Applied Economics, 1997, **29**, 1155–1161

Bootstrapping unit root tests

DANIELA DE ANGELIS*, STEFANO FACHIN** and G ALASTAIR YOUNG‡

* *MRC Biostatistics Unit, Cambridge, and PHLS AIDS Centre at CDSC, London, UK,* ** *Facoltà di Scienze Statistiche, Università di Roma 'La Sapienza', Italy and* ‡ *Statistical Laboratory, University of Cambridge, UK*

Unit root tests have been known for a long time to suffer from a variety of problems Considering that simulation methods are obvious candidates for solving some of these problems, the aim of this paper is to assess the performance of bootstrap tests of the unit root hypothesis of the Dickey–Fuller type The results obtained show that boostrap tests have empirical sizes very close to the nominal ones and deliver rejection rates generally at least as high as those obtained using simulated critical points, and are therefore a promising alternative to the latter The applications to non-standard problems such as structural stability analysis appear to be especially promising

I INTRODUCTION

Unit root tests are certainly one of the most popular tools of current applied econometric research They are, however, not free from problems Some of these, such as rather poor small sample performance, and the strong dependence of the tests on nuisance parameters, such as constant and trend terms (leading also to a certain awkwardness in the implementation of the tests due to the plethora of existing tables relevant to the different possible cases), have been evident for some time now (e g Dickey, 1984, Hylleberg and Mizon, 1989, Schmidt, 1990, Schwert, 1989) Others, such as the need to assess precisely the implications of choosing non-stationarity as the null hypothesis and dependence on the specific design of the Monte Carlo experiments that delivered the tabulated critical points, have been less stressed but are nevertheless not to be overlooked (the only reference for the latter point we are aware of is Schmidt, 1990, whereas the former has been recently examined in different ways by several authors, including, e g , Kwiatkowski *et al* , 1992 and Rudebusch, 1992) Recently some authors have used simulation methods as a means of solving some of these problems Rudebusch (1992) and Harris (1992) based their analysis on the bootstrap (Efron, 1979), whereas Blangiewicz and Charemza (1990) preferred more traditional parametric Monte Carlo methods A further important advantage of simulation methods is the ability to deliver simple and ready solutions to new problems Examples are, *inter alia*, Christiano (1992), who used the bootstrap to calculate data-specific critical points for Perron's unit root tests allowing for breaks in the deterministic kernel of the data generating process, and Fachin (1995), who used it to carry out tests of cointegration based on the Box–Tiao estimator

However, applied workers often do not seem to realize that the properties of simulation-based methods need to be thoroughly investigated (indeed, as it is routinely done with their asymptotic ones) before their results might be relied upon Now, the properties of simulation methods applied to unit root testing have not yet been properly investigated The extremely heavy computational burden involved in Monte Carlo analysis of bootstrap procedures is obviously an explanation, but by no means an excuse, for this gap in the literature The aim of this paper is thus to assess the performance of bootstrap tests of the unit root hypothesis of the Dickey Fuller (henceforth DF) type These tests, by far the most popular are also the most easily adaptable to the bootstrap Although somehow specific, our results will clearly shed some light on the more general issue of applying resampling methods to unit root testing The paper is organized as follows In the next section we shall outline the structure and rationale of the bootstrap as

0003–6846 © 1997 Routledge

applied to DF tests, in Section III we shall present the design of the Monte Carlo experiment, finally, the results obtained are discussed and some conclusions are drawn in Section IV

II BOOTSTRAPPING DICKEY-FULLER TESTS

The application of the bootstrap to DF tests is overall rather straightforward, but it does require care in some crucial points The key idea of this resampling method, proposed by Efron (1979), is that of constructing a pseudopopulation by repeated resampling from the data set under study Once a pseudopopulation has been constructed, an empirical estimate of the distribution of the statistic of interest can be trivially calculated, allowing hypothesis testing and construction of confidence intervals Unfortunately, resampling requires independence, or at least exchangeability, of the observations, an unlikely occurrence in time series A simple solution is provided by a two-stage scheme whose asymptotic validity has been proved for the AR(1) model by Bose (1988), and for the general AR(∞) case by Paparoditis (1994) The idea is simply to fit an autoregressive model to the data, resample the (empirically white noise) residuals, and construct the pseudo-series dynamically using these resampled residuals and the estimated coefficients of the AR model When the variable is suspected to be non-stationary, there is however a crucial point to be taken into account given that the bootstrap fails in the unit root case (see Basawa *et al*, 1991a, 1991b), the residuals have to be estimated imposing the null

The procedure as applied to a series is the following (see e g Harris, 1992)

(i) Calculate the empirical test-statistic (say, τ) and estimate the residuals imposing the null in the simple DF case, $e_t = x_t - c - x_{t-1}$, whereas when an augmented equation is required we shall have $e_t = x_t - c - x_{t-1} - \sum_{j=1}^{p} a_j \Delta x_{t-j}$, in both cases c and the as are OLS estimates

(ii) Draw a random sample with replacement from the residuals centred on the mean, obtaining, say, $\{e_t^*\}$ Further corrections such as dividing by the factor $\sqrt{1 - kT^{-1}}$, suggested by Wu (1986), where k is the number of variables in the autoregression and T the number of observations, may be used

(iii) Construct the pseudo-series in the DF case $x_t^* = c + x_{t-1}^* + e_t^*$, whereas in the ADF case $x_t^* = c + x_{t-1}^* + \sum_{j=1}^{p} a_j \Delta x_{t-j}^* + e_t^*$ In both cases $x_0^* = x_0$

(iv) Calculate the DF or ADF test statistic for the $\{x_t^*\}$ series (say, τ^*)

Repeating (ii), (iii) and (iv) a large number of times we shall obtain an empirical estimate of the distribution of τ^* The null hypothesis (x is non-stationary) will then be rejected at a significance level α if prop($\tau^* < \tau$) $= \alpha$, i e if the proportion of this empirical distribution falling on the left of τ, the test statistic calculated for the actual series $\{x_t\}$, is equal to α This proportion is in fact a consistent estimate of the size of the test To see why, consider the case of an NID series (the assumption of normality is just a simplification that can be relaxed asymptotically with no consequences) In such a case the resampled residuals of the fitted DF equation will also be normally distributed, and the pseudo-series constructed under the null hypothesis of a unit root (point (iv) above) thus include a sum of normally distributed increments The t-statistic for the unit root hypothesis computed on the pseudo-series will therefore converge by construction to the Dickey-Fuller distribution On the other hand, the OLS estimate $\hat{\rho}$ of the AR(1) coefficient computed on the original stationary data will satisfy $\hat{\rho} \xrightarrow{p} \bar{\rho}$, with $|\bar{\rho}| < 1$ (Stock, 1994), implying $T(\hat{\rho} - 1) \xrightarrow{p} \infty$ Given that the bootstrap test essentially compares a statistic based on $T(\hat{\rho} - 1)$ with a Dickey-Fuller distribution it is thus clearly consistent

Before proceeding, there is a further point that needs to be examined in some detail Step (ii) entails taking the first differences in order to impose the null of a unit root, when this hypothesis is actually false the differenced series will have a unit root in the MA part, and thus, strictly speaking, will not have an AR representation However, empirically this is not a problem Cochrane (1991) has shown that for any stationary process there exists a companion, non-stationary process arbitrarily close in terms of likelihood and autocorrelation structure This similarity clearly also holds for the residuals of any AR model fitted to both series, as these are simply linear combinations of the series We shall thus always be able to obtain empirically uncorrelated residuals allowing resampling [1]

As an illustration, consider the application to Italian GNP at current prices (Y) for the period 1955 91 (data from Prometeia and ISTAT) Given that a similar example is discussed at length by Harris (1992), we shall not go much into the details Applying the standard Dolado–Jenkison type testing strategy we conclude that the null of non-stationarity should not be rejected in a final equation including no constant, no trend (the exclusion of both constant and trend is suggested by the statistics $\phi_3 = 4\ 21$ and $\phi_1 = 1\ 32$, both not significant at 5% level), and three lagged differences, as the ADF statistic is 0 91 Having established that the residuals are empirically white noise by

[1] This has been confirmed by some exploratory simulations, not reported for reasons of space

means of standard LM autocorrelation tests and ARCH heteroscedasticity tests, we can proceed to produce quickly and cheaply[2] through the bootstrap an empirical estimate of the distribution of the τ test statistics with 5000 drawings the significance level turns out to be 96% (estimates accurate to the second decimal figure are actually delivered by a bootstrap with just 1000 drawings). Similar estimates could have been easily obtained for the ϕ statistics as well.

The mechanics of the testing strategy are therefore quite straightforward. There are, however, two crucial questions to be considered.

(i) Is the empirical size close to its nominal level? In other words, is it a good estimate of the 'true' size?
(ii) What are the power properties of the procedure? How do they compare with those of the standard procedure based on tabulated critical points?

Having established the framework and points of main concern, we can move to the experimental design and results.

III DESIGN OF THE MONTE CARLO EXPERIMENT

We based our experiment on a very simple design $x_t = \rho x_{t-1} + e_t$, $t = 1, \ldots, T$, in each experiment, x_0 has been generated from the same distribution as the errors (see (iii) below),

(i) Focusing especially on small sample performance, we chose $T = 20, 30, 50$.
(ii) Given that the power properties are particularly interesting in the region close to the null, $\rho = 0.8, 0.9, 1.0$.
(iii) $e_t \sim \text{IID}$ $N(0, 1)$, $t(3)$, $\chi^2(1)$ (centred on the mean), a mixture of Normals $0.5N(-2, 1) + 0.5N(2, 1)$. The first case is standard to all applications and is the basis of the existing tabulations, the $t(3)$ distribution has fatter tails than the Normal distribution, and is therefore a plausible alternative to it in an economic context, $\chi^2(1)$ is strongly skewed, and though less plausible as an error-generating mechanism over a long period, should be considered,[3] finally, the mixture of normals is bimodal.
(iv) The number of Monte Carlo replications was set to 10 000, whereas that of bootstrap drawings to 5000.[4]

Given the structure of the series, we can limit the experiment to the simpler type of DF test, without constant or trend terms. Recalling the discussion of Section II, the bootstrap procedure rejects the null of non-stationarity at a significance level α if $\text{prop}(\tau^* < \tau) = \alpha$, where τ^* are the test-statistics calculated on the bootstrapped pseudo-series and τ that calculated on the actual series. Given that in our experiment we could obviously consider only a limited set of sizes, this definition had to be generalized: the null is rejected at a significance level α if $\text{prop}(\tau^* < \tau) \leqslant \alpha$. The Monte Carlo estimate of the size of bootstrap tests, $\hat{\alpha}_b$, is simply given by the number of series generated under the null that lead to reject it, divided by the total number of series generated. Defining $p_i = \text{prop}(\tau^* < \tau)$ for the ith series

$$\hat{\alpha}_b = \text{prop}(p_i < \alpha | H_0)$$

whereas the empirical power $\hat{p}_b$ is

$$\hat{p}_b = \text{prop}(p_i < \alpha | H_1)$$

For the bootstrap tests we considered nominal sizes ranging from 1% to 10% at 1% steps, whereas the three nominal sizes for which tabulated critical points are available are obviously 1%, 5% and 10%. Defining τ_i as the DF statistic for the ith series, the Monte Carlo estimates of size and power of the standard tests based on tabulated critical points are obviously defined as, respectively,

$$\hat{\alpha}_t = \text{prop}(\tau_i < c_\alpha | H_0)$$

and

$$\hat{p}_t = \text{prop}(\tau_i < c_\alpha | H_1)$$

where c_α is the MacKinnon (1991) critical point at a significance level α.

IV RESULTS AND CONCLUSIONS

The results of the experiments, reported in Tables 1–4, suggest several considerations. First of all, consistent with those reported by Schmidt (1990), the standard procedure seems to be very robust with respect to the generating process of the errors. However, the boostrap tests seem to be somehow superior, given that, having empirical sizes always

[2]A routine written in GAUSS completed 5000 drawings in less than two minutes on a 486 PC. Popular command-driven econometric packages such as SHAZAM also include bootstrap routines.
[3]The t and χ^2 cases have been considered by Schmidt (1990), who reports critical points for these cases not very different from the standard normal case.
[4]The routines, written in FORTRAN and using NAG subroutines and random number generators, took several days of computing time on HP UNIX workstations.

Table 1 *Errors* N(*0, 1*)

		$\rho = 1\ 0$									
		Nominal (α) and empirical sizes of bootstrap ($\hat{\alpha}_b$) and standard ($\hat{\alpha}_t$) tests									
	α	0 01	0 02	0 03	0 04	0 05	0 06	0 07	0 08	0 09	0 10
$T = 20$	$\hat{\alpha}_b$	0 01	0 0187	0 0297	0 0409	0 0510	0 0642	0 0748	0 0833	0 0951	0 1057
	$\hat{\alpha}_t$	0 0093				0 0514					0 1010
$T = 30$	$\hat{\alpha}_b$	0 0091	0 0187	0 0293	0 0395	0 0480	0 0578	0 0693	0 0802	0 0898	0 0993
	$\hat{\alpha}_t$	0 0090				0 0485					0 0960
$T = 50$	$\hat{\alpha}_b$	0 0104	0 0190	0 0294	0 0376	0 0479	0 0595	0 0703	0 0808	0 0902	0 0999
	$\hat{\alpha}_t$	0 0098				0 0471					0 0968
		$\rho = 0\ 9$									
		Empirical power of bootstrap ($\hat{p}_b$) and standard ($\hat{p}_t$) tests									
	α	0 01	0 02	0 03	0 04	0 05	0 06	0 07	0 08	0 09	0 10
$T = 20$	$\hat{p}_b$	0 0275	0 0528	0 0802	0 1030	0 1266	0 1511	0 1761	0 1997	0 2212	0 2460
	$\hat{p}_t$	0 0281				0 1282					0 2363
$T = 30$	$\hat{p}_b$	0 0412	0 0772	0 1143	0 1478	0 1814	0 2156	0 2494	0 2789	0 3119	0 3421
	$\hat{p}_t$	0 0417				0 1818					0 3320
$T = 50$	$\hat{p}_b$	0 08340	0 1593	0 2237	0 2814	0 3357	0 3854	0 4301	0 4773	0 5183	0 5570
	$\hat{p}_t$	0 08520				0 3318					0 5444
		$\rho = 0\ 8$									
		Empirical power of bootstrap ($\hat{p}_b$) and standard ($\hat{p}_t$) tests									
	α	0 01	0 02	0 03	0 04	0 05	0 06	0 07	0 08	0 09	0 10
$T = 20$	$\hat{p}_b$	0 0634	0 1231	0 1724	0 2252	0 2759	0 3221	0 3613	0 3985	0 4338	0 4680
	$\hat{p}_t$	0 0645				0 2746					0 4566
$T = 30$	$\hat{p}_b$	0 1268	0 2306	0 3126	0 3886	0 4544	0 5124	0 5650	0 6140	0 6542	0 6888
	$\hat{p}_t$	0 1285				0 4522					0 6759
$T = 50$	$\hat{p}_b$	0 3570	0 531	0 6502	0 7330	0 7921	0 8383	0 8740	0 9004	0 9240	0 9403
	$\hat{p}_t$	0 3569				0 7892					0 9335

Table 2 *Errors* (*0 5*N(*−2, 1*) + *0 5*N(*2, 1*))

		$\rho = 1\ 0$									
		Nominal (α) and empirical sizes of bootstrap ($\hat{\alpha}_b$) and standard ($\hat{\alpha}_t$) tests									
	α	0 01	0 02	0 03	0 04	0 05	0 06	0 07	0 08	0 09	0 10
$T = 20$	$\hat{\alpha}_b$	0 0107	0 0198	0 0288	0 0398	0 0493	0 0617	0 0690	0 0778	0 0871	0 0984
	$\hat{\alpha}_t$	0 0116				0 0513					0 0967
$T = 30$	$\hat{\alpha}_b$	0 0099	0 0204	0 0295	0 0394	0 0496	0 0604	0 0715	0 0820	0 0932	0 1036
	$\hat{\alpha}_t$	0 01				0 0504					0 1013
$T = 50$	$\hat{\alpha}_b$	0 0109	0 0217	0 0324	0 0435	0 0522	0 0602	0 713	0 0815	0 0929	0 1039
	$\hat{\alpha}_t$	0 0116				0 0530					0 102
		$\rho = 0\ 9$									
		Empirical power of bootstrap ($\hat{p}_b$) and standard ($\hat{p}_t$) tests									
	α	0 01	0 02	0 03	0 04	0 05	0 06	0 07	0 08	0 09	0 10
$T = 20$	$\hat{p}_b$	0 0256	0 0492	0 0747	0 1003	0 1282	0 1524	0 1767	0 1992	0 2207	0 2423
	$\hat{p}_t$	0 0280				0 1320					0 2385
$T = 30$	$\hat{p}_b$	0 0382	0 07760	0 1130	0 1479	0 1865	0 2188	0 2493	0 2799	0 3101	0 3394
	$\hat{p}_t$	0 0499				0 1869					0 3319
$T = 50$	$\hat{p}_b$	0 0861	0 1606	0 2272	0 2889	0 3385	0 3889	0 4324	0 4758	0 5152	0 5559
	$\hat{p}_t$	0 0883				0 3380					0 5455
		$\rho = 0\ 8$									
		Empirical power of bootstrap ($\hat{p}_b$) and standard ($\hat{p}_t$) tests									
	α	0 01	0 02	0 03	0 04	0 05	0 06	0 07	0 08	0 09	0 10
$T = 20$	$\hat{p}_b$	0 0646	0 1208	0 1761	0 2246	0 2664	0 3069	0 3470	0 3874	0 4241	0 4604
	$\hat{p}_t$	0 0678				0 2704					0 4532
$T = 30$	$\hat{p}_b$	0 1282	0 2302	0 3148	0 3950	0 4478	0 5052	0 5548	0 6015	0 6406	0 6791
	$\hat{p}_t$	0 1333				0 4515					0 6706
$T = 50$	$\hat{p}_b$	0 3594	0 5359	0 6488	0 7338	0 7938	0 8439	0 8793	0 9069	0 9254	0 9410
	$\hat{p}_t$	0 3626				0 7922					0 9356

Table 3 *Errors centred* $\chi^2(1)$

$\rho = 1\ 0$

Nominal (α) and empirical sizes of bootstrap ($\hat{\alpha}_b$) and standard ($\hat{\alpha}_t$) tests

	α	0 01	0 02	0 03	0 04	0 05	0 06	0 07	0 08	0 09	0 10
$T = 20$	$\hat{\alpha}_b$	0 0078	0 0168	0 0264	0 0351	0 0451	0 0547	0 0660	0 0770	0 0858	0 0963
	$\hat{\alpha}$	0 0061				0 0408					0 0861
$T = 30$	$\hat{\alpha}_b$	0 0098	0 0186	0 0302	0 0391	0 0484	0 0583	0 0685	0 0789	0 0888	0 0983
	$\hat{\alpha}_t$	0 0075				0 0424					0 0884
$T = 50$	$\hat{\alpha}_b$	0 0108	0 0209	0 0316	0 0421	0 0520	0 0628	0 0723	0 0827	0 0931	0 104
	$\hat{\alpha}_t$	0 0082				0 0476					0 093

$\rho = 0\ 9$

Empirical power of bootstrap ($\hat{p}_b$) and standard ($\hat{p}_t$) tests

	α	0 01	0 02	0 03	0 04	0 05	0 06	0 07	0 08	0 09	0 10
$T = 20$	$\hat{p}_b$	0 0231	0 0496	0 0754	0 1016	0 1259	0 1515	0 1722	0 1921	0 2141	0 2357
	$\hat{p}_t$	0 0170				0 1119					0 2140
$T = 30$	$\hat{p}_b$	0 0391	0 0804	0 1180	0 1527	0 1891	0 2213	0 2530	0 2853	0 3146	0 3419
	$\hat{p}_t$	0 0293				0 1670					0 3164
$T = 50$	$\hat{p}_b$	0 0907	0 1649	0 2301	0 2856	0 3432	0 3973	0 4413	0 4849	0 5265	0 5621
	$\hat{p}_t$	0 0746				0 3130					0 5352

$\rho = 0\ 8$

Empirical power of bootstrap ($\hat{p}_b$) and standard ($\hat{p}_t$) tests

	α	0 01	0 02	0 03	0 04	0 05	0 06	0 07	0 08	0 09	0 10
$T = 20$	$\hat{p}_b$	0 0657	0 1277	0 1811	0 2272	0 2696	0 3112	0 3512	0 3867	0 4211	0 4550
	$\hat{p}_t$	0 0469				0 2421					0 4195
$T = 30$	$\hat{p}_b$	0 1417	0 2412	0 3288	0 4007	0 4645	0 5196	0 5656	0 6066	0 6478	0 6810
	$\hat{p}_t$	0 1126				0 4329					0 6496
$T = 50$	$\hat{p}_b$	0 3890	0 5604	0 6725	0 7495	0 7997	0 8402	0 8701	0 8920	0 9099	0 9253
	$\hat{p}_t$	0 3459				0 7822					0 9135

very close to the nominal ones. they deliver rejection rates generally at least as high as those obtained using the simulated critical points More specifically, when the errors are generated from a Normal distribution or from a mixture of Normals, the two methods are approximately equivalent, the bootstrap appearing to perform sometimes marginally better There is instead some difference when the errors are generated from χ^2 or t distributions in the smaller sample sizes the gain obtainable using the bootstrap can be as much as 10% of the power of the standard procedure Strategies such as that outlined in Section II, and followed, for example, by Rudebusch (1992) and Harris (1992) therefore appear to have a good asymptotic (in terms of bootstrap drawings) justification The application of the bootstrap to unit roots tests appears therefore to present two major points of interest The first is the ability to cope with non-standard situations (such as, for example, the need to augment the Dickey–Fuller test equation with exogenous variables or the sequential testing approach as required by break-searching procedures) which is granted by its data-based nature, the second is the estimation of the empirical size As far as power is concerned, we unfortunately have to conclude that the often quite poor small sample power performance of the examined tests cannot be improved much by the bootstrap when the sample size is as small as 20 or 30 observations the power can be just marginally higher than the significance level (for instance, with errors $N(0, 1)$, $T = 20$, $\rho = 0\ 9$, 0 05 significance level, the empirical rejection rates are less than 0 13 for both the standard and bootstrap procedures) Corresponding to annual series starting respectively in the early 1970s or 1960s and ending at the beginning of the present decade, such sample sizes are unfortunately almost typical in economic analysis Using quarterly data cannot be expected to improve the situation much, given that power has been shown to be largely independent of the sampling frequency (see Shiller and Perron, 1985, Pierse and Snell, 1992) Careful analysis using tests based on the null of stationarity as well (e g the LM test of Kwiatkowski *et al*, 1992) appears therefore to be essential

Table 4 *Errors* t(*3*)

	α	0 01	0 02	0 03	0 04	0 05	0 06	0 07	0 08	0 09	0 10
$\rho = 1\ 0$ — Nominal (α) and empirical sizes of bootstrap ($\hat{\alpha}_b$) and standard ($\hat{\alpha}_t$) tests											
$T = 20$	$\hat{\alpha}_b$	0 0096	0 0189	0 0285	0 0379	0 0486	0 0599	0 0707	0 0805	0 0901	0 1009
	$\hat{\alpha}_t$	0 0107				0 0473					0 0929
$T = 30$	$\hat{\alpha}_b$	0 0105	0 0201	0 0289	0 0386	0 0479	0 0579	0 0663	0 0754	0 0864	0 0975
	$\hat{\alpha}_t$	0 0094				0 0438					0 0896
$T = 50$	$\hat{\alpha}_b$	0 0122	0 0221	0 0331	0 0423	0 0519	0 0614	0 0708	0 0800	0 0894	0 0986
	$\hat{\alpha}_t$	0 0113				0 0493					0 0918
$\rho = 0\ 9$ — Empirical power of bootstrap ($\hat{p}_b$) and standard ($\hat{p}_t$) tests	α	0 01	0 02	0 03	0 04	0 05	0 06	0 07	0 08	0 09	0 10
$T = 20$	$\hat{p}_b$	0 0240	0 0475	0 0742	0 0982	0 1222	0 1476	0 1720	0 1949	0 2173	0 2401
	$\hat{p}_t$	0 0271				0 1213					0 2266
T = 30	$\hat{p}_b$	0 0408	0 0769	0 1123	0 1481	0 1840	0 2175	0 2451	0 2752	0 3078	0 3360
	$\hat{p}_t$	0 0359				0 1744					0 3139
T = 50	$\hat{p}_b$	0 0889	0 1609	0 2264	0 2871	0 3451	0 3977	0 4419	0 4886	0 5314	0 5710
	$\hat{p}_t$	0 0814				0 3237					0 5441
$\rho = 0\ 8$ — Empirical power of bootstrap ($\hat{p}_b$) and standard ($\hat{p}_t$) tests	α	0 01	0 02	0 03	0 04	0 05	0 06	0 07	0 08	0 09	0 10
$T = 20$	$\hat{p}_b$	0 0615	0 1194	0 1717	0 2216	0 2693	0 3182	0 3600	0 3987	0 4357	0 4740
	$\hat{p}_t$	0 0694				0 2663					0 4480
$T = 30$	$\hat{p}_b$	0 1313	0 2338	0 3171	0 3875	0 4520	0 5119	0 5631	0 6129	0 6557	0 6915
	$\hat{p}_t$	0 1244				0 4323					0 6620
$T = 50$	$\hat{p}_b$	0 3731	0 5606	0 6763	0 7606	0 8159	0 8599	0 8901	0 9147	0 9331	0 9474
	$\hat{p}_t$	0 3528				0 8010					0 9375

ACKNOWLEDGEMENTS

Research funded by CNR and MURST grants (research project 'Unit roots and cointegration new procedures of estimation and testing') We would like to thank Paolo Paruolo, Gianluca Cubadda and the participants at the EC(2) Conference on 'Non Parametric and Dynamic Modelling' (Humboldt University, Berlin, 1994) for helpful comments The usual disclaimers apply

REFERENCES

Basawa, I V, Mallik, A K, McCormik, W P, Reeves, J H and Taylor, R L (1991a) Bootstrapping unstable first-order autoregressive processes, *The Annals of Statistics*, **19**, 1098 101

Basawa, I V, Mallik, A K, McCormik, W P, Reeves, J H and Taylor, R L (1991b) Bootstrap tests of significance and sequential bootstrap estimation for unstable first order autoregressive processes, *Communications in Statistics Theory & Methods*, **20**, 1015 26

Blangiewicz M and Charemza, W W (1990) Cointegration in small samples empirical percentiles, drifting moments and customized testing, *Oxford Bulletin of Economics and Statistics*, **52**, 303 15

Bose, A (1988) Edgeworth Correction by Bootstrap in Autoregressions, *Annals of Statistics*, **16**, 1709–22

Christiano, L J (1992) Searching for a break in GNP, *Journal of Business and Economic Statistics*, **10**, 237–50

Cochrane, J H (1991) A critique of the application of unit root tests, *Journal of Economic Dynamics and Control* **15**, 275 84

Dickey, D A (1984) Power of unit root tests in *Proceedings of the American Statistical Association*, Business and Economic Statistics Section, pp 489-93

Efron, B (1979) Bootstrap methods another look at the jackknife *Annals of Statistics*, **7**, 1–26

Fachin, S (1995) The Finnish demand for money revisited an application of Box–Tiao cointegration analysis and bootstrap tests, *Applied Economics Letters*, **2**, 308-10

Harris, R I D (1992) Small sample testing for unit roots, *Oxford Bulletin of Economics and Statistics*, **54**, 615 25

Hylleberg, S and Mizon, G E (1989) A note on the distribution of the Least Squares estimator of a random walk with drift, *Economics Letters*, **29**, 225–30

Kwiatkowski, D, Phillips, P C B Schmidt, P and Shin, Yongcheol (1992) Testing the null hypothesis of stationarity against the alternative of a unit root, *Journal of Econometrics*, **54**, 159-78

MacKinnon, J G (1991) Critical values for cointegration tests in R F Engle and C W J Granger (eds) *Long-Run Economic Relationships*, Oxford University Press, Oxford, pp 267-77

Paparoditis, E (1994) Bootstrapping Autoregressive and Moving Average Parameter Estimates of Infinite Order Vector Autoregressive Processes, TR 11/94, Dept of Mathematics and Statistics, University of Cyprus

Pierse, R G and Snell, A J (1992) Temporal aggregation and the power of tests for a unit root, Department of Applied Economics, Cambridge, UK

Rudebusch, G D (1992) Trends and random walks in macroeconomic time series a re-examination, *International Economic Review*, **33**, 661-80

Schmidt, P (1990) Dickey-Fuller tests with drift, in T B Fomby and G F Rhodes, Jr *Cointegration, Spurious Regression and Unit Roots*, Advances in Econometrics, vol 8, JAI Press, Greenwich, CT

Schwert, G W (1989) Tests for unit roots a Monte Carlo investigation, *Journal of Business and Economic Statistics*, **7**, 147-60

Shiller, R J and Perron, P (1985) Testing the random walk hypothesis, *Economics Letters*, **18**, 381-86

Stock, J H (1994) Unit roots, structural breaks and trend, in R F Engle and D L McFadden (eds) *Handbook of Econometrics* Vol IV, Chapter 46, North-Holland, Amsterdam

Wu, C F J (1986) Jackknife, Bootstrap, and Other Resampling Methods in Regression Analysis, *The Annals of Statistics*, **14**, 1261-95