Unit-root and stationarity testing with empirical application on industrial production of CEE-4 countries

Lyócsa, Štefan and Výrost, Tomáš and Baumöhl, Eduard

Faculty of Business Economics in Košice, University of Economics in Bratislava

16 March 2011

Unit-root and stationarity testing with empirical application on industrial production of CEE-4 countries

Štefan Lyócsa[a]– Tomáš Výrost[b] – Eduard Baumöhl[c]*

Abstract

The purpose of this paper is to explain both the need and the procedures of unit-root testing to a wider audience. The topic of stationarity testing in general and unit root testing in particular is one that covers a vast amount of research. We have been discussing the problem in four different settings. First we investigate the nature of the problem that motivated the study of unit-root processes. Second we present a short list of several traditional as well as more recent univariate and panel data tests. Third we give a brief overview of the economic theories, in which the testing of the underlying research hypothesis can be expressed in a form of a unit-root / stationary test like the issues of purchasing power parity, economic bubbles, industry dynamic, economic convergence and unemployment hysteresis can be formulated in a form equivalent to the testing of a unit root within a particular series. The last, fourth aspect is dedicated to an empirical application of testing for the non-stationarity in industrial production of CEE-4 countries using a simulation based unit-root testing methodology.

Keywords: Unit-root, Stationarity, Univariate tests, Panel tests, Simulation based unit root tests, industrial production

JEL: C20, C30, E23, E60

[a] Department of Business Informatics and Mathematics

[b] Department of Finance and Accounting

[c] Department of Economics

Faculty of Business Economics in Košice,

University of Economics in Bratislava,

Slovakia, Tajovského 13, 04130 Košice, Slovakia

* Corresponding author, e-mail: stefan.lyocsa@gmail.com

Introduction

Since the first simulation published by Granger – Newbold (1974), the importance of statistical properties of time series received increasing importance in academic and empirical research. Their simulation suggested that when all (dependent and independent) time series are non-stationary, the classical regression results may be misleading. By a simple and partial reproduction of their simulation and by the review of empirical application of stationarity/unit-root tests we underlined the importance of studying the statistical properties of time series, particularly the stationarity of time series for empirical research. Our approach is not rigorous, but rather heuristic and more intuitive. The target audience of this paper are practitioners and students who engage in empirical research of time series models.

The first section of the paper reviews basic concepts of the stationarity of time series. The second section addresses testing for (non)stationarity, while the third section reviews some of the most debated applications of these tests in economics. The fourth section presents an empirical application, where we verify whether economic activity measures (Industrial Production Index, *IPP* henceforth) for Central and Eastern European Countries (Poland, Czech Republic, Hungary and Slovakia, *CEE-4*) may be regarded as stationary.

1. Non-stationarity and Spurious Regression

Before we review the standard univariate and multivariate stationarity/unit-root tests, we briefly define the stationarity property of time series and explain the intuition behind the spurious regression results. We say that a stochastic process $\{y_t\}^T_{t=1}$ is *strictly stationary* if it has a probability distribution, which is time invariant. More formally, if t is a time index and $\underline{Z}$ be a set of integers, then for any vector $(t_1,t_2,\ldots.t_h)$ of $\underline{Z}$ and any integer k: $F(y_{t_1}, y_{t_2},\ldots, y_{t_h}) = F(y_{t_1+k}, y_{t_2+k},\ldots, y_{t_h+k})$ holds, where $F(.)$ is the joint distribution function of the h values. This property implies that y_t are identically distributed. The *weak (covariance) stationarity* property focuses only on the first two moments of the stochastic process. The mean and variance of the stochastic process are finite and time invariant and the covariance between two observations depends only on the distance between those observations and not on the actual time t itself. That is: (1) $E[y_t] = \mu$, (2) $D[y_t] = \sigma^2$, (3) $cov(y_t, y_{t+k}) = cov(y_{t+h}, y_{t+h+k}) = \gamma_k$. The simple intuition behind both of the definitions is that the observations in one time series should not be from different populations. However, for most empirical applications we need another property, that y_t and y_{t+k} are almost independent as k increases. We say that if for a weakly stationary process $corr(y_t,y_{t+k}) \rightarrow 0$ for $k \rightarrow \infty$ holds, than the process is *asymptotically uncorrelated*. This property is sometimes referred to as *weak dependence* of the time series. This property is important for the use of Central

Limit Theorem and the Law of Large Numbers, thus for the Ordinary Least Squares estimators as well.

Most of the economic time series are non-stationary. As was shown in numerous studies (starting with Granger – Newbold, 1974) and also in Table 1 of this paper, the use of non-stationary variables in regressions might lead to spurious results. Analytically, the spurious regression was first explained by Phillips (1986). However, more heuristic explanations follow from simulation studies. First, let us consider the following Data Generating Process (DGP):

A) $y_t^j = \alpha + \beta^j t + \varepsilon_t^j$ (1)

Where y_t^j is the *j*-the time series, t is the time index, $t = 1,2,\ldots,T$, β^j is the slope of the deterministic trend and ε_t^j the error term (white noise). Clearly, if $\beta_1^j \neq 0$, the mean of the series y_t^j will not be constant, as the values of y_t^j deterministically depend on the time trend, thus the series is non-stationary. Such trend is called a *deterministic trend*. It's a form of variation of the series, which is predictable.

If we randomly generate two such series, $j = 1,2$, where $\varepsilon_t^j \sim N(0,1)$, $\beta^j \sim U(-0.15,+0.15)$, $T = 100$, and regress y^1 on y^2, chances are, that we will find a significant relationship, with high R^2. What we would actually find is the relationship between the two trends of the series (see Table 1). Moreover, with the increasing effect of the linear trend component $\beta^j t$ on the values of the series y_t, the signalling of such a relationship increases as well. This is clearly a spurious regression. However, after removing the trend component in both of the series, we could obtain meaningful results.

Observed economic processes are rarely that simple. A more complicated form of trend is the so called *stochastic trend*. The random component of the series at time t, ε_t^j, can directly affect all the remaining values of the series $y_{t+1}, y_{t+2},\ldots,y_T$. This introduces some form of an autocorrelation in the series. If the size of this effect is not decaying, we regard such series to have a stochastic trend. For illustration purposes, one form of the stochastic trend may be written in the following DGP:

B) $y_t^j = \sum_{i=1}^{t} \varepsilon_i^j + \delta^j \sum_{i=1}^{t-1} \varepsilon_i^j$ (2)

In this type of DGP, the error terms are cumulating, thus the value of ε_t^j will affect every subsequent observation of y_t. Such series usually resembles to series with changing trends. As before, if we randomly generate two series, $j = 1, 2$, where $\varepsilon_t^j \sim N(0,1)$, $\delta^j \sim U(-0.99, +0.99)$, $T = 100$ and regress y^1 on y^2, we might find spurious results, (see Table 1). As the effects of the error terms are cumulative, it is possible for the two series to share a temporary co-movement within the sample. This would translate into a spurious relationship that could be detected by general estimation procedures. Working with such DGPs is much more complicated than with the DGP in Eq. (1), and often we are left only to consider the transformation of the series (i.e. differencing). A special case, when both series have common stochastic trend is called *cointegration*[1].

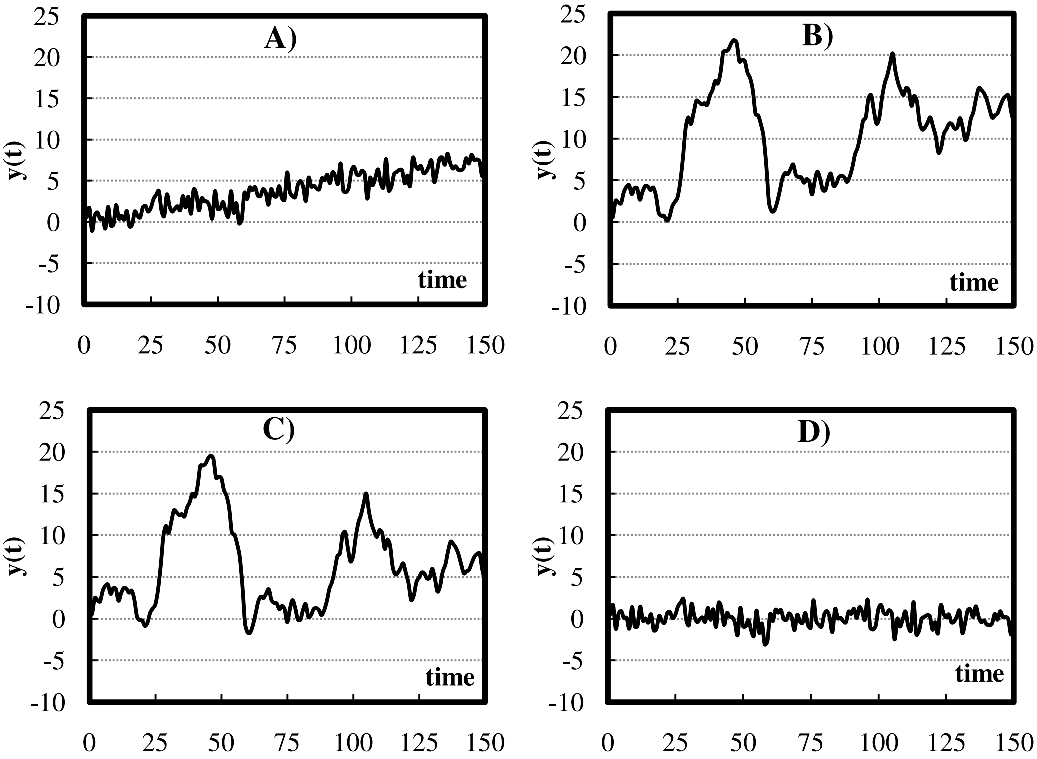

Figure 1 Sample time series plots of DGPs

Note: The figure contains four time series plots with following DGPs: A) $y_t = \alpha + \beta t + \varepsilon_t$*, B)* $y_t = \Sigma_{i=1}^{t} \varepsilon_i + \delta\Sigma_{i=1}^{t-1} \varepsilon_i$*, C)* $y_t = \beta t + \Sigma_{i=1}^{t} \varepsilon_i + \delta\Sigma_{i=1}^{t-1} \varepsilon_i$*, D)* $y_t = \varepsilon_t$*. Where* $T = 150$*,* $\varepsilon_t \sim N(0,1)$*,* $\alpha = 0$*,* $\beta = 0.05$ *and* $\delta = 0.6$*.*

A combination of the previous two DGPs is possible as well. Such series incorporates a deterministic trend and a stochastic trend. The model might look as follows:

$$\text{C)} \quad y_t^j = \beta^j t + \sum_{i=1}^{t} \varepsilon_i^j + \delta^j \sum_{i=1}^{t-1} \varepsilon_i^j \tag{3}$$

As before, regressing two randomly generated series on each other will most probably yield a spurious regression.

Frequently, the notion of spurious regression and stationarity is explained by the so called order of integration of a time series. Following Davidson and MacKinnon (2003), it can be defined in the following manner. Consider a process for which as the number of observation grows, the mean, variance and covariances tend to fixed values and covariances depend only on the lag between the observations. Thus, it can be seen that such a process is similar to the weakly

---

[1] The probably simplest type of non-stationary process is the random walk, $r_t = r_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0,1)$. Since Pearson's 1905 article in the Nature, it is also known as a drunkard's walk. Murray (1994) explains the co-integration of two series on the behaviour of a drunkard and her dog. Separately, both might seem to follow a simple random walk (their walks are still non-stationary), however they are not. Both continuously asses the gap between them and if they are too far away from each other, they close the gap. Their walks are said to be co-integrated. They both added an *error-correction* mechanism to their steps.

(covariance) stationary process defined above, except it fulfils all the requirements only asymptotically. A process obeying these requirements is called integrated of order 0, or I(0). After establishing the definition for this process, it is easy to define all other orders of integration. Specifically, a series is called I(d) if it has to be differenced d times to fulfil the requirements for a I(0) process. From this definition follows, that a series generated as a linear trend with IID Gaussian error terms would be I(1) (as the first differences would produce a series oscillating around a constant). A similar series following a quadratic trend would be I(2). For the purposes of our discussion it is important to note, that the problem of spurious regression is usually associated with series that are integrated of order higher than zero. It can be seen that our DGP (A) clearly fulfils this definition.

The Table 1 summarizes the results from a Monte Carlo simulation, where the fourth DGP (D) process is a $y_t^j = \varepsilon_t^j$, i.e. *white noise* (which is stationary). There are other possible components of a DGP which we have not considered, like structural breaks in the mean, trend or volatility of the series, or cyclical components. These could be incorporated into the simulation as well.

Table 1 Type I error with different DGPs

DGPs	A)	B)	C)	D)
A) Linear Trend	58.7% (0.483)			
B) Stochastic Trend	80.5% (0.303)	83.3% (0.240)		
C) Linear and Stochastic Trend	69.3% (0.401)	81.9% (0.294)	83.3% (0.370)	
D) White noise	5.23% (0.010)	4.74% (0.010)	4.83% (0.010)	4.96% (0.010)

Note: For each combination of DGPs we ran 10000 iterations of the simple linear regression model $y_t^1 = a_0^j + b^j y_t^2$, where the proportion of significant (at 0.05) b^j is the percentage value in the table. The value in the brackets corresponds to the average coefficient of determination. The series y_t^1 and y_t^2 was generated for each iteration according to the corresponding DGP (A, B, C, D) with sample size of T = 100. In A and B the β_t^j were generated from a uniform distribution with parameters -0.15 and 0.15. In B and C the δ_t^j were generated from a uniform distribution with parameters -0.99 and 0.99. As the error terms in each iteration are uncorrelated, one would expect the rejection rate of the H_0: $b^j = 0$ at 5% significance level to be roughly 5%, for the AA, AB, AC, BB, BC, CC this is clearly not the case.

For example when we regressed model A series on a model B series, the rejection rate of the true $b^j = 0$ was 80.5%. The highest number was for model C against model C (83.3%). As we choose 5% critical values for the rejection of the $b^j = 0$ hypothesis, we would expect that the type I error would be around 5%. Such results were found only when at least one of the series was a stationary white noise. As was shown by Banerjee et al. (1993) and consequently analytically by Marmol (1996), even if the series in the regression are not integrated of the same order (e.g. the dependent variable is stationary and the independent variable non-stationary) the results are spurious. This problem is not that serious if one of the series is stationary. Nevertheless, it is advised that we should only use variables with the same order of integration.

2. Stationarity and Unit-Root tests

A simple rule of thumb[2] that a regression generates spurious results was suggested by Granger – Newbold (1974); i.e. low Durbin – Watson statistics and a high R^2. One approach to prevent such results is *to test*, whether the time series in the regression analysis are I(1), and if so, compute the differences or other data transformations before entering the variables into the regression (Granger, 2003). An alternative option is (Granger, 2003, p. 558): "*to add lagged dependent and independent variables, until the errors appear to be white noise*" (a Durbin – Watson or Ljung – Box tests are usually carried out on the error terms)[3].

Reviewing 155 papers from 17 journals[4] (2000 – 2010, as of 17.09.2010) it seems, that the most popular strategy for testing of the stationarity property of a single time series involves using the Augmented Dickey Fuller or Dickey Fuller test (ADF and DF respectively), with 64.9% of cases where the test was used. In 9% the DF-GLS test, in 12.3% the Phillips – Perron test, in 3.3% Ng – Perron tests, in 7.1% KPSS test was used. Other tests including those which take into account structural breaks were used only rarely. More interestingly, in 66.5% cases the researchers used only one test and in 27.7% cases two tests. When the overall results were inconclusive, the researchers usually continued the analysis with a warning note, or simply assumed one of the alternatives.

The ADF test runs the following regression:

$$\Delta y_t^j = \alpha + \beta^j t + \gamma^j y_{t-1}^j + \sum_{i=1}^{k} \delta_i^j \Delta y_{t-i}^j + \varepsilon_t^j \qquad (4)$$

with an interest in the test of a null hypothesis of $\gamma^j = 0$. Clearly, if we cannot reject the null, than the series contains a unit root and the series is regarded as non-stationary. The number of lagged values Δy_{t-i} is usually selected to eliminate the autocorrelation in the error term, so that the test statistics has desired statistical properties (e.g. Ng – Perron, 1995). However, the methodology for correctly choosing the number of lags is very extensive on its own. Choosing too many lags

[2] For a simple simulation of this rule see Baumöhl – Lyócsa (2009).

[3] This procedure is not without limitations. Some tests (*F*-test of the joint null hypothesis that the coefficients under consideration are statistically not different from zero) may not be carried out using standard methodology (Hamilton, 1994, p. 562).

[4] American Economic Review, Economic Policy, Journal of Economic Growth, Journal of Economic Perspectives, Journal of Finance, Journal of Financial Economics, Journal of International Economics, Journal of Marketing Research, Journal of Monetary Economics, Journal of Money Credit and Banking, Journal of Political Economy, Journal of Public Economics, Management Science, Review of Economic Dynamics, Review of Economics and Statistics, Review of Financial Studies, World Bank Economic Review. The sample may be severely biased, although it is the belief of authors that it is representative enough to give a glance at the popularity of specific tests. One form of bias may be that in some journals one topic (e.g. PPP hypothesis) was dominant, thus researches tended to use similar tests to make connections to previous research. It was not our goal to make general inferences on the popularity of tests, rather only some insights.

usually lowers the power of the test. Various selection criteria were analysed and constructed; see Ng – Perron (1995) and Ng – Perron (2001).

For empirical researcher there are plenty of tests at hand, which may be considered as alternatives for the analysis. This may naturally cause some confusion. For example, if the series has one or more structural breaks, one might erroneously conclude (using standard tests), that the series is non-stationary, while it is stationary with structural breaks[5] (Perron, 1989). Further on, stationarity/unit-root tests have notoriously low power for near unit root processes. Unfortunately, this is often the case of economic time series. If possible, one can increase the power using panel tests, but this has its problems as well[6]. The null and the corresponding alternative hypothesis are often not that informative. For example, one might test whether all series in the panel are non-stationary against an alternative, that at least one series is stationary. As Breuer et al. (2002, p. 527) put it:"*Rejection of the null hypothesis only tells the researcher that at least one panel member is stationary, with no information about how many series, or which ones, are stationary*".

Some of the available univariate tests are reviewed in the Table 2 and the panel stationarity/unit-root tests are in the Table 3. Our list is far from exhaustive. Our goal was to provide the reader with a list of some of the widely used tests with some relatively new tests as well.

The choice of the right tests depends on the set up of the problem which is of interest to the researcher. It is difficult to follow the latest advances or to understand the nuisances between employing various tests. The following remarks should be considered as potential pitfalls for practitioners, who start time series analysis. One of the "*conjectures*" why most of the papers use similar tests is the uncertainty of practitioners over which test to use. The rule "*newer the better*" may in most cases be true. However in this case there are at least two issues. One is often left with the task to write his own codes and scripts for these tests, as these are rarely available. Even if they are, there is no common language (thus software) where tests are being written (R, GAUSS, STATA, MATLAB, RATS just to mention a few). This naturally imposes other problems. It is difficult to verify the correctness of such codes, in terms of the calculation of the test statistics and finite sample critical values (if necessary). Even if one chooses to run the "*best*" test(s), one is often left with many options which need to be considered like: the model specification, the method for choosing the number of truncation lags, the critical values to mention the traditional suspects. As the testing for (non)stationarity is often the first task in the analysis of time series data and not the ultimate goal of the analysis, it seems by itself to be a demanding process.

---

[5] A structural break in the series is usually represented with a shift in the level of the series, or a change in the rate of the trend.

[6] Contemporaneous cross-sectional dependence, possible presence of structural break(s) in the individual series, heterogeneous structure of autocorrelation if we were to just name a few.

Table 2 List of univariate unit root and stationarity tests

Unit Root tests with H_0: Series has a Unit-Root against H_1: Series is stationary
Dickey – Fuller test (1979); *DF test*
Dickey – Fuller test (1979, 1981); *ADF test*
Phillips – Perron (1988) - *PP test*
Elliott et al. (1996); *DF-GLS test*
Point optimal test; *P test*
Ng – Perron (2001); MZ^d_α test
Ng – Perron (2001); MZ^d_t test
Ng – Perron (2001); MSB^d test
Ng – Perron (2001); MP^d_T test
Stationarity tests with H_0: Series is stationary against H_1: Series has a Unit-Root
Kwiatkowski et al. (1992); KPSS test
Leybourne – McCabe (1994)
Unit root tests which allow for structural break both in H_0 and H_1
Perron (1989)[a]
Perron – Vogelsang (1992)[b]
Clemente et al. (1998)[c]
Lee – Strazicich (2003)[d]
Lee – Strazicich (2004)[e]
Kim – Perron (2009)[f]
Unit root tests which allow for structural break in H_1
Zivot – Andrews (1992)[g]; *ZA test*
Lumsdaine – Papell (1997)[h]; *LP test*
Perron (1997)[i]
Stationarity tests with structural breaks
Lee – Strazicich (2001)[j]
Carrion-i-Silvestre – Sansó (2007)[k]
Special unit-root tests
Kapetanios et al. (2003)[l]; KSS test
Elliott – Jansson (2003)[m]; EJ test

*Note: **a** - Break dates are exogenously specified. The H_0: Series has a Unit-Root with a structural break, H_1: Series is stationary with a structural break. Three structural break models were considered: A) change in the level of the series, B) change in the slope of the (linear) trend and C) simultaneous change in the level and slope of the trend. In recent works on unit root test models A and C are considered. Further on, if the structural break occurs suddenly, one assumes an additive outlier model (AO model), if it occurs gradually, than an innovation outlier model (IO model). The two models specify the transition mechanism of the structural breaks. A simple example of a model with two AO is: $y_t = \alpha + \delta_1 DU_{t,1} + \delta_2 DU_{t,2} + \varepsilon_t$, where DU_t is a dummy variable with $DU_t = 1$ for $t > Tb + 1$, and 0 elsewhere, where T_b is a date of break. This model assumes two shifts in the level of the series. An example of an IO model in this context could be: $y_t = \alpha + \omega_1 DT_{t,1} + \omega_2 DT_{t,2} + \delta_1 DU_{t,1} + \delta_2 DU_{t,2} + \rho y_{t-1} + \varepsilon_t$, where DT_t is a dummy variable with $DT_t = 1$ if $t = Tb + 1$ and 0 elsewhere and $|\rho| < 1$. **b -** Break dates are endogenously determined. Other tests with structural breaks allow some form of endogenous structural break detection. **c** - Clemente et al. (1998) considered non-trending series only (an additive and innovation outlier models) H_0: Series has a Unit-Root with two structural breaks, H_1: Series is stationary with two structural breaks. **d** - Lee – Strazicich (2001, 2003, 2004) consider model A and C. In Lee – Strazicich (2003, 2004) their methodology implies that by rejecting the null hypothesis, the series is stationary with breaks, regardless of whether structural break(s) occur under the null of Unit-Root. Thus we included both their tests to this category. H_0: Series has a Unit-Root, H_1: Series is stationary with two structural breaks. **e** - H_0: Series has a Unit-Root, H_1: Series is stationary with one structural break. **f** – H_0: Series has a Unit-Root with a structural break, H_1: Series is stationary with a structural break. **g** – H_0: Series has a Unit-Root, H_1: Series is stationary with one structural break. **h** – H_0: Series has a Unit-Root, H_1: Series is stationary with two structural breaks in trend. **i** – H_0: Series has a Unit-Root, H_1: Series is stationary with a structural break. **j** - H_0: Series is stationary with a structural break, H_1: Series has a Unit-Root. **k** - H_0: Series is stationary with two structural breaks, H_1: Series has a Unit-Root. **l** – H_0: Series has a Unit-Root, H_1: Series is globally stationary ESTAR (**E**xponential **S**mooth **T**ransition **A**uto**r**egressive) process. **m** - Test uses stationary covariates, for an example see Amara – Papell (2006), H_0: Series has a Unit-Root, H_1: Series is stationary.*

Table 3 List of panel unit root and stationarity tests

Unit Root tests with H_0: all time series have a unit root, H_1: all time series are stationary
Levin – Lin (1993); *LL test*
Levin – Lin – Chu (2002); *LLC test*
Harris – Tzavalis (1999)
Breitung – Das (2005)*[a]
Unit Root tests with H_0: all time series have a unit root, H_1: some time series are stationary
Im – Pesaran – Shin (2003); *IPS test*
Maddala – Wu (1999)
Chang (2002)*
Pesaran (2007)*
Phillips – Su (2003)*
Moon – Perron (2004)*
Choi (2006)*
Pesaran – Smith – Yamagata (2009)*
Taylor - Sarno (1998) *[b]; *MADF test*
Stationarity tests with H_0: All time series are stationary, H_1: Some series have a unit root
Choi (2001)
Hadri (2000)
Harris – Leybourne – McCabe (2004)*
Demetrescu – Hassler – Tarcolea (2010)*
Unit root tests which allow for structural breaks
Lee – Im – Tieslau (2005)*[c]; *LIT test*
Hadri – Rao (2008)*[d]
Im – Lee – Tieslau (2010)*[e]; *ILT test*
Special unit-root panel tests
Taylor – Sarno (1998)[f]*; *JLR test*
Breuer – McNown – Wallace (2001)*[g]

*Note: Test with * denote those, which take into account cross-sectional dependence in the panel, the so called 2nd generation tests.* ***a*** *– They make use of SUR estimation framework.* ***b*** *– H_0: all time series in the system have a unit root, H_1: some time series in the system are stationary.* ***c*** *– The test allows for level shifts in the series. H_0: all time series have unit-roots, H_1: some time series in the panel are stationary with structural breaks.* ***d*** *– The test allows for level and trend shifts in the series. H_0: all time series are stationary with a structural break (allow for various models of structural breaks), H_1: some time series in the panel have unit root.* ***e*** *– The test allows for level and trend shifts in the series. H_0: all time series have a unit-root, H_1: some time series in the panel are stationary with structural breaks. Similarly as LIT test, this test is invariant to the presence of structural breaks under the null hypothesis, thus rejecting the null will imply a stationary process with structural breaks and not unit-root process with structural breaks.* ***f*** *– Citing Taylor – Sarno (1998): "Engle and Granger (1987) demonstrate that, among a system of N I(1) series, there can be at most N-1 cointegrating vectors. Thus, if we reject the hypothesis that there are less than N cointegrating vectors among N series, this is equivalent to rejecting the hypothesis of non-stationarity of all of the series". This implies, that if there are N cointegrating vectors, than all series in the panel are stationary. They used this test to complement the MADF test.* ***g*** *– In their test it is possible to discriminate which particular time series in the panel is stationary and which not.*

3. Stationarity and a Review of its Empirical Applications

To give a glimpse of the empirical application of these tests, we made a short review of selected economic theories taking advantage of stationarity/unit-root tests, that is far from being comprehensive. In addition to these theories, stationarity/unit-root tests are used in most applications of time series econometrics. These were not covered as we have focused only on immediate applications of these tests.

If a time series is stationary, than any shocks that occur are transitory, their individual effects decay and eventually disappear as $t \to \infty$. If a series is non-stationary, than shocks have permanent effect on the series. This property can be easily visualized (see Figure 2).

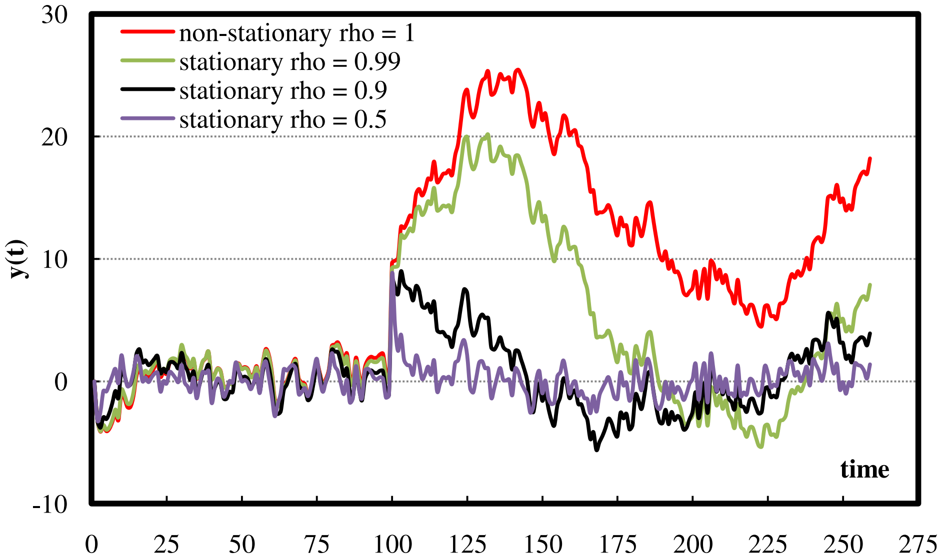

Figure 2 The non/stationary effect of shock on the time series

Note: Each of the series is an autoregressive process: $y_t = \rho y_{t-1} + \gamma D_t + \varepsilon_t$, $\varepsilon_t \sim N(0,1)$, $D_t = 1$ *for* $t = 100$, $D_t = 0$ *for* $t \neq 100$, *and* $\gamma = 10$ *denotes the size of the shock. With* $\rho = 1$ *being the non-stationary alternative, where the shock has a permanent effect. With the decreasing value of* $|\rho|$ *the transition of the shock into the following observations decreases.*

<u>Macroeconomic data</u>

One of the early empirical applications of stationarity/unit-root tests was concerned with macroeconomic time series, particularly the output of the economy. Until the seminal work of Nelson – Plosser (1982) many economists believed, that the output of the economy fluctuates around a deterministic trend. However, they were unable to reject the presence of a unit root in the GNP data (US). This finding had far reaching implications. For example, if the output is stationary, than: (1) the uncertainty about future values is by definition limited, (2)the fluctuations of output correspond to the business cycle, (3) the fluctuations may be controlled by monetary and fiscal means, (4) any effect of interventions will eventually "*vanish*". However, if the output is non-stationary than: (1) the uncertainty grows without bound, (2) the effect of monetary and fiscal policies on the output is limited while the real shocks are much more important, (3) the output is not a mean-reverting process, thus any intervention (causing real shocks to the economy) which needs to be corrected needs another intervention (another shock)[7]. The debate on the nature of the output is still open. For the results supporting the unit root in US output see Murray – Nelson (2002, 2004), Darné (2009) or Kejriwal – Lopez (2009) for nineteen OECD countries or Lyócsa et al. (2011) for Slovakia, Czech Republic, Poland and Hungary. Papell – Prodan (2004, 2007) advocate the stationarity of US output, Chang et al. (2010) for eleven Middle East Countries or Murthy – Anoruo (2009) for selected countries of Africa.

Bubbles and asset prices

If the order of integration between asset prices and their fundamentals differ, it is more likely that asset prices contain a bubble. Kirikos – Rich (1994) compared the order of integration between logarithm of CPI and M1. "*If the two series are integrated of the same order, we conclude that the price level does not exhibit bubble behaviour, which implies that the relevant exchange rate also does not exhibit bubble behaviour*" (Kirikos – Rich, 1994). Similarly Jirasakuldech et al. (2006) found that logarithms of exchange rates (US dollar against British pound, Canadian dollar, Danish krone, the Japanese yen and the South African rand) and their fundamentals (logarithms of M1, industrial production and the interest rate) are all integrated of order one. Thus existences of rational bubbles on these exchange rates were not confirmed. Cunado et al. (2007) found rational market bubbles on the S&P 500 index (1871:01–2004:06) testing for the orders of integration on the logarithm of the price to dividend ratio.

The non-stationarity property of asset prices itself is an interesting issue as well. If asset prices are non-stationary, than using only the information set containing historical values, restricts price forecasts to the last available price as the expected asset return is zero (under the assumption of normally distributed innovations). The empirical findings are not conclusive although much of the recent work supports the unit-root, thus random walk hypothesis. Narayan (2005) found that stock prices of Australian market (ASX All Ordinaries, 1960:01-2003:04) and New Zealand (NZSE Capital Index, 1967:01-2003:04) were nonlinear and non-stationary. Similar results were reported by Quian et al. (2008) for the Shanghai Stock Exchange Composite (1990:12-2007:06). Chen (2008) further strengthens these results from findings of nonlinearities and unit-roots for 11 stock markets of OECD countries.

PPP hypothesis

With regard to the applications of stationarity/unit-root tests, the probably most debated hypothesis is the *Purchasing Power Parity* (PPP) hypothesis. The amount of research is

[7] The merit of such policy is again questionable, as with nonstationary output of the economy, the effect of such an intervention would be difficult to assess, and this effect would be a permanent component of the series. These characteristics imply, that shocks have significant consequences which are difficult to assess.

overwhelming, with many methodological contributions, especially by generating new tests (lately mostly panel tests). The PPP hypothesis may be defined empirically as a following model: $q_t = s_t p_t / p_t^*$, where q_t is the real exchange rate, s_t the nominal exchange rate defined as home currency per unit of foreign and p_t, p_t^* are home and foreign price levels respectively. The equivalent specification is used with logarithms, $lq_t = ls_t - lp_t - lp_t^*$. If the PPP holds, than lq_t should be constant. The answer to the question whether lq_t is stationary is important to exchange rate modelling. As noted by Taylor – Taylor (2004): "*The question of how exchange rates adjust is central to exchange rate policy, since countries with fixed exchange rates need to know what the equilibrium exchange rate is likely to be and countries with variable exchange rates would like to know what level and variation in real and nominal exchange rates they should expect.*" However, there are various economical and statistical concerns which made analysis of this hypothesis particularly interesting. For example, researchers occupy themselves with how to incorporate various exchange rate regimes, which proxy should they use for the price level variable, how to solve possible small sample properties of used statistics or how to increase the statistical power of tests. Baum et al. (1999) summarizes that "*Due to factors such as transaction costs, taxation, subsidies, restriction on trade, the existence of non-traded goods, imperfect competition, foreign exchange market interventions, and the differential composition of market baskets and price indices across countries, one may expect PPP`s implications to emerge only in the long-run*". Current research focuses on the long-run PPP, with testing for stationarity using panel tests. For example, Narayan (2008) found strong evidence on the presence of PPP for 16 OECD countries with US dollar being the numeraire currency (on the monthly data of Main Economic Indicators of OECD from[8] 1973:01-2002:12). On quarterly dataset from 1973:1Q-2001:4Q, Lopez (2008) analyzed several panels of data. For example the largest panel of 20 countries with numeraire currency being the US dollar, Deutchemark, British Pound and Japanese Yen. Although the results are not conclusive, they suggest evidence for the PPP hypothesis. Christidou – Panagiotidis (2010) came to an interesting conclusion when they analyzed the evidence of PPP for a sample of 15 European and 12 Eurozone countries (later without Denmark, U.K. and Sweden) on monthly data from 1973:01-2009:04 using US dollar as numeraire currency. The panel tests provided evidence for the PPP on the whole sample, while surprisingly "*The introduction of the single currency has weakened the evidence in favour of the parity*" (for the Eurozone countries). Divino et al. (2009) used a series of panel stationarity/unit-root tests and provided evidence (using monthly data from 1981:01-2003:12, 26 Latin-American countries with US dollar being the numeraire currency) in favor of long-run PPP. For the US as reference currency, the PPP was not so convincing for the emerging markets of

[8] Only UK was an exception with data ending 2003:09. Other countries were: Canada, Japan, Austria, Belgium, Finland, France, Germany, Italy, Netherlands, Norway, Portugal, Spain, Sweden and Switzerland.

Eastern and Central Europe[9] (Kasman et al., 2010). However, with Deutschemark the PPP holds for Bulgaria, Croatia, Cyprus, Estonia, Romania, Slovakia, Slovenia and Turkey[10].

Interest rates

As was shown by Johnson (2006) a stationarity of real interest rates ($i_{r,t}$) is a necessary condition for the existence of the *Fisher effect*. The effect can be put formally as[11]: $i_{r,t} = i_{n,t} - \pi^e_t$, where i_n is the nominal interest rate and π^e the anticipated change in price level for the next period, from time t to $t+1$. The real interest rate can be calculated ex-post using nominal interest rates from short/long-term proxies of interest rates (Treasury bills, Government bonds), which also determines whether we are interested in the long-term or short-term Fisher effect. If Fisher effect holds, than monetary policy is sometimes interpreted as effective: real interest rate is unaffected by expected inflation (assumed to be controlled by central bank) as nominal interest rates move with accordance to the expected inflation rate. Earlier empirical results were not that convincing and many do not believe in short-run Fisher effect. However, Kapetanios et al. (2003) showed that for some countries real interest rates may be regarded as stationary, although with non-linear mean reversion what might be the reason of the previous failures to validate the Fisher hypothesis[12]. Since then, the non-linear framework used by Kapetanios et al. (2003) had been very popular and extended to other applications as well. An interesting conclusion was derived by Ito (2009) who analyzed the Fisher effect for Japan under three different monetary policy regimes. He found, that when the inflation expectations were high, nominal interest rates followed these expectations accordingly. However, when inflation rates were low, the nominal interest rates were not sensitive. This imposes some asymmetries when considering Fisher effect.

Convergence

Similar approach as the one used when testing the Fisher effect is when one is interested in whether two time series *converge*. In its simplest form, one tests for the presence of a unit root in the difference between two series. As an example, Holmes (2007), Holmes – Grimes (2008)

---

[9] Bulgaria, Croatia, Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Romania, Slovakia, Slovenia and Turkey.

[10] Cuestas (2009) found some evidence of PPP for several Eastern and Central European countries as well. Contrary to most other studies, data from these countries are rather shorter. In these two studies the monthly data were from approximately 13 – 15 years period.

[11] This form of classical Fisher effect is an approximation: $(1+ i_{n,t}) = (1+ i_{r,t})(1 + \pi^e_t) = (1+ i_{r,t} + \pi^e_t + i_{r,t}\,\pi^e_t) \approx (1+ i_{r,t} + \pi^e_t)$, from which $i_{n,t} \approx i_{r,t} + \pi^e_t$. Another form, which is used is: $li_{n,t} \approx li_{r,t} + l\pi^e_t$, where $li_{n,t}$, $li_{r,t}$ and $l\pi^e_t$ are all logarithms of the underlying variables. Also in both equations tax rate is often included.

[12] Using quarterly data from 1957:Q1-2000:Q3 of various measures of nominal interest rates (country specific) and price deflator, the countries for which the real interest rate might be considered as stationary were: Canada, France, Germany, Italy, Netherlands, New Zealand and Spain (data from 1974), while for Australia, Japan, United Kingdom and United states the real interest rates were non-stationary.

provides evidence for the convergence of regional house prices in UK. In Holmes (2007), the variables under consideration were differentials between natural logarithm of regional house prices (13 – regions) and the natural logarithm of UK house prices. More interestingly using 13 price differentials, Holmes - Grimes (2008) extracted 13 principal components and tested for the presence of unit-root in the largest principal component (LPC, largest in terms of explained variability): "*If the LPC is stationary, then all remaining principal components will also be stationary thereby confirming strong convergence among all the n regions. If the LPC based on the n house price differentials is non-stationary, strong convergence among all regional house prices is ruled out*". Another example is the economic convergence of regions. Christopoulos et al. (2007) measured the convergence between US regions in a similar way as Holmes in the previous example, however different tests were used. When testing, Christopoulos et al. (2007) searched for the unit root in the ratio between the regional real personal income per capita and the mean of all regions of US (8 regions with data covering 1921 – 2001), overall the results support convergence among regions[13].

Unemployment rates

Unemployment hysteresis is a property of unemployment where policy interventions which change the unemployment level have a tendency to sustain. These changes get "*built into the natural rate of unemployment resulting in changing the long run equilibrium"* (Belke – Polleit, 2009). Whether unemployment hysteresis is present, is an important question for policy implications, particularly regarding the labour market regulations and monetary policy (see NAIRU, **N**on-**A**ccelerating **I**nflation **R**ate of **U**nemployment). Using stationarity/unit-root tests, one can distinguish whether unemployment rate is stationary or not, thus indicating the (non)presence of hysteresis. Camarero et al. (2006) were unable to reject the hysteresis hypothesis using conventional univariate and panel tests. However, when their tests allowed for structural breaks in the series (shifts in mean and trend), the hysteresis hypothesis was rejected. The results suggest that policies have the potential to change the natural level of unemployment, but the short term shocks are mean-reverting. Their sample constituted of 19 OECD countries and 46 annual observations (1951-2001). Similar results were also found on a sample of transition countries (see Camarero et al., 2006). On a sample of Central and Eastern European countries, León-Ledesma – McAdam (2006) found some evidence against unemployment hysteresis for Lithuania, Slovak Republic, Czech Republic, Bulgaria, Hungary, Slovenia, Romania and Poland (monthly unemployment rates, with country specific length of data, starting as early as 1991:01 and ending at 2002:03). They also employed

[13] Putting the notion of convergence more formally one is interested, whether the following equation holds: $\lim E(z_{i,t} - z_{j,t}) = 0$, with $t \to \infty$. Where z is the variable under consideration (for example: real output per capita, house prices), i and j denote two different regions. For the convergence one needs that the difference be stationary, possibly trend stationary with negative trend. For more than two regions, the time trends must be identical between all differentials.

tests which take into account structural breaks, which confirmed the previous results. In contrast to previous studies, results in unemployment hysteresis hypothesis seem to support each other.

Industry dynamics

Stationarity/unit-root tests were also used for measuring industry dynamics, see e.g. Gallet – List (2001), Resende – Lima (2005), Sephton (2008), Giannetti (2009). Giannetti (2009) searched for the presence of a unit root in market shares of banks in 103 provinces in Italy with only 13 annual observations or less. The data were divided into four panels (National and North, Centre, South Italy) with result suggesting non-stationarity, perhaps except North Italy. Giannetti (2009) noted that: "*if shares turn out to be mean reverting, then would be reasonable to conclude that the industry is rather stable and that competitors had reached positions that were difficult to overcome*", thus as it seems such tests may be useful. Another application might be measuring the dynamics of industry structure using industry concentration measure.

Fiscal sustainability

There is also a wide area of empirical studies which assess the fiscal sustainability applying unit root tests. According to the so-called "*present value borrowing constraint*"[14] a stationarity of government debts is a sufficient condition for fiscal sustainability. When the analyses are conducted on the individual samples of each country data, short span of time series (mostly debt-to-GDP ratios, stock of debt or public deficits) causes low power of applied univariate unit root tests. Such studies therefore often provide mixed results (see, Wilcox, 1989; Uctum – Wickens, 2000 or Bergman, 2001). This methodological issue arises in almost every macroeconomic data. To overcome lack of power, several recent empirical papers applied more advanced techniques in a panel framework. For example, Holmes et al. (2010) over the sample period 1971 – 2006 analyzed annual budget deficits as a percent of GDP for Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Spain, Sweden and the United Kingdom. Using Hadri – Rao (2008) test which allows for cross-sectional dependence and for endogenously detected structural breaks, they conclude that EU countries exhibit fiscal stationarity over the full period, even in the subsamples 1971 – 1990 and 1991 – 2006 (pre- and post- Maastricht Treaty). Evidence against the non-stationarity is considered here as support for the strong form of fiscal sustainability insofar as satisfying the present value borrowing constraint. Similar findings are provided by

[14] For further details see e.g. Wilcox (1989), Llorca – Redzepagic (2008) and Afonso – Rault (2010).

another recent study conducted by Afonso – Rault (2010). For the EU-15 countries during the period 1970 – 2006, they found the first differences of debt-to-GDP ratio to be stationary[15].

This short review is of course not exhaustive. Among others, we have not reviewed empirical applications of stationarity/unit-root tests for the interest rate parity, inflation or labour force participation.

4. Empirical Application

For the purposes of our analysis within the CEE-4 markets, we have selected the procedure described in Kuo and Mikkola (1999), which is fairly straightforward and visually appealing. Their suggested approach is based on the fitting of two different models describing the data. The first one is a stationary ARMA model, which after an inclusion of a deterministic trend component can be thought of a trend stationary alternative. On the other hand, the second model is from the ARIMA class, which corresponds to a difference-stationarity process. The objective of the procedure is to estimate the small-sample distributions of the test statistic used in unit-root testing for the two best fitting models. By examining the test statistic from the data and comparing it to the established distributions, it is possible to draw conclusions about which distribution seems more likely to be the one generating the test statistic. This decision is equivalent to the choice of a more likely alternative (trend or difference stationarity).

We examine the series for industrial production (IPP) within the CEE-4 countries (Czech Republic, Hungary, Poland and Slovakia). We conduct our analysis on the logarithms of IPP values of each country, as is often the case in order to homogenize the variance of the series. The descriptive statistics and normality test are presented in the following table.

Table 4 Descriptive statistics and normality test of IPP

	Obs.	Mean	Median	Min.	Max.	Std. Dev.	Skewness	Ex. Kurtosis	Jarque-Bera	*p*-value
Czech rep.	237	4.413	4.380	4.088	4.846	0.216	0.323	-1.188	18.049	0.00012
Hungary	237	4.236	4.346	3.564	4.837	0.386	-0.143	-1.421	20.739	0.00003
Poland	237	4.313	4.346	3.600	4.919	0.375	-0.163	-1.072	12.402	0.00203
Slovakia	237	4.456	4.378	4.041	5.003	0.263	0.472	-0.988	18.447	0.00010

[15] Some results were mixed in this case.

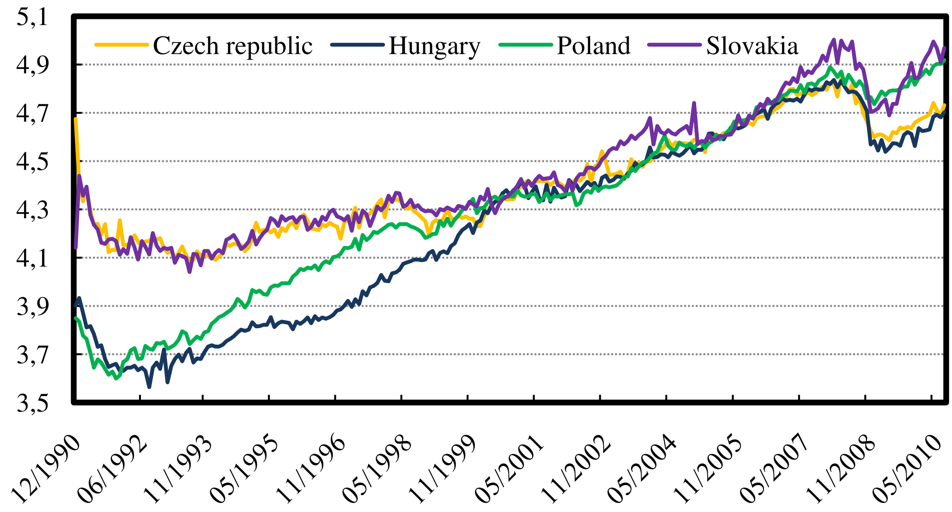

Figure 3 The logarithms of industrial production for CEE-4 countries

The first model to be estimated for each series was of the form

$$y_t = c + \beta t + u_t$$
$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right) u_t = \left(1 + \sum_{j=1}^{q} \theta_j L^j\right) \varepsilon_t \tag{5}$$

where L is the lag operator, c is the intercept term, t is a time variable (t = 1,2,…,T) y_t is the IPP series for the respective country, φ_i and θ_i are the model parameters and ε_t is the Gaussian error term. Effectively, this gives us a model with linear trend and ARMA(p,q) errors u_t. As for the selection of the order of the ARMA process, we have used the Akaike information criterion to choose from the models that have successfully removed serial correlation from the series, as indicated by the Ljung-Box portmanteau test on up to twelve lags.

A similar procedure has been performed on the alternative ARIMA model with the functional form

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right)(1 - L) y_t = \left(1 + \sum_{j=1}^{q} \theta_j L^j\right) \varepsilon_t \tag{6}$$

After estimating both models, we obtain two possible representations of the analyzed series. We treat the results in terms of the model orders and coefficients as data generating processes (DGP) to simulate 10 000 new series from the same DGPs. The basic idea of the procedure is the estimation of the distribution of the unit-root test statistic for the specific DGPs.

This leads to the question of the choice of lag orders for the tests used on the simulated series. The lag order for the original IP series was chosen with respect to the modified AIC criterion of Ng and Perron (2001), with maximum tested lag order following Schwert. In case of DF-GLS, we used the same number of lags when testing all generated series, to obtain the distribution of the test statistic. In case of point-optimal test, we used the Schwartz Bayesian information criterion with the maximal lag order specified by the Schwert's rule.

With all the above mentioned notes considered, we obtain the empirical distribution functions for the test statistics specifically for DGPs best corresponding to our sample. The results also take into account the precise number of observations in our dataset, as the generated ARMA/ARIMA series are set to the length of individual series. The empirical distributions allow in theory to decide, what distribution, and hence what process is more likely to produce the unit-root test statistics observed for industrial production for the CEE-4 countries.

Table 5 Estimated ARMA and ARIMA orders and coefficients

	Const.	AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	MA(1)	MA(2)	MA(3)	MA(4)	MA(5)	*t*
CZ	4.422	1.225	-1.073	0.848			-0.429	0.875				0.0004
(ARMA)	*(43.493)*	*(0.084)*	*(0.103)*	*(0.085)*			*(0.085)*	*(0.055)*				*(0.0019)*
HU	3.924	1.852	-0.852				-1.210	0.399	0.045			0.0029
(ARMA)	*(19.074)*	*(0.082)*	*(0.082)*				*(0.102)*	*(0.106)*	*(0.074)*			*(0.0027)*
PL	3.900	1.808	-1.211	0.701	-0.650	0.351	-1.039	0.632	-0.304	0.297	0.163	0.0045
(ARMA)	*(243.269)*	*(0.260)*	*(0.655)*	*(0.776)*	*(0.564)*	*(0.211)*	*(0.260)*	*(0.484)*	*(0.496)*	*(0.309)*	*(0.096)*	*(0.0019)*
SK	4.103	0.554	0.379	0.067								0.0030
(ARMA)	*(41.281)*	*(0.072)*	*(0.082)*	*(0.072)*								*(0.0017)*
CZ		0.481	-0.640	0.429	0.351	0.112	-0.796	0.925	-0.599			
(ARIMA)		*(0.148)*	*(0.099)*	*(0.112)*	*(0.086)*	*(0.110)*	*(0.137)*	*(0.085)*	*(0.107)*			
HU		0.897					-1.259	0.456				
(ARIMA)		*(0.056)*					*(0.076)*	*(0.063)*				
PL		1.007	0.119	-0.842	0.427		-1.199	0.126	1.010	-0.810	0.356	
(ARIMA)		*(0.226)*	*(0.186)*	*(0.174)*	*(0.191)*		*(0.234)*	*(0.228)*	*(0.155)*	*(0.263)*	*(0.088)*	
SK		-0.417										
(ARIMA)		*(0.068)*										

Note: The numbers in parentheses denote standard errors of the coefficients

Table 6 *Results from* simulation based tests

	DF-GLS test		*P-test*	
	prob.	τ_{sample}	*prob.*	τ_{sample}
Czech Republic				
$P(\tau \leq \tau_{\text{sample}} \mid f_{ARMA(p,q)}(\tau))$	0.955	(-0.663)	0.999	(136.275)
$P(\tau \leq \tau_{\text{sample}} \mid f_{ARIMA(p,d,q)}(\tau))$	0.564		0.999	
$P(\tau \leq \tau_{5\%\text{cv}}(\text{ARIMA(p,d,q)}) \mid f_{ARMA(p,q)}(\tau))$	0.630		0.033	
Hungary				
$P(\tau \leq \tau_{\text{sample}} \mid f_{ARMA(p,q)}(\tau))$	0.325	(-1.040)	0.931	(49.029)
$P(\tau \leq \tau_{\text{sample}} \mid f_{ARIMA(p,d,q)}(\tau))$	0.374		0.917	
$P(\tau \leq \tau_{5\%\text{cv}}(\text{ARIMA(p,d,q)}) \mid f_{ARMA(p,q)}(\tau))$	0.046		0.054	
Poland				
$P(\tau \leq \tau_{\text{sample}} \mid f_{ARMA(p,q)}(\tau))$	0.073	(-2.452)	0.749	(22.575)
$P(\tau \leq \tau_{\text{sample}} \mid f_{ARIMA(p,d,q)}(\tau))$	0.027		0.748	
$P(\tau \leq \tau_{5\%\text{cv}}(\text{ARIMA(p,d,q)}) \mid f_{ARMA(p,q)}(\tau))$	0.117		0.039	
Slovakia				
$P(\tau \leq \tau_{\text{sample}} \mid f_{ARMA(p,q)}(\tau))$	0.037	(-3.018)	0.394	(12.189)
$P(\tau \leq \tau_{\text{sample}} \mid f_{ARIMA(p,d,q)}(\tau))$	0.039		0.367	
$P(\tau \leq \tau_{5\%\text{cv}}(\text{ARIMA(p,d,q)}) \mid f_{ARMA(p,q)}(\tau))$	0.048		0.057	

Note: $P(\tau \leq \tau_{sample} \mid f_{ARMA(p,q)}(\tau))$ is the probability, that a randomly selected τ from the simulated distribution of test statistics generated under the ARMA model, is less than the τ_{sample}. $P(\tau \leq \tau_{sample} \mid f_{ARIMA(p,d,q)}(\tau))$ is the probability, that a randomly selected τ from the simulated distribution of test statistics generated under the ARIMA model is less than the τ_{sample}. $P(\tau \leq \tau_{5\%cv}(ARIMA(p,d,q)) \mid f_{ARMA(p,q)}(\tau))$ is the probability, that a randomly selected τ from the simulated distribution of the test statistics generated under the ARMA model is less than the 5^{th} percentile (the critical value) from the simulated distribution of τ, generated from the ARIMA model.

The visual inspection of the distributions of τ convincingly suggest, that all series under consideration are non-stationary (see Figure 4). The results for Hungary, Slovakia and Poland seem to be very similar. By design we know, that the distribution of statistics under *ARIMA* model corresponds to the non stationary series. As the simulated distributions of τ for *ARMA* and *ARIMA* models are practically overlapping each other, these results indicate that the assumption of *ARMA* model (the stationarity of the y_t series) is not valid.

These results are further supported by Table 6. All three probabilities should be interpreted together (see note under the Table 6)[16]. The simulation based results from DF-GLS tests suggest, that the behaviour of distributions are not necessarily similar, however, the conclusions are the same. We were unable to reject the unit-root hypothesis. In case of a unit-root process, one would

[16] As an example, consider the DF-GLS test, where the distribution of τ generated from the ARMA model is left of the distribution of τ of the ARIMA model. For the convenience let`s denote $P(\tau \leq \tau_{\text{sample}} \mid f_{ARIMA(p,d,q)}(\tau))$ as $P(\tau \mid {}_{ARIMA})$, $P(\tau \leq \tau_{\text{sample}} \mid f_{ARIMA(p,d,q)}(\tau))$ as $P(\tau \mid {}_{ARMA})$ and $P(\tau \leq \tau_{5\%\text{cv}}(\text{ARIMA(p,d,q)}))$ as $P({}_{ARIMA} \mid {}_{ARMA})$. The $P(\tau \mid {}_{ARIMA}) = 0.374$ (Table 6, for Hungary) indicates that conditionally on the estimated *ARIMA*(*p,d,q*) model, the probability of obtaining a DF-GLS statistics τ smaller than τ_{sample} is 0.374. We would therefore need a near 40 % significance level to reject the hypothesis, that $\tau \leq \tau_{\text{sample}} \mid f_{ARIMA(p,d,q)}(\tau)$. Larger values than the conventional significance levels suggest non-stationarity. Values around the conventional significance levels (1 %, 5 % and 10 %) are however not that indicative. As an example, consider the scenario of $P(\tau \mid {}_{ARIMA}) = 0.01$, $P(\tau \mid {}_{ARMA}) = 0.05$, which would suggest stationarity if $P({}_{ARIMA} \mid {}_{ARMA}) = 0.95$, but not if $P({}_{ARIMA} \mid {}_{ARMA}) = 0.08$. The later is actually the case of Slovakia, with overlapping distributions. On the other hand, with $P(\tau \mid {}_{ARIMA}) = 0.05$, $P(\tau \mid {}_{ARMA}) = 0.95$, if $P({}_{ARIMA} \mid {}_{ARMA}) = 0.95$, the τ_{sample} lies between the distributions and such result would alone support the rejection of the unit-root hypothesis, although the τ_{sample} is not likely to come from an ARMA DGP either. Another interesting case is the $P(\tau \mid {}_{ARIMA}) = 0.08$, $P(\tau \mid {}_{ARMA}) = 0.95$ and $P({}_{ARIMA} \mid {}_{ARMA}) = 0.05$ case, where the distribution of τ under the ARMA model lies practically between the 5 % empirical critical value and the τ_{sample}.

expect the τ_{sample} to be much closer to the centre of the simulated distribution generated under ARIMA. But the $P(\tau \leq \tau_{\text{sample}} \mid f_{ARIMA(p,d,q)}(\tau))$ are rather small. For Slovakia only 3.9 % of τ are smaller than τ_{sample} in the $f_{ARIMA(p,d,q)}(\tau)$ distribution. But also only 3.7 % of τ are smaller than τ_{sample} in the $f_{ARMA(p,d,q)}(\tau)$ distribution and finally 4.8 % of τ are smaller than $\tau_{5\%\text{cv}}(\text{ARIMA(p,d,q)})$ in the $f_{ARMA(p,d,q)}(\tau)$ distribution. Thus even though τ_{sample} is not close to the distribution corresponding to the non-stationary (unit-root) process, it is also no likely, that the τ_{sample} comes from the stationary process. The results of the *P-test* are much more homogenous, with considerably larger values of $P(\tau \leq \tau_{\text{sample}} \mid f_{ARIMA(p,d,q)}(\tau))$ which are similar to $P(\tau \leq \tau_{\text{sample}} \mid f_{ARMA(p,q)}(\tau))$. These probabilities point to the fact that the distributions under both null and alternative are very similar, which is visually confirmed. This leads us to the conclusion, that (1) although the series is not stationary, the ARIMA model is not describing the non-stationary behaviour of the series sufficiently and (2) that the true DGP might be one which is not I(1). It is possible that tests with more statistical power would be needed, or tests which allow for fractional order of integration.

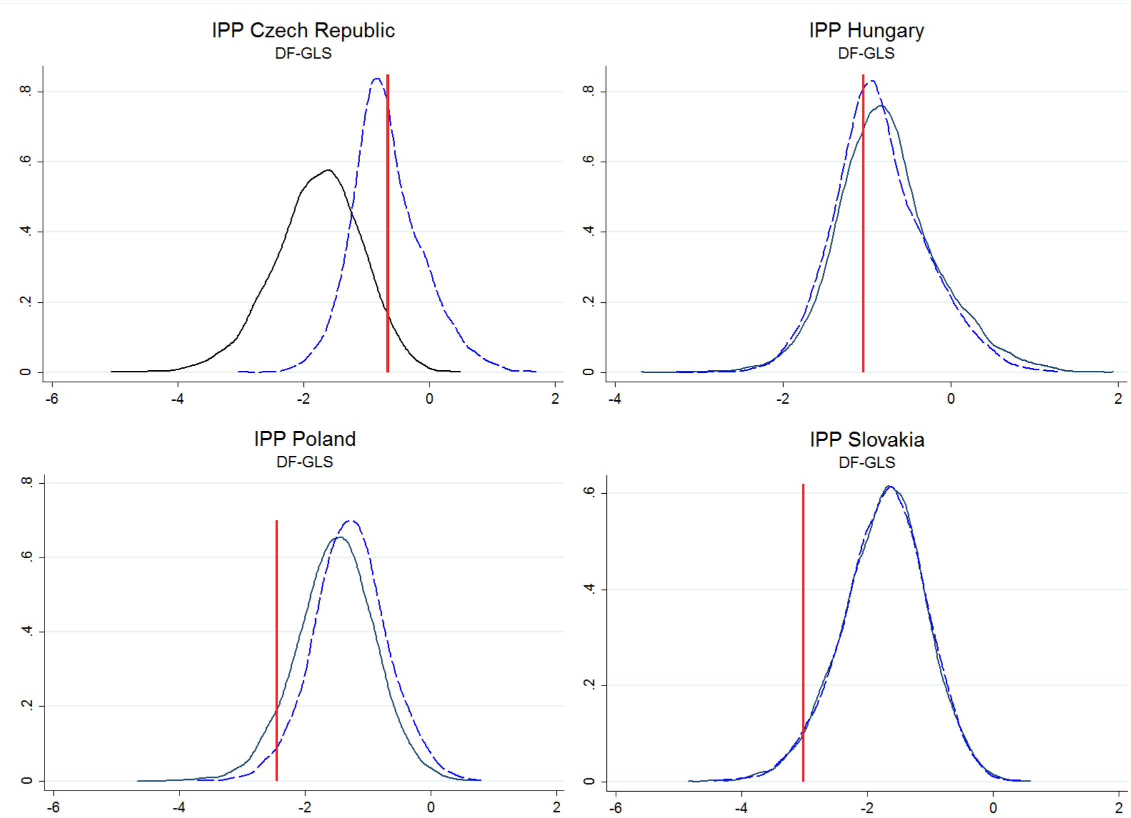

Figure 4 Kernel density distributions of simulated τ for DF-GLS tests.

Note: The solid lines are distributions of τ generated from ARMA models, the dashed lines correspond to ARIMA models, while the red bars to the test statistic calculated from the sample of IPP.

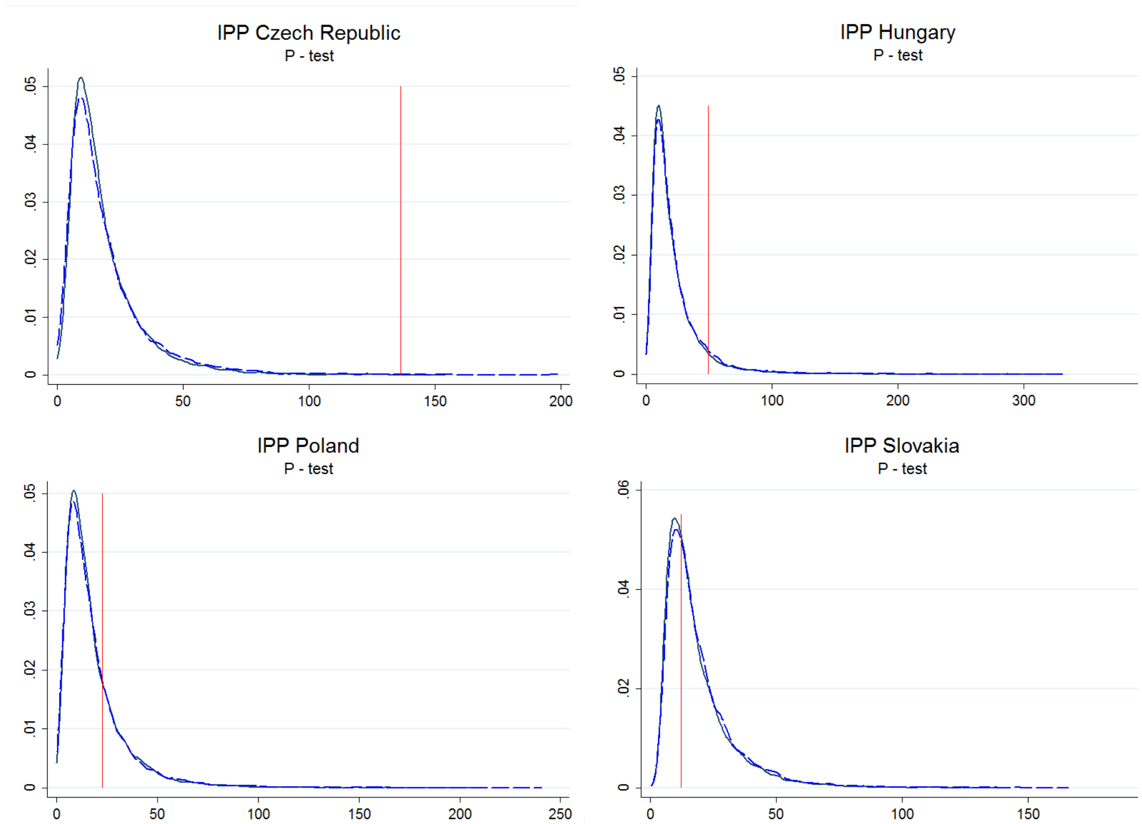

Figure 5 Kernel density distributions of simulated τ for P-tests.

Note: The solid lines are distributions of τ generated from ARMA models, the dashed lines correspond to ARIMA models, while the red bars to the test statistic calculated from the sample of IPP.

Conclusion

The topic of stationarity testing in general and unit root testing in particular is one that covers a vast amount of research. In this paper, we have been discussing the problem in four different settings.

First, we investigate the nature of the problem that motivated the study of unit-root processes. We deal with the problem of spurious regression, and by means of a simulation show the results of analysis of time-series generated by four different DGPs, namely those with deterministic and stochastic trends, their combinations as well as white noise. As previously demonstrated by others, we show that the rejection rate of the null hypothesis of no relationship is often high, despite being purely spurious.

These results demonstrate the need for proper identification of the nature of the series undergoing analysis. The years of research into the problem have produces a number of tests for stationarity. Their sheer numbers and specific conditions for their proper use may seem rather overwhelming. We therefore continued in our second part with the description of several traditional as well as more recent alternatives. Our exposition is divided into two categories of tests, namely the univariate ones and those based on panel data. We shortly discuss the benefits of the latter (namely, their potentially better power properties) as well as the new problems introduced by their use.

After describing the technical part of the testing and its alternatives, we give a brief overview of the economic theories, in which the testing of the underlying research hypothesis can be expressed in a form of a unit-root test. The issues of purchasing power parity, economic bubbles, industry dynamic, economic convergence and unemployment hysteresis can be formulated in a form equivalent to the testing of a unit root within a particular series.

The empirical part of our paper is dedicated to testing of stationarity of industrial production in CEE-4 countries. Instead of using the whole battery of tests presented in the preceding sections, we demonstrate a procedure that despite not being the mainstream solution has an interesting background and presents the problem of stationarity testing in an interesting and graphic way. This choice is made also with regard to the purpose of our paper, which was to explain both the need and the procedures of unit-root testing to a wider audience.

Acknowledgements

This paper was supported by grant from Slovak Grant Agency VEGA No. 1/0826/11.

References

[1] Afonso, A. – Rault, Ch. (2010). What Do We Really Know about Fiscal Sustainability in the EU? A Panel Data Diagnostic. *Review of World Economics*, 145(4), 731–55.

[2] Amara, J. – Papell, D. H. (2006). Testing for Purchasing Power Parity using stationary covariates. *Applied Financial Economics*, 16(1-2), 29-39.

[3] Banerjee, A. – Dolado, J. – Galbraith, J. W. – Hendry, D. F. (1993). *Co-integration, error correction, and the econometric analysis of non-stationary data*, reprint 2003. Oxford University Press, New York.

[4] Baum, Ch. F. – Barkoulas, J. T. – Caglayan, M. (1999). Long memory or structural breaks: can either explain nonstationary real exchange rates under the current float? *Journal of International Financial Markets, Institutions and Money*, 9(4), 359-76.

[5] Baumöhl, E. – Lyócsa, Š. (2009). Stationarity of Time Series and the Problem of Spurious Regression. *SSRN Working Paper Series*. Available at: <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1480682>.

[6] Belke, A. – Polleit, T. (2009). *Monetary Economics in Globalised Financial Markets*. Springer-Verlag Berlin.

[7] Breitung, J. – Das, S. (2005). Panel unit root tests under cross-sectional dependence. *Statistica Neerlandica*, 59(4), 414-33.

[8] Breuer, J. B. – McNown, R. – Wallace, M. (2002). Series-specific Unit Root Tests with Panel Data. *Oxford Bulletin of Economics and Statistics*, 64(5), 527-546.

[9] Camarero, M. – Carrion-i-Silvestre, J. L. – Tamarit, C. (2006). Testing for hysteresis in employment in OECD countries. New evidence using stationarity panel tests with breaks. *Oxford Bulleting of Economics and Statistics*, 68(2), 167-82.

[10] Carrion-i-Silvestre, J. L. – Sansó, A. (2007). The KPSS test with two structural breaks. *Spanish Economic review*, 9(2), 105-27.

[11] Chang, H.L. – Su, Ch.W. – Zhu, M. N. (2010). Is Middle East Countries Per Capita Real GDP Stationary? Evidence from Non-linear Panel Unit-root Tests. *Middle Eastern Finance and Economics*, 6, 15-19.

[12] Chang, Y. (2002). Nonlinear IV Unit Root Tests in Panels with Cross-Sectional Dependency, *Journal of Econometrics*, 110(2), 261-92.

[13] Chen, S. W. (2009). Non-stationary and Non-linearity in Stock Prices: Evidence from the OECD Countries, *Economics Bulletin*, 3(11), 1-11.

[14] Choi, I. (2001). Unit Root Tests for Panel Data, *Journal of International Money and Finance*, 20(2), 249-72.

[15] Choi, I. (2006). Combination Unit Root Tests for Cross-sectionally Correlated Panels. In: Corbae, D. – Durlauf, S. – Hansen, B. (Eds.), *Econometric Theory and Practice: Frontiers of Analysis and Applied Research*, essays in honor of Peter C. B. Phillips. Cambridge: Cambridge University Press.

[16] Christidou, M. – Panagiotidis, T. (2010). Purchasing Power Parity and the European single currency: Some new evidence. *Economic Modelling*, 27(5), 1116-123.

[17] Christopoulos, D. K. – Tsionas, E. G. (2007). Are US regional incomes converging? A nonlinear perspective. *Regional Studies*, 41(4), 525-30.

[18] Clemente, J. – Montanes, A. – Reyes, M. (1998). Testing for a unit root in variables with a double change in the mean. *Economics Letters*, 59(2), 175-182.

[19] Cuestas, J. C. (2009). Purchasing power parity in Central and Eastern European countries: an analysis of unit roots and nonlinearities. *Applied Economics Letters*, 16(1), 87-94.

[20] Cunado, J. – Gil-Alana, L. A. – Perez de Gracia, F. (2007). Testing for stock market bubbles using nonlinear models and fractional integration. *Applied Financial Economics*, 17(16), 1313-1321.

[21] Darné, O. (2009). The uncertain unit root in real GNP: A re-examination. *Journal of Macroeconomics*, 31(1), 153-66.

[22] Davidson, R. – MacKinnon, J.G. (2003). *Econometric theory and Methods.* Oxford University Press, New York. ISBN 0-19-512372-7

[23] Dickey, D. A. – Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*, 74(366), 427-31.

[24] Dickey, D. A. – Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 49(4), 1057-72.

[25] Divino, J. A. – Teles, V. K. – Andrade, J. P. (2009). On the purchasing power parity for latin-american countries. *Journal of Applied Economics*, 12(1), 33-54.

[26] Elliott, G. – Jansson, M. (2003). Testing for unit roots with stationary covariates. *Journal of Econometrics*, 115(1), 75-89.

[27] Elliott, G. – Rothenberg, T. J. – Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica*, 64(4), 813-36.

[28] Gallet, C, A. – List, J. A. (2001). Market share instability: an application of unit root tests to the cigarette industry. *Journal of Economics and Business*, 53(5), 473-80.

[29] Giannetti, C. (2009). Unit Roots and the Dynamics of Market Shares: An analysis using an Italian Banking micro-panel. Center *for Economic Research: Tilburg University*, 2008-44.

[30] Granger, C. W. J. – Newbold, P. (1974). Spurious regression in econometrics. *Journal of Econometrics*, 2(2), 111-20.

[31] Granger, C. W. J. (2003). *Spurious Regressions in Econometrics*, A companion to Theoretical Econometrics, ed. Baltagi, B. H., Blackwell Publishing, 557-61.

[32] Hadri, K. – Rao, Y. (2008). Panel stationarity test with structural break. *Oxford Bulletin of Economics and Statistics*, 70(2), 245-69.

[33] Hadri, K. – Rao, Y. (2008). Panel Stationarity Test with Structural Break. *Oxford Bulletin of Economics and Statistics*, 70(2), 245-69.

[34] Hadri, K. (2000). Testing for stationarity in heterogeneous panels. *The Econometrics Journal*, 3(2), 148-61.

[35] Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

[36] Harris, D. – Leybourne, S. – McCabe, B. (2005). Panel Stationarity Tests for Purchasing Power Parity with Cross-Sectional Dependence. *Journal of Busienss and Economic Statistics*, 23(4), 395-409.

[37] Harris, R. D. F. – Tzavalis, E. (1999). Inference for unit roots in dynamic panels where the time dimension is fixed. *Journal of Econometrics*, 91(2), 201-26.

[38] Holmes, M. – Otero, J. – Panagiotidis, T. (2010). Are EU Budget Deficits Stationary? *Empirical Economics*, 38(3), 767–78.

[39] Holmes, M. J. – Grimes, A. (2008). Is There Long-run Convergence among Regional House Prices in the UK? *Urban Studies*, 45(8), 1531-44.

[40] Holmes, M. J. (2009). How Convergent are Regional House Prices in the United Kingdom? Some New Evidence from Panel Data Unit Root Testing. *Journal of Economic and Social Research*, 9(1), 1-17.

[41] Im, K. – Pesaran, M. – Shin, Y. (2003). Testing for Unit Roots in Heterogeneous Panels. *Journal of Econometrics*, 115(1), 53–74.

[42] Im. S. K. – Lee, J. – Tieslau, M. (2010). Stationarity of Inflation: Evidence from Panel Unit Root Tests with Trend Shifts. *20th Annual Meetings of the Midwest Econometrics Group*, Oct 1-2, 2010.

[43] Ito, T. (2009). Fisher Hypothesis in Japan: Analysis of Long-term Interest Rates under Different Monetary Policy Regimes. *The World Economy*, 32(7), 1019-35.

[44] Jirasakuldech, B. – Emekter, R. – Went, P. (2006). Rational speculative bubbles and duration dependence in exchange rates: an analysis of five currencies. *Applied Financial Economics*, 16(3), 233-43.

[45] Johnson, P. A. (2006). Is it really the Fisher effect? *Applied Economics Letters*, 13(4), 201-03.

[46] Kapetanios, G. – Shin, Y. – Snell, A. (2003). Testing for a unit root in the nonlinear STAR framework. *Journal of Econometrics*, 112(2), 359-379.

[47] Kasman, S. – Kasman, A. – Ayhan, D. (2010). Testing the Purchasing Power Parity Hypothesis for the New Member and Candidate Countries of the European Union: Evidence from Lagrange Multiplier Unit Root Tests with Structural Breaks. *Emerging Markets Finance & Trade*, 46(2), 53-65.

[48] Kejriwal, M. – Lopez, C. (2010). Unit Roots, Level Shifts and Trend Breaks in Per Capita Output: A Robust Evaluation. Department of Economics: Purdue University, 1227.

[49] Kim, D. – Perron, P. Unit root tests allowing for a break in the trend function at an unknown time under both the null and alternative hypotheses, *Journal of Econometrics*, 148(1), 1-13.

[50] Kirikos, D. G. – Rich, K. (1994). Some tests for speculative exchange rate bubbles based on unit root tests, *Spoudai*, 44(1-2), 15-30.

[51] Kuo, B-S. – Mikkola, A. (1999). Re-examining Long-run Purchasing Power Parity. *Journal of International Money & Finance*, 18(2), 251-66.

[52] Kwiatkowski, D. – Phillips, P. C. B. – Schmidt, P. – Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 159-78.

[53] Lee, J. – Strazicich, M. (2001). Testing the null of stationarity in the presence of a structural break. *Applied Economics Letters*, 8(6), 377-82.

[54] Lee, J. – Strazicich, M. (2003). Minimum lagrange multiplier unit root test with two structural breaks. *The Review of Economics and Statistics*, 85(4), 1082-89.

[55] Lee, J. – Strazicich, M. (2004). Minimum LM Unit Root Test with One Structural Break, *Department of Economics, Appalachian State University*, 04-16.

[56] Lee, Junsoo, Kyung S. Im, and Margie Tieslau (2005), Panel LM unit root tests with level shifts, *Oxford Bulletin of Economics and Statistics*, 67(3), 393-419.

[57] León-Ledesma, M. – McAdam, P. (2004). Unemployment, Hysteresis and Transition. *Scottish Journal of Political Economy*, 51(3), 377-401.

[58] Levin, A. – Lin, C.F. – Chu, C.S. (2002). Unit Root Tests in Panel Data: Asymptotic and Finite Sample Properties. *Journal of Econometrics*, 108(1), 1–24.

[59] Levin, A. – Lin, C.F. (1993). Unit Root Test in Panel Data: New Results, Discussion Paper, 93-56, Department of Economics, University of California at San Diego.

[60] Leybourne, S. J. – McCabe, B. P. M. (1994). A consistent test for a Unit Root. *Journal of Business & Economic Statistics,* 12(2), 157-66.

[61] Llorca, M. – Redzepagic, S. (2008). Debt Sustainability in the EU New Member States: Empirical Evidence from a Panel of Eight Central and East European Countries. *Post-Communist Economies*, 20(2), 159–72.

[62] Lopez, C. (2008). Evidence of purchasing power parity for the floating regime period. *Journal of International Money and Finance*, 27(1), 156-164.

[63] Lumsdaine, R. L. – Papell, D. H. (1997). Multiple Trend Breaks and the Unit-Root Hypothesis, *The Review of Economics and Statistics*, 79(02), 212-218.

[64] Lyócsa, Š. – Baumohl, E. – Výrost, T. (2011). The Stock Markets and Real Economic Activity: New Evidence from CEE. *Eastern European Economics*, forthcoming.

[65] Maddala, G. – Wu, S. (1999). A Comparative Study of Unit Root Tests and a New Simple Test, *Oxford Bulletin of Economics and Statistics*, 61(0), 631-52.

[66] Marmol, F. (1996). Nonsense regressions between integrated processes of different orders. *Oxford Bulletin of Economics and Statistics*, 58(3), p.525-36.

[67] Moon, H. – Perron, B. (2004). Testing for a Unit Root in Panels with Dynamic Factors. *Journal of Econometrics*, 122(1), 8–126.

[68] Murray, Ch. J. – Nelson, Ch. R. (2002). The Great Depression and Output Persistence. *Journal of Money, Credit and Banking*, 34(4), 1090-98.

[69] Murray, Ch. J. – Nelson, Ch. R. (2004). The Great Depression and Output Persistence: A Reply to Papell and Prodan. *Journal of Money, Credit and Banking*, 36(3), 429-32.

[70] Murray, M. P. (1994). A Drunk and Her Dog: An Illustration of Cointegration and Error Correction. *The American Statistician*, 48(1), 37-39.

[71] Murthy, V. N. R. – Anoruo, E. (2009). Are Per Capita Real GDP Series in African Countries Non-stationary or Non-linear? What does Empirical Evidence Reveal? *Economics Bulletin*, 29(4), 2492-504.

[72] Narayan, P. K. (2005). Are the Australian and New Zealand stock prices nonlinear with a unit root*? Applied Economics*, 37(18), 2161-66.

[73] Narayan, P. K. (2008). The purchasing power parity revisited: New evidence for 16 OECD countries from panel unit root tests with structural breaks. *Journal International Financial Markets, Institutions and Money*, 18(2), 137-46.

[74] Nelson, Ch. R. – Plosser, Ch. R. (1982). Trend and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics*, 10(2), 139-62.

[75] Ng, S. – Perron, P. (1995). Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag. *Journal of the American Statistical Association*, 90(419), 268-81.

[76] Ng, S. – Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica*, 69(6), 1519-54.

[77] Papell, D. H. – Prodan, R. (2004). The Uncertain Unit Root in U.S. Real GDP: Evidence with Restricted and Unrestricted Structural Change. *Journal of Money, Credit and Banking*, 36(3), 423-27.

[78] Papell, D. H. – Prodan, R. (2007). Restricted structural change and the unit root hypothesis. *Economic Inquiry*, 45(4), 834-53.

[79] Perron, P. – Vogelsang, T. J. (1992). Nonstationarity and level shifts with an application to purchasing power parity. *Journal of Business & Economic Statistics*, 10(3), 301-20.

[80] Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, 57(6), 1361-401.

[81] Perron, P. (1997). Further evidence on breaking trend functions in macroeconomic variables. *Journal of Econometrics*, 80(2), 355-85.

[82] Pesaran, M.H. – Smith, L.V. – Yamagata, T. (2009). A Panel Unit Root Test in the Presence of a Multifactor Error Structure. Working paper, Cambridge University.

[83] Pesaran, M.H. (2007). A Simple Panel Unit Root Test In The Presence Of Cross Section Dependence, *Journal of Applied Econometrics*, 22(2), 265-312.

[84] Phillips, P. C. B. – Perron, P. (1988). Testing for a unit root in time series regression. *Biometrica*, 75(2), 335-46.

[85] Phillips, P. C. B. (1986). Understanding spurious regressions in econometrics. *Journal of Econometrics*, 33(3), 311-40.

[86] Phillips, P.C.B. – Sul, D. (2003). Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence, *Econometrics Journal*, 6(1), 217-59.

[87] Quian, X. Y. – Song, F. T. – Zhou, W. X. (2008). Nonlinear behaviour of the Chinese SSEC index with a unit root: Evidence from threshold unit root tests. *Physica A: Statistical Mechanics and its Applications*. 387(2-3), 503-10.

[88] Resende, M. – Lima, M. A. M. (2005). Market share instability in Brazilian industry: a dynamic panel data analysis. *Applied Economics*, 37(6), 713-718.

[89] Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics* 6 (2), 461–464.

[90] Schwert, G. W. (1989). Tests for Unit Roots: A Monte Carlo Investigation. *Journal of Business & Economic Statistics*, 7(2), 147-59.

[91] Sephton, P. (2008). Market shares and rivalry in the U.S. cigarette industry. *Applied Economics Letters*, 15(6), 417-22.

[92] Taylor, A. M. – Taylor, M. P. (2004). The Purchasing Power Parity Debate. *Journal of Economic Perspective*, 18(4), 135-58.

[93] Taylor, M, P. – Sarno, L. The behaviour of real exchange rates during the post-Bretton Woods period. *Journal of International Economics*, 46(2), 281-312.

[94] Taylor, M. P. – Sarno, L. (1998). The behavior of real exchange rates during the post-Bretton Woods period. *Journal of International Economics*, 46(2), 281-312.

[95] Wilcox, D. (1989). The Sustainability of Government Deficits: Implications of the present-value Borrowing Constraint. *Journal of Money Credit and Banking*, 21(3), 291–306.

[96] Zivot, E – Andrews, D. W. K. (1992). Further Evidence on the Great Crash, the Oil-Price Shock, ant the Unit-Root Hypothesis. *Journal of Business & Economic Statistics*, 10(03), 251-270.