Lanne, Markku; Saikkonen, Pentti

Working Paper
Reducing size distortions of parametric stationarity tests

SFB 373 Discussion Paper, No. 2000,12

Provided in Cooperation with:
Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

Reducing Size Distortions of Parametric Stationarity Tests

Markku Lanne[†]

and

Pentti Saikkonen[‡]

University of Helsinki

Abstract

The use of asymptotic critical values in stationarity tests against the alternative of a unit root process is known to lead to overrejections in finite samples when the considered process is stationary but highly persistent. We claim that in recent parametric tests this is caused by estimation errors which result when the autoregressive parameters used to describe the short-run dynamics of the process are replaced by estimators. We suggest a modification that corrects for these errors and show by simulation that the modified test works reasonably well when the persistence is moderate and there is no time trend in the model. An empirical illustration with inflation rate data is provided.

Keywords: stationarity test, inflation rate

JEL Classification: C12, C22, E31

[0] [†]Research Unit on Economic Structures and Growth, Department of Economics, P.O. Box 54 (Unioninkatu 37), FIN-00014 University of Helsinki, FINLAND.

[0] [‡]Department of Statistics, P.O. Box 54 (Unioninkatu 37), FIN-00014 University of Helsinki, FINLAND.

[0] We are grateful to SFB 373 for financial support. Part of this research was done while the first author was visiting the Institute of Statistics and Econometrics at the Humboldt University in Berlin.

1 Introduction

The problem of discriminating between a stationary process and a unit root or so-called I(1) process has received considerable attention in the recent time series and econometrics literature. Unit root tests based on the null hypothesis of an I(1) process have been used extensively for this purpose since the early eighties and during the last few years the more recent tests based on the opposite null hypothesis of stationarity have also gained popularity. By now, several stationarity tests are available. They include the tests proposed by Kwiatkowski, Phillips, Schmidt, and Shin (1992), Saikkonen and Luukkonen (1993), Choi (1994), and Leybourne and McCabe (1994, 1999) to mention a few. The tests of Saikkonen and Luukkonen (1993) and Leybourne and McCabe (1994, 1999) are parametric and derived in the context of autoregressive integrated moving average (ARIMA) models whereas the tests of Kwiatkowski *et al.* (1992) and Choi (1994) are similar semiparametric tests. A problem with all these tests is that they are asymptotic and the use of asymptotic critical values can cause serious size distortions when the considered process is stationary but highly persistent. The simulation results of Saikkonen and Luukkonen (1993) were already indicative about this and recently the problem was pointed out very clearly by Caner and Kilian (1999) for the tests of Kwiatkowski *et al.* (1992) and Leybourne and McCabe (1994).

The purpose of this paper is to develop a modification of the parametric tests of Saikkonen and Luukkonen (1993) and Leybourne and McCabe (1994). The modification is developed in the framework of Leybourne and McCabe (1994) and its application requires that the least squares estimation used to compute the test statistic in that paper is replaced by a Gaussian maximum likelihood (ML) estimation of an auxiliary regression model with moving average errors and nonlinear multiplicative constraints between the regression coefficients and moving average parameters. Thus, compared with the original test, the required computations are clearly more involved but, as will be seen, the gains achieved in the finite sample size performance are substantial. The modified test works reasonably well when the persistence of the considered series is moderate but even it has difficulties to maintain its size when the persistence gets very high. This, however, is expected because stationary but highly persistent series can be modeled by nearly integrated processes whose long-run properties are closer to those of I(1) processes than stationary processes. Although our modification reduces size distortions in a model without a time trend, a disappointing finding is that this does not occur in a model with a time trend.

The plan of the paper is as follows. Section 2 introduces the considered stationarity tests and develops the modified test. Section 3 presents results of a simulation study which show the advantages of the modification. Section 4 provides an empirical illustration with inflation rate data from thirteen OECD countries. Finally, Section 5 concludes.

2 Test Procedures

The stationarity tests to be considered in this paper can be motivated by the simple model

$$y_t = \mu + \sum_{j=1}^{t} \eta_j + \varepsilon_t, \quad t = 1, \ldots, T, \tag{1}$$

where $\varepsilon_t \sim NID\,(0, \sigma_\varepsilon^2)$ and $\eta_t \sim NID(0, \sigma_\eta^2)$. Assuming that the sequences ε_t and η_t are mutually independent, Kwiatkowski *et al.* (1992) and Leybourne and McCabe (1994) derived a locally best invariant unbiased (LBIU) test for the null hypothesis $\sigma_\eta^2 = 0$ against the alternative $\sigma_\eta^2 > 0$. Thus, under the null hypothesis the observed series is generated by Gaussian white noise while under the alternative it contains a nonstationary random walk component. As is well known, model (1) is distributionally equivalent to the ARIMA(0,1,1) model

$$(1 - L)\, y_t = (1 - \theta L)\, \zeta_t, \quad 0 < \theta \leq 1, \tag{2}$$

where L is the usual lag operator, that is, $Ly_t = y_{t-1}$ and $\zeta_t \sim NID(0, \sigma_\zeta^2)$ with $\sigma_\zeta^2 = \sigma_\varepsilon^2/\theta$. Moreover, the parameter θ is related to σ_η^2 according to $\theta = \left(r + 2 - \left(r^2 + 4r\right)^{1/2}\right)/2$ where $r = \sigma_\eta^2/\sigma_\varepsilon^2$. This means that the null hypothesis $\sigma_\eta^2 = 0$ is equivalent to the moving average unit root hypothesis $\theta = 1$ in (2). This fact provides another possibility to obtain a stationarity test, as noticed by Saikkonen and Luukkonen (1993) (see also Tanaka, 1990). The LBIU test developed by these authors for the moving average unit root hypothesis $\theta = 1$ can also be obtained from model (1) by defining $\eta_j = (1 - \theta)\, \varepsilon_{j-1}$ and $\varepsilon_0 = 0$. In this approach there is thus a perfect linear dependence between the two sequences η_t and ε_t. Despite this difference, the LBIU test obtained in this framework is based on the same test statistic as the test based on the mutual independence assumption of the sequences η_t and ε_t.

The practical usefulness of models (1) and (2) is rather limited because, under the null hypothesis, no short-run dynamics are allowed. Therefore, various extensions of the stationarity tests initially derived for these models have been devised. These extensions differ in the way the short-run dynamics are taken into account. A non-parametric approach was employed by Kwiatkowski *et al.* (1992) while Saikkonen and Luukkonen (1993) and Leybourne and McCabe (1994) used a parametric approach. In this paper we only consider the parametric approach which in the above mentioned papers was implemented in two different ways. Saikkonen and Luukkonen (1993) generalized the error term ζ_t in (2) to a stationary and invertible ARMA process while Leybourne and McCabe (1994) added a finite number of lagged values of y_t to the right hand side of model (1). The latter approach may be more attractive from a practical point of view because the short-run dynamics are modeled by a simple autoregressive model. This approach will also be used in this paper. This is partly because of its simplicity but also because tests based on it had somewhat better size properties in preliminary simulation experiments. Thus, we extend model (1) to

$$\phi(L)\, y_t = \mu + \sum_{j=1}^{t} \eta_j + \varepsilon_t, \quad t = 1, \ldots, T, \tag{3}$$

where the lag polynomial $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$ has all its roots outside the unit circle. Instead of model (2) we now have

$$\phi(L)\,(1-L)\, y_t = (1-\theta L)\, \zeta_t, \quad 0 < \theta \leq 1. \tag{4}$$

Thus, under the null hypothesis $\sigma_\eta^2 = 0$ or $\theta = 1$ the observed series is generated by a stationary AR(p) process while under the alternative a nonstationary ARIMA(p,1,1) process applies.

As far as the subsequent stationarity tests are concerned, one can assume that in (3) the processes η_t and ε_t are either totally independent or $\eta_t = (1-\theta)\,\varepsilon_{t-1}$ with $\varepsilon_0 = 0$ $(t = 1, 2, \ldots)$. With the former assumption the model and stationarity test of Leybourne and McCabe (1994) are obtained. If the latter assumption is made the model becomes equivalent to that used by Saikkonen and Luukkonen (1993) except for the treatment of initial values in the short-run dynamics. Since the treatment of such initial values has no effect on the asymptotic properties of the related stationarity tests the test procedures to be described below can also be seen as asymptotically equivalent alternatives to the corresponding test of Saikkonen and Luukkonen (1993).

A stationarity test based on model (3) can be obtained by first replacing the lag polynomial $\phi(L)$ by a consistent estimator $\phi^*(L) = 1 - \phi_1^* L - \cdots - \phi_p^* L^p$, say, and then proceeding in the same way as in the simple model (1) with y_t replaced by $\phi^*(L)\,y_t$. As explained by Saikkonen and Luukkonen (1993) and Leybourne and McCabe (1994) the consistency of the test requires that the estimated lag polynomial $\phi^*(L)$ is not based on the null model. Therefore, $\phi^*(L)$ is obtained by estimating the parameters of the ARIMA(p,1,1) model (4) by ML. The next step is to construct the series

$$u_t^* = y_t - \sum_{j=1}^{p} \phi_j^* y_{t-j}, \quad t = p+1, \ldots, T,$$

and its demeaned version $\widehat{\varepsilon}_t = u_t^* - \overline{u}^*$ $(t = p+1, \ldots, T)$ where $\overline{u}^* = (T-p)^{-1} \sum_{t=p+1}^{T} u_t^*$.

The test is then based on the test statistic

$$\widehat{S} = \hat{\varepsilon}' V_{T-p} \hat{\varepsilon} / (T-p)^2 \widehat{\sigma}_{\varepsilon}^2 \tag{5}$$

where $\hat{\varepsilon} = [\hat{\varepsilon}_{p+1} \ldots \hat{\varepsilon}_T]'$, $\widehat{\sigma}_{\varepsilon}^2 = \hat{\varepsilon}'\hat{\varepsilon}/\left(T-p\right)$, and V_{T-p} is a $(T-p) \times (T-p)$ matrix with the ijth element equal to the minimum of i and j. Under the null hypothesis,

$$\widehat{S} \xrightarrow{d} \int_0^1 \left[B\left(r\right) - rB\left(1\right)\right]^2 dr \tag{6}$$

where $B\left(r\right)$ is a standard Brownian motion and consequently $B\left(r\right) - rB\left(1\right)$ is a standard Brownian bridge (see Leybourne and McCabe, 1994). Large values of the test statistic are critical. Critical values can be found in Kwiatkowski *et al.* (1992).

A problem with the above test is that the asymptotic null distribution given by (6) can provide a poor approximation for the actual finite sample size, as the recent simulations of Caner and Kilian (1999) demonstrate. Poor size performance occurs when the lag polynomial $\phi\left(L\right)$ has a root close to unity. This is intuitively understandable because then the observed series is close to an I(1) process even under the null hypothesis. However, for a series of one hundred observations or so rather serious size distortions occur even when the roots of $\phi\left(L\right)$ appear to be reasonably far away from unity. For instance, it will be seen in the next section that in the case $p = 1$ parameter values like $\phi_1 = 0.6 - 0.7$ are already difficult. This is unfortunate because such parameter values are relevant in many applications where the test is used to discriminate between a stationary but rather persistent series and an I(1) series.

In the simple special case (1) the asymptotic null distribution of test statistic $\widehat{S}$ is known to provide a good approximation for the actual finite sample distribution. This implies that the reason for the poor size performance of the test is that estimators are used in place of the unknown parameters in the lag polynomial $\phi(L)$ or that the series $u_t = \phi(L)\, y_t$ is replaced by u_t^*. To see this, write

$$u_t^* = u_t - \sum_{j=1}^{p} (\phi_j^* - \phi_j) y_{t-j} \tag{7}$$

and assume that the null hypothesis holds. When the process y_t is persistent its variation is large and, even though $\phi_j^* - \phi_j$ converges to zero in probability, the second term on the right hand side of (7) can have a substantial effect on the series u_t^* and hence on the finite sample distribution of the test. A fairly obvious way to allow for the estimation errors on the right hand side of (7) would be to base the test statistic not on $\widehat{\varepsilon}_t$ but on the least squares residuals of a regression of u_t^* on an intercept term and $y_{t-1}, \ldots, y_{t-p}$. This, however, destroys the power of the test because under the alternative both u_t^* and y_t are affected by the same random walk component. This implies that even in large samples the above mentioned least squares residuals behave like a stationary series and not like an I(1) series as assumed by the test. (A similar argument actually explains why the estimator $\phi^*(L)$ cannot be based on the null model.) To overcome this difficulty, we shall replace (7) by an alternative representation which makes use of the well-known decomposition

$$\phi(L) = \phi(1) + \varphi(L)(1 - L)$$

where $\varphi(L) = \sum_{j=0}^{p-1} \varphi_j L^j$ with $\varphi_j = \sum_{i=j+1}^{p} \phi_i$. Using this identity we can write

$$
\begin{aligned}
-\sum_{j=1}^{p} (\phi_j^* - \phi_j) L^j &= (\phi^*(L) - \phi(L))\, \phi(1)^{-1} \phi(1) \qquad (8)\\
&= \beta(L)\, L\phi(L) - \beta(L)\, L\varphi(L)\, \Delta
\end{aligned}
$$

where $\Delta = 1 - L$ and $\beta(L)\, L = (\phi^*(L) - \phi(L))\, / \phi(1)$. A more explicit expression of $\beta(L)$ is given by $\beta(L) = \sum_{j=1}^{p} \beta_j L^{j-1} = \sum_{j=1}^{p} \phi(1)^{-1} (\phi_j^* - \phi_j) L^{j-1}$. Now, denoting $w_t = \sum_{j=1}^{t} \eta_j + \varepsilon_t$ we have $\phi(L)\, y_t = u_t = \mu + w_t$ and it follows from (8) that equation (7) can be written as

$$
u_t^* = \nu - \beta(L)\, \varphi(L)\, \Delta y_{t-1} + w_t + \beta(L)\, w_{t-1} \qquad (9)
$$

where $\nu = \mu + \beta(1)\, \mu$. Thus, we have obtained a regression model for the residual series u_t^*. The parameters in this regression model are not parameters in the usual sense, though, because they depend on the estimation errors $\phi_j^* - \phi_j$ and are therefore random variables which converge to zero in probability as the sample size tends to infinity. The regressors consist of an intercept term and lagged values of the differences Δy_t. The errors are autocorrelated and follow a conventional pth order moving average process under the null hypothesis. However, under the alternative the innovations of this moving average process contain the same random walk component as the previous models (1) and (3). There are also multiplicative constraints between the regression coefficients and the parameters of the error process.

An important feature in the regression model (9) is that, instead of lags of the levels y_t, the regressors include lags of the differences Δy_t which are stationary whether the null hypothesis is true or not. Therefore, the regressors in (9) cannot explain the I(1) type nonstationarity of the regressand which occurs under the alternative. This implies that no serious loss of power should result from using model (9) to allow for the estimation errors caused by replacing $\phi(L)\, y_t$ in (3) by $\phi^*(L)\, y_t$. Asymptotically the regression model (9) reduces to (3). Thus, our idea is to estimate the parameters in (9) and use the resulting residuals instead of $\widehat{\varepsilon}_t$ to construct a modified version of test statistic $\widehat{S}$. We assume that the null hypothesis holds and estimate the parameters in (9) by ML. Since $w_t = \varepsilon_t$ under the null hypothesis this amounts to using ML to estimate parameters of a regression model with Gaussian moving average errors. Since the lag polynomial $\beta(L)\,\varphi(L)$ is of order $2(p-1)$ model (9) is considered for $t = 2p+1, \ldots, T$ in this estimation and the multiplicative constraints between the regression coefficients and the moving average parameters are taken into account. In the special case $p = 1$ we have $\beta(L) = \beta_1$ and $\varphi(L) = \varphi_0$, and there is no need to take these constraints into account because no reduction in the dimension of the parameter space is achieved. To illustrate the constraints in the general case, suppose that $p = 2$. Then equation (9) becomes

$$\begin{aligned} u_t^* &= \nu - \beta_1\varphi_0\Delta y_{t-1} - (\beta_1\varphi_1 + \beta_2\varphi_0)\Delta y_{t-2} - \beta_2\varphi_1\Delta y_{t-3} \\ &\quad + w_t + \beta_1 w_{t-1} + \beta_2 w_{t-2} \end{aligned}$$

demonstrating that by taking the constraints into account the number of parameters is reduced by one. In general the reduction is $p-1$.

Now, let $\widetilde{\varepsilon}_t$ $(t = 2p+1, \ldots, T)$ be the residuals from the ML estimation described in the preceding paragraph. Our modified test is based on the test statistic

$$\widetilde{S} = \widetilde{\varepsilon}' V_{T-2p} \widetilde{\varepsilon} / (T-2p)^2 \widetilde{\sigma}_{\varepsilon}^2 \tag{10}$$

where $\widetilde{\varepsilon} = [\widetilde{\varepsilon}_{2p+1} \ldots \widetilde{\varepsilon}_T]'$, $\widetilde{\sigma}_{\varepsilon}^2 = \widetilde{\varepsilon}'\widetilde{\varepsilon}/(T-2p)$, and V_{T-2p} is a $(T-2p) \times (T-2p)$ matrix with the ijth element equal to the minimum of i and j. As will be discussed below, test statistics $\widetilde{S}$ and $\widehat{S}$ have the same limiting distribution under the null hypothesis. However, one would expect that the modification improves the accuracy of the asymptotic approximation because the residuals $\widetilde{\varepsilon}_t$ are obtained by modeling the estimation errors in the series u_t^* whereas the residuals $\widehat{\varepsilon}_t$ are directly affected by these estimation errors. On the other hand, the modification should not destroy the power of the test because under the alternative hypothesis the residuals $\widetilde{\varepsilon}_t$ should behave like an I(1) series.

We shall not prove in detail that the limiting null distribution of test statistic $\widetilde{S}$ is the same as that of $\widehat{S}$ but only briefly indicate how a proof can be obtained. First, one needs to establish the consistency of the ML estimators of the intercept term and the parameters in the lag polynomials $\beta(L)$ and $\varphi(L)$ in (9). This can be done by following the arguments in Hannan and Deistler (1988, Chapter 4.2) which also apply to models with multiplicative parameter constraints of the type in (9). After the consistency has been established the limiting distribution of test statistic $\widetilde{S}$ can be obtained in the same way as in the previous similar cases in Saikkonen and Luukkonen (1993, 1996) and Leybourne and McCabe (1994). In particular, one can follow the proof in Saikkonen and Luukkonen (1996) where the authors derive the limiting distribution of a moving average unit root test in a regression model with ARMA errors. Although the regressors were assumed to be strictly exogenous in that paper similar arguments apply in the case of model (9) and give the desired result.

We wish to emphasize that the parameters in (3) and (9) are supposed to be estimated by (exact) ML (with $\Delta y_2, \ldots, \Delta y_{2p}$ treated as constants in (9)). As pointed out by Leybourne and McCabe (1999), ML procedures in programs such as GAUSS and SHAZAM have no difficulty in fitting model (3) with $\theta = 1$. However, our experience with the GAUSS-ARIMA procedure is that it is not uncommon that the algorithm fails to find a global maximum of the likelihood function. Therefore, care is needed to ensure that a global maximum of the likelihood function is found when model (3) is fitted. A more detailed discussion of this point along with some practical recommendations will be given in the next section.

We close this section by noting that the above modification can also be used in a model with a linear time trend. Then a linear time trend also appears in model (9) but otherwise the modified test is obtained in the same way as before (cf. Leybourne and McCabe, 1994). However, we will not give details of this modification because simulation experiments showed that its size performance is more or less the same as that of the original test. Reasons why the modification does not help in this case are not clear to us. However, as will be explained in the next section, the estimation of the parameters of the auxiliary regression model (9) gets unreliable when the lag polynomial $\phi(L)$ has a root close to unity and, due to the large variation of the trend variable, the effect of the resulting estimation errors on the modified test may be much larger in a model with a trend than in a model without a trend.

3 Simulation Study

In this section we provide Monte Carlo simulation results to illustrate the finite sample performance of the original and modified stationarity tests. Although the modification seems to correct for overrejection in a model with no time trend, no such improvement is detected in a model including a time trend, as mentioned above. Therefore, the results for the latter case are not reported. This finding is somewhat discouraging, but on the other hand, in that case the overrejection problem is not nearly as severe as in the case with no time trend. All results are based on only 1,000 replications because the estimation involved in computing the value of the test statistic is computationally burdensome (see below). For judging the results, it may be worth taking into account that the standard error of an estimator of a true rejection probability P based on 1,000 replications is $\sqrt{P\left(1-P\right)/1,000}$. Hence, for significance levels $P = 0.01$, 0.05 and 0.1, the two-standard-error confidence intervals are $\left[0.007, 0.013\right]$, $\left[0.043, 0.057\right]$ and $\left[0.091, 0.110\right]$, respectively.

As Caner and Kilian (1999) recently pointed out, the choice of starting values in the (exact) ML estimation of the parameters of the ARIMA model (4) has a large effect on the behavior of test statistic $\widehat{S}$ because with inappropriate starting values a global maximum of the likelihood function is not likely to be found. These authors followed Leybourne and McCabe (1994) in selecting the initial value for the moving average parameter θ on a grid ranging from 0 to 1 with the autoregressive parameter(s) fixed at $\theta - 0.1$, but in addition, they also included the default starting values provided by the GAUSS-ARIMA routine among the candidate starting values. Instead of keeping the initial values of the autoregressive parameters fixed, we consider generalized least squares (GLS) estimators based on model (4) with the value of θ varying over a grid from 0 to 1. For each set of starting values obtained in this way the parameters of model (4) are estimated by ML and the value of the likelihood function is computed. After this, estimates maximizing the likelihood function are selected. It turned out that with fixed starting values of the autoregressive parameters the algorithm often fails to find a global maximum of the likelihood function, while our procedure involving the initial GLS estimation does find it.

Our experimentation with different ways of selecting the starting values suggests that for the size properties of the modified test statistic $\widetilde{S}$ finding good starting values is not equally critical. If the largest root of the lag polynomial $\phi(L)$ is not very close to unity, the real size of the modified test is not much affected by the choice of starting values. Still, the empirical size is always closest to the nominal size if the GLS estimation described above is used, especially for highly persistent processes. Hence it may be advisable to employ that procedure despite the fact that it is computationally burdensome. The exact ML estimation of the parameters of model (9) did not seem so sensitive to the choice of starting values, and the reported results are based on the default starting values provided by the GAUSS-ARIMA routine.

The size of test statistics $\widehat{S}$ and $\widetilde{S}$ is studied in the cases of AR(1) and AR(2) data generation processes (DGP) with various values of the autoregressive parameters.

Table 1 presents rejection rates for the AR(1) case, i.e., the DGP is

$$y_t = \phi_1 y_{t-1} + \varepsilon_t, \qquad t = 1, \ldots, T,$$

where $\varepsilon_t \sim NID(0,1)$. For 100 observations the original test statistic $\widehat{S}$ overrejects for all values of ϕ_1. For highly persistent series ($\phi_1 \geq 0.9$) the modification does not provide sufficient size correction but already for $\phi_1 = 0.8$ the empirical size of the modified test statistic $\widetilde{S}$ comes close to the nominal size. Notice that the relative size distortions of the modification depend on the significance level of the test which indicates that for very small samples the tail of the true distribution has a shape rather different from that of the asymptotic distribution (6). At the 5% level the distortions seem to be smallest. The rejection rates of the original test statistic $\widehat{S}$ are roughly in accordance with, although somewhat smaller than, those obtained by Caner and Kilian (1999). For 200 observations the size distortion of the modified test is only modest already for $\phi_1 = 0.9$, and for the lower values of ϕ_1 the modified test seems somewhat conservative. Intuitively, the reason why the modification does not alleviate the overrejection problem for highly persistent series, can be easily seen in the AR(1) case. From (3) it can be seen that, under the null hypothesis, Δy_t approaches ε_t as ϕ_1 approaches unity, and therefore model (9) is almost unidentified, making the estimation unreliable. For negative values of ϕ_1 (not reported) the size of the original test statistic $\widehat{S}$ is very close to the nominal size already for 100 observations, and the modification does not provide much improvement.

Rejection rates for the AR(2) case, i.e., for the following DGP,

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, \qquad t = 1, \ldots, T,$$

are presented in Table 2. We simulated AR(2) models with both roots of the autoregressive polynomial equal to 0.8, 0.7 and 0.6, corresponding to the sums of the autoregressive coefficients equal to 0.96, 0.91 and 0.84, respectively. In this context the sum of the autoregressive coefficients turned out to be the best indicator of the persistence of the series and thus also of the performance of the modification. For $\phi_1 + \phi_2 = 0.96$ the reduction in the rejection rate due to the modification turned out to be substantial but the size is not yet very close to the nominal size for even 200 observations. When the sum of the autoregressive coefficients equals 0.91, the modification is satisfactory for $T = 200$ whereas there is still room for improvement for $T = 100$. It is only when $\phi_1 + \phi_2 = 0.84$ that the modification almost reaches the nominal size with 100 observations. Notice that even in this case the original test statistic $\widehat{S}$ leads to overrejections for both sample sizes and for the smaller sample size $T = 100$ the overrejections are severe.

To show that the modification does not destroy the power of the test, Table 3 presents results of empirical power simulations. The DGP is model (1) with $\mu = 0$, and the test is based on estimating the parameters of an ARIMA(1,1,1) model (4) at the first stage. This model is thus overparameterized which should affect the power adversely. For a given value of σ_η^2 the power increases with sample size as expected. The power also initially increases with the variance of the random walk component (or equivalently, as θ gets further away from unity), but for $\sigma_\eta^2 \geq 2$ the power starts to decline. This may be due to the fact that when the value of σ_η^2 increases the ARIMA(1,1,1) model (4) becomes nearly unidentified. This implies large estimation errors which may make the estimation of the parameters of the auxiliary regression model (9) difficult and thereby impair the effectiveness of the modification. Still, the test is reasonably powerful against alternatives very close to a pure random walk with only 100 observations. We also computed the empirical power against a pure random walk, and with $\sigma_\eta^2 = 1$. At the 5% level of significance the rejection rates were 0.66 and 0.70 for samples of 100 and 200, respectively. We also tried the sample size $T = 500$ and obtained the empirical power 0.78. The corresponding rejection rates for the original test statistic $\widehat{S}$ show a similar pattern although because of the overrejection problem these figures cannot be interpreted as power. These results go contrary to the recent claim of Hobijn, Franses, and Ooms (1998) that test statistic $\widehat{S}$ is not consistent against the alternative of a pure random walk.

4 Empirical Illustration

As an empirical example we consider testing the stationarity of inflation rates. Whether inflation can be considered a stationary or unit root process has a number of economic implications, and inflation rates from different countries and time periods have been extensively subjected to unit root and stationarity tests (for references to some of the previous studies, see Culver and Papell, 1997). The data set consists of the quarterly inflation of thirteen OECD countries, computed from consumer price index data obtained from the International Financial Statistics published by the International Monetary Fund. The inflation rate is calculated by differencing the logarithm of the consumer price index. The series cover the period from 1957:II to 1997:III (162 observations). Corresponding monthly data were recently studied by Culver and Papell (1997).

The test results are presented in Table 4. In addition to test statistic $\widehat{S}$ and its modified version $\widetilde{S}$ we also report the results of the augmented Dickey–Fuller (ADF) test and the ADF-GLS test of Elliott, Rothenberg, and Stock (1996) for comparison. The latter test is expected to be more powerful against local alternatives. The lag length of the autoregressive polynomial $\phi(L)$ was selected by BIC. Time trend is not included in the models that the test are based on because a trend would not be consistent with long-run non-accelerating inflation. The irrelevance of a trend is also supported by visual inspection of the graphs of the inflation series. There seems to be predominantly evidence for stationarity of inflation. Test statistic $\widehat{S}$ rejects the null of stationarity only for Belgium, Italy and Norway at the 1% level and for the UK at the 5% level, whereas the modified test rejects only for the Netherlands at the 1% level and for Italy at the 5% level. In addition, the modified test rejects for Belgium and Germany at the 10% level. The reason that the modified test clearly rejects for the Netherlands although the original test does not reject, seems to be the sensitivity of test statistic $\widehat{S}$ for initial values. The 1957:II observation for this series is quite exceptional (the second largest observation in the series). The values of test statistics $\widehat{S}$ and $\widetilde{S}$ for the series excluding the first observation are 1.02 and 0.84, respectively, indicating rejection of the stationarity null by both tests. This result suggests that the modified test is not as sensitive to initial values as the original test. The modified test also seems to accord with the unit root tests more often than the original test. At the 10% level there is only one discrepancy between the ADF-GLS test and our modified test: for Belgium both tests reject. The $\widehat{S}$ and ADF-GLS tests, on the other hand, disagree in five cases: for Belgium, Germany, the Netherlands, Norway and the UK.

When the stationarity tests of this paper are applied it is reasonable to have an idea about the persistence of the short-run dynamics of the considered series. Therefore we computed the sums of the estimated autoregressive parameters in model (4). Typically they varied between 0.8 and 0.9. The maximum was 0.91 (Canada and France), and for Norway the sum was as small as 0.27. Judging by the simulation results in Table 2 the modification can hence be expected to work fairly well for these series.

5 Conclusion

In this paper we have modified previous stationarity tests and shown that in many cases of practical relevance the modified test has substantially smaller size distortions than its original counterpart. Compared with the original test, results of the modified test were also in closer agreement with those of unit root tests in the empirical illustration of the paper. When the persistence of the considered series is very high even the modified test was seen to suffer from size distortions, though. However, since such series can be interpreted as realizations of nearly integrated processes whose long-run properties are similar to those of I(1) processes this is simply what one can expect. Thus, even when the modified test of this paper is applied it is advisable to try to asses the persistence of the short-run dynamics of the considered series. Studying the roots of the autoregressive polynomial estimate obtained from the ML estimation of the parameters of model (4) may be useful in this respect. If roots very close to unity are found the results of the test should be interpreted with caution. Another situation where care is needed is when a time trend is included in the model because then the modification does not reduce size distortions. It is also to be emphasized that finding the maximum of the likelihood function for model (4) is not necessarily easy, and care must be taken especially in selecting the starting values. The procedure where a grid is formed for the moving average parameter and the other starting values are found by GLS estimation prior to the exact ML estimation seems to work satisfactorily.

References

Caner, M. and L. Kilian (1999). Size distortions of tests of the null hypothesis of stationarity: Evidence and implications for the PPP debate. Mimeographed.

Choi, I. (1994). Residual-based tests for the null of stationarity with applications to U.S. macroeconomic time series. Econometric Theory 10, 720-746.

Culver, S.E., and D.H. Papell (1997). Is there a unit root in the inflation rate?

Evidence from sequential break and panel data models. Journal of Applied Econometrics 12, 435-444.

Elliott, G., T.J. Rothenberg, and J.H. Stock (1996). Efficient tests for an autoregressive unit root. Econometrica 64, 813-836.

Hannan, E.J. and M. Deistler (1988). The Statistical Theory of Linear Systems. Wiley, New York.

Hobijn, B., P.H. Franses, and M. Ooms (1998). Generalizations of the KPSS-test for stationarity. Econometric Institute Report 9802/A, Erasmus University Rotterdam.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin (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 154, 159-178.

Leybourne, S.J. and B.P.M. McCabe (1994). A consistent test for a unit root. Journal of Business & Economic Statistics 112, 157-166.

Leybourne, S.J. and B.P.M. McCabe (1999). Modified stationarity tests with data-dependent model-selection rules. Journal of Business & Economic Statistics 17, 264-270.

Saikkonen, P. and R. Luukkonen (1993). Testing for a moving average unit root in autoregressive integrated moving average models. Journal of the American Statistical Association 188, 596-601.

Saikkonen, P. and R. Luukkonen (1996). Testing the order of differencing in time series regression. Journal of Time Series Analysis 17, 481-496.

Tanaka, K. (1990). Testing for a moving average unit root. Econometric Theory 16, 433-444.

Table 1: Empirical size of the original and modified Leybourne–McCabe tests when the DGP is an AR(1) model.

ϕ_1	Nominal size	T = 100 $\widehat{S}$	100 $\widetilde{S}$	200 $\widehat{S}$	200 $\widetilde{S}$
0.95	0.01	0.403	0.304	0.242	0.164
	0.05	0.453	0.385	0.286	0.218
	0.10	0.484	0.446	0.333	0.278
0.90	0.01	0.218	0.114	0.091	0.026
	0.05	0.278	0.193	0.148	0.074
	0.10	0.314	0.262	0.199	0.136
0.80	0.01	0.088	0.026	0.035	0.002
	0.05	0.159	0.076	0.086	0.045
	0.10	0.223	0.154	0.134	0.094
0.70	0.01	0.049	0.007	0.021	0.002
	0.05	0.114	0.054	0.066	0.040
	0.10	0.179	0.130	0.114	0.084
0.60	0.01	0.036	0.007	0.017	0.004
	0.05	0.095	0.045	0.061	0.039
	0.10	0.159	0.112	0.103	0.081

The figures are rejection rates based on 1,000 replications of an AR(1) model with autoregressive coefficient ϕ_1 and without drift. T is the sample size.

Table 2: Empirical size of the original and modified Leybourne–McCabe tests when the DGP is an AR(2) model.

ϕ_1	ϕ_2	$\phi_1 + \phi_2$	Nominal size	T = 100 $\widehat{S}$	100 $\widetilde{S}$	200 $\widehat{S}$	200 $\widetilde{S}$
1.60	-0.64	0.96	0.01	0.411	0.319	0.212	0.041
			0.05	0.546	0.240	0.346	0.101
			0.10	0.614	0.338	0.427	0.181
1.40	-0.49	0.91	0.01	0.145	0.026	0.043	0.006
			0.05	0.230	0.082	0.110	0.033
			0.10	0.295	0.175	0.166	0.098
1.20	-0.36	0.84	0.01	0.078	0.006	0.023	0.007
			0.05	0.145	0.054	0.073	0.030
			0.10	0.213	0.116	0.122	0.085

The figures are rejection rates based on 1,000 replications of an AR(2) model with autoregressive coefficients ϕ_1 and ϕ_2 and without drift. T is the sample size.

Table 3: Empirical power of the modified Leybourne–McCabe test.

σ_η^2	θ	T					
		100			200		
		Nominal size					
		0.01	0.05	0.10	0.01	0.05	0.10
0.001	0.969	0.048	0.152	0.240	0.204	0.347	0.457
0.010	0.905	0.328	0.482	0.586	0.646	0.784	0.840
0.100	0.730	0.647	0.785	0.863	0.873	0.937	0.963
1.000	0.382	0.835	0.916	0.950	0.970	0.982	0.989
2.000	0.268	0.817	0.880	0.907	0.965	0.980	0.983
4.000	0.172	0.739	0.795	0.823	0.899	0.917	0.928
8.000	0.101	0.641	0.704	0.746	0.776	0.798	0.820

The DGP is the following: $y_t = \sum_{j=1}^{t} \eta_j + \varepsilon_t, \quad t = 1, \ldots, T$, where $\varepsilon_t \sim NID\,(0,1)$ and $\eta_t \sim NID\left(0, \sigma_\eta^2\right)$. Equivalently the DGP can be expressed as $y_t = y_{t-1} + \zeta_t - \theta\zeta_{t-1}$ with $\zeta_t \sim NID(0, \theta^{-1})$. The test is based on estimating and ARIMA(1,1,1) model (4) in the first stage.

Table 4: Stationarity and unit root tests for quarterly inflation rates 1957:II–1997:III.

Country	Lag	ADF[a]	ADF-GLS[b]	$\widehat{S}^c$	$\widetilde{S}^c$
Belgium	3	-2.63*	-2.64***	2.35***	0.42*
Canada	4	-1.87	-1.64*	0.15	0.24
Finland	4	-2.44	-2.81**	0.17	0.16
France	4	-3.70***	-2.73***	0.29	0.16
Germany	4	-2.59*	-1.60	0.10	0.37*
Italy	3	-2.22	-1.61	1.69***	0.50**
Japan	4	-2.89**	-2.72***	0.34	0.19
Luxemburg	4	-2.36	-2.27**	0.14	0.15
Netherlands	4	-2.77*	-0.67	0.19	0.76***
Norway	4	-3.09**	-2.62***	0.74***	0.18
Spain	4	-2.48	-1.86*	0.18	0.28
United Kingdom	5	-2.64*	-2.51**	0.64**	0.16
United States	4	-2.03	-2.04**	0.12	0.12

***, ** and * indicate rejection at the 1, 5 and 10 % significance levels, respectively.

[a]The 1, 5 and 10% critical values: -3.43, -2.86, -2.57.

[b]The 1, 5 and 10% critical values: -2.58, -1.95, -1.62.

[c]The 1, 5 and 10% critical values: 0.74, 0.46, 0.35.