theta)}{h_t(\theta)}\right\| < \infty.$$

Assumption A4. For $m > k$,

$$\sum_{j=m-k+1}^{m} \tilde{\Lambda}_{\theta_0}(j)' \tilde{\Lambda}_{\theta_0}(j)$$

is positive definite.

Assumption A1 is about the DGP, where we further assume a mixing condition to justify a central limit theorem for the autocovariances of errors and their squares, apart from the martingale properties holding under H_0. Assumptions A2-A3 are standard conditions on the smoothness of f and h, see Escanciano (2008) for similar assumptions, though we also assume a natural mixing condition on the derivatives of the model residuals in view of A1 to bound the asymptotic variance of the derivatives of sample autocovariances of errors and their squares. Moments conditions on f_t and h_t hold under general moment conditions on the errors from A1 and usual identification conditions for ARMA-GARCH models due to the normalization by $h(I_{t-1}, \theta)$. Assumption A4 is an identification condition similar to Delgado and Velasco (2011) to guarantee the feasibility of the projection in the limit if enough extra sample autocorrelations are considered.

Proof of Proposition 1: Under $H_0^{(m)}$, we need to show that

$$\begin{aligned}
\sqrt{T}\hat{\rho}^{(m)}_{T\hat{\theta}_T} &= \sqrt{T}\hat{\rho}^{(m)}_{T\theta_0} + \nabla\rho^{(m)}_{\theta_0}\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) + o_p(1) \\
\sqrt{T}\hat{\delta}^{(m)}_{T\hat{\theta}_T} &= \sqrt{T}\hat{\delta}^{(m)}_{T\theta_0} + \nabla\delta^{(m)}_{\theta_0}\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) + o_p(1),
\end{aligned}$$

where $\nabla\rho^{(m)}_{\theta_0} = p\lim \frac{\partial}{\partial\theta'}\hat{\rho}^{(m)}_{T\theta_0}$ and $\nabla\delta^{(m)}_{\theta_0} = p\lim \frac{\partial}{\partial\theta'}\hat{\delta}^{(m)}_{T\theta_0}$.

Note that for $j = 1, ..., m$,

$$\frac{\partial}{\partial\theta'}\hat{\rho}_{T\theta}(j) = \frac{\partial\hat{\gamma}_{T\theta}(j)/\partial\theta'}{\hat{\gamma}_{T\theta}(0)} - \frac{\hat{\gamma}_{T\theta}(j)}{\hat{\gamma}_{T\theta}(0)}\frac{\partial\hat{\gamma}_{T\theta}(0)/\partial\theta'}{\hat{\gamma}_{T\theta}(0)}.$$

$\bar{\varepsilon}_{\theta_0} = T^{-1}\sum_{t=1}^{T}\varepsilon_{\theta_0 t} = o_p(1)$ under Assumption A1, $\hat{\gamma}_{T\theta_0}(j) = \gamma_{\theta_0}(j) + o_p(1)$, in particular $\gamma_{\theta_0}(0) = 1$, $\gamma_{\theta_0}(j) = 0$, for $j \in \mathbb{Z}$, and $\partial\hat{\gamma}_{T\theta_0}(j)/\partial\theta' = O_p(1)$, under Assumptions

A1-A3. So we conclude that

$$\sqrt{T}\hat{\rho}_{T\hat{\theta}_T}(j) = \sqrt{T}\hat{\rho}_{T\theta_0}(j) + \dot{\rho}_{T\theta_0}(j)\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) + \sqrt{T}D_T$$

where the j-th element of D_T is $\left(\hat{\theta}_T - \theta_0\right)' \ddot{\rho}^{(m)}_{T\breve{\theta}_j}(j)\left(\hat{\theta}_T - \theta_0\right)$, $\left\|\breve{\theta}_j - \theta_0\right\| \leq \left\|\hat{\theta}_T - \theta_0\right\|$,

$$\dot{\rho}_{T\theta_0}(j) = \frac{1}{T}\sum_{t=j+1}^{T} \varepsilon_{\theta_0 t}\dot{\varepsilon}_{\theta_0 t-j} + \dot{\varepsilon}_{\theta_0 t}\varepsilon_{\theta_0 t-j},$$

in which $\dot{\varepsilon}_{\theta t} = -\frac{\dot{f}_t(\theta)}{h_t(\theta)} - \left\{\frac{f(I_{t-1})-f_t(\theta)}{h_t(\theta)} + \varepsilon_{\theta t}\right\}\frac{\dot{h}_t(\theta)}{h_t(\theta)}$. Denote
$A_{T,1}(j) = \frac{1}{T}\sum_{t=j+1}^{T} \varepsilon_{\theta_0 t}\dot{\varepsilon}_{\theta_0 t-j}$, $A_{T,2}(j) = \frac{1}{T}\sum_{t=j+1}^{T} \dot{\varepsilon}_{\theta_0 t}\varepsilon_{\theta_0 t-j}$. We will show that $\nabla\rho_{\theta_0}(j) = \lim_{T\to\infty} E\left[A_{T,1}(j)\right] + \lim_{T\to\infty} E\left[A_{T,2}(j)\right]$, where $E\left[A_{T,1}(j)\right] = 0$ under H_0. Then, with $\zeta(t, t-j) = \varepsilon_{\theta_0 t}\dot{\varepsilon}_{\theta_0 t-j}$,

$$\begin{aligned}
& E\left\|A_{T,1}(j) - E\left[A_{T,1}(j)\right]\right\|^2 \\
= \; & \frac{1}{T^2}\sum_{t=j+1}^{T}\sum_{r=j+1}^{T} E\left[\zeta(t,t-j)'\zeta(r,r-j)\right] \\
\leq \; & \frac{C}{T^2}\left[E\left\|\zeta(t,t-j)\right\|^{2+\kappa} E\left\|\zeta(r,r-j)\right\|^{2+\kappa}\right]^{1/(2+\kappa)} \times \sum_{t=j+1}^{T}\sum_{r=j+1}^{T} \alpha_{t-n-j-r}^{\kappa/(2+\kappa)} \\
= \; & O\left(T^{-1}\right) = o(1),
\end{aligned}$$

using Assumptions A1 and A2, and Cauchy and Minkowski inequalities.

Similarly $\operatorname{plim}_{T\to\infty} A_{T,2}(j) = E\left[\dot{\varepsilon}_{\theta_0 t}\varepsilon_{\theta_0 t-j}\right] = -E\left[\varepsilon_{\theta_0 t-j}\frac{\dot{f}_t(\theta_0)}{h_t(\theta_0)}\right]$ under H_0, while we obtain

$$\ddot{\rho}^{(m)}_{T\breve{\theta}_j}(j) = \frac{1}{T}\sum_{t=j+1}^{T} \dot{\varepsilon}'_{\breve{\theta}_j t}\dot{\varepsilon}_{\breve{\theta} t-j} + \varepsilon_{\breve{\theta}_j t}\ddot{\varepsilon}_{\breve{\theta} t-j} + \ddot{\varepsilon}_{\breve{\theta}_j t}\varepsilon_{\breve{\theta} t-j} + \dot{\varepsilon}'_{\breve{\theta}_j t-j}\dot{\varepsilon}_{\breve{\theta} t},$$

with

$$\begin{aligned}
\ddot{\varepsilon}_{\theta t} = \; & \varepsilon_{\theta t}\left(2\frac{\dot{h}_t(\theta)'\dot{h}_t(\theta)}{h_t^2(\theta)} - \frac{\ddot{h}_t(\theta)}{h_t(\theta)}\right) - \frac{\ddot{f}_t(\theta)}{h_t(\theta)} + 2\frac{\dot{f}_t(\theta)'\dot{h}_t(\theta)}{h_t^2(\theta)} \\
& + \frac{\left(f(I_{t-1}) - f_t(\theta)\right)}{h_t(\theta)}\left(2\frac{\dot{h}_t(\theta)'\dot{h}_t(\theta)}{h_t^2(\theta)} - \frac{\ddot{h}_t(\theta)}{h_t(\theta)}\right).
\end{aligned}$$

Then using Assumption A3, we find that $E\sup_\theta \left\| \ddot{\rho}^{(m)}_{T\breve{\theta}}(j) \right\| < \infty$, so that $\ddot{\rho}^{(m)}_{T\breve{\theta}}(j) = O_p(1)$.

Similarly for $j = 1, ..., m$,

$$\frac{\partial}{\partial\theta'}\hat{\delta}^{(m)}_{T\theta}(j) = \frac{\partial\hat{\eta}_{T\theta}(j)/\partial\theta'}{\hat{\eta}_{T\theta}(0)} - \frac{\hat{\eta}_{T\theta}(j)}{\hat{\eta}_{T\theta}(0)}\frac{\partial\hat{\eta}_{T\theta}(0)/\partial\theta'}{\hat{\eta}_{T\theta}(0)},$$

with $\bar{\varepsilon}^2_{\theta_0} = T^{-1}\sum_{t=1}^T \varepsilon^2_{\theta_0 t} = 1 + o_p(1)$ under Assumption A1; $\hat{\eta}_{T\theta_0}(j) = \eta_{\theta_0}(j) + o_p(1)$ (in particular $\eta_{\theta_0}(0) = E[(\varepsilon^2_{\theta_0 t} - 1)^2]$, $\eta_{\theta_0}(j) = 0$, for $j \in \mathbb{Z}$), and $\partial\hat{\eta}_{T\theta}(j)/\partial\theta' = O_p(1)$, under Assumptions A1-A3. So we conclude that

$$\hat{\delta}^{(m)}_{T\hat{\theta}_T} = \hat{\delta}^{(m)}_{T\theta_0} + \dot{\delta}^{(m)}_{T\theta_0}\left(\hat{\theta}_T - \theta_0\right) + R_T,$$

where the j-th element of R_T is $\left(\hat{\theta}_T - \theta_0\right)' \ddot{\delta}^{(m)}_{T\breve{\theta}_j}\left(\hat{\theta}_T - \theta_0\right)$, $\left\|\breve{\theta}_j - \theta_0\right\| \le \left\|\hat{\theta}_T - \theta_0\right\|$, with

$$\begin{aligned} \dot{\delta}^{(m)}_{T\theta_0}(j) &= \frac{2}{T\eta_{\theta_0}(0)}\sum_{t=j+1}^{T}\left[\left(\varepsilon^2_{\theta_0 t} - 1\right)\dot{\varepsilon}_{\theta_0 t-j}\varepsilon_{\theta_0 t-j} + \dot{\varepsilon}_{\theta_0 t}\varepsilon_{\theta_0 t}\left(\varepsilon^2_{\theta_0 t-j} - 1\right)\right] + o_p(1), \\ &= \frac{2}{\eta_{\theta_0}(0)}\left[B_{T,1}(j) + B_{T,2}(j)\right] + o_p(1), \end{aligned}$$

say, $j = 1, 2, \cdots, m$. It is easy to prove using Assumptions A1-A2 and the same methods that $p\lim_{T\to\infty} B_{T,1}(j) = E\left[\dot{\varepsilon}_{\theta_0 t-j}\varepsilon_{\theta_0 t-j}E\left[\left(\varepsilon^2_{\theta_0 t} - 1\right)|I_{t-1}\right]\right] = 0$ under H_0 while $p\lim_{T\to\infty} B_{T,2}(j) = -E\left[\varepsilon^2_{\theta_0 t}\left(\varepsilon^2_{\theta_0 t-j} - 1\right)\frac{\dot{h}_t(\theta_0)}{h_t(\theta_0)}\right] = -E\left[\left(\varepsilon^2_{\theta_0 t-j} - 1\right)\frac{\dot{h}_t(\theta_0)}{h_t(\theta_0)}\right]$ under H_0. Finally $R_T = O_p(T^{-1})$ using the same reasoning as for D_T. $\square$

Proof of Theorem 1: The proof is similar to the reasoning in Brown, Durbin and Evans (1975) and Delgado and Velasco (2011) under Assumption A4, and it is omitted, while the central limit theorem for $\hat{\rho}^{(m)}_{T\theta_0}$ and $\hat{\delta}^{(m)}_{T\theta_0}$ follows from Assumption A1. $\square$

Proof of Theorem 2: The arguments are similar to those used in Proposition 1 for the linear expansion of sample autocovariances under H_0, just adapting the centering for the sample autocovariances under H_{1T} and exploiting the mixing properties of $\varepsilon_{\theta_0 t}$. $\square$

References

[1] Berkes, I. L., and Kokoszka, P., (2003). Asymptotics for GARCH Squared Residual Correlations. Econometric Theory, 19, 515-540.

[2] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307-327.

[3] Brillinger, D. R. (1979). Confidence intervals for the crosscovariance function. Selecta Statistic Canada 5, 3-16.

[4] Brown, R.L., Durbin, J. and Evans, J.M. (1975). Techniques for testing constancy of regression relationships over time. Journal of the Royal Statistical Society Series B. 37, 149-192.

[5] Box, G. E. P. and Pierce, D. A. (1970) Distribution of the residual autocorrelations in autoregressive integrated moving average time series models. Journal of American Statistical Association 65, 1509–26.

[6] Delgado, M.A., Velasco, C., (2011). An Asymptotically Pivotal Transform of the Residuals Sample Autocorrelations With Application to Model Checking. Journal of the American Statistical Association. 106, 946-958.

[7] Den Haan, W.J. and A. Levin (1997). A practitioner's guide to robust covariance matrix estimation. Handbook of statistics 15, G.S. Maddala and C.R. Rao, eds., Elsevier, Amsterdam, 291–341.

[8] Engle, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of variance of United Kingdom inflation. Econometrica 50, 987–1007.

[9] Escanciano, J.C., (2008). Joint and marginal specification tests for conditional mean and variance models. Journal of Econometrics, 143, 74-87.

[10] Escanciano, J.C., (2009). Quasi-Maximum Likelihood Estimation Of Semi-Strong Garch Models. Econometric Theory, 25, 561-570.

[11] Francq C., Zakoïan J-M., (2004). Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes. Bernoulli, 10, 605-637.

[12] Francq, C., Roy, R. and Zakoïan, J.-M., (2005). Diagnostic Checking in ARMA Models With Uncorrelated Errors. Journal of the American Statistical Association, 100, 532–544.

[13] Gallant A.R., Hsieh D.A., Tauchen G.,(1991). On fitting a recalcitrant series: the Pound/Dollar exchange rate, 1974–1983. W.A. Barnett, J. Powell, G. Tauchen (Eds.), Nonparametric and Semiparametric Methods in Econometrics and Statistics, Cambridge University Press, Cambridge, UK, pp. 199–240.

[14] Grenander, U., M. Rosenblatt (1957). Statistical Analysis of Stationary Time Series. Wiley, New York.

[15] Gourieroux, C. (1997). ARCH Models and Financial Applications. New York: Spring-Verlag.

[16] Halunga, A. G., Orme, C. D., (2009). First-Order Asymptotic Theory For Parametric Misspecification Tests Of Garch Models. Econometric Theory, vol. 25(02), 364-410.

[17] Hannan E. J. (1957). The Variance of the Mean of a Stationary Process. Journal of the Royal Statistical Society. Series B, 19, 282-285.

[18] Hansen, B. Autoregressive conditional density estimation. International Economic Review, 35 (1994), pp. 705–730

[19] Harvey, C., Siddique A. (1999). Autoregressive conditional skewness. Journal of Financial and Quantitative Analysis, 34, 465-487.

[20] Hong, Y and Lee, T-H., (2003). Diagnostic Checking for The Adequacy of Nonlinear Time Series Models. Econometric Theory 19, 1065-1121.

[21] Jowett, G. H. (1955). The comparison of means of sets of observations from sections of independent stochastic series. Journal of the Royal Statistical Society: Series B, 17, 208-227.

[22] Khmaladze, E. V. (1981). Martingale Approach in the Theory of Goodness-of-Fit Tests. Theory of Probability and Its Applications, 26, 240-257.

[23] Li, W. K., Mak, T. K. (1994). On the squared residual autocorrelations in nonlinear time series with conditional heteroscedasticity. Journal of Time Series Analysis, 15, 627-636.

[24] Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time series models. Biometrika 65, 297-303.

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

[26] 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.

[27] Press, H., J. W. Tukey (1956). Power spectral methods of analysis and their applications to problems in airplane dynamics. Bell Systems Monograph 2606.

[28] Wong H., Ling S., (2005). Mixed portmanteau tests for time-series models. Journal of Time Series Analysis 26, 569-579.