UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

WORKING PAPERS IN ECONOMICS

No 568

Rejection Probabilities for a Battery of Unit-Root Tests

Authors

Florin G. Maican
Richard J. Sweeney

May 2013

ISSN 1403-2473 (print)
ISSN 1403-2465 (online)

Department of Economics
School of Business, Economics and Law at University of Gothenburg
Vasagatan 1, PO Box 640, SE 405 30 Göteborg, Sweden
+46 31 786 0000, +46 31 786 1326 (fax)
www.handels.gu.se   info@handels.gu.se

Rejection Probabilities for a Battery of Unit-Root Tests

Florin G. Maican[1], Richard J. Sweeney[2*]

Draft: May 2013

[1] *University of Gothenburg and IFN, Box 640 SE 405 30, Göteborg, Sweden, Phone +46-31-786-4866, Fax: +46-31-786-4154. E-mail: florin.maican@economics.gu.se*

[2] *McDonough School of Business, Georgetown University, 37th and "O" Sts., NW, Washington, DC 20057. Phone: 1-202-687-3742. Fax: 1-202-687-4031. E-mail: sweeneyr@georgetown.edu*

If the researcher tests each model in a battery at the α% significance level, the probability that at least one test rejects is generally larger than α%. For five unit-root models, this paper uses Monte Carlo simulation and the inclusion-exclusion principle to show for α%=5% for each test, the probability that at least one test rejects is 16.2% rather than the upper-bound of 25% from the Bonferroni inequality. It also gives estimated probabilities that any combination two, three, four or five models all reject.

Keywords: Real Exchange Rates; Unit root; Monte Carlo; Break models
JEL Classification: C15, C22, C32, C33, E31, F31.

I. Introduction

The researcher can test the unit-root null against a variety of alternative models, including the Augmented Dickey Fuller test equation and several related "break models" of the type explored by Perron (1989) and Zivot and Andrews (1992) and many others thereafter. Table 1 shows examples of alternatives, the standard ADF model and four break models, B1 – B4. An author's research design may require estimating the entire five-equation battery. If the researcher rejects any model at the significance level α, she knows that the probability that at least one model rejects is generally larger than α. From the Bonferroni inequality, the researcher knows that the upper-bound probability that at least one model rejects is 5 x α, for α=5%, the upper-bound probability is 25%. This paper uses the inclusion-exclusion principle with Monte Carlo simulations to find the estimated probability, 16.2%, that at least one model rejects at the 5% level. It further provides estimates of the probability that each of the possible pairs of models rejects, that each of the possible triplets of models rejects, that each of the possible foursome of models rejects and that all five models reject.

The researcher's judgment of the cost of increasing the test's size by running multiple models depends in part on her objective. If her purpose is to discriminate between the null of a unit root and a particular alternative model in Table 1, perhaps as implied by theory, the researcher may test only the particular alternative. If the researcher is interested in whether the data contain a unit root but equally in which model corresponds to the DGP, then he will almost surely run multiple models, perhaps the entire battery. One way that the use of a battery of tests arises is from exploring the robustness of ADF results. Montañés et al. (2005, p. 43) argue that, "it is now a very extended habit to complement the results of the [Augmented] Dickey–Fuller tests with [break model] tests." In contrast, they also note that, "a number of decisions should be taken prior to their [break model] use…. [I]t is necessary to determine the most appropriate specification of the type of break for the variable being considered." There is little systematic advice, however, on how to choose appropriate break models to test beyond the ADF. Indeed, there seems to be no discussion in the literature of whether it is better, and in what sense, to try to choose "the" model rather than running the four break models. Perron (1994) suggests that the researcher start with the most general specification [B4 in Table 1] and explore the robustness of unit-root test results by comparison with results under the possible alternatives—here the other three break models. Lumsdaine and Papell (1997) explore a larger set of break models than in

Table 1; they consider as many as two mean and two trend shifts. They state (p.217) that "there is no clearly accepted way to distinguish [choose] between the [alternative] models." They raise, without explicitly endorsing, the criterion of choosing the break model that is least favorable to the unit-root null (rejects the null at the highest significance level), a criterion that makes most sense if the researcher puts great weight on avoiding Type II errors. Maican and Sweeney (2012) argue that in general the researcher might want to present results for a battery of tests to help the reader interpret results.

A major issue in running a battery of models is that, under the null, the chance that one model may reject is large though unknown. In principle, the probability that at least one model rejects may be as large as the number of models tested times the significance level for rejection, for example, five models at the 5% significance level may give a probability as high as 25%. This is the case where the intersection of the rejection regions for any model and any other model is the null set. Taking this into account, with five models the researcher may use a Bonferroni approach and adopt the rule that the model must reject the null at the 1% level for the researcher to consider that the data reject the null at the 5% level, found by dividing the nominal size for each test by the total number of specifications tested. For the five models examined here, however, use of the inclusion-exclusion principle gives an estimate of the probability that at least one model rejects of approximately 16.2%, substantially lower than the upper-bound of 25%.

A related consideration is the probability that multiple models in a battery reject. Monte Carlo simulations discussed below provide estimates under the null that say two models reject, each at the 5% significance level or better. For the five models, the number of possible combinations of two models taken two at a time is 10. The probability that two models both reject depends of the two models considered. The smallest probability under the null for two particular models is 0.646 % (ADF, B4), the largest 2.473 % (B3, B4) and the average across pairs of models 1.373 %. As a rule of thumb, the researcher might use the probability of 1.373% as the significance level if the data reject at the 5% level for any two models. For three models each rejecting, the average probability across the nine possible combinations is 0.728%, and for four models each rejecting, the average across the five possible combinations is 0.366%. The probability that the data reject the null for all five models is only 0.275%.

II. Probabilities of Rejections under the Unit-Root Null

The size of a battery of tests depends on the number of tests in the battery, the significance level required for rejection in each test, and the probability of rejecting one model conditional on rejection of another model. For a five-test battery, using a 5% significance level in every test, the logically possible range is 5.0% to 25.0%, depending on whether the unions and intersections of all five models are the same or none has an intersection with any other. For the battery of five test equations in Table 1, however, the estimated probability from Monte Carlo simulations that one of the five models rejects at the 5.0% level is a bit over 16.2% The estimate of 16.2% is found from the inclusion-exclusion principle in combinatorics, for the case where the unit-root ***null*** hypothesis is true. Let E_i be the cardinality (omitting '| |' notation) of the event that the data reject the null at the γ% level in favor of the alternative i, for i=1,5 (i.e., ADF, …, B4). Then, using "$\bigcup$" for union and "$\bigcap$" for intersection,

(1) $$\bigcup^5_{i=1} E_{\gamma i} = [\Sigma^5_{i=1} E_{\gamma i}] - [\Sigma_{i,j:1\leq i<j\leq 5} (E_{\gamma i} \bigcap E_{\gamma j})] + [\Sigma_{i,j,k:1\leq i<j<k\leq 5} (E_{\gamma i} \bigcap E_{\gamma j} \bigcap E_{\gamma k})]$$
$$- [\Sigma_{i,j,k,h:1\leq i<j<k<h\leq 5} (E_{\gamma i} \bigcap E_{\gamma j} \bigcap E_{\gamma k} \bigcap E_{\gamma h})] + (E_{\gamma 1} \bigcap E_{\gamma 2} \bigcap E_{\gamma 3} \bigcap E_{\gamma 4} \bigcap E_{\gamma 5}).$$

Taking probability,

$$P (\bigcup^5_{i=1} E_{\gamma i}) = \Sigma^5_{i=1} P (E_i) - \Sigma_{i,j:1\leq i<j\leq 5} P (E_{\gamma i} \bigcap E_{\gamma j}) + \Sigma_{i,j,k:1\leq i<j<k\leq 5} P (E_{\gamma i} \bigcap E_{\gamma j} \bigcap E_{\gamma k})$$
$$- \Sigma_{i,j,k,h:1\leq i<j<k<h\leq 5} P (E_{\gamma i} \bigcap E_{\gamma j} \bigcap E_{\gamma k} \bigcap E_{\gamma h}) + P (E_{\gamma 1} \bigcap E_{\gamma 2} \bigcap E_{\gamma 3} \bigcap E_{\gamma 4} \bigcap E_{\gamma 5}).$$

Let Π_γ denote the probability that none of the five has a test statistic significant at the γ percent level with each model tested one at a time. Then,

(2) $\Pi_\gamma = 1 - \bigcup^5_{i=1} E_{\gamma i}$. [1]

In estimating the probabilities, the Monte Carlo simulations contain 100,000 replications. In each replication, the data are generated as 170 random variables u_t ~ iid N(0, 1), with the increment in the variable r_t generated as $\Delta r_t = u_t$. The program uses the first 20 u_t to "warm up" the series. On the remaining 150 observations of the replication, the program estimates all five models and for each model notes the t-value of the slope. The program repeats for 100,000 replications. For each model, the program orders the t-values from smallest to largest in algebraic

[1]Another way of looking at Π_γ is in terms of bi-model conditional probabilities. Call $\pi_{j|i}$ the probability that model j rejects conditional on model i rejecting, $\pi_{j|i} = P(j{=}R \mid i{=}R)$ and the unconditional probability that j rejects is $P(j = R)$. At one extreme, (a) the unconditional probability of rejection in model j at say the 5% level is the same as the conditional probability for any other model i for $j \neq i$, or $P(j = R \mid i = R) = P(j = R) = 0.05$. The size of the test of the null that no model in the battery is significant at the 5.0% level is 16.2%. At the other extreme, (b), rejecting in a particular model i implies rejecting in all other models, or $P(j = R| \; i = R) = \pi_{j|i} = 1.0$, and $\Pi = 1 - \gamma$. Testing the battery of models has the same size γ as testing any single model.

value; for that model, the 5% critical value is the t-value such that including that replication and all those with smaller algebraic t-values is five percent of the 100,000 replications. Using critical values found in this way for the five models, the program then examines each replication to see if the t-value for any model rejects.[2]

The program goes through the five t-values for each replication and counts those where both the ADF and B1 models reject. It goes through again and counts each replication in which the ADF and B2 models reject. It does this for each of the ten distinct pairs of models out of the five-model battery. Table 2 shows the percentage of the 100,000 replications in which pair jointly rejects.

The program goes through the five t-values for each replication and counts those where the ADF, B1 and B2 models all reject. It goes through again and counts each replication in which the ADF, B1 and B3 models reject. It does this for each of the nine distinct triplets of models out of the five-model battery. The model also goes through for each of the five distinct sets of four models out of the five-model battery. Finally, the program goes through the replications and counts those in which all five models reject. Table 2 shows all such results. The program then uses (1) and (2) to estimate the probability that at least one of the five models reject the null at the 5% significance level.

Table 3 provides an equivalent way of looking at the results in Table 2, by showing the probability that a particular model rejects conditional that a particular subset of models all reject.

In a general framework, Harvey, Leybourne and Taylor (2009) and Smeekes and Taylor (2012) discuss implications of unit-root testing based on “union of rejection decision rule.” They argue that testing based on “union” might be difficult to be recommended for general use because it implies a trade-off. First, the appropriate critical values are much lower than the critical values resulting from a single test. This implies a reduction in the the power of the test. Second, the power of the optimal test will dominate the power of the union test if T is large. Therefore, the union test tends to have more power than the general test, i.e., B4. Unfortunately, the power function of the union test does not uniformly dominate the power function of any single test for all possible specifications. We might expect that using a battery of five tests, the gain in power is reduced by the loss of power using substantially lower critical values.

2 In the simulations, the breaks are found using Zivot and Andrews' (1992) procedures, i.e. the break is the location that gives the minimum of t-stat of α.

This paper does not suggest the use of union strategy.[3] Our aim is to discuss the implications for unit-root rejections using break model, e.g., choosing the correct model has direct implications on the speed of convergence. Using a particular specification (models with endoenous breaks), our work complements their theoretical findings.

III. Empirical Example: Real Exchange Rates in the Central and Eastern European transition countries

This section presents a summary of the empirical results using a battery of five unit-root tests for real exchange rates in Central and Eastern European transition countries (CEE). The real-exchange-rate are constructed from monthly nominal exchange rates against the Euro and consumer price indices (CPI). CPI data are from *International Financial Statistics* (CD-ROM, August 2006). Euro exchange rates are from the Reuters database and national central bank statistics. The countries included are: Bulgaria, The Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia and Slovenia. Maican and Sweeney (2013) provide a detailed analysis of unit-root testing in CEE countries.

Table 4 shows that for some CEE countries the data reject the unit-root null for a number of alternative specifications, i.e., multiple models reject the null for eight of the ten countries (no models rejects for Poland and only the ADF for Latvia). These results indicate that the estimated speed of adjustment can vary greatly across specifications. Therefore, using an incorrect specification can underestimate the speed of adjustment.

From simulations, the estimated probability of rejecting both the B3 and the B4 models is 2.47%. This is the largest for two-model pairs; the minimum is 0.646 of 1% for the ADF-B4 pair. For any country where two or more models reject at the 5% level, the battery rejects at a minimum at the 2.47% significance level and often at a much more stringent level. The estimated probability of rejection by all five models is 0.275%, e.g., Bulgaria. Using the inclusion-exclusion principle, we find that the probability under the null that at least one of the models in the battery rejects is 16.2%. Table 4 shows the significance level for the battery of tests for the seven CEE countries for which multiple models reject the null. It also shows the significance level (16.2%) for the two countries (Hungary and Latvia) where only one model

[3]To compare the union strategy with this popular pre-test approach is not the aim of this paper.

rejects.[4] Even if the size of the battery of tests is larger than the size for any one test, the probability of the battery rejecting for nine out of ten countries is extremely small. Therefore, we can "afford" to use the battery of tests with multiple series.

IV. Conclusions

The probabilities of rejections when using a battery of tests of the unit-root null have mostly gone unexplored. If the researcher tests each model one at a time at the α% significance level, the probability that at least one model rejects is generally larger than α%. The probability that at least one model in the battery of the five models in Table 1 rejects is 16.2%, and the probability at least one model in a battery of just the four break models rejects is approximately 13%.

As Tables 2 and 3 show, in interpreting test results it is valuable to know whether multiple models reject the null on the given data set. For example, under the null, the probability of two models rejecting at the 5% significance level ranges from 0.646% to 2.473%, depending on the combination, with an average of 1.373%. Put another way, under the null, if one model rejects by chance, the probability that another model rejecting ranges from 10.30% to 52.20%, with an average of 22.75% across the ten combinations of two models. Tables 2 and 3 also show results for the probabilities of each of three or four models rejecting.

To avoid testing more than one alternative, the researcher may examine the data carefully before choosing an alternative, including using various forms of statistical analysis, and may read in detail discussions of the period's history in hopes of finding clues to which alternative to choose. Of course, these data explorations use up degrees of freedom, just as does running preliminary regressions to find break points, etc. Moreover, experimentation on actual data and on simulated data shows that the researcher may still easily choose a misspecification.

These considerations suggest that it is useful to run a battery of unit-root tests. Indeed, for any series where there is a serious likelihood that a break model best fits the data, the researcher might as a matter of course run the battery and report results, much as common descriptive statistics are reported. This relieves the reader from wondering if failure to reject arises from using an inappropriate test equation. It also relieves the reader from wondering if the reported rejection is simply the best of the lot.

[4] For significance levels different from 5%, interpolation gives estimated probabilities.

References

Harvey, D.I., S.J. Leybourne, and Taylor, A.M.R. (2009) Unit root testing in practice: dealing with uncertainty over trend and initial condition (with commentaries and rejoinder), *Econometric Theory,* **25**, 587-667.

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

Maican, F. and Sweeney, R. (2013) Real Exchange Rate Adjustment In European Transition Countries*, Journal of Banking and Finance,* **37,** 907-926.

Maican, F., and Sweeney, R. (2012) Costs of Misspecification in Break-Model Unit-Root Tests. Working Papers Series, No. 536, *University of Gothenburg*.

Montañés, A., Olloqui, I. and Calvo, E. (2005) Selection of the Break in the Perron-type Tests. *Journal of Econometrics,* **129**, 41-64.

Perron, P. (1989) The Great Crash, the Oil Price Shock and the Unit Root Hypothesis, *Econometrica,* **57**, 1361–1401.

Perron, P. (1994) Trend, Unit Root and Structural Change in Macroeconomic Time Series. In: Rao, B.B. (Ed.), *Cointegration for the Applied Economists*. St. Martin's Press, New York.

Perron, P. (1997) Further Evidence on Breaking Trend Functions in Macroeconomic Variables, *Journal of Econometrics,* **80**, 355–386.

Smeekes, S. and Taylor, A.M.R. (2012) Bootstrap union tests for unit roots in the presence of nonstationary volatility, *Econometric Theory*, **28**, 422-456.

Zivot, E., and Andrews, D. (1992) Further Evidence on the Great Crash, the Oil Price Shock, and the Unit-Root Hypothesis, *Journal of Business and Economic Statistics,* **10**, 251–270.

Table 1. List of alternatives considered

Augmented Dickey-Fuller (ADF) equation:

$$\Delta r_t = \mu + \alpha\, r_{t-1} + \Sigma^{k}_{j=1}\, \gamma_j\, \Delta r_{t-j} + u_t,$$

Break Model 1 (B1): Shift in mean, no time trend

$$\Delta r_t = [\mu + D_\mu\, \theta] + \alpha\, r_{t-1} + \Sigma^{k}_{j=1}\, \gamma_j\, \Delta r_{t-j} + u_t,$$

Break Model 2 (B2): Shift in mean, time trend included

$$\Delta r_t = [\mu + D_\mu\, \theta] + \beta\, t + \alpha\, r_{t-1} + \Sigma^{k}_{j=1}\, \gamma_j\, \Delta r_{t-j} + u_t,$$

Break Model 3 (B3): Shift in coefficient on time trend

$$\Delta r_t = \mu + [\beta\, t + D_\beta\, \phi\, (t - T_\beta)] + \alpha\, r_{t-1} + \Sigma^{k}_{j=1}\, \gamma_j\, \Delta r_{t-j} + u_t,$$

General Break Model, B4, shift in mean, shift in coefficient on time trend:

$$\Delta r_t = [\mu + D_\mu\, \theta] + [\beta\, t + D_\beta\, \phi\, (t - T_\beta)] + \alpha\, r_{t-1} + \Sigma^{k}_{j=1}\, \gamma_j\, \Delta r_{t-j} + u_t$$

Notes: r_t is the variable to be tested under the unit-root null hypothesis, μ the mean, β the coefficient on the time trend t, α the coefficient on the lagged level, the γ_j the coefficients on the lagged changes, D_μ a shift-in-mean dummy, equal to zero for $t < T_\mu$, and equal to unity for $t \geq T_\mu$, D_β a shift-in-trend dummy, equal to zero for $t < T_\beta$, and equal to unity for $t \geq T_\beta$, θ and ϕ are coefficients on the parameter-shift dummies D_μ and D_β, and $T_b = T_\mu = T_\beta$ if the model includes both parameter shifts. In the standard ADF used in finance applications, $\beta = \phi = \theta = 0$. In Break Model 1, $\beta = \phi = 0$ (possible shift in mean, but no time trend or break in trend); in Break Model 2, $\phi = 0$ (possible shift in mean and possible time trend, but not shift in the time trend); in Break Model 3, $\theta = 0$ (possible time trend and shift in trend, but no shift in mean); and in Break Model 4, ϕ, β, θ are freely fit (possible trend, and possible shifts in mean and trend). In fitting the Break Models, the times of the structural breaks, T_μ and T_β, are typically taken as endogenous and are found using methods pioneered in Zivot and Andrews (1992) and advanced in Perron (1997).

Table 2. ***Joint*** **Probabilities of Rejection**

Prob[ADF B1 B2 B3 B4]	= 0.00275	**0.275 %**	
Prob[ADF B1 B2 B3]	= 0.00355	0.355 %	
Prob[ADF B2 B3 B4]	= 0.00355	0.355 %	
Prob[ADF B1 B3 B4]	= 0.00285	0.285 %	¬
Prob[ADF B2 B3 B4]	= 0.00335	0.335 %	¬
Prob[B1 B2 B3 B4]	= 0.005	0.500 %	
Σ^5		**1.830 %**	
Prob[ADF B1 B2]	= 0.00535	0.535 %	
Prob[ADF B1 B3]	= 0.0045	0.450 %	
Prob[ADF B1 B4]	= 0.00375	0.375 %	¬
Prob[ADF B2 B3]	= 0.0047	0.470 %	
Prob[ADF B2 B4]	= 0.00465	0.465 %	
Prob[ADF B3 B4]	= 0.00415	0.415 %	
Prob[B1 B2 B3]	= 0.0066	0.660 %	
Prob[B1 B2 B4]	= 0.00875	0.875 %	
Prob[B2 B3 B4]	= 0.01305	1.305 %	¬
Σ^9		**6.550 %**	
Prob[ADF B1]	= 0.0112	1.120 %	
Prob[ADF B2]	= 0.00830831	0.831 %	
Prob[ADF B3]	= 0.007207205	0.721 %	
Prob[ADF B4]	= 0.006456455	0.646 %	¬
Prob[B1 B2]	= 0.01756757	1.757 %	
Prob[B1 B3]	= 0.00930931	0.931 %	
Prob[B1 B4]	= 0.01041041	1.041 %	
Prob[B2 B3]	= 0.01771772	1.772 %	
Prob[B2 B4]	= 0.024724725	2.473 %	¬
Prob[B3 B4]	= 0.024374375	2.437 %	
Σ^{10}		**13.729 %**	

Notes: The Monte Carlo simulations consisted of 100,000 replications. Each of the five models was run on each replication. For each replication, note was made for each model of whether it rejected at the 5% level or failed to reject. [The critical values for each model were estimated in previous Monte Carlo simulations.] With these records, the number of cases in which the pseudo-data reject both for model *i* and model *j* can be found, the number of cases in which the pseudo-data reject for models *i, j* and *k* can be found, etc. Note that in some cases where models *i* and *j*, one or more other models reject. Arrows " ¬ " point out maxima and minima. Union of Rejection Regions (5% significance level), from Inclusion-Exclusion Principle: 25% - 13.729 % + 6.550 % - 1.830 % + 0.275 % = 16.236%. General Break Model, B4, shift in mean, shift in coefficient on time trend:

$$\Delta r_t = [\mu + D_\mu \theta] + [\beta t + D_\beta \phi (t - T_\beta)] + \alpha r_{t-1} + \Sigma^k_{j=1} \gamma_j \Delta r_{t-j} + u_t.$$

Table 3. Conditional probabilities of rejections under the null, from Monte-Carlo simulations

Conditional rejections, two models

Prob[ADF | B1] = 0.224	Prob[ADF | B2] = 0.1661662
Prob[ADF | B3] = 0.1441441	Prob[ADF | B4] = 0.1291291 ¬
Prob[B1 | B2] = 0.3513514	Prob[B1 | B3] = 0.1861862
Prob[B1 | B4] = 0.2082082	Prob[B2 | B3] = 0.3543544
Prob[B2 | B4] = 0.4944945 ¬	Prob[B3 | B4] = 0.4874875

Conditional rejections, three models

Prob[ADF | B1 B2] = 0.3048433	Prob[B1 B2 | ADF] = 0.1071071
Prob[ADF | B1 B3] = 0.483871	Prob[B1 B3 | ADF] = 0.09009009
Prob[ADF | B1 B4] = 0.3605769	Prob[B1 B4 | ADF] = 0.07507508 ¬
Prob[ADF | B2 B3] = 0.2655367	Prob[B2 B3 | ADF] = 0.0940941
Prob[ADF | B2 B4] = 0.1882591	Prob[B2 B4 | ADF] = 0.0930931
Prob[ADF | B3 B4] = 0.1704312 ¬	Prob[B3 B4 | ADF] = 0.08308308
Prob[B1 | ADF B2] = 0.6445783	Prob[ADF B2 | B1] = 0.107
Prob[B1 | ADF B3] = 0.625	Prob[ADF B3 | B1] = 0.09
Prob[B1 | ADF B4] = 0.5813953	Prob[ADF B4 | B1] = 0.075
Prob[B1 | B2 B3] = 0.3728814	Prob[B2 B3 | B1] = 0.132
Prob[B1 | B2 B4] = 0.354251	Prob[B2 B4 | B1] = 0.175
Prob[B1 | B3 B4] = 0.2361396	Prob[B3 B4 | B1] = 0.115
Prob[B2 | ADF B1] = 0.4776786	Prob[ADF B1 | B2] = 0.1071071
Prob[B2 | ADF B3] = 0.6527778	Prob[ADF B3 | B2] = 0.0940941
Prob[B2 | ADF B4] = 0.7209302	Prob[ADF B4 | B2] = 0.0930931
Prob[B2 | B1 B3] = 0.7096774	Prob[B1 B3 | B2] = 0.1321321
Prob[B2 | B1 B4] = 0.8413462 ¬	Prob[B1 B4 | B2] = 0.1751752
Prob[B2 | B3 B4] = 0.5359343	Prob[B3 B4 | B2] = 0.2612613 ¬
Prob[B3 | ADF B1] = 0.4017857	Prob[ADF B1 | B3] = 0.09009009
Prob[B3 | ADF B2] = 0.5662651	Prob[ADF B2 | B3] = 0.0940941
Prob[B3 | ADF B4] = 0.6434109	Prob[ADF B4 | B3] = 0.08308308
Prob[B3 | B1 B2] = 0.3760684	Prob[B1 B2 | B3] = 0.1321321
Prob[B3 | B1 B4] = 0.5528846	Prob[B1 B4 | B3] = 0.1151151
Prob[B3 | B2 B4] = 0.5283401	Prob[B2 B4 | B3] = 0.2612613 ¬
Prob[B4 | ADF B1] = 0.3348214	Prob[ADF B1 | B4] = 0.07507508 ¬
Prob[B4 | ADF B2] = 0.560241	Prob[ADF B2 | B4] = 0.0930931
Prob[B4 | ADF B3] = 0.5763889	Prob[ADF B3 | B4] = 0.08308308
Prob[B4 | B1 B2] = 0.4985755	Prob[B1 B2 | B4] = 0.1751752
Prob[B4 | B1 B3] = 0.6182796	Prob[B1 B3 | B4] = 0.1151151
Prob[B4 | B2 B3] = 0.7372881	Prob[B2 B3 | B4] = 0.2612613 ¬

Conditional rejections, four models

Prob[ADF | B1 B2 B3] = 0.5378788	Prob[B1 B2 B3 | ADF] = 0.07107107
Prob[ADF | B1 B2 B4] = 0.4057143	Prob[B1 B2 B4 | ADF] = 0.07107107
Prob[ADF | B1 B3 B4] = 0.4956522	Prob[B1 B3 B4 | ADF] = 0.05705706 ¬
Prob[ADF | B2 B3 B4] = 0.256705 ¬	Prob[B2 B3 B4 | ADF] = 0.06706707
Prob[B1 | ADF B2 B3] = 0.7553191	Prob[ADF B2 B3 | B1] = 0.071
Prob[B1 | ADF B2 B4] = 0.7634409	Prob[ADF B2 B4 | B1] = 0.071
Prob[B1 | ADF B3 B4] = 0.686747	Prob[ADF B3 B4 | B1] = 0.057 ¬
Prob[B1 | B2 B3 B4] = 0.3831418	Prob[B2 B3 B4 | B1] = 0.1
Prob[B2 | ADF B1 B3] = 0.7888889	Prob[ADF B1 B3 | B2] = 0.07107107
Prob[B2 | ADF B1 B4] = 0.9466667 ¬	Prob[ADF B1 B4 | B2] = 0.07107107
Prob[B2 | ADF B3 B4] = 0.8072289	Prob[ADF B3 B4 | B2] = 0.06706707
Prob[B2 | B1 B3 B4] = 0.8695652	Prob[B1 B3 B4 | B2] = 0.1001001 ¬
Prob[B3 | ADF B1 B2] = 0.6635514	Prob[ADF B1 B2 | B3] = 0.07107107
Prob[B3 | ADF B1 B4] = 0.76	Prob[ADF B1 B4 | B3] = 0.05705706 ¬
Prob[B3 | ADF B2 B4] = 0.7204301	Prob[ADF B2 B4 | B3] = 0.06706707
Prob[B3 | B1 B2 B4] = 0.5714286	Prob[B1 B2 B4 | B3] = 0.1001001 ¬
Prob[B4 | ADF B1 B2] = 0.6635514	Prob[ADF B1 B2 | B4] = 0.07107107
Prob[B4 | ADF B1 B3] = 0.6333333	Prob[ADF B1 B3 | B4] = 0.05705706 ¬
Prob[B4 | ADF B2 B3] = 0.712766	Prob[ADF B2 B3 | B4] = 0.06706707
Prob[B4 | B1 B2 B3] = 0.7575758	Prob[B1 B2 B3 | B4] = 0.1001001 ¬

Conditional rejections, five models

Prob[ADF || B1 B2 4 B4] = 0.55	Prob[B1 B2 B3 B4 || ADF] = 0.05505506
Prob[B1 || ADF B2 B3 B4] = 0.8208955	Prob[ADF B2 B3 B4 || B1] = 0.055
Prob[B2 || ADF B1 B3 B4] = 0.9649123	Prob[ADF B1 B3 B4 || B2] = 0.05505506
Prob[B3 || ADF B1 B2 B4] = 0.7746479	Prob[ADF B1 B2 B4 || B3] = 0.05505506
Prob[B4 || ADF B1 B2 B3] = 0.7746479	Prob[ADF B1 B2 B3 || B4] = 0.05505506

Notes: The Monte Carlo simulations consisted of 100,000 replications. Each of the five models was run on each replication. For each replication, note was made for each model of whether it rejected at the 5% level or failed to reject. [The critical values for each model were estimated in previous Monte Carlo simulations.] With these records, the number of cases in which the pseudo-data reject both for model *i* and model *j* can be found, the number of cases in which the pseudo-data reject for models *i*, *j* and *k* can be found, etc. Note that in some cases where models *i* and *j*, one or more other models reject. Arrows " ¬ " point out maxima and minima. General Break Model, B4, shift in mean, shift in coefficient on time trend:

$$\Delta r_t = [\mu + D_\mu \theta] + [\beta t + D_\beta \phi (t - T_\beta)] + \alpha r_{t-1} + \textstyle\sum^k_{j=1} \gamma_j \Delta r_{t-j} + u_t.$$

Table 4: Unit-root test results using real exchange rates in CEE: Euro Base, consumer price indices

Country	ADF	B1	B2	B3	B4	Rejection prob. for battery
Bulgaria	-0.124*	-0.392****	-0.548****	-0.503*	-0.546****	0.275%
Shift date		1997-01	1998-01	2000-12	1997-01	
Bonferroni	50%	*5%*	*5%*	50%	*5%*	
Czech Rep.	-	-	-0.294***	-0.245*	-	0.931%
Shift date			2003-05	2002-04		
Bonferroni			12.5%	50%		
Estonia	-0.022****	-0.067****	-0.093****	-	-0.081****	0.355%
Shift Date		1996-07	1996-08		2003-01	
Bonferroni	*5%*	*5%*	*5%*		*5%*	
Hungary	-	-	-0.264**	-	-	16.20%
Shift Date			2001-04			
Bonferroni			25%			
Latvia	-0.033*	-	-	-	-	32.40%
Bonferroni	50%					
Lithuania	-0.030****	-0.043***	-	-	-	1.120%
Shift Date		1994-10				
Bonferroni	*5%*	12.5%				
Poland	-	-	-	-	-	
Romania	-	-0.148***	-0.213****	-0.303***	-0.258****	0.500%
Shift Date		1997-02	1997-02	2002-01	1997-02	
Bonferroni		12.5%	*5%*	12.5%	5%	
Slovakia	-	-	-0.292**	-	-0.411****	2.472%
Shift Date			1998-05		1998-07	
Bonferroni			25%		*5%*	
Slovenia	-	-0.086*	-0.246****	-0.339****	-0.276****	0.500%
Shift Date		1994-02	1994-03	1994-08	1995-07	
Bonferroni		50%	*5%*	*5%*	*5%*	

Notes: The models are specializations of the B4:

$$\Delta r_t = [\mu + D_\mu \theta] + [\beta t + D_\beta \phi (t - T_\beta)] + \alpha r_{t-1} + \sum_{j=1}^{k} \gamma_j \Delta r_{t-j} + u_t,$$

where D_μ is a dummy, equal to zero for $t < T_b$ and equal to unity for $t \geq T_b$, where T_b is the time at which the intercept shifts. D_β is a dummy, equal to zero for $t < T_b$ and equal to unity for $t \geq T_b$, where T_b is the time at which the trend coefficient shifts. The models are as follows. ADF: $\theta=\beta=\phi=0$. B1: $\beta=\phi=0$. B2: $\phi=0$. B3: $\theta=0$. B4–no restrictions, the mean and trend shift occur at the same T_b. For each country, the first row of numbers is estimated slope coefficients on the lagged log variable r_{t-1}; this estimate is the negative of the speed of adjustment. The row indicates the significance of the slope of the lagged log real exchange rate at the 10%, 5%, 2.5% and 1% levels by *, **, ***, ****. The second line of numbers is the estimated break dates for the various break models (Zivot and Andrew, 1992). The third row is from the Bonferroni inequalities; it shows the minimum probability of each rejection, on the assumption that the rejection regions for the five break models have no intersections. The symbol "-" indicates that the slope on the lagged log real exchange rate is not significant at even the 10% level. The sample period generally starts in January 1993 and always ends in December 2005. Data begin in January 1995 for Latvia and Romania, and January 1994 for Lithuania and Poland. The number of lags k is found following Perron (1989) and Ng and Perron (1995). Approximate probabilities for a battery under the null of the pattern of rejections are from Monte Carlo simulations (Table 2). For significance levels different from 5%, interpolation gives estimated probabilities.