

Costs of misspecification in break-model unit-root tests

Florin G. Maican^{a,*} and Richard J. Sweeney^b

^aUniversity of Gothenburg and Research Institute of Industrial Economics (IFN),
 Box 640, SE40530, Goteborg, Sweden

^bMcDonough School of Business, Georgetown University, 37th and “O” Sts.,
 NW, Washington, DC, 20057, USA

This article examines power issues for the ADF and four break models (Perron, 1989; Zivot and Andrews, 1992) when the DGP corresponds to one of the break models. Choosing to test an incorrect break model can, but need not, greatly reduce the probability of rejecting the null. Break points that are relatively early in the sample period have substantial effects of increasing power. For modest shifts in time trends, simply including a time trend without shift in the model preserves power, but not for large time-trend shifts.

Keywords: unit root; Monte Carlo; break models

JEL Classification: C15; C22; C32; F31; E31

I. Introduction

The researcher can test the unit-root null against a variety of alternative models, including the Augmented Dickey–Fuller (ADF) test equation and several ‘break models’ of the type explored by Perron (1989) and Zivot and Andrews (1992) and many others after them. Below are the standard ADF and four common break models¹:

$$\text{ADF } \Delta r_t = \mu + ar_{t-1} + \sum_{j=1}^k \gamma_j \Delta r_{t-j} + u_t$$

$$\text{B1 } \Delta r_t = [\mu + D_\mu \theta] + ar_{t-1} + \sum_{j=1}^k \gamma_j \Delta r_{t-j} + u_t$$

$$\text{B2 } \Delta r_t = [\mu + D_\mu \theta] + \beta t + ar_{t-1} + \sum_{j=1}^k \gamma_j \Delta r_{t-j} + u_t$$

$$\text{B3 } \Delta r_t = \mu + [\beta t + D_\beta \phi(t - T_\beta)] + ar_{t-1} + \sum_{j=1}^k \gamma_j \Delta r_{t-j} + u_t$$

$$\text{B4 } \Delta r_t = [\mu + D_\mu \theta] + [\beta t + D_\beta \phi(t - T_\beta)] + ar_{t-1} + \sum_{j=1}^k \gamma_j \Delta r_{t-j} + u_t$$

In formulating her research design, the analyst may test only a single model or may test a battery of models. This article supposes that one of the break models is the Data Generation Process (DGP) and explores the consequences if the researcher chooses to use an incorrect model, that is, misspecifies the test equation.

Power experiments show that if the null is false but a misspecified model is tested, the probability of a Type II error may, but need not, be large. As expected, the costs

*Corresponding author. E-mail: florin.maican@economics.gu.se

¹ 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 ; a 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 $t \leq T_\beta$ 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 B1, $\beta = \phi = 0$; in B2, $\phi = 0$; in B3, $\theta = 0$; and in B4, ϕ , β and θ are freely fit. In fitting the break models, the times of the structural breaks, T_μ and T_β , are typically estimated, using methods pioneered in Zivot and Andrews (1992) and advanced in Perron (1997).

depend on the misspecification relative to the true model: in some cases, the probability of rejection falls little or actually rises compared to the true model; in other cases, the probability falls drastically. Further, the costs of selecting a misspecified model depend on the location of the true break in the sample period. The researcher who decides to test only a particular specification is testing the null that the data contain a unit root against the joint alternative that the data do not contain a unit root and that the researcher has correctly selected the model to which the DGP corresponds or a different model with close likelihood of rejecting.

The research on real exchange rates contains many unit-root tests of purchasing power parity (PPP).² A part of this literature finds that including one parameter shift generally gives higher adjustment speeds than without breaks. Hegwood and Papell (1998) investigate PPP in long annual data series and find substantially faster adjustment speeds in specifications that allow for mean shifts using the Lee (1976) and Lothian and Taylor (1996) data sets. Accounting for structural breaks, Maican and Sweeney (2013a) examine PPP in 10 Central and Eastern European transition countries and find substantially faster mean-reversion speeds than those found for major industrialized countries. If the DGP of the real exchange rate corresponds to one of the break models, the likelihood of correctly rejecting the unit-root null depends on the correct specification of the alternative.

The researcher's choice between models and between estimating one versus multiple models depends on her objective. If her purpose is to discriminate between the unit-root null and a particular alternative model above, 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 she may estimate multiple models.³ Even if the results convince the researcher that the data do not contain a unit root, she may find it very difficult to decide among several plausible alternatives, depending on which parameter shifts, the size of the shift, the speed of adjustment, the point in the sample period at which the shift occurs and the model that corresponds to the DGP.

A number of papers have discussed issues of power, but with different perspectives from this article's. Perron (1989, 1997) and Zivot and Andrews (1992) discuss the loss of power if the researcher must search the data for the break point that gives the optimum value of some criterion rather than knowing the break *ex ante*. They note that if the DGP is an ADF, then using a break model as the alternative reduces power. They do not, however, discuss

power in the case where the DGP is one break model, but the researcher uses a different break model as the alternative. Kim and Perron (2009) analyse unit-root tests where the DGP has a break in time trend at an unknown point under both null and alternative hypotheses. They show that taking account of this trend break improves power. Hecq and Urbain (1993) show that using an incorrect break date in break-model tests causes loss of power (and size distortion). Muller and Elliot (2003) discuss the effect on unit-root tests of initial values: an initial value far different from what might be expected from the DGP, after 'warming up' for say 20 periods, substantially affects the test's power. Montañés *et al.* (2005) 'study the consequences of an incorrect selection of the break model' when the DGP corresponds to a break model. They explore only two break models as DGPs; they use such a fast speed of adjustment and large shifts in trend that their simulation results are unrevealing for many practical purposes, for example, evaluating real exchange rates.

Tables 1–3 show power results for various break-model DGPs, adjustment speeds, parameter shifts and location of break in the sample period. (i) The power of models if the time of the break is early in the sample period, $T_b = 30$, is often larger, sometimes much larger, than if $T_b = 75$, $T_b = 120$. Reporting the estimated date of the break relative to the beginning and end of the sample period is important to aid the reader's evaluation. (ii) If the time trend shifts, models that do not include a time trend, ADF and B1 models, may have quite low power. (iii) Conditional on the speed of adjustment, multiple models are likely to reject when the parameter change is a time-trend shift and the model contains a time trend (B2, B3 and B4). As the shift in the time trend increases above a certain level, however, the B2 model's power falls sharply because B2 contains a time trend but does not allow for a shift in trend. As might be expected, (iv) *ceteris paribus*, power increases with increases in the speed of adjustment, and (v) multiple models are likely to reject when the speed of adjustment is large, as sometimes occurs in practice – see below.

II. Small Speed of Adjustment

The simulation results are presented in Table 1. The Monte Carlo simulations use 100 000 replications of 170 observations for each DGP. The first 20 observations warm up the model and are discarded. In the DGP, the breaks in the remaining 150 observations are at $T_b = 30$, $T_b = 75$ or

² Taylor (2006) surveys recent literature on PPP.

³ Maican and Sweeney (2013b) study rejection probabilities for a battery of unit-root tests using the inclusion-exclusion principle with Monte Carlo simulations.

Table 1. Power of five models across DGPs for break models, in per cent^a

Case	B1	B2	B3a	B3b	B4a	B4b	Av.
θ	1.00	1.00	–	–	1.00	1.00	of row
ϕ	–	–	0.10	0.01	0.10	0.01	$\sum/6$
α	–0.05	–0.05	–0.05	–0.05	–0.05	–0.05	
$T_b = 30$							
ADF	17.06	7.13	0.00	0.00	0.00	0.28	6.12
B1	16.74	1.51	0.00	0.00	0.00	0.03	4.57
B2	4.95	4.84	84.74	10.49	90.55	5.29	6.39
B3	18.50	18.91	87.38	9.58	90.13	17.51	16.13
B4	5.86	6.17	64.86	8.45	71.30	6.48	6.65
$T_b = 75$							
ADF	0.00	0.00	0.00	0.00	0.00	0.00	0.00
B1	14.97	3.85	0.00	0.00	0.00	0.45	4.87
B2	5.97	6.17	0.40	6.57	10.65	3.37	5.52
B3	0.68	0.63	61.47	10.30	89.28	1.31	3.23
B4	3.53	3.42	24.55	8.11	100.00	8.15	4.52
$T_b = 120$							
ADF	0.00	0.00	0.00	0.00	0.00	0.00	0.00
B1	7.00	0.51	0.00	0.00	0.00	0.31	1.96
B2	7.16	6.80	0.00	7.40	0.00	2.26	5.91
B3	3.35	3.29	25.46	10.30	64.11	6.17	5.55
B4	3.39	3.50	4.29	7.81	21.73	4.19	4.72

Notes: Each entry in the block is the percentage of the simulations in which the model rejects the null at the 5% significance level for the values of θ , ϕ used. The simulations are for 50 000 replications. Each replication follows one of the break models with no lags, a speed of adjustment of $\alpha = -0.05$ (root of $(1 + \alpha) = 0.95$), a mean of zero ($\mu = 0$), a time trend of 1% per month ($\beta = 0.010$), a shift in mean of θ and a shift in trend of ϕ in the column heading. The increments are $u_t \sim N(0, 1)$, each series is 170 observations, the series is ‘warmed up’ with 20 observations and the models are estimated over the remaining 150 observations. For each replication, each model is estimated without lags because the DGP contains no lags. The row for the model that corresponds to the DGP is in **bold** in each block; the largest entry in each column in the bloc is in **bold italic** (unless the DGP model’s entry is largest). Each entry in the column ‘Av. of row ($\sum/6$)’ is the simple average of the frequencies in that row for rejection under the four DGPs, B1, B2, B3b and B4b. The entry in the \sum column, compared to the entries in the row, allows the reader to see the DGPs that are more or less likely to generate a rejection by the model in that row.

^aShift in mean: One SD of 1.00. Shift in time trends: 0.10 \rightarrow 10%/month, 0.01 \rightarrow 1%/month. General break model, B4: shift in mean, shift in coefficient on time trend:

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

Table 2. Power of various alternatives, in per cent. DGP: B1

Case	1	2	3	4	5
θ	1.00	1.00	1.00	2.00	2.00
α	–0.05	–0.30	–0.50	–0.30	–0.50
$T_b = 30$					
ADF	17.06	100.00	100.00	99.32	100.00
B1	16.74	97.85	100.00	99.53	100.00
B2	4.95	97.44	100.00	98.83	100.00
B3	18.50	98.78	100.0	99.31	100.00
B4	5.86	94.95	100.00	97.06	100.00
$T_b = 75$					
ADF	0.00 \leftarrow	81.10	100.00	0.05 \leftarrow	90.18
B1	14.97	97.80	100.00	99.46	100.00
B2	5.97	97.18	100.00	98.49	100.00
B3	0.68 \leftarrow	86.86	100.00	26.73 \leftarrow	99.70
B4	3.53	94.44	100.00	96.50	100.00

(continued)

Table 2. Continued

Case	1	2	3	4	5
$T_b = 120$					
ADF	0.00	77.12	100.00	0.00	83.62
B1	7.00	97.97	100.00	99.33	100.00
B2	7.16	97.74	100.00	98.90	100.00
B3	3.35	88.32	100.00	31.29 ←	99.78
B4	3.39	94.17	100.00	94.98	100.00

Notes: Each entry in the block is the percentage of the simulations in which the model rejects the null at the 5% significance level for each of the values of α used. The simulations are for 50 000 replications. Each replication follows the B1 with no lags, a mean of zero ($\mu = 0$), a time trend of zero ($\beta = 0$), a shift in mean of θ in the column heading and the α in the column heading. The increments are $u_t \sim N(0, 1)$, each series is 170 observations, the series is ‘warmed up’ with 20 observations and the models are estimated over the remaining 150 observations. For each replication, each model is estimated without lags because the DGP contains no lags. The row for the DGP model is in bold in each block; the largest entry in each column in the bloc is in **bold italic** (unless the DGP model’s entry is largest).

Mean shifts: One SD of 1.00. General break model, B4: shift in mean, shift in coefficient on time trend:

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

Table 3. Power of various alternatives, in per cent. DGP: B3

Case	1	2	3	4	5	6	7
ϕ	0.100	0.001	0.00001	0.00001	0.0001	0.001	0.005
α	-0.05	-0.05	-0.05	-0.30	-0.20	-0.20	-0.20
$T_b = 30$							
ADF	0.00	0.01	0.02	86.10	11.96	6.17	0.15
B1	0.00	0.06	0.19	93.02	34.34	28.75	7.37
B2	84.74	10.32	10.06	98.27	73.39	73.18	73.46
B3	87.38	8.37	8.16	98.81	73.11	72.55	72.42
B4	64.86	9.16	9.11	97.28	67.94	68.08	67.57
$T_b = 75$							
ADF	0.00	0.00	0.01	86.77	12.58	7.35	0.50
B1	0.00	0.14	0.10	92.81	35.08	30.64	14.27
B2	0.40	9.93	10.09	98.36	73.47	73.84	73.72
B3	61.47	8.42	8.33	98.76	72.78	72.81	72.98
B4	24.55	9.11	9.15	97.04	68.03	68.28	67.64
$T_b = 120$							
ADF	0.00	0.00	0.00	86.08	12.55	10.80	3.60
B1	0.00	0.08	0.12	92.95	35.57	33.59	22.68
B2	0.00	9.77	10.08	98.66	73.76	74.02	73.76
B3	25.46	8.48	8.18	98.95	73.06	72.95	72.76
B4	4.29	8.54	9.33	97.60	67.24	67.73	67.93

Notes: Each entry in the block is the percentage of the simulations in which the model rejects the null at the 5% significance level for each of the values of α used. The simulations are for 50 000 replications. Each replication follows the B1 with no lags, a mean of zero ($\mu = 0$), a time trend of zero ($\beta = 0$), a shift in mean of ϕ in the column heading and the α in the column heading. The increments are $u_t \sim N(0, 1)$, each series is 170 observations, the series is ‘warmed up’ with 20 observations and the models are estimated over the remaining 150 observations. For each replication, each model is estimated without lags because the DGP contains no lags. The row for the DGP model is in bold in each block; the largest entry in each column in the bloc is in **bold italic** (unless the DGP model’s entry is largest).

Time-trend shifts: 0.100 (10%/month), 0.005 (0.5%/month), 0.001 (0.1%/month), 0.0001 (0.01%/month), 0.00001 (0.001%/month). General break model, B4: Shift in mean, shift in coefficient on time trend:

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

Downloaded by [University of Gothenburg] at 23:24 07 November 2013

$T_b = 120$. Following Perron (1989) and others, for B4, the shifts in mean and time trend occur on the same date.

The power results in Table 1 are representative for the case researchers often face, a slow adjustment speed. The initial mean is $\mu = 0$ and the mean shift is one standard deviation of the error process. The initial time trend is $\beta = 0.01$, meaning 1%/month or 12%/year. One time-trend shift is 0.01 (1%/month or 12%/year), or a doubling of trend. A second is 0.10 (10%/month or 120%/year), a value unlikely save in crises; results for this case are useful, however, for exploring effects of large changes. First, for each DGP – corresponding to B1, B2 and two versions each of B3 and B4 models – the total percentage of rejections falls as T_b rises from period 30 to 75 to 120. There is a trivial reversal of B2 for $T_b = 75$ and $T_b = 120$ periods. Differences between the estimated and true break points have important effects; see Kim *et al.* (2000) and Montañés *et al.* (2005). This article has a different focus, however. In general, for all break models power is sensitive to where the break occurs in the data $T_b = 30$, $T_b = 75$, $T_b = 120$, but the models are sensitive in different ways. Second, the probability of the ADF rejecting is zero across all DGPs in Table 2 for $T_b = 75$ or $T_b = 120$. With a small $-\alpha$, even a modest parameter shift destroys the power of the ADF (to anticipate Tables 2 and 3, however, the ADF can maintain power if $-\alpha$ is larger). Third, for the DGPs B3 and B4, the B1 model seldom rejects. Note that ADF and B1 models contain neither a shift in time trend *nor* a time trend; even with only a modest shift in time trend of 0.01 – or 1%/month, implying a doubling of the time trend of 1%/month – models without a time trend have essentially no power. The B2 model, which contains a time trend, has relatively good power across DGPs – save for $T_b = 75$ and $T_b = 120$ for the DGP B3a with a large time-trend shift of 0.10, or 10%/month. Finding multiple rejections is more likely if $T_b = 30$ than if $T_b = 75 = 120$. In cases where $T_b = 30$ and a model other than the ADF rejects, the researcher may choose to believe the DGP corresponds to the model that rejects, though perhaps assigning a B2 rejection to a DGP corresponding to B3a, and suspecting that a rejection by B3 arises from a DGP corresponding to say B1, B2 or B4b.

Rejection frequencies for $T_b = 75$ and $T_b = 120$

For $T_b = 75$, for any DGP that corresponds to B1, B2, B3b or B4b, the maximum occurs for the same model as the

DGP. This suggests that in the absence of other information the researcher might take rejection by, say, the B1 model at face value as implying the DGP corresponds to B1. For $T_b = 120$, however, in a number of cases, the break model corresponding to the DGP has lower power than other models. If the DGP corresponds to B1, then the rejection frequencies are 7.00% for B1 but 7.16% for B2. If the DGP corresponds to B4b, then the B4 frequencies are 6.17% for B3 and 4.19% for B4. In other cases, the model corresponding to the DGP has the highest rejection rate but other models have somewhat similar rates. Thus, the researcher should be wary of assuming that the model that rejects the null – or rejects the null most strongly in multiple rejections – corresponds to the DGP.

Effects of the location of T_b on the probability of rejection

(i) Across the four DGPs and five estimating models in Table 1, frequencies tend to vary strongly as T_b increases; for example, if the DGP corresponds to B1, then the power for B1 is 16.74%, 14.97% and 7.00% for T_b rises from 30 to 75 to 120.⁴ (ii) Related, when the DGP corresponds to B2, the power for B3b is 18.91%, 0.63% and 3.29% as T_b rises; the researcher hence faces smaller danger of having the B3b model ‘steal’ power from the B2 model if $T_b = 75$ or $T_b = 120$. (iii) If the DGP corresponds to B3b, then rejection frequencies for the B3b model for $T_b = 30 = 75 = 120$ are close to each other. On balance, the researcher who finds a rejection by the B3b model for T_b of 75 or 120 might take this at face value. (iii) When the DGP corresponds to B4b, the B4 model has low frequencies of rejection, 6.48%, 8.15% and 4.19% for T_b rises from 30 to 75 to 120. Further, in the two ‘extreme’ panels – for B3a and B4a – B3 has greater power than B4 for $T_b = 30$ and $T_b = 120$. Thus, in finite samples, the data often ‘prefer’ a simpler model even if the true underlying model is B4b.⁵

III. Results for Ranges of Adjustment Speeds and Parameter Shifts

In Table 2 (shift in mean) and Table 3 (shift in the time trend), the anomalous results in Table 1 persist for $T_b = 30$. For $T_b = 75$ or $T_b = 120$, new results arise. First, from Table

⁴ Related, Kim *et al.* (2000) discuss how break points early in the sample lead to size distortions. Often researchers trim samples to avoid problems from early and late break points. Note that if the sample is trimmed by 15% at each end, it becomes approximately 23–128, including both $T_b = 30$ and $T_b = 120$.

⁵ The small rejection rates for B4 and $T_b = 120$ recur in repeated simulations. Rejections rates of less than 5% appear to arise because the iterative search for the break point is not very accurate. Montañés *et al.* (2005) discuss, with theoretical and simulation results, how selection of an incorrect break date affects the asymptotic and finite-sample distributions of the test statistic under the alternative. They use (a) a 100% speed of adjustment and (b) generally extreme parameter shifts; their simulations are unrevealing for many practical purposes, for example, evaluating real exchange rates.

2, if the DGP is the B1 model, for a relatively small shift in mean of one standard deviation, the power results depend crucially on α . For $\alpha = -0.05$, the power of the B1 alternative is only 14.97% for $T_b = 75$ and 7.00% for $T_b = 120$. For a DGP corresponding to B1, other models have substantially lower power, save for the B2 model for $T_b = 120$. Second, for $\alpha = -0.30$ or (even more strongly) $\alpha = -0.50$, all models have good power. Taken together, these two results suggest that multiple rejections in the face of mean shifts are unlikely unless $-\alpha$ is relatively large. As is intuitive, the ADF has no or minimal power in Table 2 unless $-\alpha$ is large enough to offset the misspecification from omitting the shift in mean. This is not quite conclusive because the important issue is the probability that, conditional on a model rejecting, one or more additional models reject. Table 4 addresses the issue of conditional probabilities for $\alpha = -0.05$.

Table 3 focuses on a DGP corresponding to the B3 model and examines three time-trend shifts: very large 0.10 (10%/month, 120%/year), small 0.005 or 0.001 (0.50%/month, 6%/year or 0.1%/month, 1.2%/year) and very small 0.00001 (0.001%/month, 0.012%/year). First, for $\alpha = -0.05$ models that do not include a time trend (ADF and B1) have very poor power no matter the magnitude of the time-trend shift. The B2 model contains a time trend, but has low power for the largest time-trend shift for $T_b = 75$ and $T_b = 120$ (but substantial power for $T_b = 30$). For the two smaller time-trend shifts, B2 has power comparable to B3 or B4 for all three T_b values. Second, for larger α values of -0.20 and -0.30 , power is good for all models if the time-trend shift is quite small, 0.00001 (0.001%/month,). As the time-trend shift increases from 0.0001 (0.01%/month) to 0.001 (0.10%/month) to 0.005 (0.50%/month), for $\alpha = -0.20$, the power of the ADF and B1 models declines monotonically, but the power of the B2, B3 and B4 models stays much the same.

Adjustment speeds of 20% to 50% per period

Often the researcher faces slow adjustment speeds. Relatively large adjustment speeds sometimes arise in practice, however. For example, Maican and Sweeney (2013a) report significant estimated adjustment speeds for nine Central and Eastern European countries, where the largest significant estimate for each country for $-\alpha$ is as follows:⁶

Bulgaria	Czech R.	Estonia	Hungary	Latvia	Lithuania
0.548	0.294	0.093	0.264	0.033	0.043
Romania	Slovakia	Slovenia			
0.303	0.411	0.339			

Six of the nine estimates correspond to values of $-\alpha$ explored in Tables 2 and 3. For a selection of cases, examples of the effects of fast adjustment speeds ($-\alpha$ large) or large mean shifts (θ large) are shown in Table 2. Once again, the time at which the break occurs (T_b) often has important effects on results. In cases (4) and (5) in Table 2, the shift in mean is the base-case shift 1.00 (equal to one standard deviation of the error process). The decrease in α from -0.30 to -0.50 raises the rejection rates for all of the models, but particularly those which were not already close to 100%. In models B2 and B4, $\alpha = -0.30$, but the shift in mean is 1.00 in case (4) and 2.00 in case (2). The larger shift in mean raises the rejection rate in the three models which include time trends (B2, B3 and B4), but reduces it for B1 and ADF; the lack of a time trend becomes more serious, the larger is the mean shift. Still, in going from case (2) to case (3), the decrease in α gives 100% rejection rates for all models; when the shift in mean is 2.00, the substantial misspecification of the ADF is more than offset by the smaller $\alpha = -0.50$.

For shifts in time trend, comparing the results for a small shift of 0.00001 for $\alpha = -0.05$ and $\alpha = -0.30$ shows a huge effect on rejection frequencies. For Cases 5, 6 and 7 in Table 3, where $\alpha = -0.20$ and the shifts in trend increase as 0.0001, 0.001 to 0.005, the rejection rates are essentially the same, as 72.78, 72.81 to 72.98 for $T_b = 75$ (and the same is true for $T_b = 30$ and $T_b = 120$). The speed of adjustment dominates the results, not the size of the trend shift. Across the same increases in the trend shift, for $T_b = 75$, the rejection rates for ADF fall as 12.58, 7.35 to 0.50 and for the B1 as 35.08, 30.64 to 14.27. Both ADF and B1 omit time trends and shifts in trends; as the shift increases, the large speed of adjustment is less and less able to offset the effect on rejections frequencies of the omission of the time trend and shift in trend.

Extreme adjustment speeds and parameter shifts in other work

In power investigations, Montañés *et al.* (2005, pp. 52–55) assume $\alpha = -1.00$, that is, 100% adjustment per period and a root of $(1 + \alpha) = 0$. This extreme α and extreme values in three of their four parameter shifts render many of their power simulation results uninformative for case commonly encountered in practice where $-\alpha$ is small and shifts are not extreme. In their simulations, when the DGP corresponds to B3 but B2 is instead fit, there are very few rejections (pp. 54–55) because of the interaction of $\alpha = -1.00$ and the shifts in the time trend are so extreme (either 100% or 300% per period, as opposed to this article's shifts of 1% or 10% per period). When the DGP

⁶ As is frequent in the literature, these results are unadjusted for the well-known downward bias in estimates of roots and thus upward bias in adjustment speed, or for bias from the time aggregation of prices.

corresponds to B4, both B2 and B3 have low power, again because of the interaction of $\alpha = -1.00$ and the extreme shifts in time trend and, in one case, a shift in mean of 10 standard deviations of the error process. At the other extreme in Montañés *et al.* (2005), when the DGP corresponds to B2 and the mean shift is one standard deviation of 1.0, the B3 and B4 break models reject at very high levels, over 70% in most cases and over 90% in some cases (Table 1, p. 53). Their mean shift corresponds to that in Table 3 here, where rejection rates are much lower.⁷

IV. Conditional Rejection: Rejections by Multiple Alternative Models

Results in Table 4 show how frequencies of conditional rejection for $\alpha = -0.05$ and $T_b = 75$ depend on the model to which the DGP corresponds. For the DGP corresponding to B1, the power of B1 is 14.97% (see the (power) row), or the null is rejected in 14.97% of the replications. The conditional frequency of B2 rejecting is 25.38%. Similarly, for the DGP corresponding to B2, the B2 model has power of 6.17% and the B4 model rejects in 36.63%, of the replications in which the B2 rejects. For the B3 models, B4 conditional rejections range from

approximately 40% to 50%; B2 rejections are negligible for B3a but approximately 25% for B3b. For B4 models, B3 conditional rejections are 66.02% and 5.767% for B4a and B4b, and B2 conditional rejections are 9.97% and 10.80%.

V. Conclusions

If the DGP corresponds to one of the break models, the likelihood of correctly rejecting the unit-root null in many cases turns importantly on correct specification of the alternative. Using the correct specification is more likely to be important in ‘hard cases’ frequently encountered in empirical work, where the speed of adjustment is small (the root is large) and the shift in mean or in time trend is moderate. The researcher might choose to test the null for only one alternative in hope of preserving size, but can go far astray unless the alternative chosen corresponds to the DGP. On the one hand, if the researcher misspecifies the alternative – a mistake unknowable in advance – then he may severely reduce power and thus importantly increase the probability of a Type II error. On the other hand, the model corresponding to the DGP may not have maximum power. In interpreting test results, it is valuable to know

Table 4. Frequency of conditional rejections across five models, for DGPs corresponding to break models ($T_b = 75$)

Case	B1	B2	B3a	B3b	B4a	B4b
θ	1.00	1.00	–	–	1.00	1.00
ϕ	–	–	0.10	0.01	0.10	0.01
α	-0.05	-0.05	-0.05	-0.05	-0.05	-0.05
ADF	0.00	0.00	0.00	0.00	0.00	0.00
B1	[14.97]≠	12.48	0.00	0.00	0.00	0.00
B2	25.38	[6.17]≠	0.76	24.85	9.97	10.80
B3	0.00	4.05	[61.47]≠	[10.30]≠	66.02	66.02
B4	11.49	36.63	39.60	48.48	[100.00]≠	[8.15]≠

Notes: Each entry in the block is the percentage of the simulations in which the model corresponding to the DGP rejects the null at the 5% significance level and the other models, taken one-by-one, also rejects the null. In the case where the DGP corresponds to the model B1, The entry in this column in brackets [14.97] is the percentage of times (from Table 2) that the B1 model rejects. In the column for the B1 model, conditional on the B1 model rejecting, the B2 model rejects 25.38% of the time. The simulations are for 50 000 replications. Each replication follows one of the four break models with no lags, a mean of zero ($\mu = 0$) and a time trend of 1%/month ($\beta = 0.01$); a shift in mean θ , a shift in the time trend ϕ and $\alpha = -0.05$. The break point is at date $T_b = 75$. The increments are $u_t \sim N(0, 1)$, each series is 170 observations, the series is ‘warmed up’ with 20 observations and the models are estimated over the remaining 150 observations. For each replication, each model is estimated without lags because the DGP contains no lags.

≠ Figures in brackets are the percentages of rejections from Table 2 for $T_b = 75$. **Time-trend shifts:** 0.10 (10%/month) for B3a and B4a and 0.01 (1.0%/month) for B3b and B4b. **Mean shifts:** One SD of 1.00.

General break model, B4: shift in mean, shift in coefficient on time trend:

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

⁷ For a mean shift of 10, the rejection rates in Montañés *et al.* (2005) depend strongly on the location of the break point used for estimation versus the true break point, measured by their ξ .

whether multiple models reject the null on the given data set. Some pairs of rejections are more likely than others when the null is false, depending on parameter shifts, time of the parameter shift in the sample period, speed of adjustment, etc.

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, with the costs that entails.

These considerations suggest that it is useful to run a battery of unit-root tests, to avoid overlooking the DGP model and to achieve good power. If the researcher does so, of course he should carefully document and discuss which models he explored and any pre-test exploration done to select alternatives. 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.

References

- Hecq, A. and Urbain, J. P. (1993) Misspecification tests, unit-root and level shifts, *Economics Letters*, **43**, 129–35.
- Hegwood, N. and Papell, D. (1998) Quasi purchasing power parity, *International Journal of Finance and Economics*, **3**, 279–89.
- Kim, D. and Perron, P. (2009) Unit roots tests allowing for a break in the trend function at an unknown time under both the null and alternative hypotheses, *Journal of Econometrics*, **148**, 1–13.
- Kim, T. H., Leybourne, S. J. and Newbold, P. (2000) Spurious rejections by Perron tests in the presence of a break, *Oxford Bulletin of Economics and Statistics*, **62**, 433–44.
- Lee, M. H. (1976) *Purchasing Power Parity*, Marcel Dekker, New York.
- Lothian, J. and Taylor, M. P. (1996) The recent float from the perspective of the past two centuries, *Journal of Political Economy*, **104**, 488–509.
- Maican, F. and Sweeney, R. (2013a) Real exchange rate adjustment in European transition countries, *Journal of Banking and Finance*, **37**, 907–26.
- Maican, F. and Sweeney, R. (2013b) Rejection Probabilities for a Battery of Unit-Root Tests, Working Papers Series, No. 568, 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.
- Muller, U. and Elliot, G. (2003) Tests for unit roots and the initial condition, *Econometrica*, **71**, 1269–86.
- Perron, P. (1989) The great crash, the oil price shock and the unit root hypothesis, *Econometrica*, **57**, 1361–401.
- Perron, P. (1997) Further evidence on breaking trend functions in macroeconomic variables, *Journal of Econometrics*, **80**, 355–86.
- Taylor, M. P. (2006) Real exchange rates and purchasing power parity: mean-reversion in economic thought, *Applied Financial Economics*, **16**, 1–17.
- 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**, 251–70.