CIENCIA ergo-sum, Revista Científica Multidisciplinaria de Prospectiva
ISSN: 1405-0269
ISSN: 2395-8782
ciencia.ergosum@yahoo.com.mx
Universidad Autónoma del Estado de México
México

Testing financial time series for autocorrelation: Robust Tests

Muriel Torrero, Nelson Omar
Testing financial time series for autocorrelation: Robust Tests
CIENCIA ergo-sum, Revista Científica Multidisciplinaria de Prospectiva, vol. 27, no. 3, 2020
Universidad Autónoma del Estado de México, México
Available in: https://www.redalyc.org/articulo.oa?id=10463384006
DOI: https://doi.org/10.30878/ces.v27n3a6

Ciencias Exactas y Aplicadas

Testing financial time series for autocorrelation: Robust Tests

Autocorrelación en series de tiempo financieras: pruebas robustas

Nelson Omar Muriel Torrero
Universidad Iberoamerican, México
nelson.muriel@ibero.mx
http://orcid.org/0000-0002-7760-7826

DOI: https://doi.org/10.30878/ces.v27n3a6
Redalyc: https://www.redalyc.org/articulo.oa?id=10463384006

Received: 16 January 2019
Accepted: 10 April 2019

Abstract:

Two modified Portmanteau statistics are studied under dependence assumptions common in financial applications which can be used for testing that heteroskedastic time series are serially uncorrelated without assuming independence or Normality. Their asymptotic distribution is found to be null and their small sample properties are examined via Monte Carlo. The power of the tests is studied under the MA and GARCH-in-mean alternatives. The tests exhibit an appropriate empirical size and are seen to be more powerful than a robust Box-Pierce to the selected alternatives. Real data on daily stock returns and exchange rates is used to illustrate the tests.

Keywords: Nonlinear Dependence, Sample Autocorrelation, Portmanteau Statistics, Robust Tests.

Resumen:

Se estudian dos estadísticos de Portmanteau modificados bajo supuestos de dependencia comunes en aplicaciones financieras que pueden utilizarse para comprobar que series de tiempo heterocedásticas son serialmente incorreladas sin suponer independencia o normalidad. Se encuentra que su distribución asintótica es nula y se examinan sus propiedades de muestras pequeñas usando Monte Carlo. El poder de las pruebas se estudia para alternativas MA y GARCH en la media. Las pruebas exhiben un tamaño muestral apropiado y se comprueba que son más poderosas que la prueba robusta de Box-Pierce para alternativas selectas. Ilustramos las pruebas usando datos diarios de retornos financieros y de tipos de cambio.

Palabras clave: dependencia no lineal, autocorrelación muestral, estadísticos de Portmanteau, pruebas robustas.

Introduction

Testing for zero autocorrelation is a frequently encountered problem in applied financial econometrics. Customarily, testing the random walk hypothesis for log-returns of financial assets is done with the Portmanteau statistic of Box & Pierce (1970) or its size-corrected version of Ljung & Box (1978), both of which are based in the vector or empirical autocorrelations

$$\widehat{\rho_X} = (\widehat{\rho_X}(1), \ldots, \widehat{\rho_X}(k))',$$

and whose limiting null distribution is found using Bartlett's formula of Bartlett (1946). This result heavily depends on the hypothesis that the underlying series is not only uncorrelated; but independent. Thus, in a sense, Portmanteau tests based on Bartlett's formula are not only tests for the absence of autocorrelation; but also, and mainly, tests of independence. One way in which this may not be adequate in practice is if the observed series appears to be uncorrelated; but certain functions of it do not. Certainly, if the series $\{X_t\}_t$ is of independent random variables, then so is the transformed series $\{f(X_t)\}_t$ for any function *f*. A usual finding in financial econometrics is the presence of a significant correlation in the squared and absolute log-returns which signals a nonlinear dependence ignored by the usual Portmanteau tests. Since these tests are not robust to nonlinear dependencies, serious consequences arise in terms of both, size distortion and inappropriate power when using them in financial applications.

As argued among other authors (Campbell *et al*., 1996) financial econometrics is more concerned with the absence of autocorrelation than it is with independence in dealing with log-returns. The reason for this is that the models that best suit the examination of the Efficient Market Hypothesis are precisely those in which the log-prices are either a martingale or a random walk with uncorrelated increments. Since martingales in discrete time can be represented as a random walk with martingale difference increments, and given that the property of being a martingale difference entails uncorrelatedness; the random walk with uncorrelated increments is the most general one for dealing with log-returns, according to financial theory.

For this reason, the study of Portmanteau statistics for uncorrelated, dependent data has received some attention in the literature. One of the main strategies in this direction is to correct Bartlett's formula in different ways. For instance, Diebold (1986) focuses on the ARCH(q) case showing that the correct limiting variance for $\hat{\rho}_X(j)$ depends on the autocovariance function of the squared process, $\{X_t^2\}$. Franq *et al*. (2005) and Franq & Zakoïan (2009) also obtain modified versions of Bartlett's formula for weak white noise. A fundamental reference in this direction is Romano & Thombs (1996), where the asymptotic normality of the sample autocorrelation function of weakly dependent processes is proven. In all cases, moment conditions set the stage for generalized laws of large numbers and central limit theorems as those of Ibragimov & Linnik (1971).

Applying a particular version of these results, Lobato *et al*. (2001) propose a simple modification of the Portmanteau statistic of Box and Pierce, Q^*, which is suitable for financial analysis. In fact, Escnaciano and Lobato (2009, p. 978) recommend this modified statistic (or its Ljung-Box analog) to be "routinely computed for financial data instead of the standard Q_q [of Box and Pierce]". A generalization of this test can be found in Lobato *et al*. (2002), and an alternative version in Lobato (2001).

On the other hand, research on improving the power of Portmanteau tests against different alternatives has been fueled by their use in univariate model selection. For example, the test by Monti (1994) uses the empirical partial autocorrelation function, $\hat{\pi}(\cdot)$, instead of $\hat{\rho}(\cdot)$ to test $H_0^{(k)}$. More specifically, the test statistic, Q_M, is exactly that of Ljung & Box (1978); but with $\hat{\rho}(j)$ substituted by $\hat{\pi}(j)$, and is more powerful to AR alternatives. The asymptotic null distribution of Q_M is established by using the fact that $\hat{\pi}_k$ satisfies the same limit theorem as $\hat{\rho}_k$: Bartlett's formula. Other tests have been devised based on completely different ideas. For example, Lin & McLeod (2006) and Peña & Rodríguez (2002, 2006) develop tests based on a general measure of multivariate dependence. Their main idea is that the estimated residuals from an ARMA fit can be viewed as a sample from a multivariate distribution, so that testing for zero autocorrelation amounts to testing for proportionality of their correlation matrix to the identity, in other words, testing whether or not the correlation matrix is diagonal. In a similar vein, Fisher & Gallagher (2012) are inspired in high-dimensional data analysis to derive new weighting schemes for the Portmanteau statistic. All the resulting tests are weighted sums of the empirical autocorrelation and partial autocorrelation functions as explained in Gallagher & Fisher (2015). The asymptotic null distribution under is, in all cases, a linear combination of k independent $X^2(1)$ variates. This follows from the fact that under the assumption of independence, the classical results for the empirical autocorrelation functions apply so that $n\hat{\rho}(j)^2$ converges in distribution to a $X^2(1)$ variate.

In this paper, we combine these two approaches, namely, modifying the Portmanteau statistic to achieve a better power and, simultaneously, making the tests robust to heteroskedasticity and weak dependence as is required for financial applications. We study a modified version of the test of Peña & Rodríguez (2006) and the test of Fisher & Gallagher (2012) under the assumptions of weak dependence of Lobato *et al*. (2001) or (Franq & Zakoïan, 2009). These assumptions are satisfied by the usual models in the GARCH and Stochastic Volatility families and are suitable for financial applications. The limiting null distribution will be obtained as a linear combination of independent $\chi^2(1)$ with coefficients which depend on the fourth order properties of the underlying process. A feasible version of the test, obtained by estimating such moments in a $\sqrt{n}$-consistent fashion, is proposed, and its small sample properties are examined.

We focus on the GARCH process of Bollerslev (1986) and the Long Memory Stochastic Volatility of Hurvich & Soulier (2009) to study the small sample properties of our tests. The reader is referred to Breidt *et al*. (1998), Ding *et al*. (1993), Franq & Zakoïan (2010), Harvey (1998), Lobato *et al*. (2001) for further justification of our choice.

A Monte Carlo experiment is presented which includes seven models representing different degrees and forms of persistence in volatility. For each model, we use sample sizes ranging from small (n = 100) to relatively large (n = 1 000) and different lags (j = 1, 5, 10). We estimate the empirical size of the test and the power under two alternatives: the weak MA and the GARCH in mean. We compare the new tests with the one by Lobato *et al*. (2001) and find that these new statistics offer more powerful tests without a significant size bias.

1. The tests and their asymptotic null distribution

Let $\{\varepsilon_t\}$ be a stochastic process with autocorrelation function $\rho(\#)$ and autocovariance function $\gamma(\#)$. The dependence structure of $\{\varepsilon_t\}$ is limited in two ways: First we impose a mild mixing condition with fast enough a decay and, second, we impose a symmetry condition on the fourth order moments. Outliers are also restricted by fourth order conditions. Specifically, the following set of assumptions is maintained throughout the paper.

Assumptions A1

The following assumptions hold for the underlying stochastic process $\{\varepsilon_t\}$:

1. $\{\varepsilon_t\}$ is stationary,
2. $E[\varepsilon_1^4] < \infty$, and $E[|\varepsilon_1 \varepsilon_2|^{2+\delta}] < \infty$ for some $\delta > 0$,
3. $\{\varepsilon_t\}$ is α–mixing with mixing coefficients satisfying $\sum_n \alpha(n)^{\frac{\delta}{2+\delta}} < \infty$, and
4. $E(\varepsilon_t - \mu)(\varepsilon_{t+i} - \mu)(\varepsilon_{t+d} - \mu)(\varepsilon_{t+d+j} - \mu) = 0$ for $i, j = 1,\ldots,k$; for all d when $i \neq j$ and for $d = 0$ if $i = j$.

Let $\hat{R}$ be the Toeplitz matrix of sample autocorrelations of order $k \geq 1$, that is,

$$R_k = \begin{bmatrix} 1 & \hat{\rho}(1) & \hat{\rho}(2) & \cdots & \hat{\rho}(k) \\ \hat{\rho}(1) & 1 & \hat{\rho}r(1) & \cdots & \hat{\rho}(k-1) \\ \vdots & \vdots & \vdots & & \vdots \\ \hat{\rho}(k) & \hat{\rho}(k-1) & \hat{\rho}(k-2) & \cdots & 1 \end{bmatrix}$$

Also, let $|\hat{R}_k|$ denote its determinant. The statistics of Fisher & Gallagher (2012) and Peña & Rodríguez (2006), hereafter abbreviated as *PR* and FG, for the null hypothesis

$$H_0^{(k)} : \rho(1) = \cdots = \rho(k) = 0,$$

as we will use them throughout the paper are defined as

$$PR = -\frac{n}{k}\log(|\widehat{R_k}|) \text{ and } FG = n(n+2)\sum_{j=1}^{k}\frac{k-j+1}{k}\frac{\hat{\rho}^2(j)}{n-j}.$$

The *PR* statistic is originally normalized by $k + 1$ instead of k, the number of autocorrelations in the test. Nonetheless, Peña & Rodríguez (2006) use this normalization only as a matter of preferred interpretation. We normalize by k in order to make the weight schemes of both statistics the same. Assuming $\{\varepsilon_t\}$ is strong white noise, these statistics share a common limiting distribution which can be written as that of:

$$Z = \sum_{i=1}^{k}\frac{k-i+1}{k}\chi^2(1), \tag{1}$$

where the $\chi^2(1)$ variates are independent. To understand this asymptotic equivalence, let $(\hat{\rho}_k = (\hat{\rho}(1), \hat{\rho}(2),\ldots,\hat{\rho}(k))^T$, $\pi_k = (\pi(1), \pi(2),\ldots,\pi(k))^T$ and M the diagonal matrix

$$M = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \frac{k-1}{k} & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & 0 & 0 & \frac{1}{k} \end{bmatrix}. \tag{2}$$

As one can see from the Appendix to Peña & Rodríguez (2002), two applications of the delta method imply that the *PR* statistic is asymptotically equivalent to $n\hat{\pi}_k^T M\hat{\pi}_k$; while it's clear that the *FG* statistic is asymptotically equivalent to $n\hat{\rho}_k^T M\hat{\rho}_k$. It must be emphasized that this asymptotic equivalence does not rest upon the dependence structure of $\{\varepsilon_t\}$ but only in the form of the test statistics. Furthermore, $\hat{\pi}_k = \hat{\rho}_k + O_P(n^{-1})$ as shown, for instance, in Monti (1994), which gives the asymptotic equivalence of $\sqrt{n}\,\hat{\rho}_k$ and $\sqrt{n}\,\hat{\pi}_k$. This is further explained in the following lemma whose proof we give for completeness of exposition.

Lemma 1: Let $\{\varepsilon_t\}$ satisfy assumptions A1. Then, if the *PR* statistic is asymptotically distributed as X, so is the *FG* statistic.

Proof: As shown in Peña & Rodríguez (2002), the log-determinant of $\hat{R}_k$ can be written as

$$|\hat{R}_k|^{1/k} = \prod_{i=1}^{k} (1 - \hat{\pi}(i)^2)^{(k-i+1)/k},$$

which implies that $-n\log|\hat{R}_k|^{1/k} = -n\sum_{i=1}^{k} \frac{k-i+1}{k} \log(1 - \hat{\pi}(i)^2)$. Apply the δ-method as it is explained, for instance, in Theorem 11.2.14 of Lehmann & Romano (2005) to the function

$$f(x_1,\ldots,x_k) = -\sum_{i=1}^{k} \frac{k-i+1}{k} \log(1 - x_i),$$

to obtain that if $n\hat{\pi}_k^2$ converges in distribution to X then $-n\log|\hat{R}_k|^{1/k}$ does so to $\#f(0)X$, where $\#f$ is the gradient of f and can be easily evaluated to $\nabla f(0) = \left(1, \frac{k-1}{k}, \ldots, \frac{1}{k}\right)$. Now, *FG* is asymptotically equivalent $n\hat{\rho}_k^T M\hat{\rho}_k = n\sum_{j=1}^{k} \frac{k-i+1}{k} \hat{\rho}(i)^2 = n\nabla f(0)\hat{\rho}_k^2$. Since $\sqrt{n}\,\hat{\rho}_k$ shares its asymptotic distribution with $\sqrt{n}\,\hat{\pi}_k$, it follows that *FG* converges in distribution to $\#f(0)X$ and the proof is complete.

Assumptions A1 are a particular case of those used by Romano & Thombs (1996) to derive a Central Limit Theorem for the sample autocorrelations and is a simple generalization of the assumptions of Theorem 18.5.3 in Ibragimov & Linnik (1971). Under A1 and $H_0^{(k)}$, Theorem 3.2 in Romano & Thombs (1996) states that

$$n^{1/2}\,\hat{\rho}_k \Rightarrow N(0,W), \tag{3}$$

where $W(i,j) = \gamma(0)^{-2}\,(C_{i,j} - \rho(I)C_{0,j} - \rho(j)C_{0,i} + \rho(I)\rho(j)C_{0,0})$. The last condition in assumptions A1 helps us in simplifying W to $W(i,j) = \gamma(0)^{-2}\,C_{i,j}$. Furthermore, $C_{i,j} = 0$ for $i \neq j$ and $C_{i,i} = E\,[(\varepsilon_t - \mu)^2\,(\varepsilon_{t+i} - \mu)^2]$, so that W is diagonal. This is the same strategy followed in Lobato (2001) and as mentioned in the Introduction, includes the usual GARCH and LMSV models. As a consequence, we have the following result.

Theorem 1: Let $\{\varepsilon_t\}$ be a stationary stochastic process for which assumptions A1 hold. Then, under $H_0^{(k)}$, the limit in distribution of the statistics *PR* and *FG* can be written as $X = \sum_{i=1}^{k} \lambda_i \tau_i \chi^2(1)$. The $\chi^2(1)$ variables are independent, $\lambda_i = (k + i - 1)/k$, and

$$\tau_i = \frac{E\left[(\varepsilon_t - \mu)^2 (\varepsilon_{t+i} - \mu)^2\right]}{E\left[(\varepsilon_t - \mu)^2\right]^2}$$

Proof: First, we know from Lemma 1, that both statistics have the same limiting distribution, so we focus on the *PR* case. From the proof to Lemma 1, we also know that $PR = n\hat{\pi}_k^T M\hat{\pi}_k$. Since $\{\varepsilon_t\}$ satisfies assumptions A1, we know from (Romano, & Thombs, 1996) that $\hat{\pi}_k$ is asymptotically Normal with asymptotic covariance matrix W/n. The results in Box (1954) imply that $PR \Rightarrow \sum_{j=1}^{k} \nu_i \chi^2(1)$, where ``#" signifies convergence in distribution, the $\chi^2(1)$ variates are independent, and ν_i are the eigenvalues of matrix W/M. Finally, since both, W and M, are diagonal, with respective diagonal elements τ_i and λ_i, the result follows.

The asymptotic distribution depends on the unknown quantities $\{\tau_i\}$ which need to be estimated. Consistent estimators of τ_i are readily available making this a simple procedure to implement. To make the robust *PR* and *FG* statistics feasible, we estimate τ_i consistently as

$$\hat{\tau}_i = \frac{\frac{1}{n}\sum_{t=i+1}^{T} (\varepsilon_t - \bar{\varepsilon})^2 (\varepsilon_{t+i} - \bar{\varepsilon})^2}{\left(\frac{1}{n}\sum_{t=i}^{T} (\varepsilon_t - \bar{\varepsilon})^2\right)^2} \tag{4}$$

We follow Fisher & Gallagher (2012) and Peña & Rodríguez (2002, 2006) in approximating the limiting distribution with the methods of Box (1954) and Satterthwaite (1941, 1946). Thus, we use the distribution of an $a\chi^2(b)$ variable as an approximation to the distribution of $\sum_{i=1}^{k} \lambda_i \hat{\tau}_i \chi^2(1)$. The constants a and b are chosen to equate the first two moments of these distributions and are, in our particular case, given by

$$a = \frac{\sum_{i=1}^{k} \lambda_i^2 \hat{\tau}_i^2}{\sum_{i=1}^{k} \lambda_i \hat{\tau}_i}, \qquad b = \frac{\left(\sum_{i=1}^{k} \lambda_i \hat{\tau}_i\right)^2}{\sum_{i=1}^{k} \lambda_i^2 \hat{\tau}_i^2}. \tag{5}$$

Remark: Under assumptions A1, Bartlett's formula may hold either exactly or approximatively. If the squared process $\{\varepsilon_t^2\}$ is uncorrlated, we have $\tau_i = 1$ in Theorem 1, which implies that Bartlett's formula holds exactly in this case. On the other hand, if the lag, i, is large then $W(i, I) \approx 1$. To see why, define $z_t = \varepsilon_t - \mu$, so that

$$\tau_i = \frac{E[z_t^2 z_{t+1}^2]}{E[z_t^2]^2} = \frac{\gamma_{z^2}(i) + E[z_t^2]^2}{E[z_t^2]^2} = \frac{\gamma_{z^2}(i)}{\gamma_z(0)^2} + 1.$$

Now, under assumptions A1, $\{\varepsilon_t\}$ is short memory, and thus so are $\{z_t\}$ and $\{z_t^2\}$. Therefore $\gamma_z{}^2(I) \to 0$ and Bartlett's formula holds asymptotically.

Thus, the need for modifications to the usual Portmanteau tests for zero autocorrelation is driven by the autocorrelation in the squared process, which as explained in Cont (2001), Granger & Ding (1995) and Granger *et al*. (2000), is common in financial data.

2. Design of the Monte Carlo experiments

We focus on two families of stochastic processes that are common in financial applications, namely the GARCH and LMSV models with Gaussian innovations, which satisfy assumptions A1. Thus, we simulate from the mul- tiplicative models $\varepsilon_t = \sigma_t Z_t$, where the specification of σ_t is either GARCH or LMSV. We specify our GARCH models to have the orders (1, 1) and thus $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$. The parameters (ω, α, β) are required to satisfy $\omega > 0$, to avoid the trivial stationary solution, $1 - \alpha - \beta > 0$, which implies second order

stationarity, and $1 - 3\alpha^2 - \beta^2 - 2\alpha\beta > 0$, which is necessary for the fourth order moment to exist. For the LMSV, we focus on the fractional AR(1) specification

$$\log(\sigma_t^2) = (1 - \phi L)^{-1} (1 - L)^{-d} \eta_t := h_t,$$

where $|\varphi| < 1$, so that the process is stationary, $0 < d < 0.5$, so that there is persistence in the volatility process,$\{\eta_t\}$ and is strong Gaussian white noise with variance σ_n^2. The process $\{\varepsilon_t\}$ can then be written as

$$\varepsilon_t = \exp(h_t) Z_t,$$

and since h_t is a Gaussian process, ε_t follows a lognormal distribution which implies the existence of its moments. In particular

$$\text{Var}(\varepsilon_t) = \exp(\sigma_h^2/2),$$
$$E[\varepsilon_t^4] = 3\exp(2\sigma_h^2),$$

with $\sigma_h^2 = \gamma(1 - 2d)/\gamma(1 - d)^2 \sigma_n^2$, the variance of h_t. See, for example Harvey (1998).

We conduct our experiments for the hypotheses $H_0^{(k)}$ for $k = 1, 5, 10$ and for the sample sizes of $n = 100, 200, 500$, and 1 000. Even if these sample sizes are relatively small, when the availability of long streams of daily log-returns is considered, we choose them to illustrate the speed of convergence of the test to its nominal size. A total of $M = 10\,000$ independent paths are generated from each of the models in table 1. The particular parameters chosen are intended to represent the usual range found in financial applications and are similar to those used by Lobato *et al.* (2001) to study the Q^* statistic. The robust test based on the *PR* and *FG* statistics are then performed as outlined in the previous Section, and the size is estimated as the empirical rejection probability, that is, the ratio of the number of rejections to the number of simulated paths.

TABLE 1
Simulated processes under the Null

Model	Parameters		
GARCH	α	β	ω
1	0.09	0.90	0.01
2	0.01	0.97	0.01
3	0.10	0.90	0.01
LMSV	ϕ	σ_n^2	d
1	0.97	0.005	0.25
2	0.97	0.005	0.45
3	0.90	0.003	0.25
4	0.90	0.003	0.45

Source: Own elaboration

To study the power of the tests, we use two alternative hypotheses. In the first, we use the GARCH and LMSV models in table 1 as innovations in an MA(1) process. Thus, under this alternative we have $\{\varepsilon_t\}$ a GARCH or LMSV process and simulate from $u_t = \varepsilon_t + \vartheta\varepsilon_{t-1}$, allowing ϑ to vary in (0.01, 0.35). Under the second alterna- tive, we have a specification in the mean given by $u_t = \mu + c\sigma_t^2 + \varepsilon_t$, where $\{\varepsilon_t, \sigma_t\}$ are simulated with each one of the model specifications of table 1. Parameter c is allowed to vary in [-0.9, 0.9], whereas parameter μ is fixed to 0.005 which does not pose any problem with the tests since μ is just a location parameter. As in the study of the empirical size, we generate $M = 10\,000$ independent paths of each process and apply the robust testing pro- cedures each time. We report the empirical rejection probability as an estimate of

the power of the tests. All the simulations for these experiments were carried out in the Julia language of Bezanson *et al*. (2017).

3. Monte Carlo Results

We begin by observing that the GARCH 3 specification is an IGARCH(1, 1) model which therefore does not meet the hypotheses A1, in particular its strongly stationary solution is not second-order stationary. Nonetheless, as can be seen in the tables below, the size of the tests is quite well achieved, which suggests that the result in Theorem 1 can be proven under even milder assumptions. Tracing this fact to the minimal set of assumptions required for the robust statistics to perform as expected is out of the scope of this paper; but may be an interesting line of future research.

Figures 1, 2, and 3 show the deviations of the empirical size from the nominal one for the robust tests under the chosen GARCH and LMSV specifications. We include the Q^* test of Lobato *et al*. (2001), labeled *LNS*, which is a robust Box test with good empirical properties, as a reference. The first thing to notice in these Figures, is that there is no overall winner for every given sample size, model, and nominal size. Nonetheless, when it comes to large sample sizes, illustrated here with $n = 1\,000$, the *PR* and *FG* tests exhibit a smaller deviation from its intended size most of the times. The scales of the vertical axis in the figures show that the deviation is always small, though it increases slightly with the lag. The *FG* test appears to be more sensitive to persistence for smaller sample sizes than the *PR* test; but as the sample size grows, the situation is reversed in many cases. Overall, we can say that for sample sizes larger than, say, $n = 500$ the empirical size is certainly reasonable for applications. For the smaller sample sizes, some distortion will occur, but not a drastic one. In any case, for smaller sample sizes one could bootstrap the test to make its size more accurate.

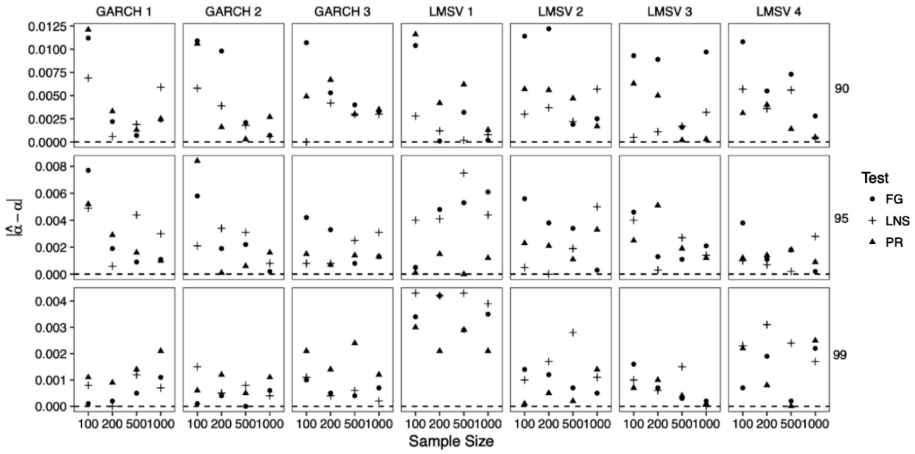

FIGURE 1
Empirical size of the *PR* and *FG* tests at lag 1. Included, as a reference, is the LNS test. The horizontal axis shows sample size, while the vertical axis shows the empirical size. The dotted line indicates the nominal size.
Source: Own elaboration

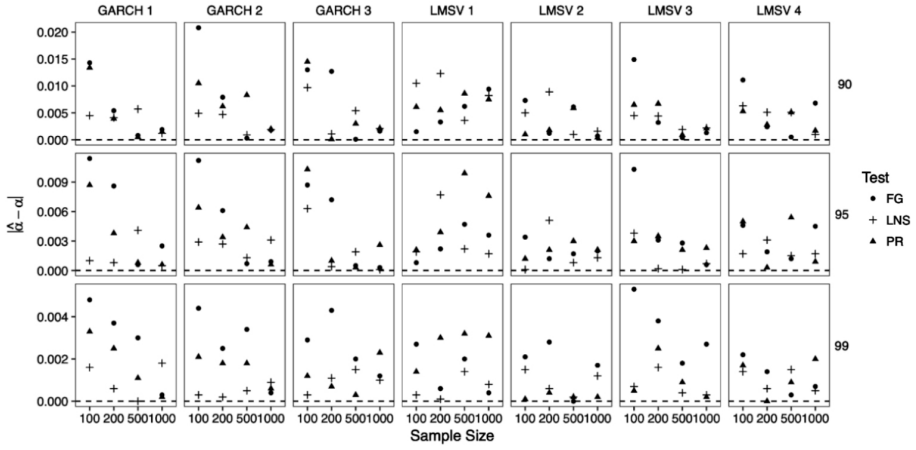

FIGURE 2
Empirical size of the *PR* and *FG* tests at lag 5. Included, as a reference, is the LNS test. The horizontal axis shows sample size, while the vertical axis shows the empirical size. The dotted line indicates the nominal size.
Source: Own elaboration

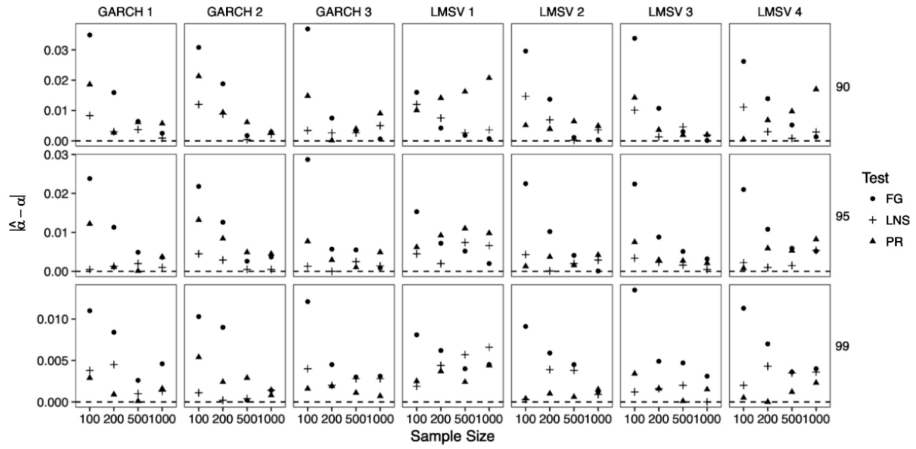

FIGURE 3
Empirical size of the *PR* and *FG* tests at lag 10. Included, as a reference, is the LNS test. The horizontal axis shows sample size, while the vertical axis shows the empirical size. The dotted line indicates the nominal size.
Source: Own elaboration

The power against the MA(1) alternative is presented graphically for the GARCH models in figure 4. We choose the nominal level of the test as $\alpha_0 = 0.95$ in all cases. As it can be appreciated, the persistence plays a role in the power function, making it grow slower, if slightly. The difference in persistence between the specifications is consistently 0.01 being 1 for GARCH3, 0.99 for GARCH 1 and 0.98 for GARCH2. The decrease in the steepness of the power function with increasing persistence does not prevent the test from being highly sensitive even in the IGARCH model. As it can be seen, the test for $H_0^{(1)}$ is, for all practical purposes, equivalent under all the tests. Nonetheless, for $H_0^{(5)}$ and $H_0^{(10)}$ the *PR* and *FG* tests exhibit a considerably higher power.

Figure 5 shows the same kind of results for the MA(1) model with LMSV innovations that the GARCH counterparts of figure 4. Again, the greater the persistence is in the volatility process, the lower the power of the test for any given value of ϑ. Here, the effect of persistence is combined in the parameters d, for the fractional noise, and φ for the AR(1) part of the process. The upper part of the figure concerns models LMSV

1 and 2, in which $\varphi = 0.97$, whereas the lower part depicts the power function for models with $\varphi = 0.90$. On the other hand, the lefthand side of the figure includes models with $d = 0.25$, while $d = 0.45$ can be seen in the righthand side. The loss of power is evident in both directions, when φ grows from 0.90 to 0.97 and d is fixed, or when φ is fixed and d grows from 0.25 to 0.45. Indeed, the upper right figure, which corresponds to $\varphi = 0.97$ and $d = 0.45$ shows a much slower increase in power than the one in the lower left where $\varphi = 0.90$ and $d = 0.25$. Notice that in this case the *PR* and *FG* tests are also more powerful than the LNS test, the difference in power being more pronounced for more persistent processes.

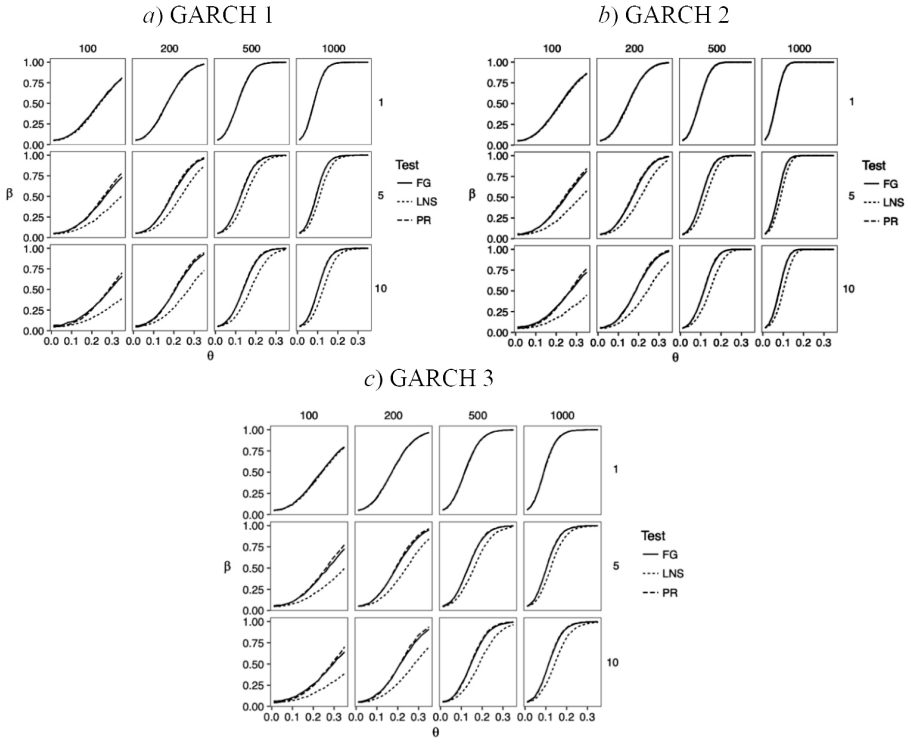

FIGURE 4
Power function for the MA(1) alternative under GARCH innovations. The figures are faceted by sample size and lag. The horizontal axis shows the value of $\vartheta \# [0.01, 0.35]$ and the verticalaxis shows the power function
Source: Own elaboration

Figures 6 and 7 illustrate the situation for the second alternative of processes with a volatility effect in the mean. Again, we see the strong impact that persistence has on the power of the tests. This is, of course, not a surprise since the correlation in the series $\{\varepsilon_t\}$ is introduced by that of the volatility process $\{\sigma_t^2\}$. In this case all the tests seem to have a virtually equivalent power. The *FG* test is slightly more powerful than the other procedures in most of the instances when the powers differ. Figure (b) suggests that with a persistence as high as 0.98, the test may have a very low power.

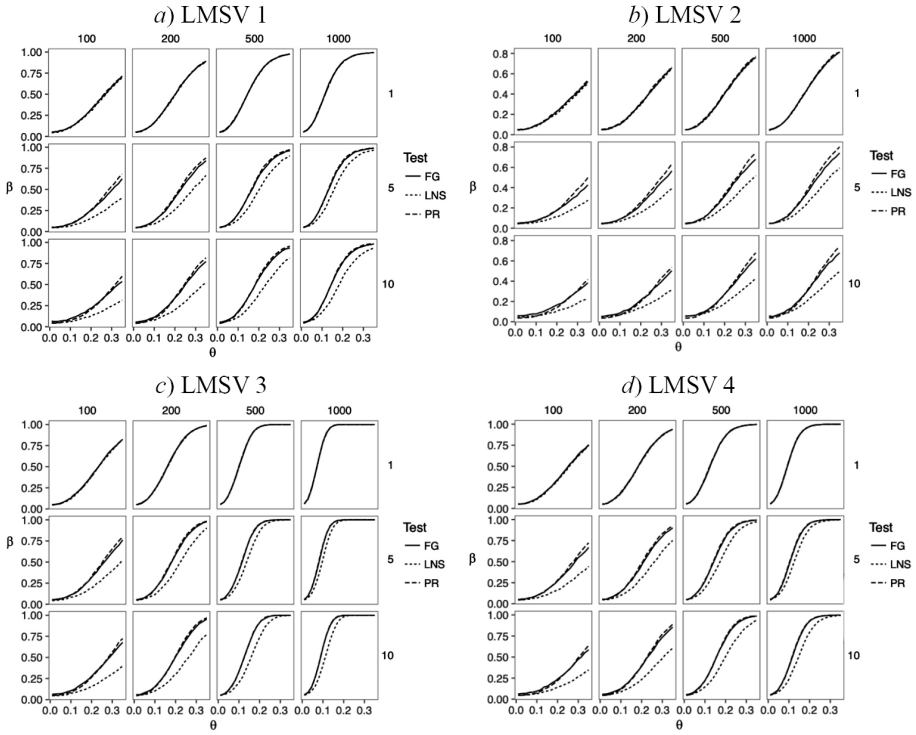

FIGURE 5
Power function for the MA(1) alternative under LMSV innovations. The figures are faceted by sample size and lag. The horizontal axis shows the value of $\vartheta \# [0.01, 0.35]$ and the vertical axis shows the power function
Source: Own elaboration

Figures 6 and 7 illustrate the situation for the second alternative of processes with a volatility effect in the mean. Again, we see the strong impact that persistence has on the power of the tests. This is, of course, not a surprise since the correlation in the series $\{\varepsilon_t\}$ is introduced by that of the volatility process $\{\sigma_t^2\}$. In this case all the tests seem to have a virtually equivalent power. The *FG* test is slightly more powerful than the other procedures in most of the instances when the powers differ. Figure (b) suggests that with a persistence as high as 0.98, the test may have a very low power. Remember that under the GARCH 1 and GARCH 3 specifications, the persistence is 0.99 and 1.0 respectively. However, this low power is rather a function of parameter α in the specification of the model, namely,

$$\varepsilon_t = \sigma_t Z_t,$$
$$\sigma_t^2 = 0.01 + \alpha\varepsilon_t^2 + (0.98 - \alpha)\sigma_t^2,$$
$$\alpha \in [0.01, 0.20].$$

Finally, figure (6-d) depicts the power function of the family of GARCH-M models with $\mu = 0.005$ and $c = 0.9$ and a rapid and persistent increase of power can be seen for all testing procedures.

Observe that the persistence in volatility is fixed to 0.98 just as in the GARCH 2 specification. The reason for this behavior is that the correlation in the series comes from the correlation in the volatilities, which is an increasing function of α. When $\alpha = 0.01$, the correlation between u_t and u_{t-1} is as low as 0.00083 and that between u_t and u_{t-10} as low as 0.00069.

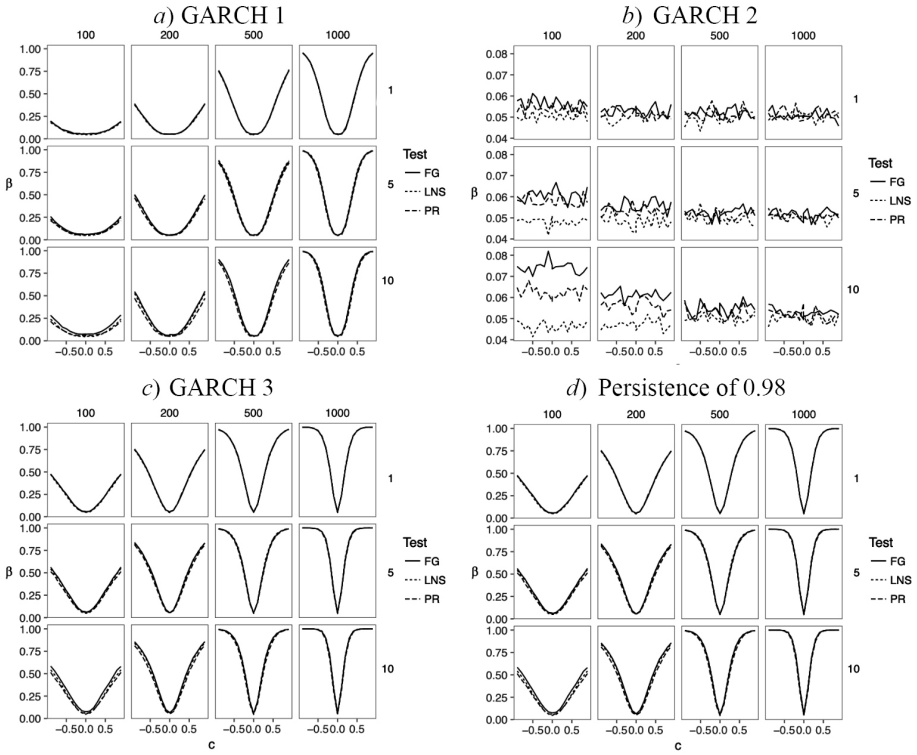

FIGURE 6
Power function for the in-mean alternative in GARCH specifications. The figures are facetted by sample size and lag. The horizontal axis shows the value of the parameter *c* in [-0.9, 0.9] and the vertical axis shows the power function. Figure (d) helps better explain what we see in Figure (b).
Source: Own elaboration

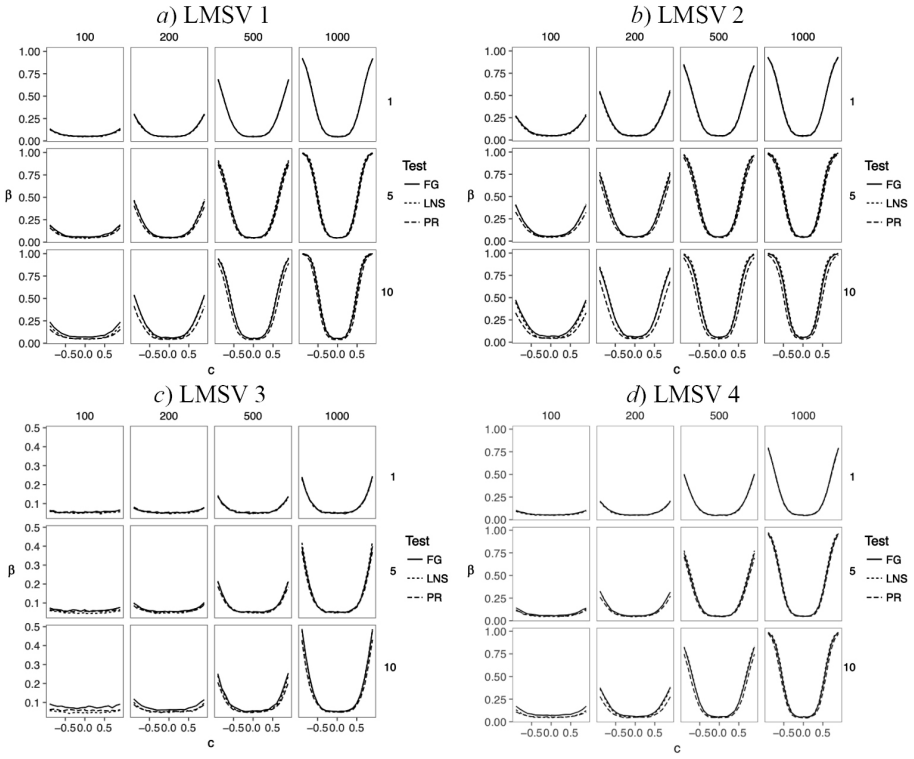

FIGURE 7
Power function for the in-mean alternative in LMSV specifications. The figures are facetted by sample size and lag. The horizontal axis shows the value of the parameter *c* in [-0.9, 0.9] and the vertical axis shows the power function.
Source: Own elaboration

4. Empirical application

In this section, we consider the log-returns of different stock and indexes, and the growth rate of some exchange rates. We test for zero autocorrelation at lags 1, 5, and 10 with the usual Box-Pierce and Ljung-Box tests, and with the robust tests studied in this paper: LNS, PR, FG. The empirical quantiles are determined, as in our Monte Carlo experiment, with the $a\chi^2(b)$ approximation using (8) with $\hat{\tau}_i$ as in (7). We report the test statistic and its significance, coded with stars, as usual. The data for stocks and indexes has been downloaded from Yahoo! Finance and includes the daily log-returns of the S&P500, NASDAQ composite, CAC-40, DAX, NIKKEI-225, Exxon Mobile Corporation (XOM), Bank of America Corporation (BAC) and Apple Inc (AAPL). The data spans the period from 2007-01-03 to 2018-06-15. The daily exchange rates were downloaded from the Federal Reserve Bank of Saint Louis, through FRED, and include the following exchanges: USD / EUR, CNY / USD, JPY / USD, USD / GBP, MXN / USD. The data covers the period starting 2013 -01- 03 and ending 2018 -06 -15.

Table 2 shows the results of our application. The first thing to notice is that both traditional tests employed,Box-Pierce and Ljung-Box, tend to reject at high significance levels where the robust tests do not. For example,with the S&P500 index $H_0^{(1)}$ is rejected at 99% with the traditional tests; but all robust tests fail to reject. In other instances, the robust tests lower the level of rejection from 99% to 90% as is the case of $H_0^{(5)}$ also for S&P500. In practical terms, since empirical work is usually carried out at least at a 95% confidence level, this is equivalent to going from a sound rejection to a non-rejection. This discrepancy is due to the size distortion experimented by the traditional tests under nonlinear dependence.

TABLE 3
Tests for zero autocorrelation for stocks and indexes. Included are the tests of Box-Pierce (BP), Ljung-Box (LB), Lobato-Nankervis-Savin (LNS), Robsustified Peña-Rodríguez(RPR) and Robistified Fisher Gallagher (RFG).

Lag	Test	S&P500	NASDAQ	CAC40	DAX	NIKKEI	XOM	BAC	APL
1	BP	11.948***	29.947***	0.389	0.082	1.516	189.116***	2.889*	4.047**
	LB	11.950***	29.954***	0.389	0.083	1.516	18.121***	2.890*	4.048**
	LNS	2.288	6.709***	0.173	0.038	0.381	2.367	0.267	1.590
	RPR	11.948	29.947***	0.389	0.082	1.516	18.116	2.889	4.047
	RFG	11.950	29.954***	0.389	0.082	1.516	18.121	2.890	4.048
5	BP	47.541***	44.267***	44.400***	21.006***	14.140**	93.174***	51.123***	26.224***
	LB	47.552***	44.279***	44.436***	21.023***	14.145**	93.207***	51.155***	26.237***
	LNS	7.414	9.969*	16.981***	7.987	3.722	18.070***	6.356	11.955**
	RPR	39.274*	40.548**	20.010**	9.206	10.732	75.792***	20.177	16.637*
	RFG	38.125*	38.986**	19.555**	9.188	19.655	69.630***	20.381	16.637*
10	BP	57.640***	50.995***	50.552***	24.743***	27.623***	103.150***	77.689***	34.786***
	LB	57.657***	51.015***	50.597***	24.765***	27.637***	103.190***	77.753***	33.810***
	LNS	10.126	11.701	19.722**	9.564	8.506	21.601**	10.123	16.697*
	RPR	47.701*	44.785*	34.705***	16.348	17.857	96.563***	38.113	23.563*
	RFG	45.934	43.123*	33.729***	16.300	17.688	85.009***	39.210	23.612*

Source: Own elaboration

Comparing the robust tests among them is more interesting, since it is these tests that are corrected for dependence. In some cases, the tests offer no practical difference. For example, working at 5% with the S&P500 series, all three tests agree that there is no correlation; or that there is correlation in NASDAQ or order 1. Lag 5 of NASDAQ is an example of the robust *PR* and *FG* tests offering a different appreciation than the robust Box-Pierce (LNS). In this case *LNS* does not reject $H_0^{(5)}$ whereas both, *PR* and *FG*, do at 5%. A similar situation is seen in lag 10 of CAC-40 and XOM where rejection goes from 95% in LNS to 99% in both, *PR* and *FG*. Conversely, lag 5 of AAPL shows an instance in which the new tests do not reject $H_0^{(5)}$ at 5% but *LNS* does. Similarly, for lag 5 of CAC-40, *LNS* rejects at 99% when the robust *PR* and *FG* do only at 95%.

We can see how the traditional tests of Box & Pierce (1970) and Ljung & Box (1978) lead to spurious autocorrelation. Apparently, the robust Box-Pierce test of Lobato *et al.* (2001) may be subject to the opposite error when the actual correlation is small enough; but the *PR* and *FG* test seem quite sensitive to deviations from the null. Thus, when using the traditional tests, we will be led, usually, to overparameterization of the conditional mean of the series under study, that is, to overparameterized ARMA models with conditionally heteroskedastic innovations. Of course, this is not desirable since the overparameterization would cause the point forecasts not to be optimal (in mean squared error) and the confidence intervals for them to be misleading. Another reason why spurious autocorrelation is to be avoided concerns a common use of the econometric model: The estimation of risk measures in the context of dynamic risk management. In this case, as explained in Chapter 4 of Embrechts *et al.* (2005), conditional risk measures are estimated by means of the representation

$$r_t = \mu_t + \sigma_t Z_t,$$

where r_t is the (negative) log-return and (μ_t, σ_t) stand for the conditional mean and volatility of the series. Unnecessarily including an ARMA component would lead to setting a working model for μ_t when in fact μ_t

= 0, which would impact the estimation of the associated risk measures. For instance, the Value at Risk and Expected Shortfall for period $t + 1$ based on the information available up to time t would be computed as

$$VaR_{\alpha}^{t} = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} VaR_{\alpha}(Z),$$

$$ES_{\alpha}^{t} = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} ES_{\alpha}(Z),$$

where $\hat{\mu}_{t+1}$ and $\hat{\sigma}_{t+1}$ are the one-step-ahead forecasts for the mean and the volatility, and Z is a random variable with the same distribution as Z_t . Incorrectly deciding to model $\{\mu_t\}$ as an ARMA process would lead to biased estimators of both measures which implies that the VaR will not have the required level and the expected shortfall will not accurately depict the conditional distribution of losses exceeding the VaR_{α}^{t}.

Table 3 shows the results of our application to exchange rates. We see that persistence in volatility and nonlinear dependence does not seem to affect the traditional tests in some cases. For example, the null hypothesis of zero autocorrelation is not rejected for the USD / EUR at all lags $k = 1, 5, 10$ and all the tests. This also happens with the exchange JPY/USD for lags 1 and 5. Another scenario is illustrated by CNY/USD and USD/GBP, where volatility and nonlinear dependence alter the size of the traditional tests, both of which favor rejection at 5%; but not for the robust versions which do not reject the null. Lastly, for MXN/USD we see that all tests favor rejection at some level for lags 5 and 10. In the first case, the robust Box-Pierce test (*LNS*) rejects at 5% whereas the newly proposed tests (*PR* and FG) do only at 10%. In practical terms, where decisions are usually taken at least at the 5% level, *LNS* favors rejection while *PR* and *FG* do not, so that the later tests support the hypothesis that the exchange rate follows a random walk (when applied with five lags). Lag 10 for this same exchange shows that the *LNS* and *FG* tests favor rejection at 5% whereas *PR* does only at 10%. In practical terms, *LNS* and *FG* would imply, overall, that the exchange MXN/USD has a statistically significant (short) memory, whereas *PR* would advise otherwise.

Finally, two things should be noticed about the proposed tests: First, when applied to strong white noise they are equivalent to their non-robust counterparts. Indeed, since for strong white noise we have $\gamma_{\varepsilon^2}(I) = 0$ the limiting distribution of $\hat{\rho}_k$ is given by Bartlett's formula so that the results in Fisher & Gallagher (2012) and Peña & Rodríguez (2006) apply. As noted earlier, this is not limited to strong white noise; but it is valid whenever the squared process is uncorrelated. Second, the robust *PR* and *FG* tests exhibit a similar power which is greater, in many cases, than that of the *LNS* test. Since the *FG* statistic is computationally simpler, we rephrase the advice of Escanciano & Lobato (2009) and recommend to routinely compute the modified *FG* statistic when testing for zero autocorrelation in financial applications.

Conclusions

Testing for autocorrelation is deeply connected with some of the most common hypothesis in financial theory, such as the Efficient Market Hypothesis, so that empirical financial econometrics require reliable tests for the hypothesis of zero autocorrelation. Even though the existing literature provides with different test statistics for autocorrelation, most of them are developed under the hypothesis of independence or even Normality of the sequence under study. These hypotheses are inconsistent with financial data, making it necessary to develop tests for autocorrelation specifically suited for financial applications. In this paper we proposed and studied two such tests and the results make their use promising.

To begin with, both tests are based on recently proposed Portmanteau statistics which are more powerful than the traditional tests of Box & Pierce (1970); Ljung & Box (1978) which suggests that their robustification will also be more powerful. This intuition is corroborated by our Monte Carlo study at least for two common alternatives, namely, the moving average and the GARCH in mean. It should also be noted that the three tests that are compared are almost identical for lag one; but as the lag being tested increases, the new tests are more sensitive than the modified Box test. Finally, since the limiting distribution can

be approximated with a simple transformation of a χ^2 distribution, its applicability does not impose high computational costs.

The proposed generalizations differ from the one given in Lobato *et al*. (2001) in that it is not the statistic that we compute differently, but rather the limiting distribution. It should be realized that there is a double approximation in this process: First, the linear combination of χ^2 random variable is an asymptotic distribution and, second, the terms appearing in this linear combination are, themselves, approximations –being, as they are, consistent estimators of the actual terms involved. This imprecision is counteracted, as usual, by larger sample sizes. In other areas, such as macroeconomics, requiring large samples can be problematic; but not in finance, so that the tests are applicable without small-sample corrections. Future research includes the generalization of our results to general weighted Portmanteau statistics, where the weight functions may be fixed or random.

References

Bartlett, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time series. *Journal of the Royal Statistical Society, Supplement*, *8*(1), 27-41.

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, *59*(1), 65-98.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, *31*, 307-327.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems I: Effect on the inequality of variance in the one-way clasification. *Annals of Mathematical Statistics*, *25*, 290-302.

Box, G. E. P., & Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive integrated moving average time series models. *Journal of the American Statistical Association*, *65*, 1509-1526.

Breidt, F. J., Crato, N., & de Lima, P. (1998). The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics*, *83*, 325-348.

Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1996). *The econometris of financial markets*. Princeton University Press.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, *1*, 223-236.

Diebold, F. X. (1986). Testing for serial correlation in the presence of ARCH. In *Proceedings of the Business and Economics Statistics Section* (pp. 323-328).

Ding, Z., Granger, C. W. J., & Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, *1*, 83-106.

Embrechts, P., McNeil, A. J., & Frey, R. (2005). *Risk management: Concepts, techniques and tools*. Princeton University Press.

Escanciano, J. C., & Lobato, I. N. (2009). Palgrave handbook of econometrics: Applied Econometrics. In T. C. Mills & K. Patterson (Eds.) (pp. 972-1003). Palgrave MacMillan.

Fisher, T. J., & Gallagher, C. M. (2012). New weighted Portmanteau statistics for time series goodness of fit testing. *Journal of the American Statistical Association*, *107*(498), 777-787.

Franq, C., & Zakoïan, J.-M. (2009). Bartlett's formula for a general class of nonlinear processes. *Journal of Time Series Analysis*, *30*(4), 449-465.

Franq, C., & Zakoïan, J.-M. (2010). *GARCH models: Structure, statistical inference and financial application*s. Wiley.

Franq, C., Zakoïan, J.-M., & Roy, R. (2005). Diagnostic checking in ARMA Models with uncorrelated errors. *Journal of the American Statistical Associatio*n, *100*(470), 532-544.

Gallagher, C. M., & Fisher, T. J. (2015). On weighted Portmanteau tests for time-series goodnes-of-fit. *Journal of Time Series Analysis*, *36*, 67-83.

Granger, C. W. J., & Ding, Z. (1995). Some properties of absolute returns: An alternative measure of risk. *Annales d'économie et de Statistique*, *40*, 67-95.

Granger, C. W. J., Spear, S., & Ding, Z. (2000). Stylized facts on the temporal and distributional properties of absolute returns: An update. In W.-S. Chan, W. K. Li, & H. Tong (Eds.), *Statistics and finance: An interface* (pp. 97-120). Imperial College Press, London.

Harvey, A. C. (1998). Long memory in stochastic volatility. In J. Knight & S. Satchell (Eds.), *Forecasting volatility in financial markets.* Butterworth-Heinemann, London.

Hurvich, C. M., & Soulier, P. (2009). Stochastic volatility models with long memory. In T. Mikosch, J.-P. Kreiß, R. A. Davis, & T. G. Andersen (Eds.), *Handbook of financial time series* (pp. 345-354). Berlin: Springer Berlin Heidelberg.

Ibragimov, I., & Linnik, Y. (1971). *Independent and stationary sequences of random variables*. Wolters-Noordhoff Publishing Groningen.

Lehmann, E. L., & Romano, J. P. (2005). *Testing statistical hypotheses*. Springer Texts in Statistics.

Lin, J.-W., & McLeod, A. I. (2006). Improved Peña-Rodriguez portmanteau test. *Computational Statistics & Data Analysis*, *51*, 1731-1738.

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, *62*(2), 297-303.

Lobato, I. N. (2001). Testing that a dependent process is uncorrelated. *Journal of the American Statistical Association*, *96*(455), 1066-1076.

Lobato, I. N., Nankervis, J. C., & Savin, N. E. (2001). Testing for autocorrelation using a modified box--pierce Q Test. *International Economic Review*, *42*(1), 187-205.

Lobato, I. N., Nankervis, J. C., & Savin, N. E. (2002). Testing for zero autocorrelation in the presence of statistical dependence. *Econometric Theory*, *18*(3), 730-743.

Monti, A. C. (1994). A proposal for a residual autocorrelation test in linear models. *Biometrika*, *81*(4), 776-780.

Peña, D., & Rodríguez, J. (2002). A powerful portmanteau test of lack of fit for time series. *Journal of the American Statistical Association*, *97*(458), 601-610.

Peña, D., & Rodríguez, J. (2006). The log of the determinant of the autocorrelation matrix for testing goodness of fit in time series. *Journal of Statistical Planning and Inference*, *136*, 2706-2718.

Romano, J. P., & Thombs, L. A. (1996). Inference for autocorrelations under weak assumptions. *Journal of the American Statistical Association*, *91*(434), 590-600.

Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrica*, *6,* 309-316.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, *2*, 110-114.

Alternative link

https://cienciaergosum.uaemex.mx/article/view/11758 (html)