UPPSALA
UNIVERSITET

Working Paper 2013:6

Department of Statistics

Bootstrap versus Bartlett type correction of the Dickey-Fuller test

Johan Lyhagen and Jianxin Wei

Working Paper 2013:6
June 2013
Department of Statistics
Uppsala University
Box 513
SE-751 20 UPPSALA
SWEDEN

Title: Bootstrap versus Bartlett type correction of the Dickey-Fuller test

Author: Johan Lyhagen and Jianxin Wei
E-mail: johan.lyhagen@statistics.uu.se, jianxin.wei@statistics.uu.se

Bootstrap versus Bartlett type correction of the Dickey-Fuller test

Johan Lyhagen, Jianxin Wei

Abstract

The bad small sample performance of unit root tests is well known and depends partly on the existence of nuisance parameters. To deal with this Johansen (2004) presents a small sample correction of the Dickey-Fuller test. The correction factor depends on biased parameter estimates. The purpose of this study is to, through simulation, compare the small sample correction with a bootstrap approach as well as to analyze the effect of bias adjusting both the small sample correction and the bootstrap. The bootstrap works considerably better than the bias adjusted small sample correction when considering size but not for power.

Keywords: Bartlett type correction, unit root test, Monte Carlo simulation, bootstrap

1. Introduction

More than 30 years after the seminal work of Fuller (1976) and Dickey and Fuller (1979, 1981), the issue of whether a time series is generated by a unit root process or not is still an area of great interest. One reason for this is the policy implications. If e.g. unemployment has a unit root then it is possible for interventions to have a permanent effect while if there is no unit root then the effect is temporary. Unit roots are also necessary for cointegration, which can be interpreted as a long run economic equilibrium. Although a number of tests have been proposed, the ADF-test is still the one most commonly used. One reason for this is the appealing simplicity of the test, simple regressions are enough. The drawback is that the empirical size (i.e. the rejection frequencies calculated under the null) of the test is far from the nominal if the asymptotic critical values are used. This is clear from inspection of the tables for the small sample critical values in e.g. Fuller (1976) and

Preprint submitted to Elsevier *June 3, 2013*

the simulations performed in DeJong et al. (1992a,b). Adjusting the critical values, or the test statistic, such that the empirical size becomes closer to the nominal can principally be done in three ways. Firstly, the small sample critical values can be simulated as in e.g. Fuller (1976) and MacKinnon (1996), but these are mostly based on a very simplistic process. Hence, the influence of nuisance parameters is not considered. Secondly, the bootstrap has been found to be useful, see Harris (1992) and Nankervis and Savin (1996) for early papers, improving the properties of the empirical size. For a theoretical motivation why the bootstrap does work in the case of unit root tests, see Park (2003). Thirdly, Johansen (2004) derives a Bartlett type correction of the ADF-test. Although it is not possible to show theoretically that the correction works, see Jensen and Wood (1997), it is found to work in practice as shown by e.g. Nielsen (1997). A closer inspection of the results of Johansen (2004) indicates that there is room for improvements as the correction factor depends on the least squares estimates of the short run dynamics which are biased. The aim of this paper is to, by means of Monte Carlo simulations, compare the following: i) the test statistic using asymptotic critical values, ii) the Bartlett type corrected test statistic, iii) the Bartlett type corrected test statistic using bias adjusted parameter estimates, iv) the bootstrapped test statistic and v) the bootstrap test statistic using bias adjusted parameter estimates. The bias adjustment factor is based on the results in e.g. Tjöstheim and Paulsen (1983) and van Giersbergen (2005) where the bias of the parameters in an AR(p) process is derived.

In our simulation study we consider an AR(1) process as well as an AR(2) process. The parameters are varied extensively and various sample sizes are considered. The results are that, when considering size, both bootstrap based tests outperform the other tests and there are only minor differences between the two bootstrap tests. As noted in previous studies, using the asymptotic critical values yields a too high empirical size. The performance of the Bartlett type correction depends on the parameter space. When considering size adjusted power, there is no clear winner as the estimated power curves cross each other. In general it seems that the Bartlett type corrected test statistic with bias adjustment outperforms the others. When considering power, using the asymptotic critical values yields a high power, and the performance of the two bootstrap tests is no worse than for the Bartlett type correction.

The paper is organized as follows: In Section 2 the ADF-test, the Bartlett type correction, the bias adjustment and the bootstrap procedure are revisited. The design of the Monte Carlo simulation and the results are in Section 3 while a conclusion ends the paper.

2. The test and the correction

2.1. The model and the test statistic

Suppose a time series $X_{-k+1}, \ldots, X_{-1}, X_0, X_1, \ldots, X_T$ of length $T + k$ is observed and consider the univariate AR(k) model in the form,

$$\Delta x_t = \pi x_{t-1} + \beta t^d + \sum_{i=1}^{k-1} \gamma_i \Delta x_{t-i} + \sum_{i=0}^{d-1} \beta_i t^i + \varepsilon_t \tag{1}$$

where k is the number of lags, d determines the forms of the deterministic components and $\varepsilon_t \sim N(0, \sigma^2)$ are simultaneously independent. The null $H_0 : \pi = \beta = 0$ vs $H_1 : H_0 isnottrue$ makes sure that the order of the trend is the same under the null as under the alternative.

The likelihood ratio test statistic for the joint hypothesis is based on the likelihood ratio

$$\lambda = \left(\frac{RSS_U}{RSS_R} \right)^{T/2}$$

where RSS_U and RSS_R are the residual sum of squares resulting from the unrestricted model (1)and the restricted model

$$\Delta x_t = \sum_{i=1}^{k-1} \gamma_i \Delta x_{t-i} + \sum_{i=0}^{d-1} \beta_i t^i + \varepsilon_t.$$

The test statistic is $-2 \ln(\lambda) = -T(\frac{RSS_U}{RSS_R})$ and, under H_0, it has the Dickey Fuller type asymptotic distribution

$$\int_0^1 (dB) F' \left(\int_0^1 F F' du \right)^{-1} \int_0^1 F (dB)'$$

where B is a standard Brownian motion and F is

$$F(u) = \left(\begin{array}{c} B(u) \\ u^d \end{array} \Big| 1, \ldots, u^{d-1} \right)$$

i.e. the Brownian motion and the highest trend component are corrected for lower order trend components (Johansen, 1995). The asymptotic distribution does not depend on nuisance parameters γ_i and β_i, but only on the type of deterministic components used. The test statistic is the same as the Johansen's cointegration rank trace test for $p = 1$, the critical value can be found in, among others, Johansen (1995), table 15.1 to 15.5 for different types of deterministic terms.

2.2. The bootstrap

The bootstrap procedure we use is the residual bootstrap, the algorithm can be described as follows:

1. Fit the restricted model

$$\Delta x_t = \sum_{i=1}^{k-1} \gamma_i \Delta x_{t-i} + \sum_{i=0}^{d-1} \beta_i t^i + \varepsilon_t \tag{2}$$

   and obtain the estimates of short run parameters γ_i. Check that the parameters are inside the stationary region[1];

2. Fit the unrestricted model (1) and obtain the residuals $\{\hat{\varepsilon}_t\}$;

3. Apply iid bootstrap to $\{\hat{\varepsilon}_t\}$ and obtain the bootstrap residuals $\{\varepsilon_t^*\}$. If an intercept is not included in the regression, the residuals need to be centered in order to make $E\varepsilon_t^* = 0$, i.e. bootstrap from $\{\hat{\varepsilon}_t - \bar{\hat{\varepsilon}}_.\}$, where $\bar{\hat{\varepsilon}}_. = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t$;

4. Use the estimated parameter $\hat{\gamma}_i$ from step 1, the bootstrap residuals $\{\varepsilon_t^*\}$ from step 3 and the starting values X_{-k+1}, X_{-1}, X_0 to construct the bootstrap sample y_t^* under the null hypothesis;

5. Calculate the bootstrap statistic from the bootstrap sample;

6. Repeat step 3-5 B times to get an empirical distribution of the bootstrap statistic.

The distribution in step 6 can be considered as an approximation to the actual finite sample distribution, and the bootstrap test rejects the null if the proportion of the bootstrap statistics that is greater than the observed test

[1]The Yule-Walker method is not used here due to its finite sample bias problem, which may outweigh the benefits of ensuring stationarity.

statistic is less than the significant level.
Remarks: In step 3, the residuals could also be obtained from the restricted model (2). However, it is not a good choice when the null does not hold, and in practice this is likely to have negative consequences on power. Some discussion can be found in Paparoditis and Politis (2003) and Parker et al. (2006). In step 4, the bootstrap sample must be constructed under the null, i.e. $\pi = \beta = 0$, this is crucial in the bootstrap procedure.

The asymptotic distribution of the test statistic is asymptotic pivotal, so the bootstrap may provide an asymptotic refinement, i.e. the bootstrap distribution is corrected up to the second order terms in the expansion and consequently gives better size in the test. For a theoretical discussion, see Park (2003).

2.3. Bartlett type correction

Bartlett correction is also a technique to obtain an accurate finite sample distribution. In a seminal paper Bartlett (1937), Bartlett proposed an improved LR statistic. The basic idea was the following, suppose the expectation of the LR statistic is $E(LR) = q(1 + \frac{b}{T} + O(T^{-2}))$, where q is the number of degrees of freedom of the limiting chi-square distribution and b is a constant which can be estimated. Then the expected value of $LR/(1 + \frac{b}{T})$ is closer to q than $E(LR)$, and $1 + \frac{b}{T}$ is the Bartlett correction factor. Distributionally,

$$\Pr(S_B \leq x) = \Pr(\chi_p^2 \leq x) + O(T^{-2}). \tag{3}$$

In the unit root context, the limiting distribution of the LR test statistic is not chi-square, but non-standard, and moreover, Jensen and Wood (1997) show that the order property (3) is violated. However, Nielsen (1997) found that the correction works in practice. Johansen (2004) suggests a Bartlett type correction factor and as a consequence the corrected distribution will shift to the left and this should improve the size properties. Defining $a_T(d) = 1 + a_1(d)T^{-1} + a_2(d)T^{-2}$, the correction factor is expressed as

$$a_T(d)\left(1 + \frac{1}{T}\left[\left(k - 1 + 2\frac{\sum_{i=1}^{k-1} i\gamma_i}{1 - \sum_{i=1}^{k-1}\gamma_i}\right)m(d) + \frac{1}{2}\left((-1)^{k-1} - 1\right)g(d)\right]\right),$$

where $a_1(d)$, $a_2(d)$, $a_T(d)$, $m(d)$ and $g(d)$ are tabulated for $d = 0, 1, 2$, and for the model without deterministic terms, all can be found in Johansen (2004).

2.4. Biased adjusted estimator

It is well known that the OLS estimates of the short run parameters $\hat{\gamma}_i$ are biased and this might affect the correction factor in a negative way. Tjöstheim and Paulsen (1983) and van Giersbergen (2005) derive the bias of the OLS estimator of the parameters of an AR(p) process. For example, for the case of $k = 2$ and $d = 1$ the bias of $\hat{\gamma}$ is

$$E(\hat{\gamma}) - \gamma = -\frac{1+3\gamma}{T} + o\left(T^{-1}\right),$$

for $k = 3$,

$$E\left(\begin{bmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{bmatrix}\right) - \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} = -\begin{bmatrix} \frac{1+\gamma_1+\gamma_2}{T-1} \\ \frac{2-\gamma_1+2\gamma_2}{T-1} \end{bmatrix} + o\left(T^{-1}\right).$$

These results can be used to bias adjust the estimators. To be specific, for the case of $k = 2$ and $d = 1$,

$$\hat{\gamma}_1^u = \frac{\hat{\gamma}_1 + \frac{1}{T}}{\left(1 - \frac{3}{T}\right)},$$

for $k = 3$

$$\begin{bmatrix} \hat{\gamma}_1^u \\ \hat{\gamma}_2^u \end{bmatrix} = \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} \frac{1}{T-1} & \frac{1}{T-1} \\ -\frac{1}{T-1} & \frac{2}{T-1} \end{bmatrix}\right)^{-1} \left(\begin{bmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{bmatrix} + \begin{bmatrix} \frac{1}{T-1} \\ \frac{2}{T-1} \end{bmatrix}\right).$$

The adjusted estimators $\hat{\gamma}_1^u$, $\begin{bmatrix} \hat{\gamma}_1^u \\ \hat{\gamma}_2^u \end{bmatrix}$ can be considered unbiased if ignoring the $o(T^{-1})$ terms. In some cases, the bias adjusted estimates will fall out of the stationary region. When this happens, we can use either the unadjusted estimates or a predetermined value, for example, in AR(1), ± 0.99.

Once we obtain the bias adjusted estimators, it is straight forward to obtain the bias adjusted Bartlett correction factor and the bias adjusted bootstrap statistic by simply replacing the short run estimates by the bias adjusted version.

3. The Monte Carlo simulation and results

3.1. General Monte Carlo design

We consider the following DGP for the simulation study.

$$\Delta x_t = \pi x_{t-1} + \beta t + \gamma_1 \Delta x_{t-1} + \gamma_2 \Delta x_{t-2} + \beta_0 + \varepsilon_t. \tag{4}$$

For all setups we set $\varepsilon_t \sim N(0,1)$, $\beta_0 = 0$. The number of replicates is 10000 while the number of bootstrap replicates is $B = 199$ and the significance level considered is 5%. The test statistics evaluated are, with abbreviations in parentheses: i) the test statistic using asymptotic critical values (Asy), ii) the Bartlett corrected (BC), iii) the Bartlett corrected using bias adjusted parameters (BC(Ad)), iv) the bias adjusted bootstrap (Boot(Ad)) and v) the bootstrap (Boot). For the bootstrap tests, any replications violating the stationarity check conditions were discarded and the experiment continued until 10000 valid replications were obtained. The violation frequency is reported.

3.2. Size properties

Under the null hypothesis $\pi = \beta = 0$, the data is generated from the model:

$$\Delta x_t = \gamma_1 \Delta x_{t-1} + \gamma_2 \Delta x_{t-2} + \varepsilon_t.$$

We investigate both the cases AR(1) process and AR(2) process with different parameter values. For AR(1), the parameter values are

$$\gamma_1 = \{-0.9, -0.7, \ldots, 0.7, 0.9\}.$$

For AR(2), we consider two sets of parameters. The first set of parameters is the same as used by Johansen (2004):

γ_1	-0.9	-0.3	0.3	0.9	0	0	0	0
γ_2	0	0	0	0	-0.9	-0.3	0.3	0.9

The second set of parameters is

γ_1	-1.8	-1.2	-0.6	0.6	1.2	1.8	1.2	0.6	-0.6
γ_2	-0.9	-0.3	0.3	0.3	-0.3	-0.9	-0.9	-0.9	-0.9

These parameters are close to the boundary of the stationarity triangle of AR(2) as illustrated in Figure 1 below. The sample sizes for AR(1) are $T = \{10, 20, 50\}$ while for the AR(2) $T = \{20, 50\}$. To make sure that the initial conditions do not matter $100 + T + k$ observations are generated and the first 100 are discarded.

In Figure A.2 the empirical size is shown as a function of γ_1 for the AR(1) and sample size $T = 10$. Clearly, when using the asymptotic critical values

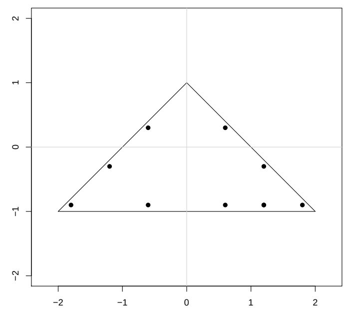

Figure 1: Parameter values for AR(2). All are close to the boundary of the stationary triangle of AR(2).

the empirical size increases significantly to about 0.2 as γ_1 increases. Using the Bartlett type corrected test statistic improves the performance substantially. Using the bias adjusted parameter values improves the size properties, but for γ_1 larger than 0.25 the test becomes undersized. This agrees with the result in Johansen (2004), since the correction factor becomes larger as γ increases and that will over correct the rejection frequency. In contrast, the two bootstrap test statistics are less sensitive to the parameter value. Overall, the two bootstrap test statistics perform best with an empirical size close to the nominal. Bias adjusting renders only a minor improvement.

In Figures A.3 and A.4 the empirical size is plotted as a function of sample size for $\gamma_1 = 0.9$ and $\gamma_1 = -0.9$. As expected, a fairly large sample is needed for the empirical size to be close to the nominal when using the asymptotic critical values. For $\gamma_1 = 0.9$, the bias adjusted Bartlett type corrected test statistic is quite undersized and not bias adjusting shifts the curve towards the nominal size but not far enough. However, for $\gamma_1 = -0.9$, the Bartlett correction performs much better. The bootstrap tests are robust to the parameter values, as in Figure A.2 the bootstrap tests perform better.

Tables A.2 and A.3 show the empirical size for AR(2) for the two sets of parameters outlined above. As in the figures the bootstrap outperforms the Bartlett type small correction which in turn is better than using the asymptotic distribution. The bias adjustment shifts the distribution to the left and lowers the empirical size. For the point $\gamma_1 = 1.8, \gamma_2 = -0.9$, which is closest to $I(2)$, even the Bartlett corrections have a very high size, but the two bootstrap tests work well. The Bartlett type small sample correction overcompensates in a few cases, yielding a too low empirical size, for example (0,0.9). This is because the Bartlett correction factor is very large, and shifts the distribution too much to the left. As in AR(1), the bootstrap tests are more robust to the true parameter values. The results for the Bartlett type small correction and the asymptotic in Table A.2 and $T = 10$ agree well with Table 2 of Johansen (2004).

3.3. Power properties

Statistical properties of tests include power as well as size. We analyze the size adjusted power and the raw power for the test statistics outlined above. The data generating process is model (4) for two situations. The first is an AR(1) with $\gamma_1 = 0.9$ and the second is an AR(2) with $\gamma_1 = 0.6$ and $\gamma_2 = 0.3$. The sample size is $T = 20$ for both cases. We focus on the effect of changing π and fix $\beta = 0$, the π values are

$$\pi = \{-1, -0.98, -0.96, \ldots, 0\}.$$

For the test using asymptotic critical values, Bartlett correction and the Bartlett correction with bias adjusted parameters, the size adjusted critical value is obtained by 10000 simulations. For the two bootstrap tests, the critical value is already size adjusted. To save the computational cost, we use the method in Davidson and MacKinnon (2006) to approximate the rejection probability. To be specific, for each replication:

1 Generate data under the alternative;

2 Calculate the statistic;

3 Generate ONE bootstrap sample under the null;

4 Calculate the bootstrap statistic $-T(\frac{RSS_U^*}{RSS_R^*})$.

After B replications, we have B statistics under the alternative and B bootstrap statistics. We consider the B bootstrap statistics as the approximation to the bootstrap distribution and carry out the bootstrap test for each of the statistics under the alternative. The difference is that the B bootstrap tests use the same bootstrap distribution while in a usual bootstrap test the B bootstrap distributions are generated individually.

In Figures A.5 and A.7 the size adjusted power is displayed as a function of π. The patterns from the two figures are the same, but the magnitudes differ. For values of π close to the null there is not much difference but it seems that the bias adjusted Bartlett type corrected test statistic has the lowest power. Increasing π decreases the power, especially for the bias adjusted Bartlett type corrected test statistic which has the highest power when $\pi < -0.5$. There does not seem to be any significant difference between the bootstrapped and the bias adjusted bootstrap test statistics, both having lower power than the Bartlett type corrected test statistics. Using asymptotic critical values yields the lowest size adjusted power. For the AR(2) the differences in power are much larger than for the AR(1). This might be due to that for the AR(2), two of the characteristic roots are close to the unit circle, see Figure A.9.

We also simulated the raw power to give a complete comparison. As shown in Figures A.6 and A.8, for both AR(1) and AR(2), the asymptotic test has the highest estimated power. This is not surprising if we take the high size into account. In AR(1) the bootstrap tests outperform the Bartlett corrections. The biased adjusted Bartlett correction has the lowest power, the two bootstrap tests are equally good. In AR(2), the power curves of the four methods cross each other, but the power of the bootstrap tests is only slightly smaller than the power of Bartlett correction.

4. Conclusions

In this paper an extensive Monte Carlo simulation has been performed to analyze the small sample properties of the ADF-test for unit roots. The versions compared are the Bartlett type correction of Johansen (2004), the bootstrap test statistic and using bias adjusted versions of the two. Further, as a base line, the properties of the test using the asymptotic critical values are also investigated. The bootstrap test statistic has the best size properties but not uniformly the best power. Overall the bootstrap is to be preferred.

Acknowledgements

Financial support from the Jan Wallander and Tom Hedelius Foundation, research grant number P2005-0117:1, is gratefully acknowledged.

References

Bartlett, M. S., 1937. Properties of sufficiency and statistical tests. Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences 160, 268–282.

Davidson, R., MacKinnon, J. G., 2006. The power of bootstrap and asymptotic tests. Journal of Econometrics 133, 421 – 441.

DeJong, D. N., Nankervis, J. C., Savin, N. E., Whiteman, C. H., 1992a. Integration versus trend stationary in time series. Econometrica 60, 423–433.

DeJong, D. N., Nankervis, J. C., Savin, N. E., Whiteman, C. H., 1992b. The power problems of unit root tests in time series with autoregressive errors. Journal of Econometrics 53, 323–343.

Dickey, D. A., Fuller, W. A., 1979. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association 74, 427–431.

Dickey, D. A., Fuller, W. A., 1981. Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica 49, 1057–1072.

Fuller, W. A., 1976. Introduction to statistical time series. Wiley, New York.

Harris, R. I. D., 1992. Small sample testing for unit roots. Oxford Bulletin of Economics and Statistics 54, 615–625.

Jensen, J. L., Wood, A. T., 1997. On the non-existence of a Bartlett correction for unit root tests. Statistics and Probability Letters 35, 181–187.

Johansen, S., 1995. Likelihood-based Inference in Cointegrated Vector Autoregressive Models. Oxford University Press.

Johansen, S., 2004. A small sample correction of the Dickey-Fuller test. In: New Directions in Macromodelling. Elsevier, pp. 49–68.

MacKinnon, J. G., 1996. Numerical distribution functions for unit root and cointegration tests. Journal of Applied Econometrics 11, 601–618.

Nankervis, J. C., Savin, N. E., 1996. The level and power of the bootstrap t-test in the AR(1) modell with trend. Journal of Business and Economic Statistics 14, 161–168.

Nielsen, B., 1997. Asymptotic results for cointegration tests in non-stable cases. Biometrika 84, 500–504.

Paparoditis, E., Politis, D. N., 2003. Residual-based block bootstrap for unit root testing. Econometrica 71, 813–855.

Park, J. Y., 2003. Bootstrap unit root tests. Econometrica 71, 1845–1895.

Parker, C., Paparoditis, E., Politis, D. N., 2006. Unit root testing via the stationary bootstrap. Journal of Econometrics 133, 601 – 638.

Tjöstheim, D., Paulsen, J., 1983. Bias of some commonly-used time series estimates. Biometrika 70, 389–399.

van Giersbergen, N. P., 2005. On the effect of deterministic terms on the bias in stable ar models. Economics Letters 89, 75–82.

Appendix A. Tables and Figures

Table A.1: Empirical size in percent when nominal size is 5%. The data generating process is an AR(1). 'Er' denotes the fraction of times the estimated parameters fall outside the stationary region.

			$T = 10$			
γ	Asy	BC	BC(Ad)	Boot	Boot(Ad)	Er
0.9	18.1	3.5	1.3	6.0	5.1	4.1
0.3	12.7	5.0	3.5	5.5	4.7	0.1
-0.3	10.4	5.5	5.3	5.3	5.2	0.2
-0.9	8.6	5.2	5.3	5.3	5.5	15.6
			$T = 20$			
γ	Asy	BC	BC(Ad)	Boot	Boot(Ad)	Er
0.9	12.9	2.9	1.2	5.9	5.0	1.7
0.3	7.8	4.5	3.9	5.0	4.7	0
-0.3	6.9	5.0	4.9	4.9	4.9	0
-0.9	6.3	4.9	4.9	5.1	5.3	5.2

Table A.2: Empirical size of in percent when nominal size is 5%. The data generating process is an AR(2).

$T = 20$							
γ_1	γ_2	Asy	BC	BC(Ad)	Boot	Boot(Ad)	Er
-0.9	0.0	7.0	5.6	5.5	4.9	4.8	4.6
-0.3	0.0	7.2	5.2	4.8	5.0	4.7	0
0.3	0.0	8.5	5.2	4.4	5.1	4.7	0
0.9	0.0	14.6	4.0	2.1	5.7	5.0	2.3
0.0	-0.9	6.4	5.6	5.6	4.9	4.9	9.2
0.0	-0.3	7.1	5.4	5.1	4.9	4.8	0.0
0.0	0.3	8.9	4.8	3.6	5.2	4.6	0.0
0.0	0.9	13.4	3.2	2.1	6.1	5.7	14.7

$T = 50$							
γ_1	γ_2	Asy	BC	BC(Ad)	Boot	Boot(Ad)	Er
-0.9	0.0	5.7	5.2	5.2	5.2	5.1	0.3
-0.3	0.0	5.7	4.8	4.7	5.0	5.0	0
0.3	0.0	6.1	4.9	4.7	4.9	4.9	0
0.9	0.0	10.3	4.0	2.5	5.8	5.1	0.2
0.0	-0.9	5.6	5.5	5.5	5.3	5.3	0.5
0.0	-0.3	5.6	5.0	4.9	5.1	5.0	0
0.0	0.3	6.4	4.7	4.5	4.9	4.7	0
0.0	0.9	10.8	3.1	1.7	5.8	5.0	2.7

Table A.3: Empirical size in percent of unit root tests when nominal size is 5%. The parameter values of the AR(2) are chosen such that they are close to the non-stationary boundary.

				$T = 20$			
γ_1	γ_2	Asy	BC	BC(Ad)	Boot	Boot(Ad)	Er
-1.8	-0.9	5.5	4.9	4.9	5.0	5.0	9.4
-1.2	-0.3	7.0	5.8	5.8	5.2	5.1	2.9
-0.6	0.3	7.8	5.4	5.0	4.1	4.9	6.2
0.6	0.3	14.6	3.7	1.8	6.0	5.2	2.7
1.2	-0.3	16.2	5.1	3.2	5.7	4.8	2.0
1.8	-0.9	27.3	18.6	19.7	5.2	4.7	10.2
1.2	-0.9	9.0	7.8	7.9	4.6	4.5	9.3
0.6	-0.9	6.6	5.8	5.9	4.6	4.6	9.1
-0.6	-0.9	5.9	5.4	5.4	5.2	5.1	9.2
-1.2	-0.9	5.9	5.4	5.3	5.5	5.4	9.0
				$T = 50$			
γ_1	γ_2	Asy	BC	BC(Ad)	Boot	Boot(Ad)	Er
-1.8	-0.9	5.2	5.1	5.1	5.0	5.0	0.6
-1.2	-0.3	6.0	5.6	5.6	5.3	5.2	0.1
-0.6	0.3	6.0	5.1	5.0	5.2	5.1	0.7
0.6	0.3	10.2	3.1	2.0	5.4	4.8	0.3
1.2	-0.3	8.9	4.1	2.9	5.0	4.5	0.1
1.8	-0.9	9.7	8.0	8.2	5.6	5.4	0.7
1.2	-0.9	5.5	5.3	5.3	5.0	5.0	0.6
0.6	-0.9	5.4	5.2	5.2	4.9	4.9	0.6
-0.6	-0.9	5.8	5.6	5.6	5.5	5.5	0.4
-1.2	-0.9	4.9	4.8	4.8	4.8	4.8	0.7

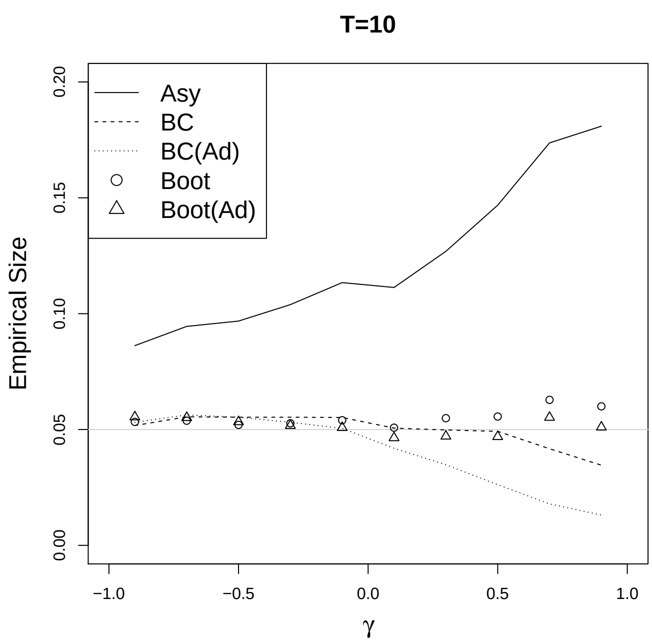

Figure A.2: Empirical size of Dickey-Fuller unit root test as a function of the AR(1) short run parameter γ. Nominal size is 5% and sample size is $T = 10$.

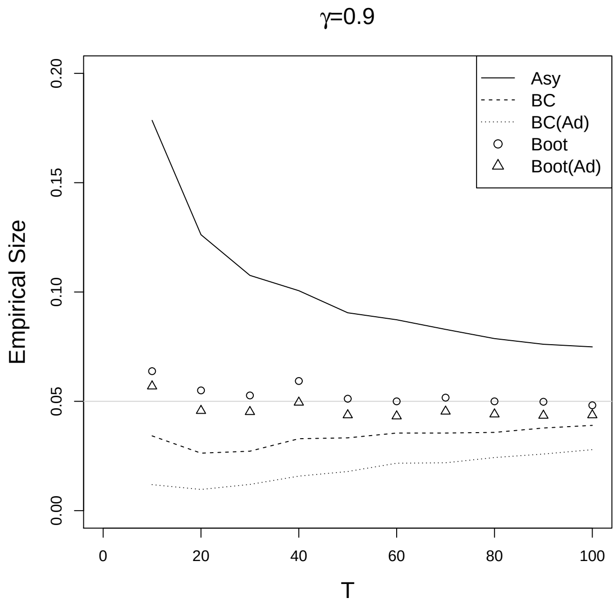

Figure A.3: Empirical size of Dickey-Fuller unit root test as a function of sample size. Nominal size is 5% and short run AR(1) parameter is $\gamma = 0.9$.

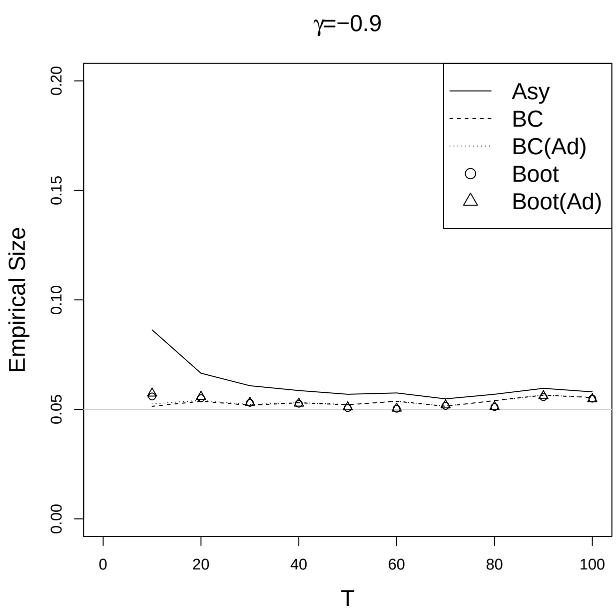

Figure A.4: Empirical size of Dickey-Fuller unit root test as a function of sample size. Nominal size is 5% and short run AR(1) parameter is $\gamma = -0.9$.

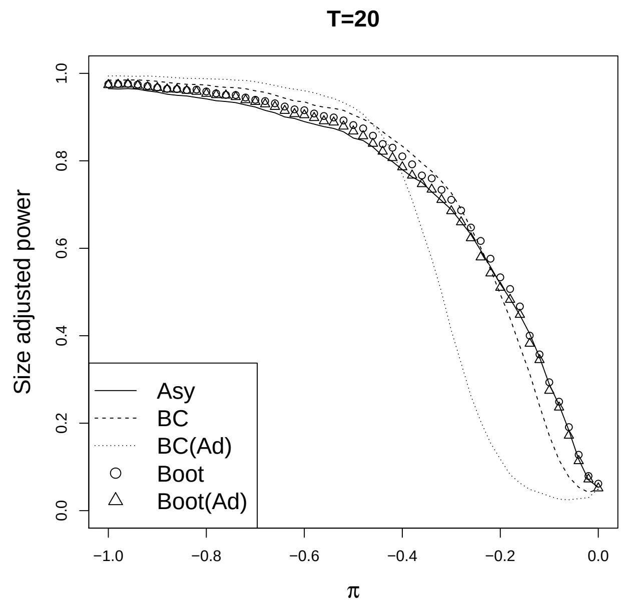

Figure A.5: Size adjusted power of Dickey-Fuller unit root test. Nominal size is 5% and the AR(1) short run parameter is $\gamma = 0.9$.

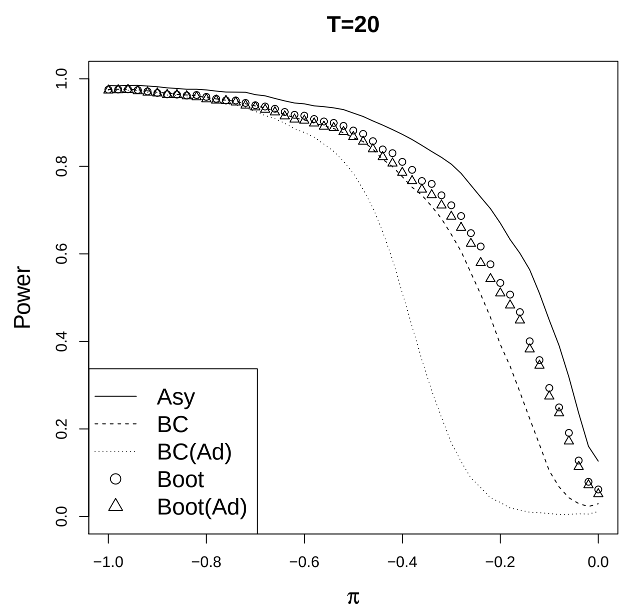

Figure A.6: Estimated rao power of Dickey-Fuller unit root test. Nominal size is 5% and the AR(1) short run parameter is $\gamma = 0.9$.

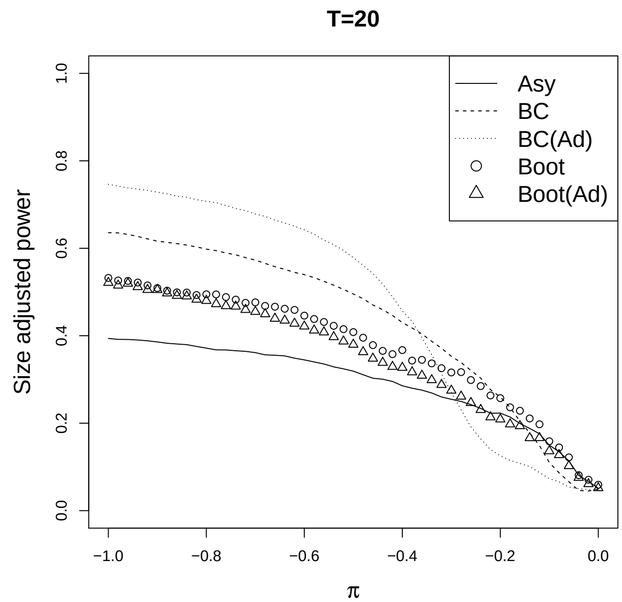

Figure A.7: Size adjusted power of Dickey-Fuller unit root test. Nominal size is 5% and the AR(2) short run parameters are $\gamma_1 = 0.6$ and $\gamma_2 = 0.3$.

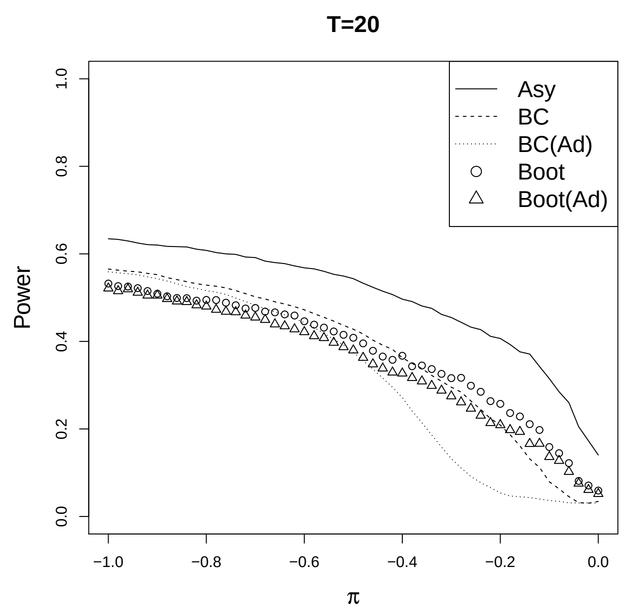

Figure A.8: Estimated raw power of Dickey-Fuller unit root test. Nominal size is 5% and the AR(2) short run parameters are $\gamma_1 = 0.6$ and $\gamma_2 = 0.3$.

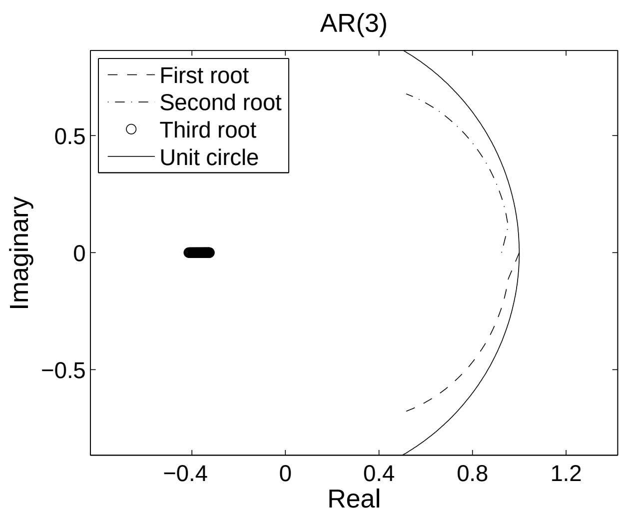

Figure A.9: Characteristic roots for the ADF-model with two lags.