NÁRODNÁ BANKA SLOVENSKA
EUROSYSTÉM

Testing the Validity of Assumptions of UC-ARIMA Models for Trend-Cycle Decompositions

Marián Vávra

Working Paper 4/2016

www.nbs.sk
Imricha Karvaša 1
813 25 Bratislva

research@nbs.sk

September 2016

ISSN 1337-5830

Testing the Validity of Assumptions of UC-ARIMA Models for Trend-Cycle Decompositions[1]

Working paper NBS

Marián Vávra[2]

Abstract

This article tests the validity of underlying assumptions (i.e. linearity and normality) of UC-ARIMA models for trend-cycle decompositions using macroeconomic variables from 16 OECD countries. Clear and overwhelming evidence of non-normality and non-linearity is found. Our results thus cast doubts on the adequacy of the filtered cyclical component from this type of model.

JEL classification: C12; C22; E32.
Key words: Normality; Lobato-Velasco test; Linearity; Portmanteau Q test; Trend-cycle decomposition; UC-ARIMA models.

[1] I would like to thank Ron Smith and Simon van Norden for useful comments and interesting suggestions. All remaining errors are only mine.

[2] Marián Vávra, Research Department of the NBS.

1. INTRODUCTION

The decomposition of economic variables (e.g. real GDP) into trend and cycle components plays a fundamental role in economics. Although various methods have been developed in the literature, Gaussian linear unobserved component ARIMA (UC-ARIMA) models of Harvey (1985) have become very popular (see also Watson (1986), Clark (1987), Harvey and Jaeger (1993), Kuttner (1994), Orphanides and van Norden (2002), Morley, Nelson, and Zivot (2003), Harvey and Trimbur (2003)).[3] However, there is no guarantee (in general) that Gaussian linear models can generate the kind of stochastic properties that a particular economic series exhibits. A model misspecification can, in turn, give rise to misleading inference from both economic and econometric standpoints. Examples include:

1. The improperly filtered cyclical component (the output gap in case of real GDP) can provide noisy information about the state of the economy, something which is of much practical importance for central bankers when setting the policy rates (see, e.g., Taylor (1993)).

2. The Gaussian linear output gap estimates can provide misleading information about the degree of business cycle coherence (measured, for instance, by the Pearson correlation coefficient), something which represents a fundamental problem for any monetary union since if the business cycle movements of the member states are actually not sufficiently coherent, then the common monetary policy will not be optimal for all members within the union (see, e.g., Frankel and Rose (1998) and De Haan, Inklaar, and Jong-A-Pin (2008)).

3. Overlooking the distributional aspects of economic variables can lead to output gap estimates which might be systematically biased due to a potential bias of the steady-state growth of real GDP.

4. It is a well known fact that the Kalman filter, used in UC-ARIMA models for updating the likelihood function and filtering the unobserved states, is optimal among all linear filters when the noise processes are Gaussian (see Anderson and Moore (1979, Chapter 5)). However, this is not generally true for non-linear non-Gaussian models (see Harvey (1991, Chapter 3.7), for which more complicated (e.g. Monte Carlo) filters might be preferred.[4]

In contrast, non-linear time series models can capture empirically observed phenomena (e.g. business cycle asymmetry, conditional volatility, etc.) without breaking theoretical concepts or imposing unrealistic assumptions (see, e.g., Proietti (1998), Kim and Nelson (1999), Kuan, Huang, and Tsay (2005), or Sinclair (2010)). On the other hand, all modelling steps (i.e. model selection, identification, estimation, bias correction) of non-linear models are far more complex

[3]The main reason why this class of models has become so popular in practice is that, after imposing some restrictions, the UC-ARIMA model encompasses other routinely used methods for a trend-cycle decomposition such as a linear trend method or a mechanical Hodrick-Prescott filter (see Harvey and Jaeger (1993, p. 233) for details).

[4]It might be worth noting that the Kalman filter also provides the optimal MSE estimates of the unobserved states for non-linear non-normal processes but these estimates are typically not the maximum likelihood ones.

and complicated as compared to linear counterparts (see, e.g., Boldin (1996), Breunig and Pagan (2004), or Psaradakis (1998)).[5] Therefore, using appropriate testing procedures is often desirable in order to establish the adequacy or otherwise of a Gaussian linear data representation before exploring more complicated (non-linear) models. The aim of this note is to fill the gap in the literature and test the validity of two key assumptions (i.e. linearity and normality) of UC-ARIMA models using a set of OECD macroeconomic variables.

The paper is organized as follows. Section 2 briefly describes UC-ARIMA models used for trend-cycle decompositions in the literature. Section 3 explains two test statistics used to assess various hypotheses of interest. The empirical results for OECD macroeconomic variables are presented in Section 4. Section 5 summarizes and concludes.

2. Unobserved Component ARIMA Models

Following the UC-ARIMA model of Harvey (1985), each observed economic variable (y) is decomposed into a trend component (μ) and a cyclical component (ψ) according to

$$y_t = \mu_t + \psi_t + a_t, \tag{1}$$

where $a_t \sim$ NID$(0, \sigma_a^2)$ is a noise component (e.g., due to measurement errors). The trend component is assumed to be defined as

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \tag{2a}$$
$$\beta_t = \beta_{t-1} + \zeta, \tag{2b}$$

where β_t is a (stochastic) slope of the trend component, $\eta_t \sim$ NID$(0, \sigma_\eta^2)$ and $\zeta_t \sim$ NID$(0, \sigma_\zeta^2)$.[6] The cyclical component is given by

$$\psi_t = \rho\cos(\lambda)\psi_{t-1} + \rho\sin(\lambda)\psi^*_{t-1} + u_t, \tag{3a}$$
$$\psi^*_t = -\rho\cos(\lambda)\psi_{t-1} + \rho\sin(\lambda)\psi^*_{t-1} + u^*_t, \tag{3b}$$

where $0 \leq \rho \leq 1$ denotes a damping factor, λ is the frequency of the (business) cycle (in radians), $u_t \sim$ NID$(0, \sigma_u^2)$ and $u^*_t \sim$ NID$(0, \sigma_u^{*2})$. The error terms are assumed to be mutually independent.[7] The interested reader is referred to Watson (1986), Clark (1987), Harvey and

[5] Additionally, the results of Kim and Nelson (1999, p. 325) and Sinclair (2010, p. 13), among others, clearly show that non-linear models can generate output gap estimates which might be difficult to justify.

[6] Some authors use an autoregressive representation instead of a random walk for the stochastic slope in (2b) (see, e.g., Proietti, Musso, and Westermann (2007)). This specification has no effect on our results.

[7] Note that although this assumption can be weakened it helps to identify the model (see Morley, Nelson, and Zivot (2003) for an example).

Jaeger (1993), Kuttner (1994), Orphanides and van Norden (2002), Morley, Nelson, and Zivot (2003), Harvey and Trimbur (2003), among others, for various modifications of linear Gaussian UC-ARIMA models used for a trend-cycle decomposition of economic variables.

It can be easily shown that the model in (1)–(3) can be given, after appropriate differencing (depending on the stochastic trend specification), a stationary **Gaussian linear** finite-order ARMA representation. For example, the model in (1)–(3) reduces, after double differencing, to an ARMA(2,4) process (see Harvey (1985, p. 220)). Without loss of generality (and for the purpose of hypothesis testing discussed in Section 3), we restrict our attention to the model written in the following form

$$x_t = c + \sum_{j=0}^{\infty} \psi_j(\boldsymbol{\delta})\, \varepsilon_{t-j}, \qquad t \in \mathbb{Z}, \tag{4}$$

$\{\psi_j(\boldsymbol{\delta})\}$ is an absolutely summable sequence of weights, assumed to be known functions of a finite-dimensional vector $\boldsymbol{\delta}$ of unknown parameters, $\{\varepsilon_t\}$ is strictly stationary white noise.

3. Testing for Normality and Linearity

The main objective of this article is to test the hypothesis that $\{x_t\}$ in (4) is a **linear** stochastic process with one-dimensional **Gaussian** marginal distribution F, that is

$$H_0^{NL} : \{x_t\} \sim \text{ linear with } F(x) = N(c, \sigma^2), \text{ for all } x \in \mathbb{R}, \tag{5}$$

where $N(c, \sigma^2)$ denotes a normal distribution with mean c and variance σ^2.[8]

Two statistics are applied to test the joint hypothesis H_0^{NL}: the Lobato-Velasco normality test and the generalized portmanteau test.

Testing for normality: Lobato and Velasco (2004) proposed a simple statistic for testing normality based on the sample coefficients of skewness ($\hat{\tau}$) and kurtosis ($\hat{\kappa}$). The LV statistic takes the form as follows

$$LV = n\left(\frac{\hat{\tau}^2}{6\hat{G}_3} + \frac{(\hat{\kappa}-3)^2}{24\hat{G}_4}\right) \xrightarrow{d} \chi^2(2), \tag{6}$$

where $\hat{\tau} = n^{-1}\sum_{t=1}^{n}[(X_t - \hat{\mu})/\hat{\sigma}]^3$ and $\hat{\kappa} = n^{-1}\sum_{t=1}^{n}[(X_t - \hat{\mu})/\hat{\sigma}]^4$ are sample coefficients of skewness and kurtosis calculated from $\mathbf{X}_n$, and $\hat{G}_k = \sum_{j=1-n}^{n-1}\hat{\rho}_j^k$, for $k \in \{3, 4\}$, with $\hat{\rho}_j$ is the estimated autocorrelation at lag j calculated from $\mathbf{X}_n$. The main advantage of the LV test is the estimation of the long-run variances of skewness and kurtosis which (in contrast to other

[8] The alternative hypothesis is set in an obvious way.

tests such as Bai and Ng (2005)) does not involve any bandwidth selection.[9]

Testing for linearity: Psaradakis and Vávra (2016) modified the portmanteau Q statistic for testing linearity of the stochastic process in (4).[10] When an estimator $\hat{\boldsymbol{\delta}}$ of $\boldsymbol{\delta}$ is available, one may use residuals $\{\hat{\varepsilon}_t\}$ in place of the unobservable noise $\{\varepsilon_t\}$. The weak form of the i.i.d. property can be tested using the generalized correlations of the residuals defined for lag k and $(r, s) \in \{(1, 2), (2, 1), (2, 2)\}$ as follows

$$\hat{\rho}_{rs}(k) = \frac{\hat{\gamma}_{rs}(k)}{\sqrt{\hat{\gamma}_{rr}(0)\hat{\gamma}_{ss}(0)}}, \qquad k = 0, \pm 1, \dots, \pm(n-1),$$

where $\hat{\gamma}_{rs}(k) = n^{-1}\sum_{t=1}^{n-k} f_r(\hat{\varepsilon}_t)f_s(\hat{\varepsilon}_{t+k})$ for $k \geqslant 0$, $\hat{\gamma}_{rs}(k) = \hat{\gamma}_{sr}(-k)$ for $k < 0$, and $f_b(\xi_t) = \xi_t^b - n^{-1}(\xi_1^b + \cdots + \xi_n^b)$ for any collection of random variables $\{\xi_t\}$ and $b \in \mathbb{N}$. The null of linearity implies that $\rho_{rs}(k) = 0$ for all $k \neq 0$ and given (r, s) parameters which can be formally tested using the portmanteau statistic defined as

$$Q_{rs}(m) = n \sum_{k=1}^{m} \hat{\rho}_{rs}^2(k) \xrightarrow{d} \chi^2(m), \tag{7}$$

where the integer $m << n$. The main advantage of a triplet of the Q_{rs} tests (compared to other non-linearity tests such as the Tsay test, the ARCH test, or the BDS test.) is that it has power against a large variety of non-linear models such as regime-switching models, bilinear models, or conditional volatility models (see Psaradakis and Vávra (2016) for details).

When the null hypothesis of normality and linearity is tested using, for instance, the quaternion of tests (i.e. one LV test and three Q_{rs} tests), the overall probability of Type I error (i.e. the probability that one or more tests lead to a false rejection of the null) is inflated. Some P-value adjustment is thus desirable in order to avoid the problem of spurious inference (see Psaradakis (2000)).[11] A simple adjustment for multiple testing based on Simes (1986) is implemented here: Let $P_{(1)} \leq P_{(2)} \leq P_{(3)} \leq P_{(4)}$ denote the ordered quaternion of (asymptotic) P-values associated with the set of test statistics under consideration. Multiplicity-adjusted P-values are then calculated as $\tilde{P}_{(i)} = \min\{4P_{(i)}/i, 1\}$, $i \in \{1, 2, 3, 4\}$ and the joint null hypothesis H_0^{NL} of normality and linearity is rejected at overall level $\alpha \in (0, 1)$ if $\min_{i\in\{1,2,3,4\}} \tilde{P}_{(i)} \leq \alpha$.

Note that as a byproduct of the above testing procedure, two individual hypotheses can be set and tested as well: (i) The process $\{x_t\}$ in (4) has the one-dimensional Gaussian marginal distribution F, that is $H_0^N : F(x) = N(c, \sigma^2), \text{ for all } x \in \mathbb{R}$; (ii) The process $\{x_t\}$ in (4) is a linear

[9] Note that using the original Jarque-Bera statistic (see Jarque and Bera (1980)) is inappropriate here due to dependence of observations.

[10] A stochastic process $\{x_t\}$ is typically characterized as linear if it admits the moving-average representation (4) with $\{\varepsilon_t\}$ being independent and identically distributed (i.i.d.) random variables. This is the notion of linearity considered by McLeod and Li (1983), Lawrance and Lewis (1985, 1987), Berg, Paparoditis, and Politis (2010), and Giannerini, Maasoumi, and Dagum (2015), among many others, and is the one adopted in this paper.

[11] Although there exist many adjusting methods in the literature (see Westfall and Young (1993, Chapter 2)), simple adjusting methods (e.g. the Simes method) work no worse than computationally expensive bootstrap methods even for mutually dependent statistics. See also Appendix A for the Monte Carlo results.

stochastic process, that is $H_0^L : \{x_t\} \sim \text{linear}$. The H_0^N hypothesis is tested by the LV statistic in (6) and the H_0^L is tested by the Q_{rs} statistics in (7) with the Simes adjusted P-values.

The finite sample properties of the LV and Q_{rs} tests for the null hypotheses of interest are examined via the Monte Carlo experiments. The results are reported in Appendix A.

4. Empirical Results

The LV test in (6) and the Q_{rs} test in (7) are applied to a set of 48 series – 3 macroeconomic variables for each of 16 OECD countries.[12] Attention is paid to real gross domestic product (denoted as GDP), industrial production (denoted as IP), and the unemployment rate (denoted as UR) – economic variables predominantly used for a trend-cycle decomposition in practice. Data from the following 16 OECD countries are employed in the study: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Italy, Japan, the Netherlands, Portugal, Spain, Sweden, Switzerland, the United Kingdom, United States. Most of the series span the period 1973q1 – 2014q4 (i.e. 168 observations).[13]

GDP and IP are transformed using both the first (log) differences and the second (log) differences, which corresponds to widely used specifications of the stochastic trend for non-stationary variables, whereas UR is transformed using the first differences only. The lag order of the AR-sieve, to obtain the residuals for computing the Q_{rs} tests, is taken to be the minimizer of the Hannan-Quinn information criterion over the range $1 \leq p \leq \lfloor 5 \log_{10} n \rfloor$, where $\lfloor \cdot \rfloor$ denotes the greatest-integer function.[14] The lag order m of the Q_{rs} test is equal to the selected lag order p of the AR-sieve.

The (Simes adjusted) P-values of the test statistics are reported in Table 1. The results suggest the following:

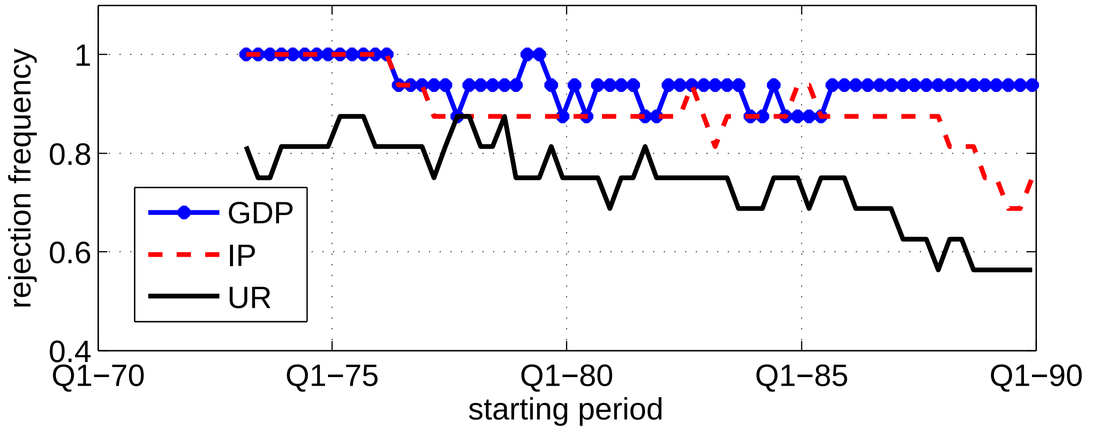
(i) The joint null hypothesis of normality and linearity H_0^{NL} is rejected at 0.05 nominal level in almost 93% of cases. No significant differences in rejecting the joint null are observed for $I(1)$ or $I(2)$ specifications of the stochastic trend of output variables.

(ii) From the individual null hypotheses about normality H_0^N and linearity H_0^L, it can be concluded that the presence of non-Gaussian stochastic features dominates non-linear features in the data. In particular, the normality hypothesis is rejected at 0.05 nominal level in almost 90% of cases whereas linearity in "only" 66% of cases. Higher rejection rates of the normality hypothesis are to be expected since non-linearity features often imply non-normality but not vice versa.

[12] The dataset is downloaded from the OECD Database.
[13] Note that observations after 2014 are not used in order to minimize the effect of statistical revisions on the results.
[14] Note that the HQ is a little bit more benevolent in determining the lag order of AR models as compared to the BIC. Additional lags may eliminate remaining serial correlation in residuals which is desirable when using neglected nonlinearity tests (see Lumsdaine and Ng (1999) for details).

(iii) It can be also concluded from the individual test results that there is no country in our set for which all selected variables can be considered as linear and/or Gaussian processes. This result confirms widespread evidence of non-linear and non-normal features in economic series in OECD countries.

Figure 1: Empirical Rejection Frequencies of Tests



Although the joint null hypothesis H_0^{NL} of normality and linearity is clearly rejected for almost all OECD macroeconomic variables, it may still be interesting to assess the stability of the test results over time. For this purpose, a recursive approach is applied here. The method is based on quarter-by quarter shortening of the full sample consisting of 168 observations (i.e. 1973q1–2014q4) to a sample of only 100 observations (i.e. 1990q1–2014q4). The LV and Q tests are then applied to each of 69 data-windows. The empirical rejection frequencies of the joint null hypothesis at the nominal level 0.05 for the macroeconomic variables with the $I(1)$ configuration of the stochastic trend are reported in the graphical form in Figure 1.[15]

It can be concluded from the results that the empirical rejection frequencies are reasonably stable over time for all three macroeconomic variables – some power loss is to be expected as a result of shortening the sample from 168 to only 100 observations (see Table 3 for Monte Carlo evidence). Stronger and more stable evidence of non-normality and non-linearity is found for real GDP and IP than for UR. Stability of the empirical rejection frequencies clearly indicates that non-linearity and non-normality seem to be a characteristic features of this type of macroeconomic variables.

[15]The rejection frequency is calculated as $\frac{1}{16}\sum_{i=1}^{16} I(\tilde{\alpha}_i \leq 0.05)$, where $\tilde{\alpha}_i$ is the overall Simes adjusted P-value obtained for the i-th country in a given time-window, and $I(\cdot)$ is an indicator function. So, the rejection frequency equals to 1.0 means that the procedure rejects the joint null hypothesis of normality and linearity in all countries in a given data-window.

Table 1: (Simes Adjusted) P-values of LV and Q Tests

	H_0^{NL} : Normality and Linearity					H_0^{N} : Normality					H_0^{L} : Linearity				
	$I(1)$			$I(2)$		$I(1)$			$I(2)$		$I(1)$			$I(2)$	
	GDP	IP	UR	GDP	IP	GDP	IP	UR	GDP	IP	GDP	IP	UR	GDP	IP
Australia	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.56	0.80	0.69	0.00
Austria	0.00	0.01	0.78	0.00	0.12	0.00	0.01	0.42	0.00	0.38	0.00	0.01	0.59	0.00	0.08
Belgium	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.48	0.00	0.04	0.44	0.70
Canada	0.00	0.00	0.00	0.00	0.90	0.00	0.00	0.00	0.00	0.90	0.11	0.43	0.00	0.25	0.86
Denmark	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.00	0.00	0.02	0.30	0.00	0.00	0.24
Finland	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.06	0.00	0.00
France	0.00	0.00	0.14	0.00	0.00	0.00	0.00	0.26	0.00	0.00	0.43	0.00	0.10	0.86	0.00
Italy	0.00	0.00	0.20	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.00	0.00	0.15	0.00	0.00
Japan	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
Netherlands	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.34	0.00	0.00	0.51
Portugal	0.00	0.05	0.00	0.00	0.23	0.46	0.01	0.00	0.00	0.06	0.00	0.67	0.41	0.00	0.75
Spain	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sweden	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Switzerland	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.25	0.19	0.11	0.05
United Kingdom	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.01	0.04	0.00	0.02
United States	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00

5. Conclusion

This article has focused on testing the validity of underlying assumptions (i.e. normality and linearity) of UC-ARIMA models using OECD macroeconomic variables. Clear and overwhelming evidence of non-normality and non-linearity is found in the vast majority of OECD indicators. Our results thus cast doubts on the adequacy of routinely used Gaussian linear UC-ARIMA models for a trend-cycle decomposition. We are of the opinion that our results call for implementing simple, yet flexible, non-linear non-Gaussian models accompanied by appropriate economic restrictions.

References

ANDERSON, B., AND J. MOORE (1979): "Optimal Filtering," *Prentice-Hall Information and System Sciences Series*.

BAI, J., AND S. NG (2005): "Tests for skewness, kurtosis, and normality for time series data," *Journal of Business and Economic Statistics*, 23, 49–60.

BERG, A., E. PAPARODITIS, AND D. N. POLITIS (2010): "A bootstrap test for time series linearity," *Journal of Statistical Planning and Inference*, 140, 3841–3857.

BOLDIN, M. (1996): "A check on the robustness of Hamiltons Markov switching model approach to the economic analysis of the business cycle," *Studies in Nonlinear Dynamics and Econometrics*, 1, 35–46.

BREUNIG, R. V., AND A. PAGAN (2004): "Do Markov-switching models capture nonlinearities in the data?: Tests using nonparametric methods," *Mathematics and computers in simulation*, 64, 401–407.

CLARK, P. (1987): "The cyclical component of US economic activity," *The Quarterly Journal of Economics*, 102, 797–814.

DE HAAN, J., R. INKLAAR, AND R. JONG-A-PIN (2008): "Will business cycles in the euro area converge? A critical survey of empirical research," *Journal of Economic Surveys*, 22, 234–273.

FRANKEL, J. A., AND A. ROSE (1998): "The endogenity of the optimum currency area criteria," *The Economic Journal*, 108, 1009–1025.

GIANNERINI, S., E. MAASOUMI, AND E. DAGUM (2015): "Entropy testing for nonlinear serial dependence in time series," *Biometrika*, 102, 661–675.

HAMILTON, J. (2005): "What's real about the business cycle?," *NBER Working Paper*, 11161.

HARVEY, A. (1985): "Trends and cycles in macroeconomic time series," *Journal of Business & Economic Statistics*, 3, 216–227.

——— (1991): *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

HARVEY, A., AND A. JAEGER (1993): "Detrending, stylized facts and the business cycle," *Journal of applied econometrics*, 8, 231–247.

HARVEY, A., AND T. TRIMBUR (2003): "General model-based filters for extracting cycles and trends in economic time series," *Review of Economics and Statistics*, 85, 244–255.

JARQUE, C., AND A. BERA (1980): "Efficient tests for normality, homoscedasticity and serial independence of regression residuals," *Economics Letters*, 6, 255–259.

KIM, C., AND C. NELSON (1999): "Friedman's plucking model of business fluctuations: tests and estimates of permanent and transitory components," *Journal of Money, Credit and Banking*, 31, 317–334.

KUAN, C., Y. HUANG, AND R. TSAY (2005): "An unobserved-component model with switching permanent and transitory innovations," *Journal of Business & Economic Statistics*, 23, 443–454.

KUTTNER, K. (1994): "Estimating potential output as a latent variable," *Journal of Business & Economic Statistics*, 12, 361–368.

LAWRANCE, A., AND P. LEWIS (1985): "Modelling and residual analysis of nonlinear autoregressive time series in exponential variables," *Journal of the Royal Statistical Society*, 47, 165–202.

——— (1987): "Higher-order residual analysis for nonlinear time series with autoregressive correlation structures," *International Statistical Review*, 55, 21–35.

LOBATO, I., AND C. VELASCO (2004): "A simple test of normality for time series," *Econometric Theory*, 20, 671–689.

LUMSDAINE, R., AND S. NG (1999): "Testing for ARCH in the presence of a possibly misspecified conditional mean," *Journal of Econometrics*, 93, 257–279.

MCLEOD, A., AND W. LI (1983): "Diagnostic checking ARMA time series models using squared-residual autocorrelations," *Journal of Time Series Analysis*, 4, 269–273.

MORLEY, J., C. NELSON, AND E. ZIVOT (2003): "Why are the Beveridge-Nelson and unobserved-components decompositions of GDP so different?," *Review of Economics and Statistics*, 85, 235–243.

ORPHANIDES, A., AND S. VAN NORDEN (2002): "The unreliability of output-gap estimates in real time," *Review of Economics and Statistics*, 84, 569–583.

PROIETTI, T. (1998): "Characterizing asymmetries in business cycles using smooth-transition structural time-series models," *Studies in Nonlinear Dynamics & Econometrics*, 3, 141–156.

PROIETTI, T., A. MUSSO, AND T. WESTERMANN (2007): "Estimating potential output and the output gap for the euro area: a model-based production function approach," *Empirical Economics*, 33, 85–113.

PSARADAKIS, Z. (1998): "Bootstrap-based evaluation of markov-switching time series models," *Econometric Reviews*, 17, 275–288.

——— (2000): "P-value adjustments for multiple tests for nonlinearity," *Studies in Nonlinear Dynamics & Econometrics*, 4, 1–14.

PSARADAKIS, Z., AND M. VÁVRA (2016): "Portmanteau tests for linearity of stationary time series," *National Bank of Slovakia Working Paper*, 1/2016.

SIMES, R. J. (1986): "An improved Bonferroni procedure for multiple tests of significance," *Biometrika*, 73, 751–754.

SINCLAIR, T. (2010): "Asymmetry in the business cycle: Friedman's plucking model with correlated innovations," *Studies in Nonlinear Dynamics & Econometrics*, 14, 1–31.

TAYLOR, J. (1993): "Discretion versus policy rules in practice," *Carnegie-Rochester conference series on public policy*, 39, 195–214.

WATSON, M. (1986): "Univariate detrending methods with stochastic trends," *Journal of monetary economics*, 18, 49–75.

WESTFALL, P. H., AND S. S. YOUNG (1993): *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. John Wiley & Sons, New York.

A. Simulation Study

In this section, we present and discuss the results of a simulation study examining the small-sample properties of the LV and Q_{rs} tests under various data-generating mechanisms. The following linear and non-linear data-generating processes (DGPs) are used in the simulations:

M1: $X_t = 0.5X_{t-1} + \varepsilon_t,$

M2: $X_t = 0.7X_{t-1} - 0.3\varepsilon_{t-1} + \varepsilon_t,$

M3: $X_t = 0.5X_{t-1} - 0.3X_{t-1}\varepsilon_{t-1} + \varepsilon_t,$

M4: $X_t = -0.5X_{t-1}I(X_{t-1} \leqslant 1) + 0.4X_{t-1}I(X_{t-1} > 1) + \varepsilon_t,$

M5: $X_t = 0.5X_{t-1} + \sigma_t\varepsilon_t, \quad \ln\sigma_t^2 = 0.01 + 0.3\{|\varepsilon_{t-1}| - \mathbb{E}(|\varepsilon_{t-1}|)\} - 0.8\varepsilon_{t-1} + 0.9\ln\sigma_{t-1}^2,$

$\{\varepsilon_t\}$ are i.i.d. zero-mean random variables with unit variance. The distribution of ε_t is either Gaussian (labeled N in what follows) or a member of generalized lambda distributions having inverse distribution function $F_\varepsilon^{-1}(u) = \beta_1 + \beta_2^{-1}\{u^{\beta_3} - (1-u)^{\beta_4}\}$; the parameter values used in the experiments are taken again from Bai and Ng (2005) and can be found in Table 2. The distributions S1 – S2 are symmetric (yet leptokurtic), whereas A1 – A2 are asymmetric.

For each design point, 1000 independent realizations of $\{X_t\}$ of length $100 + n$, with $n \in \{100, 200\}$, are generated. The first 100 data points of each realization are then discarded in order to eliminate start-up effects and the remaining n data points are used to compute the value of the test statistics of interest. The lag order is taken to be the minimizer of the Hannan-Quinn information criterion over the range $1 \leq p \leq \lfloor 5\log_{10} n \rfloor$, where $\lfloor\cdot\rfloor$ denotes the greatest-integer function. The lag order m of the Q_{rs} test is equal to the selected lag order p of the AR-sieve.

The Monte Carlo rejection frequencies of the above defined set of hypotheses (i.e. H_0^{NL}, H_0^N, and H_0^L, see Section 3 for details) using the LV and Q tests at nominal level 0.05 are reported in Table 3. The results clearly suggest that the test statistics have good size and power properties for all hypotheses of interest even in the smallest sample considered.

Table 2: Noise Distributions

	λ_1	λ_2	λ_3	λ_4	skewness	kurtosis
N	–	–	–	–	0.0	3.0
S1	0.000000	-1.000000	-0.080000	-0.080000	0.0	6.0
S2	0.000000	-0.397912	-0.160000	-0.160000	0.0	11.6
A1	0.000000	-1.000000	-0.007500	-0.030000	1.5	7.5
A2	0.000000	-1.000000	-0.100900	-0.180200	2.0	21.1

Table 3: Empirical Rejection Frequencies of Normality and Linearity Tests

		H_0^{NL} : Normality and Linearity					H_0^N : Normality					H_0^L : Linearity				
sample	distr.	M1	M2	M3	M4	M5	M1	M2	M3	M4	M5	M1	M2	M3	M4	M5
$n = 100$	N	0.04	0.04	0.79	0.74	0.98	0.03	0.03	0.38	0.04	0.95	0.04	0.04	0.78	0.76	0.83
	S1	0.31	0.31	0.93	0.84	0.98	0.37	0.40	0.69	0.47	0.97	0.02	0.03	0.91	0.75	0.68
	S2	0.53	0.54	0.96	0.90	0.99	0.60	0.59	0.79	0.67	0.98	0.03	0.03	0.94	0.76	0.62
	A1	0.69	0.69	0.90	0.93	0.95	0.80	0.79	0.66	0.88	0.92	0.04	0.03	0.86	0.47	0.45
	A2	0.67	0.66	0.94	0.94	0.96	0.74	0.75	0.76	0.81	0.96	0.03	0.03	0.91	0.59	0.43
$n = 200$	N	0.05	0.04	0.99	0.98	1.00	0.04	0.04	0.66	0.04	1.00	0.05	0.03	0.99	0.99	0.97
	S1	0.55	0.54	1.00	0.99	1.00	0.62	0.62	0.91	0.70	1.00	0.04	0.04	1.00	0.98	0.90
	S2	0.76	0.78	1.00	1.00	1.00	0.82	0.83	0.97	0.90	1.00	0.04	0.04	1.00	0.97	0.83
	A1	0.94	0.94	1.00	1.00	1.00	0.97	0.97	0.88	1.00	1.00	0.03	0.04	0.99	0.88	0.72
	A2	0.91	0.92	1.00	1.00	1.00	0.95	0.95	0.94	0.98	1.00	0.04	0.04	1.00	0.90	0.65