Title

Tests for Unit Roots and the Initial Observation

Authors
Muller, Ulrich
Elliott, Graham

Publication Date
2001-12-03

2001-19

UNIVERSITY OF CALIFORNIA, SAN DIEGO

DEPARTMENT OF ECONOMICS

TESTS FOR UNIT ROOTS AND THE INITIAL OBSERVATION

BY

ULRICH K. MÜLLER

AND

GRAHAM ELLIOTT

DISCUSSION PAPER 2001-19
DECEMBER 2001

TESTS FOR UNIT ROOTS AND THE INITIAL OBSERVATION

By Ulrich K. Müller and Graham Elliott[1]

First version: April 10, 2001
This revision: December 3, 2001

The paper analyzes the impact of the initial observation on the problem of testing for unit roots. To this end, we derive a family of optimal tests that maximize a weighted average power criterion with respect to the initial observation. We then investigate the relationship of this optimal family to unit root tests in an asymptotic framework. We find that many popular unit root tests are closely related to specific members of the optimal family, but the corresponding members employ very different weightings for the initial observation. The popular Dickey-Fuller tests, for instance, are closely related to optimal tests which put a large weight on extreme deviations of the initial observation from the deterministic component, whereas other popular tests put more weight on moderate deviations. At the same time, the power of the various unit root tests varies dramatically with the initial observation. This paper therefore helps to explain the results of comparative power studies of unit root tests, and allows a much deeper understanding of the merits of particular tests in specific circumstances.

KEYWORDS: Unit root tests, point optimal tests, weighted average power, asymptotic distributions

[1]Part of the paper was written during a visit of Müller at the University of California at San Diego. He thanks UCSD for its hospitality and the Swiss National Fund for generous financial support. Elliott acknowledges support in part from National Science Foundation Grant SES-01-11238.

1. INTRODUCTION

Few papers in the large unit root testing literature explicitly address the role of the beginning of the data series y_t. In analyzing the properties of these tests, all papers must make an assumption on the deviation of y_0 from its modelled deterministic part — call this deviation ξ. As is known from Monte Carlo studies and shown analytically below, the way the nuisance parameter ξ is dealt with affects profoundly the power performance and the relevant asymptotic theory, as well as optimality results for testing procedures. The motivation behind this paper is to develop optimality theory with regards to testing a unit root under various possible assumptions on the initial condition, and also to show the implicit assumptions made by various tests currently in use for testing for a unit root. In this way we gain a deeper understanding of the properties of these now ubiquitous tests. In addition we derive a general family of feasible tests that have known optimality properties.

There has been great emphasis on 'getting the deterministic terms right' in the unit root testing literature, both theoretically and in practice. This is important for both the null and alternative model. In one sense, the initial condition has an effect similar to these deterministic terms. A fixed nonzero initial condition introduces a term that is indistinguishable from a mean shift when there is a unit root and a term which asymptotes to zero when the root is less than one. The primary difference between the typical deterministic terms and the initial condition is this differential behavior under the null and alternative hypotheses. The usual method for dealing with deterministic terms, invariance, is not appropriate in this situation. Instead, we derive the implication that for any unit root test one must take either implicitly or explicitly take a stand on what the initial condition is.

The idea that one must take a stand on the initial condition seems at odds with the majority of the literature. Typically, most papers derive their tests under the assumption that ξ is constant or comes from a bounded distribution. In terms of the convergence of y_t suitably standardized to an Ornstein-Uhlenbeck process such assumptions have no effect as ξ similarly transformed disappears at rate $T^{1/2}$. Thus under this assumption, one can ignore this term. Secondly, in terms of optimality under assumptions such as normality of the underlying error process, this term in the likelihood similarly disappears asymptotically and so can be ignored. However in fixed sample sizes this term may well be relatively large compared to the terms that do not disappear asymptotically, and hence the typical functional central limit theorem results may not give good approximations to the small sample distributions and power functions of the tests. Tests which are theoretically optimal for large samples when we ignore this term may well be suboptimal in practice.

We deal with both of these implications for the initial condition. In terms of optimality, we relax the assumptions on ξ to include the cases where it disappears but also cases where ξ does not disappear asymptotically. The second of these allows the presence of the initial condition to remain asymptotically, and hence provide potentially more relevant asymptotic approximations for models where ξ is relatively large. As we cannot appeal to invariance to rid ourselves of this nuisance parameter, we instead derive a family of tests which are optimal in the sense of maximizing a weighted average power over different values of ξ.

By comparison to the optimal family, this analysis allows us to infer what implicit weighting is made by commonly applied tests for a unit root. For some tests we are able to show that they are indeed members of the optimal family for a certain weighting function. Whilst the Dickey-Fuller statistics do not belong to this optimal family, they exhibit a very close relationship to test statistics which optimally test a standard mean reverting model against a special integrated model. For other popular tests, we are able to identify a specific member of the optimal family such that the asymptotic power of the test and the corresponding optimal test becomes very much comparable.

These correspondences allows us to much deeper understand the merits of various unit root tests with respect to their handling of ξ. The implicit weightings explain much of the behavior in terms of power of different unit root tests. We find for the popular augmented Dickey Fuller test that it places very high weight on extremely large values of the initial condition relative to the variance of the error process. For other tests which fare well in Monte Carlo studies, like tests based on Weighted Symmetric estimators, the implicit weighting is concentrated on much smaller values of ξ. Given that both tests are close to optimal for different assumptions concerning ξ, the lower power of the Dickey Fuller tests is a natural consequence of the typical set-up of Monte Carlo studies which use relatively moderate values for the initial condition.

In the next section we build the basic model and discuss various methods how to deal with the nuisance parameter ξ. We then consider a family of tests which maximize weighted average power over different initial conditions, where the weighting function is a given distribution function in both small and large samples. Section four relates commonly employed unit root tests to members of the optimal family. The relationships allow us to explain the differences in robustness to the initial condition between the tests and also the power trade-offs implicit in the tests. Finally, we derive a family of feasible tests which are asymptotically equivalent to the family of optimal tests.

2. HYPOTHESIS TESTING AND THE INITIAL CONDITION

We will consider the following general model in this paper

$$\begin{aligned} y_t &= X_t'\beta + \mu + w_t \qquad t = 0, 1, \cdots, T \\ w_t &= \rho w_{t-1} + \nu_t \qquad t = 1, \cdots, T \\ w_0 &= \xi \end{aligned} \tag{2.1}$$

where X_t is a predetermined vector which has no constant element, $X_0 = 0$ and μ, β and ξ are unknown. We also assume that the regressor matrix $X = (X_1, \cdots, X_T)'$ has full column rank.

We are interested in distinguishing the two hypotheses

$$\begin{aligned} H_0 &: \rho = 1 \\ H_1 &: \rho < 1. \end{aligned} \tag{2.2}$$

This model has received a great deal of attention. Test statistics typically do not have approximate normal distributions, and much of the intuition from the stationary world as to which tests are optimal does not hold for this testing situation. Many feasible test statistics have been suggested, the most famous being Dickey and Fuller's (1979) t-test and $\hat{\rho}$-test. Monte Carlo evidence leads Pantula, Gonzalez-Farias, and Fuller (1994) to promote tests based on Weighted Symmetric regressions. Leybourne (1995) suggests a test based on forward and reverse Dickey-Fuller regressions. None of these tests have known optimality properties.

Less work is concerned with the derivation of optimal tests. Dufour and King (1991) derive the Point Optimal Test and Locally Best Test for independent Gaussian disturbances ν_t and an independent zero-mean normal ξ for various ρ. Elliott, Rothenberg, and Stock (1996) derive the family of asymptotically optimal tests against a fixed alternative $\rho < 1$ when ξ is bounded in probability and (possibly correlated) Gaussian ν_t. Rothenberg and Stock (1997) extend this to alternate distributions on the error terms. Elliott (1999) derives the family of asymptotically optimal tests for independent ν_t when ξ is drawn from its unconditional distribution under the fixed alternative.

In most parts of this paper, we will consider ξ to be an unknown and fixed nuisance parameter. With $y = (y_1, \cdots, y_T)'$, $w = (w_1, \cdots, w_T)$ and e a $T \times 1$ vector of ones, we can write the last T observations of the model more compactly as

$$y = X\beta + \mu e + w.$$

Note that the expectation of w is not zero, but rather $E[w] = \xi R(\rho)$, where $R(r)$ is the $T \times 1$ vector $R(r) = (r, r^2, \cdots, r^T)'$. Define the $T \times T$ matrix $A(r)$ as

$$A(r) = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -r & 1 & 0 & \ldots & 0 & 0 \\ 0 & -r & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -r & 1 \end{pmatrix}$$

and let V be the variance-covariance matrix of $\nu = (\nu_1, \cdots, \nu_T)'$. Then $A(\rho)(w - \xi R(\rho)) = \nu$, and the variance-covariance matrix $\Sigma(\rho)$ of w satisfies $A(\rho)\Sigma(\rho)A(\rho)' = V$. Finally, it will frequently be useful to consider the properties of the vector $u = w - \xi e$. With $\Upsilon(r) = R(r) - e$, we find that $E[u] = \xi\Upsilon(\rho)$.

We will use tildes to denote vectors and matrices which describe the model for all $T+1$ observations: Let $\tilde{y} = (y_0, y')'$, $\tilde{X} = \begin{pmatrix} 1 & 0 \\ e & X \end{pmatrix}$, $\tilde{e}$ a $(T+1) \times 1$ vector of ones, $\tilde{w} = (\xi, w)'$, $\tilde{R}(r) = (1, R(r))'$, $\tilde{u} = (0, u')'$, $\tilde{\Upsilon}(r) = \tilde{R}(r) - \tilde{e}$ and $\tilde{\beta} = (\mu, \beta')'$. We can then write $\tilde{y} = \tilde{X}\tilde{\beta} + \tilde{w}$, $E[\tilde{w}] = \xi\tilde{R}(\rho)$, and the variance-covariance matrix of $\tilde{w}$ is given by $\tilde{\Sigma}(\rho) = diag(0, \Sigma(\rho))$.

From a statistical perspective, the initial condition ξ is just an additional nuisance parameter besides the variance-covariance matrix of ν, β and μ. We are not primarily interested in its value, but we must be concerned about its impact on the Data Generating Process in order to construct useful tests and evaluate their performance.

Under the null hypothesis ($\rho = 1$) different values of ξ induce simple mean shifts in the data. Hence testing methods which are invariant to the mean μ will be invariant also to ξ. Indeed, in this model we cannot separately identify the mean μ and ξ. This follows as under the null hypothesis $\tilde{y} = \tilde{X}\tilde{\beta} + \xi\tilde{R}(1) + \tilde{w} = \tilde{X}\tilde{\beta} + \xi\tilde{e} + \tilde{w}$ and $\tilde{e}$ is in the column space of $\tilde{X}$, so that the estimated mean will estimate $(\mu + \xi)$. Under the alternative hypothesis ($\rho < 1$), however, altering ξ amounts to adding a geometrically decaying series $\Delta\xi\rho^t$ to the data. It is interesting to note that the exact form of the series $\Delta\xi\rho^t$ depends on ρ — the very value we are conducting inference on. Clearly invariance to the mean does not prevent statistics to change for different values of ξ. Statistics which are invariant to the mean are necessarily similar for ξ, but their distribution under the alternative and hence their power is affected by ξ.

It is long known that the power of unit root tests depends on the initial condition in small samples, see Evans and Savin (1981, 1984), for instance, and Stock (1994), p. 2777, for additional references. Typical asymptotic results, however, imply that the power of various test statistics remains unaffected by the initial condition for large samples. Typically, the asymptotic analyses assume ξ either fixed or random but bounded in probability. Under these assumptions, the asymptotic distributions of tests statistics become independent of ξ. The reason for this is that a Functional Central Limit Theorem applies to terms like

$$\frac{1}{\omega\sqrt{T}}w_t = \rho^t \frac{\xi}{\omega\sqrt{T}} + \frac{1}{\omega\sqrt{T}}\sum_{s=1}^{t}\rho^{t-s}\nu_s.$$

So long as ξ is bounded the term involving ξ is disappearing at rate $T^{1/2}$, whereas the second term remains.

But consider the unconditional variance of a stationary autoregressive process, which is proportional to $(1-\rho^2)^{-1}$. Useful asymptotics for the unit root testing problem require ρ to become ever closer to unity as the sample size T increases. Following the analysis of Chan and Wei (1987) and Phillips (1987b), the right rate of convergence of ρ to one is achieved by setting $\rho = 1-\gamma T^{-1}$ for a fixed γ. In this local-to-unity framework, $(1-\rho^2)^{-1} = T(2\gamma)^{-1} + o(T)$, so that the unconditional variance is of order T. An assumption of $\xi = O_p(1)$ therefore makes the initial condition an order of magnitude smaller than the square root of the unconditional variance of a comparable stationary process.

An appropriate way to capture the effects of the initial condition is hence to treat ξ as an $O_p(T^{1/2})$ variable. This is carried out for a special case in Elliott (1999), where ξ is assumed to stem from the unconditional distribution of a stationary series with $\rho = r = 1-gT^{-1}$ for a fixed g. This assumption of an initial observation being drawn from the unconditional distribution has also been popular in small sample Monte Carlo analyses of unit root tests.

So if ξ is a nuisance parameter of some impact, why not handling it the same way as the other nuisance parameters of the model? Two possible strategies come to mind.

A first approach is the 'plug-in' method. Assuming a certain distribution for ν_t, we could develop an optimal statistic for the hypothesis test for ξ known, and then plug-in an estimator $\hat{\xi}$. If it were possible to estimate ξ accurately enough (which means here up to a $o_p(T^{1/2})$ term), the asymptotic distributions of the optimal statistic for ξ known would remain unaffected by this replacement, and, at least asymptotically, we would not have foregone any power. Unfortunately, as we will show below, such an estimator of ξ does not exist, so that we lose the optimality of the procedure.

Alternatively, we could try to develop tests which are invariant to ξ even under the alternative. Invariance is a popular concept in the literature, and it is by appealing to this principle that the deterministic terms like the mean and the time trend are dealt with (both the common OLS detrending as well as GLS detrending yield tests that are numerically invariant to translations of the form of additive constants and time trends). But as mentioned above, the group of transformations

$$y_t \to y_t + \rho^t x \tag{2.3}$$

induced by different x depends on the true ρ.

From a conceptual point of view, the idea of invariance is that a test should be symmetric with respect to a certain group of translations, because these translations are not informative about the parameter of interest. This is clearly not the case here — the translation (2.3) depends on the parameter of interest ρ and is informative for ρ. If we had a very large ξ, there would be a pronounced arc in the data only if ρ were less than one, but not if it were equal to one. But an invariant test is precluded from exploiting this information. We conclude that invariance is not the correct concept for removing ξ.

Rather, we will explore an alternative way to deal with the nuisance parameter ξ. We will derive tests that maximize weighted average power over various values of ξ, where the weight function is a prespecified distribution function. Since tests which are invariant to the mean are unaffected by different values of ξ under the null, we only need to specify the weight function under the alternative. In this respect, the situation here is very similar to Andrews and Ploberger's (1994) analysis of optimal asymptotic tests for the general testing problem when a nuisance parameter is present only under the alternative. We cannot directly draw on their results, however, since several of their assumptions are not satisfied for the testing problem here.

3. A FAMILY OF OPTIMAL TESTS

3.1. **Small Sample Analysis.** In this section, we will develop optimal procedures for the hypothesis test above for Gaussian disturbances ν_t, where we assume that the nonsingular $T \times T$ variance-covariance matrix V of ν_t is known. This assumption is, of course, not likely to be met in practice. In section 5, however, we will show that it is possible to obtain the same asymptotic power functions with feasible tests which do not require such knowledge.

Our aim is to develop an optimal procedure for the unit root testing problem (2.2). For three reasons, a direct application of the Neyman-Pearson Lemma is not possible: (i) β and μ are unknown, (ii) the alternative is composite and (iii) there is an additional nuisance parameter ξ, that is individually identified only under the alternative.

To deal with the first problem, we will restrict attention to tests which are invariant to the group of transformations

$$\tilde{y} \to \tilde{y} + \tilde{X}\tilde{b} \; \forall \tilde{b}, \tag{3.1}$$

i.e. the requirement that a test statistic $T^*(\tilde{y})$ has the property $T^*(\tilde{y}+\tilde{X}\tilde{b}) = T^*(\tilde{y})$ for all $\tilde{b}$. This has been the dominant strategy in the unit root literature for the treatment of the unknown β and μ, and we will follow this approach. As already noted above, invariance to the mean also makes the test statistic automatically independent of ξ under the null of $\rho = 1$.

The composite nature of the alternative is indeed a problem for unit root testing. The value of ρ under the alternative enters the likelihood in such a way that there does not exist a uniformly most powerful test, even asymptotically (cf. Elliott, Rothenberg, and Stock (1996)). Dufour and King (1991) have thus derived small sample point optimal tests that maximize power at a specific alternative $\rho = \bar{\rho} < 1$, and Elliott, Rothenberg, and Stock (1996) have extended these results in a local-to-unity asymptotic framework. We will follow this lead in this paper.

Finally, in order to deal with the nuisance parameter ξ, we will derive tests which maximize a weighted average power criterion. Specifically, let $F(\xi)$ be a probability measure on ξ. We will refer to a test $\varphi_0(\tilde{y}; r, F)$ as an optimal test if for a given significance level α_0, $\varphi_0(\tilde{y}; r, F)$ maximizes weighted average power at the alternative $\rho = r < 1$

$$\int_{-\infty}^{\infty} P(\varphi(\tilde{y}) \text{ rejects } |\rho = r, \xi = x) dF(x) \tag{3.2}$$

over all tests $\varphi(\tilde{y})$ of size α_0. F may be seen as representing the importance a researcher attaches to the test being able to distinguish the two hypotheses for various values of ξ. In this perspective, the weighting F is a device to derive tests with a certain power characteristic as a function of ξ.

This treatment of a nuisance parameter is very similar to the approach of Andrews and Ploberger (1994) for the general testing problem where a nuisance parameter is present only under the alternative. Note, however, that their weighted average power criterion not only averages over various values of the nuisance parameter (here ξ), but also over various alternatives (here ρ). The criterion (3.2) puts all mass on a single alternative $\rho = r$. Conceptionally, it is easy to generalize (3.2) accordingly and to derive optimal tests in the larger class (cf. Müller (2002)). The reasons we stick with maximizing power at a single alternative $\rho = r$ in this paper are threefold: First of all, the resulting optimal tests are of a much simpler form, which in turn are easier to interpret. Also the main focus of the paper is on the effect of the initial observation on unit root testing, and we think that we serve this purpose best by keeping additional dimensions of the problem as simple as possible. Finally, it turns out that if the alternative r is chosen local-to-unity $r = 1 - g/T$, then the asymptotic properties of the optimal test are relatively insensitive to the value of g. An additional averaging over this parameter would thus change little in the properties of the tests.

The following Theorem is the most general result in this section. It provides an optimal procedure for (2.2) for general X with an arbitrary weighting function F for ξ.

Theorem 1. *Consider the Data Generating Process (2.1) when the disturbance vector ν is known to be multivariate Gaussian $N(0, V)$. Then the test of $H_0 : \rho = 1$ against $H_1 : \rho = r$ which rejects for small values of the statistic*

$$S(r, F) = \tilde{y}' [G_1 - G_0] \tilde{y} - 2 \ln \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} \left[x^2 \tilde{R}_1' G_1 \tilde{R}_1 - 2x \tilde{R}_1' G_1 \tilde{y} \right] \right\} dF(x)$$

maximizes (3.2) under all tests of the same size which are invariant to the transformations (3.1), where F is any cumulative distribution function, $G_i = \tilde{\Sigma}_i^- - \tilde{\Sigma}_i^- \tilde{X} (\tilde{X}' \tilde{\Sigma}_i^- \tilde{X})^{-1} \tilde{X}' \tilde{\Sigma}_i^-$, $\tilde{\Sigma}_i^- = \begin{pmatrix} 1 + e'\Sigma_i^{-1} e & -e'\Sigma_i^{-1} \\ -\Sigma_i^{-1} e & \Sigma_i^{-1} \end{pmatrix}$, $\Sigma_0 = \Sigma(1)$, $\Sigma_1 = \Sigma(r)$ and $\tilde{R}_1 = \tilde{R}(r)$.

Terms of the form $\tilde{a}' G_i \tilde{a}$ might be recognized as the weighted sum of squared residuals of a GLS regression of $\tilde{a}$ on $\tilde{X}$ using $\tilde{\Sigma}_i^-$ as weighting matrix.

The form of the optimal test where power is maximized at a single value $\xi = x$ follows from Theorem 1:

Corollary 1. *Under the assumptions of Theorem 1, the optimal test of* $H_0 : \rho = 1$ *against* $H_1 : \rho = r$ *which is invariant to the transformations (3.1) and maximizes power at the specific alternative* $\xi = x$ *rejects for small values of the statistic*

$$\text{Env}(r, x) = \tilde{y}' \left[G_1 - G_0\right] \tilde{y} + x^2 \tilde{R}_1' G_1 \tilde{R}_1 - 2x\tilde{R}_1' G_1 \tilde{y}.$$

At the point (r, x) in the alternative space, $\text{Env}(r, x)$ reaches the maximal power any test of the unit root hypothesis (2.2) can achieve. The family of tests $\text{Env}(r, x)$ therefore enables us to derive the power envelope of the testing problem in both the ξ and ρ dimension.

For the subsequent discussion, it turns out to be useful to focus on mean zero Gaussian weighting function F. From Theorem 1, we find from carrying out the integration

Corollary 2. *Under the assumptions of Theorem 1, the test of* $H_0 : \rho = 1$ *against* $H_1 : \rho = r$ *which is invariant to the transformations (3.1) and maximizes (3.2) with* F *being the cumulative density function of a zero mean normal with variance* λ *rejects for small values of the statistic*

$$Q(r, \lambda) = \tilde{y}' \left[G_1 - G_0\right] \tilde{y} - \frac{\lambda(\tilde{R}_1' G_1 \tilde{y})^2}{1 + \lambda \tilde{R}_1' G_1 \tilde{R}_1}.$$

As discussed above, the weighting function F may be seen as simple device to construct a family of optimal tests $\varphi_0(\tilde{y}; r, F)$ with a different power characteristic in the ξ dimension. Alternatively, one might interpret these tests in a Bayesian manner. The restriction to tests that are invariant to the transformations (3.1) implies that any test statistic must be a function of a maximal invariant (cf. Lehmann (1986), p. 285). It is easy to check that a maximal invariant for this group of transformations is given by $\tilde{z} = \tilde{M}\tilde{y}$, where $\tilde{M} = \tilde{I} - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'$ and $\tilde{I}$ is the $(T+1) \times (T+1)$ identity matrix. Denote with $f(\tilde{z}; r, x)$ the density of $\tilde{z}$ when $\rho = r$ and $\xi = x$ in the Gaussian model (2.1). Note that, since $\tilde{z}$ is independent of the value of ξ under the null of $\rho = 1$ ($\tilde{M}\tilde{e} = 0$), the density $f(\tilde{z}; 1, x)$ is independent of x. But now it is easy to establish that maximizing (3.2) is equivalent to finding the most powerful test of H_0 : the density of $\tilde{z}$ is given by $f(\tilde{z}; 1, 0)$ against the alternative H_1 : the density of $\tilde{z}$ is given by $\int_{-\infty}^{\infty} f(\tilde{z}; r, x) dF(x)$.

Theorem 1 above thus has an additional interpretation as providing a test statistic which optimally (subject to the invariance restriction) discriminates $H_0 : \rho = 1$ with arbitrary ξ against $H_1 : \rho = r$ and $\xi \sim F$, where ξ is a random variable independent of ν. As long as there is no stochastic dependence between ξ and ν, the optimal unit root test for any *distributional* assumption F on ξ under the single alternative $\rho = r < 1$ is the maximal weighted average power test based on $S(r, F)$. It is this second interpretation that makes our nonstochastic assumption of ξ much less restrictive than it looks.

In order to make the link between the two perspectives more explicit, consider how one would usually approach optimal unit root testing if ξ was assumed random and independent of ν. For a zero mean Gaussian ξ with variance λ, a test that maximizes power against the alternative $\rho = r$ has to optimally discriminate between two variance-covariance matrices in a normal linear model: With the $(T+1) \times (T+1)$

matrix

$$\tilde{A}(r) = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -r & 1 & 0 & \cdots & 0 & 0 \\ 0 & -r & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -r & 1 \end{pmatrix}$$

clearly $\tilde{A}(\rho)\tilde{w} = (\xi, \nu)'$, so that the variance-covariance matrix $\tilde{\Omega}(\rho)$ of $\tilde{w}$ satisfies $\tilde{A}(\rho)\tilde{\Omega}(\rho)\tilde{A}(\rho)' = diag(\lambda, V) \equiv \tilde{V}$. The form of the optimal statistic for this problem is usually written as

$$\hat{y}(r)'\tilde{\Omega}(r)^{-1}\hat{y}(r) - \hat{y}(1)'\tilde{\Omega}(1)^{-1}\hat{y}(1) \tag{3.3}$$

where $\hat{y}(c)$ is the residual vector of a GLS regression of $\tilde{y}$ on $\tilde{X}$, using variance-covariance matrices $\tilde{\Omega}(c)$. The equivalence between the distributional assumption on ξ and the weighting in Theorem 1 implies that tests based on (3.3) and $Q(r, \lambda)$ should be identical. Some matrix algebra shows that $Q(r, \lambda)$ can indeed be rewritten in the form of (3.3). See Appendix for details.

Dufour and King's (1991) Point Optimal Invariant statistics are also very much related to $Q(r, \lambda)$. They consider the special case where $V = \sigma^2 I$, but impose invariance to the larger group of transformations of the form $\tilde{y} \to a\tilde{y} + \tilde{X}\tilde{b}$ for any positive a and all vectors $\tilde{b}$. The additional invariance to scale makes the resulting tests independent of σ^2. The test statistic for this problem is given by the ratio rather than difference of the weighted sum of squared residuals in (3.3). We focus in this paper on an asymptotic analysis, and since σ^2 can be estimated consistently, the formulation of Dufour and King (1991) and (3.3) lead to the same asymptotic power functions.

When ν_t are independent with variance σ^2 and λ is chosen to be $\sigma^2/(1-r^2)$, then $\tilde{w}$ becomes stationary for $\rho = r$. $Q(r, \sigma^2/(1-r^2))$ is hence the most powerful invariant test of the unit root hypothesis against the stationary alternative with $\rho = r$. When ν_t is stationary and autocorrelated, however, a random ξ that makes $\tilde{w}$ stationary under the alternative cannot be stochastically independent of ν_t. The optimal test statistic in this case is hence not member of the family $Q(r, \lambda)$. The following Theorem provides the optimal test for this case.

Theorem 2. *Consider the Data Generating Process (2.1) when $\{\nu_t\}_{-\infty}^{\infty}$ is a stationary Gaussian process with known autocovariances $E[\nu_t \nu_{t-k}] = \gamma(k)$ and assume that $\sum_{k=0}^{\infty} |\gamma(k)| < \infty$. The statistic that optimally tests $H_0 : \rho = 1$ against $H_1 : \rho = r$ and $\xi = \sum_{s=0}^{\infty} r^s \nu_{-s}$ which is invariant to the transformations (3.1) and rejects for small values is*

$$\bar{Q}(r) = \tilde{y}'\bar{J}_1\tilde{y} - \tilde{y}'\bar{J}_0\tilde{y},$$

where $\bar{J}_0 = G_0$, $\bar{J}_1 = \bar{\Omega}_1^{-1} - \bar{\Omega}_1^{-1}\tilde{X}(\tilde{X}'\bar{\Omega}_1^{-1}\tilde{X})^{-1}\tilde{X}'\bar{\Omega}_1^{-1}$, $\tilde{A}(r)\bar{\Omega}_1\tilde{A}(r)' = \bar{V}$, $\bar{V} = \begin{pmatrix} v_0 & \eta' \\ \eta & V \end{pmatrix}$, $v_0(r) = \mathrm{Var}\left[\sum_{s=0}^{\infty} r^s \nu_{-s}\right]$ *and the* $T \times 1$ *vector* η *is given by* $\eta = [\eta_t] = \left[\sum_{i=0}^{\infty} r^i \gamma(t+i)\right]$.

Whilst in small samples there is a distinction between this test and the family $Q(r, 1)$, we show in the next section that this distinction is irrelevant asymptotically.

3.2. **Asymptotic Analysis.** The subsequent discussion of these tests and their relationship with other known unit root tests will concentrate on their asymptotic distributions. For the asymptotic theory, we require a somewhat stronger condition than a known nonsingular variance-covariance matrix of the Gaussian disturbances ν_t.

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

The asymptotics are developed in the local-to-unity framework introduced above, i.e. we investigate the limiting distribution of the test statistics as the sample size T goes to infinity and $\gamma = T(1-\rho) \geq 0$ is a fixed constant. The natural alternative r in the optimal family of tests based on $S(r, F)$ is then also of the form $r = 1 - gT^{-1}$ for fixed $g > 0$. Furthermore, as mentioned in Section 2, we will investigate the case that the initial condition has the same order of magnitude as the standard deviation of the unconditional distribution of the stationary process with $\rho = 1 - \gamma T^{-1}$, $\gamma > 0$, which is given by $v_0(\rho)^{1/2} = \omega T^{1/2}(2\gamma)^{-1/2} + o(T^{1/2})$, where $v_0(\rho)$ is defined in Theorem 2 and ω^2 is the 'long-run' variance of ν_t, $\omega^2 = 2\pi f_\nu(0)$. Denote by α the such scaled version of the initial condition, $\alpha = \xi\omega^{-1}T^{-1/2}(2\gamma)^{1/2}$, so that $\xi = O(T^{1/2})$ has the right order of magnitude to matter asymptotically. In order to simplify notation, we introduce 'asymptotic' versions $S_a(g, F_a)$, $\mathrm{Env}_a(g, a)$, $Q_a(g, k)$ and $\bar{Q}_a(g)$ of the tests above. They are simple reparameterizations of the original tests, where $g = T(1-r)$, $a = xT^{-1/2}\omega^{-1}(2g)^{1/2}$, $k = 2gT^{-1}\lambda\omega^{-2}$ and F_a describes the weighting function in terms of α rather than ξ. Given the equivalence between the maximal weighted average power tests and optimal tests for a certain distributional assumption on ξ, the case where F_a is given by the cumulative density function of a standard normal random variable corresponds to Elliott's (1999) analysis, and a degenerate F_a which puts all mass on $\alpha = 0$ corresponds to the case considered by Elliott, Rothenberg, and Stock (1996).

For the asymptotic distributions, we extensively use the following version of a Functional Central Limit Theorem:

Lemma 1. *Let w_t be generated by the Data Generating Process (2.1). Under condition 1, $T(1-\rho) = \gamma \geq 0$ fixed and, when $\gamma > 0$, $\xi = \alpha\omega(2\gamma)^{-1/2}T^{1/2}$, as $T \to \infty$*

$$T^{-1/2}u_{[Ts]} \Rightarrow \begin{cases} \omega W(s) \text{ for } \gamma = 0 \\ \omega\alpha(e^{-\gamma s} - 1)(2\gamma)^{-1/2} + \omega \int_0^s e^{-\gamma(s-\lambda)} dW(\lambda) \text{ else} \end{cases}$$
$$\equiv \omega M(s)$$

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. Further, $M(s)$ is continuous at $\gamma = 0$.

The asymptotic distributions of $S_a(g, F_a)$, $\mathrm{Env}_a(g, x)$, $Q_a(g, k)$ and $\bar{Q}_a(g)$ for general deterministics (subject to a smoothness condition) can be found in Theorem 8 in the Appendix. In the main text, we restrict ourself to the two most popular cases: the mean only case without X, which will be denoted by a superscript μ, and the mean and trend case $X = \tau \equiv (1, 2, \cdots, T)'$, denoted with a superscript τ.

For the time trend case, it is useful to write the asymptotic distributions in terms of the asymptotic projection of $M(s)$ off s, denoted $M^\tau(s)$. It is the weak limit of the residuals u^τ from an OLS regression of u on τ, i.e. $T^{-1/2}u^\tau_{[Ts]} \Rightarrow M(s) - 3s\int \lambda M(\lambda)d\lambda \equiv M^\tau(s)$. The following Theorem states the asymptotic distributions of $S^i_a(g,F_a)$, $\mathrm{Env}^i_a(g,x)$, $Q^i_a(g,k)$ and $\bar{Q}_a(g)$ in terms of $M^\mu(s) \equiv M(s)$ and $M^\tau(s)$. For notational convenience, the limits of integration are understood to be zero and one, if not stated otherwise.

Theorem 3. *Under the conditions of Lemma 1, for $i = \mu, \tau$*
(i)

$$S^i_a(g,F_a) \;\Rightarrow\; s^i_0 + s^i_1 M^i(1)^2 + s^i_2\int M^i(s)^2 ds$$
$$-2\ln\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}\left[s^i_3 a\int M^i(s)ds + s^i_4 a M^i(1) + s^i_5 a^2\right]\right\}dF_a(a)$$

where $s^\mu_0 = -g$, $s^\mu_1 = g$, $s^\mu_2 = g^2$, $s^\mu_3 = \sqrt{2}g^{3/2}$, $s^\mu_4 = (2g)^{1/2}$, $s^\mu_5 = g/2$ *and* $s^\tau_0 = -g$, $s^\tau_1 = g^2(1+g)/(3+3g+g^2)$, $s^\tau_2 = g^2$, $s^\tau_3 = \sqrt{2}g^{3/2}$, $s^\tau_4 = -g^{3/2}(3+g)/[\sqrt{2}(3+3g+g^2)]$, $s^\tau_5 = g^3/[8(3+3g+g^2)]$
(ii)

$$\mathrm{Env}^i_a(g,a) \Rightarrow s^i_0 + s^i_1 M^i(1)^2 + s^i_2\int M^i(s)^2 ds$$
$$+\, s^i_3 a\int M^i(s)ds + s^i_4 a M^i(1) + s^i_5 a^2$$

(iii)

$$Q^i_a(g,k) \Rightarrow q^i_0 + q^i_1 M^i(1)^2 + q^i_2\left(\int M^i(s)ds\right)^2$$
$$+\, q^i_3 M^i(1)\int M^i(s)ds + q^i_4\int M^i(s)^2 ds$$

where $q^\mu_0 = -g$, $q^\mu_1 = g - gk/(2+gk)$, $q^\mu_2 = -g^3k/(2+gk)$, $q^\mu_3 = -2g^2k/(2+gk)$, $q^\mu_4 = g^2$ *and*
$q^\tau_0 = -g$, $q^\tau_1 = (8g^2+8g^3-3g^3k+g^4k)/(24+24g+8g^2+g^3k)$, $q^\tau_2 = -4g^3(3+3g+g^2)k/(24+24g+8g^2+g^3k)$, $q^\tau_3 = 4g^3(3+g)k/(24+24g+8g^2+g^3k)$, $q^\tau_4 = g^2$
(iv) $\bar{Q}^i_a(g)$ follows the same asymptotic distribution as $Q^i_a(g,1)$.

It is at this point that we can conclude that the 'plug-in' method for ξ mentioned in Section 2 does not yield the optimal test independent of ξ. Any 'point-point' optimal test based on $\mathrm{Env}_a(g,a)$ is admissible, but its asymptotic distribution depends on the value of a. If there was an estimator $\hat{\alpha}$ of α such that $\mathrm{Env}_a(g,\hat{\alpha})$ was asymptotically equivalent to $\mathrm{Env}_a(g,\alpha)$, then clearly $\mathrm{Env}_a(g,\hat{\alpha})$ would dominate $\mathrm{Env}_a(g,a)$ for all values of $\alpha \neq a$. Therefore such an $\hat{\alpha}$ cannot exist.

Noting that $\xi = y_0 - \mu$, a natural estimator for α is $T^{-1/2}\omega^{-1}(2\gamma)^{1/2}(y_0 - \hat{\mu}^i)$, where $\hat{\mu}^i$ is some estimator of μ. Let $\hat{\alpha}^i_{GLS}$ be this estimator when μ is estimated by a GLS regression under the fixed alternative $\rho = r = 1 - gT^{-1}$, using $\tilde{\Omega}(r)$ as the variance-covariance matrix. Substituting α in $\mathrm{Env}^i_a(g,\alpha)$ with $\hat{\alpha}^i_{GLS}$, we find that the resulting test is asymptotically equivalent to $Q^i_a(g,\tilde{k}^i)$, where $\tilde{k}^\mu = 2k + k^2g/2$ and $\tilde{k}^\tau = 2k + k^2g^3/[8(3+3g+g^2)]$. The 'plug-in' method hence does yield optimal tests, even though their interpretation in terms of $Q^i_a(g,k)$ seems more straightforward. Similarly, letting $\hat{\alpha}^i_{ML}$ be the Maximum Likelihood estimator of

α under the fixed alternative with $\rho = 1 - gT^{-1}$ yields $\mathrm{Env}_a^i(g, \hat{\alpha}_{ML}^i)$ which are equivalent to $Q_a^i(g, \infty)$. See Appendix for details.

The statistics of Theorem 3 were all computed with the knowledge of the variance-covariance matrix V of ν_t. But their asymptotic distributions do not depend on the specific form of the autocorrelations of ν_t. This — maybe surprising — result has already been established by Elliott, Rothenberg, and Stock (1996) for the statistic $Q(g, 0)$ in our notation. It carries over to more general assumptions concerning the initial condition, as well as to the optimal statistic against the stationary model $\bar{Q}_a(g)$. The result implies that it is impossible to exploit autocorrelations in ν_t to devise unit root tests which have higher asymptotic local power than optimal tests for independent ν_t.

As a last step in the analysis of this section, we derive the characteristic functions of the asymptotic distributions of $\mathrm{Env}_a^i(g, a)$ and $Q_a^i(g, k)$, following the exposition of Tanaka (1996). The characteristic functions will enable us to calculate critical values and the power function with a much higher precision compared to Monte Carlo methods. Note that $\mathrm{Env}_a^i(g, x)$ and $Q_a^i(g, k)$ are a weighted sum of functionals of $M(s)$. The desired characteristic functions are therefore easily derived once the following Lemma is established:

Lemma 2. *Let* $Z^\mu = (M(1), \int M(s)ds)'$ *and* $Z^\tau = (Z^{\mu\prime}, \int sM(s)ds)'$ *and define* $V^i(\gamma) = E[Z^i Z^{i\prime}]$, *where* γ *is the non-negative parameter in the definition of* $M(s)$ *in Lemma 1. Let* $T^i = l_0^i + l_1^i \int M(s)^2 ds + Z^{i\prime}\Lambda^i Z^i + \lambda^{i\prime} Z^i$, *where the symmetric matrix* Λ^i, *the vector* λ^i, l_0^i *and* l_1^i *are nonstochastic. Then the characteristic function of* T^i *for* $i = \mu, \tau$ *is given by*

$$\phi^i(\theta) = \left| I - 2V^i(\delta)\tilde{\Lambda}^i \right|^{-1/2} \exp\left\{ \tilde{l}_0^i + \frac{1}{2}\tilde{\lambda}^{i\prime}(V^i(\delta)^{-1} - 2\tilde{\Lambda}^i)^{-1}\tilde{\lambda}^i - \frac{1}{4}\alpha^2\gamma \right\},$$

where $\delta = \sqrt{\gamma^2 - 2l_1^i\theta\mathbf{i}}$, $\tilde{l}_0^i = \theta l_0^i \mathbf{i} - \frac{1}{2}(\delta - \gamma)$, $\tilde{\Lambda}^\mu = \theta\Lambda^\mu\mathbf{i} + diag((\delta - \gamma)/2, 0)$, $\tilde{\Lambda}^\tau = \theta\Lambda^\tau\mathbf{i} + diag((\delta - \gamma)/2, 0, 0)$, $\tilde{\lambda}^\mu = \theta\lambda^\mu\mathbf{i} - \alpha\frac{\sqrt{2}}{2}(\gamma^{1/2},\ \gamma^{3/2})'$ *and* $\tilde{\lambda}^\tau = \theta\lambda^\tau\mathbf{i} - \alpha\frac{\sqrt{2}}{2}(\gamma^{1/2},\ \gamma^{3/2}, 0)'$.

The variance-covariance matrices $V^i(\gamma)$ follow after some stochastic calculus and are given in the Appendix.

Figure 1 depicts the asymptotic power of $Q_a^i(g, k)$ for various g and $k = 0, 1$ with $\gamma = 5, 10, 15$, 20 and 25 as a function of α. All power curves in this paper are for a level of 5%. For $k = 0$, the values $g = 7$ and $g = 13.5$ are those suggested by Elliott, Rothenberg, and Stock (1996), and for the case $k = 1$ Elliott (1999) suggested using $g = 10$ and $g = 15$. For large enough $|\alpha|$, the power of all considered tests drops to zero. The tests with $k = 0$ achieve the maximal power at $\alpha = 0$ (in fact, $Q_a(g, 0) = \mathrm{Env}_a(g, 0)$), but their power drops to zero for $|\alpha| > 2$ in the mean case and for $|\alpha| > 3$ in the trend case for all considered values of γ. The tests with $k = 1$ have an asymptotic power which decreases in $|\alpha|$, too, but at a considerably slower rate than the tests with $k = 0$. A comparison with the asymptotic power for the larger values of g show that the test performance is relatively insensitive to this parameter. For $k = 1$, larger g seem to moderately increase the power for large $|\alpha|$, at the cost of lower power for small values of $|\alpha|$.

Figure 2 shows the asymptotic power envelopes, derived by the test $\mathrm{Env}_a(g, a)$. Clearly, the larger the *known* initial condition, the easier it becomes to distinguish the two competing hypotheses. It is interesting to note that the difference between

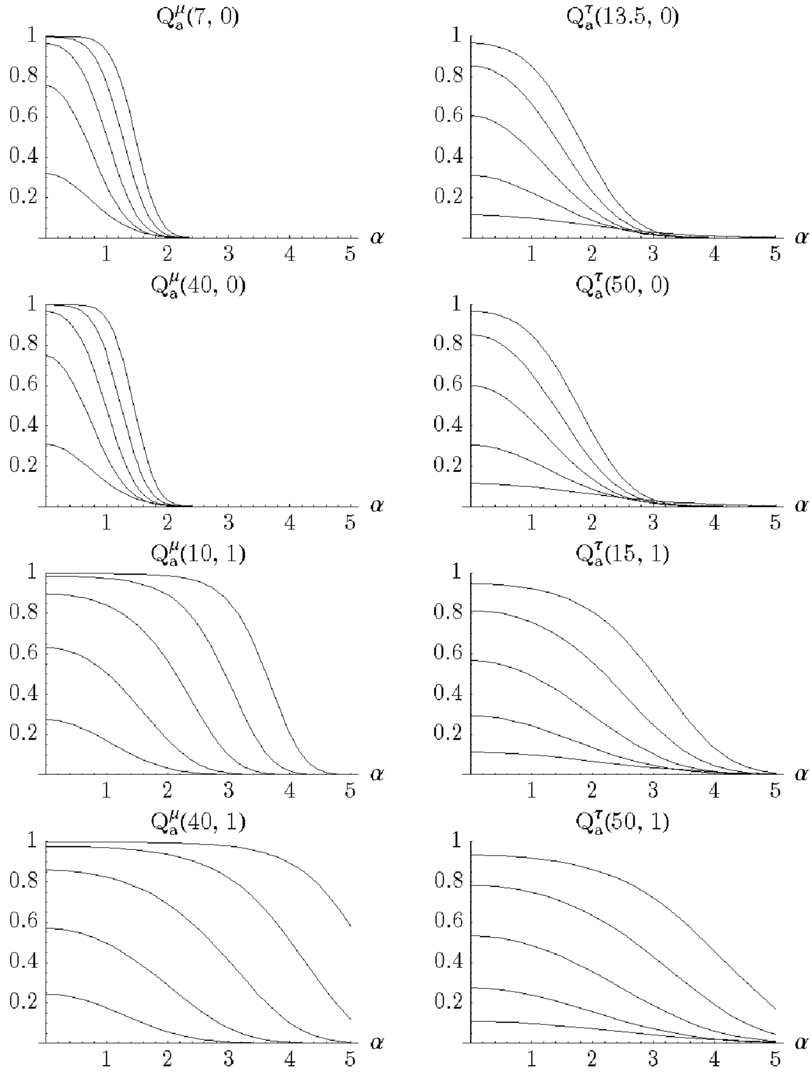

FIGURE 1. Asymptotic power as a function of α for $\gamma = 5$, 10, 15, 20 and 25

the power envelopes for the mean only case and the time trend case is very pronounced. The intuitive explanation for this phenomenon is that the arc generated by a non-zero α for moderate γ is similar to a time trend. Invariance to the time trend of $Q_a^{\tau}(g, k)$ makes it difficult to distinguish the two possible origins, and hence the smaller power.

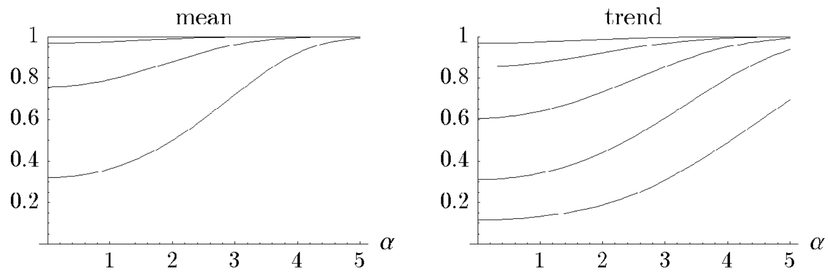

FIGURE 2. Asymptotic power envelope for $\gamma = 5$, 10, 15, 20 and 25

4. RELATION TO SOME POPULAR UNIT ROOT TESTS

In this section, we will explore the relationship of the optimal test based on $Q_a^i(g,k)$ with other unit root tests suggested in the literature. Following Stock (2000), we classify unit root test statistics by their asymptotic distributions, written as some function $h : C[0,1] \longmapsto \mathbb{R}$ of $M(\cdot)$, where $C[0,1]$ is the space of square integrable continuous functions on the unit interval. Various tests lead to a multitude of functions h, and many other tests could be devised. However, for most classes of tests and resulting functions h nothing is known about their optimality.

The asymptotic distribution of $Q_a^i(g,k)$ in Theorem 3 allows us to shed some light on the optimality properties of many popular unit root tests. The classes of tests considered in this paper fall into four categories: First, classes of tests which are asymptotically equivalent to an optimal test based on $Q_a^i(g_0,k_0)$ for a particular choice of g_0 and k_0. Second, the classes of statistics $\hat{\rho}^{\mathrm{DF}}$ and $\hat{\tau}^{\mathrm{DF}}$ are not directly member of the family $Q_a^i(g,k)$, but they have a close relationship to the test statistics $\breve{Q}_a^i(g,k)$ which optimally discriminate $\rho < 1$ in (2.1) against a different null model — see Theorems 5 and 6 below. Third, classes of tests whose asymptotic distribution is a function of $M^i(1)^2$, $\left(\int M^i(s)ds\right)^2$, $M^i(1)\int M^i(s)ds$ and $M^i(s)^2ds$, just as the $Q_a^i(g,k)$ statistics, but no exact equivalence prevails. For these tests we are still able to identify particular values g_0 and k_0 such that a test based on $Q_a^i(g_0,k_0)$ has asymptotic power as a function of γ and α which is very much comparable. The value of k_0 in these correspondences reveals the implicit weight the classes of tests put on α. Finally, some classes of tests have asymptotic distributions which depend on $M(\cdot)$ through other functions than the asymptotic distribution of the optimal tests. These tests have low power over a wide range of values for γ and α.

Table 1 shows the classes of tests considered in this paper. Members of the $\hat{\rho}^{\mathrm{DF}}$-class and $\hat{\tau}^{\mathrm{DF}}$-class include the statistic suggested by Dickey and Fuller (1979) as well as those of Phillips (1987a) and Phillips and Perron (1988), members of the N-class and R-class include the (appropriately scaled) N_1, N_2 and R_1, R_2 statistics of Bhargava (1986) as well as the t-statistics suggested by Schmidt and Phillips (1992) and Schmidt and Lee (1991), members of the LB-class include the (appropriately scaled) Locally Best Invariant test for the mean case as derived in Dufour and King (1991) and the Locally Best Unbiased Invariant test for the trend case as derived by Nabeya and Tanaka (1990), members of the P_T-class, $\hat{\rho}^{\mathrm{DFGLS}}$-class and

TABLE 1. Classes of Unit Root Tests

#	class	asymptotic distribution
1	N	$\left[\int M^{i,N}(s)^2 ds\right]^{-1}$
2	R	$\left[\int M^{i,R}(s)^2 ds\right]^{-1}$
3	LB	$\begin{cases} M(1)^2 & \text{mean case} \\ \int M^{\tau,N}(s)^2 ds & \text{trend case} \end{cases}$
4	$P_T(\bar{c})$	$\bar{c}M^{i,P}(1)^2 + \bar{c}^2 \int M^{i,P}(s)^2 ds$
5	$Q_T(\bar{c})$	same asymptotic distribution than $Q_a(\bar{c},1) + \bar{c}$
6	$\hat{\rho}^{\text{DFGLS}}(\bar{c})$	$\dfrac{M^{i,P}(1)^2 - M^{i,P}(0)^2 - 1}{2\int M^{i,P}(s)^2 ds}$
7	$\hat{\rho}^{\text{DF}}$	$\dfrac{M^{i,\text{OLS}}(1)^2 - M^{i,\text{OLS}}(0)^2 - 1}{2\int M^{i,\text{OLS}}(s)^2 ds}$
8	$\hat{\rho}^{WS}$	$\dfrac{M^{i,\text{OLS}}(1)^2 + M^{i,\text{OLS}}(0)^2 - 1 - 2\int M^{i,\text{OLS}}(s)^2 ds}{2\int M^{i,\text{OLS}}(s)^2 ds}$
9	$\hat{\tau}^{\text{DFGLS}}$	$\dfrac{M^{i,P}(1)^2 - M^{i,P}(0)^2 - 1}{2\sqrt{\int M^{i,P}(s)^2 ds}}$
10	$\hat{\tau}^{WS}$	$\dfrac{M^{i,\text{OLS}}(1)^2 + M^{i,\text{OLS}}(0)^2 - 1 - 2\int M^{i,\text{OLS}}(s)^2 ds}{2\sqrt{\int M^{i,\text{OLS}}(s)^2 ds}}$
11	$\hat{\tau}^{Ley}$	$\dfrac{\lvert M^{i,\text{OLS}}(1)^2 - M^{i,\text{OLS}}(0)^2\rvert - 1}{2\sqrt{\int M^{i,\text{OLS}}(s)^2 ds}}$
12	$\hat{\tau}^{\text{DF}}$	$\dfrac{M^{i,\text{OLS}}(1)^2 - M^{i,\text{OLS}}(0)^2 - 1}{2\sqrt{\int M^{i,\text{OLS}}(s)^2 ds}}$
13	R/S	$\sup_{s\in(0,1)} M^{i,\text{OLS}}(s) - \inf_{s\in(0,1)} M^{i,\text{OLS}}(s)$
14	J	$\begin{cases} \left[\int M^{\tau,\text{OLS}}(s)^2 ds\right]^{-1}\left[\int M^{\mu,\text{OLS}}(s)^2 ds\right] & \text{mean case} \\ \left[\int M^{q,\text{OLS}}(s)^2 ds\right]^{-1}\left[\int M^{\tau,\text{OLS}}(s)^2 ds\right] & \text{trend case} \end{cases}$

with $M^{\mu,\text{OLS}}(s) = M(s) - \int M(\lambda)d\lambda$, $M^{\tau,\text{OLS}}(s) = M^{\tau}(s) - 4\int M^{\tau}(\lambda)d\lambda + 6s\int M^{\tau}(\lambda)d\lambda$, $M^{q,\text{OLS}}(s) = M(s) - 3(3-12s+10s^2)\int M^{\tau}(\lambda)d\lambda - 30(1-6s+6s^2)\int \lambda^2 M^{\tau}(\lambda)d\lambda$, $M^{\mu,N}(s) = M(s)$, $M^{\tau,N}(s) = M^{\tau}(s) - sM^{\tau}(1)$, $M^{\mu,R}(s) = M^{\mu,\text{OLS}}(s)$, $M^{\tau,R}(s) = M^{\tau}(s) - (s-\frac{1}{2})M^{\tau}(1) - \int M^{\tau}(\lambda)d\lambda$, $M^{\mu,P}(s) = M(s)$ and $M^{\tau,P}(s) = M^{\tau}(s) - s(\bar{c}+1)(\frac{1}{3}\bar{c}^2+\bar{c}+1)^{-1}M^{\tau}(1)$, $\bar{c} > 0$.

$\hat{\tau}^{\text{DFGLS}}$-class, indexed by a positive parameter $\bar{c}$, include the statistics proposed in Elliott, Rothenberg, and Stock (1996), members of the Q_T-class, also indexed by the parameter $\bar{c}$, include the statistic suggested by Elliott (1999), members of the $\hat{\rho}^{\text{WS}}$- and $\hat{\tau}^{\text{WS}}$-class include the Weighted Symmetric Estimator of Pantula, Gonzalez-Farias, and Fuller (1994), members of the $\hat{\tau}^{Ley}$-class include the statistic suggested by Leybourne (1995), members of the R/S-class include the range statistic suggested by Mandelbrot and Ness (1968), and members of the J-class include Park's (1990) variable addition tests $J(0,1)$ in the mean case and $J(1,2)$ in the trend case where appropriate corrections for correlated disturbances are employed for all test statistics. See Stock (1994) for details regarding these corrections. As already pointed out by Nabeya and Tanaka (1990), note that the asymptotic distribution of the class of the Locally Best tests LB^{τ} in the trend case is a monotonic transformation of the asymptotic distribution of the N^τ-class, so that tests based on these statistics are asymptotically equivalent.

In the following derivations, we assume that the optimal statistics of Theorem 3 and the members of the classes of test statistics in Table 1 converge jointly to their respective weak limits; it is the same $M(s)$ that describes all asymptotic distributions. It can be shown using the Continuous Mapping Theorem that this assumption is fulfilled for all standard members of the classes described above — see Müller (2002).

We first show that tests based on the classes of statistics 1–8 are equivalent to tests based on statistics with an asymptotic distribution that closely parallels the asymptotic distribution of $Q_a(g,k)$. To this end, note that the detrended $M^{i,m}(s)$ for the classes of statistics 1–8 can be expressed as $M^{i,m}(s) = M^i(s) + a_1^{i\prime} C^i + s a_2^{i\prime} C^i$, where $C^i = (\int M^i(\lambda)d\lambda, M^i(1))'$, a_j^i are 2×1 vectors of constants and $a_2^\mu = 0$. This makes it possible to express their asymptotic distributions in terms of a constant, $M^i(1)^2$, $\left(\int M^i(s)ds\right)^2$, $M^i(1)\int M^i(s)ds$ and $\int M^i(s)^2 ds$, since

$$\begin{aligned} M^{i,m}(0)^2 &= (a_1^{i\prime} C^i)^2 \\ M^{i,m}(1)^2 &= [(a_1^i + a_2^i + d_2)' C^i]^2 \\ \int M^{i,m}(s)^2 ds &= \int M^i(s)^2 ds + C^{i\prime}[2a_1^i d_1' + a_1^i a_1^{i\prime} + a_1^i a_2^{i\prime} + \tfrac{1}{3} a_2^i a_2^{i\prime}]C^i \end{aligned}$$

where $d_1 = (1,0)'$ and $d_2 = (0,1)'$. Moreover, note that many of the classes of statistics have asymptotic distributions of the form A/B, where $B > 0$ (at least with probability 1). But $A/B \leq \text{cv}$ if and only if $A - \text{cv}\, B \leq 0$, where cv stands for the asymptotic critical value used in the test. With these insights, it is straightforward to show that tests based on statistics in the classes 1–8 with asymptotic critical value cv are asymptotically equivalent to tests based on statistics with asymptotic distribution[2]

$$(4.1)\quad \lambda_0^i + \lambda_1^i M^i(1)^2 + \lambda_2^i \left(\int M^i(s)ds\right)^2 + \lambda_3^i M^i(1) \int M^i(s)ds + \lambda_4^i \int M^i(s)^2 ds$$

that reject for negative values, where some of the weights λ_j^i depend on the critical value cv.

The weights λ_j^i for the various classes of unit root tests are given in Tables 2 and 3. Note that the asymptotic distribution of $Q_a^i(g,k)$ has the same form (4.1), a weighted sum of a constant, $M^i(1)^2$, $\left(\int M^i(s)ds\right)^2$, $M^i(1)\int M^i(s)ds$ and $\int M^i(s)^2 ds$. From the joint convergence of the classes of statistics 1–8 in Table 1 and $Q_a^i(g,k)$ to their respective weak limits, rejection or nonrejection in the limit as $T \to \infty$ of tests based on these statistics can hence be described in terms of inequalities of weighted sums (4.1). If for a certain choice of g and k these inequalities turn out to be identical, then the tests are asymptotically equivalent.

Consider a test based on a statistic which is member of a class 1–8 in Table 1. Suppose that the test is carried out at a level α_0, which implies an asymptotic critical value $\text{cv}(\alpha_0)$. From the discussion above, the test is asymptotically equivalent to a test based on a statistic T_λ with asymptotic distribution (4.1) and that

[2] There is a zero probability set of events for which the statement is not true (for example, if any of the denominators becomes zero). All similar statements in the sequel are understood with this measure theoretic qualifier.

TABLE 2. Weights in the alternative description (4.1) of the asymptotic critical region of some classes of unit root tests – mean case

	λ_0^μ	λ_1^μ	λ_2^μ	λ_3^μ	λ_4^μ
R	-1	0	$-\text{cv}$	0	cv
N	-1	0	0	0	cv
LB	$-\text{cv}$	1	0	0	0
$\hat\rho^{\text{DF}}$	$-\frac{1}{2}$	$\frac{1}{2}$	cv	-1	$-\text{cv}$
$\hat\rho^{\text{DFGLS}}$	$-\frac{1}{2}$	$\frac{1}{2}$	0	0	$-\text{cv}$
$\hat\rho^{\text{WS}}$	$-\frac{1}{2}$	$\frac{1}{2}$	$2+\text{cv}$	-1	$-\text{cv}-1$
P_T	$-\text{cv}$	$\bar c$	0	0	$\bar c^2$
Q_T	$-\text{cv}$	$\frac{\bar c(1+\bar c)}{2+\bar c}$	$-\frac{\bar c^3}{2+\bar c}$	$-\frac{2\bar c^2}{2+\bar c}$	$\bar c^2$
$Q_a(g,k)$	$q_0^\mu - \text{cv}$	q_1^μ	q_2^μ	q_3^μ	q_4^μ

TABLE 3. Weights in the alternative description (4.1) of the asymptotic critical region of some classes of unit root tests – trend case

	λ_0^τ	λ_1^τ	λ_2^τ	λ_3^τ	λ_4^τ
R	-1	$\frac{1}{12}\,\text{cv}$	$-\text{cv}$	cv	cv
N	-1	$\frac{1}{3}\,\text{cv}$	0	0	cv
LB	$-\text{cv}$	$\frac{1}{3}$	0	0	1
$\hat\rho^{\text{DF}}$	$-\frac{1}{2}$	$\frac{1}{2}$	$-6+4\,\text{cv}$	2	$-\text{cv}$
$\hat\rho^{\text{DFGLS}}$	$-\frac{1}{2}$	$\frac{\bar c^4 - 6\,\text{cv} - 12\bar c\ \text{cv} - 6\bar c^2\,\text{cv}}{2(3+3\bar c+\bar c^2)^2}$	0	0	$-\text{cv}$
$\hat\rho^{\text{WS}}$	$-\frac{1}{2}$	$\frac{1}{2}$	$14+4\,\text{cv}$	2	$-1-\text{cv}$
P_T	$-\text{cv}$	$\frac{\bar c^2(1+\bar c)}{3+3\bar c+\bar c^2}$	0	0	$\bar c^2$
Q_T	$-\text{cv}$	$\frac{\bar c^2(\bar c^2+5\bar c+8)}{\bar c^3+8\bar c^2+24\bar c+24}$	$-\frac{4\bar c^3(\bar c^2+3\bar c+3)}{\bar c^3+8\bar c^2+24\bar c+24}$	$\frac{4\bar c^3(\bar c+3)}{\bar c^3+8\bar c^2+24\bar c+24}$	$\bar c^2$
$Q_a(g,k)$	$q_0^\tau - \text{cv}$	q_1^τ	q_2^τ	q_3^τ	q_4^τ

rejects for negative values. Now suppose that the weight vector $(\lambda_1^i, \lambda_2^i, \lambda_3^i, \lambda_4^i)'$ of the statistic satisfies

$$
\begin{pmatrix} \lambda_1^i \\ \lambda_2^i \\ \lambda_3^i \\ \lambda_4^i \end{pmatrix} = l_0 \begin{pmatrix} q_1^i(g_0,k_0) \\ q_2^i(g_0,k_0) \\ q_3^i(g_0,k_0) \\ q_4^i(g_0,k_0) \end{pmatrix} \tag{4.2}
$$

for some $l_0 > 0$, g_0 and k_0, where q_j^i are defined in Corollary 3. Apart from a constant, the statistics T_λ and $l_0 Q_a^i(g_0,k_0)$ then converge jointly to the same asymptotic distribution. It hence suffices to show (4.2) for some $l_0 > 0$, g_0 and k_0 in order to establish the asymptotic equivalence of a test based on a statistic in the classes 1–8 of Table 1 with an optimal test based on $Q_a^i(g_0,k_0)$.

When solving for an asymptotically equivalent test, the four nonlinear equations (4.2) must be satisfied by the three unknowns g_0, k_0 and l_0. Therefore, it is not surprising that not all classes of tests in Tables 2 and 3 are asymptotically equivalent to a particular $Q_a^i(g_0,k_0)$. But for classes 1–6, equivalence can indeed be established.

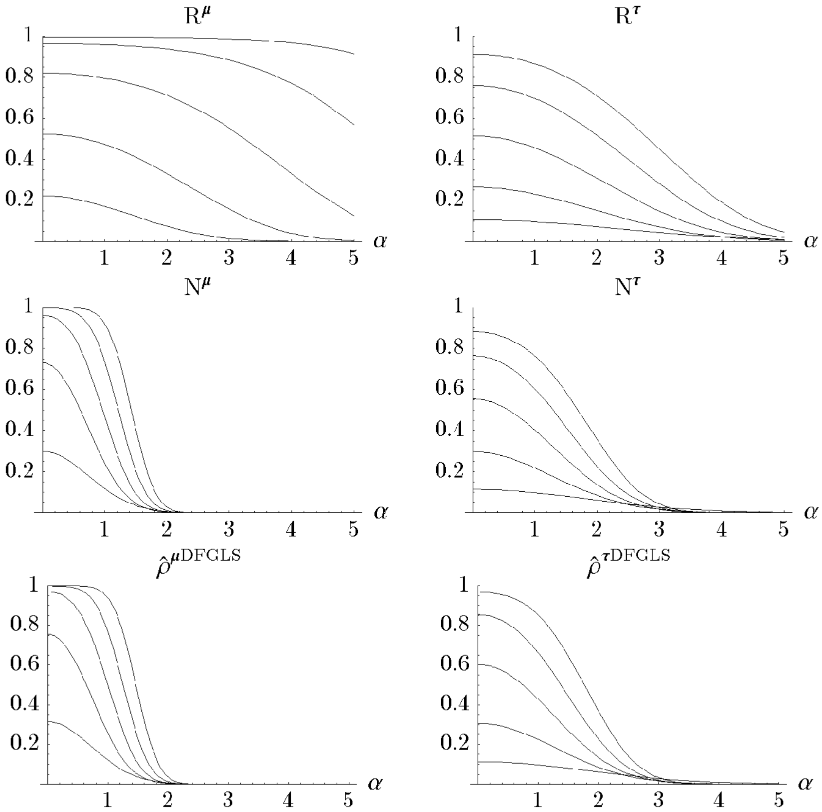

FIGURE 3. Asymptotic power as a function of α for $\gamma = 5$, 10, 15, 20 and 25

Theorem 4. *Under the conditions of Lemma 1 the classes of unit root tests 1–6 of Table 1 are asymptotically equivalent to optimal tests based on $Q_a^i(g,k)$ for a particular choice of g and k:*

	mean		*mean and trend*	
	g	k	g	k
R	∞	$\neq 0$	$\to 0$	$2/g$
N	∞	0	$\to 0$	*arbitrary constant*
$\hat{\rho}^{\text{DFGLS}}$	$-2\,\text{cv}$	0	$g^{\tau,\text{DFGLS}}$	0
LB	$\to 0$	$k \neq 2$	$\to 0$	*arbitrary constant*
$P_T(\bar{c})$	$\bar{c}$	0	$\bar{c}$	0
$Q_T(\bar{c})$	$\bar{c}$	1	$\bar{c}$	1

where the equivalence for $\hat{\rho}^{\text{DFGLS}}$ in the trend case holds provided $g^{\tau,\text{DFGLS}} = \frac{1-3a+(1-2a-3a^2)^{1/2}}{2a}$ *with* $a = -\frac{\bar{c}^4 - 6\,\text{cv} - 12\bar{c}\,\text{cv} - 6\bar{c}^2\,\text{cv}}{2(3+3\bar{c}+\bar{c}^2)^2\,\text{cv}}$ *is real.*

Given the equivalence of a distributional assumption concerning ξ and the weighting approach (3.2), the asymptotic equivalence of tests based on $P_T(\bar{c})$ and $Q_T(\bar{c})$

with tests based on $Q_a(\bar{c}, 0)$ and $Q_a(\bar{c}, 1)$, respectively, is warranted 'by construction'. The locally best tests which make up the LB class were derived by Dufour and King (1991) and Nabeya and Tanaka (1990) under the assumption that the variance of ξ is a fixed number. This corresponds to the case $k = 0$, and so again by construction the LB-class of tests is asymptotically equivalent to a test based on the (appropriately scaled) limit of $Q_a(g, 0)$ as $g \to 0$. But Theorem 4 additionally implies that this limit is independent of k, proving that the LB-class of tests are also asymptotically locally optimal against an alternative with a zero-mean, normal ξ independent of ν_t with variance $O(\gamma^{-1})$. Additionally, since the asymptotic distribution of $\bar{Q}_a(g)$ is the same than that of $Q_a(g, 1)$, this is also true if the initial observation is drawn from the unconditional distribution under the alternative.

The (uncorrected) R and N statistics were constructed as approximations to the locally best tests of the unit root hypothesis for independent disturbances ν_t against the stationary model ($\xi = \sum_{s=0}^{\infty} \rho^s \nu_{-s}$) and nonstationary model with $\xi = \nu_0$ in the neighborhood of $\rho = 1$, respectively. Their derivation by Sargan and Bhargava (1983) and Bhargava (1986) uses the Anderson approximation to the variance-covariance matrices in the Gaussian densities. Interestingly, the different assumption concerning the initial in the derivation of N and R leads to asymptotically different approximate locally best tests, in contrast to the exact locally best tests based on LB. As already pointed out by Nabeya and Tanaka (1990), the N and R statistics generally do not — even asymptotically — correspond to the locally best test statistics when the exact densities are used. In fact, the R^μ and N^μ statistics are optimal for a ρ that is just smaller than any alternative considered in the local-to-unity framework, and a test based on R^τ statistic is locally optimal against the alternative that the initial stems from a normal with a variance which is an order of magnitude larger than the variance of the unconditional distribution.

The $\hat{\rho}^{i,\text{DFGLS}}$-class of tests are asymptotically equivalent to a test based on $Q_a^i(g, 0)$ where g depends on the level of the test. Fixing $\bar{c}$ at 13.5 (the value suggested by Elliott, Rothenberg, and Stock (1996)), we find that the 5% critical values of $\hat{\rho}^{\mu,\text{DFGLS}}$ and $\hat{\rho}^{\tau,\text{DFGLS}}$ are given by -8.039 and -16.591, respectively, which correspond to $g = 16.08$ and $g = 29.20$, whereas the 1% critical values are -13.694 and -23.576, which correspond to $g = 27.39$ and $g = 36.14$. The reduction of the level therefore yields tests which are optimal for alternatives which are easier to distinguish.

The asymptotic power as a function of α and γ of R^i, N^i and $\hat{\rho}^{\mu,\text{DFGLS}}$ are depicted in Figure 3. Tests based on the R^i statistic have about the same robustness against large $|\alpha|$ as tests based on $Q_a^i(g, 1)$ for a relatively large g. The N^i-class and $\hat{\rho}^{i,\text{DFGLS}}$-class of tests have similar asymptotic power characteristics, underlining the observation that the choice of g is especially unimportant for the case $k = 0$. As their construction would suggest, both tests concentrate their power at moderate values of $|\alpha|$.

Figure 4 shows asymptotic power as a function of γ and α of the LB-class of tests. Especially in the mean case, power is very low for most of the considered values of γ and α. In the mean case, the exact locally best approach hence does not yield a test with good power for relevant values of γ and α, in contrast to the approximation to this test employed by Sargan and Bhargava.

We now turn to the $\hat{\rho}^{DF}$ and $\hat{\tau}^{DF}$ classes of tests. While the asymptotic critical region of the $\hat{\rho}^{DF}$-class can be written in the form (4.1), there are no values l_0,

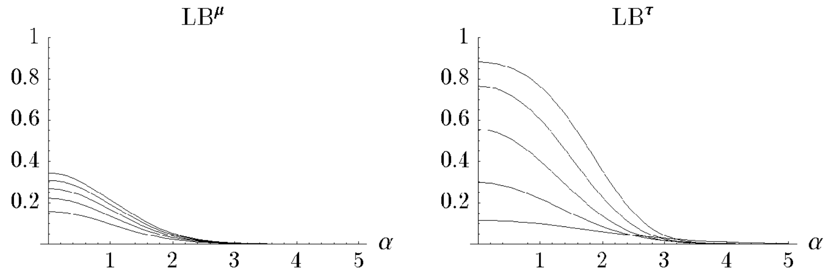

FIGURE 4. Asymptotic power as a function of α for $\gamma = 5$, 10, 15, 20 and 25

g_0 and k_0 such that (4.2) is satisfied. The asymptotic critical region of the $\hat{\tau}^{DF}$-class is fundamentally nonlinear in $M^i(1)^2$, $\left(\int M^i(s)ds\right)^2$, $M^i(1)\int M^i(s)ds$ and $\int M^i(s)^2 ds$. But consider the model

$$(4.3) \qquad \begin{cases} \Delta y_t = T^{-1/2}\zeta + \nu_t & \text{mean case} \\ \Delta y_t = \beta^\tau + T^{-3/2}\zeta t + \nu_t & \text{trend case} \end{cases}$$

which includes the additional parameter ζ over (2.1) with $\rho = 1$.

Theorem 5. *Consider the Data Generating Processes (4.3) and (2.1) with regressor matrix $X = X^i$ when the disturbance vector ν is multivariate Gaussian $N(0, V)$ and ζ is random and independent of ν. Then*

$$\breve{Q}^i_\kappa(r,\lambda) = Q^i(r,\lambda) + \frac{\kappa(\tilde{y}' G_0^i \tilde{Z}^i_\zeta)^2}{1 + \kappa \tilde{Z}^{i\prime}_\zeta G_0^i \tilde{Z}^i_\zeta}$$

is a statistic to optimally test H_0 : $\tilde{y}$ stems from model (4.3) with $\zeta \sim N(0,\kappa)$ against H_1 : $\tilde{y}$ stems from model (2.1) with $\rho = r$ in the sense of (3.2) with a zero mean Gaussian weighting function of variance λ which is invariant to the transformations (3.1), where $\tilde{Z}^\mu_\zeta = T^{-1/2}\tilde{\tau}$ and $\tilde{Z}^\tau_\zeta = T^{-3/2}(0, 1, 2^2, \cdots, T^2)'$.
Furthermore, let $\breve{Q}^i(r,\lambda)$ be the limit of these statistics as $\kappa \to \infty$, and denote with $\breve{Q}^i_a(g,k)$ the asymptotic version $\breve{Q}^i_a(g,k) = \breve{Q}^i(1 - gT^{-1}, \omega^2 k(2g)^{-1}T)$. Then under the assumptions of Lemma 1, a test based on a member of the $\hat{\rho}^{i,DF}$-class with critical value $\text{cv}^{i,DF}$ is asymptotically equivalent to a test based on $\breve{Q}^i_a(-2\,\text{cv}^{i,DF}, \infty)$.

Members of the $\hat{\rho}^{i,DF}$-class hence are (the limit of) asymptotically optimal tests, but only under the assumption that the null hypothesis is given by (4.3). This somewhat strange property may be understood from the fact that the coefficients in a Dickey-Fuller regressions of y on (y_{-1}, e, X) have an altogether different interpretation under the null of $\rho = 1$ compared to $\rho < 1$ (cf. Bhargava (1986)). In the mean case, for instance, the coefficient on e estimates the transformed mean $\mu(1-\rho)$ under the alternative, but a time trend under the null of $\rho = 1$. Some authors, like Phillips and Xiao (1998), have conjectured that this introduction of a superfluous regressor under the null is the reason for the reportedly low power of the $\hat{\rho}^{DF}$-class compared to other unit root tests.

But Theorem 5 reveals another striking feature of the statistics in the $\hat{\rho}^{DF}$-class: the weight function for α puts very much weight on very large $|\alpha|$. At the 5% level, the asymptotic critical values of the $\hat{\rho}^{DF}$-class are -14.1 and -21.7 in the

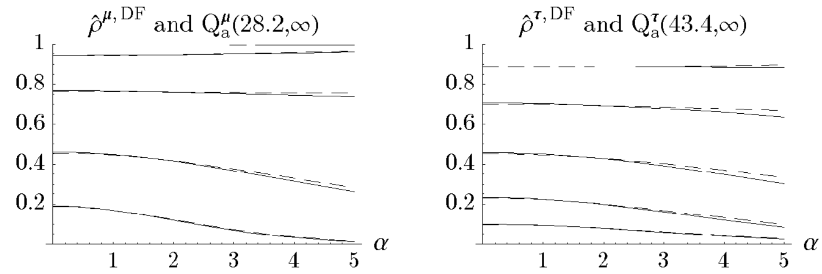

FIGURE 5. Asymptotic power as a function of α for $\gamma = 5$, 10, 15, 20 and 25

mean and time trend case, respectively, so that the $\hat{\rho}^{i,DF}$ class is asymptotically equivalent to tests based on $\breve{Q}^{\mu}(28.2,\infty)$ and $\breve{Q}^{\tau}(43.4,\infty)$ for this level. Figure 5 depicts the asymptotic power of the 5% tests based on the $\hat{\rho}^{i,DF}$-class (solid line), along with the asymptotic power of a test based on $Q_a^i(-2\,\mathrm{cv}^{i,DF},\infty)$ (dashed line). The difference in asymptotic power between the two tests is small for all considered values of γ and α. This implies that it is the extreme weighting of large values of $|\alpha|$ (rather than the additional parameter regressor under the null) that is the deeper reason for the low power of tests in the $\hat{\rho}^{DF}$-class at small and moderate values of $|\alpha|$ — the range of initial values typically considered in Monte Carlo studies.

Theorem 6. *Let* $\breve{Q}_a^i(g,k)$ *as in Theorem 5. Then*
(i) $\hat{g}^{i,DF} = \mathrm{argmin}_g\, \breve{Q}_a^i(g,\infty)$ *exists with probability 1*
(ii) under the assumptions of Lemma 1, a test based on a statistic of the $\hat{\tau}^{DF}$*-class is asymptotically equivalent to a test which rejects for small values of*

$$\mathrm{sign}(\hat{g}^{i,DF})\breve{Q}_a^i(\hat{g}^{i,DF},\infty).$$

Tests based on statistics in the $\hat{\tau}^{DF}$-class can hence be thought of as estimating γ, and then using the estimated value for g in the statistic $\breve{Q}_a^i(g,\infty)$. This makes the $\hat{\tau}^{DF}$-class comparable to a generalized likelihood ratio statistic, especially given that $\breve{Q}_a^i(g,\infty)$ is asymptotically equivalent to $E\breve{n}v_a^i(g,\hat{\alpha}_{ML}^i)$, where $E\breve{n}v_a^i(g,a)$ is defined analogously to $Env_a^i(g,a)$. Since the tests reject first if the estimated γ is larger than zero, it is a 'signed' version of a generalized likelihood ratio statistic. While not proving any optimally property of the $\hat{\tau}^{DF}$-class, Theorem 6 relates this class to the optimal statistics $\breve{Q}_a^i(g,\infty)$, which put an extreme weighting on large values of $|\alpha|$.

Figure 6 depicts the asymptotic power of a test based on $\hat{\tau}^{DF}$ (solid line), along with a test based on

$$gLR^i = \mathrm{sign}(\hat{g}^i)Q_a^i(\hat{g}^i,\infty)$$

where $\hat{g}^i = \mathrm{argmin}_g\, Q_a^i(g,\infty)$ (dashed line). Note that $Q_a^i(\hat{g}^i,\infty)$ is asymptotically equivalent to the likelihood ratio statistics Φ_2 and Φ_3 of Dickey and Fuller (1981). As one would expect, the asymptotic power of both tests is large for large values of $|\alpha|$. In fact, in stark contrast to all other popular tests (apart from the $\hat{\rho}^{DF}$-class) power increases in $|\alpha|$ for all considered values of γ. The 'right' null model in the construction of gLR^i leads to an even steeper gradient.

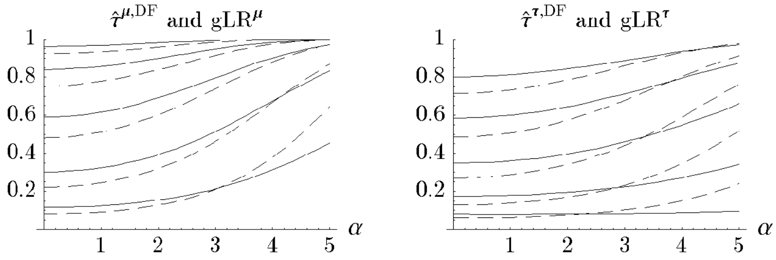

FIGURE 6. Asymptotic power as a function of α for $\gamma = 5$, 10, 15, 20 and 25

TABLE 4. Classes of Unit Root Tests and values of g and k of a comparable test based on $Q_a(g, k)$

	mean			mean and trend		
	cp	g_0	k_0	cp	g_0	k_0
$\hat{\tau}^{\text{WS}}$	.983	10.8	.988	.991	15.9	1.01
$\hat{\rho}^{\text{WS}}$	.994	23.3	.999	.998	36.4	1.00
$\hat{\tau}^{Ley}$	.901	25.8	.973	.913	37.4	1.08
$\hat{\tau}^{\text{DFGLS}}$	.996	8.36	0.00	.999	17.7	0.00

Whilst the asymptotic distribution of the classes of test statistics 8–11 depend on the same functionals of $M(s)$ than $Q_a(g, k)$, there does not seem to exist such a close link to a particular member of the family $Q_a(g, k)$ compared to the test statistics considered so far. But it would still be insightful to identify particular values of g_0 and k_0 such that tests based on a class of statistics 8–11 become very much comparable to tests based on $Q_a(g_0, k_0)$. One measure of 'comparability' of tests of the same level is the probability that the two tests either both reject or both do not reject. By (approximately) maximizing this probability over g and k under the null hypothesis of $\rho = 1$ for T large, we found the results depicted in Table 4. The column cp is the (estimated) conditional asymptotic probability that the 5% level test based on $Q_a(g_0, k_0)$ rejects given that the 5% level test based on the the statistic in the first column rejects for $\rho = 1$. See the Appendix for how we conducted the search for suitable values of g_0 and k_0.

The large values of cp imply that the behavior of the classes of test statistics 8–11 can be closely mimicked by members of the optimal family $Q_a(g, k)$, only for the test based on $\hat{\tau}^{Ley}$ cp is below 98%. This close correspondence is confirmed in Figure 7 which depicts the asymptotic power the tests of Table 4 (solid line) along with the corresponding tests based on $Q_a(g, k)$ (dashed line). The asymptotic power of the popular tests is hardly distinguishable from the corresponding optimal tests over a wide range of values of γ and α. The classes of tests $\hat{\tau}^{\text{WS}}$, $\hat{\rho}^{\text{WS}}$ and $\hat{\tau}^{Ley}$ are hence very much comparable to tests based on $Q_a(g, 1)$ for some g. The implicit weighting of different α of these tests almost corresponds to the optimal weighting if the initial value is drawn from the unconditional distribution. This explains why $\hat{\tau}^{\text{WS}}$, $\hat{\rho}^{\text{WS}}$ and $\hat{\tau}^{Ley}$ fared well in Monte Carlo studies which employed such an

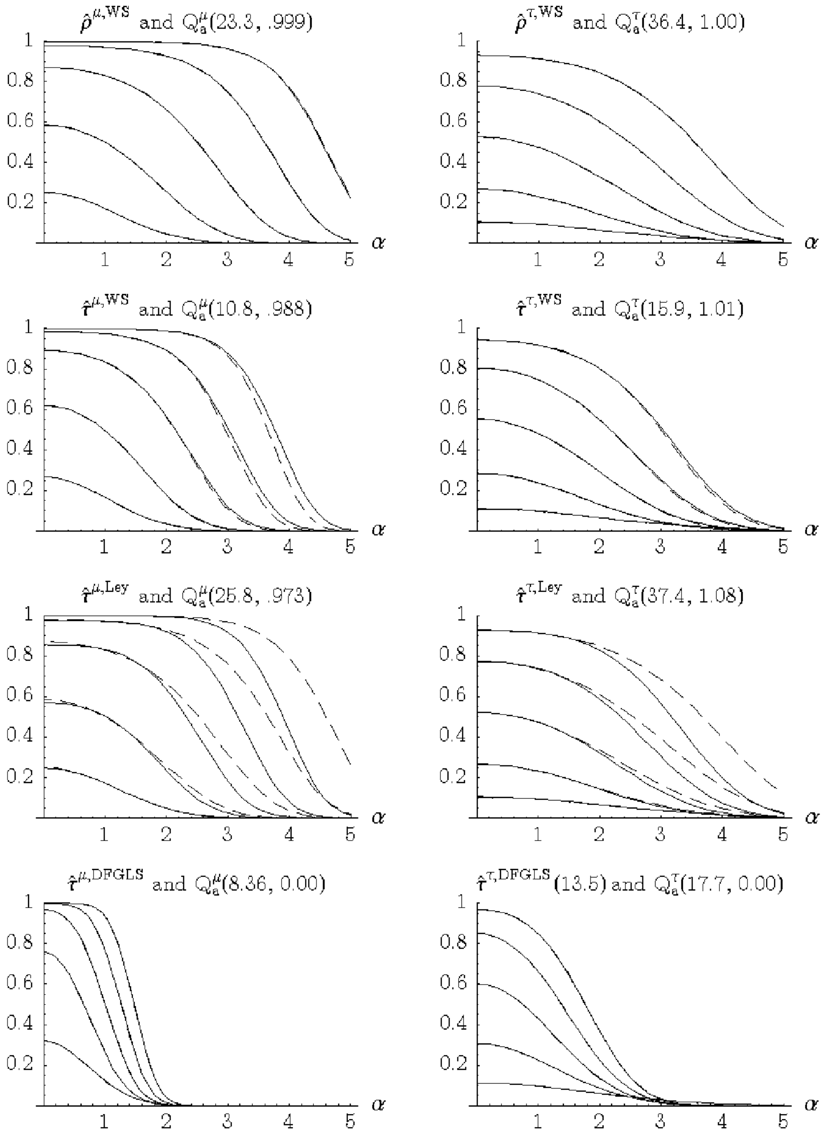

FIGURE 7. Asymptotic power as a function of α for $\gamma = 5$, 10, 15, 20 and 25

assumption — see Pantula, Gonzalez-Farias, and Fuller (1994), Leybourne (1995) and Elliott (1999).

The asymptotic distributions of the J and R/S statistics depend on functionals of $M(s)$ which do not appear in any of the optimal families $S_a(g, F_a)$, $\mathrm{Env}_a(g, a)$ and $Q_a(g, k)$. But the asymptotic distribution of $\mathrm{Env}_a(g, a)$ is composed of an exhaustive list of functionals of $M(s)$ which appear in the asymptotic likelihood of a first order autoregressive process. An 'asymptotic sufficiency' argument therefore suggests that the J and R/S statistics are inefficient unit root test statistics. This would explain their poor performance in Monte Carlo experiments as provided by Stock (1994). This reasoning implies that the low power of these statistics is likely to be general to any particular alternative or assumption on the initial condition.

5. Asymptotically Equivalent Feasible Tests for the Optimal Family

The optimal statistics derived in section 3 depend not only on the assumption of the weighting F but also on the assumption that the disturbance vector ν is Gaussian with known variance-covariance matrix V. The results of the last section imply that some popular tests, which are constructed without the knowledge of V, are asymptotically equivalent to specific members of the family of optimal test statistics $Q_a(g,k)$. In this section we extend this feasibility by deriving families of test statistics which do not require the knowledge of V either and that are asymptotically equivalent to the optimal families of statistics $S_a(g,F_a)$, $\mathrm{Env}_a(g,a)$ and $Q_a(g,k)$ of section 3. Furthermore, we show that tests based on these statistics achieve the same asymptotic power under very general assumptions on the distribution of the error process as under the assumption of normality.

We assume the following condition on the Data Generating Process

Condition 2. *The zero mean process $\{\nu_t\}$ is covariance-stationary and ergodic with finite autocovariances $\gamma(k) = E[\nu_t \nu_{t-k}]$ such that*
(a) $\omega^2 = \sum_{k=-\infty}^{\infty} \gamma(k)$ is finite and nonzero
(b) the scaled partial-sum process $T^{-1/2}\sum_{t=1}^{[sT]} \nu_t \Rightarrow \omega W(s)$.

The suggested statistics for the mean case ($i=\mu$) and trend case ($i=\tau$) are

$$\hat{S}_a^i(g,F_a) = s_0^i + s_1^i\left(\hat{\omega}^{-1}T^{-1/2}y_T^i\right)^2 + s_2^i\left(\hat{\omega}^{-1}T^{-3/2}\sum_{t=0}^{T} y_t^i\right)^2$$
$$-2\ln\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}\left[s_3^i a\hat{\omega}^{-1}T^{-3/2}\sum_{t=0}^{T} y_t^i + s_4^i a\hat{\omega}^{-1}T^{-1/2}y_T^i + s_5^i a^2\right]\right\}dF_a(a)$$

$$\widehat{\mathrm{Env}}_a^i(g,a) = s_0^i + s_1^i\left(\hat{\omega}^{-1}T^{-1/2}y_T^i\right)^2 + s_2^i\left(\hat{\omega}^{-1}T^{-3/2}\sum_{t=0}^{T} y_t^i\right)^2$$
$$+ s_3^i a\hat{\omega}^{-1}T^{-3/2}\sum_{t=0}^{T} y_t^i + s_4^i a\hat{\omega}^{-1}T^{-1/2}y_T^i + s_5^i a^2$$

$$\hat{Q}_a^i(g,k) = q_0^i + q_1^i\left(\hat{\omega}^{-1}T^{-1/2}y_T^i\right)^2 + q_2^i\left(\hat{\omega}^{-1}T^{-3/2}\sum_{t=0}^{T} y_t^i\right)^2$$
$$+ q_3^i\hat{\omega}^{-2}T^{-2}y_T^i\sum_{t=0}^{T} y_t^i + q_4^i\hat{\omega}^{-2}T^{-2}\sum_{t=0}^{T}(y_t^i)^2$$

where s_j^i and q_j^i (for $i = \mu,\ \tau,\ j = 0,1,2,3,4$) are defined in Theorem 3 and $\hat{\omega}$ is an estimator of the long run variance of ν_t. The detrended data y_t^i is generated according to

(i) $i = \mu$ then $y_t^\mu = y_t - y_0$

(ii) $i = \tau$ then $y_t^\tau = y_t^\mu - \hat{\beta} y_t^\mu$, where $\hat{\beta}$ is the OLS estimate from a regression of y_t^μ on τ.

The construction of these tests follows the 'modified' test statistic approach suggested by Stock (2000), and extended in Ng and Perron (2001). They suggest the above method for obtaining a test which is asymptotically equivalent to the Elliott, Rothenberg, and Stock (1996) P_T test, which corresponds to $Q_a(g,0)$ test in our notation, so that for $k=0$ the above tests are equivalent (apart again from the constant) to the MP_T test in Ng and Perron (2001).

Theorem 7. *Assume that the data is generated under Condition 2 and that $\hat{\omega} \xrightarrow{p} \omega$ under the null and local alternatives. Then $\hat{S}_a^i(g,F_a)$, $\widehat{\mathrm{Env}}_a^i(g,a)$ and $\hat{Q}_a^i(g,k)$ have the same asymptotic distribution than $S_a^i(g,F_a)$, $\mathrm{Env}_a^i(g,a)$ and $Q_a^i(g,k)$ under Condition 1, i.e. the distributions described in Theorem 3.*

Thus tests can be constructed that have asymptotically the same power as the optimal tests under the normality assumption. Note, however, that the tests are not generally optimal if the disturbances ν_t stem from a nonnormal distribution. See Rothenberg and Stock (1997) for a discussion of how one might construct more powerful unit root tests in this situation.

A great number of potential estimators for $\hat{\omega}$ are available — each will affect the small sample properties of the test but not the large sample properties. Potential estimators include 'sums of covariances' type estimators such as derived and examined in Newey and West (1987) and Andrews (1991) or alternatively autoregressive estimators discussed in Haan and Levin (2000) and employed for special cases of the above unit root tests in Elliott, Rothenberg, and Stock (1996) and Elliott (1999). Stock (1994) discusses choices available. We will follow the suggestions in Ng and Perron (2001) with a modification in the first step in accordance with the points of this paper and examine the properties of autoregressive estimators constructed according to the following scheme.

Step 1: Calculate y_t^i according to their definitions above

Step 2: Run the regression $\Delta y_t^i = \theta_0 y_{t-1}^i + \sum_{j=1}^p \theta_j \Delta y_{t-j}^i + e_{p,t}$ for a range of possible lag lengths $p = 0, \cdots, p_{\max}$.

Step 3: Choose p^* according to minimize the MAIC criterion of Ng and Perron (2001). This criterion is given by

$$\mathrm{MAIC} = \ln(\hat{\sigma}_p^2) + \frac{2\left(\tau(p)+p\right)}{T - p_{\max}}$$

where $\hat{\sigma}_p^2 = (T-p_{\max})^{-1} \sum_{t=p_{\max}}^T e_{p,t}^2$ and $\tau(p) = (\hat{\sigma}_p^2)^{-1} \hat{\theta}_0^2 \sum_{t=p_{\max}}^T (y_{t-1}^i)^2$.

Step 4: For the chosen p^*, rerun the regression in Step 2 and construct the estimate

$$\hat{\omega}^2 = \frac{\hat{\sigma}_{p^*}^2}{\left(1-\hat{\theta}(1)\right)^2}$$

where $\hat{\theta}(1) = \sum_{j=1}^{p^*} \hat{\theta}_j$.

TABLE 5. Asymptotic critical values and small sample size for $T = 100$, independent normal innovations

	critical values 1%	critical values 5%	size		critical values 1%	critical values 5%	size
$\hat{Q}^{\mu}_a(7,0)$	-5.035	-3.694	0.044	$\hat{Q}^{\tau}_a(7,0)$	-5.882	-5.403	0.031
$\hat{Q}^{\mu}_a(10,0)$	-6.110	-3.513	0.044	$\hat{Q}^{\tau}_a(10,0)$	-7.764	-6.814	0.030
$\hat{Q}^{\mu}_a(7,1)$	-5.428	-4.585	0.040	$\hat{Q}^{\tau}_a(7,1)$	-5.964	-5.552	0.028
$\hat{Q}^{\mu}_a(10,1)$	-6.94	-5.354	0.039	$\hat{Q}^{\tau}_a(10,1)$	-7.945	-7.152	0.027
$\hat{Q}^{\mu}_a(7,2)$	-5.618	-4.920	0.038	$\hat{Q}^{\tau}_a(7,2)$	-6.040	-5.678	0.025
$\hat{Q}^{\mu}_a(10,2)$	-7.245	-5.874	0.037	$\hat{Q}^{\tau}_a(10,2)$	-8.090	-7.380	0.024

Notes: Size values are for 5% nominal level and are based on 60000 Monte Carlo replications with $p_{\max} = 0$.

The modification to the Ng and Perron (2001) procedure is in Step 1 where in that paper the detrending of $\tilde{y}$ is carried out according to the assumptions of Elliott, Rothenberg, and Stock (1996).

To implement the tests, asymptotic critical values were computed from inverting the characteristic function for the $Q^i_a(g,k)$ tests constructed via Lemma 2. Table 5 reports 1% and 5% critical values for mean version of the tests where we have set $g = 7, 10$ and $k = 0, 1, 2$. Recall that different choices for g simply maximize the power at different points in the γ dimension and different choices for k accord to a different weighting of possible initial conditions, where larger values indicate more weight on larger initial conditions. Also reported are Monte Carlo results for size using the asymptotic 5% critical value when the model is generated according to (2.1) with $T = 100$, no serial correlation and normally distributed errors. The statistics are all a little undersized, although this is minor.

The asymptotic theory and results presented so far suggest that there is a trade-off between power for small and large initial conditions. We have shown this analytically through showing that optimal tests are different for different assumptions on the initial condition, and numerically that increasing k lowers asymptotic power when the initial condition is small. We have also shown that many popular unit root tests have a close correspondence to a member of the family of optimal tests $Q^i_a(g,k)$ for a certain choice of g and k. These tests too are thus making implicitly this trade-off in their power functions.

Tables 6 and 7 show these points in a Monte Carlo exercise where the model is as for the size calculation in Table 5 and we have varied the initial condition. The two panels show power against $\rho = 0.9$ (as $T = 100$ this corresponds to the asymptotic results where $\gamma = 10$) for the mean case and the trend case, respectively. We can see clearly that the ranking in terms of power for both cases accords with the asymptotic results. Concentrating on the $\hat{Q}^{\mu}(g,k)$ test, we have for $\alpha = 0$ that the power falls as the weighting function puts more mass on larger initial conditions (increase k). For example, with $g = 10$, we have power equal to 64% when $k = 0$, this drops to 62% for $k = 1$ and 54% when $k = 2$. However, for all of these tests power falls as $|\alpha|$ increases. At $\alpha = 1.5$ the power rankings of the tests is reversed. When $k = 0$ power has fallen to just 11%, for $k = 1$ it is 35% and for $k = 2$ power has only fallen by 11% to 43%. The same qualitative behavior can also be

Table 6. Size corrected small sample power of various tests for $\rho = 0.90$, $T = 100$ and independent normal innovations in the mean case

test	0	0.5	1	1.5	α 2	2.5	3	3.5	4
$\hat{\rho}^{\mu,\mathrm{DF}}$	.47	.46	.45	.44	.42	.40	.37	.34	.32
$\hat{\mu}^{\mu,\mathrm{DF}}$	.30	.31	.33	.35	.40	.45	.52	.60	.68
N^{μ}	.74	.57	.24	.05	.00	.00	.00	.00	.00
R^{μ}	.51	.51	.48	.43	.36	.26	.18	.10	.05
P_T^{μ}	.75	.63	.33	.09	.01	.00	.00	.00	.00
$\hat{\rho}^{\mu,\mathrm{DFGLS}}$	.73	.64	.38	.13	.02	.00	.00	.00	.00
$\hat{\mu}^{\mu,\mathrm{DFGLS}}$	.73	.64	.39	.14	.02	.00	.00	.00	.00
$\hat{\rho}^{\mu,\mathrm{WS}}$	.57	.56	.51	.42	.30	.17	.07	.02	.00
$\hat{\mu}^{\mu,\mathrm{WS}}$	.61	.59	.51	.38	.22	.09	.03	.00	.00
$\hat{Q}_a^{\mu}(7,0)$	.76	.60	.26	.05	.00	.00	.00	.00	.00
$\hat{Q}_a^{\mu}(10,0)$	.77	.60	.26	.05	.00	.00	.00	.00	.00
$\hat{Q}_a^{\mu}(7,1)$	.64	.61	.51	.34	.16	.04	.01	.00	.00
$\hat{Q}_a^{\mu}(10,1)$	.62	.60	.51	.37	.20	.07	.01	.00	.00
$\hat{Q}_a^{\mu}(7,2)$	.52	.52	.50	.45	.39	.31	.22	.14	.08
$\hat{Q}_a^{\mu}(10,2)$	.52	.52	.50	.45	.38	.30	.21	.13	.07
$\hat{Q}_T^{\mu}$ $(\bar{c}=10)$	.62	.59	.51	.37	.21	.08	.02	.00	.00

Notes: Based on 60000 Monte Carlo replications with $p_{max} = 0$.

Table 7. Size corrected small sample power of various tests for $\rho = 0.90$, $T = 100$ and independent normal innovations in the trend case

test	0	0.5	1	1.5	α 2	2.5	3	3.5	4
$\hat{\rho}^{\tau,\mathrm{DF}}$	.24	.24	.23	.22	.20	.18	.16	.14	.12
$\hat{\tau}^{\tau,\mathrm{DF}}$	.18	.18	.19	.20	.21	.23	.25	.28	.31
N^{τ}	.30	.28	.22	.15	.08	.04	.02	.01	.00
R^{τ}	.26	.25	.23	.20	.16	.12	.08	.05	.03
P_T^{τ}	.32	.30	.24	.17	.10	.05	.02	.01	.00
$\hat{\rho}^{\tau,\mathrm{DFGLS}}$	.30	.29	.24	.18	.11	.06	.03	.01	.00
$\hat{\tau}^{\tau,\mathrm{DFGLS}}$	.31	.29	.24	.18	.12	.06	.03	.01	.00
$\hat{\rho}^{\tau,\mathrm{WS}}$	.28	.27	.24	.21	.16	.12	.07	.04	.02
$\hat{\tau}^{\tau,\mathrm{WS}}$	.29	.28	.24	.20	.15	.10	.06	.03	.01
$\hat{Q}_a^{\tau}(7,0)$	.31	.29	.23	.15	.09	.04	.02	.01	.00
$\hat{Q}_a^{\tau}(10,0)$	.32	.29	.23	.16	.09	.04	.02	.01	.00
$\hat{Q}_a^{\tau}(7,1)$	.31	.29	.25	.19	.13	.07	.04	.02	.01
$\hat{Q}_a^{\tau}(10,1)$	.30	.29	.24	.19	.13	.08	.04	.02	.01
$\hat{Q}_a^{\tau}(7,2)$	.27	.26	.24	.21	.17	.13	.09	.06	.03
$\hat{Q}_a^{\tau}(10,2)$	.27	.26	.24	.22	.18	.14	.10	.07	.04
$\hat{Q}_T^{\tau}$ $(\bar{c}=10)$	.30	.28	.24	.19	.13	.08	.04	.02	.01

Notes: See notes of Table 6.

observed for the trend case in Table 7. Such results clearly show that the trade-off between power for small and large initial conditions accords with the analytical results presented earlier.

The robustness of the Dickey Fuller t-statistic also comes through clearly in the Monte Carlo results. This test has the lowest power amongst the tests by a fairly large margin when $|\alpha|$ is small. However, this test is the only test that has power that increases as $|\alpha|$ gets larger. For very large $|\alpha|$ it is the dominant test in terms of power. This accords directly to the analytical results that show that although this test is not asymptotically equivalent to a member of the $Q_a(g,k)$ family, it is intimately linked to the optimal family $\check{Q}_a(g,\infty)$ which puts very much weight on very large $|\alpha|$.

6. CONCLUSION

In choosing a test based on asymptotic properties the choice comes down to choosing a power function. In employing unit root tests in practice the choice of a particular test is inherently more difficult as there is no uniformly most powerful test even in the case of assuming a zero initial condition. In this paper we show that for more general assumptions on the initial condition, the choice of an asymptotic power function depends also on the initial condition of the process. This is true for both the set of tests that are derived to exploit information on the initial condition in some optimal way (e.g. Elliott, Rothenberg, and Stock (1996)) and also commonly applied tests that do not explicitly consider the initial condition (e.g. the Dickey and Fuller (1979) statistics). This paper shows how power is affected analytically and empirically. Since a researcher must make a choice based on some view of the initial condition, it makes sense to make this choice explicit. We derive families of statistics that are optimal in the sense of maximizing weighted average power, where the weighting is a distribution function of the initial condition. Such statistics contain previous efficient tests for a unit root such as those in Elliott, Rothenberg, and Stock (1996) and Elliott (1999). We relate other statistics, that do not have such optimality properties, to this optimal class as a method of understanding the implicit assumptions they make on the initial condition. Through this we are able to gain a much fuller understanding of these statistics and their properties, as well as a deeper understanding as to the differences in power properties found in previous Monte Carlo studies (which usually have the initial condition being drawn from its unconditional distribution under the alternative, which is identical to simply dropping off 'enough' observations to 'get rid' of the initial condition). The results then give researchers fuller information on the power curve choice.

The results of this paper have implications for larger dimension systems, and may be extendable in these directions. First, a multivariate analog of unit root testing is testing for cointegrating rank. Such tests usually assume away the initial condition as is often the case in the univariate literature. Such tests will have some relationship between power and robustness to this assumption. Second, a recent literature has developed around pooling unit root tests for a number of variables in a panel setting. The panel setting will also be affected in a similar way to the univariate tests, although in a panel setting more may be possible in terms of using the data to discern between potential assumptions on the initial condition.

Department of Economics, University of St. Gallen, Dufourstr. 48, 9000 St. Gallen, Switzerland; ulrich.mueller@unisg.ch

and

Department of Economics, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508; gelliott@weber.ucsd.edu

APPENDIX

Proof of Theorem 1:
From the theory about invariant tests, we know that an optimal invariant test statistic is given by the Likelihood Ratio statistic of a maximal invariant (cf. Lehmann (1986), pp. 282). Let $\tilde{M} = \tilde{I} - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'$, where $\tilde{I}$ is the $(T+1)\times(T+1)$ identity matrix. Then $\tilde{M}\tilde{y}$ is such a maximal invariant. Maximizing the weighted average power criterion (3.2) is equivalent to maximizing power against the simple alternative H_1^* : the density of $\tilde{y}$ is given by $\int_{-\infty}^{\infty} f(\tilde{y}|x)dF(x)$, where $f(\tilde{y}|x)$ is the density of $\tilde{y}$ given $\xi = x$. Under H_0 and for any ξ, $\tilde{M}\tilde{y} \sim N(0, \tilde{M}\tilde{\Sigma}_0\tilde{M})$. Under H_1 and ξ fixed to x, $\tilde{M}\tilde{y} \sim N(\tilde{M}\tilde{R}_1 x, \tilde{M}\tilde{\Sigma}_1\tilde{M})$, such that the distribution of $\tilde{M}\tilde{y}$ under H_1^* is given by a mixture of these normals. Note that the common null space of $\tilde{M}\tilde{\Sigma}_i\tilde{M}$ is the column space of $\tilde{X}$. The density of $\tilde{M}\tilde{y}$ restricted to the hyperplane which is orthogonal to $\tilde{X}$ is then proportional to $\exp\{-\frac{1}{2}\tilde{y}'\tilde{M}(\tilde{M}\tilde{\Sigma}_0\tilde{M})^{-}\tilde{M}\tilde{y}\}$ under H_0 and proportional to $\exp\{-\frac{1}{2}(\tilde{y}-\tilde{R}_1x)'\tilde{M}(\tilde{M}\tilde{\Sigma}_1\tilde{M})^{-}\tilde{M}(\tilde{y}-\tilde{R}_1x)\}$ under H_1 and ξ fixed to any x, where $(\ )^{-}$ is any generalized inverse (cf. Rao and Mitra (1971), p. 204).

We now prove that $G_i = \tilde{\Sigma}_i^{-} - \tilde{\Sigma}_i^{-}\tilde{X}(\tilde{X}'\tilde{\Sigma}_i^{-}\tilde{X})^{-1}\tilde{X}'\tilde{\Sigma}_i^{-}$ are generalized inverses of $\tilde{M}\tilde{\Sigma}_i\tilde{M}$. (Recall that a generalized inverse G of the matrix A has the property $AGA = A$). Note that $\tilde{X}'\tilde{\Sigma}_i^{-}\tilde{X}$ is necessarily nonsingular, since

$$\det\tilde{\Sigma}_i^{-} = \det\Sigma_i^{-1}\det[1 + e'\Sigma_i^{-1}e - e'\Sigma_i^{-1}e] \neq 0.$$

Furthermore, $\tilde{M}G_i\tilde{M} = G_i$, because $G_i\tilde{X} = \tilde{X}'G_i = 0$, so that G_i has a column space no larger than the projection matrix $\tilde{M}$. We hence find

$$\begin{aligned}
\tilde{M}\tilde{\Sigma}_i\tilde{M}G_i\tilde{M}\tilde{\Sigma}_i\tilde{M} &= \tilde{M}\tilde{\Sigma}_i G_i\tilde{\Sigma}_i\tilde{M} \\
&= \tilde{M}\tilde{\Sigma}_i\left[\tilde{\Sigma}_i^{-} - \tilde{\Sigma}_i^{-}\tilde{X}(\tilde{X}'\tilde{\Sigma}_i^{-}\tilde{X})^{-1}\tilde{X}'\tilde{\Sigma}_i^{-}\right]\tilde{\Sigma}_i\tilde{M} \\
&= \tilde{M}\tilde{\Sigma}_i\tilde{M} - \tilde{M}\tilde{\Sigma}_i\tilde{\Sigma}_i^{-}\tilde{X}(\tilde{X}'\tilde{\Sigma}_i^{-}\tilde{X})^{-1}\tilde{X}'\tilde{\Sigma}_i^{-}\tilde{\Sigma}_i\tilde{M} \\
&= \tilde{M}\tilde{\Sigma}_i\tilde{M}
\end{aligned}$$

where the last line follows from

$$\begin{aligned}
\tilde{X}'\tilde{\Sigma}_i^{-}\tilde{\Sigma}_i\tilde{M} &= \begin{pmatrix} 1 & e' \\ 0 & X' \end{pmatrix}\begin{pmatrix} 1+e'\Sigma_i^{-1}e & -e'\Sigma_i^{-1} \\ -\Sigma_i^{-1}e & \Sigma_i^{-1} \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 0 & \Sigma_i \end{pmatrix}\tilde{M} \\
&= \begin{pmatrix} 1 & e' \\ 0 & X' \end{pmatrix}\begin{pmatrix} 0 & -e' \\ 0 & I \end{pmatrix}\tilde{M} \\
&= \begin{pmatrix} 0 & 0 \\ 0 & X' \end{pmatrix}\tilde{M} = 0
\end{aligned}$$

Putting everything together yields

$$\begin{aligned} LR &= \frac{\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}(\tilde{y}-\tilde{R}_1 x)'\tilde{M}(\tilde{M}\tilde{\Sigma}_1\tilde{M})^{-}\tilde{M}(\tilde{y}-\tilde{R}_1 x)\right\} dF(x)}{\exp\left\{-\frac{1}{2}\tilde{y}'\tilde{M}(\tilde{M}\tilde{\Sigma}_0\tilde{M})^{-}\tilde{M}\tilde{y}\right\}} \\ &= \frac{\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}(\tilde{y}-\tilde{R}_1 x)'G_1(\tilde{y}-\tilde{R}_1 x)\right\} dF(x)}{\exp\left\{-\frac{1}{2}\tilde{y}'G_0\tilde{y}\right\}} \\ &= \exp\left\{-\frac{1}{2}\tilde{y}'\left[G_1-G_0\right]\tilde{y}\right\}\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}\left[x^2\tilde{R}_1'G_1\tilde{R}_1-2x\tilde{R}_1'G_1\tilde{y}\right]\right\}dF(x). \end{aligned}$$

Taking logarithms proves the Theorem.

Proof of Corollary 2:
The proof for $\lambda = 0$ is a direct consequence of Corollary 1. For $\lambda > 0$, we need to find

$$Q(r,\lambda) = \tilde{y}'\left[G_1-G_0\right]\tilde{y} - 2\ln\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}\left[x^2(\tilde{R}_1'G_1\tilde{R}_1+\lambda^{-1})-2x\tilde{R}_1'G_1\tilde{y}\right]\right\}dx + c$$

where c stands for a constant that does not depend on $\tilde{y}$. Carrying out the integration yields the result.

Different form of $Q(r,\lambda)$:
Let $\tilde{\Omega}_1 = \tilde{\Omega}(r)$, $\tilde{\Omega}_0 = \tilde{\Omega}(1)$, $\tilde{R}_0 = \tilde{e}$ and $\tilde{R}_1 = \tilde{R}(r)$. Then $\hat{y}(1)'\tilde{\Omega}_0^{-1}\hat{y}(1)$ and $\hat{y}(r)'\tilde{\Omega}_1^{-1}\hat{y}(r)$ are equal to $\tilde{y}'[\tilde{\Omega}_i^{-1} - \tilde{\Omega}_i^{-1}\tilde{X}(\tilde{X}'\tilde{\Omega}_i^{-1}\tilde{X})^{-1}\tilde{X}'\tilde{\Omega}_i^{-1}]\tilde{y}$ with $i = 0, 1$, respectively. It is clearly sufficient to show that

$$\tilde{y}'G_i\tilde{y} - \frac{(\tilde{R}_i'G_i\tilde{y})^2}{\lambda^{-1}+\tilde{R}_i'G_i\tilde{R}_i} = \tilde{y}'\left[\tilde{\Omega}_i^{-1} - \tilde{\Omega}_i^{-1}\tilde{X}(\tilde{X}'\tilde{\Omega}_i^{-1}\tilde{X})^{-1}\tilde{X}'\tilde{\Omega}_i^{-1}\right]\tilde{y}$$

for $i = 0, 1$. Since both statistics are invariant to the mean and $G_i\tilde{e} = 0$ and $\tilde{e}'G_i = 0$ we can substitute $\tilde{y}^\mu = \tilde{y} - y_0\tilde{e}$ for $\tilde{y}$ and $\tilde{\Upsilon}_i \equiv \tilde{R}_i - \tilde{e}$ for $\tilde{R}_i$ in the above expression. Note that $\tilde{\Upsilon}_0 = 0$. Recall that $G_i = \tilde{\Sigma}_i^- - \tilde{\Sigma}_i^-\tilde{X}(\tilde{X}'\tilde{\Sigma}_i^-\tilde{X})^{-1}\tilde{X}'\tilde{\Sigma}_i^-$. Let $\tilde{d} = (0, d')'$ and $\tilde{b} = (0, b')'$ with b and d arbitrary $T \times 1$ vectors. Then $\tilde{d}'\tilde{\Sigma}_i^-\tilde{b} = d'\Sigma_i^{-1}b$, and with $\tilde{X} = \begin{pmatrix} 1 & 0 \\ e & X \end{pmatrix}$, we find

$$\begin{aligned} \tilde{X}'\tilde{\Sigma}_i^-\tilde{X} &= \begin{pmatrix} 1 & 0 \\ e & X \end{pmatrix}'\begin{pmatrix} 1+e'\Sigma_i^{-1}e & -e'\Sigma_i^{-1} \\ -\Sigma_i^{-1}e & \Sigma_i^{-1} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ e & X \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & X'\Sigma_i^{-1}X \end{pmatrix}. \end{aligned}$$

It follows that $\tilde{d}'\tilde{\Sigma}_i^-\tilde{X}(\tilde{X}'\tilde{\Sigma}_i^-\tilde{X})^{-1}\tilde{X}'\tilde{\Sigma}_i^-\tilde{b} = d'\Sigma_i^{-1}X(X'\Sigma_i^{-1}X)^{-1}X'\Sigma_i^{-1}b$. Define $\Psi_i = X'\Sigma_i^{-1}X$, $D_i = \lambda^{-1} + \Upsilon_i'\Sigma_i^{-1}\Upsilon_i - \Upsilon_i'\Sigma_i^{-1}X\Psi_i^{-1}X'\Sigma_i^{-1}\Upsilon_i$, e_1 the $T \times 1$ vector $(1, 0, \cdots, 0)'$ and

$$\Phi_i = D_i^{-1}\begin{pmatrix} 1 & \Upsilon_i'\Sigma_i^{-1}X\Psi_i^{-1} \\ \Psi_i^{-1}X'\Sigma_i^{-1}\Upsilon_i & \Psi_i^{-1}X'\Sigma_i^{-1}\Upsilon_i\Upsilon_i'\Sigma_i^{-1}X\Psi_i^{-1} + D_i\Psi_i^{-1} \end{pmatrix}.$$

With these definitions,

$$\tilde{y}^{\mu\prime}G_i\tilde{y}^\mu - \frac{(\tilde{\Upsilon}_i'G_i\tilde{y}^\mu)^2}{\lambda^{-1}+\tilde{\Upsilon}_i'G_i\tilde{\Upsilon}_i} = y^{\mu\prime}\Sigma_i^{-1}y^\mu - \begin{pmatrix} -\Upsilon_i'\Sigma_i^{-1}y^\mu \\ X'\Sigma_i^{-1}y^\mu \end{pmatrix}'\Phi_i\begin{pmatrix} -\Upsilon_i'\Sigma_i^{-1}y^\mu \\ X'\Sigma_i^{-1}y^\mu \end{pmatrix}.$$

Then

$$\begin{aligned}
\Phi_i^{-1} &= \begin{pmatrix} \lambda^{-1} + \Upsilon_i'\Sigma_i^{-1}\Upsilon_i & -\Upsilon_i'\Sigma_i^{-1}X \\ -X'\Sigma_i^{-1}\Upsilon_i & \Psi_i \end{pmatrix} \\
&= \tilde{X}' \begin{pmatrix} \lambda^{-1} + r^2 e_1' V^{-1} e_1 & -r e_1' V^{-1} A_i \\ -r A_i' V^{-1} e_1 & \Sigma_i^{-1} \end{pmatrix} \tilde{X} \\
&= \tilde{X}'\tilde{\Omega}_i^{-1}\tilde{X}
\end{aligned}$$

and

$$\begin{pmatrix} -\Upsilon_i'\Sigma_i^{-1}y^\mu \\ X'\Sigma_i^{-1}y^\mu \end{pmatrix} = \tilde{X}'\tilde{\Omega}_i^{-1}\tilde{y}^\mu$$

and finally

$$y^{\mu\prime}\Sigma_i^{-1}y^\mu = \tilde{y}^{\mu\prime}\tilde{\Omega}_i^{-1}\tilde{y}^\mu$$

so that the equality is established.

Proof of Theorem 2:
We first prove that $v_0(r)$ is necessarily finite. We have

$$\begin{aligned}
v_0(r) &= \mathrm{Var}\left[\sum_{s=0}^{\infty} r^s \nu_{-s}\right] \\
&= \sum_{s=0}^{\infty} r^{2s}\gamma(0) + 2\sum_{s=0}^{\infty} r^{2s+1}\gamma(1) + \cdots + 2\sum_{s=0}^{\infty} r^{2s+j}\gamma(j) + \cdots \\
&= \frac{\gamma(0) + 2\sum_{k=1}^{\infty} r^k\gamma(k)}{1-r^2} \leq 2\frac{\sum_{k=0}^{\infty}|\gamma(k)|}{1-r^2} < \infty
\end{aligned}$$

Now just as in the proof of Theorem 1, $\tilde{M}\tilde{y}$ is a maximal invariant. The aim is thus to optimally discriminate between the two multivariate normals $\tilde{M}\tilde{y}|H_1 \sim N(0, \tilde{M}\bar{\Omega}_1\tilde{M})$ and $\tilde{M}\tilde{y}|H_0 \sim N(0, \tilde{M}\tilde{\Omega}_0\tilde{M})$. The following statistic optimally discriminates between two multivariate normals with singular variance-covariance matrices which share the same column space (Rao and Mitra (1971), p. 206)

$$\bar{Q}(r) = \tilde{y}'\tilde{M}(\tilde{M}\bar{\Omega}_1\tilde{M})^-\tilde{M}\tilde{y} - \tilde{y}'\tilde{M}(\tilde{M}\tilde{\Omega}_0\tilde{M})^-\tilde{M}\tilde{y}$$

where $(\ )^-$ is any generalized inverse. The proof of Theorem 1 shows that $\bar{J}_0 = G_0$ is a generalized inverse of $\tilde{M}\tilde{\Omega}_0\tilde{M}$, and the same line of argument also shows that $\bar{J}_1$ is a generalized inverse of $\tilde{M}\bar{\Omega}_1\tilde{M}$.

Proof of Lemma 1:
The Lemma is similar to Lemma 2 in Elliott (1999). By recursive substitution,

$$u_t = \sum_{j=1}^{t} \rho^{t-j}\nu_j + \alpha(2\gamma)^{-1/2}\omega T^{1/2}(1-\rho^t).$$

But $\alpha(2\gamma)^{-1/2}\omega(1-\rho^{[Ts]}) \to \omega\alpha(1-e^{-\gamma s})(2\gamma)^{-1/2}$ uniformly in s and by Phillips (1987b) and the Continuous Mapping Theorem

$$T^{-1/2}\sum_{j=1}^{[Ts]} \rho^{t-j}\nu_j \Rightarrow \omega\int_0^s e^{-\gamma(s-\lambda)}dW(\lambda),$$

so that the result for $\gamma > 0$ follows from the Continuous Mapping Theorem. Furthermore, as $\gamma \to 0$, we find $\omega\alpha(1 - e^{-\gamma s})(2\gamma)^{-1/2} \to 0$ uniformly in s by applying l'Hôpital's rule.

Asymptotic Theory:
Notation: If A is a $T \times q$ matrix $A = [a_{tj}]$, then let A_{-1} be the $T \times q$ matrix $A_{-1} \equiv B = [b_{tj}]$, where for $1 < t \leq T$, $b_{tj} = a_{t-1,j}$ and $b_{1j} = 0$. Furthermore, let $\Delta A \equiv A - A_{-1}$. If not stated otherwise, '$\to$' and '$\xrightarrow{p}$' are understood to denote the limit and limit in probability as $T \to \infty$.

The proof of the asymptotic results consists of two major steps: First, we investigate the effect of the (known) variance-covariance matrix V of ν_t. Elliott, Rothenberg, and Stock (1996) have shown that for bounded ξ (which correspond to $\alpha = o_p(1)$), the asymptotic distributions of the various terms which make up the optimal test statistic in this case only depend on V through the variance $\gamma(0)$ and the long run variance ω^2 of ν_t. As we will show, the same is true for the test statistics considered here. The second step then consists of applying Lemma 1 to derive the limiting distributions of the various test statistics in terms of $M(s)$.

Lemma 3. *Let $V = [V_{ij}] = [\gamma(i-j)]$ and $\Psi = [\Psi_{ij}] = [\rho(i-j)]$ be $T \times T$ Toeplitz matrices from $\gamma(k)$ and $\rho(k)$, the Fourier coefficients of the spectral density function $f_\nu(\cdot)$ of ν_t, $2\pi f_\nu(\lambda)$ and $[2\pi f_\nu(\lambda)]^{-1}$, respectively. If the elements of the $T \times 1$ vector b are bounded in absolute value, then, under Condition 1,*

$$T^{-1}b'(V^{-1} - \Psi)b \to 0$$

Proof. See Elliott, Rothenberg, and Stock (1996), Lemma A1. □

Lemma 4. *Let $d = (d_1, \cdots, d_T)'$ and $b = (b_1, \cdots b_T)'$ be two $T \times 1$ non-stochastic vectors such that for all $s \in [T^{-1}, 1]$, $d_{[Ts]} = O(1)$ and $b_{[Ts]} = O(1)$, and that for all $s \in [2T^{-1}, 1]$, $T\Delta d_{[Ts]} = O(1)$. Then under the conditions of Lemma 1*

(i)
$$T^{-1}d'(V^{-1} - \omega^{-2}I)b \to 0$$

(ii)
$$T^{-3/2}d'(V^{-1} - \omega^{-2}I)u_{-1} \xrightarrow{p} 0$$

(iii)
$$T^{-1/2}d'(V^{-1} - \omega^{-2}I)\nu \xrightarrow{p} 0$$

(iv)
$$T^{-1/2}d'(V^{-1} - \omega^{-2}I)\Delta u \xrightarrow{p} 0$$

(v)
$$T^{-2}u_{-1}'(V^{-1} - \omega^{-2}I)u_{-1} \xrightarrow{p} 0$$

(vi)
$$2T^{-1}\Delta u'V^{-1}u_{-1} + 1 - \omega^{-2}\left[2T^{-1}\Delta u'u_{-1} + \gamma(0)\right] \xrightarrow{p} 0$$

Proof. The proof of the Theorem draws heavily on the Appendix of Elliott, Rothenberg, and Stock (1996). For any matrix B, let $|B| = tr^{1/2}[B'B]$ and note that for real conformable matrices B and C, $|tr(BC)| \leq |B||C|$, and $|BC| \leq |B|r(C) \leq |B||C|$, where $r(C)$ is the largest characteristic root of C. Define the $T \times T$ matrices $\Lambda \equiv \omega V^{-1/2} - \omega^{-1}V^{1/2}$, Ψ as in Lemma 3 and $F^c = [F^c_{ij}]$ with $F^c_{ij} = c^{i-j}$ if $i > j$ and 0 otherwise. Furthermore, for $\gamma > 0$, let $\hat{u} = w - \xi R(\rho) = u - \xi\Upsilon(\rho) =$

$u - \alpha\omega(2\gamma)^{-1/2}T^{1/2}\Upsilon(\rho)$ such that $\hat{u}$ is the disturbance vector purged of the influence of ξ. Elliott, Rothenberg, and Stock (1996) show in their Appendix that under Condition 1, $0 < \omega < \infty$, $\omega^2 = \sum_{k=-\infty}^{\infty}\gamma(k)$, $\omega^{-2} = \sum_{k=-\infty}^{\infty}\rho(k)$, both $\gamma(k)$ and $\rho(k)$ are absolutely summable, $T^{-1}|\Lambda F^c| \to 0$ for all $0 < c \leq 1$, $r(V) = O(1)$, $r(V^{-1}) = O(1)$, $T^{-2}\hat{u}'_{-1}(V^{-1} - \omega^{-2}I)\hat{u}_{-1} \xrightarrow{p} 0$ and $2T^{-1}\Delta\hat{u}'V^{-1}\hat{u}_{-1} + 1 - \omega^{-2}\left[2T^{-1}\Delta\hat{u}'\hat{u}_{-1} + \gamma(0)\right] \xrightarrow{p} 0$.

We proceed by first proving that $T^{-1/2}|\Lambda d| \to 0$. To this end, let ι be the $T \times 1$ vector $\iota = [\iota_t]$, where for $1 \leq t < T$, $\iota_t = d_{t+1} - d_t$ and $\iota_T = 0$, so that $d = F^1\iota + d_1 e$, where d_1 is the initial element of d. Note that $\iota = O(T^{-1})$. Now

$$\begin{aligned} T^{-1}|\Lambda d|^2 &= T^{-1}d'\Lambda^2 d \\ &\leq T^{-1}|(d + d_1 e)'\Lambda^2 F^1\iota| + T^{-1}d_1^2 e'\Lambda^2 e \end{aligned}$$

and with $\Lambda = \omega V^{-1/2} - \omega^{-1}V^{1/2}$,

$$\begin{aligned} T^{-1}|\Lambda d|^2 &\leq T^{-1}\omega|(d + d_1 e)'V^{-1/2}\Lambda F^1\iota| + T^{-1}\omega^{-1}|(d + d_1 e)'V^{1/2}\Lambda F^1\iota| \\ &\quad + T^{-1}d_1^2 e'\Lambda^2 e \\ &\leq T^{-1}\omega r(V^{-1/2})|\Lambda F^1||\iota(d + d_1 e)'| + T^{-1}\omega^{-1}r(V^{1/2})|\Lambda F^1||\iota(d + d_1 e)'| \\ &\quad + T^{-1}d_1^2|e'(\omega^{-2}V + \omega^2 V^{-1} - 2I)e|. \end{aligned}$$

But $r(V^{-1/2}) = r^{1/2}(V^{-1}) = O(1)$, $r(V^{1/2}) = r^{1/2}(V) = O(1)$, $|\iota(d + d_1 e)'| = [\iota'\iota(d + d_1 e)'(d + d_1 e)]^{1/2} = O(1)$ and $T^{-1}|\Lambda F^1| \to 0$ by the result of Elliott, Rothenberg, and Stock (1996), so that we are left to show that the final element in the sum above converges to zero, too. Now using $\omega^{-2} = \sum_{k=-\infty}^{\infty}\rho(k)$ and $\omega^2 = \sum_{k=-\infty}^{\infty}\gamma(k)$, we find

$$\begin{aligned} T^{-1}|e'(\Psi - \omega^{-2}I)e| &= T^{-1}\left|2\sum_{k=1}^{T-1}(T - k)\rho(k) - 2T\sum_{k=1}^{\infty}\rho(k)\right| \\ &= 2\left|T^{-1}\sum_{k=1}^{T-1}k\rho(k) + \sum_{k=T}^{\infty}\rho(k)\right| \\ &\leq 2\sum_{k=1}^{\infty}\min\left(\frac{k}{T}, 1\right)|\rho(k)| \to 0 \end{aligned}$$

and

$$T^{-1}|e'(V - \omega^2 I)e| \leq 2\sum_{k=1}^{\infty}\min\left(\frac{k}{T}, 1\right)|\gamma(k)| \to 0$$

from the absolute summability of the sequences $\rho(k)$ and $\gamma(k)$. The result now follows from Lemma 3.

(i)

$$\begin{aligned} T^{-1}|b'(V^{-1} - \omega^{-2}I)d| &= T^{-1}\omega^{-1}|b'V^{-1/2}\Lambda d| \\ &\leq T^{-1}\omega^{-1}|b'|r(V^{-1/2})|\Lambda d| \\ &= T^{-1}\omega^{-1}(b'b)^{1/2}r(V^{-1/2})|\Lambda d| \to 0 \end{aligned}$$

since $b'b = O(T)$.

(ii) We first treat the case $\gamma > 0$. With $u_{-1} = \rho^{-1} F^{\rho} \nu + T^{1/2} \alpha \omega (2\gamma)^{-1/2} \Upsilon(\rho)_{-1}$, we have that

$$\begin{aligned} T^{-3/2} d'(V^{-1} - \omega^{-2} I) u_{-1} &= T^{-3/2} \rho^{-1} d'(V^{-1} - \omega^{-2} I) F^{\rho} \nu \\ &\quad + T^{-1} \alpha (2\gamma)^{-1/2} \omega d'(V^{-1} - \omega^{-2} I) \Upsilon(\rho)_{-1}. \end{aligned}$$

But $\Upsilon(\rho)_{-1}$ satisfies the assumptions made on the vector b, so that the second term goes to zero by (i). The variance of the first term is

$$\begin{aligned} \mathrm{Var}[T^{-3/2} \rho^{-1} d'(V^{-1} - \omega^{-2} I) F^{\rho} \nu] &= T^{-3} \rho^{-2} \omega^{-2} d' V^{-1/2} \Lambda F^{\rho} V F^{\rho\prime} \Lambda V^{-1/2} d \\ &\leq T^{-3} \rho^{-2} \omega^{-2} |\Lambda F^{\rho}|^2 |dd'| r^2(V^{-1/2}) r(V) \to 0 \end{aligned}$$

since $T^{-1}|dd'| = T^{-1} d'd = O(1)$. For $\gamma = 0$, $u_{-1} = F^1 \nu$, so that the result follows from setting $\rho = 1$ in (6.1).

(iii) We prove convergence in mean square

$$\mathrm{Var}[T^{-1/2} d'(V^{-1} - \omega^{-2} I) \nu] = T^{-1} \omega^{-2} d' \Lambda^2 d \to 0.$$

(iv) For $\gamma = 0$, $\Delta u = \nu$, and the result follows immediately from (iii). For $\gamma > 0$, we find $\Delta u = \nu - \gamma T^{-1}(u_{-1} + \alpha \omega (2\gamma)^{-1/2} T^{1/2} e)$, so that we find the desired convergence by applying (i), (ii) and (iii).

(v) and (vi): For $\gamma = 0$, $\hat{u} = u$, and we are back to the analysis of Elliott, Rothenberg, and Stock (1996). For $\gamma > 0$, we have $u_{-1} = \hat{u}_{-1} + \alpha \omega (2\gamma)^{-1/2} T^{1/2} \Upsilon(\rho)_{-1}$ and $\Delta u = \Delta \hat{u} + \alpha \omega (2\gamma)^{-1/2} T^{1/2} \Delta \Upsilon(\rho) = \Delta \hat{u} - \alpha \omega \gamma (2\gamma)^{-1/2} T^{-1/2} \rho^{-1} R(\rho)$. But $\Upsilon(\rho)_{-1}$ and $R(\rho)$ both satisfy the necessary assumptions for applying parts (i), (ii) and (iv) to the respective pieces, so that the result follows. □

Lemma 5. *Under Condition 1, if b is a $T \times 1$ $O_p(1)$ vector, then $b'V^{-1}\eta = O_p(1)$.*

Proof. The proof will be carried out in the framework developed in the Appendix of Elliott, Rothenberg, and Stock (1996) and already employed in Lemma 4. Define the $T \times T$ matrix $D = I - \Psi V$. For a real $T \times p$ matrix $B = [b_{tj}]$, let $||B|| = \sum_{t=1}^{T} \sum_{j=1}^{p} b_{tj}$. Then $|B| \leq ||B||$. Furthermore, Elliott, Rothenberg, and Stock (1996) argue at the beginning of their Appendix that under Condition 1, $\sum_{k=0}^{\infty} |k\gamma(k)| < \infty$ and $||D|| = O(1)$. Note that these inequalities imply that the sequence η_t is absolutely summable, since

$$\begin{aligned} \sum_{t=1}^{\infty} \sum_{k=0}^{\infty} r^k \gamma(k+t) &\leq \sum_{t=1}^{\infty} \sum_{k=0}^{\infty} |\gamma(k+t)| \\ &= \sum_{k=0}^{\infty} |k\gamma(k)| < \infty. \end{aligned}$$

We can write

$$\begin{aligned} |b'V^{-1}\eta| &= |b'(\Psi + DV^{-1})\eta| \\ &\leq |b'\Psi\eta| + |\eta' V^{-1} D' b|. \end{aligned}$$

Now since any element of b is bounded in probability and the sequence $\rho(k)$ is absolutely summable, we have that every element of $b'\Psi$ is bounded in probability. But the absolute summability of the sequence η_t then implies that $|b'\Psi\eta|$ is $O_p(1)$, too.

For the second term, first note that boundedness in probability of b_t together with boundedness of $||D||$ implies boundedness in probability of $||D'b||$. Furthermore, the absolute summability of the sequence η_t implies boundedness of $\eta'\eta$. We hence find

$$\begin{aligned}|\eta'V^{-1}D'b| &\le |\eta'V^{-1}||D'b| \\ &\le |\eta'|r(V^{-1})||D'b|| \\ &= (\eta'\eta)^{1/2}r(V^{-1})||D'b|| = O_p(1).\end{aligned}$$

□

Lemma 6. *Let b and d be $T\times 1$ $O_p(1)$ vectors, and define $\tilde b = (b_1, b')'$ and $\tilde d = (d_1, d')'$. Then, as $T\to\infty$,*

$$\begin{pmatrix} T^{1/2}b_1 \\ T^{-1/2}b \end{pmatrix}' (\bar V^{-1} - \tilde V^{-1}) \begin{pmatrix} T^{1/2}d_1 \\ T^{-1/2}d \end{pmatrix} \xrightarrow{p} 0,$$

where $\tilde V = \operatorname{diag}(T\omega^2(2g)^{-1}, V)$ and $\bar V$ is defined in Theorem 2.

Proof. From the formula for partitioned inverses, we have

$$\begin{aligned}\bar V^{-1} &= \begin{pmatrix} v_0(r) & \eta' \\ \eta & V \end{pmatrix}^{-1} \\ &= \delta^{-1}\begin{pmatrix} 1 & -\eta'V^{-1} \\ -V^{-1}\eta & \delta V^{-1} + V^{-1}\eta\eta'V^{-1} \end{pmatrix}\end{aligned}$$

where $\delta = v_0(r) - \eta'V^{-1}\eta$. Furthermore

$$\begin{aligned}&\delta^{-1}\begin{pmatrix} T^{1/2}b_1 \\ T^{-1/2}b \end{pmatrix}' \begin{pmatrix} 1 & -\eta'V^{-1} \\ -V^{-1}\eta & \delta V^{-1} + V^{-1}\eta\eta'V^{-1} \end{pmatrix}\begin{pmatrix} T^{1/2}d_1 \\ T^{-1/2}d \end{pmatrix} \\ &= \delta^{-1}\begin{pmatrix} T^{1/2}b_1 \\ T^{-1/2}b \end{pmatrix}' \begin{pmatrix} T^{1/2}d_1 - T^{-1/2}\eta'V^{-1}d \\ -T^{1/2}d_1V^{-1}\eta + T^{-1/2}\delta V^{-1}d + T^{-1/2}V^{-1}\eta\eta'V^{-1}d \end{pmatrix} \\ &= \delta^{-1}\left[Tb_1d_1 - d_1\eta'V^{-1}b - b_1\eta'V^{-1}d + T^{-1}\delta b'V^{-1}d + T^{-1}b'V^{-1}\eta\eta'V^{-1}d\right].\end{aligned}$$

Apply Lemma 5 and a direct calculation to find $\delta = v_0(r) - \eta'V^{-1}\eta = \frac{\omega^2}{2g}T + o(T)$. Furthermore, use Lemma 5 again in order to show that $T^{-1}\eta'V^{-1}b \xrightarrow{p} 0$, $T^{-1}\eta'V^{-1}d \xrightarrow{p} 0$ and $T^{-1}b'V^{-1}\eta\eta'V^{-1}d \xrightarrow{p} 0$. But

$$\begin{aligned}&\begin{pmatrix} T^{1/2}b_1 \\ T^{-1/2}b \end{pmatrix}' \tilde V^{-1}\begin{pmatrix} T^{1/2}d_1 \\ T^{-1/2}d \end{pmatrix} \\ &= \begin{pmatrix} T^{1/2}b_1 \\ T^{-1/2}b \end{pmatrix}' \begin{pmatrix} T\omega^2(2g)^{-1} & 0 \\ 0 & V \end{pmatrix}^{-1}\begin{pmatrix} T^{1/2}d_1 \\ T^{-1/2}d \end{pmatrix} \\ &= \frac{2g}{\omega^2}b_1d_1 + T^{-1}b'V^{-1}d + o(1)\end{aligned}$$

so that the result follows. □

Lemma 7. *Let $h(s) = (h_1(s), \cdots, h_q(s))'$ be a bounded, twice differentiable vector function $[0,1]\to\mathbb{R}^q$ with $h(0) = (0,\cdots,0)'$ and bounded first and second derivatives $h'(s) = (h_1'(s),\cdots,h_q'(s))'$ and $h''(s) = (h_1''(s),\cdots,h_q''(s))'$. Define $\tilde h(s,c) = ch(s) + h'(s)$ and $\tilde H(s,c) = ch'(s) + h''(s)$. Let $Z = T\Delta X + cX_{-1}$ and $\varphi = \Delta u + cT^{-1}u_{-1}$. If X_t, $t = 1,\cdots,T$, is such that for all $s\in[T^{-1},1]$, $X_{[sT],j}\to h_j(s)$ and $T\Delta X_{[sT],j}\to h_j'(s)$, and for all $s\in[2T^{-1},1]$, $T^2\Delta^2X_{[sT],j}\to h_j''(s)$, then*

under the conditions of Lemma 1
(i)

$$T^{-1}\omega^2 Z'V^{-1}Z \to L(c) = [L(c)_{ij}] = [\int \tilde{h}_i(s,c)\tilde{h}_j(s,c)ds]$$

(ii)

$$T^{-1}\omega^2 cZ'V^{-1}e \to l_e(c) = c\int \tilde{h}(s,c)ds$$

(iii)

$$T^{-1/2}\omega Z'V^{-1}\varphi \Rightarrow m(c) = \tilde{h}(1,c)M(1) - \int [\tilde{H}(s,c) - c\tilde{h}(s,c)]M(s)ds$$

(iv)

$$T^{-1/2}\omega ce'V^{-1}\varphi \Rightarrow m_e(c) = c^2\int M(s)ds + cM(1)$$

Proof. (i) Apply Lemma 4 to find $T^{-1}Z'V^{-1}Z \to T^{-1}\omega^2 Z'Z$. But $Z_{[Ts],j} \to \tilde{h}_i(s,c)$, so that the result follows.

(ii) Follows directly from Lemma 4.

(iii) Use Lemma 4 to establish $T^{-1/2}Z'V^{-1}\varphi - T^{-1/2}\omega^{-2}Z'\varphi \xrightarrow{p} 0$. Then use the identity $Z'\Delta u = Z'u - Z'_{-1}u_{-1} - \Delta Z'u_{-1}$ to write $T^{-1/2}Z'\varphi = T^{-1/2}Z'(\Delta u + cT^{-1}u_{-1}) = T^{-1/2}Z'_T u_T - T^{-1/2}\Delta Z'u_{-1} + cT^{-3/2}Z'u_{-1}$, where Z_T and u_T are the last rows of Z and u, respectively. By definition, $T^{-1/2}\Delta Z'u_{-1} = T^{1/2}(\Delta^2 X)'u_{-1} + cT^{-1/2}\Delta X'u_{-1}$. The first row of $T^2\Delta^2 X$ is possibly not $O(1)$, but since the first element of u_{-1} is zero this is inconsequential. The result now follows from Lemma 1 and the Continuous Mapping Theorem.

(iv) Proceed as in (iii), with Z replaced by ce. □

Theorem 8. *Suppose that $L(0)$ and $L(g)$ of Lemma 7 are nonsingular. Then under the conditions of Lemma 7,*
(i)

$$\begin{aligned} S_a(g,F_a) \Rightarrow\ & P - m(g)'L(g)^{-1}m(g) + m(0)'L(0)^{-1}m(0) \\ & - 2\ln\int_{-\infty}^{\infty} \exp\left\{-\tfrac{1}{2}\left[a^2(2g)^{-1}(g^2 - l_e(g)'L(g)^{-1}l_e(g))\right.\right. \\ & \left.\left. +2a(2g)^{-1/2}(m_e(g) - l_e(g)'L(g)^{-1}m(g))\right]\right\} dF_a(a) \end{aligned}$$

(ii)

$$\begin{aligned} \mathrm{Env}_a(g,a) \Rightarrow\ & P - m(g)'L(g)^{-1}m(g) + m(0)'L(0)^{-1}m(0) \\ & + a^2(2g)^{-1}(g^2 - l_e(g)'L(g)^{-1}l_e(g)) + 2a(2g)^{-1/2}(m_e(g) - l_e(g)'L(g)^{-1}m(g)) \end{aligned}$$

(iii)

$$\begin{aligned} Q_a(g,k) \Rightarrow\ & P + m(0)'L(0)^{-1}m(0) \\ & - \begin{pmatrix} k^{1/2}m_e(g) \\ m(g) \end{pmatrix}' \begin{pmatrix} 2g + g^2k & k^{1/2}l_e(g)' \\ k^{1/2}l_e(g) & L(g) \end{pmatrix}^{-1} \begin{pmatrix} k^{1/2}m_e(g) \\ m(g) \end{pmatrix} \end{aligned}$$

(iv) the asymptotic distribution of $\bar{Q}_a(g)$ is the same as the asymptotic distribution of $Q_a(g,1)$

where $P = gM(1)^2 - g + g^2 \int M(s)^2 ds$, *and* $m(\cdot)$, $l_e(\cdot)$ *and* $m_e(\cdot)$ *are defined in Lemma 7.*

Proof. First note that since $\tilde{G}_i\tilde{X} = 0$ we can set $\tilde{y} = \tilde{u}$ and substitute $\tilde{R}_1$ by $\tilde{\Upsilon}_1 = \tilde{\Upsilon}(r)$ in the expressions of the test statistics without loss of generality.
We first establish that $\tilde{u}'\tilde{\Sigma}_1^- \tilde{u} - \tilde{u}'\tilde{\Sigma}_0^- \tilde{u} \Rightarrow P$. But the initial element of $\tilde{u}$ is zero, and since $\Sigma_i^{-1} = A_i'V^{-1}A_i$, we find $u'\Sigma_1^{-1}u - u'\Sigma_0^{-1}u = 2gT^{-1}\Delta u'V^{-1}u_{-1} + g^2T^{-2}u_{-1}'V^{-1}u_{-1}$. Applying Lemma 4 yields $2gT^{-1}\Delta u'V^{-1}u_{-1} + g^2T^{-2}u_{-1}'V^{-1}u_{-1} + g - \omega^{-2}[2gT^{-1}\Delta u'u_{-1} + g\gamma(0) + g^2T^{-2}u_{-1}'u_{-1})] \xrightarrow{p} 0$. But $2T^{-1}\Delta u'u_{-1} = T^{-1}u_T^2 - T^{-1}\sum_{t=1}^T (\Delta u_{t-1})^2 \Rightarrow \omega^2 M(1)^2 - \gamma(0)$, so that the result follows from Lemma 1 and the Continuous Mapping Theorem.
(i) and (ii): We need to calculate the asymptotic distributions of $\tilde{d}'\tilde{\Sigma}_i^{-1}\tilde{b}$ and $\tilde{d}'\tilde{\Sigma}_i^- \tilde{X}(\tilde{X}'\tilde{\Sigma}_i^- \tilde{X})^{-1}\tilde{X}'\tilde{\Sigma}_i^- \tilde{b}$, where $\tilde{d} = (0, d')'$ and $\tilde{b} = (0, b')'$ and d and b stand for any combination of u and $\omega T^{1/2}\Upsilon_1$. With $\tilde{\Sigma}_i^- = \begin{pmatrix} 1 + e'\Sigma_i^{-1}e & -e'\Sigma_i^{-1} \\ -\Sigma_i^{-1}e & \Sigma_i^{-1} \end{pmatrix}$ and $\tilde{X} = \begin{pmatrix} 1 & 0 \\ e & X \end{pmatrix}$, we find $\tilde{d}'\tilde{\Sigma}_i^- \tilde{b} = d'\Sigma_i^{-1}b$, $\tilde{X}'\tilde{\Sigma}_i^{-1}\tilde{X} = \begin{pmatrix} 1 & 0 \\ 0 & X'\Sigma_i^{-1}X \end{pmatrix}$ and $\tilde{X}'\tilde{\Sigma}_i^- \tilde{d} = \begin{pmatrix} 0 \\ X'\Sigma_i^{-1}d \end{pmatrix}$, so that

$$\tilde{d}'\tilde{\Sigma}_i^- \tilde{X}(\tilde{X}'\tilde{\Sigma}_i^- \tilde{X})^{-1}\tilde{X}'\tilde{\Sigma}_i^- \tilde{b} = d'\Sigma_i^{-1}X(X'\Sigma_i^{-1}X)^{-1}X'\Sigma_i^{-1}b.$$

Since $\Sigma_i^{-1} = A_i'V^{-1}A_i$, $d'\Sigma_i^{-1}b = \delta_i'V^{-1}\beta_i$ and $d'\Sigma_i^{-1}X(X'\Sigma_i^{-1}X)^{-1}X'\Sigma_i^{-1}b = \delta_i'V^{-1}Z_i(Z_i'V^{-1}Z_i)^{-1}Z_i'V^{-1}\beta_i$, where $Z_0 = T\Delta X$, $Z_1 = T\Delta X + gX_{-1}$, $\delta_0 = \Delta d$, $\beta_0 = \Delta b$, $\delta_1 = \Delta d + gT^{-1}d_{-1}$ and $\beta_1 = \Delta b + gT^{-1}b_{-1}$. Note that for $d = \omega T^{1/2}\Upsilon_1$, δ_1 becomes $\omega T^{1/2}\Delta\Upsilon_1 + \omega g T^{-1/2}(\Upsilon_1)_{-1} = -\omega g T^{-1/2}e$. The function $f(x,y) = \int_{-\infty}^{\infty} \exp\{ay - a^2x\}dF_a(a)$ is continuous in x and y for any cumulative distribution function $F_a(\cdot)$. Now apply Lemma 1, 4 and Lemma 7 with $c = g$ and $c = 0$ to the various terms with d equal to u or $T^{1/2}\omega\Upsilon_1$ and b equal to u or $T^{1/2}\omega\Upsilon_1$, and parts (i) and (ii) follow from their joint convergence and the Continuous Mapping Theorem.

(iii) We first consider the case $k > 0$, and we rely on the alternative expression (3.3) for $Q_a(g,k)$. With $u_0 = 0$, clearly $\hat{u}(1)'\tilde{\Omega}(1)^{-1}\hat{u}(1)$ is equal to $\tilde{u}'G_0\tilde{u}$ so that we can rely on the derivations above for this part. Define $\tilde{\Omega}_1 = \tilde{\Omega}(r)$ and $\tilde{A}_1 = \tilde{A}(r)$. Then $\hat{u}(r)'\tilde{\Omega}(r)^{-1}\hat{u}(r) = \tilde{u}'[\tilde{\Omega}_1^{-1} - \tilde{\Omega}_1^{-1}\tilde{X}(\tilde{X}'\tilde{\Omega}_1^{-1}X)^{-1}\tilde{X}'\tilde{\Omega}_1^{-1}]\tilde{u}$ and $\tilde{\Omega}_1^{-1} = \tilde{A}_1'\tilde{V}^{-1}\tilde{A}_1$, where $\tilde{V} = \begin{pmatrix} k\omega^2(2g)^{-1}T & 0 \\ 0 & V \end{pmatrix}$. From $u_0 = 0$, we find $\tilde{u}'\tilde{\Omega}_1^{-1}\tilde{u} = u'\Sigma_1^{-1}u$. Furthermore $T^{1/2}\tilde{A}_1\tilde{X} = \begin{pmatrix} T^{1/2} & 0 \\ T^{-1/2}ge & T^{-1/2}Z_1 \end{pmatrix}$ and $\tilde{A}_1\tilde{u} = \begin{pmatrix} 0 \\ \varphi_1 \end{pmatrix}$, where $\varphi_1 = \Delta u + gT^{-1}u_{-1}$, so that

$$T\tilde{X}'\tilde{\Omega}_1^{-1}\tilde{X} = \begin{pmatrix} 2gk^{-1}\omega^{-2} + g^2T^{-1}e'V^{-1}e & gT^{-1}e'V^{-1}Z_1 \\ gT^{-1}Z_1'V^{-1}e & T^{-1}Z_1'V^{-1}Z_1 \end{pmatrix}$$

and

$$T^{1/2}\tilde{X}'\tilde{\Omega}_1^{-1}\tilde{u} = \begin{pmatrix} gT^{-1/2}e'V^{-1}\varphi_1 \\ T^{-1/2}Z_1'V^{-1}\varphi_1 \end{pmatrix}.$$

From Lemma 4 and 7, $T^{-1}e'V^{-1}e \to \omega^{-2}$, $T^{-1}Z_1'V^{-1}e \to l_e(g)$, $T^{-1/2}e'V^{-1}\varphi_1 \Rightarrow m_e(g)$ and $T^{-1/2}Z_1'V^{-1}\varphi_1 \Rightarrow m(g)$. The result for $k > 0$ now follows from Lemma 1, the joint convergence of these pieces and the Continuous Mapping Theorem.

Clearly, $Q_a(g,0)$ corresponds to $\text{Env}_a(g,0)$. But setting $k=0$ in the general expression for $Q_a(g,k)$ yields $\text{Env}_a(g,0)$, so that the expression for $Q_a(g,k)$ holds true also at $k=0$.

(iv) The only difference between $\bar{Q}_a(g)$ and $Q_a(g,1)$ would arise from the presence of the variance-covariance matrix $\bar{V}$ rather than $\tilde{V}$ with $\lambda = T\omega^2(2g)^{-1}$ in part (iii). But Lemma 6 shows that this difference does not matter asymptotically, so that the two asymptotic distributions are identical. □

In the light of this Theorem, Theorem 3 in the main text becomes a corollary. For the mean case, there is no X, so that the result follows directly after substituting all expressions:

$$\begin{aligned} S_a^\mu(g,F_a) &\Rightarrow gM(1)^2 - g + g^2 \int M(s)^2 ds \\ &\quad - 2\ln \int \exp\left\{-\frac{1}{2}\left[\frac{1}{2}a^2 g + \sqrt{2}a\left(g^{3/2}\int M(s)ds + g^{1/2}M(1)\right)\right]\right\} dF_a(a) \end{aligned}$$

$$\text{Env}_a^\mu(g,a) \Rightarrow gM(1)^2 - g + g^2 \int M(s)^2 ds + \frac{1}{2}a^2 g + \sqrt{2}a\left(g^{3/2}\int M(s)ds + g^{1/2}M(1)\right)$$

$$Q_a^\mu(g,k) \Rightarrow gM(1)^2 - g + g^2 \int M(s)^2 ds - \frac{gk\left(g\int M(s)ds + M(1)\right)^2}{2+gk}$$

For the time trend case we can set $X = T^{-1}\tau$, so that for $s \in [T^{-1},1]$, $X_{[Ts]} = T^{-1}[Ts] \to s = h(s)$ and $T\Delta X_{[Ts]} = 1 = h'(s)$, and for $s \in [2T^{-1},1]$, $T^2\Delta^2 X_{[Ts]} = 0 = h''(s)$. Therefore $\tilde{h}(s,c) = ch(s) + h'(s)$ and $\tilde{H}(s,c) = ch'(s) + h''(s)$ become $\tilde{h}(s,c) = cs+1$ and $\tilde{H}(s,c) = c$. We find $L(c) = \int \tilde{h}(s,c)^2 ds = \int (c^2s^2 + 2cs + 1)ds = c^2/3 + c + 1$, $l_e(c) = c\int \tilde{h}(s,c)ds = c\int(cs+1)ds = c^2/2 + c$ and $m(c) = \tilde{h}(1,c)M(1) - \int[\tilde{H}(s,c) - c\tilde{h}(s,c)]M(s)ds = (c+1)M(1) + c^2\int sM(s)ds$. Since all considered statistics are invariant to the time trend, we might substitute $M(s)$ with its projection off s, $M^\tau(s) \equiv M(s) - 3s\int \lambda M(\lambda)d\lambda$. Clearly then $\int sM^\tau(s)ds = 0$. Plugging in these terms in the expressions of Theorem 8 yields

$$\begin{aligned} S_a^\tau(g,F_a) &\Rightarrow (g+1)M^\tau(1)^2 - g + g^2\int M^\tau(s)^2 ds - \frac{(g+1)^2 M^\tau(1)^2}{\frac{1}{3}g^2 + g + 1} \\ &\quad - 2\ln \int_{-\infty}^{\infty} \exp\left\{-\tfrac{1}{2}\left[a^2(2g)^{-1}\left(g^2 - \frac{(\frac{1}{2}g^2+g)^2}{\frac{1}{3}g^2+g+1}\right)\right.\right. \\ &\quad \left.\left. +\sqrt{2}ag^{-1/2}\left(g^2\int M^\tau(s)ds + gM^\tau(1) - \frac{(\frac{1}{2}g^2+g)(g+1)M^\tau(1)}{\frac{1}{3}g^2+g+1}\right)\right]\right\} dF_a(a) \end{aligned}$$

$$\begin{aligned} \text{Env}_a^\tau(g,a) &\Rightarrow (g+1)M^\tau(1)^2 - g + g^2\int M^\tau(s)^2 ds - \frac{(g+1)^2 M^\tau(1)^2}{\frac{1}{3}g^2+g+1} \\ &\quad + a^2(2g)^{-1}\left(g^2 - \frac{(\frac{1}{2}g^2+g)^2}{\frac{1}{3}g^2+g+1}\right) \\ &\quad + \sqrt{2}ag^{-1/2}\left(g^2\int M^\tau(s)ds + gM^\tau(1) - \frac{(\frac{1}{2}g^2+g)(g+1)M^\tau(1)}{\frac{1}{3}g^2+g+1}\right) \end{aligned}$$

$$Q_a^\tau(g,k) \Rightarrow (g+1)M^\tau(1)^2 - g + g^2 \int M^\tau(s)^2 ds -$$
$$\begin{pmatrix} k^{1/2}(g^2 \int M^\tau(s)ds + gM^\tau(1)) \\ (g+1)M^\tau(1) \end{pmatrix}' \begin{pmatrix} 2g+g^2k & k^{1/2}(\frac{1}{2}g^2+g) \\ k^{1/2}(\frac{1}{2}g^2+g) & \frac{1}{3}g^2+g+1 \end{pmatrix}^{-1}$$
$$\times \begin{pmatrix} k^{1/2}(g^2 \int M^\tau(s)ds + gM^\tau(1)) \\ (g+1)M^\tau(1) \end{pmatrix}$$

The coefficients given in Theorem 3 in the main text follow after a considerable amount of algebra.

It is now easy to analyze the GLS estimator $\hat{\alpha}_{GLS} = (2g)^{1/2}\omega^{-1}T^{-1/2}(y_0 - \hat{\mu})$ of α mentioned in the main text. Define the vector $s_1 = (1, 0, \cdots, 0)'$ which has the same number of rows as $\tilde{X}$ has columns. Then

$$\hat{\mu} - \mu - \xi = s_1'(\tilde{X}'\tilde{\Omega}_1^- \tilde{X})^{-1}\tilde{X}'\tilde{\Omega}_1^- \tilde{u}$$

so that

$$\hat{\alpha}_{GLS} = -(2g)^{1/2}\omega^{-1}T^{-1/2}s_1'(\tilde{X}'\tilde{\Omega}_1^- \tilde{X})^{-1}\tilde{X}'\tilde{\Omega}_1^- \tilde{u}.$$

Let $d \equiv 2g + kg^2 - kl_e(g)'L(g)^{-1}l_e(g)$. From steps very similar to those used in the derivations of Theorem 8, we find

$$\begin{aligned}
\hat{\alpha}_{GLS} &\Rightarrow -(2gk)^{1/2}s_1' \begin{pmatrix} 2g+g^2k & k^{1/2}l_e(g)' \\ k^{1/2}l_e(g) & L(g) \end{pmatrix}^{-1} \begin{pmatrix} k^{1/2}m_e(g) \\ m(g) \end{pmatrix} \\
&= -(2gk)^{1/2}d^{-1}s_1' \begin{pmatrix} 1 & -k^{1/2}l_e(g)'L(g)^{-1} \\ -k^{1/2}L(g)^{-1}l_e(g) & kL(g)^{-1}l_e(g)l_e(g)'L(g)^{-1} + dL(g)^{-1} \end{pmatrix} \\
&\qquad \times \begin{pmatrix} k^{1/2}m_e(g) \\ m(g) \end{pmatrix} \\
&= -(2g)^{1/2}\frac{k(m_e(g) - l_e(g)'L(g)^{-1}m(g))}{2g + k(g^2 - l_e(g)'L(g)^{-1}l_e(g))}.
\end{aligned}$$

Since $\hat{\alpha}_{GLS}$ and the various parts in $\text{Env}_a(g,\alpha)$ converge jointly to their respective weak limits, we can apply the Continuous Mapping Theorem to $\text{Env}_a(g, \hat{\alpha}_{GLS})$. With $S \equiv g^2 - l_e(g)'L(g)^{-1}l_e(g)$ we find

$$\text{Env}_a(g,\hat{\alpha}_{GLS}) \Rightarrow P - m(g)'L(g)^{-1}m(g) + m(0)'L(0)^{-1}m(0)$$
$$+ \frac{k^2S(m_e(g) - l_e(g)'L(g)^{-1}m(g))^2}{(2g+kS)^2} - \frac{2k(m_e(g) - l_e(g)'L(g)^{-1}m(g))^2}{2g+kS}.$$

Furthermore, using the original form for $Q(r,\lambda)$ of Corollary 2, an alternative expression for the asymptotic distribution of $Q_a(g,k)$ is

$$Q_a(g,k) \Rightarrow P - m(g)'L(g)^{-1}m(g) + m(0)'L(0)^{-1}m(0)$$
$$- \frac{k(m_e(g) - l_e(g)'L(g)^{-1}m(g))^2}{2g+kS}.$$

We conclude that $Q_a(g,\tilde{k})$ and $\text{Env}_a(g,\hat{\alpha})$ converge to the same limiting distribution when

$$\tilde{k} = \frac{k(4g+kS)}{2g}.$$

Substituting out for S yields the expressions in the main text.

Similarly, from the expression for $\text{Env}(r,x)$ in Corollary 1, let

$$\hat{\xi}_{ML} = \frac{\tilde{R}_1' G_1 \tilde{y}}{\tilde{R}_1' G_1 \tilde{R}_1} \quad \text{and} \quad \hat{\alpha}_{ML} = \omega^{-1} T^{-1/2} (2g)^{1/2} \hat{\xi}_{ML}$$

be the Maximum Likelihood estimators of ξ and α under the fixed alternative with $\rho = r = 1 - gT^{-1}$, respectively. From the expressions derived in the proof of Theorem 8 we find

$$\hat{\alpha}_{ML} \Rightarrow -(2g)^{1/2} \frac{m_e(g) - l_e(g)' L(g)^{-1} m(g)}{g^2 - l_e(g)' L(g)^{-1} l_e(g)}.$$

Noting that this asymptotic distribution corresponds to the distribution of $\hat{\alpha}_{GLS}$ as $k \to \infty$, we conclude from the discussion above that $\text{Env}_a(g, \hat{\alpha}_{ML})$ is equivalent to $Q_a(g, \infty)$.

Proof of Lemma 2:

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

We have

$$\begin{aligned}
\phi^i(\theta) &= E_\gamma \left[\exp\left\{\theta T^i \mathbf{i}\right\}\right] \\
&= E_\delta[\exp\{\theta l_0^i \mathbf{i} - \frac{1}{2}(\delta - \gamma) - \frac{1}{4}\alpha^2 \gamma \\
&+ \left(\theta l_1^i \mathbf{i} - \frac{\gamma^2 - \delta^2}{2}\right) \int M(s)^2 ds + Z^{i\prime} \tilde{\Lambda} Z^i + \tilde{\lambda}' Z^i\}] \\
&= \exp\left\{\theta l_0 \mathbf{i} - \frac{1}{2}(\delta - \gamma) - \frac{1}{4}\alpha^2 \gamma\right\} E_\delta \left[\exp\left\{Z^{i\prime} \tilde{\Lambda}^i Z^i + \tilde{\lambda}^{i\prime} Z^i\right\}\right],
\end{aligned}$$

where $\delta = \sqrt{\gamma^2 - 2l_1\theta\mathbf{i}}$ and the subscript of the expectation operator denotes the value of γ in the definition of $M(\cdot)$ with respect to which the expectation has to be computed. To justify the change in measure used in the second line, consider the likelihood of u from (2.1) with standard normal innovations, $\gamma = g$ and $\alpha = a$

$$\begin{aligned}
L_g^a &\propto \exp\left\{-\frac{1}{2}\left[\sum_{t=1}^T \left(u_t - ru_{t-1} + T^{1/2} a (2g)^{-1/2}(1-r)\right)^2\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\sum_{t=1}^T \left(\Delta u_t + gT^{-1} u_{t-1} + \tfrac{\sqrt{2}}{2} a g^{1/2} T^{-1/2}\right)^2\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\sum_{t=1}^T \left(\Delta u_t^2 + 2gT^{-1}\Delta u_t u_{t-1} + \sqrt{2} a T^{-1/2} g^{1/2} \Delta u_t \right.\right.\right. \\
&\quad \left.\left.\left. + g^2 T^{-2} u_{t-1}^2 + \sqrt{2} a T^{-3/2} g^{3/2} u_{t-1} + \tfrac{1}{2} a^2 g T^{-1}\right)\right]\right\}
\end{aligned}$$

and the likelihood when $\gamma = d$ and $\alpha = 0$

$$L_d^0 \propto \exp\left\{-\frac{1}{2}\left[\sum_{t=1}^T \left(\Delta u_t^2 + 2\frac{d}{T}\Delta u_t u_{t-1} + \frac{d^2}{T^2} u_{t-1}^2\right)\right]\right\}.$$

Therefore

$$\left.\frac{L_g^a}{L_d^0}\right|_\gamma = \exp\left\{\frac{1}{2}\sum_{t=1}^{T}\left[2\frac{d-g}{T}\Delta u_t u_{t-1} - \sqrt{2}aT^{-1/2}g^{1/2}\Delta u_t\right.\right.$$
$$\left.\left. + \frac{d^2-g^2}{T^2}u_{t-1}^2 - \sqrt{2}aT^{-3/2}g^{3/2}u_{t-1} - \frac{a^2gT^{-1}}{2}\right]\right\}$$
$$\Rightarrow \exp\left\{\frac{1}{2}(d-g)(M(1)^2-1) + \frac{1}{2}(d^2-g^2)\int M(s)^2ds\right.$$
$$\left. -(2g)^{-1/2}a[gM(1) + g^2\int M(s)ds] - \frac{a^2g}{4}\right\}.$$

Note that rather changing the measure to a process with an identical α, as is done in Tanaka (1996), we change it to a process with $\alpha = 0$, which somewhat simplifies our calculations.

The remaining expectation in $\phi^i(\theta)$ may be computed 'by completing the square'

$$E_\delta\left[\exp\left\{Z^{i\prime}\tilde{\Lambda}^i Z^i + \tilde{\lambda}^{i\prime} Z^i\right\}\right]$$
$$= (2\pi)^{-d^i/2}\det{}^{-1/2}V^i(\delta)\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}\left[Z^{i\prime}(V^i(\delta)^{-1} - 2\tilde{\Lambda}^i)Z^i - 2\tilde{\lambda}^{i\prime}Z^i\right]\right\}dZ^i$$
$$= \det{}^{-1/2}\left[I - 2V^i(\delta)\tilde{\Lambda}^i\right]\exp\left\{\frac{1}{2}\tilde{\lambda}^{i\prime}(V^i(\delta)^{-1} - 2\tilde{\Lambda}^i)^{-1}\tilde{\lambda}^i\right\}$$

where $d^\mu = 1$ and $d^\tau = 2$. The last line follows after noting that $E_\delta[Z^i] = 0$ and, with $m^i = (V^i(\delta)^{-1} - 2\tilde{\Lambda}^i)^{-1}\tilde{\lambda}^i$, that $(Z^i - m^i)'(V^i(\delta)^{-1} - 2\tilde{\Lambda}^i)(Z^i - m^i) = Z^{i\prime}(V^i(\delta)^{-1} - 2\tilde{\Lambda}^i)Z^i - 2\tilde{\lambda}^{i\prime}Z^i + \tilde{\lambda}^{i\prime}(V^i(\delta)^{-1} - 2\tilde{\Lambda}^i)^{-1}\tilde{\lambda}^i$.

$V^\mu(\delta)$ and $V^\tau(\delta)$:

The definition of $M(s)$ and some stochastic calculus reveal that with $\alpha = 0$ and $\gamma = \delta$

$$\begin{pmatrix} M(1) \\ \int M(s)ds \\ \int sM(s)ds \end{pmatrix} = \int\begin{pmatrix} \exp[-\delta(1-s)] \\ \{1 - \exp[-\delta(1-s)]\}/\delta \\ \{1 + \delta s - \exp[-\delta(1-s)](1+\delta)\}/\delta^2 \end{pmatrix}dW(s)$$
$$\equiv \int\begin{pmatrix} g_1(s) \\ g_2(s) \\ g_3(s) \end{pmatrix}dW(s).$$

Hence $V^i(\delta) = [v_{j,k}] = \left[\int g_j(s)g_k(s)ds\right]$. Carrying out the integration yields $v_{1,1} = (1 - e^{-2\delta})/(2\delta)$, $v_{1,2} = (1 - e^{-\delta})^2/(2\delta^2)$, $v_{2,2} = (-3 + 2\delta + 4e^{-\delta} - e^{-2\delta})/(2\delta^3)$, $v_{1,3} = (\delta - 1 + (1+\delta)e^{-2\delta})/(2\delta^3)$, $v_{2,3} = (\delta^2 - (1+\delta)(1 - e^{-\delta})^2)/(2\delta^4)$ and $v_{3,3} = (3 - 3\delta^2 + 2\delta^3 - 3(1+\delta)^2e^{-2\delta})/(6\delta^5)$.

Proof of Theorem 5:

The statistic $\breve{Q}_\kappa^i(r,\lambda)$ differs from the statistic $Q^i(r,\lambda)$ only in the density of y under the null hypothesis. Using the notation of the proof of Theorem 1, we have $\tilde{M}\tilde{y}|H_0, \zeta = z \sim N(z\tilde{M}\tilde{Z}_\zeta, \tilde{M}\tilde{\Sigma}_0\tilde{M})$, so that the density of $\tilde{M}\tilde{y}|H_0$ on the hyperplane $\tilde{y}'\tilde{X} = 0$ is proportional to

$$\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left[(\tilde{y} - z\tilde{Z}_\zeta)'\tilde{M}(\tilde{M}\tilde{\Sigma}_0\tilde{M})^{-}\tilde{M}(\tilde{y} - z\tilde{Z}_\zeta) + z^2\kappa^{-1}\right]\right\} dz$$
$$= c\exp\left\{-\frac{1}{2}\left[\tilde{y}'G_0\tilde{y} - \frac{(\tilde{y}'G_0\tilde{Z}_\zeta)^2}{\kappa^{-1} + \tilde{Z}_\zeta' G_0\tilde{Z}_\zeta}\right]\right\}$$

where c is a constant that does not depend on $\tilde{y}$. The optimal test statistic hence $\breve{Q}^i_\kappa(r, \lambda)$ becomes

$$\breve{Q}^i_\kappa(r,\lambda) = Q^i(r,\lambda) + \frac{(\tilde{y}'G^i_0\tilde{Z}^i_\zeta)^2}{\kappa^{-1} + \tilde{Z}^{i\prime}_\zeta G^i_0\tilde{Z}^i_\zeta}$$

and

$$\begin{aligned} \breve{Q}^i(r,\lambda) &= \lim_{\kappa\to\infty} \breve{Q}^i_\kappa(r,\lambda) \\ &= Q^i(r,\lambda) + \frac{(\tilde{y}'G^i_0\tilde{Z}^i_\zeta)^2}{\tilde{Z}^{i\prime}_\zeta G^i_0\tilde{Z}^i_\zeta}. \end{aligned}$$

From the joint convergence of the two pieces of $\breve{Q}^i(r,\lambda)$ and the Continuous Mapping Theorem, the asymptotic distribution of $\breve{Q}^i_a(g,k)$ is given by sum of the asymptotic distribution of $Q^i_a(g,k)$ and the asymptotic distribution of $(\tilde{y}'G^i_0\tilde{Z}^i_\zeta)^2/\tilde{Z}^{i\prime}_\zeta G^i_0\tilde{Z}^i_\zeta$. Since $\breve{Q}^i(r,\lambda)$ is invariant to $\tilde{X}$, we can replace $\tilde{y}$ with $\tilde{u}$. Let $h^\mu(s) = 0$ and $h^\tau(s) = s$. Clearly, $T^{-1/2}\tilde{Z}^\mu_{\zeta,[Ts]} \to s \equiv g^\mu(s)$ and $T^{-1/2}\tilde{Z}^\tau_{\zeta,[Ts]} \to s^2 \equiv g^\tau(s)$, so that from part (i) and (iii) of Lemma 7 and the reasoning in the proof of Theorem 8

$$\begin{aligned} \omega\tilde{u}'G^i_0\tilde{Z}^i_\zeta &= \omega\tilde{u}'[\tilde{\Sigma}_0^{-1} - \tilde{\Sigma}_0^{-}\tilde{X}^i(\tilde{X}^{i\prime}\tilde{\Sigma}_0^{-}\tilde{X}^i)^{-1}\tilde{X}^{i\prime}\tilde{\Sigma}_0^{-}]\tilde{Z}^i_\zeta \\ &\Rightarrow g^{i\prime}(1)M(1) - \int g^{i\prime\prime}(s)M(s)ds \\ &- \left[\left(h^{i\prime}(1)M(1) - \int h^{i\prime\prime}(s)M(s)ds\right)\left(\int h^{i\prime}(s)^2ds\right)^{-1}\int h^{i\prime}(s)g^{i\prime}(s)ds\right] \end{aligned}$$

and

$$\omega^2\tilde{Z}^{i\prime}_\zeta G^i_0\tilde{Z}^i_\zeta \Rightarrow \int g^{i\prime}(s)^2ds - \left(\int h^{i\prime}(s)g^{i\prime}(s)ds\right)^2\left(\int h^{i\prime}(s)^2ds\right)^{-1}.$$

Carrying out the calculation for $i = \mu$ and $i = \tau$ yields

$$\frac{(\tilde{u}'G^i_0\tilde{Z}^i_\zeta)^2}{\tilde{Z}^{i\prime}_\zeta G^i_0\tilde{Z}^i_\zeta} \Rightarrow \begin{cases} M(1)^2 & \text{for } i = \mu \\ 3M^\tau(1)^2 + 12\left(\int M^\tau(s)ds\right)^2 - 12M^\tau(1)\int M^\tau(s)ds & \text{for } i = \tau \end{cases}$$

Combining these results with the asymptotic distribution of $Q^i_a(g,k)$ of Theorem 3, we find that the weights $\breve{q}^i_j$ in the asymptotic distribution of $\breve{Q}^i_a(g,k)$ on 1, $M^i(1)^2$, $\left(\int M^i(s)ds\right)^2$, $M^i(1)\int M^i(s)ds$ and $\int M^i(s)^2ds$ are given by $\breve{q}^i_0 = q^i_0$, $\breve{q}^\mu_1 = q^\mu_1 + 1$, $\breve{q}^\mu_2 = q^\mu_2$, $\breve{q}^\mu_3 = q^\mu_3$, $\breve{q}^i_4 = q^i_4$, and $\breve{q}^\tau_1 = q^\tau_1 + 3$, $\breve{q}^\tau_2 = q^\tau_2 + 12$, $\breve{q}^\tau_3 = q^\tau_3 - 12$, where q^i_j are defined in Theorem 3. With these weights, it is straightforward to establish the solution $g_0 = -2\,\mathrm{cv}^{i,DF}$ and $k_0 \to \infty$ of (4.2).

Proof of Theorem 6:

First note that minimizing $\breve{Q}^i(r,\lambda)$ with respect to r is equivalent to minimizing $\tilde{y}'G_1\tilde{y} - \frac{(\tilde{R}_1'G_1\tilde{y})^2}{\lambda^{-1}+\tilde{R}_1'G_1\tilde{R}_1}$ — only the density under the alternative depends on r. Expression (3.3) allows to rewrite

$$(6.2)\quad \tilde{y}'G_1\tilde{y}-\frac{(\tilde{R}_1'G_1\tilde{y})^2}{\lambda^{-1}+\tilde{R}_1'G_1\tilde{R}_1} = \tilde{y}'[\tilde{\Omega}(r)^{-1}-\tilde{\Omega}(r)^{-1}\tilde{X}(\tilde{X}'\tilde{\Omega}(r)^{-1}\tilde{X})^{-1}\tilde{X}'\tilde{\Omega}(r)^{-1}]\tilde{y},$$

and note that $\tilde{\Omega}(r)^{-1} = \tilde{A}(r)'diag(\lambda^{-1}, V^{-1})\tilde{A}(r)$. Clearly, $k \to \infty$ implies $\lambda^{-1} = 2g\omega^{-2}T^{-1}k^{-1} \to 0$. Now $\lim_{\lambda^{-1}\to 0}\tilde{\Omega}(r)^{-1} = \tilde{A}(r)'diag(0,V^{-1})\tilde{A}(r) = \tilde{B}(r)'\tilde{V}_0^{-}\tilde{B}(r)$, where $\tilde{B}(r) = \tilde{A}(r) - diag(1,0,\cdots,0)$ and $\tilde{V}_0^{-} = diag(0,V^{-1})$. By standard regression algebra, (6.2) as $\lambda^{-1}\to 0$ is hence equal to the sum of squared residuals of a GLS regression of $\tilde{B}(r)\tilde{u} = \Delta\tilde{u} - gT^{-1}\tilde{u}_{-1}$ on $\tilde{B}(r)\tilde{X}$ with $\tilde{V}_0^{-}$ as the inverse of the variance-covariance matrix. For the regressors in the mean and time trend case, we find $\tilde{B}(r)\tilde{e} = gT^{-1}(0,1,\cdots,1)'$ and $\tilde{B}(r)\tilde{\tau} = (0,1,1+gT^{-1},1+2gT^{-1},\cdots,1+(T-1)gT^{-1})'$. Let $\tilde{H}^\mu = (0,1,\cdots,1)'$ and $\tilde{H}^\tau = (\tilde{H}^\mu,\tilde{\tau})$, so that the last T rows of $\tilde{H}^i$ become $H^\mu = e$ and $H^\tau = (e,\tau)$ and note that the same column space is spanned by the regressors $\tilde{B}(r)\tilde{X}^i$ than by $\tilde{H}^i$. We hence find

$$\begin{aligned}
&\lim_{\lambda^{-1}\to 0} \tilde{y}'[\tilde{\Omega}(r)^{-1} - \tilde{\Omega}(r)^{-1}\tilde{X}^i(\tilde{X}^{i\prime}\tilde{\Omega}(r)^{-1}\tilde{X}^i)^{-1}\tilde{X}^{i\prime}\tilde{\Omega}(r)^{-1}]\tilde{y}\\
&= (\Delta\tilde{u} - gT^{-1}\tilde{u}_{-1})'\tilde{V}_0^{-}(\Delta\tilde{u} - gT^{-1}\tilde{u}_{-1})\\
&\qquad -(\Delta\tilde{u} - gT^{-1}\tilde{u}_{-1})'\tilde{V}_0^{-}\tilde{H}^i(\tilde{H}^{i\prime}\tilde{V}_0^{-}\tilde{H}^i)^{-1}\tilde{H}^{i\prime}\tilde{V}_0^{-}(\Delta\tilde{u} - gT^{-1}\tilde{u}_{-1})\\
&= (\Delta u - gT^{-1}u_{-1})'V^{-1}(\Delta u - gT^{-1}u_{-1})\\
(6.3)\quad &\qquad -(\Delta u - gT^{-1}u_{-1})'V^{-1}H^i(H^{i\prime}V^{-1}H^i)^{-1}H^{i\prime}V^{-1}(\Delta u - gT^{-1}u_{-1})
\end{aligned}$$

It is now an easy matter to minimize this expression with respect to g. Note that the coefficient of g^2 is positive almost surely, so that

$$\hat{g}^{i,DF} = \frac{T^{-1}\left(u_{-1}'V^{-1}\Delta u - u_{-1}'V^{-1}H^i(H^{i\prime}V^{-1}H^i)^{-1}H^{i\prime}V^{-1}\Delta u\right)}{T^{-2}\left(u_{-1}'V^{-1}u_{-1} - u_{-1}'V^{-1}H^i(H^{i\prime}V^{-1}H^i)^{-1}H^{i\prime}V^{-1}u_{-1}\right)}$$

minimizes (6.3) and exists with probability 1. From part (i) and (iii) of Lemma 7 and the reasoning in the proof of Theorem 8, we find that the asymptotic distribution of $\hat{g}^{i,DF}$ corresponds to the asymptotic distribution of the statistics in the $\hat{\rho}^{i,DF}$-class. Clearly then, as $T\to\infty$, the sign of $\hat{g}^{i,DF}$ corresponds to the sign of the Dickey-Fuller t-test type statistics.

Furthermore, substituting $\hat{g}^{i,DF}$ for g in (6.3) yields

$$\begin{aligned}
&\Delta u'V^{-1}\Delta u - \Delta u'V^{-1}H^i(H^{i\prime}V^{-1}H^i)^{-1}H^{i\prime}V^{-1}\Delta u\\
&\qquad - \frac{T^{-2}\left(u_{-1}'V^{-1}\Delta u - u_{-1}'V^{-1}H^i(H^{i\prime}V^{-1}H^i)^{-1}H^{i\prime}V^{-1}\Delta u\right)^2}{T^{-2}\left(u_{-1}'V^{-1}u_{-1} - u_{-1}'V^{-1}H^i(H^{i\prime}V^{-1}H^i)^{-1}H^{i\prime}V^{-1}u_{-1}\right)}.
\end{aligned}$$

A direct calculation shows that the first part (excluding the fraction) of this expression is equal to $\tilde{u}'G_0^i\tilde{u} + \frac{(\tilde{u}'G_0^i\tilde{Z}_\varsigma^i)^2}{\tilde{Z}_\varsigma^{i\prime}G_0^i\tilde{Z}_\varsigma^i}$ in the notation of the proof of Theorem 5 above. Therefore, applying part (i) and (iii) of Lemma 7 and the reasoning in the proof of Theorem 8 one finds

$$\breve{Q}_a^i(\hat{g}^{i,DF},\infty) \Rightarrow -\frac{\left[M^{i,\mathrm{OLS}}(1)^2 - M^{i,\mathrm{OLS}}(0)^2 - 1\right]^2}{4\int M^{i,\mathrm{OLS}}(s)^2 ds},$$

which is negative almost surely. $\text{sign}(\hat{g}^{i,DF})\breve{Q}_a^i(\hat{g}^{i,DF},\infty)$ has hence an asymptotic distribution that is a monotone transformation of the asymptotic distribution of the $\hat{\tau}^{i,DF}$ class, which concludes the proof.

Method to find a comparable $Q_a(g,k)$ for a given class of tests:
Denote with r_n the decision of a given class of tests of size 5% to reject ($r_n = 1$) or not to reject ($r_n = 0$) the n-th draw $W_n^i(s)$ of a random sample $n = 1, \cdots, N$ of a Wiener process. Consider a nonlinear logit regression of r_n on a constant and a scaled weighted sum of $W_n^i(1)^2$, $\left(\int W_n^i(s)ds\right)^2$, $W_n^i(1)\int W_n^i(s)ds$ and $\int W_n^i(s)^2ds$, where the weights depend on g and k and are given by the weights $q_j^i(g,k)$ of $Q_a^i(g,k)$ (cf. Theorem 3)

$$r_n = L(l_0 + l_1 Q_{a,n}^i(g,k)) + e_n$$

where $L(x)$ is the logistic function $L(x) = 1/(1+e^{-x})$. Then the estimated values of g and k in this regression may serve as approximations to the values of g_0 and k_0 which maximize the asymptotic probability that the two tests both reject or do not reject under the null hypothesis of $\rho = 1$. The values of Table 4 were calculated with $N = 80'000$.

Proof of Theorem 7:
Since all considered tests are invariant to the respective deterministics, we can set $\tilde{y} = \tilde{u}$ without loss of generality. By recursive substitution $u_t = \sum_{j=1}^t \rho^{t-j}\nu_j + \alpha(2\gamma)^{-1/2}\omega T^{1/2}(1-\rho^t)$. Under Condition 2 and fixed $\gamma = T(1-\rho) \geq 0$ we find from the Continuous Mapping Theorem that $T^{-1/2}\tilde{u}_{[Ts]} \Rightarrow \omega M(s)$. Recall that $M^\mu(s) = M(s)$ and $M^\tau(s) = M(s) - 3s\int \lambda M(\lambda)d\lambda$. The result now follows from the consistency of $\hat{\omega}$, the Continuous Mapping Theorem and Theorem 3.

References

Andrews, D. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

Andrews, D., and W. Ploberger (1994): "Optimal Tests When a Nuisance Parameter Is Present Only under the Alternative," *Econometrica*, 62, 1383–1414.

Bhargava, A. (1986): "On the Theory of Testing for Unit Roots in Observed Time Series," *Review of Economic Studies*, 53, 369–384.

Chan, N., and C. Wei (1987): "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *The Annals of Statistics*, 15, 1050–1063.

Dickey, D., and W. Fuller (1979): "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427–431.

——— (1981): "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica*, 49, 1057–1072.

Dufour, J.-M., and M. King (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., T. Rothenberg, and J. Stock (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, 813–836.

Elliott, G., and J. Stock (2001): "Confidence Intervals for Autoregressive Coefficients Near One," *Journal of Econometrics*, 103, 155–181.

Evans, G., and N. Savin (1981): "Testing for Unit Roots: 1," *Econometrica*, 49, 753–779.

——— (1984): "Testing for Unit Roots: 2," *Econometrica*, 52, 1241–1269.

HAAN, W. D., AND A. LEVIN (2000): "Robust Covariance Matrix Estimation with Data Dependent VAR Prewhitening Order," *UCSD WP2000-11.*

LEHMANN, E. (1986): *Testing Statistical Hypotheses.* Wiley, New York, second edn.

LEYBOURNE, S. (1995): "Testing for Unit Roots Using Forward and Reverse Dickey-Fuller Regressions," *Oxford Bulletin of Economics and Statistics*, 57, 559–571.

MANDELBROT, B., AND J. V. NESS (1968): "Fractional Brownian Motions, Fractional Noise and Applications," *SIAM Review*, 10, 422–437.

MÜLLER, U. K. (2002): "Tests for Unit Roots and the Initial Observation," Ph.D. thesis, University of St. Gallen.

NABEYA, S., AND K. TANAKA (1990): "Limiting Power of Unit-Root Tests in Time-Series Regression," *Journal of Econometrics*, 46, 247–271.

NEWEY, W., AND K. WEST (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

NG, S., AND P. PERRON (2001): "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power," *Econometrica*, 69, 1519–1554.

PANTULA, S., G. GONZALEZ-FARIAS, AND W. FULLER (1994): "A Comparison of Unit-Root Test Criteria," *Journal of Business & Economic Statistics*, 12, 449–459.

PARK, J. (1990): "Testing for Unit Roots and Cointegration by Variable Addition," in *Advances in Econometrics: Co-Integration, Spurious Regressions and Unit Roots*, ed. by T. Fomby, and G. Rhodes. JAI Press, Greenwich, CT.

PHILLIPS, P. (1987a): "Time Series Regression with a Unit Root," *Econometrica*, 55, 277–301.

——— (1987b): "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535–547.

PHILLIPS, P., AND P. PERRON (1988): "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75, 335–346.

PHILLIPS, P., AND Z. XIAO (1998): "A Primer on Unit Root Testing," *Journal of Economic Surveys*, 12, 423–469.

RAO, C., AND S. MITRA (1971): *Generalized Inverse of Matrices and its Applications.* Wiley, New York.

ROTHENBERG, T., AND J. STOCK (1997): "Inference in a Nearly Integrated Autoregressive Model with Nonnormal Innovations," *Journal of Econometrics*, 80, 269–286.

SARGAN, J., AND A. BHARGAVA (1983): "Testing Residuals from Least Squares Regression for Being Generated by the Gaussian Random Walk," *Econometrica*, 51, 153–174.

SCHMIDT, P., AND J. LEE (1991): "A Modification of the Schmidt-Phillips Unit Root Test," *Economics Letters*, 36, 285–289.

SCHMIDT, P., AND P. PHILLIPS (1992): "LM Tests for a Unit Root in the Presence of Deterministic Trends," *Oxford Bulletin of Economics and Statistics*, 54, 257–287.

STOCK, J. (1994): "Unit Roots, Structural Breaks and Trends," in *Handbook of Econometrics*, ed. by R. Engle, and D. McFadden, vol. 4, pp. 2740–2841. North Holland, New York.

——— (2000): "A Class of Tests for Integration and Cointegration," in *Cointegration, Causality, and Forecasting — A Festschrift in Honour of Clive W.J. Granger*, ed. by R. Engle, and H. White, pp. 135–167. Oxford University Press.

TANAKA, K. (1996): *Time Series Analysis — Nonstationary and Noninvertible Distribution Theory.* Wiley, New York.