Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Econometrics 135 (2006) 285–310
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Minimizing the impact of the initial condition on testing for unit roots

Graham Elliott[a,*], Ulrich K. Müller[b]

[a] University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA
[b] Princeton University, Economics Department, Princeton, NJ 08544, USA

Available online 31 August 2005

Abstract

The outcome of popular unit root tests depends heavily on the initial condition, i.e. on the difference between the initial observation and the deterministic component. In some applications it is difficult to rule out small or large values of the initial condition a priori, so this dependence can be quite difficult to deal with in practice. We explore a number of methods for constructing unit root tests whose properties are less affected by the initial condition. We show that no nontrivial test can remain completely unaffected, and instead derive an asymptotically efficient unit root test whose power varies relatively little as a function of the initial condition.

JEL classification: C12; C22

Keywords: Invariance; Characteristic function; Efficient tests

1. Introduction

There exist by now many tests for examining whether or not a time series has a unit root. A classical (frequentist) approach for choosing between tests amounts to choosing an appropriate power profile. Power should be maximized against alternatives of interest and the outcome of the test should depend as little as possible on nuisance

*Corresponding author. Tel.: +1 858 534 4481; fax: +1 858 534 7040.
E-mail addresses: gelliott@weber.ucsd.edu (G. Elliott), umueller@princeton.edu (U.K. Müller).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2005.07.024

parameters of the problem. For the unit root testing problem, an important nuisance parameter is the initial condition: the deviation of the first observation from its deterministic component—the initial condition—strongly influences the power of popular tests, up to the point of reversing the power ranking of different tests. See Evans and Savin (1981, 1984), and Stock (1994) for additional references.

Müller and Elliott (2003) (abbreviated ME in the following) extend results in Elliott et al. (1996) and derive a family of efficient tests that allow for various possible weightings on the initial condition. A clear trade-off emerges between power for small and large[1] initial conditions, hence the decision of which test to choose is not trivial. For situations where the researcher has some idea of the possible size of the initial condition, these tests or their asymptotic equivalents are likely to be good choices. Specifically, in many testing situations it would not be expected that the initial observation of the series is 'unusual', where unusual starting observations imply a very large initial condition. Often the data collected and available for testing has a beginning date dictated by how far back records can be found, or they date back to periods when governments started collecting and disseminating data on a regular basis. In many of these situations a researcher can make the highly plausible assumption that the absolute value of the initial condition is relatively small, which suggests the use of a unit root test that exploits this knowledge.

In other testing situations, the researcher may not have any reason to believe that the initial condition is large or small. Such a situation could arise if the data is chosen to start after a perceived break in the series, or due to the instigation of a new market or institution. In these cases a desirable test would be one whose power is as little affected by the size of the unknown and essentially arbitrary initial condition as possible. This paper examines the possibility of obtaining unit root tests with relatively flat power as a function of the initial condition.

The need to either take a stand on the size of the initial condition or alternatively choose a test with good power over many possible values for the initial condition would disappear if this choice had no effect in practice. But in practice the decisions provided by most popular tests do depend strongly on the size of the initial condition. This means that different conclusions can be reached with samples of the same data that differ only in the date at which the observations begin. This effect of the initial condition can be illustrated using data on the real exchange rate. Fig. 1 shows the annual real exchange rate between the British pound and the US dollar from 1791 until 1990 (the data are from Lothian and Taylor (1996)).

What can be seen from the graph is that there are many 'unusual' values that can be identified ex post. For example, the peak around the time of the first world war, or similarly the very low value after the end of the second world war deviate substantially from the sample average. In the 19th century there are many such peaks and troughs. Supposing that the real exchange rate is mean reverting, an idea fully supported by economic theory, then power of typically applied unit root tests will be affected substantially if one chooses one of these unusual values as the starting date of the series.

[1]Descriptions of the magnitude of the initial condition here and in the following refer to the absolute value of the initial condition.

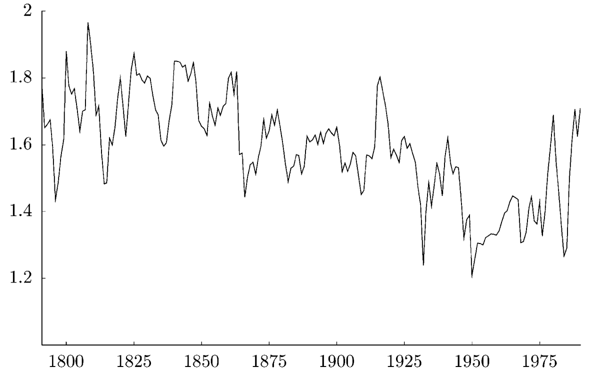

Fig. 1. Real exchange rate between the British pound and US dollar.

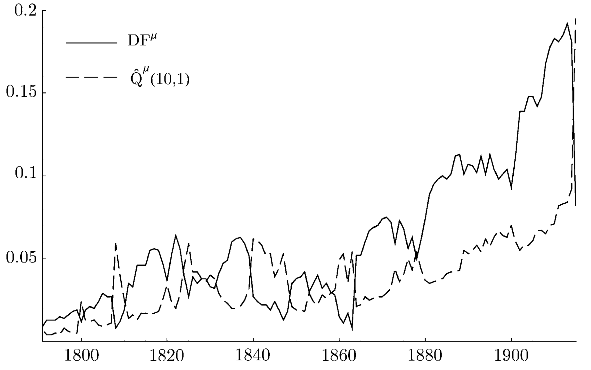

Fig. 2. p-values of unit root tests of the British pound/US dollar real exchange rates as a function of the starting date of the sample.

We can examine the effect of varying the starting date of the series by considering the outcome of two unit root tests that differ in their treatment of the initial condition for all start dates ranging from 1791 to 1915, while keeping the endpoint of the sample fixed at 1990. The last test is hence based on 76 observations. The effect of the changing start dates on the outcome of the unit root tests is shown by calculating the p-value for each sample. In Fig. 2, these p-values are plotted against the beginning date for the augmented Dickey and Fuller (1979) t-statistic test (denoted DF$^\mu$) and a modified (along the lines of Stock (2000)—this corresponds to the statistic $\hat{Q}^\mu(10,1)$ of ME, see Section 3) version of the Q_T statistic of Elliott (1999) (the dotted line).[2] Both tests tend to reject for the longer data series (earlier start dates), but there are vast periods of disagreement. For the first half of the sample we can see that the p-values, apart from their common upward drift, seem to

[2]The long-run variance estimator for this test was constructed as described in Section 4; the appropriate lag-length was determined once for the whole sample based on MAIC of Ng and Perron (2001).

be moving in opposite directions. Notice that around 1800, 1808 and 1825 there are abrupt shifts in the p-values of the two tests in opposite directions, where the p-value for the $\hat{Q}^{\mu}(10,1)$ statistic jumps up just as the p-value for the DF^{μ} test drops. These periods correspond precisely to local peaks in the exchange rate. Alternatively, the relatively more tranquil period around 1870 for the real exchange rate (no big peaks or troughs) is associated with the $\hat{Q}^{\mu}(10,1)$ statistic rejecting whereas DF^{μ} does not.

These periods are emblematic of the points we are raising in this paper. When initial values are chosen from a 'tranquil' period, so initial conditions for the series are close to zero, the $\hat{Q}^{\mu}(10,1)$ test is relatively efficient and has much better power than the DF^{μ} test. Thus it is much more likely to reject if the real exchange rate is mean reverting. However, when unusual values are employed at the start of the sample—dates where it is likely that the first observation is far from its deterministic component—the power ranking between the tests can be reversed. Indeed, most of the dramatic differences in the p-values in Fig. 2 directly correspond to periods where the real exchange rate changed suddenly by a relatively large margin. Overall, however, 'most' starting values tend to be close to the deterministic components. For this example, the DF^{μ} test rejects the relatively preposterous null of a unit root in the real exchange rate for half of the starting dates at the 5% level, whereas the $\hat{Q}^{\mu}(10,1)$ statistic rejects for about 84% of the samples.

The aim of this paper is to provide a careful analysis of the role of the initial condition in the unit root testing problem. To this end, the next section (i) explores in which circumstances one might be concerned about large initial conditions, (ii) shows that it is impossible to derive useful test statistics that do not depend numerically on the value of the initial condition and (iii) discusses methods that one could think of using or that have been suggested previously in the literature to lessen the impact of the initial condition on the power of unit root tests. Section 3 contains an asymptotic analysis. Building on the work of ME, we identify an efficient unit root test statistic whose power does not vary (much) as a function of the initial condition. Its local asymptotic power is then compared to an alternative approach based on partial invariance to the initial condition. Section 4 contains small sample Monte Carlo simulations of the size and power properties of these tests. In the Conclusion we return to the real exchange rate data and show that the test statistic proposed in Section 3 reduces the variation in p-values over the different starting values, suggesting that its outcome is less dependent on whether the starting point comes from an unusual or tranquil period. Proofs are collected in an appendix.

2. The dependence of unit root tests on the initial condition

In this paper, we consider the general model

$$
\begin{aligned}
y_t &= X_t'\beta + \mu + w_t \quad t = 0, 1, \ldots, T, \\
w_t &= \rho w_{t-1} + v_t \quad t = 1, \ldots, T, \\
w_0 &= \xi,
\end{aligned}
\tag{1}
$$

where X_t is a deterministic vector with no constant element, $X_0 = 0$ and μ, β and ξ are unknown and $\{v_t\}$ are mean zero disturbances that satisfy a functional central limit theorem (FCLT). We are interested in distinguishing the two hypotheses

$$\mathrm{H}_0 : \rho = 1 \quad \text{vs} \quad \mathrm{H}_1 : \rho < 1. \tag{2}$$

A large number of test statistics have been suggested for this testing problem. All of these tests share the property of invariance to the group of transformations

$$\{y_t\}_{t=0}^T \to \{y_t + X_t'b + m\}_{t=0}^T \quad \forall b, m. \tag{3}$$

Note that different values of β and μ will induce changes of the form (3), so that restricting attention to invariant tests solves the problem of lack of knowledge of μ and β. In addition, the invariance to translations ($\{y_t\} \to \{y_t + m\}$) renders unit root tests independent of the value of the initial condition ξ under the null hypothesis. In order to see this, note that alternative values of ξ in general induce changes of the data of the form $\{y_t\} \to \{y_t + x\rho^t\}$. So when $\rho = 1$, different values of ξ induce translations of the form $\{y_t\} \to \{y_t + x\}$, which are identical to the transformation (3) for $b = 0$. The numerical value, and hence size of translation invariant unit root tests does not depend on ξ.

This clearly cannot be said of the power of unit root tests. Under the alternative of $\rho < 1$, different values of ξ induce changes which in general will alter the value of the test statistics. In fact, as demonstrated by ME, the power of all popular unit root tests substantially varies with respect to ξ. Dickey and Fuller (1979) t-tests, for instance, have power under local alternatives that increases in $|\xi|$, whereas the efficient tests of Elliott et al. (1996) have power that drops to zero for moderately large $|\xi|$. In these descriptions, the magnitude of ξ is measured against values of ξ that one would expect if $\{w_t\}$ was a (second-order) stationary process under the alternative, i.e. where ξ stems from the unconditional distribution of a stationary process with largest autoregressive root $\rho < 1$.

One might argue that one can obtain good power for all values of the initial condition by making the choice of test dependent on an estimator of ξ. In particular, one might be tempted to try to obtain optimal inference for any ξ by 'plugging' the estimator $\hat{\xi}$ into an efficient test for known ξ. By the Neyman–Pearson lemma, a point-optimal test for known ξ against the alternative $\rho = r$ would be based on the difference in the log-likelihoods of a maximal invariant to (3) under the null and alternative hypothesis. Unfortunately, it is not possible to estimate ξ with sufficient precision for such a procedure to work, even asymptotically.

In order to see why, recall that efficient unit root tests for different (stochastic) assumptions on ξ have different power in the dimension of the initial condition. For concreteness, consider the P_T statistic of Elliott et al. (1996) and the Q_T statistic of Elliott (1999). By construction, P_T is an (asymptotically) point-optimal statistic for a very small initial condition, whereas Q_T is point-optimal for an initial condition that stems from the unconditional distribution of the stationary process under the point-alternative. Due to their efficiency, both tests are admissible tests by construction, that is there exists no test that dominates the power of either of the two over all values of the initial condition. Now clearly, if it was possible to estimate ξ accurately enough to yield as good inference as if ξ was known, then the resulting test would generate a power greater or equal to any test constructed for a specific stochastic assumption on the initial condition. This implies that its power could never be lower of either P_T and Q_T, for any given value of the initial condition. Since P_T and Q_T do not have the same asymptotic power properties, a test that has at least as much power than either of the two would dominate both. But this is impossible, so that no estimator of ξ with high enough precision can exist. While it is possible to exploit the information contained in estimators of the initial condition (cf. Harvey and Leybourne (2005), for instance), such an approach will not in general result in an efficient test.

In practice, then, the choice of unit root test has to be based on what is known about the initial condition ξ a priori. Knowledge about the initial condition is not as far-fetched as it might appear at first. If the process under investigation stems from an autoregressive process with stable largest autoregressive root (mean reverting or not) and has been running quite some time prior to the start of the sample, then in the mean reverting case, likely initial conditions will have a distribution that is well approximated by the unconditional distribution. Elliott's (1999) unit root tests are designed to be optimal for such instances. Often the beginning of the sample does not coincide with the beginning of the process under study, so that the assumption of the initial condition stemming from the unconditional distribution is quite plausible.

At the same time, this makes the alternative one of (second-order) stationarity, rather than one of mere mean reversion. Depending on the application, any lack of power of Elliott's (1999) unit root test against mean reverting processes with large initial conditions might be not such a bad thing, after all. Consider the case where a sample stems from a mean reverting process, and $|\xi|$ is large. If a unit root test 'correctly' rejects the null hypothesis of integration, a researcher might be tempted to use methods for stationary time series in his subsequent statistical analysis. But it might very well be that these methods perform poorly for large $|\xi|$, and that statistical procedures for integrated series would yield more adequate results. Unfortunately, little is known about the impact of large initial conditions in time series methods, so that it seems very difficult to decide what kind of behavior of unit root tests is desirable from this perspective.

In a considerable number of applications, however, the relevant alternative hypothesis is one of mean reversion, and there is little reason to believe the initial condition to be small in absolute value. If the beginning of the sample coincides with the start of the process itself or with an event with a profound impact on the series, starting values far off the equilibrium (if one exists) are quite plausible. Think of macroeconomic time series that were not collected during the second world war, or German data after the reunification in 1991. For such series, the inability of efficient tests such as those derived in Elliott et al. (1996) and Elliott (1999) to reject the null hypothesis of integration for large $|\xi|$ can be a major drawback.

This is even more true in empirical studies of whether series are 'convergent', which by definition implies a starting point substantially different from the eventual equilibrium.[3] The analysis of such a hypothesis via unit root tests is obviously only

[3]Contrast this to an empirical investigation of whether two series have already converged over a certain sample, which would be adequately carried out with Elliott's (1999) statistic.

sensible with tests that reject for mean reverting processes with large $|\xi|$—which leads Harvey and Bates (2003), for instance, to promote the Dickey–Fuller test for such applications.

But as demonstrated in ME and elsewhere, the power of the Dickey–Fuller test is quite considerably lower than other tests when the initial condition is small, no doubt to a large extent due to the fact that its power is increasing in $|\xi|$. What is being called for, then, is the construction of an efficient unit root test whose power as a function of the initial condition does not vary in the $|\xi|$ dimension. The efficiency of the test implies maximum possible power for small initial conditions under the restriction that the test's ability to reject for large $|\xi|$ is comparable. Such a test would be naturally suited for applications where the researcher does not want to rule out the possibility of a very large initial condition under the alternative, while sacrificing the least possible power in the case where the initial condition is small in absolute value.

Furthermore, such a test potentially eliminates the large impact of the (quite often somewhat arbitrary) choice of the start of the sample even in cases where strict stationarity is the more interesting alternative. A test that is insensitive to the magnitude of the initial condition would considerably lessen the scope for data mining along this dimension.

A number of approaches come to mind how one might attempt to construct a unit root test with good power for all ξ. One method consists of generalizing the invariance argument to the effect of the initial condition under the alternative. After all, under the alternative of $\rho<1$, ξ in model (1) might be thought of as just another parameter that describes the importance of the deterministic term $\{\rho^t\}$ in $\{y_t\}$. The relevant transformations are of the form

$$\{y_t\} \to \{y_t + xr^t\} \quad \forall x, r \in (0,1). \tag{4}$$

A test statistic that is invariant to (4) is by construction independent of ξ for all values of $\rho<1$, so that the parameter ξ has no impact whatsoever on the outcome (and hence power) of the test. The following theorem, however, proves that the only such test is the trivial test that does not depend on the data.

Theorem 1. *Any test of* (2) *which is invariant to the transformations* (4) *has trivial power.*

Intuitively, invariance to transformations of the data that remove any possible dependence on ξ for any ρ removes the entire variation of the data, leaving only tests that do not depend on the data. The result continues to hold even when the set of transformations (4) is reduced to transformations of the same form which restricts r to be in an arbitrarily small open set. The slightest uncertainty about the value of ρ under the alternative therefore makes it impossible to construct useful tests which numerically do not depend on ξ.

Dufour and King (1991) suggested a weaker form of invariance to the initial condition, which holds only for a particular value of $\rho = r < 1$:

$$\{y_t\} \to \{y_t + xr^t\} \quad \forall x. \tag{5}$$

A test that is invariant to both (3) and (5) will be invariant to the value of ξ only under the alternative $\rho = r$. For other values of $\rho < 1$, the test will still depend on ξ, so that the impact of ξ on the hypothesis test $\mathrm{H}_0 : \rho = 1$ vs $\mathrm{H}_1 : \rho < 1$ is only reduced in this restricted way.

In order to find an efficient test that is invariant to both (3) and (5), one can draw on the methods developed for the construction of efficient tests invariant to (3) only. The reason is simply that with r fixed, (5) can be cast as a special case of (3), where X_t is substituted by $X_t^R = (X_t', r^t)'$. Invariance to (5) can hence be imposed by adding an artificial regressor in model (1) with tth element r^t.

Specifically, following Dufour and King (1991), we consider the point-optimal invariant test that maximizes power against the alternative $\rho = r$ for Gaussian disturbances $\{v_t\}$ with known covariance–matrix.[4] Proceeding as in ME shows that the point-optimal invariant test statistic P_{inv} (invariant to (3) and (5)) is given by the difference in weighted squared residuals of a GLS regression of $\{y_t\}$ on $\{(1, X_t', r^t)\}$, where the GLS weighting corresponds to the variance–covariance matrices of $\{w_t\}$ under the null hypothesis of $\rho = 1$ and the alternative hypothesis of $\rho = r$. The construction of this test requires knowledge of the variance–covariance matrix of $\{v_t\}$, which is typically not available in practice. But in the next section, drawing on the results of Elliott et al. (1996) and ME, we show that under weak assumptions it is possible to construct a test statistic that is asymptotically as powerful as the small sample optimal statistic and that does not require such knowledge.

It is instructive to explore the form of P_{inv} for the special case where $\{v_t\}$ is distributed i.i.d. standard normal and there is no X_t in model (1)—such an assumption on the deterministics corresponds to the standard 'mean only' case in the unit root literature. Let $\{u_t\} = \{y_t - y_0\} = \{w_t - \xi\}$, a maximal invariant to the group of translations $\{y_t\} \to \{y_t + m\} \quad \forall m$. After some tedious algebra, one finds

$$P_{\text{inv}} = \sum_{t=1}^{T} [2(1-r)\Delta u_t u_{t-1} + (1-r)^2 u_{t-1}^2] - T^{-1}\left[u_T + (1-r)\sum_{t=1}^{T} u_{t-1}\right]^2 + \left[r^T u_T + (1-r)\sum_{t=1}^{T} r^{t-1} u_{t-1}\right]^2 \Big/ \sum_{t=1}^{T} r^{2t}. \tag{6}$$

P_{inv} hence depends on the data through five statistics, $\sum_{t=1}^{T} \Delta u_t u_{t-1}$, $\sum_{t=1}^{T} u_{t-1}^2$, u_T, $\sum_{t=1}^{T} u_{t-1}$ and $\sum_{t=1}^{T} r^t u_{t-1}$. At the same time, the log-likelihood of the maximal invariant $\{u_t\}$ as a function of ξ is, omitting constants, given by

$$l(u|\rho, \xi) = -\frac{1}{2}\sum_{t=1}^{T} [\Delta u_t^2 + 2(1-\rho)\Delta u_t u_{t-1} + (1-\rho)^2 u_{t-1}^2 + 2(1-\rho)\xi \Delta u_t + 2(1-\rho)^2 \xi u_{t-1} + (1-\rho)^2 \xi^2]. \tag{7}$$

[4] Dufour and King (1991) consider the slightly larger group of transformations that yield an additional invariance to scale. Their test statistic hence only requires knowledge of the variance–covariance matrix of $\{v_t\}$ up to a scalar multiple.

By the factorization theorem, one set of sufficient statistics hence consists of the four statistics $\sum_{t=1}^{T} \Delta u_t u_{t-1}, \sum_{t=1}^{T} u_{t-1}^2, u_T$ and $\sum_{t=1}^{T} u_{t-1}$. The optimal statistic P_{inv} that is invariant to ξ under the alternative $\rho = r$ therefore depends on the data not only through the sufficient statistics, but additionally through $\sum_{t=1}^{T} r^{t-1} u_{t-1}$. The dependence on $\sum_{t=1}^{T} r^{t-1} u_{t-1}$ arises through the inclusion of the 'artificial' regressor $\{r^t\}$, which has no counterpart in the data generating process under the null hypothesis. The imposition of invariance to the initial condition even only under a single alternative $\rho = r$ might hence come at the cost of yielding an inadmissible test, a possibility we further investigate in Section 3.

A second approach that might seem to limit the influence of the initial condition on power is to 'condition' on the first observation, y_0, when constructing likelihood based tests. The Dickey and Fuller tests, for instance, are based on OLS regressions of $\{y_t\}$ on $\{(y_{t-1}, 1, X_t')'\}$. Standard arguments show that the coefficient estimator of ρ of such a regression is equivalent to the conditional maximum likelihood estimator of ρ assuming Gaussian i.i.d. disturbances $\{v_t\}$, where the contribution of y_0 to the likelihood is omitted. The difficulty that arises here is that the initial condition is not equal to the initial observation y_0, but it is the difference between y_0 and the deterministic component at $t = 0$: $\xi = y_0 - \mu - X_0'\beta$. When the deterministic part of the model is known, y_0 is sufficient for the initial condition ξ and conditioning on y_0 would indeed be the same as conditioning on ξ. However, when μ and β are unknown this is no longer true, and ξ is unobserved. In fact, without knowledge of the deterministic part of the model, knowledge of y_0 alone has no implication whatsoever for the value of ξ, as $y_0 = \xi + \mu + X_0'\beta$. Therefore, the conditioning approach does not limit the effect of the initial condition by construction.

Finally, one might hope that the approach of plugging an estimator of ξ into the efficient test statistic for known ξ limits the impact of ξ on the power of the test. While the discussion above demonstrates that such an approach cannot yield as good inference as if ξ was known, even asymptotically, one might still think that power is at least reasonable for all values of ξ. Intuitively, the estimation of the initial condition should give the data more flexibility under the alternative, so that no value of ξ prevents the test from rejecting. When ξ is estimated by maximum likelihood under the fixed alternative $\rho = r$ (which amounts to finding the value of ξ that maximizes (7) with $\rho = r$), and the estimator is plugged into the efficient test statistic for known ξ (which is simply the difference of (7) evaluated at $\rho = r$ and $\rho = 1$), one obtains the statistic

$$-\frac{1}{2}\sum_{t=1}^{T} [2(1-r)\Delta u_t u_{t-1} + (1-r)^2 u_{t-1}^2] + \frac{1}{2} T^{-1} \left[u_T + (1-r) \sum_{t=1}^{T} u_{t-1} \right]^2 \tag{8}$$

for the mean case model with i.i.d. Gaussian disturbances. Interestingly, some straightforward algebra reveals that this test statistic is a (limiting) member of the family of efficient tests derived by ME: it corresponds to $Q^{\mu}(r, \infty)$ in ME's notation. But contrary to the intuition alluded to above, the power of this test varies drastically as a function of ξ—see Section 4. Specifically, power is very low for small and moderate values of $|\xi|$ compared to other tests, but quickly approaches unity as $|\xi|$ becomes large.

3. Asymptotic analysis

3.1. An efficient test with good power for arbitrary initial conditions

It is hence quite challenging to derive a unit root test whose power does not vary with the initial condition by construction. While the partial invariance approach of Dufour and King (1991) necessarily leads to constant power as a function of the initial condition at least for one alternative, this might very well come at the cost of inadmissibility of the test. An alternative way to approach this problem is to start out with a family of efficient tests, where different members of the family have different power properties in the dimension of the initial condition, and to determine a member of the efficient family that has close to equal power for various initial conditions.

The family of unit root tests derived by ME are perfectly suited for that enterprise. ME derive tests that maximize a weighted average power criterion, where the weighting is over the power of the unit root tests against a specific alternative for various values of the initial condition. Members of the family differ in their weighting function for the initial condition. In this section, we will analytically determine a specific member of this family (i.e. a specific weighting function) that leads to roughly constant power over a wide range of initial conditions. Being a member of the optimal family, the resulting test is guaranteed to be admissible. The comparison between this particular test and the optimal invariant test in the spirit of Dufour and King (1991) reveals that the former dominates the latter.

Since in practice, i.i.d. disturbances are highly implausible, we follow Elliott et al. (1996) and ME and consider the data generating process (1) with potentially correlated Gaussian disturbances $\{v_t\}$.

Condition 1. *The stationary sequence $\{v_t\}$ has a strictly positive spectral density function $f_v(\lambda)$; it has a moving average representation $v_t = \sum_{s=0}^{\infty} \delta_s \varepsilon_{t-s}$ where the ε_t are independent standard normal random variables and $\sum_{s=0}^{\infty} s|\delta_s| < \infty$.*

The correlation structure of $\{v_t\}$ is typically not known in practice, so the question arises how to implement optimal tests—based on the Neyman–Pearson Lemma—under this condition. But it turns out that it is possible to derive feasible unit root tests without such knowledge that have the same asymptotic power as the infeasible tests that exploit the exact correlation structure of $\{v_t\}$.

Asymptotic power for the hypothesis test (2) is nontrivial only in a neighborhood of the null hypothesis of $\rho = 1$ of the form where $\rho = 1 - \gamma/T$, where $\gamma > 0$ is fixed. This is the local-to-unity framework developed by Phillips (1987) and Chan and Wei (1987). Under such asymptotics, it is reasonable to measure the initial condition $|\xi|$ in multiples of the square root of the unconditional variance of a stationary process for $\gamma > 0$, which is equal to $\omega(2\gamma)^{-1/2}T^{1/2} + o(T^{1/2})$, where $\omega^2 = 2\pi f_v(0)$. We hence denote $\alpha = (2\gamma)^{1/2}\omega^{-1}T^{-1/2}\xi$. In this notation, we find for $\{u_t\} = \{w_t - \xi\}$ (cf. Elliott (1999))

$$T^{-1/2}u_{[Ts]} \Rightarrow \begin{cases} \omega W(s) & \text{for } \gamma = 0 \\ \omega\alpha(\mathrm{e}^{-\gamma s} - 1)(2\gamma)^{-1/2} + \omega \int_0^s \mathrm{e}^{-\gamma(s-\lambda)}\,\mathrm{d}W(\lambda) & \text{else} \end{cases} \tag{9}$$

$$\equiv \omega M(s), \tag{10}$$

where '$\Rightarrow$' denotes weak convergence of the underlying probability measures, $W(s)$ is a standard Brownian motion and $[\cdot]$ indicates the greatest lesser integer function. The deterministic part of $M(s)$, $\alpha(\mathrm{e}^{-\gamma s} - 1)(2\gamma)^{-1/2}$ for $\gamma < 0$, captures the effect of the initial condition.

ME derive small sample optimal tests that maximize the weighted average power criterion

$$\int P(\varphi(y) \text{ rejects} \,|\, \rho = r, \alpha = a)\,\mathrm{d}F(a) \tag{11}$$

over all tests $\varphi(y)$ that satisfy the level constraint. Such a test has the largest average power under the alternative $\rho = r < 1$, where the averaging over possible initial conditions α is according to F. Specifically, they employ a Gaussian weighting for different values of the initial condition α, so that F in (11) is the cumulative distribution function of a $\mathrm{N}(0, k)$ random variable. The variance of this Gaussian weighting function determines the relative weight of small and large initial conditions: the larger the variance, the larger the relative weight of large $|\alpha|$.

Given our interest in finding feasible, asymptotically equivalent tests, it is natural to set $r = 1 - g/T$ for some fixed g, in accordance with the asymptotic thought experiment concerning ρ under local-to-unity asymptotics. Only data generating processes with ρ such that $\gamma = T(1-\rho)$ is a moderate number between zero and, say, 50 lead to a challenging testing problem. When γ is larger than 50, ρ is so far from unity that reasonable tests will reject with probability very close to one. A good test should hence maximize its power in this region, and we follow Elliott (1999) by choosing $g = 10$ in the mean case and $g = 15$ in the trend case, choices that lead to optimal power at alternatives where 5% tests have power roughly equal to one half.

ME derive the asymptotic distribution of small sample optimal invariant test statistics $Q^i(g,k)$ (derived under Condition 1 with knowledge of the correlation structure of $\{v_t\}$ as described by $\{\delta_t\}$), where $i = \mu, \tau$ stands for the mean and trend cases, respectively. The invariance of $Q^i(g,k)$ refers to the group of transformations (3) only. A test that rejects for small values of $Q^i(g,k)$ maximizes weighted average power against the alternative $\rho = 1 - g/T$, where the averaging over the initial conditions in (11) is carried out according to the cumulative distribution function of a zero mean Gaussian variate with a variance that is k times the variance of the unconditional distribution of a stationary AR(1) process with $\rho = 1 - g/T$. The statistics P_T of Elliott et al. (1996) and Q_T of Elliott (1999) are therefore asymptotically equivalent to $Q^i(g,0)$ and $Q^i(g,1)$, respectively. Let $M^\mu(s) = M(s) - \int M(u)\,\mathrm{d}u$ and $M^\tau(s) = M^\mu(s) - 12(s - \frac{1}{2})\int uM^\mu(u)\,\mathrm{d}u$, where integrals here and in the following are understood to have delimiters zero and one, if not indicated otherwise. The processes $M^\mu(s)$ and $M^\tau(s)$ are hence continuous time least squares projections of $M(s)$ off a constant and off $(1, s)$, respectively. In this representation, Theorem 3 of ME implies that

$$Q^i(g,k) \Rightarrow q_0^i + q_1^i M^i(0)^2 + q_2^i M^i(1)^2 + q_3^i M^i(0)M^i(1) + q_4^i \int M^i(s)^2\,\mathrm{d}s, \qquad (12)$$

where, $q_0^\mu = -g$, $q_1^\mu = -g(1+g)(k-2)/(2+gk)$, $q_2^\mu = (2g - gk + g^2k)/(2+gk)$, $q_3^\mu = 2g(k-2)/(2+gk)$, $q_4^\mu = g^2$ and $q_0^\tau = -g$, $q_1^\tau = (8g^2 + g^3(8-3k) - g^4(k-2))/(24 + 24g + 8g^2 + g^3k)$, $q_2^\tau = (8g^2 + g^3(8-3k) + g^4k)/(24+24g+8g^2+g^3k)$, $q_3^\tau = (8g^2 + 2g^3(4-3k))/(24+24g+8g^2+g^3k)$, $q_4^\tau = g^2$.

There does not seem to be a way of analytically investigating the power of unit root tests based on $Q^i(g,k)$ as a function of α. A tractable question, however, is how these tests behave under local alternatives for very large initial conditions. In order to see this, recall that $M(s)$ can be decomposed in a sum of two elements: $\alpha(\mathrm{e}^{-\gamma s} - 1)(2\gamma)^{-1/2}$ and $\int_0^s \mathrm{e}^{-\gamma(s-\lambda)}\,\mathrm{d}W(\lambda)$. As α becomes large, the deterministic part dominates the overall shape of the process $M(\cdot)$. Typically, the deterministic part of the asymptotic distribution of $Q^i(g,k)$ diverges to either $+\infty$ or $-\infty$ as $|\alpha|$ grows without bound under the fixed alternative $\gamma = g$. Given that the critical value of a test based on $Q^i(g,k)$ is finite for any level, this implies that the power as a function of $|\alpha|$ converges to either zero or unity for large enough $|\alpha|$. The power of such tests hence necessarily varies substantially for some $|\alpha|$. In an attempt to limit the influence of the initial condition on the power of the test, one might hence try to pick $k = k_*$ in a judicious way for this not to happen. A test based on $Q^i(g,k_*)$ for such a k_* has asymptotic power under the local alternative $\gamma = g$ which does not converge to either zero or one even when $|\alpha|$ is huge.

Preventing $Q^i(g,k)$ from diverging under $\gamma = g$ as $|\alpha| \to \infty$ requires the deterministic element $K(s) = \alpha(\mathrm{e}^{-gs} - 1)(2g)^{-1/2}$ of $M(s)$ to cancel in the asymptotic distribution. In other words, when $M^i(s)$ is replaced by $K^i(s)$ with $\gamma = g$ in (12), the weighted sum must equal zero, where $K^\mu(s) = K(s) - \int K(u)\,\mathrm{d}u$ and $K^\tau(s)$ is the residual of a continuous time regression of $K(s)$ on $(1, s)$. The reason is simply that if this weighted sum is positive (negative), the asymptotic distribution of $Q^i(g,k)/\alpha^2$ converges to a positive (negative) constant as $\alpha^2 \to \infty$, which in turn implies $Q^i(g,k)$ to diverge as $|\alpha| \to \infty$.

After a considerable amount of tedious algebra, one finds that the appropriate choice of $k = k_*^\mu$ in the mean case is given by

$$k_*^\mu = \frac{4g - 2 + 2\mathrm{e}^{-2g}}{(1 - \mathrm{e}^{-2g})g} \qquad (13)$$

and in the trend case

$$k_*^\tau = \frac{2}{\mathrm{e}^g - 1} - \frac{2(2+g)^2}{\mathrm{e}^g(g-2)^2 + g^2 - 4} + \frac{48 + 24g - 8g^2 - 8g^3 + 4g^4}{g^3(g-2)} \qquad (14)$$

which evaluates to $k_*^\mu = \frac{38}{10}$ for $g = 10$ and $k_*^\tau = 3.968\ldots$ for $g = 15$. See the appendix for development.

$Q^i(g,k)$ is a small sample optimal statistic for known covariance matrix $\mathrm{E}[vv']$. Its asymptotic distribution (12) under Condition 1, however, does not depend on $\mathrm{E}[vv']$

in any particular form. This makes it possible to construct test statistics which share the same asymptotic distribution (and hence asymptotic power) without knowledge of E[vv']—see Elliott et al. (1996). One possible choice is

$$\hat{Q}^i(g,k) = q_0^i + q_1^i(T^{-1/2}\hat{y}_0^i)^2 + q_2^i(T^{-1/2}\hat{y}_T^i)^2 + q_3^i T^{-1}\hat{y}_0^i\hat{y}_T^i + q_4^i T^{-2}\sum_{t=1}^{T}(\hat{y}_{t-1}^i)^2 \tag{15}$$

with $i = \mu, \tau$. In these expressions, $\hat{y}_t^\mu = \hat{\omega}^{-1}(y_t - T^{-1}\sum_{s=0}^{T} y_s)$ with $\hat{\omega}^2$ some consistent estimator of the long-run variance of $\{v_t\}$, and $\{\hat{y}_t^\tau\}$ are the residuals of a regression of $\{\hat{y}_t^\mu\}_{t=0}^T$ on $\{1, t\}$.

Application of the FCLT and continuous mapping theorem (CMT) yields that the asymptotic distribution of $\hat{Q}^i(g,k)$ for all $\gamma \geqslant 0$ is as given in (12), such that the asymptotic (local) power of tests based on $\hat{Q}^i(g,k)$, and in particular $\hat{Q}^i(g,k_*^i)$, is the same as the asymptotic power of tests based on the efficient but typically unfeasible tests $Q^i(g,k)$ and $Q^i(g,k_*^i)$. In fact, $\hat{Q}^i(g,k)$ follows the asymptotic distribution (12) under much more general assumptions than Condition 1: whenever $\{v_t\}$ satisfies a FCLT and $\hat{\omega}$ is a consistent estimator of the scale of the limiting Wiener process, the asymptotic distribution and asymptotic power of the statistic $\hat{Q}^i(g,k)$ are equivalent to those obtained under Condition 1. While the normality of the disturbances is crucial for the optimality claim, the validity and asymptotic behavior of tests based on $\hat{Q}^i(g,k)$ is therefore as described by (12) under very general conditions on the disturbances.

3.2. Asymptotic properties of P_{inv}

It is interesting to compare the performance of a test based on $Q^i(g,k_*^i)$ with the approach of Dufour and King (1991) that imposes invariance to the initial condition under $\rho = r = 1 - g/T$. To this end we extend the results of ME by deriving the asymptotic distribution of the small sample optimal invariant Dufour and King statistic $P_{\mathrm{inv}}^i(g)$, generalized to a Condition 1 data generating processes. A test based on $P_{\mathrm{inv}}^i(g)$ maximizes the power under the alternative $\rho = r = 1 - g/T$ under the restriction of invariance of the test to transformations of data of the form (3) and (4). Since the group of transformations (4) corresponds to all possible initial conditions for $\rho = r$, the test is numerically independent of the true value of the initial condition under the alternative $\rho = r = 1 - g/T$. Any weighting over ξ as in (11) does hence generate the same statistic.

From the reasoning in Dufour and King (1991) and ME, under Condition 1 P_{inv} is given by

$$P_{\mathrm{inv}}^i(g) = y'M_Z[M_Z\Sigma_1 M_Z]^- M_Z y - y'M_Z[M_Z\Sigma_0 M_Z]^- M_Z y, \tag{16}$$

where the tth row of the $T+1$ rows of Z is given by $(1, r^t)$ for $i = \mu$ and by $(1, t, r^t)$ for $i = \tau$, $M_Z = I_{T+1} - Z(Z'Z)^{-1}Z'$, $\Sigma_1 = \mathrm{E}[uu']$ for $\rho = r = 1 - g/T$ and $\xi = 0$ and $\Sigma_0 = \mathrm{E}[uu']$ for $\rho = 1$ and $\xi = 0$, u is the $(T+1) \times 1$ vector with tth element $w_t - \xi$

and $[\cdot]^-$ is any generalized inverse. The following theorem establishes the asymptotic distribution of $P^i_{\text{inv}}(g)$.

Theorem 2. *Let $M^{i_R}(s)$ be the continuous time projection of $M^i(s)$ off $(1-e^{-gs})$. Under Condition 1, for $i=\mu,\tau$*

$$P^i_{\text{inv}}(g) \Rightarrow \kappa^i_0 + \kappa^i_1 M^{i_R}(0)^2 + \kappa^i_2 M^{i_R}(1)^2 + \kappa^i_3 M^{i_R}(0)M^{i_R}(1) + \kappa^i_4 \int M^{i_R}(s)^2\, \mathrm{d}s, \tag{17}$$

where $\kappa^\mu_0 = -g$, $\kappa^\mu_1 = \kappa^\mu_2 = (1+g+e^{2g}(g-1))/(e^{2g}-1)$, $\kappa^\mu_3 = (2e^{2g}-2-4ge^g)/(e^{2g}-1)$, $\kappa^\mu_4 = g^2$ and $\kappa^\tau_0 = -g$, $\kappa^\tau_1 = \kappa^\tau_2 = (8-16e^g+5g+g^2+e^{2g}(8-5g+g^2))/\lambda^\tau$, $\kappa^\tau_3 = (8-2e^{2g}(g-4)+2g-4e^g(4+g^2))/\lambda^\tau$, $\kappa^\tau_4 = g^2$ with $\lambda^\tau = (e^g-1)(2+g+e^g(g-2))$.

It is interesting to note that the weights κ^i_1 and κ^i_2 on $M^{i_R}(0)^2$ and $M^{i_R}(1)^2$ coincide in (17). This simply reflects that with the invariance restriction (4), the efficient test for unit roots is invariant to reversing the time scale of the observations. The only other way to achieve this reversibility is to make the data strictly stationary under the alternative—as is reflected in the equality of the weights q^i_i and q^i_2 in (17) when $k=1$.

Theorem 2 shows that the asymptotic distribution of $P^i_{\text{inv}}(g)$ does not depend on the correlation structure of v. Just as for $Q^i(g,k)$, one can hence construct a feasible statistic $\hat{P}^i_{\text{inv}}(g)$ with the same asymptotic distribution and hence power by taking a weighted sum of the small sample analogs to the integrals of M^{i_R}. Details are omitted for brevity.

3.3. Local asymptotic power

We now turn to a numerical description of the asymptotic power of tests based on $Q^i(g,k)$ and $P^i_{\text{inv}}(g)$ (or, equivalently, of $\hat{Q}^i(g,k)$ and $\hat{P}^i_{\text{inv}}(g)$). Rather than basing this comparison on Monte Carlo results, we follow Nabeya and Tanaka (1990) and Tanaka (1996) and numerically invert the characteristic functions of the asymptotic distributions (12) and (17). The following Lemma establishes the characteristic function of a class of random variables general enough for our purposes.

Lemma 1. *In the notation of (10), let $H = (M(1), \int M(s)\,\mathrm{d}s, \int g_3(s)\,\mathrm{d}W(s), \ldots, \int g_\ell(s)\,\mathrm{d}W(s))'$ with $g_j(\cdot)$, $j=3,\ldots,\ell$, square integrable functions on the unit interval (possibly dependent on γ) and define $V(\gamma) = \mathrm{E}[HH']$. Let $\Psi = l_0 + l_1 \int M(s)^2\,\mathrm{d}s + H'\Lambda H + \lambda' H$, where the matrix Λ, the vector λ, l_0 and l_1 are nonstochastic. Then the characteristic function of Ψ is given by*

$$\phi(\theta) = |I_\ell - 2V(\delta)\tilde{\Lambda}|^{-1/2} \exp\left[\tilde{l}_0 + \tfrac{1}{2}\tilde{\lambda}'(V(\delta)^{-1} - 2\tilde{\Lambda})^{-1}\tilde{\lambda} - \tfrac{1}{4}\alpha^2\gamma\right], \tag{18}$$

where $\delta = \sqrt{\gamma^2 - 2l_1\theta\mathbf{i}}$, $\tilde{l}_0 = \theta l_0 \mathbf{i} - \frac{1}{2}(\delta-\gamma)$, $\tilde{\Lambda} = \theta\Lambda\mathbf{i} + \mathrm{diag}((\delta-\gamma)/2, 0I_{\ell-1})$, $\tilde{\lambda} = \theta\lambda\mathbf{i} - \alpha\sqrt{2}/2(\gamma^{1/2}, \gamma^{3/2}, O'_{\ell-2})'$ and $\mathbf{i} = \sqrt{-1}$ and $O_{\ell-2}$ is a $(\ell-2)\times 1$ vector of zeros. Furthermore, the ij element of $V(\gamma)$ is given by $\int g_i(s)g_j(s)\,\mathrm{d}s$, where $g_1(s) = \exp[-\gamma(1-s)]$ and $g_2(s) = \{1-\exp[-\gamma(1-s)]\}/\gamma$.

Table 1 contains the asymptotic critical values of the unit root tests considered in this paper, and Fig. 3 depicts the local asymptotic power for tests with size 5% based

Table 1
Asymptotic critical values

Test	Critical value		
	1%	5%	10%
$\hat{Q}^{\mu}(10,1)$	−6.94	−5.34	−4.06
$\hat{Q}^{\mu}(10,3.8)$	−7.70	−6.40	−5.37
$\hat{Q}^{\mu}(10,\infty)$	−10.01	−7.58	−6.46
$\hat{P}^{\mu}_{\mathrm{inv}}(10)$	−7.520	−6.34	−5.40
$\hat{Q}^{\tau}(15,1)$	−10.53	−8.85	−7.61
$\hat{Q}^{\tau}(15,3.968)$	−11.24	−9.77	−8.70
$\hat{Q}^{\tau}(15,\infty)$	−12.97	−11.44	−10.09
$\hat{P}^{\tau}_{\mathrm{inv}}(15)$	−11.11	−9.73	−8.73

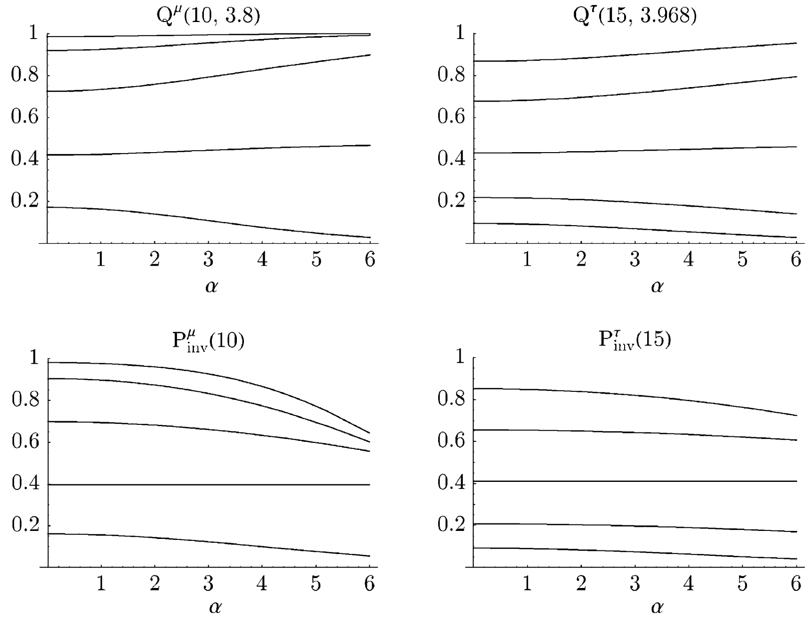

Fig. 3. Asymptotic power for $\gamma = 5, 10, 15, 20, 25$ as a function of the initial condition α.

on $Q^{\mu}(10,3.8)$, $Q^{\tau}(15,3.968)$, $P^{\mu}_{\mathrm{inv}}(10)$ and $P^{\tau}_{\mathrm{inv}}(15)$ for $\gamma = 5, 10, 15, 20, 25$ as a function of α. Curves further from the horizontal axis correspond to more distant alternatives. The local asymptotic power of both tests remains relatively flat as a function of α, especially for alternatives against which power is approximately 50%. The numerical independence of $P^{\mu}_{\mathrm{inv}}(10)$ and $P^{\tau}_{\mathrm{inv}}(15)$ of the value of α at the alternatives $\gamma = 10$ and $\gamma = 15$, respectively, implies constant power for all values of

α. But the tests based on $Q^{\mu}(10, 3.8)$ and $Q^{\tau}(15, 3.968)$ also have power very close to constant against these alternatives, although their construction only implies that their power does not converge to zero or one as α becomes large. Interestingly, the power of $Q^{\mu}(10, 3.8)$ and $Q^{\tau}(15, 3.968)$ is uniformly higher than the power of $P^{\mu}_{\text{inv}}(10)$ and $P^{\tau}_{\text{inv}}(15)$ at the alternatives $\gamma = 10$ and $\gamma = 15$, respectively, i.e. at the alternative against which these tests have optimality properties. Numerical results not reported here show that this remains true for any α. Tests based on $P^{\mu}_{\text{inv}}(10)$ and $P^{\tau}_{\text{inv}}(15)$ are hence inadmissible for testing the point alternative they were constructed for. This numerical result validates the intuition developed in Section 2 which suggested that the dependence of $P^{i}_{\text{inv}}(g)$ on the data other than through sufficient statistics might lead to an inadmissible test. In contrast, the construction of $Q^{i}(g, k^{i}_{*})$ renders tests on this statistic necessarily admissible, i.e. no unit root test can have higher power over all alternatives and initial conditions.

Even taking other alternatives into account, the overall asymptotic power properties of $Q^{i}(g, k^{i}_{*})$ are almost everywhere superior to those of $P^{i}_{\text{inv}}(g)$. The approximately flat power profile of $Q^{i}(g, k^{i}_{*})$ makes tests on this statistic potentially attractive in instances where a large initial condition cannot be ruled out a priori.

4. Monte Carlo evidence

The asymptotically efficient test based on $\hat{Q}^{i}(g, k^{i}_{*})$ proposed in the previous section has the property that its power converges to neither zero nor one as the initial condition gets large for the alternative $T(1-\rho) = \gamma = g$. In this section we examine the small sample properties of this test statistic, showing size control properties as well as examining what happens to power when the sample size is not large.

In addition, we will examine some other tests that have been used or suggested for testing for a unit root. We examine the Dickey and Fuller (1979) t-statistic, denoted DF^{i}, augmented with lagged dependent variables to account for serial correlation. For comparison purposes we include $\hat{Q}^{\mu}(10, 1)$ and $\hat{Q}^{\tau}(15, 1)$, which are asymptotically efficient against the strictly stationary alternative, i.e. where the initial stems from the unconditional distribution under the point alternative. We also consider the small sample performance of two other tests: The $\hat{Q}^{i}(g, \infty)$ statistic that—as explained in Section 2—arises if the maximum likelihood estimator of the initial condition is plugged into the efficient unit root statistic for a known initial condition, and the $\hat{P}^{i}_{\text{inv}}(g)$ statistic derived from a test that is invariant to data transformations $\{y_t\} \to \{y_t + xr^t\} \quad \forall x$, where $r = 1 - g/T$. The power of a test based on $\hat{P}^{i}_{\text{inv}}(g)$ is therefore independent of the value of the initial condition under the alternative of $\rho = r$.

We consider the model in (1) where we set μ and β to zero without loss of generality, since all tests are numerically independent of the value of these parameters. The model is generated with

$$(1 - \phi L)v_t = (1 + \theta L)\varepsilon_t \quad t = 1, \ldots, T, \tag{19}$$

Table 2
Size of various 5% level unit root tests

$\phi =$	0	0	0.3	−0.3	0	0
$\theta =$	0	0	0	0	0.3	−0.3
DF^{μ}	0.056	0.050	0.044	0.056	0.044	0.064
$\hat{Q}^{\mu}(10,1)$	0.039	0.036	0.054	0.063	0.055	0.074
$\hat{Q}^{\mu}(10,\infty)$	0.039	0.038	0.050	0.050	0.048	0.056
$\hat{Q}^{\mu}(10,k_*^{\mu})$	0.034	0.032	0.052	0.058	0.052	0.068
$\hat{P}_{\text{inv}}^{\mu}(10)$	0.036	0.033	0.054	0.058	0.054	0.070
DF^{τ}	0.055	0.044	0.030	0.054	0.029	0.065
$\hat{Q}^{\tau}(15,1)$	0.026	0.026	0.053	0.078	0.050	0.097
$\hat{Q}^{\tau}(15,\infty)$	0.019	0.020	0.030	0.038	0.031	0.046
$\hat{Q}^{\tau}(15,k_*^{\tau})$	0.022	0.022	0.048	0.067	0.045	0.086
$\hat{P}_{\text{inv}}^{\tau}(15)$	0.023	0.022	0.052	0.067	0.049	0.087

Based on 10 000 Monte Carlo replications with $p_{\max} = 0$ for the first column and $p_{\max} = 4$ for the remaining columns.

where $\varepsilon_t \sim$ i.i.d. $N(0,1)$ and v_0 is drawn from the unconditional stationary distribution. For $\rho < 1$ ($\gamma > 0$) the initial condition is chosen as $\xi = T^{1/2}\alpha\omega(2\gamma)^{-1/2}$ where $\omega = (1+\theta)/(1-\phi)$, such that α measures the magnitude of the initial condition in (approximate) standard deviations of the stationary process $\{w_t\}$. The data is for observations 0 through T, and we consider 5% level tests throughout.

Size results are presented in Table 2 for various parameterizations of the serial correlation in v_t for $T = 100$. The serial correlation was accounted for using slightly different methods for the Dickey–Fuller test compared to the other tests. We have endeavoured to keep the tests on an equal footing here. In particular, we use the MAIC method of Ng and Perron (2001) to choose a lag length that is used for all tests in adjusting for serial correlation. The maximum lag length considered by the method is denoted by $p_{\max}$. The augmented Dickey–Fuller statistic DF^i adds lagged first differences in y_t where the number of lags chosen is as determined by MAIC. The test statistics which require a 'plug-in' estimator of $\hat{\omega}^2$ use the autoregressive estimator for this parameter based on the regression used to compute the DF^i test.[5]

Size is well controlled across all of the considered models when only a nonzero mean is allowed for. The tests tend to be a little oversized when there is negative serial correlation or a negative moving average component in the residuals. For the tests invariant to a constant and a time trend, the size distortions are qualitatively similar but somewhat more pronounced. Such results are typical for unit root tests.

We present size adjusted power for comparison with the asymptotic results examined in Section 3 in Tables 3 and 4, where $\{w_t\}$ is an $AR(1)$ with independent

[5] The ADF regression is $\Delta y_t^i = a y_{t-1}^i + \sum_{j=1}^{p} b_j \Delta y_{t-j}^i + e_t$ and the estimator $\hat{\omega}^2 = \hat{\sigma}^2/(1-\sum_{j=1}^{p} b_j)^2$ with $\hat{\sigma}^2 = T^{-1}\sum \hat{e}_t^2$. Here y_t^i is the detrended data where the detrending method is by OLS. For $P_{\text{inv}}^i(g)$ the detrending includes the additional regressor described in Section 2.

Table 3
Size corrected small sample power in the mean case

Test	α						
	0	1	2	3	4	5	6
$\rho = 0.95$							
DF^{μ}	0.11	0.12	0.15	0.21	0.30	0.44	0.61
$\hat{Q}^{\mu}(10,1)$	0.27	0.16	0.03	0.00	0.00	0.00	0.00
$\hat{Q}^{\mu}(10,\infty)$	0.05	0.09	0.19	0.29	0.38	0.43	0.48
$\hat{Q}^{\mu}(10,k_*^{\mu})$	0.17	0.16	0.13	0.09	0.06	0.03	0.02
$\hat{P}^{\mu}_{\mathrm{inv}}(10)$	0.16	0.15	0.14	0.12	0.10	0.08	0.06
$\rho = 0.90$							
DF^{μ}	0.30	0.32	0.39	0.51	0.67	0.83	0.94
$\hat{Q}^{\mu}(10,1)$	0.64	0.51	0.16	0.01	0.00	0.00	0.00
$\hat{Q}^{\mu}(10,\infty)$	0.14	0.21	0.40	0.62	0.80	0.91	0.96
$\hat{Q}^{\mu}(10,k_*^{\mu})$	0.43	0.43	0.43	0.43	0.42	0.42	0.41
$\hat{P}^{\mu}_{\mathrm{inv}}(10)$	0.39	0.39	0.39	0.39	0.39	0.39	0.39
$\rho = 0.85$							
DF^{μ}	0.60	0.63	0.70	0.80	0.91	0.97	0.99
$\hat{Q}^{\mu}(10,1)$	0.90	0.85	0.55	0.08	0.00	0.00	0.00
$\hat{Q}^{\mu}(10,\infty)$	0.33	0.43	0.64	0.85	0.96	0.99	1.00
$\hat{Q}^{\mu}(10,k_*^{\mu})$	0.75	0.75	0.77	0.79	0.82	0.85	0.88
$\hat{P}^{\mu}_{\mathrm{inv}}(10)$	0.71	0.70	0.68	0.66	0.62	0.59	0.54
$\rho = 0.80$							
DF^{μ}	0.85	0.87	0.91	0.95	0.98	1.00	1.00
$\hat{Q}^{\mu}(10,1)$	0.99	0.98	0.89	0.43	0.02	0.00	0.00
$\hat{Q}^{\mu}(10,\infty)$	0.61	0.69	0.85	0.96	0.99	1.00	1.00
$\hat{Q}^{\mu}(10,k_*^{\mu})$	0.93	0.93	0.95	0.96	0.97	0.98	0.99
$\hat{P}^{\mu}_{\mathrm{inv}}(10)$	0.91	0.90	0.87	0.83	0.76	0.67	0.58

Based on 10 000 Monte Carlo replications with $p_{\max} = 0$.

Gaussian disturbances. Results are presented for various α with four panels of results corresponding to $\rho = 0.95, 0.9, 0.85$ and 0.8, respectively.

The Dickey and Fuller t-statistic test has the unusual power profile of power increasing in $|\alpha|$. The larger the initial value the higher the power, a property that comes at the expense of having relatively low power when the initial condition is small. When $\alpha = 0$ and $\rho \geqslant 0.9$ power for DF^{μ} is less than half that attained by $\hat{Q}^{\mu}(10,1)$.

The statistics presented in Sections 2 and 3, each attempting to deal with obtaining inference that is not too much affected over a wide range of values for α, achieve this goal with varying degrees of success. As found above, relying on the efficient test for a known initial condition ξ with ξ replaced by its maximum likelihood estimator under the fixed alternative $\rho = r = 1 - g/T$ results in a test that is in the efficient

Table 4
Size corrected small sample power in the trend case

Test	α						
	0	1	2	3	4	5	6
$\rho = 0.95$							
DF^{τ}	0.08	0.08	0.09	0.09	0.10	0.12	0.14
$\hat{Q}^{\tau}(15,1)$	0.10	0.09	0.06	0.03	0.01	0.00	0.00
$\hat{Q}^{\tau}(15,\infty)$	0.06	0.07	0.10	0.13	0.17	0.20	0.23
$\hat{Q}^{\tau}(15,k_*^{\tau})$	0.10	0.09	0.08	0.07	0.05	0.03	0.02
$\hat{P}^{\tau}_{\text{inv}}(15)$	0.09	0.09	0.09	0.08	0.07	0.05	0.04
$\rho = 0.90$							
DF^{τ}	0.19	0.19	0.21	0.25	0.31	0.39	0.49
$\hat{Q}^{\tau}(15,1)$	0.29	0.24	0.12	0.04	0.01	0.00	0.00
$\hat{Q}^{\tau}(15,\infty)$	0.10	0.13	0.21	0.33	0.46	0.58	0.69
$\hat{Q}^{\tau}(15,k_*^{\tau})$	0.23	0.23	0.21	0.19	0.16	0.13	0.10
$\hat{P}^{\tau}_{\text{inv}}(15)$	0.22	0.22	0.21	0.21	0.20	0.19	0.18
$\rho = 0.85$							
DF^{τ}	0.38	0.39	0.44	0.51	0.61	0.72	0.82
$\hat{Q}^{\tau}(15,1)$	0.59	0.50	0.29	0.09	0.01	0.00	0.00
$\hat{Q}^{\tau}(15,\infty)$	0.22	0.27	0.40	0.58	0.75	0.87	0.94
$\hat{Q}^{\tau}(15,k_*^{\tau})$	0.47	0.47	0.46	0.45	0.43	0.41	0.39
$\hat{P}^{\tau}_{\text{inv}}(15)$	0.43	0.43	0.43	0.43	0.43	0.43	0.43
$\rho = 0.80$							
DF^{τ}	0.64	0.66	0.70	0.77	0.85	0.91	0.96
$\hat{Q}^{\tau}(15,1)$	0.84	0.78	0.55	0.22	0.03	0.00	0.00
$\hat{Q}^{\tau}(15,\infty)$	0.42	0.48	0.63	0.80	0.92	0.97	0.99
$\hat{Q}^{\tau}(15,k_*^{\tau})$	0.73	0.73	0.73	0.74	0.75	0.75	0.76
$\hat{P}^{\tau}_{\text{inv}}(15)$	0.70	0.69	0.69	0.68	0.66	0.64	0.62

Based on 10 000 Monte Carlo replications with $p_{\max} = 0$.

class derived by ME with $k \to \infty$. This test, denoted $\hat{Q}^{i}(g,\infty)$, has power that increases as α gets larger just as for the DF^{i} statistic, however it starts from a much lower level power when α is small and increases at a faster rate as α gets larger. Hence if a researcher believed that the initial value was far from the deterministic component, this test is preferable as it has higher power when the initial is indeed large. The test does not, however, possess attractive power properties for the case where the initial condition is plausibly either small or large.

The test $\hat{P}^{\mu}_{\text{inv}}(10)$, designed to be invariant to ξ at $\gamma = 10$ (which corresponds to $\rho = 0.9$ for $T = 100$), does achieve to some extent the goal of providing a reliable testing procedure for a wide range of values for ξ. Against the alternative that $\rho = 0.9$, the test has power of 39% everywhere, a result that matches the asymptotic results given in Section 3. For closer and more distant alternatives, however, just as the asymptotic theory predicts, power falls over the range of α considered. This drop in power is quite dramatic at $\rho = 0.80$.

Finally, $\hat{Q}^{\mu}(10, k_*^{\mu})$ does achieve the aim of having relatively flat power over different initial conditions. For $\rho = 0.9$ the power of the statistic is between 41% and 43%. Compared to $\hat{Q}^{\mu}(10, 1)$, it is making a trade-off reducing power for small initial conditions but gaining in power for larger values of α. For alternatives closer to the null power falls as α gets large, and the reverse is true for more distant alternatives. However, over nearly all of the range of α power of $\hat{Q}^{\mu}(10, k_*^{\mu})$ exceeds that of $\hat{P}_{\text{inv}}^{\mu}(10)$. The power gains of $\hat{Q}^{\mu}(10, k_*^{\mu})$ over $\hat{P}_{\text{inv}}^{\mu}(10)$ are more pronounced for more distant alternatives (i.e., small ρ) and larger initial conditions. For such models, $\hat{Q}^{\mu}(10, k_*^{\mu})$ has power getting larger as the initial condition gets larger whereas the power of $\hat{P}_{\text{inv}}^{\mu}(10)$ goes to zero.

The comparison between DF^{μ} and $\hat{Q}^{\mu}(10, k_*^{\mu})$ shows the unusual trade-off made by the former test. It has greater power than $\hat{Q}^{\mu}(10, k_*^{\mu})$ only when the initial condition is quite far from zero, again at a fairly substantial cost to power when the initial condition is small. Whilst DF^{μ} does maintain some semblance of power against all possible initial conditions, the less pronounced variations in power of $\hat{Q}^{\mu}(10, k_*^{\mu})$ makes it a more attractive test statistic when it is not known whether the initial condition is large or small.

The same models are considered when the tests are invariant to a time trend in addition to a mean shift. These results are given in Table 4. Essentially the same comments apply, although as is typical of results in this literature magnitudes are all smaller when a time trend is considered. For more distant alternatives and large values for α the differences between $\hat{Q}^{\tau}(15, k_*^{\tau})$ over $\hat{P}_{\text{inv}}^{\tau}(15)$ are less pronounced than in the mean case.

The development of the family of tests $Q^{i}(g, k)$ relied on constructing likelihood ratios under the assumption that the shocks driving the process are normally distributed. All of the asymptotic properties of the tests—asymptotic critical values and power properties—were derived under wider distributional assumptions on these shocks. The effect of departures from normality such as skewness and kurtosis are explored in Table 5. We report size adjusted power for $\hat{Q}^{\mu}(10, k_*^{\mu})$ in the model examined in Table 3 with the exception that we now have three different generating processes for v_t. The first is the $N(0, 1)$ results from Table 3, the second model has $v_t \sim \chi_1^2 - 1$, where χ_1^2 is a centered chi-squared random variable with one degree of freedom, and in the third $v_t \sim t_5$, a student-t distribution with 5 degrees of freedom. Thus for the second model the shocks are skewed and for the third there is excess kurtosis over the standard normal. Size for both the second and third models is 3.2% and 3.9%, respectively, similar to the case of a standard normal. The table shows clearly that there is little to no effect of the distributional assumption for i.i.d. innovations with this number of observations.

Overall the small sample results validate our asymptotic results of Section 3. For situations with potentially large initial conditions, we hence recommend basing

Table 5
Size corrected small sample power of a test based on $\hat{Q}^{\mu}(10, k_*^{\mu})$ for i.i.d. innovations of various distributions in the mean case

α	0	1	2	3	4	5	6
$\rho = 0.95$							
N(0, 1)	0.17	0.16	0.13	0.09	0.06	0.03	0.02
$\chi_1^2 - 1$	0.19	0.18	0.16	0.12	0.09	0.06	0.04
t_5	0.18	0.17	0.15	0.12	0.09	0.07	0.04
$\rho = 0.90$							
N(0, 1)	0.43	0.43	0.43	0.43	0.42	0.42	0.41
$\chi_1^2 - 1$	0.47	0.47	0.47	0.47	0.46	0.46	0.45
t_5	0.43	0.44	0.44	0.43	0.43	0.43	0.42
$\rho = 0.85$							
N(0, 1)	0.75	0.75	0.77	0.79	0.82	0.85	0.88
$\chi_1^2 - 1$	0.80	0.80	0.81	0.82	0.84	0.85	0.87
t_5	0.75	0.75	0.76	0.78	0.80	0.82	0.84
$\rho = 0.80$							
N(0, 1)	0.93	0.93	0.95	0.96	0.97	0.98	0.99
$\chi_1^2 - 1$	0.95	0.95	0.96	0.96	0.97	0.97	0.98
t_5	0.94	0.94	0.94	0.95	0.96	0.97	0.98

Based on 10 000 Monte Carlo replications with $p_{\max} = 0$.

inference on $\hat{Q}^{\mu}(10, k_*^{\mu})$ or $\hat{Q}^{\tau}(15, k_*^{\tau})$, which have approximately flat power as a function of the initial condition while controlling size reasonably well.

5. Conclusion

The power of all unit root tests suggested in the literature so far depends on the initial condition. We show that numerical independence of the outcome of a unit root test on the value of the initial condition can only be achieved by the trivial test that does not depend on the data. A requirement of no influence of the initial condition on unit root tests is hence not a reasonable proposition in practice. This means that researchers are forced to consider the influence of the initial condition on the power of their tests. In classical testing theory the choice between tests reduces to the choice between power profiles, and the appropriate choice of test reduces to choosing the power profile most appropriate to the problem at hand. Whilst this usually means considering power as a function in the direction of the stated alternative, in the case of unit root tests this also includes consideration of power as a function of various initial conditions. For problems where it is unlikely that the initial condition is large, it makes sense to choose a test that has maximal power when the initial condition is small or moderate. On the other hand, if the data is such that the initial condition is quite possibly large, it makes sense to choose a test that has good power over many possible values for the initial condition. In either case, the choice should be an efficient test, i.e. a test whose power cannot be dominated over all alternatives and values of the initial condition.

In some circumstances, it makes sense not to make the assumption of a necessarily small initial condition and to rely on unit root tests with equal power for general initial conditions, including large ones (in absolute value). We derive an asymptotically efficient unit root test statistic that possesses this property, at least to a fair degree. The test is one specific member of the family of efficient tests derived by ME: in the case where the deterministics are given by a constant only, it is $\hat{Q}^{\mu}(10, 3.8)$ and in the linear trend case it is $\hat{Q}^{\tau}(15, 3.968)$. In contrast to the DF^{i} test (whose power especially in the mean case becomes quite low when the initial condition is small) and tests that are optimal for small initial conditions (whose power becomes very low when the initial is large), tests based on our statistic have power that is less sensitive to the size of the initial condition. We hence suggest relying on the statistic derived in this paper when a researcher has no knowledge or expectation that the initial value is large or small.

The effect of using this test in the empirical example described in the introduction can be seen in Fig. 4, which depicts the p-values of unit root tests based on the DF^{μ} statistic, $\hat{Q}^{\mu}(10, 1)$ and $\hat{Q}^{\mu}(10, 3.8)$ as a function of the sample starting date. The figure shows that the variability of the p-values of $\hat{Q}^{\mu}(10, 3.8)$ is considerably reduced compared to the other tests, especially compared to the DF^{μ} statistic. The differences are stark in the first half of the 19th century. For starting dates in this period, a period of high volatility of the real exchange rate, the p-values for the DF^{μ} test exhibit large variations, exceeding the 5% threshold for some short periods. Contrast this with the p-values of the $\hat{Q}^{\mu}(10, 3.8)$ test, which—despite being positively correlated with those of the DF^{μ} test—are considerably smoother over time and do not lead to these sudden failures to reject the null hypothesis. The most striking feature in the comparison with $\hat{Q}^{\mu}(10, 1)$ is that the test based on $\hat{Q}^{\mu}(10, 3.8)$ does not lead to peaks in the p-values aligned with peaks in the data, as we discussed in the Introduction. Overall the outcome of $\hat{Q}^{\mu}(10, 3.8)$ statistic is most stable as a function of the starting date, with a relatively smooth evolution of its p-values.

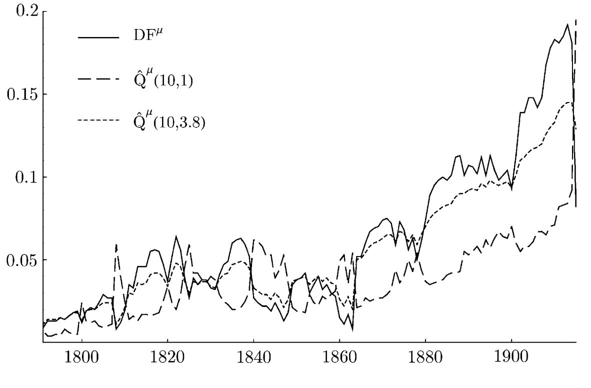

Fig. 4. p-values of unit root tests of the British pound/US dollar real exchange rates as a function of the starting date of the sample.

We noted that the differing performance of unit root tests for different initial values allowed data mining of the form of choosing the start date to obtain the desired outcome. A researcher that wanted to use cointegration methods with the real exchange rate using the DF^{μ} statistic, say, could choose a starting date of 1820 in order to ensure that the unit root is not rejected, so as to justify the subsequent cointegration analysis. Basing unit root testing inference on $\hat{Q}^{\mu}(10, 3.8)$ minimizes this possibility, as the outcomes of this test are much less volatile.

Either as safeguard against this form of data mining, or as the appropriate test in situations where nothing definite can be said about the magnitude of the initial condition, the tests developed in this paper should be a useful addition to the unit root testing toolbox: They are asymptotically efficient, control size reasonably well and have the appropriate power profile with respect to the initial condition for such circumstances.

Acknowledgements

We thank David Harvey, two referees and the editor for helpful comments.

Appendix

Proof of Theorem 1. We show that any y can be reduced to $(0, \ldots, 0)'$ by an appropriate sequence of $T+1$ transformations of the form $y \to y + a_i R(r)$, where $y = (y_0, \ldots, y_T)'$, $R(r) = (1, r, r^2, \ldots, r^T)'$ and a_i is a scalar. Let $A = (a_0, a_1, \ldots, a_T)'$ and $B = (R(\rho_0),\ R(\rho_1), \ldots, R(\rho_T))$. An appropriate sequence must exist if for some choice of $\rho_0, \ldots, \rho_T$, B has rank $T+1$, as in this case

$$0 = y + BA \ \Leftrightarrow \ A = -B^{-1}y. \tag{20}$$

We now prove by contradiction that there must exist $\rho_0, \ldots, \rho_T$ such that B has rank $T+1$. Suppose otherwise. Let $r_0, r_1, \ldots, r_N$, $N < T$, be such that $B = (R(r_0), R(r_1), \ldots, R(r_N))$ has rank $N+1$ and $R(s)$ lies in the column space of B for all $s \in (0, 1)$. A least-squares regression of $R(s)$ on B must then yield residuals that are a $T+1$ column vector of zeros

$$0 = R(s) - B(B'B)^{-1}B'R(s) \tag{21}$$

for any $s \in (0, 1)$. Differentiation of (21) with respect to s yields

$$0 = \frac{\partial R(s)}{\partial s} - B(B'B)^{-1}B'\frac{\partial R(s)}{\partial s}. \tag{22}$$

It follows that $\partial R(s)/\partial s$ lies in the column space of B, too. Reapplication of the same argument with $\partial R(s)/\partial s$ in place of $R(s)$ shows that derivatives of $R(s)$ of any order must lie in the column space of B. But clearly the jth derivative of $R(s)$ has zeros in the first j rows and no zeros in all other rows, so that the derivatives of order zero to T form a basis of $\mathbb{R}^{T+1}$. It cannot be true that they all lie in the column space of B at the same time.

Proof of Theorem 2. The proof is similar to the proof of Theorem 3 of ME, we only need to consider the larger column space of $X = (X_0, \dots, X_T)'$, which we generate by considering the additional regressor $\{r^t - 1\}$. We focus on the time trend case, so that we define Z as the $(T+1) \times 3$ matrix with tth row $(1, t, r^t - 1)$. The mean case is proved analogously.

First note that the results (i) and (iii) of Lemma 1 of ME extend to the $T \times 4$ matrix $\tilde{B}$ with tth row $(1, t, \rho^t, r^t)$ with the same proof, simply redefine $\tilde{C}$ as the $T \times 4$ matrix with tth row $C_t = (0, 1, -\gamma\rho^t, -gr^t)$ and proceed as in ME.

Define Σ_0^-, Σ_1^-, $\tilde{V}^{-1}$, c and $\tilde{\varphi}(c)$ as in the proof of Theorem 3 of ME. Let $\tilde{r} = 1 - c/T$, $\tilde{H}(c) = (\tilde{e} + cT^{-1}\tilde{\tau}_{-1}, cT^{-1}\tilde{e} + (c-g)T^{-1}\tilde{R}(\tilde{r})/\tilde{r})'$, $\Upsilon_1 = \mathrm{diag}(T^{1/2}, T^{-1/2}, T^{1/2})$ and $\Upsilon_0 = \mathrm{diag}(1, T^{-1/2}, T^{1/2})$. Then $\Upsilon_1 Z' \Sigma_1^- Z \Upsilon_1 = \mathrm{diag}(2g\omega^{-2}, T^{-1}\tilde{H}(g)'\tilde{V}^{-1}\tilde{H}(g))$, $\Upsilon_0 Z' \Sigma_0^- Z \Upsilon_0 = \mathrm{diag}(1, T^{-1}\tilde{H}(0)'\tilde{V}^{-1}\tilde{H}(0))$, $\Upsilon_1 Z' \Sigma_1^- u = (0, T^{-1/2}\tilde{H}(g)'\tilde{V}^{-1}\tilde{\varphi}(g))'$ and $\Upsilon_0 Z' \Sigma_0^- u = (0, T^{-1/2}\tilde{H}(0)'\tilde{V}^{-1}\tilde{\varphi}(0))'$. Proceeding as in ME, we find by applying (the slightly generalized) Lemma 1 of ME, (10) and the continuous mapping theorem that

$$\begin{aligned}
P_{\mathrm{inv}}(g) \Rightarrow{} & g\tilde{M}^{\tau_R}(1)^2 - g + g^2 \int \tilde{M}^{\tau_R}(s)^2\,\mathrm{d}s \\
& - \begin{pmatrix} (g+1)\tilde{M}^{\tau_R}(1) \\ g\tilde{M}^{\tau_R}(1) + g^2 \int \tilde{M}^{\tau_R}(s)\,\mathrm{d}s \end{pmatrix}' \begin{pmatrix} \int (1+gs)^2\,\mathrm{d}s & \int g(1+gs)\,\mathrm{d}s \\ \int g(1+gs)\,\mathrm{d}s & g^2 \end{pmatrix}^{-1} \\
& \times \begin{pmatrix} (g+1)\tilde{M}^{\tau_R}(1) \\ g\tilde{M}^{\tau_R}(1) + g^2 \int \tilde{M}^{\tau_R}(s)\,\mathrm{d}s \end{pmatrix} + \begin{pmatrix} \tilde{M}^{\tau_R}(1) \\ g^2 \int \tilde{M}^{\tau_R}(s) + g\mathrm{e}^{-g}\tilde{M}^{\tau_R}(1) \end{pmatrix}' \\
& \times \begin{pmatrix} 1 & -g\int \mathrm{e}^{-gs}\,\mathrm{d}s \\ -g\int \mathrm{e}^{-gs}\,\mathrm{d}s & g^2 \int \mathrm{e}^{-2gs}\,\mathrm{d}s \end{pmatrix}^{-1} \begin{pmatrix} \tilde{M}^{\tau_R}(1) \\ g^2 \int \tilde{M}^{\tau_R}(s) + g\mathrm{e}^{-g}\tilde{M}^{\tau_R}(1) \end{pmatrix},
\end{aligned} \tag{23}$$

where $\tilde{M}^{\tau_R}(s)$ is the continuous time projection of $M(s)$ off $(s, 1 - \mathrm{e}^{-gs})$, such that $\int \tilde{M}^{\tau_R}(s)(1 - \mathrm{e}^{-gs})\,\mathrm{d}s = 0$ implies $\int \tilde{M}^{\tau_R}(s)\mathrm{e}^{-gs}\,\mathrm{d}s = \int \tilde{M}^{\tau_R}(s)\,\mathrm{d}s$. The result now follows from the relationship between $\tilde{M}^{\tau_R}(s)$ and $M^{\tau_R}(s)$ after some straightforward but tedious algebra.

Proof of Lemma 1. The proof follows closely the method developed in Nabeya and Tanaka (1990) and Tanaka (1996, p. 109). A similar result may also be found in Elliott and Stock (2001). We have to take care, however, how α enters the picture.

By Girsanov's Theorem, the Radon–Nikodym derivative of the measure of the Itô Process G with $G(0) = 0$ and

$$\mathrm{d}G(s) = -\vartheta G(s) - \tfrac{\sqrt{2}}{2} a\vartheta^{1/2} + \mathrm{d}W(s) \tag{24}$$

with respect to the measure of the Wiener process W, evaluated at G, is given by

$$\exp\left[-\int (\vartheta G(s) + \frac{\sqrt{2}}{2} a\vartheta^{1/2})\,\mathrm{d}G(s) - \frac{1}{2}\int (\vartheta G(s) + \frac{\sqrt{2}}{2} a\vartheta^{1/2})^2\,\mathrm{d}s \right]. \tag{25}$$

The Radon–Nikodym derivative of the measure of M as defined in (10), i.e. the process G with $a = \alpha$ and $\vartheta = \gamma$, with respect to the measure of the process $\mathrm{d}\tilde{M}(s) = -\delta\tilde{M}(s) + \mathrm{d}W(s)$, i.e. G with $a = 0$ and $\vartheta = \delta$, evaluated at $\tilde{M}$ is hence given by

$$\exp\left[(\delta-\gamma)\int \tilde{M}(s)\,\mathrm{d}\tilde{M}(s) + \frac{1}{2}(\delta^2-\gamma^2)\int \tilde{M}(s)^2\,\mathrm{d}s \right.$$
$$\left. -(2\gamma)^{-1/2}\alpha\left[\gamma\tilde{M}(1) + \gamma^2\int \tilde{M}(s)\,\mathrm{d}s\right] - \frac{1}{4}\alpha^2\gamma\right].$$

Therefore

$$\phi(\theta) = \mathrm{E}[\exp\{\theta\Psi\mathbf{i}\}], \tag{26}$$

$$= \tilde{\mathrm{E}}\left[\exp\left\{\theta l_0\mathbf{i} - \frac{1}{2}(\delta-\gamma) - \frac{1}{4}\alpha^2\gamma \right.\right.$$
$$\left.\left. + \left(\theta l_1\mathbf{i} - \frac{1}{2}(\gamma^2-\delta^2)\right)\int M(s)^2\,\mathrm{d}s + H'\tilde{\Lambda}H + \tilde{\lambda}'H\right\}\right], \tag{27}$$

$$= \exp\{\theta l_0\mathbf{i} - \tfrac{1}{2}(\delta-\gamma) - \tfrac{1}{4}\alpha^2\gamma\}\tilde{\mathrm{E}}[\exp\{H'\tilde{\Lambda}H + \tilde{\lambda}'H\}], \tag{28}$$

where $\delta = \sqrt{\gamma^2 - 2l_1\theta\mathbf{i}}$ and $\tilde{\mathrm{E}}$ denotes the expectation of M with respect to the measure of $\tilde{M}$.

But under the measure of $\tilde{M}$, H is Gaussian, and by 'completing the square'

$$\tilde{\mathrm{E}}[\exp\{H'\tilde{\Lambda}H + \tilde{\lambda}'H\}] = (2\pi)^{-\ell/2}|V(\delta)|^{-1/2}$$
$$\times\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}[H'(V(\delta)^{-1} - 2\tilde{\Lambda})H - 2\tilde{\lambda}'H]\right]\mathrm{d}H, \tag{29}$$

$$= |I_\ell - 2V(\delta)\tilde{\Lambda}|^{-1/2}\exp\left[\frac{1}{2}\tilde{\lambda}'(V(\delta)^{-1} - 2\tilde{\Lambda})^{-1}\tilde{\lambda}\right]. \tag{30}$$

The last line follows after noting that $\tilde{\mathrm{E}}[H] = 0$ and, with $\tilde{m} = (V(\delta)^{-1} - 2\tilde{\Lambda})^{-1}\tilde{\lambda}$, that $(H-\tilde{m})'(V(\delta)^{-1} - 2\tilde{\Lambda})(H-\tilde{m}) = H'(V(\delta)^{-1} - 2\tilde{\Lambda})H - 2\tilde{\lambda}'H + \tilde{\lambda}'(V(\delta)^{-1} - 2\tilde{\Lambda})^{-1}\tilde{\lambda}$.

For the second claim of the Lemma, note that the definition of $\tilde{M}(\cdot)$ and some stochastic calculus yields for $\alpha = 0$ that $H \sim \int(g_1(s),\ldots,g_\ell(s))'\,\mathrm{d}W(s)$, and the result follows.

Derivation of k_*^μ and k_*^τ.

We focus on k_*^μ, the derivations for k_*^τ are analogous but highly tedious and are omitted for brevity.

From direct calculations

$$K^\mu(0) = \alpha\frac{g - 1 + \mathrm{e}^{-g}}{\sqrt{2}g^{3/2}}, \tag{31}$$

$$K^\mu(1) = \alpha\frac{\mathrm{e}^{-g}(1+g) - 1}{\sqrt{2}g^{3/2}}, \tag{32}$$

$$\int K^\mu(s)^2\,\mathrm{d}s = \alpha^2\frac{4\mathrm{e}^{-g} + g - 2 - \mathrm{e}^{-2g}(g+2)}{4g^3}, \tag{33}$$

$q_0^\mu = -g$, $q_1^\mu = -g(1+g)(k-2)/(2+gk)$, $q_2^\mu = (2g-gk+g^2k)/(2+gk)$, $q_3^\mu = 2g(k-2)/(2+gk)$, $q_4^\mu = g^2$. Hence

$$q_1^\mu K^\mu(0)^2 + q_2^\mu K^\mu(1)^2 + q_3^\mu K^\mu(0)K^\mu(1) + q_4^\mu \int K^\mu(s)^2\,\mathrm{d}s \tag{34}$$

$$= \alpha^2\Bigg[-\frac{g(1+g)(k-2)}{2+gk}\left(\frac{g-1+\mathrm{e}^{-g}}{\sqrt{2}g^{3/2}}\right)^2 + \frac{2g-gk+g^2k}{2+gk}\left(\frac{\mathrm{e}^{-g}(1+g)-1}{\sqrt{2}g^{3/2}}\right)^2$$

$$+\left(\frac{2g(k-2)}{2+gk}\right)\left(\frac{g-1+\mathrm{e}^{-g}}{\sqrt{2}g^{3/2}}\right)\left(\frac{\mathrm{e}^{-g}(1+g)-1}{\sqrt{2}g^{3/2}}\right) + g^2\frac{4\mathrm{e}^{-g}+g-2-\mathrm{e}^{-2g}(g+2)}{4g^3}\Bigg],$$

$$= \frac{1}{4}\alpha^2\left(-1+\mathrm{e}^{-2g}+\frac{4g}{2+gk}\right) \tag{35}$$

from the definition of q_i^μ in (12). Setting this expression to zero and solving for k yields k_*^μ.

References

Chan, N., Wei, C., 1987. Asymptotic inference for nearly nonstationary AR(1) processes. The Annals of Statistics 15, 1050–1063.

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

Dufour, J.-M., King, M., 1991. Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary or nonstationary AR(1) errors. Journal of Econometrics 47, 115–143.

Elliott, G., 1999. Efficient tests for a unit root when the initial observation is drawn from its unconditional distribution. International Economic Review 40, 767–783.

Elliott, G., Rothenberg, T., Stock, J., 1996. Efficient tests for an autoregressive unit root. Econometrica 64, 813–836.

Elliott, G., Stock, J., 2001. Confidence intervals for autoregressive coefficients near one. Journal of Econometrics 103, 155–181.

Evans, G., Savin, N., 1981. Testing for unit roots: 1. Econometrica 49, 753–779.

Evans, G., Savin, N., 1984. Testing for unit roots: 2. Econometrica 52, 1241–1269.

Harvey, A., Bates, D., 2003. Multivariate unit root tests and testing for convergence. Cambridge Working Paper in Economics 0301.

Harvey, D., Leybourne, S., 2005. On testing for unit roots and the initial observation. The Econometrics Journal 8, 97–111.

Lothian, J., Taylor, M., 1996. Real exchange rate behavior: the recent float from the perspective of the past two centuries. Journal of Political Economy 104, 488–509.

Müller, U., Elliott, G., 2003. Tests for unit roots and the initial condition. Econometrica 71, 1269–1286.

Nabeya, S., Tanaka, K., 1990. Limiting power of unit-root tests in time-series regression. Journal of Econometrics 46, 247–271.

Ng, S., Perron, P., 2001. Lag length selection and the construction of unit root tests with good size and power. Econometrica 69, 1519–1554.

Phillips, P., 1987. Towards a unified asymptotic theory for autoregression. Biometrika 74, 535–547.

Stock, J., 1994. Unit roots, structural breaks and trends. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, vol. 1. North-Holland, New York, pp. 2740–2841.

Stock, J., 2000. A class of tests for integration and cointegration. In: Engle, R., White, H. (Eds.), Cointegration, Causality, and Forecasting—A Festschrift in Honour of Clive W.J. Granger. Oxford University Press, Cambridge, pp. 135–167.

Tanaka, K., 1996. Time Series Analysis—Nonstationary and Noninvertible Distribution Theory. Wiley, New York.