Testing normality for heteroscedastic macroeconomic variables

Preliminary version

Hamdi Raïssi*

May 30, 2017

Abstract: In this paper the testing of normality for unconditionally heteroscedastic macroeconomic time series is studied. It is underlined that the classical Jarque-Bera test (JB hereafter) for normality is inadequate in our framework. On the other hand it is found that the approach which consists in correcting the heteroscedasticity by kernel smoothing for testing normality is justified asymptotically. Nevertheless it appears from Monte Carlo experiments that such methodology can noticeably suffer from size distortion for samples that are typical for macroeconomic variables. As a consequence a bootstrap methodology for correcting the problem is proposed. The innovations distribution of a set of inflation measures for the U.S., Korea and Australia are analyzed.

Keywords: Unconditionally heteroscedastic time series; Jarque-Bera test.

JEL: C12, C15, C18

---

*Instituto de Estadística, PUCV. Address: Av. Errazuriz 2734, Valparaíso, V región, Chile. E-mail: hamdi.raissi@pucv.cl. This paper was supported by CONICYT-FONDECYT under grant $N°1160527$.

1 Introduction

In the econometric literature, the Jarque Bera (1980) test is routinely used to assess the normality of variables. The properties of this test are well documented for stationary conditionally heteroscedastic processes. For instance Fiorentini, Sentana and Calzolari (2003), Lee, Park and Lee (2010) and Lee (2012) investigated the test developed by JB test in the context of GARCH models. However few studies are available on the distributional specification of the errors in presence of unconditional heteroscedasticity. Considering S&P500 returns Drees and Stărică (2002) and Mikosch and Stărică (2004) investigated the distribution of the errors corrected from unconditional heteroscedasticity. Fryźlewicz (2005) pointed out that unconditionally heteroscedastic financial time series can exhibit a sample kurtosis that may lead to consider spuriously leptokurtic distributions. Note that non constant variance constitutes an important pattern for time series in general, and macroeconomic variables in particular. Reference can be made to Sensier and van Dijk (2004) who found that most of the 214 U.S. macroeconomic time series they studied have a time-varying variance. In this paper we aim to provide a reliable tool for testing normality for small samples time series with non constant unconditional variance.

The structure of the paper is as follows. In Section 2 we first set the dynamics ruling the observed process. In particular the unconditional heteroscedastic structure of the errors is given. The inadequacy of the standard JB test in our framework is highlighted. The approach consisting in correcting the errors from the heteroscedasticity for building a JB test is presented. We then introduce a bootstrap procedure that is intended to correct the inaccuracy of the above JB test. In Section 3 numerical experiments are conducted to shed some light on the finite sample behavior of the studied tests. We illustrate our outputs examining the distributional properties of the U.S., Korean and Australian GDP implicit price deflators.

The following notations are used. The convergence in distribution is denoted by $\Rightarrow$. For a random variable v we define $\|v\|_q = (E|v|^q)^{1/q}$, with $E|v|^q < 1$ and $q \geq 1$.

2 Testing normality in presence of unconditional heteroscedasticity

We consider processes $(y_{t,n})$ which can be written as

$$\begin{aligned} y_{t,n} &= \omega_0 + x_{t,n}, \\ x_{t,n} &= \sum_{i=1}^{p} a_{0i} x_{t-i,n} + u_{t,n}, \end{aligned} \tag{2.1}$$

where $y_{1,n}, \dots, y_{n,n}$ are available, n being the sample size and $E(x_{t,n}) = 0$. The conditional mean of $x_{t,n}$ is driven by the autoregressive parameters $\theta_0 = (a_{01}, \dots, a_{0p})'$. We make the following assumption on the conditional mean.

Assumption A0: The $a_{0j} \in \mathbb{R}$, $1 \le j \le p$, are such that $\det(a(z)) \neq 0$ for all $|z| \le 1$, with $a(z) = 1 - \sum_{j=1}^{p} a_{0j} z^j$.

In the assumption **A1** below, the well known rescaling device introduced by Dahlhaus (1997) is used to specify the errors process $(u_{t,n})$.

Assumption A1: We assume that $u_{t,n} = h_{t,n}\epsilon_t$ where:

(i) $h_{t,n} \ge c > 0$ for some constant $c > 0$, and satisfies $h_{t,n} = g(t/n)$, where $g(r)$ is a measurable deterministic function on the interval $(0, 1]$, such that $\sup_{r \in (0,1]} |g(r)| < \infty$. The function $g(.)$ satisfies a Lipschitz condition piecewise on a finite number of some sub-intervals that partition $(0, 1]$.

(ii) The process (ϵ_t) is iid and such that $E(\epsilon_t) = 0$, $E(\epsilon_t^2) = 1$, and $(E(\|\epsilon_t\|^{8\nu}) < \infty$ for some $\nu > 1$.

The non constant variance induced by **A1(i)** allows for a wide range of non stationarity patterns commonly faced in practice, as for instance abrupt shifts, smooth changes or cyclical behaviors. Note that in the zero mean AR case, tools needed to carry out the Box and Jenkins specification-estimation-validation modeling cycle, are provided in Patilea and Raïssi (2013) and Raïssi (2015). For $\omega_0 \neq 0$ define the estimator $\hat{\omega} = n^{-1}\sum_{t=1}^{n} y_{t,n}$, and $x_{t,n}(\omega) = y_{t,n} - \omega$ for any $\omega \in \mathbb{R}$. Writing $\hat{\omega} - \omega_0 = n^{-1}\sum_{t=1}^{n} x_{t,n}$, it can be shown that

$$\sqrt{n}(\hat{\omega} - \omega_0) = O_p(1), \tag{2.2}$$

using the Beveridge-Nelson decomposition. Now let

$$\hat{\theta}(\omega) = \left(\Sigma_{\underline{x}}(\omega)\right)^{-1} \Sigma_{x}(\omega), \tag{2.3}$$

where

$$\Sigma_{\underline{x}}(\omega) = n^{-1}\sum_{t=1}^{n} \underline{x}_{t-1,n}(\omega)\underline{x}_{t-1,n}(\omega)' \quad \text{and} \quad \Sigma_{x}(\omega) = n^{-1}\sum_{t=1}^{n} \underline{x}_{t-1,n}(\omega)x_{t-1,n}(\omega),$$

with $\underline{x}_{t-1,n}(\omega) = (x_{t-1,n}(\omega), \dots, x_{t-p,n}(\omega))'$. With these notations define the OLS estimator $\hat{\theta}(\hat{\omega})$ and the unfeasible estimator $\hat{\theta}(\omega_0)$. Straightforward computations give $\sqrt{n}(\hat{\theta}(\hat{\omega}) - \hat{\theta}(\omega_0)) = o_p(1)$, so that using the results of Patilea and Raïssi (2012) we have

$$\sqrt{n}(\hat{\theta}(\hat{\omega}) - \theta_0) = O_p(1). \tag{2.4}$$

Once the conditional mean is filtered in accordance to (2.2) and (2.4), we can proceed to the test of the following hypotheses:

$$H_0 : \epsilon_t \sim \mathcal{N}(0,1) \quad \text{vs.} \quad H_1 : \epsilon_t \text{ has a different distribution},$$

with the usual slight abuse of interpretation inherent of the use JB test for normality testing. Clearly the skewness and kurtosis of $u_{t,n}$ correspond to those of ϵ_t. However in practice $E(u_{t,n}^3) = 0$ and $E(u_{t,n}^4) = 3$ is checked using the JB test statistic:

$$Q_{JB}^{u} = n\left[Q_{JB}^{S,u} + Q_{JB}^{K,u}\right], \tag{2.5}$$

where

$$Q_{JB}^{S,u} = \frac{\hat{\mu}_3^2}{6\hat{\mu}_2^3} \quad \text{and} \quad Q_{JB}^{K,u} = \frac{1}{24}\left(\frac{\hat{\mu}_4}{\hat{\mu}_2^2} - 3\right)^2,$$

with $\hat{\mu}_j = n^{-1}\sum_{t=1}^n (\hat{u}_{t,n} - \bar{\hat{u}})^j$ and $\bar{\hat{u}} = n^{-1}\sum_{t=1}^n \hat{u}_{t,n}$. The $\hat{u}_{t,n}$'s are the residuals obtained from the estimation of the conditional mean. If we suppose the process (u_t) homoscedastic ($g(.)$ is constant), then the standard result $Q_{JB}^u \Rightarrow \chi_2^2$ is retrieved (see Yu (2007), Section 2.2). However under **A0** and **A1** with $g(.)$ non constant (the unconditionally heteroscedastic case) we have:

$$Q_{JB}^{K,u} = \frac{1}{24}\left[\kappa_2\left(E(\epsilon_t^4)\right) - 3\right) + 3\left(\kappa_2 - 1\right)\right] + o_p(1), \tag{2.6}$$

where $\kappa_2 = \frac{\int_0^1 g^4(r)dr}{\left(\int_0^1 g^2(r)dr\right)^2}$. Hence if we suppose the errors process unconditionally heteroscedastic with $E(\epsilon_t^4) = 3$, we have $Q_{JB}^u = Cn + op(n)$ for some strictly positive constant C. As a consequence, the classical JB test will tend to detect spuriously departures from the null hypothesis of a normal distribution in our framework. This fact was used by Fryźlewicz (2004) to underline that unconditionally heteroscedastic specifications can cover time series that exhibit an excess of kurtosis.

In order to assess the distribution of S&P500 returns, Drees and Stărică (2002) considered errors corrected from heteroscedasticity using a kernel estimator of the variance. We will follow this approach in the sequel considering

$$\hat{h}_{t,n}^2 = \sum_{i=1}^n w_{ti}(\hat{u}_{i,n} - \bar{\hat{u}})^2, \qquad 1 \le t \le n,$$

with $w_{ti} = \left(\sum_{j=1}^n K_{tj}\right)^{-1} K_{ti}$ and $K_{ti} = K((t-i)/nb)$ if $t \neq i$ and $K_{ii} = 0$, where $K(\cdot)$ is a kernel function on the real line and b is the bandwidth. The following assumption is needed for our variance estimator.

Assumption A2: (i) The kernel $K(\cdot)$ is a bounded density function defined on the real line such that $K(\cdot)$ is nondecreasing on $(-\infty, 0]$ and decreasing on $[0, \infty)$ and $\int_{\mathbb{R}} v^2 K(v)dv < \infty$. The function $K(\cdot)$ is differentiable except a finite number of points and the derivative $K'(\cdot)$ satisfies $\int_{\mathbb{R}} |xK'(x)|dx < \infty$. Moreover, the Fourier Transform $\mathcal{F}[K](\cdot)$ of $K(\cdot)$ satisfies $\int_{\mathbb{R}} |s|^{\tau} |\mathcal{F}[K](s)| \, ds < \infty$ for some $\tau > 0$.

(ii) The bandwidth b is taken in the range $\mathfrak{B}_n = [c_{min} b_n, c_{max} b_n]$ with $0 < c_{min} < c_{max} < \infty$ and $nb_n^{4-\gamma} + 1/nb_n^{2+\gamma} \to 0$ as $n \to \infty$, for some small $\gamma > 0$.

Let $\hat{\epsilon}_t = (\hat{u}_{t,n} - \bar{\hat{u}})/\hat{h}_{t,n}$. We are ready to introduce the following JB test statistic:

$$Q^{\epsilon}_{JB} = n \left[Q^{S,\epsilon}_{JB} + Q^{K,\epsilon}_{JB} \right],$$

where

$$Q^{S,\epsilon}_{JB} = \frac{\hat{\nu}_3^2}{6\hat{\nu}_2^3} \quad \text{and} \quad Q^{K,\epsilon}_{JB} = \frac{1}{24} \left(\frac{\hat{\nu}_4}{\hat{\nu}_2^2} - 3 \right)^2,$$

with $\hat{\nu}_j = n^{-1} \sum_{t=1}^{n} \hat{\epsilon}_t^j$. The following proposition gives the asymptotic distribution of Q^{ϵ}_{JB}.

Proposition 1. *Under the assumptions* **A0**, **A1** *and* **A2**, *we have as* $n \to \infty$

$$Q^{\epsilon}_{JB} \Rightarrow \chi^2_2, \tag{2.7}$$

uniformly with respect to $b \in \mathfrak{B}_n$.

Proposition 1 can be proved using the same arguments given in Yu (2007), together with those of the proof of Proposition 4 in Patilea and Raïssi (2014). Therefore we skip the proof. For building a test using the above result, we will consider the normal kernel, and a bandwidth chosen minimizing the cross-validation (CV) criterion (see Wasserman (2006,p69-70)). The test obtained using (2.7) and the above settings will be denoted by T_{cv}. The standard test, that do not take into account the unconditional heteroscedasticity, is denoted by T_{st}.

For high frequency time series it is reasonable to suppose that the approximation (2.7) is satisfactory when the bandwidth is carefully chosen. Nevertheless considering the above sophisticated procedure for small n is questionable. Therefore we propose to apply the following bootstrap algorithm inspired from Francq and Zakoïan (2010,p335).

1- Generate $\epsilon_t^{(b)} \sim \mathcal{N}(0,1)$, $1 \leq t \leq n$, build the bootstrap errors $u_{t,n}^{(b)} = \epsilon_t^{(b)} \hat{h}_{t,n}$, and the bootstrap series $y_t^{(b)}$ using (2.1), but with $\hat{\omega}$ and $\hat{\theta}(\hat{\omega})$ (see (2.2) and (2.3)).

2- Estimate the autoregressive parameters and a constant as in (2.1), but using the $y_t^{(b)}$'s. Build the kernel estimators $\hat{h}_{t,n}^{(b)}$ from the resulting residuals $\hat{u}_{t,n}^{(b)}$.

3- Compute $\hat{\epsilon}_{t,n}^{(b)} = \hat{u}_{t,n}^{(b)} / \hat{h}_{t,n}^{(b)}$ for $t = 1, \dots, n$. Compute $Q_{JB}^{\epsilon,(b)}$.

4- Repeat the steps 1 to 3 B times for some large B. Use the $Q_{JB}^{\epsilon,(b)}$'s to compute the p-values of the bootstrap JB test.

The test obtained using the above bootstrap procedure is denoted by T_{boot}.

3 Numerical illustrations

The finite sample properties of the T_{st}, T_{cv} and T_{boot} tests are first examined by means of Monte Carlo experiments. The distribution of the GDP implicit price deflator for the U.S., Korea and Australia is then investigated. Throughout this section the asymptotic nominal level of the tests is 5%. In the sequel, we fixed $B = 499$.

3.1 Monte Carlo experiments

We simulate $N = 1000$ trajectories of AR(1) processes:

$$y_{t,n} = a_0 y_{t-1,n} + u_{t,n}, \tag{3.1}$$

where $a_0 = 0.4$ and $u_{t,n} = h_{t,n}\epsilon_t$ with ϵ_t iid(0,1). Under the null hypothesis we set $\epsilon_t \sim \mathcal{N}(0,1)$. On the other hand under the alternative hypothesis $\epsilon_t = \cos(\delta)v_t + \sin(\delta)w_t$ is taken, with $v_t \sim \mathcal{N}(0,1)$, $(\sqrt{2}w_t + 1) \sim \chi_1^2$, and $0 \leq \delta \leq \frac{\pi}{2}$. In order to study the case where the series are actually homoscedastic, we set $h_{t,n} = 1$. For the heteroscedastic case, the variance structure is given by

$$h_{t,n} = 1 + 2\exp\left(t/n\right) + 0.3(1 + t/n)\sin\left(5\pi t/n + \pi/6\right).$$

In such situation the variance structure exhibits a global monotone behavior together with a cyclical/seasonal component that is common in macroeconomic data (see e.g. Trimbur, and Bell (2010) for seasonal effects in the variance). In all our experiments, the mean in (3.1) is treated as unknown. More precisely the AR parameter in (3.1) is estimated using $y_{t,n} - \hat{\omega}$, where $\hat{\omega}$ is given in (2.2), and then the resulting centered residuals are used to compute the test statistics.

The outputs obtained under the null hypothesis are first analyzed. The results are given in Table 1 for the homoscedastic case and in Table 2 for the heteroscedastic case. Noting that macroeconomic time series with noticeable heteroscedasticity are relatively large but smaller than $n = 400$ in general, a special emphasis will be put on interpreting results for samples $n = 100, 200, 400$. Since $N = 1000$ processes are simulated, and under the hypothesis that the finite sample size of a given test is 5%, the relative rejection frequency should be between the significant limits 3.65% and 6.35% with probability 0.95. The outputs outside these confidence bands will be displayed in bold type.

From Table 1, it appears that the T_{cv} is oversized for small samples ($n = 100$ and $n = 200$). This could be explained by the fact that this test is too much sophisticated for the standard case. When the samples are increased the relative rejection frequencies become close to the 5% ($n = 400$ and $n = 800$). On the other hand the T_{st} and T_{boot} tests have good results for all the samples. Of course if there is no evidence of heteroscedasticity, the simple

T_{st} should be used. However Table 1 reveals that in case of doubt about the presence of heteroscedasticity in the data, the use of the T_{boot} is a good alternative.

In the heteroscedastic case, it is seen from Table 2 that the T_{st} test fails to control the type I error as $n \to \infty$. This was expected from (2.6). Next it seems that the relative rejection frequencies of the T_{cv} test are somewhat for from the nominal level 5%, even when $n = 800$. From Table 2 it also emerges that the T_{boot} control reasonably well the type I error. It is important to point out that experiments not displayed here show good results for the T_{cv} test for larger samples ($n > 1000$). If high frequency time series are analyzed, the T_{cv} should certainly be preferred to the computationally intensive T_{boot}. Nevertheless we can draw the conclusion that the T_{boot} ensures a correct implementation of the JB test for samples that are typical for heteroscedastic macroeconomic variables.

Now we turn to the analysis of the behavior of the tests under the alternative hypothesis. The outputs of the simulations we carried out are displayed in Figure 1. For a fair comparison we only studied the T_{st} and T_{boot} in the homoscedastic case. The sample size $n = 100$ is fixed and recall that the parameter δ defines the departures from the null hypothesis. The results we obtained show that the T_{boot} test does not suffer from a lack of power in comparison to the T_{st}. In conclusion it turns out that the T_{boot} improves the distribution analysis, in the sense that it ensures a good control of the type I error, but without entailing noticeable loss of power.

3.2 Real data analysis

The inflation measures data are commonly used to analyze economic facts. Reference can be made to the numerous empirical papers studying the relation between price levels and money supply (see e.g. Jones and Uri (1986)). In such kind of investigations clearly the distributional analysis can help to build a model for the data. For instance following Engle (1982), authors aimed to detected ARCH effects assessing asymmetry and/or leptokurticity in inflation variables (see e.g. Broto and Ruiz (2008p22), among others).

In this part we will study the normality of the log differences of the (quarterly) GDP implicit price deflators for the U.S., Korea and Australia from 10/01/1983 to 01/01/2017 ($n = 132$). More precisely we use $y_{t,n} = 100 \log (GDP_{t,n}/GDP_{t-1,n})$. The data can be downloaded from the webpage of the economic research of the federal reserve bank of Saint Louis: https://fred.stlouisfed.org. The studied variables plotted in Figure 2 seem to show cyclical heteroscedasticity. In the case of Korea we can suspect a global decreasing behavior leading to a stabilization after the Asian crisis. The times series are first filtered according to (2.1). The non correlation of the residuals is tested using the adaptive portmanteau test of Patilea and Raïssi (2013). Once the linear dynamics of the series seem captured in an appropriate way, the tests considered in this paper are applied to the residuals. The results are given in Table 3. When the null hypothesis of normality is rejected at the 5% level, the p-value is displayed in bold type. It emerges that the outputs of the T_{boot} test are in general clearly different from those of the T_{cv} and T_{boot} tests. The p-values of the T_{cv} are all lower than those of the T_{boot}. Note that in the case of the U.S. GDP implicit price deflator, the difference between the T_{boot} and T_{st}, T_{cv} tests lead to different conclusions. In view of the outputs obtained from the simulations experiments, it is reasonable to decide that the normality assumption cannot be rejected for the U.S. data.

References

Broto, C., and Ruiz, E. (2008) Testing for conditional heteroscedasticity in the components of inflation. Working document.

Dahlhaus, R. (1997) Fitting time series models to nonstationary processes. *Annals of Statistics* 25, 1-37.

Drees, H., and Stărică, C. (2002) A simple non stationary model for stock returns. Preprint. Universität des Saaland.

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987-1007.

Fiorentini, G., Sentana, E., Calzolari, G. (2004) On the validity of the Jarque-Bera normality test in conditionally heteroscedastic dynamic regression models. *Economic Letters* 83, 307-312.

Francq, C., and Zakoïan, J-M. (2010) *GARCH models : structure, statistical inference, and financial applications.* Wiley.

Fryźlewicz, P. (2005) Modelling and forecasting financial log-returns as locally stationary wavelet processes. *Journal of Applied Statistics* 32, 503-528.

Jarque, C.M., and Bera, A.K. (1980) Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters* 6, 255-259.

Jones, J.D., and Uri, N. (1986) Money, inflation and causality (another look at the empirical evidence for the USA, 1953-84). *Applied Economics* 19, 619-634.

Lee, S., Park, S., and Lee, T. (2010) A note on the JarqueâĂŞBera normality test for GARCH innovations. *Journal of the Korean Statistical Society* 39, 93-102.

Lee, T. (2010) A note on JarqueâĂŞBera normality test for ARMAâĂŞGARCH innovations. *Journal of the Korean Statistical Society* 41, 37-48.

Mikosch, T., and Stărică, C. (2004) Stock market risk-return inference. An unconditional non-parametric approach. Research report, the Danish national research foundation: Network in Mathematical Physics and Stochastics.

Patilea, V., and Raïssi, H. (2012) Adaptive estimation of vector autoregressive models with time-varying variance: application to testing linear causality in mean. *Journal of Statistical Planning and Inference* 142, 2891-2912.

Patilea, V., and Raïssi, H. (2013) Corrected portmanteau tests for VAR models with time-varying variance. *Journal of Multivariate Analysis* 116, 190-207.

Patilea, V., and Raïssi, H. (2014) Testing second order dynamics for autoregressive processes in presence of time-varying variance. *Journal of the American Statistical Association* 109, 1099-1111.

Sensier, M., and van Dijk, D. (2004) Testing for volatility changes in U.S. macroeconomic time series. *Review of Economics and Statistics* 86, 833-839.

Trimbur, T.M., and Bell, W.R. (2010) Seasonal heteroscedasticity in time series data: modeling, estimation, and testing. Working paper.

Wasserman, L. (2006) *All of Nonparametric Statistics.* Springer, New-York.

Raïssi, H. (2015) Autoregressive order identification for VAR models with non-constant variance. *Communications in Statistics: Theory and Methods* 44, 2059-2078.

Yu, H. (2007) High moment partial sum processes of residuals in ARMA models and their applications. *Journal of Time Series Analysis* 28, 72-91.

Tables and Figures

Table 1: Empirical size (in %) of the studied tests for normality. The homoscedastic case.

n	100	200	400	800
T_{st}	4.0	5.2	4.9	4.2
T_{cv}	**7.2**	**7.5**	5.6	5.0
T_{boot}	4.5	5.6	5.0	4.7

Table 2: Empirical size (in %) of the studied tests for normality. The heteroscedastic case.

n	100	200	400	800
T_{st}	**8.7**	**13.0**	**11.5**	**19.1**
T_{cv}	**9.4**	**9.2**	**8.3**	**7.8**
T_{boot}	4.4	**6.5**	6.3	6.3

Table 3: The p-values of the tests for normality for GDP implicit price deflators of the U.S., Korea and Australia.

	U.S.	Korea	Australia
T_{st}	**3.8**	16.4	50.9
T_{cv}	**2.3**	82.0	21.0
T_{boot}	8.2	87.0	49.0

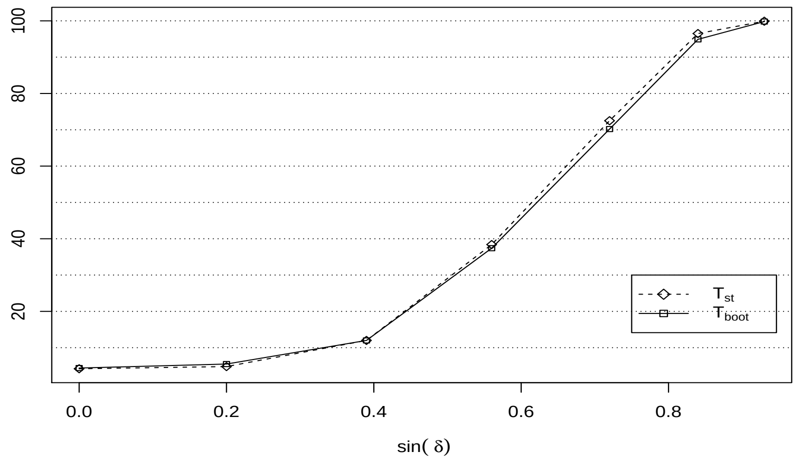

Figure 1: Empirical power (in %) of the T_{st} and T_{boot} tests.

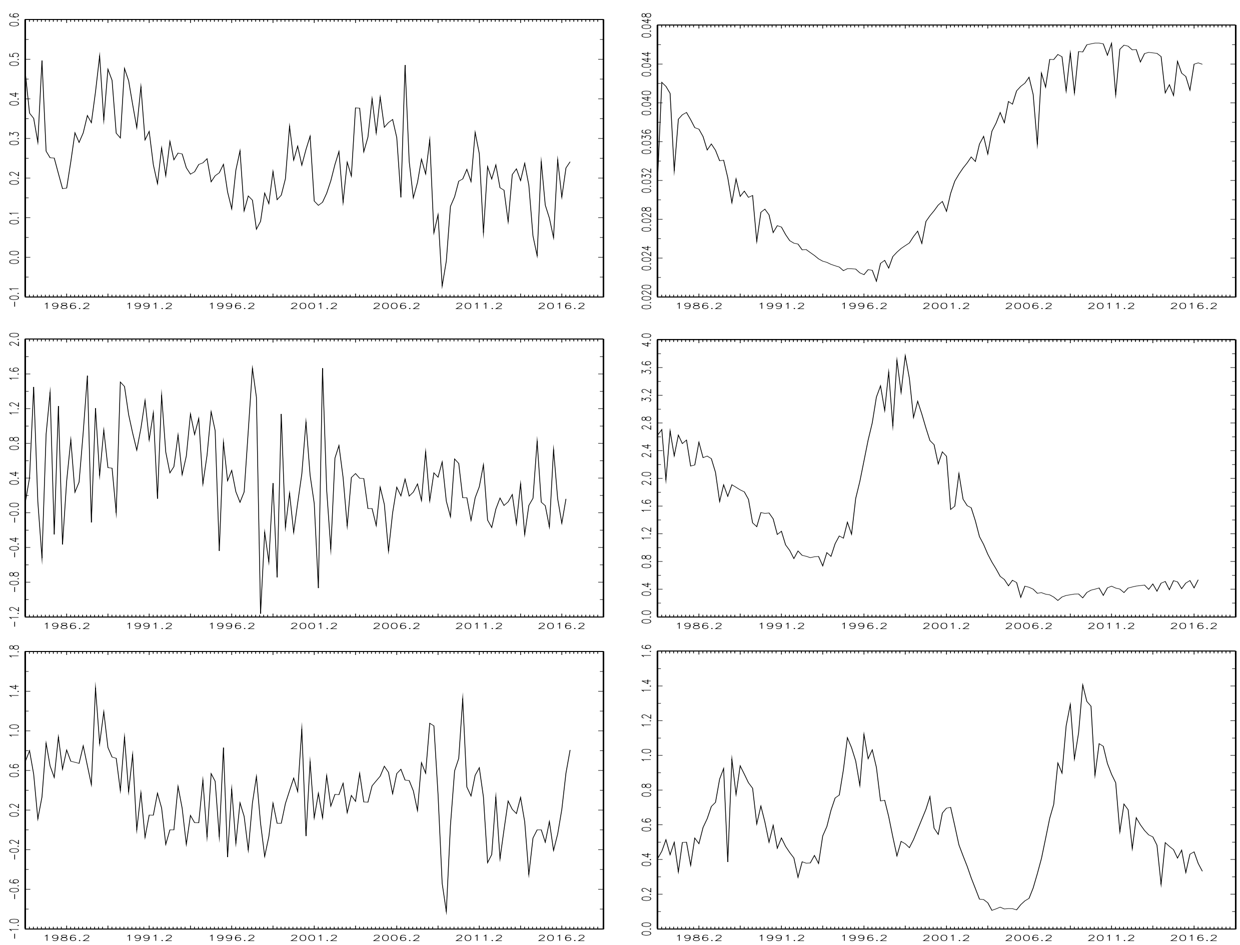

Figure 2: The log differences of the quarterly U.S. (top left panel), Korean (middle left panel) and Australian (bottom left panel) GDP implicit price deflators from 10/01/1983 to 01/01/2017 ($n = 132$), and the corresponding estimations of the innovations variance on the right. Data source: The research division of the federal reserve bank of Saint Louis, fred.stlouisfed.org.