1 Introduction

Tests of distributional symmetry about a known or unknown centre are used to answer substantive questions in economics (see e.g. Kilian and Demiroglu, 2000). A case in point is the analysis of time-reversibility (TR) that hinges on the distributional properties of time series. TR holds if the statistical properties of a time series are not affected when it is observed in reverse time. Tests of TR have been applied to the analysis of business cycle symmetry (DeLong and Summers, 1986), the existence of Edgeworth price cycles (Beare and Seo, 2014) and tests of the mixture-of-distribution hypothesis (Fong, 2003). TR can be assessed relying on the coefficient of skewness of the j-th order difference of a time series (DeLong and Summers, 1986).

We consider a test of the null hypothesis of distributional symmetry based on L-moments and its application to time series data. L-moments are linear functions of expectations of order statistics that characterize the shape of any distribution with finite mean (Hosking, 1990).

We make four contributions to the literature on tests of symmetry for dependent data. First, we introduce the test of symmetry based on L-moments due to Harri and Coble (2011) in the time series econometrics literature. Second, we study the comparative performance of tests of symmetry based on the coefficient of skewness and on L-moments with Monte Carlo experiments. Third, we introduce a bootstrap version of the test, suitable for time series applications. Fourth, we analyse the symmetry of business cycles for the G7 countries.

The rest of the paper is organized as follows. Section 2 discusses tests of symmetry and their bootstrap implementation. Section 3 presents the simulation study. Section 4 is devoted to the empirical application and Section 5 concludes.[1]

2 Symmetry tests and their bootstrap implementation

Let $\{X_t\}_{t=1}^{T}$ be a sample from a strictly stationary time series $\mathcal{X} = \{X_t\}_{t=-\infty}^{\infty}$ with mean μ and r-th central moment $\mu_r = E\left[(X_t - \mu)^r\right]$. We focus on two tests of symmetry of the

[1] Additional results and details are available as Supplementary material.

probability distribution of X_t about its centre.[2]

Test based on moments ($\hat{\tau}_3$). If X_t is symmetrically distributed, $\mu_3 = 0$ and its coefficient of skewness $SK \equiv \mu_3/\mu_2^{3/2} = 0$. When X_t is iid normally distributed, a test of symmetry based on the squares of sample skewness, $\widehat{SK}$, is:

$$\hat{\tau}_3 = T\frac{\widehat{SK}^2}{6} \xrightarrow{d} \chi^2_1 \tag{1}$$

The null hypothesis of the test, $H_0 : SK = 0$, is rejected whenever $\hat{\tau}_3$ is greater than the upper critical value of a χ^2_1.

Test based on L-moments ($\hat{\tau}_{3,L}$). The first four L-moments of X_t are (Hosking, 1990):

$$\ell_1 = \int_0^1 Q(u)\mathrm{d}u \tag{2}$$
$$\ell_2 = \int_0^1 Q(u)(2u-1)\mathrm{d}u \tag{3}$$
$$\ell_3 = \int_0^1 Q(u)(6u^2-6u+1)\mathrm{d}u \tag{4}$$
$$\ell_4 = \int_0^1 Q(u)(20u^3-30u^2+12u-1)\mathrm{d}u \tag{5}$$

where $Q(\alpha)$ is the quantile function, while ℓ_1 and ℓ_2 are measures of location and scale. Population L-skewness (SK_L, for $r=3$) and L-kurtosis (KR_L, for $r=4$) are defined as ℓ_r/ℓ_2 for $r=3,4$. For a standard Normal variate $SK_L = 0$ and $KR_L = 0.1226$. Since $|\ell_r/\ell_2| < 1$ for $r \geq 3$, SK_L and KR_L are bounded on the unit interval. This property makes their interpretation somehow easier than conventional skewness and kurtosis that can take arbitrarily large values. A test of the null of symmetry $H_0 : SK_L = 0$ relies on the squares of $\widehat{SK}_L$ (Harri and Coble, 2011):

$$\hat{\tau}_{3,L} = \frac{\widehat{SK}_L^2}{(0.1866T^{-1} + 0.8000T^{-2})} \xrightarrow{d} \chi^2_1 \tag{6}$$

Bootstrap inference. The asymptotic distributions of $\hat{\tau}_3$ and $\hat{\tau}_{3,L}$ depend on the assumption

[2]Symmetry tests of regression residuals also play an important role in macroeconomics and finance, but such analysis is beyond the scope of this article.

that the underlying data are iid. To accommodate serially correlated data generating processes (DGPs), we rely on a bootstrap approximation of the null sampling distribution of $\hat{\tau}_3$ and $\hat{\tau}_{3,L}$. Following Psaradakis (2003), we implement a symmetrized version of the sieve bootstrap of Bühlmann (1997) that has some advantages over alternative bootstrap schemes when small samples or linear DGPs are involved (Bühlmann, 2002; Psaradakis and Vávra, 2019). The assumption underlying the sieve bootstrap is that $\mathcal{X}$ can be described by the following DGP:

$$X_t - \mu = \sum_{j=1}^{\infty} \phi_j \left(X_{t-j} - \mu \right) + \varepsilon_t \tag{7}$$

where $\{\phi_j\}_{j=1}^{\infty}$ is a square-summable sequence and ε_t is an iid symmetrically distributed zero-mean random variable. This DGP encompasses a large set of stochastic processes, including Autoregressive Moving Average (ARMA) models. Notice that the symmetry of the distribution of X_t is implied by the symmetry of the distribution of the error term ε_t. Ensuring that the bootstrap pseudo-data satisfy symmetry is key to guarantee that $\hat{\tau}_3$ and $\hat{\tau}_{3,L}$ have reasonable power against the departures from the null hypothesis.

The sieve bootstrap approximates the DGP in Equation (7) with an AR(p) model, with p increasing at a slower rate than the sample size. In practice, denoting with $\mathcal{T}$ either $\hat{\tau}_3$ or $\hat{\tau}_{3,L}$, we implement the following algorithm to estimate its asymptotic distribution under the null hypothesis:

1. Select the order p an AR(p) model for a sample $\{X_t\}_{t=1}^{T}$ minimizing the Akaike Information Criterion in the range $1 \leq p \leq \lfloor \log(T)^2 \rfloor$;

2. Get least-squares estimates of the coefficients $\hat{\phi}_1, \ldots, \hat{\phi}_p$ of the AR(p) model $X_t - \hat{\mu} = \sum_{j=1}^{p} \phi_j \left(X_{t-j} - \hat{\mu} \right) + \varepsilon_t$ where $\hat{\mu}$ is the sample average of X_t;

3. Construct the residuals $\hat{\varepsilon}_t = (X_t - \hat{\mu}) - \sum_{j=1}^{p} \hat{\phi}_j \left(X_{t-j} - \hat{\mu} \right)$ for $t = p+1, \ldots, T$;

4. Draw a random sample $\{\varepsilon_t^*\}_{t=1}^T$ from the empirical distribution function of $\tilde{\varepsilon}_t$ where:

$$\tilde{\varepsilon}_t = \begin{cases} \hat{\varepsilon}_t & \text{if } t = p+1, \ldots, T \\ -\hat{\varepsilon}_t & \text{if } t = T+1, \ldots, 2T-p \end{cases}$$

5. Generate bootstrap replicates $\{X_t^*\}_{t=1}^T$ relying on:

$$X_t^* - \hat{\mu} = \sum_{j=1}^{p} \hat{\phi}_j \left(X_{t-j}^* - \hat{\mu}\right) + \varepsilon_t^*$$

6. Construct the bootstrap analog of $\mathcal{T}$, denoted as $\mathcal{T}^*$, applying a given test of symmetry to the bootstrap time series $\{X_t^*\}_{t=1}^T$;

7. Repeat steps 4-6 a large number of times to obtain a sample of size B of bootstrap symmetry tests, $\{\mathcal{T}_b^*\}_{b=1}^B$.

The empirical distribution of $\{\mathcal{T}_b^*\}_{b=1}^B$ is used as a bootstrap approximation of the null sampling distribution of $\mathcal{T}$.

3 Simulation study

3.1 Experimental design

Our main results focus on a simple AR(1) DGP with different degrees of persistence:

$$\mathcal{M}_1: \quad X_t = \rho X_{t-1} + \varepsilon_t \quad \text{for } \rho = 0.0, 0.5, 0.9 \tag{8}$$

To investigate the size of tests we rely on $\varepsilon_t \overset{iid}{\sim} N(0,1)$ and on symmetric parametrizations of the Generalized Lambda Family (GLF) of Ramberg and Schmeiser (1974). The GLF encompasses symmetric (S1-S3) and asymmetric (A1-A4) distributions. Symmetric distributions S1-S3 all have kurtosis in excess of the Normal distribution. Moreover, recall that the $\hat{\tau}_3$ test requires six finite moments, while S3, A2 and A4 possess at most five moments. All error distributions have been standardized to have zero mean and unit variance.

We consider three different sample sizes – 40, 160 and 480 – that would correspond to 40 years of yearly, quarterly or monthly data. We use 100 burn-in observations to minimize dependence of the AR process on initial conditions and rely on 199 bootstrap samples.

3.2 Size

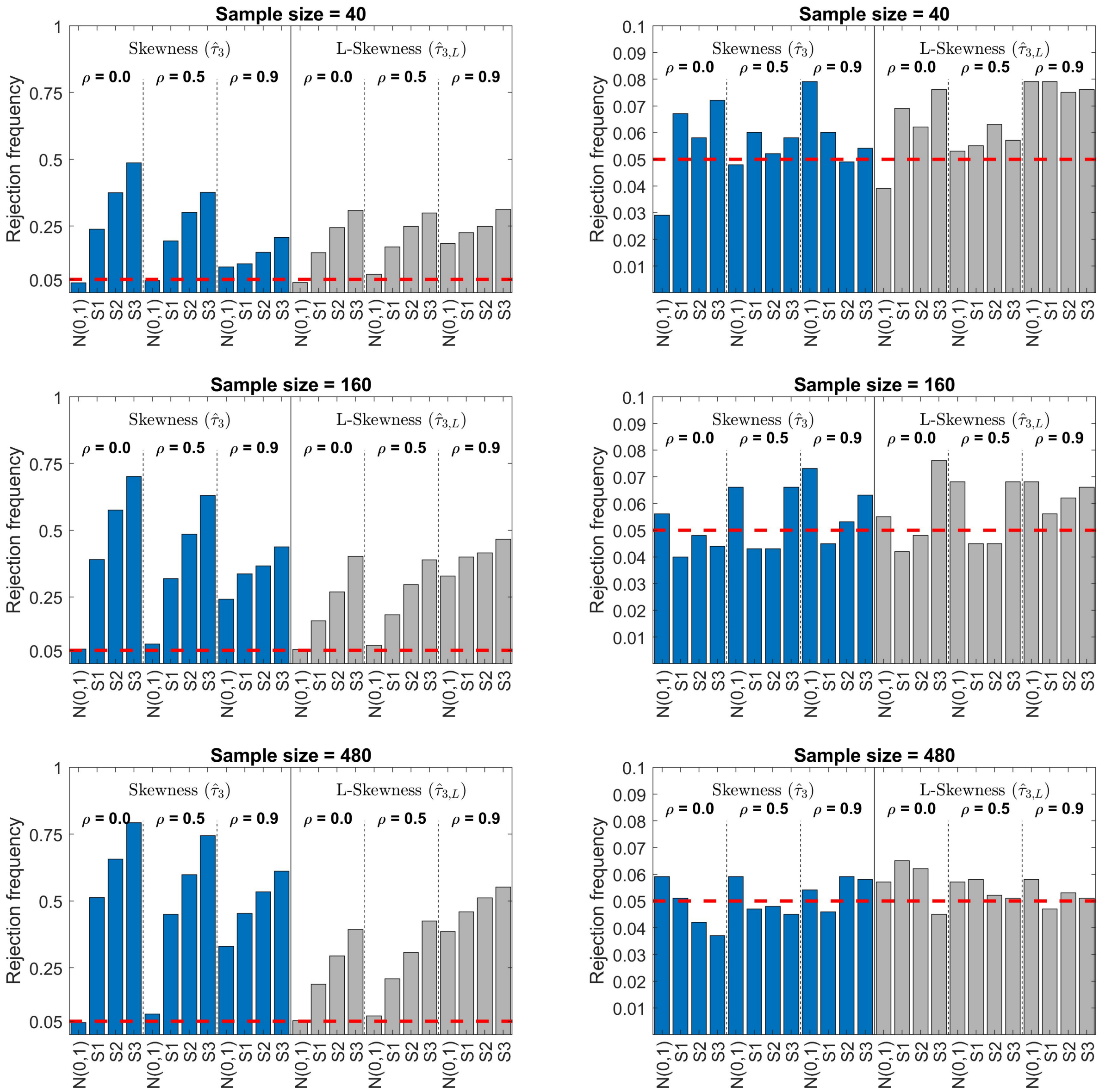
The size of the $\hat{\tau}_3$ and $\hat{\tau}_{3,L}$ tests is investigated comparing the Monte Carlo rejection frequency against the nominal size of the test, set to 5%. Figure 1a shows results based on the asymptotic distribution of tests. For the Normal distribution both tests have empirical size close to the nominal level only when the AR parameter does not exceed 0.5. In the remaining cases both tests are strongly oversized and their performance deteriorates as the degree of excess kurtosis (i.e. moving from S1 to S3) and/or the serial correlation increases. All in all, Figure 1a highlights that implementing the two tests of symmetry based on their asymptotic distribution is not advisable when data feature serial correlation or excess kurtosis.

In Figure 1b we investigate the size of tests when their null sampling distribution is approximated with the bootstrap. Both procedures now feature empirical rejection frequencies close to the nominal level. Once again, both tests tend to be slightly oversized as the degree of excess kurtosis and or the serial correlation increases. However, the empirical rejection frequency gets closer to 5% as the sample size increases.

3.3 Power

To study the power of tests we rely on four asymmetric distributions – A1 to A4 – with increasing degree of asymmetry and excess kurtosis. As documented in Section 3.2, both procedures are oversized when relying on their asymptotic distributions, therefore we present only results based on bootstrap critical values. Power analysis highlights that $\hat{\tau}_{3,L}$ is superior to $\hat{\tau}_3$ for what concerns its ability to detect asymmetries. Power gains from L-moments are highest in smaller samples and increase as the degree of excess kurtosis and asymmetry rise.

Figure 1: Size: empirical rejection frequency of tests of symmetry



(a) Asymptotic distribution

(b) Sieve bootstrap

Notes: the figure shows the empirical rejection frequency of symmetry tests for each sample size, symmetric distribution and autoregressive parameter, ρ. Blue bars denote the $\hat{\tau}_3$ test, while gray bars identify the $\hat{\tau}_{3,L}$ test. A well sized test should have empirical rejection frequency close to its nominal size, 0.05 (red dahsed line).

3.4 Further results

In Table 1 we analyse the behaviour of tests when data are generated with alternative DGPs.

To check whether the bootstrap procedure is capable of approximating other DGPs, we generate data from MA(1) and ARMA(1,1) models. Independently of the size of the sample and of the DGP, under normality both tests have empirical size close to the nominal level. Similarly, for both DGPs tests are correctly sized in samples of size 480 also under S1-S3. In samples of size 40 and 160 and under S1-S3 distributions, the MA(1) model is more challenging for the bootstrap approximation than the ARMA(1,1), especially as far as $\hat{\tau}_3$ is concerned. In fact, under both DGPs and independently of which symmetric distribution is used, $\hat{\tau}_{3,L}$ has empirical size closer to 5% than $\hat{\tau}_3$ when the sample size is 160. For these linear DGPs power analysis confirms that $\hat{\tau}_{3,L}$ has higher power than $\hat{\tau}_3$.

Next, we simulate data relying on a Generalized Autoregressive Conditionally Heteroskedastic (GARCH) model. Notice that this DGP does not admit the representation in Equation (7); moreover, in this case the distribution of X_t is symmetric and leptokurtic even when errors are normally distributed. While under normality of the errors both tests are well sized independently of sample size, their performance deteriorates moving from S1 to S3 particularly in smaller samples. Under S1-S3, $\hat{\tau}_{3,L}$ has empirical size closer to 5% than $\hat{\tau}_3$ in samples of size 160 or larger. Also in this case $\hat{\tau}_{3,L}$ has higher power than $\hat{\tau}_3$.

Lastly, we consider a Logistic Smooth Transition Autoregressive (LSTAR) DGP to further assess ability of the sieve bootstrap algorithm to identify asymmetries generated by nonlinear models. In this case, we limit the analysis to a DGP with normally distributed errors and show that, independently of sample size, $\hat{\tau}_{3,L}$ has higher power than $\hat{\tau}_3$.

All in all, Table 1 confirms that the sieve bootstrap delivers symmetry tests with appropriate empirical size when the sample size is moderately large. In the case of sample size equal to 40, we see that both tests are oversized when the error distribution is non-Gaussian. Moreover, we can notice that for $\hat{\tau}_{3,L}$ the empirical size is closer to the nominal 5% level than for $\hat{\tau}_3$.

4 The symmetry of business cycles of the G7 economies

A test of TR can be implemented by focusing on the symmetry of the distribution of the j-difference of a time series (DeLong and Summers, 1986). TR implies that the joint distributions of (X_t, X_{t-j}) and (X_{t-j}, X_t) are equal for all t and all $j = 1, 2, \ldots$ In the presence of TR, $\Delta^j X_t \equiv X_t - X_{t-j}$ has a symmetric distribution and hence $P(\Delta^j X_t > 0) = P(\Delta^j X_t < 0) = \frac{1}{2}$.

Table 1: Empirical rejection frequencies of tests of symmetry under alternative Data Generating Processes – Sieve Bootstrap

		Skewness, $\hat{\tau}_3$								L-Skewness, $\hat{\tau}_{3,L}$							
DGP	T	N(0,1)	S1	S2	S3	A1	A2	A3	A4	N(0,1)	S1	S2	S3	A1	A2	A3	A4
$\mathcal{M}_2$	40	0.0871	0.1612	0.2322	0.2503	0.4014	0.3303	0.6306	0.6787	0.0851	0.1361	0.1922	0.2172	0.4615	0.3313	0.8348	0.8969
	160	0.0430	0.1001	0.1231	0.1542	0.7978	0.5035	0.9109	0.9039	0.0541	0.0751	0.0921	0.1131	0.9339	0.6276	1.0000	1.0000
	480	0.0440	0.0591	0.0691	0.0551	0.9880	0.7768	0.9710	0.9620	0.0450	0.0450	0.0651	0.0561	1.0000	0.9620	1.0000	1.0000
$\mathcal{M}_3$	40	0.0621	0.1652	0.1732	0.2102	0.5686	0.3524	0.7908	0.7938	0.0601	0.1031	0.1251	0.1461	0.6657	0.3584	0.9770	0.9820
	160	0.0450	0.0721	0.0831	0.0741	0.9429	0.5676	0.9369	0.8849	0.0521	0.0681	0.0691	0.0831	1.0000	0.7708	1.0000	1.0000
	480	0.0470	0.0551	0.0450	0.0430	0.9940	0.8378	0.9710	0.9499	0.0561	0.0581	0.0501	0.0581	1.0000	0.9980	1.0000	1.0000
$\mathcal{M}_4$	40	0.0601	0.1421	0.1882	0.2202	0.5596	0.3544	0.8368	0.8208	0.0631	0.1031	0.1351	0.1542	0.6517	0.3423	0.9720	0.9730
	160	0.0661	0.1151	0.1121	0.1211	0.9289	0.5666	0.9540	0.9219	0.0531	0.0741	0.0731	0.0851	0.9870	0.7307	0.9990	1.0000
	480	0.0611	0.0751	0.0751	0.0911	0.9800	0.7828	0.9750	0.9720	0.0551	0.0541	0.0641	0.0731	1.0000	0.9770	1.0000	1.0000
$\mathcal{M}_5$	40	0.4605	–	–	–	–	–	–	–	0.4905	–	–	–	–	–	–	–
	160	0.9129	–	–	–	–	–	–	–	0.9550	–	–	–	–	–	–	–
	480	1.0000	–	–	–	–	–	–	–	1.0000	–	–	–	–	–	–	–

Notes: Data Generating Processes (DGPs) are as follows: $\mathcal{M}_2$ is a MA(1) model; $\mathcal{M}_3$ is an ARMA(1,1) model; $\mathcal{M}_4$ is a GARCH(1,1) model; $\mathcal{M}_5$ is a LSTAR(2) model. The distribution of data is symmetric under $\mathcal{M}_2 - \mathcal{M}_4$ when error terms are either Normally distributed or follow one the S1-S3 distributions. In these cases a well sized test should have empirical rejection frequency close to its nominal size, 0.05 in our analysis. The distribution of data is asymmetric under $\mathcal{M}_2 - \mathcal{M}_4$ when error terms follows one of the A1-A4 distributions and under $\mathcal{M}_5$ even if the errors are Normally distributed. In these cases, higher rejection frequencies indicate higher power.

Table 2: Symmetry test - real GDP

	Skewness ($\hat{\tau}_3$)			L-Skewness ($\hat{\tau}_{3,L}$)		
Country	$\widehat{SK}$	$j=1$	$j=1,\ldots,4$	$\widehat{SK}_L$	$j=1$	$j=1,\ldots,4$
Canada	-0.2182	0.4645	0.1041	-0.0286	0.4975	0.3203
France	-0.4897	0.1982	0.4324	-0.0016	0.9710	1.0000
Germany	-0.6645	0.3914	0.0841	-0.0121	0.8068	1.0000
Italy	-0.7051	0.0541	0.0000	-0.0199	0.7157	0.1522
Japan	-1.6380	0.0581	0.0000	-0.1421	0.0851	0.0080
UK	0.1448	0.7838	0.0721	-0.0372	0.5155	0.1001
US	-0.2985	0.4925	0.3844	-0.0425	0.3934	0.4765

Notes: the table shows sample skewness ($\widehat{SK}$) and L-skewness ($\widehat{SK}_L$) for Δy_t and in columns 3-4, 5-7 the p-values of tests of symmetry. The null hypothesis of the test is that the distribution of $\Delta^j x_t = \ln(GDP_t/GDP_{t-j})$ is symmetric. We report the Bonferroni p-value for the joint null hypothesis that the distributions of $\Delta^j x_t$ for $j = 1,\ldots,4$ are symmetric.

The concept of TR has been widely used to investigate the so-called Mitchell–Keynes hypothesis, which posits that expansions are more gradual than recessions (see e.g. Neftçi, 1984; DeLong and Summers, 1986). We focus on quarterly real GDP for the G7 economies over the period 1970:Q1-2019:Q4.[3] We analyse whether $\Delta^j x_t = \ln(GDP_t/GDP_{t-j})$ for $j = 1,\ldots,4$ is TR. Notice that in this application the sample size ranges from 104 to 200, therefore the small sample size distortions highlighted in the Monte Carlo analysis should be largely attenuated.

Table 2 shows that, consistently with the Mitchell–Keynes hypothesis, the sample skewness of real GDP growth is negative for all countries except UK. Interestingly, if we omit the single largest observation in absolute value, $\widehat{SK}$ turns negative also for UK, which is expected, given that the low degree of resistance to outliers is a feature of conventional moments (see e.g. Bastianin, 2020). On the contrary, $\widehat{SK}_L$ for UK preserves its sign even when omitting the largest observation in absolute value.

In Table 2 we also present the p-values of the $\hat{\tau}_3$ and $\hat{\tau}_{3,L}$ test for Δx_t and the Bonferroni p-values for the joint null hypothesis that the distributions of $\Delta^j x_t$ for $j = 1,\ldots,4$ are symmetric. We have consistent evidence against symmetry only for Japan. On the other hand, we can reject the null of symmetry for Italy when using the $\hat{\tau}_3$ test, but not when relying on $\hat{\tau}_{3,L}$. A robustness analysis using Industrial Production data provides evidence against

[3]For Italy and Japan we have shorter sample periods.

TR for the Japanese and US business cycles. All in all, the Mitchell–Keynes hypothesis seems to be strongly supported only for Japan.

5 Conclusions

Our Monte Carlo simulations show that symmetry tests based on L-moments have better size properties and more power than tests based on sample skewness. Tests of symmetry can be applied to serially correlated and persistent time series, provided that their distributions under the null hypothesis is approximated with an an appropriate bootstrap algorithm. In fact, asymptotic results for iid data cannot applied to time series data in that they yield badly sized tests. A byproduct our paper extends the results of Psaradakis (2003) showing that the symmetrized version of the sieve bootstrap works well also for tests based on L-moments. Our results highlight that when the DPGs are symmetric, but heavily depart from the linear Gaussian case, the sieve bootstrap delivers tests with correct empirical size in moderately large samples. In small samples both tests tend to be oversized, but the procedure based on L-moments much less so.

Acknowledgments

We thank an anonymous referee and the editor for constructive and helpful comments. Usual disclaimers apply.

References

Bastianin, A. (2020). Robust measures of skewness and kurtosis for macroeconomic and financial time series. *Applied Economics*, 52(7):637–670.

Beare, B. K. and Seo, J. (2014). Time Irreversible Copula-Based Markov Models. *Econometric Theory*, 30(5):923–960.

Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli*, 3(2):123–148.

Bühlmann, P. (2002). Bootstraps for time series. *Statistical science*, 17(1):52–72.

DeLong, J. B. and Summers, L. H. (1986). Are business cycles symmetrical. In Gordon, R. J., editor, *The American Business Cycle: Continuity and Change*, pages 166–178. University of Chicago Press.

Fong, W. M. (2003). Time reversibility tests of volume–volatility dynamics for stock returns. *Economics Letters*, 81(1):39–45.

Harri, A. and Coble, K. H. (2011). Normality testing: two new tests using L-moments. *Journal of Applied Statistics*, 38(7):1369–1379.

Hosking, J. R. (1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society. Series B*, pages 105–124.

Kilian, L. and Demiroglu, U. (2000). Residual-based tests for normality in autoregressions: asymptotic theory and simulation evidence. *Journal of Business & Economic Statistics*, 18(1):40–50.

Neftçi, S. N. (1984). Are economic time series asymmetric over the business cycle? *Journal of Political Economy*, 92(2):307–328.

Psaradakis, Z. (2003). A bootstrap test for symmetry of dependent data based on a Kolmogorov–Smirnov type statistic. *Communications in Statistics-Simulation and Computation*, 32(1):113–126.

Psaradakis, Z. and Vávra, M. (2019). Bootstrap-assisted tests of symmetry for dependent data. *Journal of Statistical Computation and Simulation*, 89(7):1203–1226.

Ramberg, J. S. and Schmeiser, B. W. (1974). An approximate method for generating asymmetric random variables. *Communications of the ACM*, 17(2):78–82.