A Simple Semiparametrically Efficient Rank-Based Unit Root Test

Marc Hallin [*], Ramon van den Akker [†], and Bas J.M. Werker [‡]

Université Libre de Bruxelles, Tilburg University, and Tilburg University

PRELIMINARY VERSION

Abstract

We propose a simple rank-based test for the unit root hypothesis. Our test is semiparametrically efficient in the ubiquitous case that the model contains a non-zero drift. Validity of the test, in terms of exact finite sample size, is guaranteed irrespective of a constant drift. Compared to the appropriate Dickey-Fuller test, our test's efficiency attains up to 45% in case the actual underlying innovation distribution is double exponential.

AMS 1980 subject classification : 62G10 and 62G20.
JEL codes : C12, C22.
Key words and phrases : Dickey-Fuller test, Local Asymptotic Normality.

1 Introduction

There exists an abundant econometric and statistical literature dealing with near unit root asymptotics in time series models. Not only does a unit root generally lead to non-standard rates of convergence for statistical inference, but (policy) implications of economic models often depend crucially on whether the model contains a unit root or, alternatively, is strictly stationary. Analysis of least-squares estimators in zero-mean nonstationary autoregressive processes started with White (1958), but gained more attention after publication of Dickey and Fuller (1979). The unit root testing problem was first studied in detail in Dickey and Fuller (1981).

We restrict in this paper attention to the simplest possible setting of a univariate unit root model with i.i.d. innovations. Extensions to multivariate settings and heteroskedastic innovations fall within the general ideas of the present

[*] Member of the Royal Academy of Belgium.

[†] Econometrics Group, Tilburg University, P.O. Box 90153, 5000 LE, Tilburg, The Netherlands.

[‡] Econometrics and Finance Group, Tilburg University, P.O. Box 90153, 5000 LE, Tilburg, The Netherlands. E-mail: Werker@TilburgUniversity.nl

paper but their technical implications are not pursued here. As an example of these extensions we mention Phillips (1987), Chan and Wei (1988) and Phillips and Perron (1988).

As in Jansson (2008) we are interested in optimal inference in a univariate first-order autoregressive model with a (near) unit root. Formally, we observe $(Y_1, \ldots, Y_n)$ generated from

$$Y_t = \rho Y_{t-1} + \mu + \varepsilon_t, \tag{1}$$

where $Y_0 = 0$ and (ε_t) is a sequence of i.i.d. zero-mean innovations from a distribution with distribution function F admitting a density f. We stress that we do not assume that f admits a finite variance. We do assume that f is absolutely continuous with derivative f' and finite Fisher information for location

$$I_f := \int (f'/f)^2 \, \mathrm{d}F < \infty. \tag{2}$$

Our interest lies in testing the unit root hypothesis, formally we want to test $H_0 : \rho = 1$ against $H_1 : \rho < 1$.

We will be interested in *optimal* inference concerning the unit root hypothesis. Optimality of unit root tests has been studied in Elliot, Rothenberg, and Stock (1996) and Jansson (2008). However, both papers primarily deal with zero mean unit root process, i.e., $\mu = 0$ in the notation of (1). We will see below that the situation is quite different in case the process has a non-zero mean. For cointegrated systems, efficient inference has been studied in Phillips (1991), compare also Jansson and Moreira (2006). We have in common with these papers that we assume the underlying errors to be i.i.d. This is required in order to define optimality of tests in a meaningful way. However, although in none of the papers mentioned above or this paper, i.i.d.-ness of the innovations is very essential as (parametric) forms of heteroskedasticity can be dealt with using the same techniques. Normality of the innovations is not required and as such our results complement those in Rothenberg and Stock (1997).

While, e.g, Elliot, Rothenberg, and Stock (1996), Rothenberg and Stock (1997) and Jansson (2008) focuss on zero-mean AR(1) models, our model of interest (1) contains a possibly non-zero constant μ. In many situations where the unit root hypothesis is of economic interest, inclusion of a constant μ is habitual, for instance in interest rate, inflation, or GDP modeling. As Perron (1988) puts it: "Model (B) [(1)] is likely to be the relevant one for most macroeconomic time series, for which we suspect the presence of a unit root; these series usually have a definite tendency to increase over time." In case μ is known to be zero, the model (1) is Locally Asymptotically Brownian Functional, see Jeganathan (1995). Characterization of optimal inference in these experiments is not fully understood yet, although some results in this direction have been obtained in Gushchin (1996) and Ploberger (2004). Ploberger (2008) considers (non)admissibility of tests for this situation.

In contrast to the above, the model (1) remains Locally Asymptotically Normal (LAN) for $\mu \neq 0$ (be it with a nonstandard convergence rate of $n^{3/2}$).

The fact that unit root models with a constant remain, near the unit root, in the class of LAN models is indicated at various places in the literature, but does not seem to have been stated very explicitly before. In any case, the main contribution of the present paper is to exploit this result to construct a rank-based unit root test, with exact size α for any μ and attaining the semiparametric efficiency bound in case $\mu > 0$ (for notational simplicity we consider only the empirically more relevant case of positive values for μ).

Let us introduce the statistic we propose. Our statistic is based on the ranks R_t of the increments $\Delta Y_t := Y_t - Y_{t-1}$. Let g be a given (so-called *reference*) density, not necessarily equal to the true underlying density f. We assume throughout that g has finite Fisher information for location $I_g < \infty$. As usual, G denotes the corresponding distribution function. Our statistic is now defined as

$$T_g^{(n)} := \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\frac{t}{n+1} - \frac{1}{2} \right) \varphi_g \left(\frac{R_t}{n+1} \right), \tag{3}$$

with $\varphi_g(u) := -g'\left(G^{-1}(u)\right)/g\left(G^{-1}(u)\right),\ u \in (0,1)$. Before going into the asymptotic analysis below, let us already observe the following. The distribution of the ranks (R_t) is, under $H_0:\ \rho = 1$, invariant with respect to both μ and F. As a result, the test statistic $T_g^{(n)}$ is distribution-free. In particular, this implies that its exact size, for finite n, can be easily simulated. Also, this implies that our test does not rely on any moment conditions for ε_t: it even remains valid if first moments would not exist.

Section 2 below discusses the asymptotic behavior of the test statistic T_g. In particular, the asymptotic null-distribution is derived. This limiting distribution is normal and, by the invariance property of the ranks mentioned above, under the null, does not depend on either the true value of μ nor on that of the underlying density f. The limiting null distribution only depends on the chosen reference density g. In order to claim optimality of our test statistic for any strictly positive value of μ, we derive its local power for testing the unit root hypothesis and compare this to the bound obtained from the LAN property to be derived in Section 2.2. The (asymptotic and local) power of our test does depend on both the reference density g and the underlying density f. We show that a correctly specified reference density $g = f$ leads to a test that achieves the lower bound and thus is parametrically efficient. As a result, while our test is valid irrespective of which reference density is used, it is efficient for correctly specified reference density. This situation is tantamount to quasi or pseudo maximum likelihood estimation. Choosing a (Gaussian) reference density leads to an estimator that is consistent even if the reference density is misspecified. The limiting variance of the estimator, however, depends on both the true and the reference density. Our test has a comparable interpretation, with the important exception that we may use *any* density as reference density, while quasi or pseudo likelihood procedures are generally restricted to using a Gaussian reference density: when using another reference density the estimator does not remain consistent under misspecification of the innovation distribution.

The are several other papers that use rank-based methods in unit root analysis; although in different settings. Campbell and Dufour (1995) and Campbell and Dufour (1997) consider testing orthogonality restrictions using sign and rank-based tests instead of regression based approaches. These methods are based on zero-median or symmetry assumptions and are shown, using extensive simulation, to beat regression-based tests. Hasan and Koenker (1997) extend these results using regression rank-scores in order to deal with the nuisance parameter problem. Their focus of interest is again a zero-mean unit root model. Hasan (2001) extends this work to allow for infinite variances. Neither Hasan and Koenker (1997) or Hasan (2001) provide a formal optimality analysis. Finally, we must mention Breitung and Gouriéroux (1997) that essentially proposes to test for a unit root in the ranks of an observed time series. The underlying hypothesis in that case is that *some* transformation of the process exhibits a unit root.

The remainder of the paper is organized as follows. Section 2 provides a full analysis of the limiting properties of our test: size, (local) power, and asymptotic relative efficiency. Our test, when based on a Gaussian reference density, beats the appropriate Dickey-Fuller test uniformly in the actual underlying density g with asymptotic efficiency gains up to 45%. To indicate finite sample performance of our test, we provide some simulations in Section 3. Section 4 concludes, while proofs are gathered in AppendixA.

2 Asymptotic Theory

It turns out that it is easier to prove results on (asymptotic) size and power of our test when it is slightly redefined. We maintain (3), but instead of using φ_g itself we use

$$\tilde{\varphi}_g(u) := \mathrm{E}_G \left\{ \varphi_g \left(G\left(\varepsilon_t\right)\right) \middle| R_t = \lfloor u(n+1) \rfloor \right\}, \quad u \in (0,1). \tag{4}$$

Clearly, the statistic based on φ is simpler to compute, although the function $\tilde{\varphi}_g$ is easily simulated using distribution freeness of the ranks. When n is large and conditionally on the rank of ε_t being $R_t = i$, $G\left(\varepsilon_t\right)$ is approximately equal to $i/(n+1)$. This intuitively explains that asymptotically both φ_g and $\tilde{\varphi}_g$ are identical as formalized by the following result.

Lemma 2.1 *If the score function φ_g is continuous almost everywhere, nonconstant, and satisfies*

$$\frac{1}{n}\sum_{i=1}^{n} \varphi_g \left(\frac{i}{n+1} \right)^2 \to I_g = \int_0^1 \varphi_g(u)^2 du, \ n \to \infty, \tag{5}$$

we have, as $n \to \infty$,

$$T_g^{(n)} \quad = \quad \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\frac{t}{n+1} - \frac{1}{2} \right) \varphi_g \left(\frac{R_t}{n+1} \right) \tag{6}$$

$$= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\frac{t}{n+1} - \frac{1}{2} \right) \tilde{\varphi}_g \left(\frac{R_t}{n+1} \right) + o_p(1).$$

PROOF: This is a well-known result on the equivalence of the so-called *approximate score* functions based on φ_g and the *exact score* functions based on $\tilde{\varphi}_g$ in (4). It is proved at various places, for instance, in Van der Vaart (2000), Theorem 13.5. □

Remark 2.1 *A consequence of the "distribution freeness" of ranks is that (6) is valid under any (continuous) distribution of the innovations ε_t. Next to this, a consequence of the Local Asymptotic Normality result that is established below in Lemma 2.2 is contiguity of the probability measures at the unit root ($\rho_0 = 1$) and near the unit root ($\rho_n = 1 - O(n^{-3/2})$). Asymptotic negligibility in the sense of $o_p(1)$ is preserved under contiguous probability measures (see, e.g., Van der Vaart (2000), Chapter 6). As a result, in expressions like in Lemma 2.1 we do not have to worry whether these are taken at the unit root or near the unit root. This contiguity result will be used throughout the paper without further notice.*

Condition (5) on φ_g is satisfied for all standard reference densities g. Under this condition, the asymptotic equivalence in (6) implies that all results concerning asymptotic size, power (under contiguous alternatives), and efficiency carry over from one statistic to the other. Note already that in our setup the average of the regression constants $t/(n+1) - 1/2$ in (3) equals zero, for each n. The average of the exact scores (4) also equals zero as $\mathrm{E}\left\{\tilde{\varphi}_g(R_t/(n+1))\right\} = \mathrm{E}_G\left\{\varphi_g\left(G(\varepsilon_t)\right)\right\} = 0$.

2.1 Size

Under the null hypothesis of a unit root, the ranks $(R_t)_{t=1}^n$ are those of i.i.d. random variables $(\varepsilon_t)_{t=1}^n$. As a result, their distribution, and consequently that of our statistic $T_g^{(n)}$, does not depend on the true underlying innovation density f. One could, therefore, easily simulate exact null distributions and construct tests with exact finite sample size. Asymptotically, appropriate critical values are obtained from a normal distribution with variance $I_g/12$.

Theorem 2.1 *Let $(\varepsilon_1, \ldots, \varepsilon_n)$ be i.i.d. from a continuous distribution with distribution function F and denote by R_t the rank of ΔY_t. Let the reference density g have finite Fisher information for location $I_g < \infty$. Then, under $H_0: \ \rho_0 = 1$ and as $n \to \infty$,*

$$T_g^{(n)} \Rightarrow N(0, I_g/12). \tag{7}$$

PROOF: First recall that a finite Fisher information for location implies that $\int_{u=0}^1 \varphi_g(u)\mathrm{d}u = 0$ and $\int_{u=0}^1 \varphi_g(u)^2\mathrm{d}u = I_g$. Moreover, under $H_0: \ \rho_0 = 1$, we have $\Delta Y_t = \varepsilon_t$. Now, using $\tilde{W}^{(n)}_{\varphi_g}$ as defined in (18), Lemma A.1 and (6) with $U_t = F(\varepsilon_t)$, we find the following representation

$$T_g^{(n)} = \int_{u=0}^{1} \left(u - \frac{1}{2}\right) \mathrm{d}\tilde{W}^{(n)}_{\varphi_g}(u) + o_p(1). \tag{8}$$

From the continuous mapping theorem, we obtain consequently that $T_g^{(n)}$ is asymptotically distributed as

$$\int_{u=0}^{1} \left(u - \frac{1}{2}\right) \mathrm{d}\tilde{W}(u) \sim N\left(0, \int_{u=0}^{1} \left(u - \frac{1}{2}\right)^2 I_g\right) = N\left(0, \frac{I_g}{12}\right). \tag{9}$$

□

As we have seen before, the null distribution of our test statistic does not depend on the underlying distribution of the innovations in the model, nor on the (unknown) constant μ. As a result, our test remains valid (in terms of size) for any reference density chosen. In particular, the null distribution of our test statistic does not require moments of the innovations to exist. We will see in Section 3 that, for instance in case of Cauchy distributed innovations, our test statistic behaves appropriately, while the Dickey-Fuller test statistic breaks down.

The power of our test does depend on the reference density. We will see in the remainder of the paper that a choice close to the actual underlying density improves efficiency up to the semiparametric efficiency bound.

2.2 Limit experiment and efficient inference

As mentioned in the introduction, the limiting experiment for (near) unit root behavior in the model (1) crucially depends on the value of μ. In case it is known that $\mu = 0$, the limit experiment is Locally Asymptotically Brownian Functional with rate of convergence $n{-}1$ as shown by Jeganathan (1995). This is exploited in Jansson (2008) to derive power envelopes for unit root tests.

The situation is quite different in case the a priori knowledge $\mu = 0$ is not available. This is the content of the next result, formulated for μ arbitrary, but positive.

Lemma 2.2 *Consider the model (1) with innovation density f having finite Fisher information as defined in (2). This model is Locally Asymptotically Normal at any $\mu_0 > 0$ and $\rho_0 = 1$ with respect to the alternatives $\mu_n = \mu_0 + h_\mu n^{-1/2}$ and $\rho_n = 1 + h_\rho n^{-3/2}$ and central sequences*

$$\begin{aligned} \begin{bmatrix} \Delta_\mu^{(n)} \\ \Delta_\rho^{(n)} \end{bmatrix} &= \begin{bmatrix} n^{-1/2} \sum_{t=1}^{n} \frac{-f'}{f}(\varepsilon_t) \\ \mu n^{-1/2} \sum_{t=1}^{n} \frac{t}{n+1} \frac{-f'}{f}(\varepsilon_t) \end{bmatrix} \\ &\Rightarrow N\left(0, I_f \begin{bmatrix} 1 & \mu/2 \\ \mu/2 & \mu^2/3 \end{bmatrix}\right). \end{aligned} \tag{10}$$

To be more precise, with $P^{(n)}_{\mu,\rho,f}$ denoting the probability measure generated by (1), we have, under $P^{(n)}_{\mu_0,\rho_0,f}$ and for $n \to \infty$,

$$\begin{aligned}\log \frac{dP^{(n)}_{\mu_n,\rho_n,f}}{dP^{(n)}_{\mu_0,\rho_0,f}} &= h_\mu n^{-1/2} \sum_{t=1}^{n} \frac{-f'}{f}(\varepsilon_t) + h_\rho \mu_0 n^{-1/2} \sum_{t=1}^{n} \frac{t}{n+1} \frac{-f'}{f}(\varepsilon_t) \\ &\quad - \frac{I_f}{2}\left(h_\mu^2 + \mu_0 h_\mu h_\rho + \frac{\mu_0^2}{3} h_\rho^2\right) + o_p(1).\end{aligned}$$

PROOF: The proof is analogous to that in Drost, Klaassen, and Werker (1997) for a pure location model. The rates of convergence obviously have to be adapted. The Fisher information matrix follows from the observation that, under the null hypothesis and $\mu_0 > 0$, $n^{-3}\sum_{t=1}^{n} Y_t^2 \to \mu_0^2/3$ (*a.s.*) as the drift $\mu_0 t$ dominates the stochastic part

$$\frac{1}{n^3}\sum_{t=1}^{n}\left(\sum_{s=1}^{t}\varepsilon_s\right)^2 \leq \frac{1}{n}\sum_{t=1}^{n}\left(\frac{1}{t}\sum_{s=1}^{t}\varepsilon_s\right)^2 \to 0 \ (a.s.),$$

by a Cesàro mean argument and the strong law of large numbers. □

Remark 1 *Incidentally, note that the Local Asymptotic Normality result on Lemma 2.2 does not require $h_\rho \leq 0$. As a result, all claims in this paper can easily be rephrased when testing $H_0: \ \rho = 1$ against $H_1: \ \rho > 1$.*

Knowledge of the asymptotic structure of a statistical experiment as in the previous lemma automatically induces the structure of optimal test statistics using the Hájek and Le Cam theory on limits of experiments, see, e.g., Van der Vaart (2000), Section 15.3. Using Lemma 2.2, we find that an optimal test for $H_0: \ \rho = 1$, considering μ a nuisance parameter and (for the moment) for given innovation density f, should be based on the statistic

$$I_f^{-1}\left(\Delta_\rho^{(n)} - \frac{\mu}{2}\Delta_\mu^{(n)}\right) = \frac{\mu}{I_f} n^{-1/2} \sum_{t=1}^{n}\left(\frac{t}{n+1} - \frac{1}{2}\right)\frac{-f'}{f}(\varepsilon_t). \tag{11}$$

Clearly, in order to construct a test statistic, the constant factor μ/I_f can be ignored. As a result, we find an interesting candidate test statistic. However, as this statistic is based on the true underlying density f (and in general the score $-f'/f$ is not centered under alternative distributions so that martingale central limit theorems do not apply), this statistic will not be asymptotically normal in case $q \neq f$. This is why we proposed the rank-based statistic (3) and this is done without sacrificing any efficiency. Section 2.4 discusses the details.

2.3 Local power

The power of our test statistic $T_g^{(n)}$ under near unit root alternatives follows directly from the so-called Le Cam's third lemma.

Theorem 2.2 *Let $(\varepsilon_1, \dots, \varepsilon_n)$ be i.i.d. from a continuous distribution with distribution function F and denote by R_t the rank of ΔY_t. Let the reference density g have finite Fisher information for location $I_g < \infty$. Then, under distribution F, under $H_1^{(n)}: \ \rho_n = 1 + h_\rho n^{-3/2}$, and as $n \to \infty$,*

$$T_g^{(n)} \Rightarrow N(h_\rho \mu_0 I_{fg}/12, I_g/12), \tag{12}$$

where

$$I_{fg} := \int_{u=0}^{1} \varphi_g(u)\varphi_f(u)du = \int_{u=0}^{1} (-f'/f)\left(F^{-1}(u)\right)(-g'/g)\left(G^{-1}(u)\right) du. \tag{13}$$

Proof: Using the notation of Lemma A.1 and Hájek's Representation Theorem (see, e.g., Van der Vaart (2000), Theorem 13.5) we find another representation of our statistic, that is

$$T_g^{(n)} = \int_{u=0}^{1} \left(u - \frac{1}{2}\right) \mathrm{d}W_{\varphi_g}^{(n)}(u) + o_p(1), \tag{14}$$

with $W_{\varphi_g}^{(n)}$ as in (18). From Lemma 2.2 we see as well that $\log \mathrm{dP}_{\mu_0,\rho_n,f}^{(n)}/\mathrm{dP}_{\mu_0,\rho_0,f}^{(n)}$ is asymptotically equal to $h_\rho \mu_0 \int_{u=0}^{1} u\mathrm{d}W_{\varphi_f}^{(n)}(u) - h_\rho^2 \mu_0^2 I_f/6$. As a result, the statistic $T_g^{(n)}$ and the log likelihood ratio are asymptotically jointly normally distributed with limiting covariance $h_\rho \mu_0 \int_{u=0}^{1} u(u-1/2)\mathrm{d}u I_{fg} = h_\rho \mu_0 I_{fg}/12$. Le Cam's third lemma, see, e.g., Van der Vaart (2000), Section 6.7, now readily implies (12). □

Our test has power against alternatives that are at distance $n^{-3/2}$ from the unit root. This is, of course, much more than the usual rate $n^{-1/2}$. We also see that for $\mu_0 = 0$ the statistic has no local power at this rate of convergence. However, in that case, our statistic still does have local power at, for that case optimal, rate n^{-1}.

It is interesting to compare the power of our test statistic to that of the classical Dickey-Fuller test. For our comparison we choose the Dickey-Fuller test based on the least-squares estimate for ρ in the regression (1). The asymptotic properties of this classical Dickey-Fuller statistic are well-known and we have the following corollary.

Corollary 2.1 *The Asymptotic Relative Efficiency of our test based on $T_g^{(n)}$ for the unit root hypothesis $H_0: \ \rho = 0$ with respect to the Dickey-Fuller test based on the least-squares estimate $\hat{\rho}_n^{DF}$ of ρ in the model (1) is given by*

$$ARE(G|F) = \frac{I_{fg}^3 \sigma_f^3}{I_g^{3/2}}, \tag{15}$$

where g denotes the reference density in $T_g^{(n)}$ and f the actual underlying density.

Proof: The asymptotic of the Dickey-Fuller test statistic are well studied. For instance using Hamilton (1994), Chapter 17, we find, with $\bar{Y}_n = n^{-1}\sum_{t=1}^n Y_{t-1}$,

$$
\begin{aligned}
n^{3/2}\left(\hat{\rho}_n^{DF} - 1\right) &= \frac{n^{-3/2}\sum_{t=1}^n \left(Y_{t-1} - \bar{Y}_n\right)\Delta Y_{t-1}}{n^{-3}\sum_{t=1}^n \left(Y_{t-1} - \bar{Y}_n\right)^2} \\
&= n^{-1/2}\frac{12}{\mu_0}\sum_{t=1}^n \left(\frac{t}{n+1} - \frac{1}{2}\right)\varepsilon_t + o_p(1).
\end{aligned}
\tag{16}
$$

The null limiting distribution of $n^{3/2}\left(\hat{\rho}_n^{DF} - 1\right)$ thus equals $N(0, 12\sigma_f^2/\mu_0^2)$. As in Theorem 2.2 its limiting distribution under the near unit root alternatives $H_1^{(n)}:\ \rho_n = 1 + h_\rho n^{-3/2}$ follows from Le Cam's third lemma as $N(h_\rho, 12\sigma_f^2/\mu_0^2)$ using that $E_f(-f'/f)(\varepsilon_t)\varepsilon_t = 1$. Incidentally, this shows that the least-squares estimator is regular in this situation (as well). □

Remark 2 *Our test and the Dickey-Fuller test have local power at rate* $n^{3/2}$. *We define the ARE in (15) as the factor of observations more that is needed by the Dickey-Fuller test in order to match the performance (in terms of power) of our rank-based test. This explains the exponent* 3 *in (15).*

Our test depends on a reference density to be chosen by the researcher. First of all note that our test statistic $T_g^{(n)}$ in (3) is homogeneous in the scale of the reference distribution. Consequently, an investigator does not have to worry about choosing an appropriate scale. Besides the scale, the form of the reference density does influence the local power of our test via the quantity I_{fg} in (13). We will discuss the (optimal) choice of the reference density in more detail in Section 2.4, but some well-known reference densities will be discussed now.

An obvious first choice is a Gaussian reference density $g(x) \propto \exp(-x^2/2)$ leading to the so-called Van der Waerden scores. In this case, $I_g = 1$ and the ARE in (15) reduces to

$$
ARE(\Phi|F) = \left(\int_{u=0}^1 \frac{-f'}{f}\left(F^{-1}(u)\right)\Phi^{-1}(u)\right)^3 \sigma_f^3, \tag{17}
$$

where Φ denotes the standard normal distribution function. A celebrated result in Chernoff and Savage (1958) shows that $ARE(\Phi|F)$ is always larger than one, except under Gaussian densities, where it takes value one. Consequently, a Gaussian reference density constitutes a safe choice as it always leads to an improvement over the Dickey-Fuller test. The magnitude of the improvement is all the more sizeable in our situation due to the faster rate of convergence $n^{3/2}$; see the first row in Table 1. For instance, true underlying double exponentially distributed innovations lead to 45% efficiency gain.

Table 1 shows that an incorrect choice of the reference density can also lead to a performance which is worse than Dickey-Fuller's. However in contrast to standard likelihood inference, choosing an incorrect innovation distribution does

	Actual density		
Reference density	Gaussian	Logistic	Double Exponential
Gaussian	1.00	1.07	1.45
Logistic	0.93	1.15	1.84
Double exponential	0.51	0.75	2.83

Table 1: The Asymptotic Relative Efficiency (ARE) of our test statistic in (3) with respect to the Dickey-Fuller test for various choices of reference density and given actual density.

not affect validity of the test (its size remains correct), but only its power. See Section 2.4 for more discussion on this.

It is useful to note that we nowhere imposed that the innovations need finite variances. Our test remains valid and its local power follows from Theorem 2.2. The relative efficiency with respect to the Dickey-Fuller test can, of course, not meaningfully be defined if the innovation's variance is infinite as in that case the Dickey-Fuller test has no guaranteed (asymptotic) size. Formula (15) would lead to an infinite ARE in this case.

2.4 Efficiency and Adaptivity

Although we have seen in the previous section that a Gaussian reference density is always a safe choice leading to guaranteed efficiency improvements with respect to the classical Dickey-Fuller test. However, in some situations information may be available about the form of the underlying distribution of the innovations.

Clearly, given the power as derived in Theorem 2.2 and using Cauchy-Schwarz, it's easy to see that maximum power is achieved when the reference density equals the actual density, up to a possible scale transformation. In that case, our statistic asymptotically coincides with the optimal central sequence as derived in (11). Consequently, our test attains the parametric efficiency bound at $g = f$. As our test is valid irrespective of f, and as such is a test in a semiparametric model, the test obviously is also semiparametrically efficient and the inference problem is actual adaptive: not knowing the innovation density in additional to not knowing its mean does not complicate the inference about ρ further.

Let us stress once more that, in contrast to classical pseudo-likelihood procedures that generally loose consistency when choosing a non-Gaussian reference density, our test remains valid irrespective of the reference density chosen. Optimality of the choice of reference density pertains to power only.

A final advantage of our test we mention here is that, in order to choose a reference density, an investigator may estimate the density of the ΔY_t's and use the resulting estimated scores. This will not have any impact on the distribution-freeness of the resulting test statistics or the size of the resulting test. In particular, if (conditionally) exact α-critical points are computed for the estimated-score version of (3), conditional size, hence also the unconditional one, is exactly α too. The reason for this is simple. A density estimate is a function of the order statistics of ΔY_t only, while our test statistic is a function of the ranks of ΔY_t only. As ranks and order statistics are independent, all results about our test statistic remain valid conditional on the order statistic. A simple way to implement this would be to use a kernel density estimator to find a suitable reference density.

3 Simulations

Section to be completed.

4 Conclusions

We provide optimal rank-based tests of the unit root hypothesis. Our tests offer the standard advantages of rank-based tests: "distribution freeness", exact finite sample size, and robustness. Moreover, our tests are flexible in the sense that a reference density can be chosen. We stress that our tests have correct size irrespective of the reference density chosen. A reference density can either be chosen, or even estimated, without affecting the null distribution of the test statistic. Moreover, choosing a Gaussian reference density guarantees that our test is more powerful that the optimal Dickey-Fuller test in our model. Efficiency gains may run up to 45%. Choosing a reference density close to the true underlying innovation density improves the power of our test and, at the underlying innovation density, our test even attains the (semi)parametric efficiency bound.

The present paper focusses on the simplest setting possible. In particular, we assume the underlying innovations of the process to be i.i.d. This is needed in order to define optimality of testing procedures. However, extensions to models that allow for, e.g., parametric forms of heteroskedasticity are easily imagined.

A Proofs

For ease of reference, we first provide a lemma on the joint convergence of a partial sum process and its rank-based version. This lemma formalizes the consequences of the fact that, for uniformly distributed i.i.d. random variables U_t, we have $R_t/(n+1) \approx U_t$. Although based on existing results in the literature, this lemma as such does not seem to have been provided. The bottom line is that, where the partial sum process converges to a Brownian motion, its rank-based version converges to the Brownian bridge generated by that Brownian motion.

Lemma A.1 *Let $(U_1, \ldots, U_n)$ be i.i.d. uniformly distributed random variables and denote by R_t the rank of U_t. Let $\varphi : [0, 1] \to \mathbb{R}$ be a measurable function* *satisfying* $\int_0^1 \varphi(v)dv = 0$ *and* $\int_0^1 \varphi(v)^2 dv < \infty$. *Define the partial sum processes* $W_\varphi^{(n)}$ *and* $\tilde{W}_\varphi^{(n)}$*, both on* $[0,1]$*, by*

$$W_\varphi^{(n)}(u) = \frac{1}{\sqrt{n}} \sum_{t=1}^{un} \varphi(U_t) \text{ and } \tilde{W}_\varphi^{(n)}(u) = \frac{1}{\sqrt{n}} \sum_{t=1}^{un} E\{\varphi(U_t)|\, R_t\}. \tag{18}$$

Then, we have

$$\begin{bmatrix} W_\varphi^{(n)} \\ \tilde{W}_\varphi^{(n)} \end{bmatrix} \Rightarrow \begin{bmatrix} W \\ \tilde{W} \end{bmatrix}, \tag{19}$$

where W *denotes a zero-drift Brownian motion with variance* $\int_0^1 \varphi(v)^2 dv$ *per unit of time and* $\tilde{W}$ *its associated Brownian bridge:* $\tilde{W}(u) = W(u) - uW(1)$, $u \in [0,1]$. *The convergence in (19) is on* $D^2[0,1]$ *equipped with the uniform topology.*

Proof: It is well-known that weak convergence in $D^2[0,1]$ under the uniform topology follows from establishing convergence of marginals and asymptotic tightness, see, for example, Van der Vaart and Wellner (1993), Theorem 1.5.4.

Convergence of marginals for the partial sum process $W_\varphi^{(n)}$ is easily obtained from the central limit theorem. This implies also (joint) convergence of the marginals of its rank-based version $\tilde{W}_\varphi^{(n)}$ using what is sometimes known as Hájek's representation theorem:

$$\tilde{W}_\varphi^{(n)}(u) = W_\varphi^{(n)}(u) - uW_\varphi^{(n)}(1) + o_p(1), \tag{20}$$

see Van der Vaart (2000), Theorem 13.5. In the notation of Van der Vaart (2000), we have $i = t$, $N = n$, $C_{Ni} = I\{t \le un\}$, and $a_{Ni} = \mathrm{E}\{\varphi(U_t)|\, R_t = i\}$. From $\int_0^1 \varphi(v)\mathrm{d}v = 0$ we find $\bar{a}_N = 0$. Moreover, we have $\bar{c}_N = \lfloor un \rfloor / n \to u$.

Asymptotic tightness in $D^2[0,1]$ under the uniform topology is still to be established. □

References

Breitung, J., and C. Gouriéroux (1997). Rank tests for unit roots. *Journal of Econometrics* **81** 7–27.

Campbell, B., and J.M. Dufour (1995). Exact Nonparametric Orthogonality and Random Walk Tests. *The Review of Economics and Statistics* **77** 1–16.

Campbell, B., and J.M. Dufour (1997). Exact Nonparametric Tests of Orthogonality and Random Walk in the Presence of a Drift Parameter. *International Economic Review* **38** 151–173.

Chan, N.H., and C.Z. Wei (1988). Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes. *Annals of Statistics* **16** 367–401.

Chernoff, H., and I.R. Savage (1958). Asymptotic Normality and Efficiency of Certain Nonparametric Test Statistics. *The Annals of Mathematical Statistics* **29** 972–994.

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

Dickey, D.A., and W.A. Fuller (1981). Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica* **49** 1057–1072.

Drost, F.C., C.A.J. Klaassen, and B.J.M. Werker (1997). Adaptive Estimation in Time-Series Models. *The Annals of Statistics* **25** 786–817.

Elliott, G., T.J. Rothenberg, and J.H. Stock (1996). Efficient Tests for an Autoregressive Unit Root. *Econometrica* **64** 813–836.

Gushchin, A.A. (1996). Asymptotic Optimality of Parameter Estimators under the LAQ Condition. *Theory of Probability and its Applications* **40** 261–272.

Hamilton, J.D. (1994). *Time Series Analysis* (1 ed.). Princeton University Press.

Hasan, M.N. (2001). Rank tests of unit root hypothesis with infinite variance errors. *Journal of Econometrics* **104** 49–65.

Hasan, M.N., and R.W. Koenker (1997). Robust Rank Tests of the Unit Root Hypothesis. *Econometrica* **65** 133–161.

Jansson, M. (2008). Semiparametric Power Envelopes for Tests of the Unit Root Hypothesis. *Econometrica* **76** 1103–1142.

Jansson, M., and M.J. Moreira (2008). Optimal Inference in Regression Models with Nearly Integrated Regressors. *Econometrica* **74** 681–714.

Jeganathan, P. (1995). Some Aspects of Asymptotic Theory with Applications to Time Series Models. *Econometric Theory* **11** 818-887.

Perron, P. (1988). Trends and Random Walks in Macroeconomic Time Series. *Journal of Economics Dynamics and Control* **12** 297–332.

Phillips, P.C.B. (1987). Time Series Regression with a Unit Root. *Econometrica* **55** 277–301.

Phillips, P.C.B. (1991). Optimal Inference in Cointegrated Systems. *Econometrica* **59** 283–306.

Phillips, P.C.B., and P. Perron (1988). Testing for a Unit Root in Time Series Regression. *Biometrika* **75** 335–346.

Ploberger, W. (2004). complete class of tests when the likelihood is locally asymptotically quadratic. *Journal of Econometrics* **118** 67–94.

Ploberger, W. (2008). Admissible and nonadmissible tests in unit-root-like situations. *Econometric Theory* **24** 15–42.

Rothenberg, T.J., and J.H. Stock (1997). Inference in a Nearly Integrated Autoregressive Model with Nonnormal Innovations. *Journal of Econometrics* **80** 268–286.

Van der Vaart, A. (2000). *Asymptotic Statistics* (1 ed.). Cambridge University Press.

Van der Vaart, A. and J. Wellner (1993). *Weak convergence and empirical processes* (2nd ed.). Springer-Verlag.

White, J.S. (1958). The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case. *The Annals of Mathematical Statistics* **29** 1188–1197.