Using the Bootstrap technique and the sample autocorrelation function for the identification of unit root processes

Panagiotis Mantalos
Department of Statistics
Lund University
Sweden

ABSTRACT

By using a Bootstrap procedure and the function of sample autocorrelations of a time series, we are able to distinguish between a non-stationary integrated series of order one, I(1), from a stationary AR(1) process. We investigated our Bootstrap procedure by using a Monte Carlo method and by an empirical example. The results show that by using the Bootstrap average of the sample autocorrelations function, we can easily identify a non-stationary I(1).

Keywords: Bootstrap, Co-integration, Sample Autocorrelations, Unit Root.

Communications should be sent to:
Panagiotis Mantalos, Department of Statistics, P.O.Box 743, S-22007 Lund, Sweden.
E-mail : Panagiotis.Mantalos@stat.lu.se

Using the Bootstrap technique and the sample autocorrelation function for the identification of unit root processes

By

Panagiotis Mantalos

Department of Statistics

Lund University

1. Introduction

Consider an observed time series, denoted by $\{y_1, y_2, \ldots, y_T\}$, which give rise to a set of sample autocorrelations[1], r_k, which followed a slow linear decline from a value close to unity at zero lag.

This indicates that we may be dealing with a non-stationary process, and by the Box-Jenkins principle, one might suggest differencing the series; that is, we might apply the simplifying operator, (1-B), to the observed time series, and transform the original series into a stationary one. Note that B denotes the usual backshift operator, $\mathrm{B}y_t \equiv y_{t-1}$.

In a similar way, we can apply the simplifying operator, (1-j B), to the observed time series, if the sample autocorrelations follow a slow, geometric decay from a value close to unity at zero lag, in this case, the j will first need to be estimated.

[1] The kth sample autocorrelations is: $r_k = g_k / g_0 \quad k = 0,1,\ldots,T-1$,

where $g_k = \dfrac{1}{T}\sum_{t=1}^{T-k}\left((y_t - \bar{y})(y_{t+k} - \bar{y})\right)$, In both situations, the original correlation structure indicates that there is a dominant autoregressive, AR, factor in the process, which shows up as *slow decline* in the non-stationary case, (1-B), or as *slow decay* in the nearly-non-stationary case, (1- j B).

However, whether the case is non-stationary or nearly-non-stationary, the shape of the sample autocorrelation is similar, and it is difficult to distinguish between them by just investigating the Sample Autocorrelation Function (SACF).[2] As many economic time series appear to be non-stationary, an efficient inference in time series analysis requires this to be taken into account.

Unit root techniques has become a rapidly growing focus in statistical literature (see Dickey and Fuller (1979), Said and Dickey (1984), Phillips and Perron (1988), Kwiatkowski *et al.* (1992)), as have Bootstrap techniques. Horowitz (1994), Mammen (1993), Mantalos and Shukur (1998), Mantalos (1999), and Mantalos and Shukur (1999) are just some of the studies in which the use of Bootstrap techniques has improved the critical values of different tests, so that the true size of the test approaches its nominal value.

By using the Künsch (1989) modified blocks Bootstrap technique in estimating the distribution of sample autocorrelations of a time series, Aczel and Josephy (1992) were able to identify the non-stationary integrated series of order one, I(1), from the nearly-non-stationary AR(1) case. However, the determination of the size of the modified blocks Bootstrap procedure is still not automatic; for example, one has to carefully examine the data in order to determine the size of the blocks, and the most important point is that the blocks Bootstrap technique is designed for a stationary series.

The purpose of this paper is to introduce a Bootstrap procedure that will distinguish between a non-stationary I(1) process from a stationary AR(1) process, using a function of the

[2] SACF is a plot of the r_k against the k lags.

sample autocorrelations of a time series. We will use the Monte Carlo method to study the Bootstrap procedure.

The paper is arranged as follows. In the next section, we study the behaviour of the SACF for an AR(1) process. In Section 3, we present the Bootstrap identification procedure, while in Section 4 we describe the results concerning the size and power of our technique. Section 5 contains an empirical example, and, finally, a brief summary and conclusions are presented in Section 6.

2. The behaviour of the SACF for an AR(1) process

In order to study the behaviour of the SACF for an AR(1) process, we have to consider the data generated process, DGP, that is given by the AR(1) model :

$$y_t = j y_{t-1} + u_t , \tag{1}$$

where $u_t \sim IN(0, s^2)$, and $|j| \leq 1$.When $j = 1$, we have a non-stationary I(1), and when $-1 < j < 1$, we have a stationary process. We can define the k^th sample autocorrelations as: $r_k = g_k / g_0 \quad k = 0,1,\ldots,T-1$.

Although it is not easy to obtain the r_k distribution, Wichern (1973) has provided some insight into the behaviour of the r_k by examining the ratio of $E(g_k)/E(g_0)$. Wichern (1973) also provided the function of the ratio $E(g_k)/E(g_0)$ for DGP (1) with $j = 1$ as

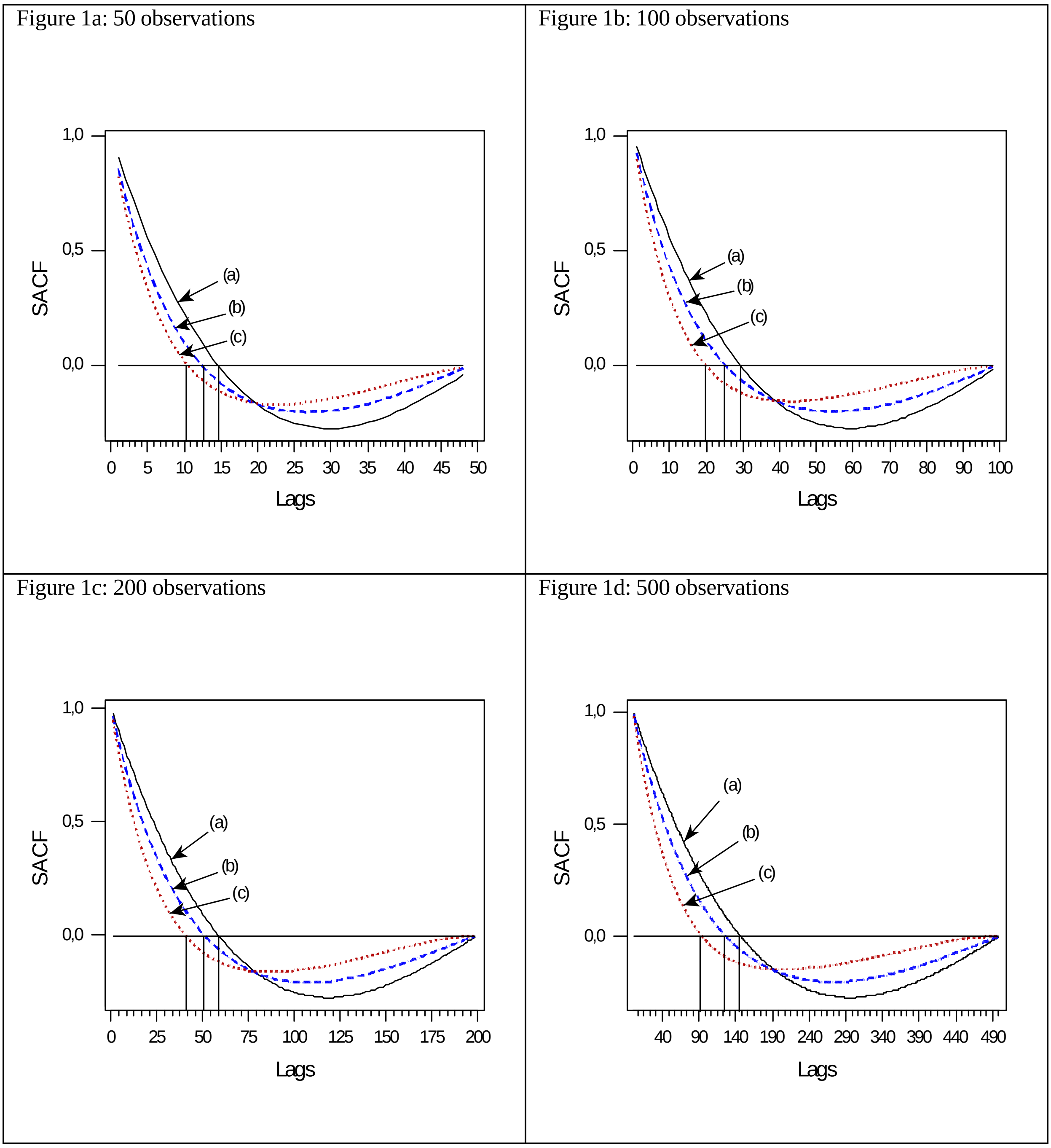
$$\frac{E(g_k)}{E(g_0)} = \frac{(T-k)(T^2 + 2k^2 - 4kT - 1)}{T(T^2 - 1)} \tag{2}$$

Function (2) is zero at approximately $k = T/\sqrt{2} + 2$. In Figure 1, curve (a) shows Function 2 for different series samples (50, 100, 200, and 500). However, the ratio $E(g_k)/E(g_0)$ is not equal to the average value of the sample autocorrelations, $E(r_k)$. By using a Monte-Carlo experiment to simulate 5000 series, and by taking the average value of the sample autocorrelations for each of the four different samples (50, 100, 200, and 500), we can show that the ratio $E(g_k)/E(g_0)$ is larger than $E(r_k)$. In Figure 1, curve (b) is from a random walk model showing the average of the sample autocorrelations for the simulated series. It clearly shows the differences between the ratio $E(g_k)/E(g_0)$ and the average value of the sample autocorrelations. The average of this function seems to be zero at approximately $k = T/3.96$, which is less than the $k = T/\sqrt{2} + 2$. Finally, we simulate 5000 near-integrated series for the four different samples (50, 100, 200, and 500) with $j = 1 - \frac{T^{1/3}}{T}$.

Curve (c) in Figure 1 shows the function of the average of the sample autocorrelations for the simulated near-integrated series. We can see that even when $j = 0.985$, as in the case of the 500 observations, the near-integrated series has much less average sample autocorrelations than the integrated process, and it is quite easy to distinguish this from the I(1) series.

Using the behaviour of the average sample autocorrelations and the Bootstrap technique, we show in the next section the possibility of identifying an integrated time series by using the sample autocorrelations.

Figure 1. The Average Sample Autocorrelations



I. Curve (a) shows the ratio $E(g_k)/E(g_0)$ function.

II. Curve (b) shows the average of the sample autocorrelations function for the integrated process.

III. Curve (c) shows the average of the sample autocorrelations function for the near-integrated process.

3. The Bootstrap procedure

When studying the SACF for ARIMA models, Wichern (1973) and Anderson (1981) pointed out the importance of studying the serial correlations for all possible lags. However, whether the case is non-stationary or nearly-non-stationary, the shape of the sample autocorrelation is similar and it is difficult to distinguish between them by just investigating the SACF, even for series with large sample sizes. The (a) curves in Figures 2, 3, 4, and 5 demonstrate this difficulty. Note that all the series are simulated random walks, and all clearly show that the function is not always zero at the T/3.96 lag, as the average of the sample autocorrelations function for the I(1) process was shown to be in our Monte Carlo experiment.

To overcome this problem we now introduce a method of using the Bootstrap technique to make it possible to identify the SACF of the series from the I(1) of the stationary series. We describe our Bootstrap procedure by the following steps:

(1) first, we use the Said and Dickey (1984) approach, which involves the approximation of the true process by an auto-regression in which the number of lags increases with sample size, and the OLS to estimate: that is, we use:

$$\Delta y_t = r y_{t-1} + \sum_{j=1}^{k} a_j \Delta y_{t-j} + d_t \, , \tag{3}$$

where k is large enough to allow the d_t to approximate to white noise. Note that in our Monte Carlo experiment, even if we know the real DGP, we still use k = $T^{1/3}$, because when we investigate the observed series, the real order of the AR and MA polynomials that the series might follow, are unknown.

(2) We use the adjusted OLS residuals $(\hat{d}_t - \bar{d})$ i = 1,...,T, to draw i.i.d $d_1^*, \ldots, d_T^*$ data and define three different series:

I. $$\Delta y_t^* = \sum_{j=1}^{k} \hat{a}_j \Delta y_{t-j}^* + d_t^* \, , \tag{4}$$

that is, we restrict the y_t^* series to follow a unit root process, and then use the y_t^* series to estimate the SACF:

II. $\Delta z_t^* = \hat{r} z_{t-1}^* + \sum_{j=1}^{k} \hat{a}_j \Delta z_{t-j}^* + d_t^*$, (5)

that is, the z_t^* series is an unrestricted ARIMA process, and estimates the SACF for the z_t^* series.

Finally, the series:

III. $\Delta w_t^* = f w_{t-1}^* + \sum_{j=1}^{k} \hat{a}_j \Delta w_{t-j}^* + d_t^*$ (6)

where $f = 1 - \frac{T^{1/3}}{T}$; that is, we restrict the series to follow a near-integrated stationary process, and then we estimate the SACF for w_t^*.

(3) By repeating this step N_b times, and then by taking the average of the Bootstrap Sample Autocorrelation function (ASACF), we form the ASACF0 for the series y_t^*, ASACF1 for series z_t^*, and ASACF2 for series w_t^*. The size of the Bootstrap sample used to estimate the average of the ASACF was N_b =200.

(4) Finally, by plotting in the same figure the SACF, ASACF0, ASACF1, and ASACF2, we are able to distinguish the unit root series from the stationary.

The principle is easy to understand. The series in which the ASACF1 is larger than the ASACF2 (the ASACF1 is zero at larger lag values than the ASACF2) are integrated, and series in which the ASACF1 is smaller than the ASACF2, are stationary.

Figure 2. The Bootstrap ASACF for 50 observations

Figure 2a: (Near stationary border)

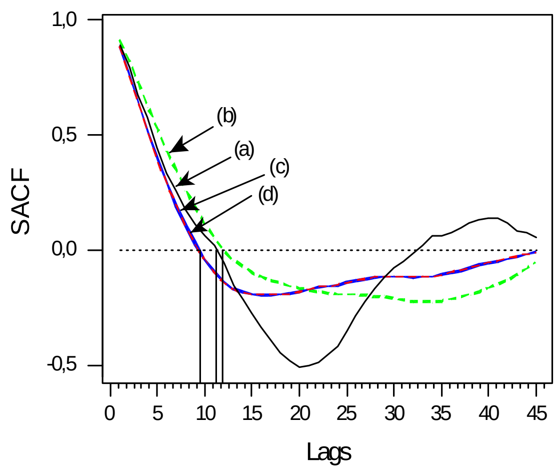


Figure 2b: (Near unit root border)

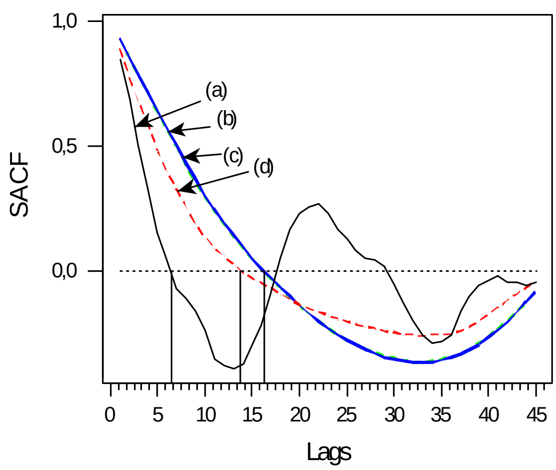


Figure 2c: (Between the upper and lower border)

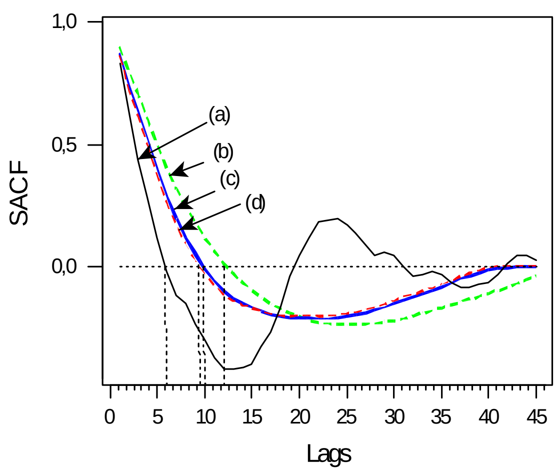
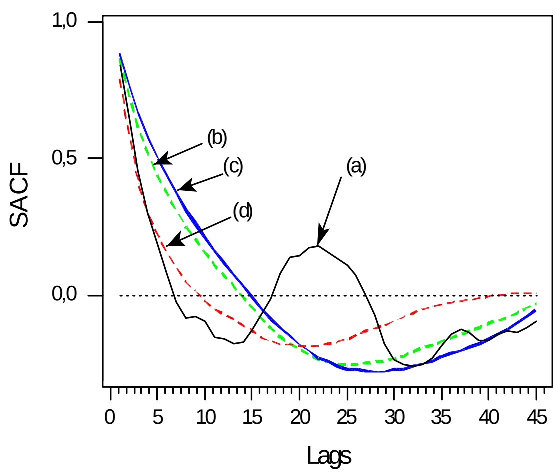


Figure 2d: (Above the upper border)



Curve (a) shows the original SACF function of the series, curve (b) shows the ASACF0, curve (c) shows the ASACF1, and curve (d) shows the ASACF2.

Figure 3. The Bootstrap ASACF for 100 observations

Figure 3a: (Near stationary border)

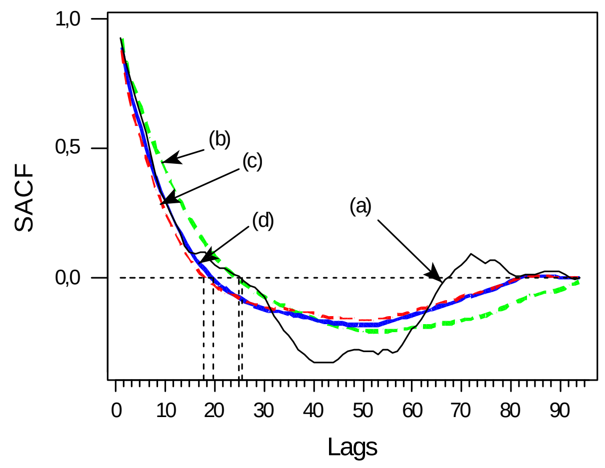


Figure 3b: (Near unit root border)

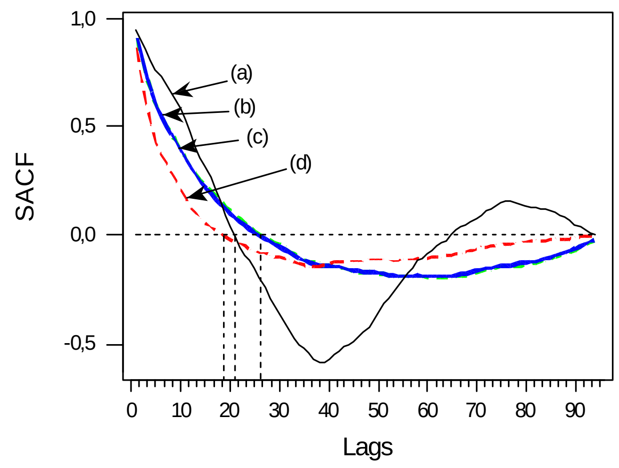


Figure 3c: (Between the upper and lower border)

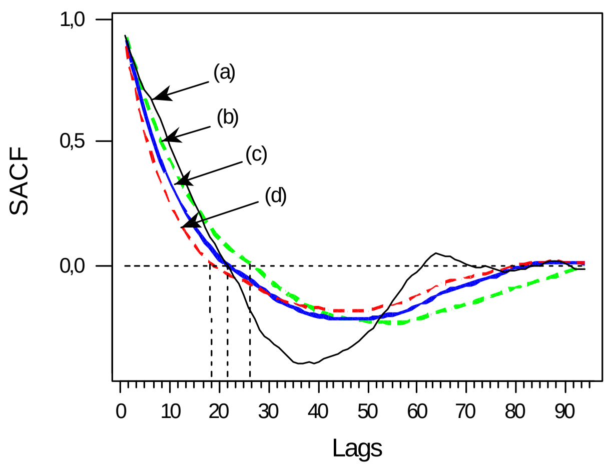
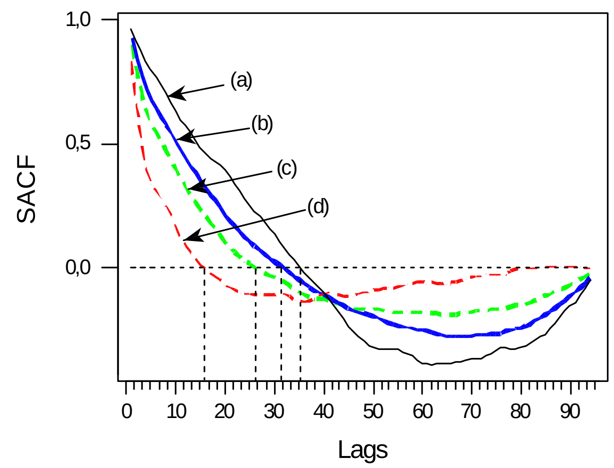


Figure 3d: (Above the upper border)



Curve (a) shows the original SACF function of the series, curve (b) shows the ASACF0, curve (c) shows the ASACF1, and curve (d) shows the ASACF2.

Figure 2 shows four distinguishing cases of an I(1) series for 50 observations: Figure 2(a) shows a series that has an ASACF1 near the stationary border, ASACF2; Figure2(b) shows a series that has an ASACF1 near the unit root border, ASACF0; Figure 2(c) shows a series that has an ASACF1 between the stationary border, ASACF2, and the unit root border, ASACF0; and, finally, Figure 2(d) shows a series that has an ASACF1 above the unit root border, ASACF0. The same conditions are shown in Figure 3 for 100 observations, in Figure 4 for 200 observations, and in Figure 5 for 500 observations.

By using our Bootstrap procedure, we are able to easily identify the integrated series from the stationary, but there are two questions to answer: (1) what is the size of the test procedure, that is, how many times the unit root is rejected in repeated samples when it is real a unit root process; and (2) how powerful is our method? In the next section, we investigate these questions using a small Monte Carlo experiment.

4. Results of the size and power of the Bootstrap Method.

In this section, we illustrate the use of our Monte Carlo experiment to identify the I(1) series from the stationary series by using the SACF. We calculate the *estimated size* by simply observing how many times the null is rejected in repeated samples, under conditions where the null is true. The number of replications per model chosen was 1000 for the size, and 1000 for the power of the tests. The calculations were performed using the GAUSS 3.2 package.

We generated the DGP (1) with $j = 1$;

$$y_t = y_{t-1} + u_t, \tag{7}$$

and then used the Bootstrap procedure described in the previous section to identify the series that has an ASACF1 less than the ASACF2.

Figure 4. The Bootstrap ASACF for 200 observations

Figure 4a: (Near stationary border)

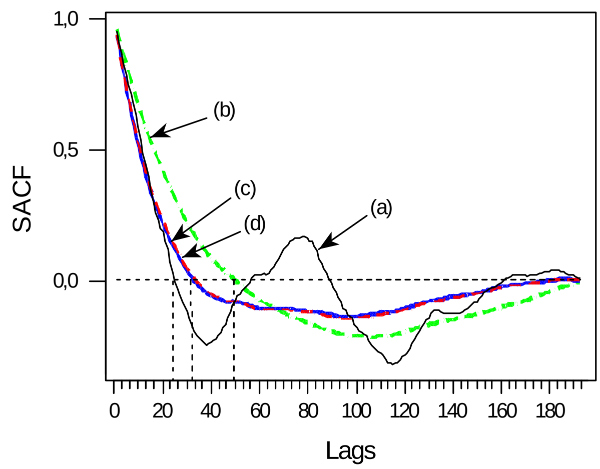


Figure 4b: (Near unit root border)

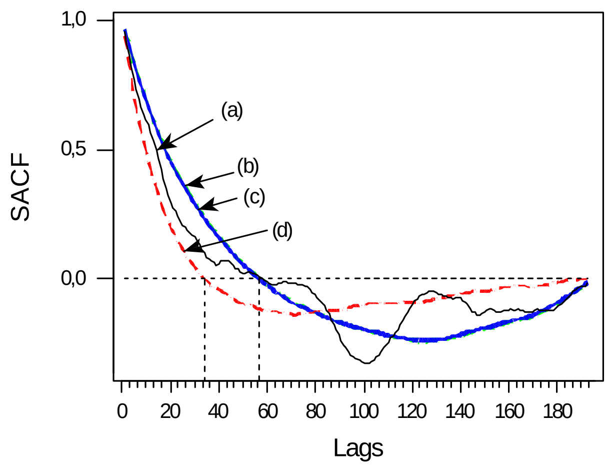


Figure 4c: (Between the upper and lower border)

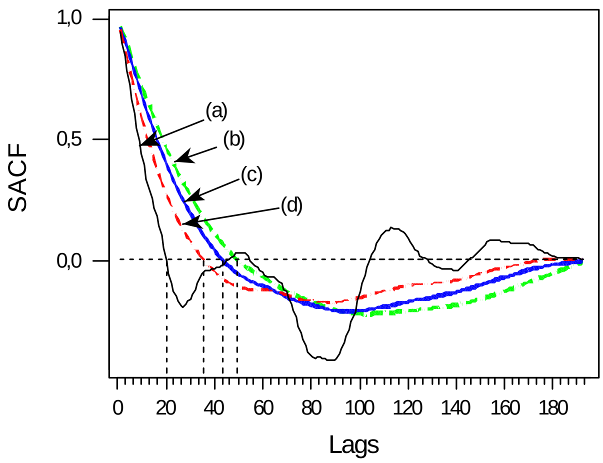
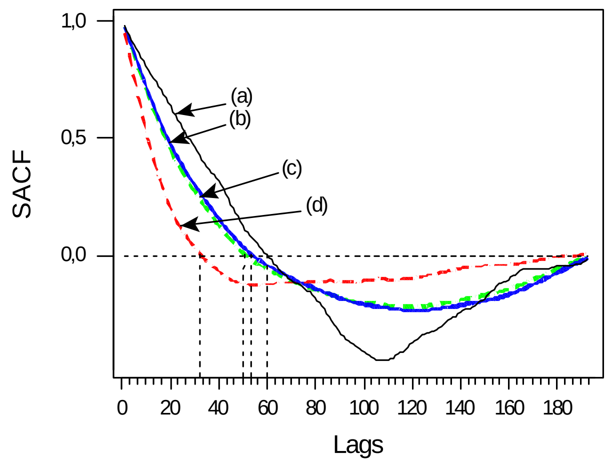


Figure 4d: (Above the upper border)



Curve (a) shows the original SACF function of the series, curve (b) shows the ASACF0, curve (c) shows the ASACF1, and curve (d) shows the ASACF2.

Figure 5. The Bootstrap ASACF for 500 observations

Figure 5a: (Near stationary border)

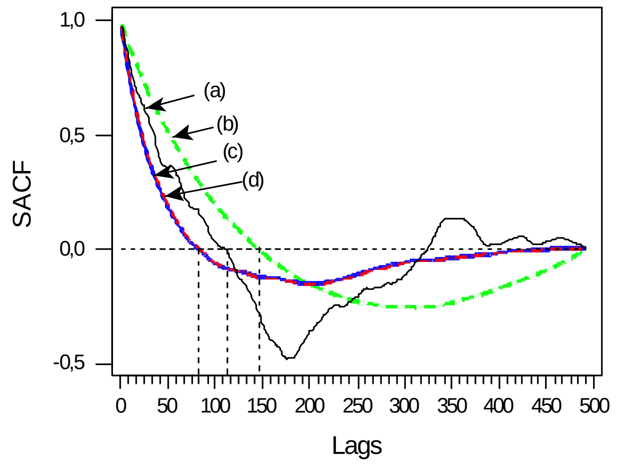


Figure 5b: (Near unit root border)

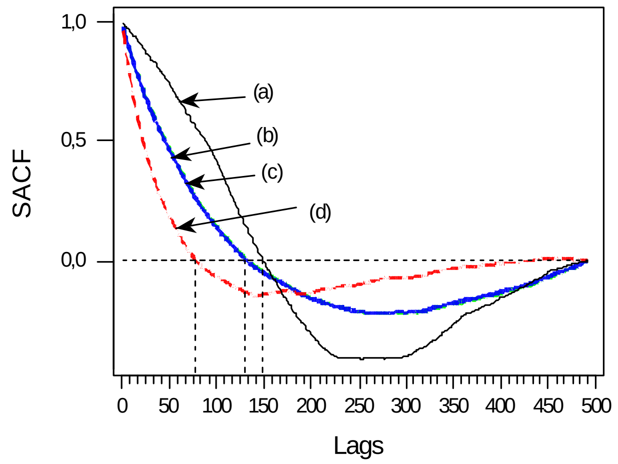


Figure 5c: (Between the upper and lower border)

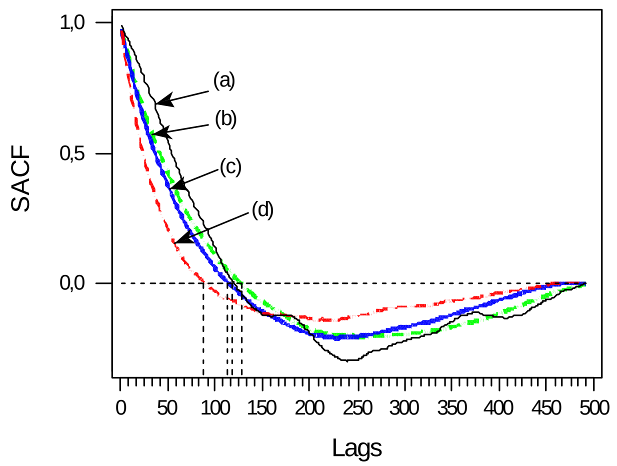
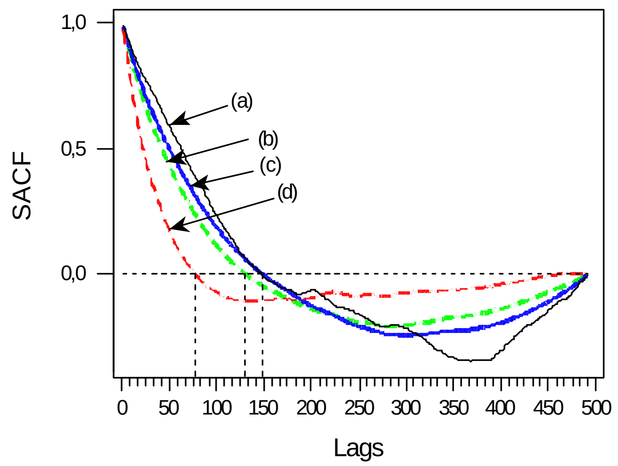


Figure 5d: (Above the upper border)



Curve (a) shows the original SACF function of the series, curve (b) shows the ASACF0, curve (c) shows the ASACF1, and curve (d) shows the ASACF2.

For each time series, 50 pre-sample values are generated with zero initial conditions, taking net sample sizes of T = 50, 100, 200, and 500. This was large enough to give a good insight on the small and large sample properties. Table 1 shows the relative rejection frequencies of the four different samples, (50, 100, 200, and 500).

Table 1. Estimated size for the Bootstrap procedure

	Sample size			
	50	100	200	500
Rejection frequencies	0.066	0.053	0.056	0.037

To judge the reasonability of the results, we used an approximated 95% confidence interval for the actual size (π) as:

$$\hat{p} \pm 2\sqrt{\frac{\hat{p}(1-\hat{p})}{N}}\ , \tag{8}$$

where $\hat{p}$ is the estimated size and N is the number of replications. This confidence interval, when using a 5% significance level, will be expected to lie between 3.6% and 6.4% in our test procedure.

Looking at the results displayed in Table 1, we can see that the test performed satisfactorily, except in the case of the small sample number, in which the relative rejection frequencies are outside the 95% confidence interval. Table 2a shows the lag value for which the ASACF1 is zero for a 5% significance level; that is, the lag in which only 5% of the repeated samples have an ASACF zero in less than the note lag. Finally, by studying the results shown in Table 2a, as a rule-of-thumb, we consider as an I(1) series those series that have zero value at lags larger than the T/6 lag. The size, by using the rule-of-thumb, is shown in Table 2b. Note that this rule-of-thumb is reasonable in approximating the relative rejection frequencies that are inside the 95% confidence interval.

Table 2a. Lags for a 5% size for the Bootstrap procedure and size by using the rule-of-thumb

	Sample size			
	50	100	200	500
Lags with a 5% significance level	8	17	33	80

Table 2b. Size by using the rule-of-thumb

	Sample size			
	50	100	200	500
Size by using the rule-of-thumb	0.042(<9)	0.059(<18)	0.056(<34)	0.064(<83)

The power function is estimated by calculating the rejection frequencies in 1000 replications, using the value of $j = 1 - \frac{T^{1/3}}{T}$, in the generated DGP (1)

$$y_t = j y_{t-1} + u_t, \qquad (1)$$

We followed the same process as used for the size investigation to evaluate the ASCFs by using the same sequence of random numbers that we used to estimate the size of the test. Table 3 shows the power when we used the Bootstrap procedure with a stationary border and with the rule-of-thumb. Note that even here, using the rule-of-thumb is good, as the power is greater than in the stationary border procedure.

Table 3. Power for the Bootstrap procedure

	Sample size			
	50	100	200	500
Power	0.468	0.488	0.543	0.548
Power by using the rule-of-thumb	0.669 (≤ 9)	0.659(≤18)	0.597(≤34)	0.856(≤83)

The conclusion to our size and power investigation is that our Bootstrap procedure is a powerful instrument to identify whether a series is an I(1) process.

5. An empirical example.

Mantalos (1999) studied co-movement by using the daily stock index data for France and Germany, covering the period 1 Nov. 1997 to 31 Mar. 1999. By Bootstrapping (a) the ECM-co-integration test procedure and (b) Johansen's co-integration procedure, he found that there is indeed a co-movement in the period 1 Nov. 1997 to 1 Nov. 1998, but then there is a tendency in the period Nov. 1998 to Mar. 1999 for the series to move apart.

Figure 6. Daily returns covering the period 1 Nov. 1997 to 31 Mar. 1999

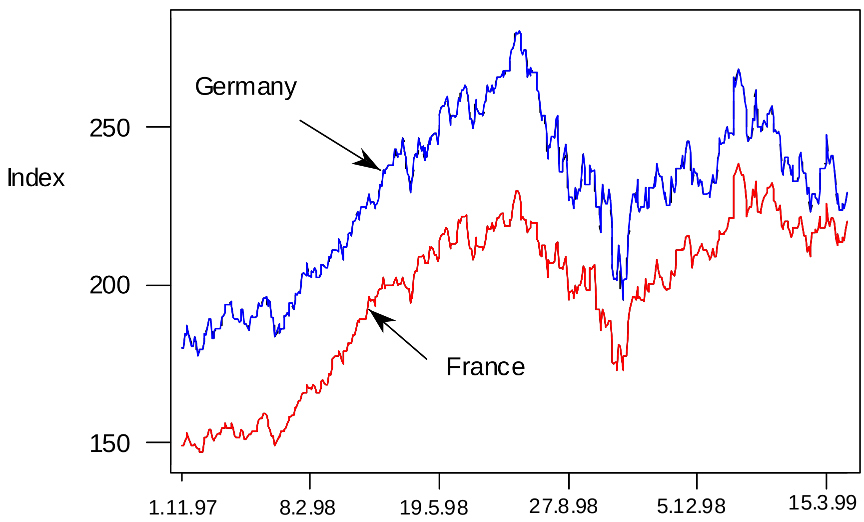


Figure 6 shows 373 observations of the International Dow Jones daily close-to-close index returns (Dow Jones Equity Market Index, DOW), covering the period 1 Nov. 1997 to 31 Mar. 1999, for the French and German markets.

Here we use the Granger and Engle (1987) residual based test, which is designed to test the null hypothesis of no co-integration by testing the null hypothesis that there is a unit root in the residuals against the alternative, that the root is less than unity. If the null hypothesis of unit root is rejected, then the null hypothesis of no co-integration is also rejected.

By using our Bootstrap ASACF principle, if the residuals show that the ASACF1 is larger than the ASACF2, then there is no co-integration between the series, and if the residuals show that the ASACF1 is less than the ASACF2, then the series are stationary, and probably there is co-integration between the French and German index returns.

We use two periods to study the co-movement: Period I, the first half of the data, for 1 Nov. 1997 to 16 Jul. 1998; and Period II, the second half of the data, for 17 July 1998 to 31 Mar. 1999.

Let us define Fra_t as the series of the French stock exchange index, and Ger_t as the series of the German stock exchange index. We use OLS to estimate the residuals from the co-integrating regression $Ger_t = a + b\,Fra_t + u_t$, and then identify whether the residuals $\hat{u}_t$ appear to be I(0) or not.

Figure 7a shows that the Bootstrap ASACF (ASACF1) has a zero at about the 60 lag, which is larger than the zero lag for the stationary border (ASACF2), and is even larger with the rule-of-thumb (308/6 = 51). The residuals are I(1). This result seems to corroborate the Mantalos (1999) study, and shows that there is evidence against co-integration in Period I.

Figure 7. The ASACF of the Residuals

Figure 7a: The ASACF for Period I

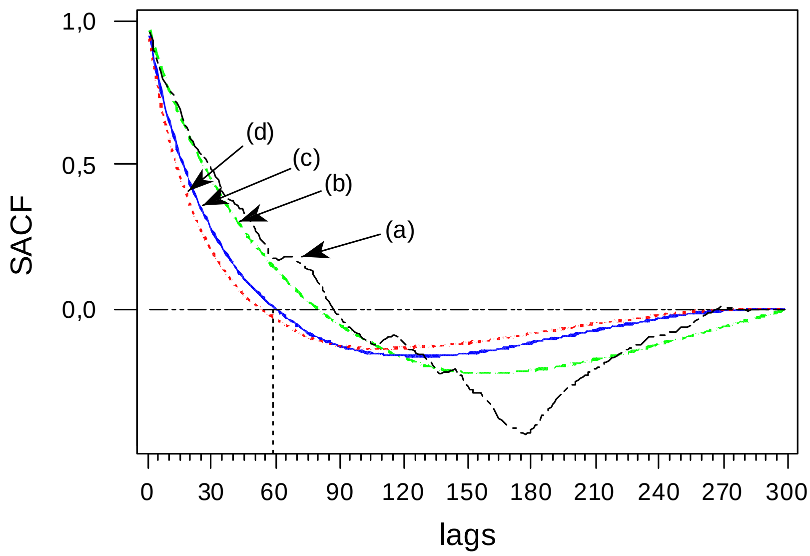
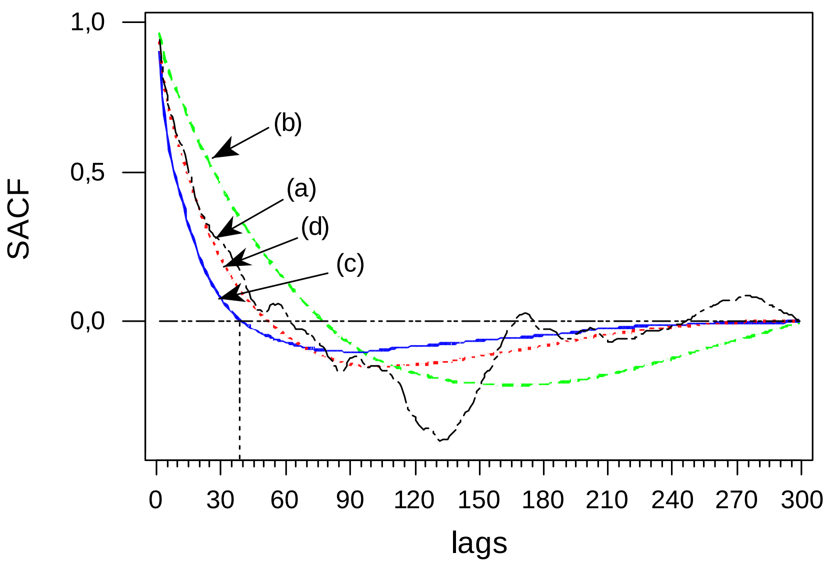


Figure 7b: The ASACF for Period II



Curve (a) shows the original SACF function of the series, curve (b) shows the ASACF0, curve (c) shows the ASACF1, and curve (d) shows the ASACF2.

Figure 7b shows that the Bootstrap ASACF (ASACF1) for Period II residuals is zero about the 39th lag, that is, much less than the zero lag for the stationary border (ASACF2). This result shows that there is no doubt that the residuals are a stationary series, and is strong evidence that the series are co-integrated in Period II.

6. Conclusions and Summary.

We have introduced a new Bootstrap procedure to distinguish between a non-stationary I(1) from a stationary AR(1) process by using a function of sample autocorrelations of a time series. We studied our Bootstrap procedure using a Monte Carlo method, and with an empirical example. The results show that by use of the Bootstrap average of the sample autocorrelations function, we can easily identify a non-stationary I(1). The advantages of our procedure are that it is easy to implement and easy to interpret. The principle is easily understood: series that have an ASACF1 larger than the ASACF2 (the ASACF1 function has a zero at a larger lag than the ASACF2) are integrated, and series that have an ASACF1 smaller than an ASACF2 are stationary. The combined use of this procedure that has the right size and high power gives us a visual representation, and extends the Box-Jenkins tradition in the identification of a non-stationary I(1), from a near-non-stationary AR(1) process.

We have studied the behaviour of the SACF for an AR(1) process, and by observation it showed that the near-integrated series has much less average sample autocorrelations than the integrated process, which makes it quite easy to distinguish it from an I(1) series, using our Bootstrap procedure. We use this Bootstrap procedure to define three different series. One Bootstrap series was confined to follow a unit root process, another, unrestricted, followed an ARIMA process, and a final series was confined to follow a near-integrated stationary process. Then we estimated the SACF for all three series by taking the average of the Bootstrap ASACF. Finally, by plotting the SACF in the same figure, we were able to distinguish the unit root series from the stationary process. Furthermore, our Monte Carlo experiment shows that the estimated size is about 5%, that is, about 50 out of the 1000 of the unit root processes are rejected in repeated samples under conditions where the unit root process is true. Moreover, the power of our procedure is adequate.

The empirical example used confirms that our procedure is an easy and powerful instrument in distinguishing between a non-stationary I(1) from the stationary AR(1) process.

References

Aczel, A. and Josephy, N. (1992), 'Using the Bootstrap for Improved ARIMA Model Identification', *Journal of Forecasting,* 11, 71-80.

Anderson, O.D. (1981), 'Covariance structure of Sample Correlations from ARUMA Models', in *TIME SERIES ANALYSIS* edited by O.D Anderson and M.R. Perryman, North-Holland, New York 3-26.

Dickey, D.A and 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.

Horowitz, J. L. (1994), 'Bootstrap-based critical values for the information matrix test', *Journal of Econometrics,* 61, 395-411.

Künsch, H.R (1989), 'The jackknife and the bootstrap for general stationary observations', *The Annals of Statistics*, 17, 1217-1241.

Kwiatkowski, D., Phillips, P.C.B. and Schmidt, P. (1992), 'Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root: How Sure Are We that Economic Time Series Have a Unit Root', *Journal of Econometrics*, 54, 159-178

Mammen, E. (1993), 'Bootstrap and wild bootstrap for high dimensional linear models', *The Annals of Statistics*, 21, 255-285.

Mantalos, P. (1998), 'A graphical Investigation of the Size and Power of the Granger-causality tests in Integrated-Cointegrated VAR Systems', Working paper 1998:2. Department of Statistics, University of Lund, Sweden. (*Studies in Nonlinear Dynamics and Econometrics*, 4(1), in press).

Mantalos, P. and G. Shukur (1998), 'Size and Power of the Error Correction Model (ECM) Cointegration Test- A Bootstrap Approach', *Oxford Bulletin of Economics and Statistics*, 60, 249-255.

Mantalos, P. and G. Shukur (1999), 'Testing for Cointegrating Relations - A Bootstrap Approach.' Working paper, Department of Statistics, University of Lund and University of Göteborg. (*Journal of Statistical Computation and Simulation*, in press).

Phillips, P. and P. Perron (1988), 'Testing for a Unit Root in Time series Regression', *Biometrika*, 75, 335-346.

Said, S.E. and Dickey, D.A (1984), 'Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order', *Biometrika*, 71, 599-607.

Wichern, D. (1973), 'The behaviour of the sample autocorrelation function for an integrated moving average process', *Biometrika*, 60, 235-239.