Applied Economics, 2005, **37**, 607–617

Rank-based unit root testing in the presence of structural change under the null: simulation results and an application to US inflation

Steven Cook

Department of Economics, University of Wales Swansea, Singleton Park, Swansea SA2 8PP
E-mail: s.cook@swan.ac.uk

The size distortion of the Dickey-Fuller (*Journal of the American Statistical Association*, **74**, pp. 427–31, 1979) unit root test is examined in the presence of structural changes in both the level and variance of integrated time series. In contrast to previous studies, the empirically relevant situation in which such breaks occur simultaneously is examined. It is shown that the severe distortion observed for the Dickey-Fuller test can be dramatically reduced via application of a simple rank-based method. The simulation results presented are supported by an empirical examination of the integrated nature of US inflation where differing inferences are drawn using the Dickey-Fuller test and the rank-based Dickey-Fuller test.

I. Introduction

Examination of the order of integration of time series data has become a familiar feature of applied research in economics and finance, with the Dickey-Fuller (DF) (1979) test frequently employed. However, following the work of Perron (1989) it has long been recognized that the DF test can have low power when applied to series which are stationary but subject to a break in either level or trend. It is therefore possible that non-integrated economic and financial time series subject to structural change may be mistakenly classified as unit root processes on the basis of a DF test. In response to this failure to correctly reject the unit root hypothesis, a large literature has emerged examining the issue of unit root testing in the presence of structural change or regime shifts (see, *inter alia*, Banerjee *et al*., 1992; Perron, 1989, 1990; Zivot and Andrews, 1992). More recently, a converse Perron phenomenon has been proposed, whereby structural change under the null of a unit root results in the size distortion of the DF test. While the results of Leybourne *et al.* (1998) show the DF test to spuriously reject the null when applied to unit root processes subject to breaks in level or drift, the findings of Kim *et al.* (2002) show this phenomenon is also apparent in the presence of breaks in variance. These results are of obvious importance to practitioners given the prevalence of such structural change in economic and financial data (see, *inter alia*, Aggarwal *et al.*, 1999; Kim and Nelson, 1999; McConnell and Quirez, 2000; Sensier and van Dijk, 2004). In this study the impact of structural change under the null upon unit root tests is revisited, with two particular issues focused upon. First, the properties of the DF test are examined when breaks in level and variance occur simultaneously. The empirical relevance of this 'joint break' hypothesis has been noted by Busetti and Taylor (2003) who recognize that changes in the volatility of time series are often observed in conjunction with changes in mean.

Applied Economics ISSN 0003–6846 print/ISSN 1466–4283 online © 2005 Taylor & Francis Group Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/0003684042000337370

Second, the properties of the rank-based DF test of Granger and Hallman (1991) are examined. The use of rank-based procedures is recognized in the statistics literature as a means of achieving robust inference. In this study it is examined whether application of a rank-based procedure results in robust unit root testing in the present circumstances.

This paper proceeds as follows. In Section II Monte Carlo experimentation is undertaken to examine the properties of the DF test and the rank-based DF test in the presence of structural change under the null. To provide an empirical illustration to support the simulation results obtained, Section III presents an empirical examination of the integrated nature of a measure of US inflation. Section IV concludes.

II. Monte Carlo Analysis

The familiar DF τ_μ statistic tests for the presence of a unit root via examination of the null hypothesis $H_0: \beta = 0$ in the following model:[1]

$$\Delta y_t = \alpha + \beta y_{t-1} + \xi_t \qquad t = 1, \ldots, T \qquad (1)$$

The rank-based DF test proposed by Granger and Hallman (1991) simply involves replacing y_t with r_t, where r_t is the rank of y_t in $y_0, \ldots, y_T$. Testing of the unit root hypothesis is then achieved via examination of the null hypothesis $H_0: \beta^* = 0$ in the model below:

$$\Delta r_t = \alpha^* + \beta^* r_{t-1} + \xi_t^* \qquad t = 1, \ldots, T \qquad (2)$$

This test will be denoted here as τ_r. The τ_r test is of interest as in addition to possessing greater power than τ_μ test (see Granger and Hallman 1991, p. 219), the use of ranked data might be expected to result in robust inference. To explore possible robustness to breaks under the null, Monte Carlo experimentation is employed using the following data generation process (DGP):

$$y_t = \alpha s_t(\lambda) + \varepsilon_t \qquad t = 1, \ldots, T \qquad (3)$$

$$\varepsilon_t = \varepsilon_{t-1} + \eta_t \qquad (4)$$

$$\eta_t \sim i.i.d.\ \mathrm{N}(0, \sigma_t^2) \qquad (5)$$

$$s_t(\lambda) = \begin{cases} 0 & \text{for} \quad t \le \lambda T \\ 1 & \text{for} \quad t > \lambda T \end{cases} \qquad \lambda \in (0,1) \qquad (6)$$

$$\sigma_t^2 = \begin{cases} \sigma_1^2 & \text{for} \quad t \le \lambda T \\ \sigma_2^2 & \text{for} \quad t > \lambda T \end{cases} \qquad \lambda \in (0,1) \qquad (7)$$

The error series $\{\eta_t\}$ is generated using the RNDNS procedure in the Gauss programming language. All experiments are performed over 10 000 replications using a sample size of 100 observations with $\varepsilon_0 = 0$. The use of a sample of $T = 100$ follows the seminal analysis of Leybourne *et al.* (1998).[2] Denoting the break fraction as λ, breaks in level and/or variance can be imposed after observation λT via selection of appropriate values of α and $\{\sigma_1, \sigma_2\}$. In the present paper, previous research examining break in variance is followed with $\sigma_1 = 1$ imposed, and the break in variance set according to the ratio of the post- and pre-break standard deviations ($\delta = \sigma_2/\sigma_1$). Four sets of experimental designs are considered using alternative combinations of the design parameters $\{\alpha, \delta\}$. The first set of experiments relate to changes in level only ($\alpha \neq 0$, $\delta = 1$), while the second set considers breaks in variance only ($\alpha = 0, \delta \neq 1$). Although the performance of the τ_μ test in these circumstances has been considered previously in the literature, the properties of the τ_r test have not been examined. The final two sets of experiments examine the impact of simultaneous changes in level and variance ($\alpha \neq 0, \delta \neq 1$) upon the τ_μ and τ_r tests. Despite the empirical relevance of simultaneous changes in level and volatility (see Busetti and Taylor (2003) for discussion of this phenomenon), the literature has yet to consider the properties of either test in such circumstances. The two sets of experiments for joint breaks differ according to the size of the breaks considered. For all experiments, empirical rejection frequencies are reported at the 5% nominal level of significance. To the extent that these rejection frequencies differ substantially from 0.05, size distortion is present.

Level breaks

The properties of the τ_μ and τ_r tests in the presence of breaks in level only are examined by imposing ($\alpha \neq 0, \delta = 1$) in the above simulation DGP. The results obtained using the values $\alpha = \{0.5, 1, 4, 8\}$ are presented graphically in Figs 1–4.[3] The results show that the size distortion of the tests depends

[1] In this paper the DF test is considered when an intercept is included in the testing equation, as this specification is available for the rank-based test.

[2] Results for a single representative sample size are reported in the interests of brevity. Further similar results for alternative sample sizes are available from the author upon request.

[3] It was found that the absolute value of the break in level is of importance, with increasing and decreasing breaks of the same magnitude found to have the same impact upon rejection frequencies.

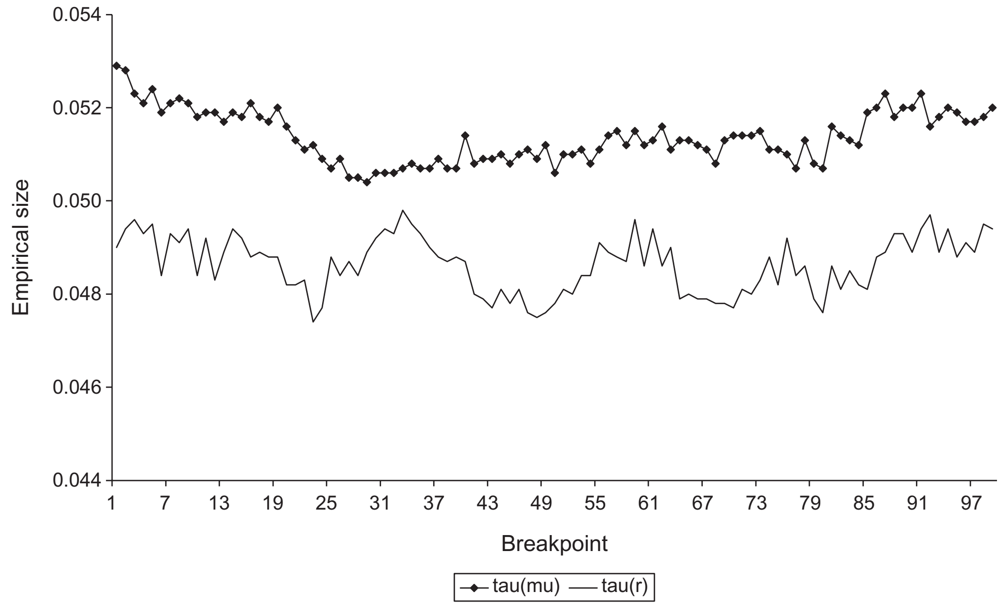

Fig. 1. Empirical sizes of the τ_μ and τ_r tests ($\alpha = 0.5, \delta = 1$).

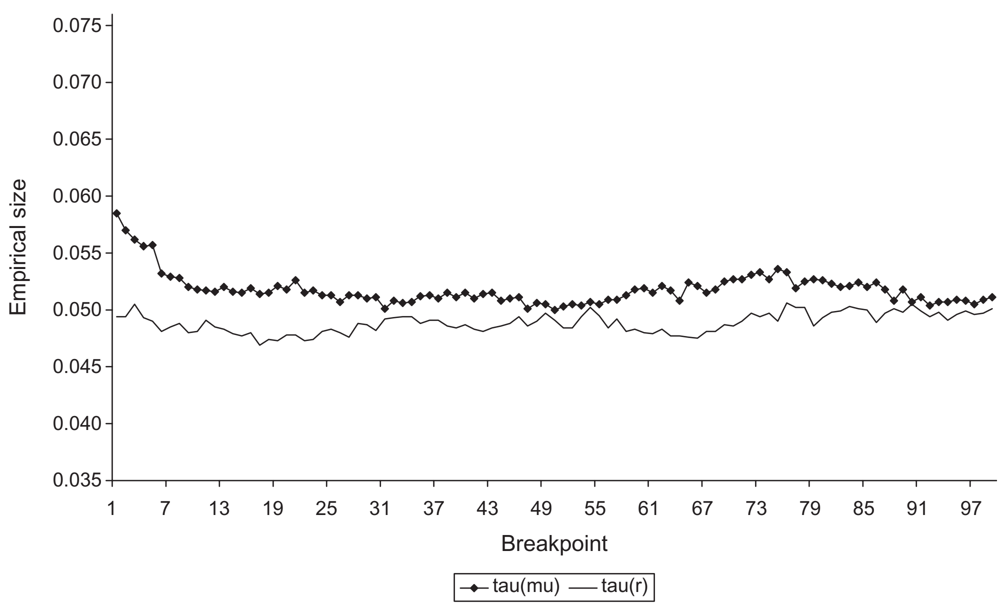

Fig. 2. Empirical sizes of the τ_μ and τ_r tests ($\alpha = 1, \delta = 1$).

upon the size of the break imposed. While Fig. 1 shows little evidence of spurious rejection by either test in the presence of a small break ($\alpha = 0.5$), the largest break ($\alpha = 8$) leads to substantial false rejection of the null. The results also show that the τ_μ suffers greater size distortion than the rank-based τ_r test, and that this distortion is maximized when the break in level occurs after the first observation in the sample period. Both of these issues are illustrated by the maximum rejection frequencies for the two tests, these being 39.2% for τ_μ and 9.6% for τ_r when $\alpha = 8$ and $\lambda T = 1$.

Variance breaks

The properties of the τ_μ and τ_r tests in the presence of variance breaks of differing sizes are presented in Figs 5–8. The sizes of breaks considered are given by $\delta = \{0.25, 0.5, 2, 4\}$. While the first two values impose decreasing breaks in variance $\left(\sigma_2^2 < \sigma_1^2\right)$,

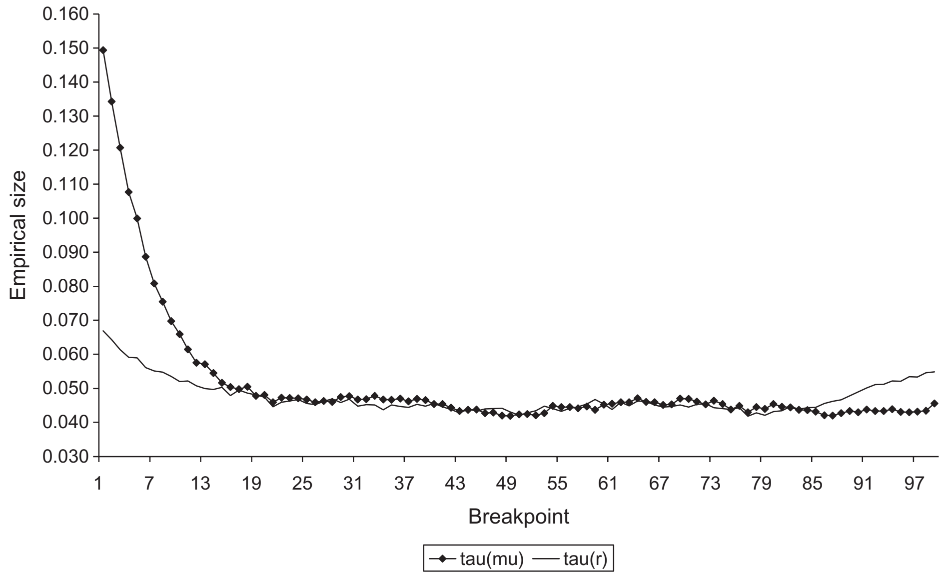

Fig. 3. Empirical sizes of the τ_μ and τ_r tests ($\alpha = 4, \delta = 1$).

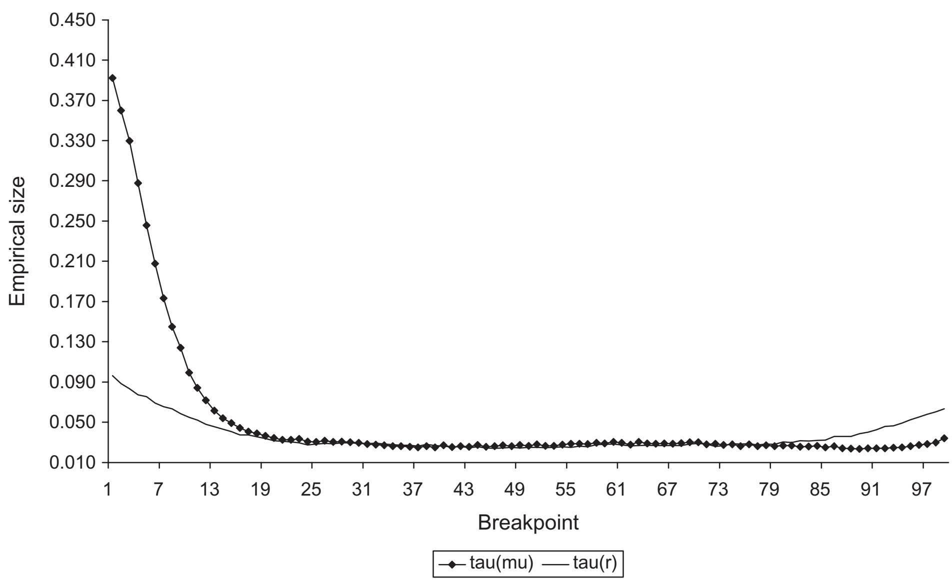

Fig. 4. Empirical sizes of the τ_μ and τ_r tests ($\alpha = 8, \delta = 1$).

the latter two impose increasing breaks $\left(\sigma_2^2 > \sigma_1^2\right)$. From inspection of Figs 5–8 it is apparent that decreasing and increasing breaks have very different effects. Considering decreasing breaks, the findings of Kim *et al.* (2002) are replicated with dramatic size distortion observed for the τ_μ test for breaks early in the sample period. Again, the extent of distortion depends upon the magnitude of the break and its position. Under the current experimental design, size distortion of the τ_μ test is maximized when $\delta = 0.25$ and $\lambda T = 14$, the observed rejection frequency being 42.9%. The results for the τ_r test show a similar pattern is followed in the presence of decreasing breaks in variance, albeit to a much lesser extent, with the corresponding maximum rejection frequency for this test being 15.76% when $\delta = 0.25$ and $\lambda T = 20$. Turning the results for increasing breaks in variance ($\delta > 1$), breaks early in the sample period now generate undersizing, this again being more apparent for the larger breaks considered. However, this undersizing becomes oversizing as the breakpoint is imposed later in the sample period. Interestingly, it can be seen that the τ_r test suffers

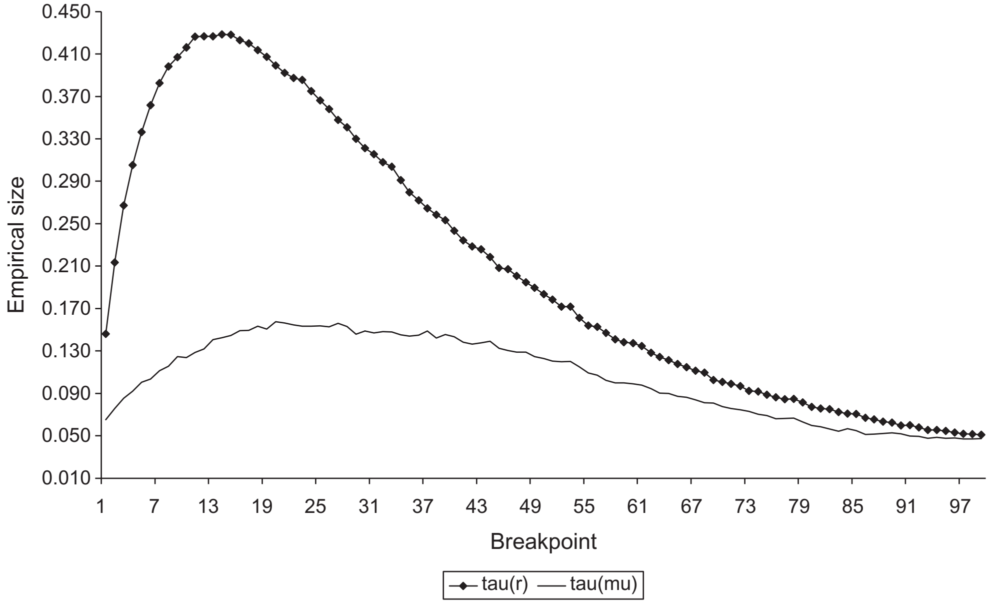

Fig. 5. Empirical sizes of the τ_μ and τ_r tests ($\alpha = 0, \delta = 0.25$).

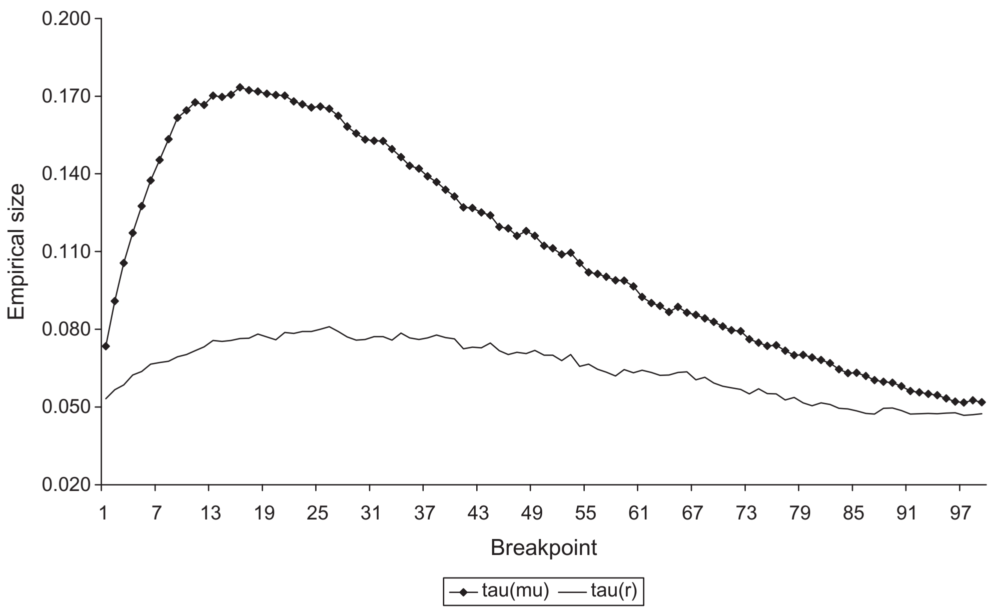

Fig. 6. Empirical sizes of the τ_μ and τ_r tests ($\alpha = 0, \delta = 0.5$).

greater size distortion than the τ_μ for these later breaks.

Level and variance breaks I

In Figs 9–12 empirical rejection frequencies are reported for simultaneous breaks in level and variance when the sizes of the breaks are related according to $\alpha = 2\delta$, where $\delta = \{0.25, 0.5, 2, 4\}$. In Figs 9 and 10 breaks in level are imposed jointly with decreasing breaks in variance. From inspection of these graphs, the timing of the break is again crucial with distortion greater for breaks relatively early in the sample period. Under these circumstances, the τ_r test is more robust than the τ_μ test with less distortion apparent. As an example of this, the maximum rejection frequencies of the two tests are 43.2% (τ_μ) and 15.6% (τ_r), these occurring under the more extreme joint break of $(\alpha, \delta) = (0.5, 0.25)$. These figures and the overall pattern of results in

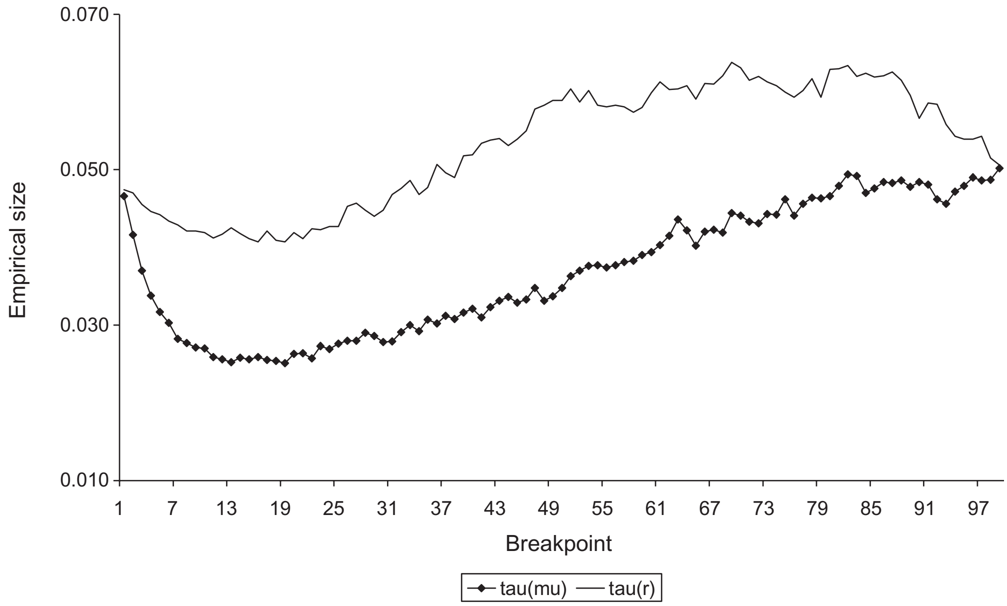

Fig. 7. Empirical sizes of the τ_μ and τ_r tests ($\alpha = 0, \delta = 2$).

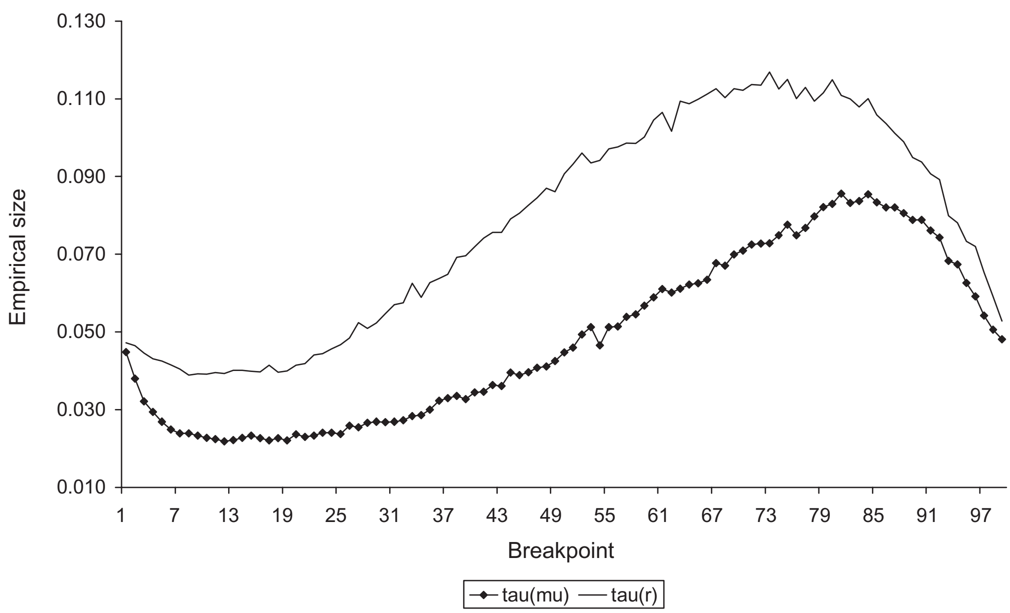

Fig. 8. Empirical sizes of the τ_μ and τ_r tests ($\alpha = 0, \delta = 4$).

Figs 9 and 10 are very similar to those obtained under breaks in variance alone, suggesting the variance break to have a greater impact than the break in level.

Turning to the results containing decreasing breaks in variance presented in Figs 11 and 12, the findings are again very similar to those presented previously for breaks in variance alone. In particular, early breaks are found to cause undersizing while later breaks result in oversizing. Interestingly, the τ_r test again exhibits greater oversizing than the τ_μ test although the extent of oversizing is substantially smaller than that reported for the τ_μ test in the presence of decreasing breaks in variance.

Level and variance breaks II

To allow further analysis of the impact of joint breaks in level and variance, Figs 13–16 present empirical rejection frequencies when the sizes of

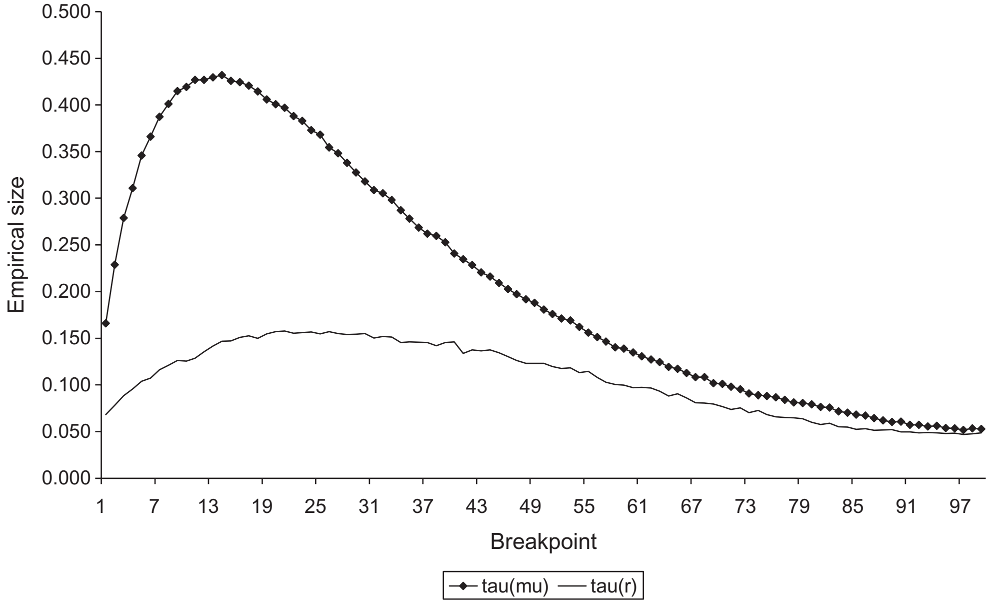

Fig. 9. Empirical sizes of the τ_μ and τ_r tests ($\alpha = 0.5, \delta = 0.25$).

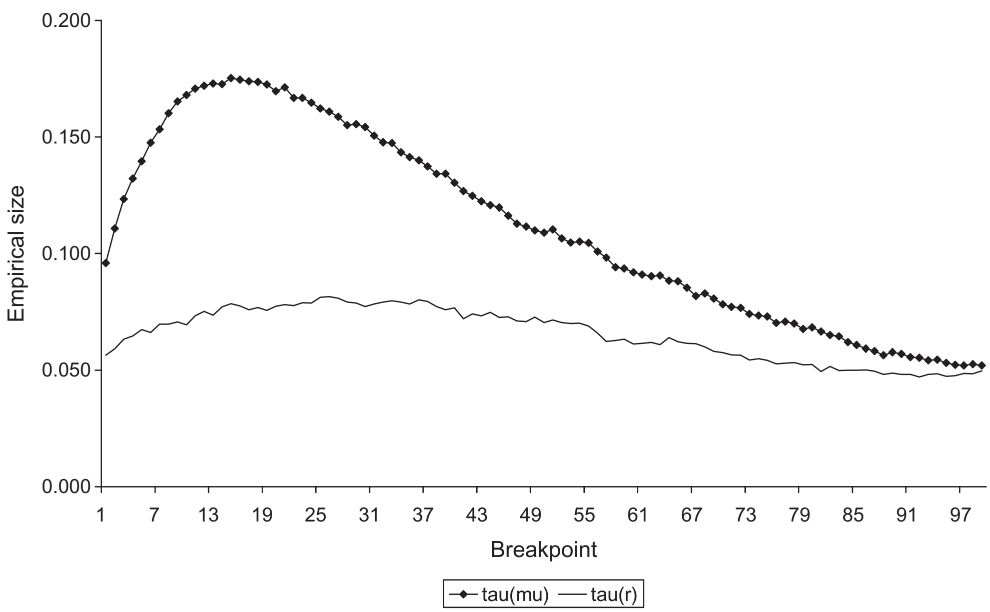

Fig. 10. Empirical sizes of the τ_μ and τ_r tests ($\alpha = 1, \delta = 0.5$).

the breaks are related according to $\alpha = 2/\delta$, where $\delta = \{0.25, 0.5, 2, 4\}$. This allows increasing breaks in variance to be matched with smaller breaks in level, while decreasing breaks in variance are associated with larger breaks in level, thereby reversing the above analysis of joint breaks. Considering the results in Figs 13 and 14, it can be seen that despite the presence of a break in level, similar results are above the design considering increasing breaks in variance in isolation. However, when the decreasing breaks in variance presented in Figs 15 and 16 are considered, very different results are obtained. In particular it can be seen that the previously obtained findings for level breaks are exacerbated, with the τ_μ test suffering severe size distortion. Indeed, for the most extreme case considered with $(\alpha, \delta) = (8, 0.25)$, the τ_μ test has an empirical rejection frequency of 98.6% when $\lambda T = 1$. In contrast, the τ_r test is more

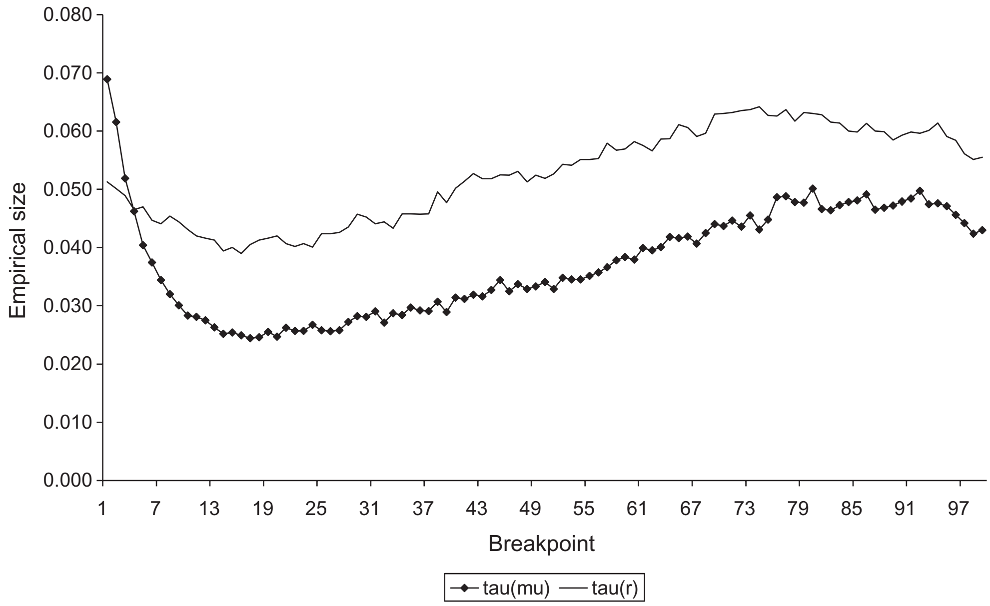

Fig. 11. Empirical sizes of the τ_μ and τ_r tests ($\alpha = 4, \delta = 2$).

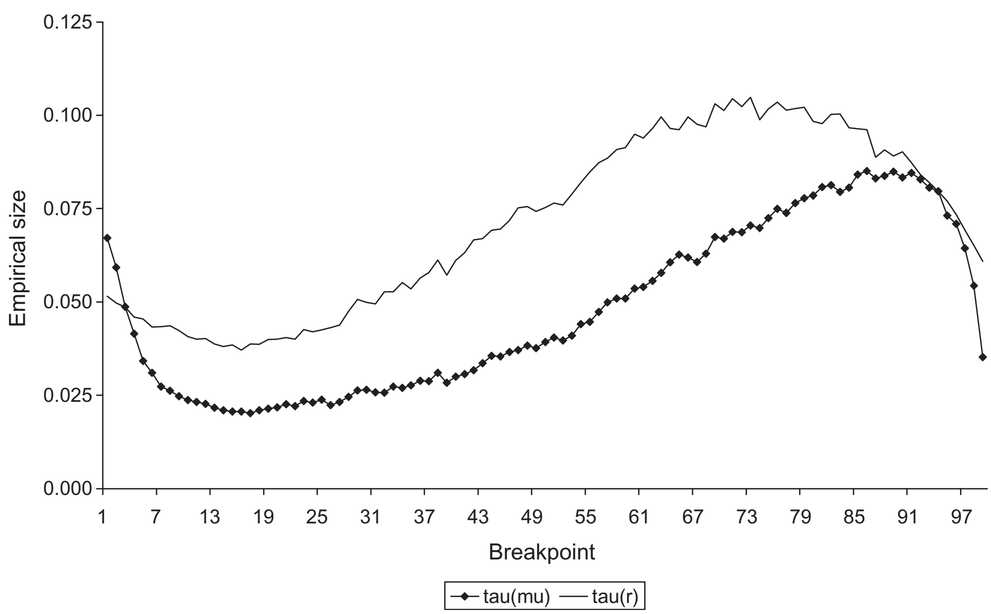

Fig. 12. Empirical sizes of the τ_μ and τ_r tests ($\alpha = 8, \delta = 4$).

robust with a maximum rejection frequency of 11.7% observed.

III. Empirical Analysis of US Inflation

To provide an empirical illustration of the above simulation results, the integrated nature of a measure of US inflation is examined. The data employed are seasonally adjusted monthly observations on the producer price index for the period January 1967 to December 1996 giving a total of 360 observations.[4] Denoting natural logarithm of this series as p_t, the inflation series Δp_t is analysed. This series has been selected as the results of Busetti and Taylor (2003) show that it possesses a clear break,

[4] The precise series used is the producer price index of finished goods excluding food.

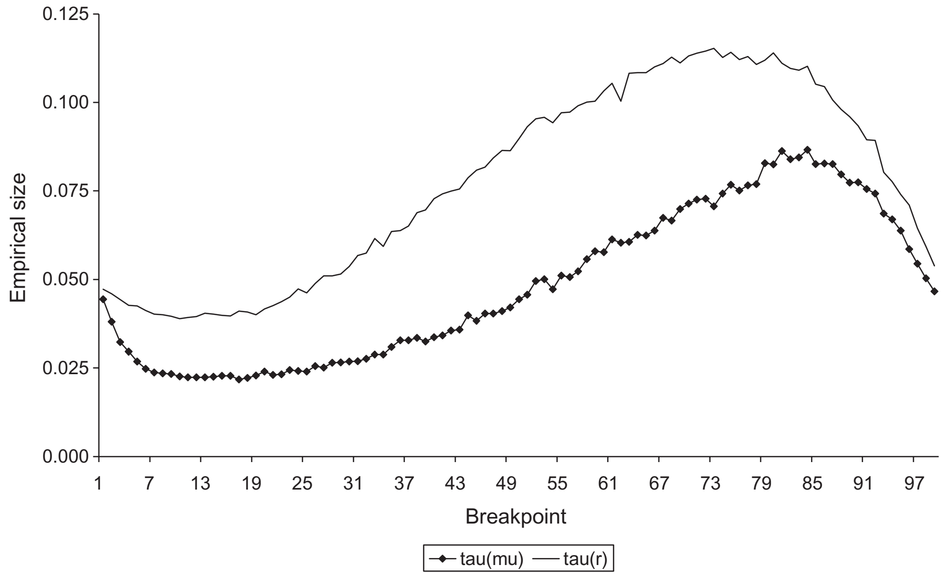

Fig. 13. Empirical sizes of the τ_μ and τ_r tests ($\alpha = 0.5, \delta = 4$).

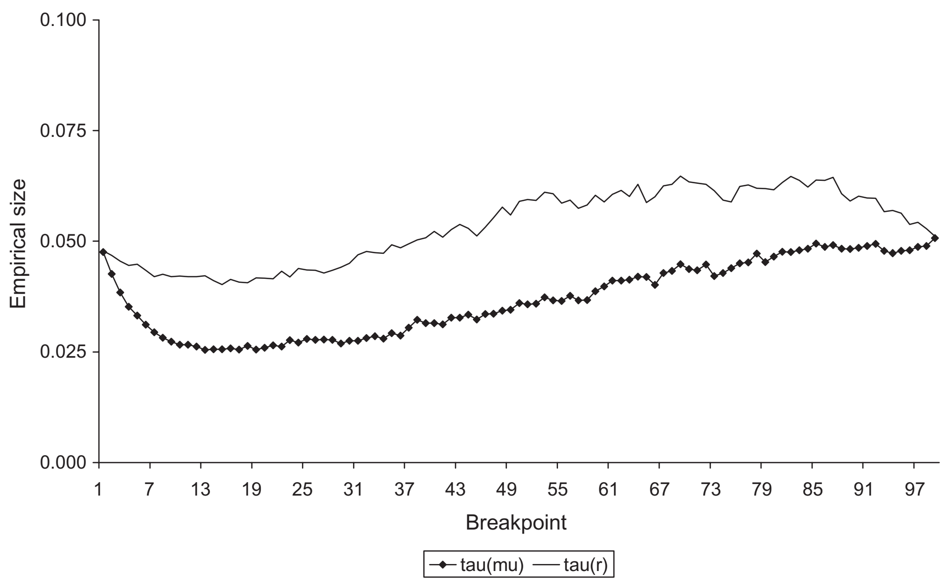

Fig. 14. Empirical sizes of the τ_μ and τ_r tests ($\alpha = 1, \delta = 2$).

but overwhelming rejects the null of stationarity when subject to break-robust stationarity tests. Alternatively expressed, the series appears to possess a unit root when the distortionary effects of level and variance breaks are allowed for. To examine whether the unit root hypothesis can be rejected for Δp_t, the above τ_μ and τ_r tests are employed in their augmented forms, using 12 lags to overcome the potential problem of serial correlation. The critical values to employ for these tests are presented in Table 1. The reported critical values are obtained from Monte Carlo experimentation for the augmented tests for the current sample size.[5] The calculated statistics obtained are −2.89 for the τ_μ test and −2.11 for the τ_r test. Comparing these values to the critical values in Table 1, the unit root hypothesis is rejected when the familiar τ_μ test is applied, but the null cannot be rejected under

[5] The reported critical values result from 50 000 replications of a standard unit root data generation process and application of the described augmented unit root tests.

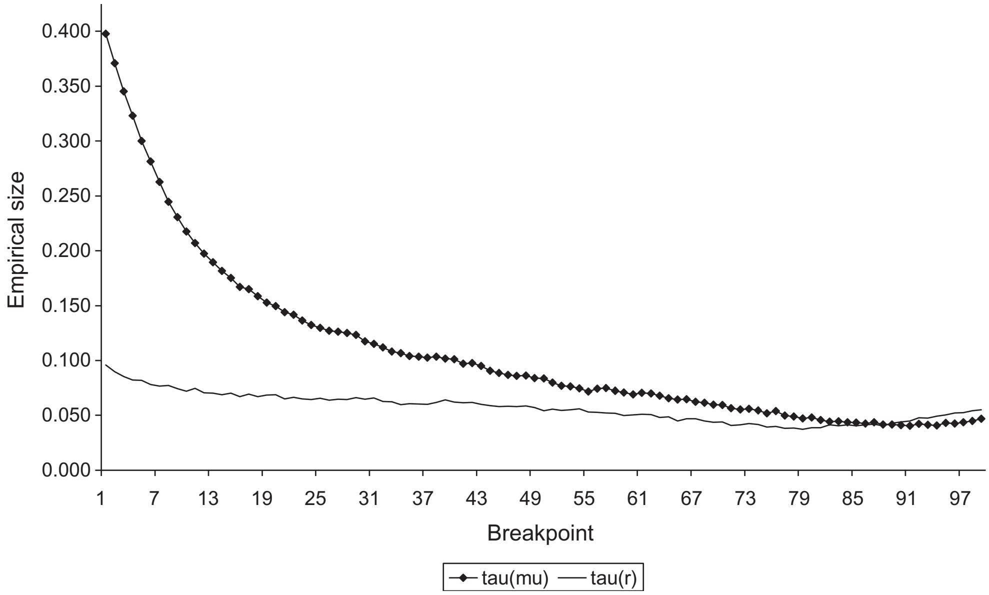

Fig. 15. Empirical sizes of the τ_μ and τ_r tests ($\alpha = 4, \delta = 0.5$).

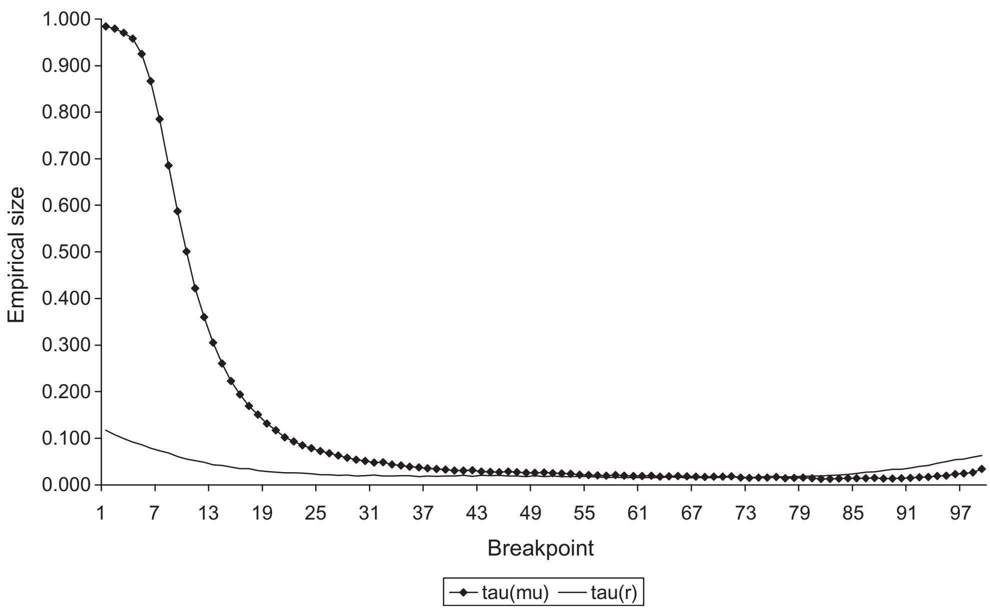

Fig. 16. Empirical sizes of the τ_μ and τ_r tests ($\alpha = 8, \delta = 0.25$).

application of the τ_r test, not even at the 10% level. Indeed, the simulation experimentation employed to calculate the critical values shows the rank-based test to have a *p*-value of 20.3%.

IV. Conclusion

In this study the size properties of the Dickey-Fuller (1979) and rank-based Dickey-Fuller tests have been examined in the presence of structural changes in both level and variance. The existing literature on the behaviour of the Dickey-Fuller test has been extended by considering the finite-sample distribution of the test in the empirically relevant situation when breaks in level and variance occur simultaneously. In addition, results have been presented for the previously unconsidered rank-based test. The simulation findings obtained show the application of the proposed rank-based method of Granger

Table 1. Critical values for the unit root tests

	Significance level		
Test	0.01	0.05	0.10
τ_μ	−3.43	−2.84	−2.54
τ_r	−3.25	−2.70	−2.43

and Hallman (1991) results in a substantially more robust test in the presence of breaks under the null. To illustrate the differing properties of the two tests, an empirical analysis of the integrated nature of US inflation was undertaken. The findings obtained showed the unit root to be rejected by the Dickey-Fuller test, but not the rank-based Dickey-Fuller test. These results support the conclusions drawn from the simulation analysis as the series examined was found to possess a unit root by Busetti and Taylor (2003) when subject to break-robust stationarity tests. In summary, the findings obtained in the present analysis show the use of a simple ranking procedure to increase the robustness of the simple Dickey-Fuller test. Combined with its known greater power, the results of the present study suggest this easily applied test may be of interest to applied researchers in economics and finance.

References

Aggarwal, R., Inclan, C. and Leal, R. (1999) Volatility in emerging stock markets, *Journal of Financial and Quantitative Analysis*, **34**, 33–55.

Banerjee, A., Lumsdaine, R. and Stock, J. (1992) Recursive and sequential tests of the unit root and trend break hypotheses: theory and international evidence, *Journal of Business and Economic Statistics*, **10**, 271–87.

Busetti, F. and Taylor, A. (2003) Variance shifts, structural breaks and stationarity tests, *Journal of Business and Economic Statistics*, **21**, 510–31.

Dickey, D. and Fuller, W. (1979) Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association*, **74**, 427–31.

Granger, C. and Hallman, J. (1991) Nonlinear transformations of integrated time series, *Journal of Time Series Analysis*, **12**, 207–18.

Kim, C. and Nelson, C. (1999) Has the US economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle, *Review of Economics and Statistics*, **81**, 608–16.

Kim, T.-H., Leybourne, S. and Newbold, P. (2002) Unit root tests with a break in innovation variance, *Journal of Econometrics*, **102**, 365–87.

Leybourne, S. (1995) Testing for unit roots using forward and reverse Dickey-Fuller regressions, *Oxford Bulletin of Economics and Statistics*, **57**, 559–71.

Leybourne, S., Mills, T. and Newbold, P. (1998) Spurious rejections by Dickey-Fuller tests in the presence of a break under the null, *Journal of Econometrics*, **87**, 191–203.

Leybourne, S. and Newbold, P. (2000) Behaviour of the standard and symmetric Dickey-Fuller-type tests when there is a break under the null hypothesis, *Econometrics Journal*, **3**, 1–15.

McConnell, M. and Quiros, G. (2000) Output fluctuations in the United States: what has changed since the early 1980s?, *American Economic Review*, **90**, 1464–76.

Perron, P. (1989) The Great Crash, the oil price shock and the unit root hypothesis, *Econometrica*, **57**, 1361–401.

Perron, P. (1990) Testing for a unit root in time series with a changing mean, *Journal of Business and Economic Statistics*, **8**, 153–62.

Sensier, M. and van Dijk, D. (2004) Testing for volatility changes in US macroeconomic time series, *Review of Economics and Statistics*, **86**(3), 833–9.

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.