SEMIPARAMETRICALLY OPTIMAL HYBRID RANK TESTS FOR UNIT ROOTS

Bo Zhou[1], Ramon van den Akker[1] and Bas J.M. Werker[1,2]

Abstract

Preliminary and incomplete draft · Please do not cite or circulate.

We propose a new class of unit root tests that exploits invariance properties in the Locally Asymptotically Brownian Functional limiting experiment of the unit root model. These invariance structures naturally suggest tests based on the ranks of the increments of the observations, their mean, and an assumed reference density for the innovations. The tests are semiparametric in the sense that the reference density need not equal the true innovation density. For correctly specified reference density, the asymptotic power curve of our test is point-optimal and nearly efficient (in the sense of Elliott, Rothenberg, and Stock (1996)). When using a Gaussian reference density, our test performs as well as commonly used tests under true Gaussian innovations and better under other distributions, e.g., fat-tailed or skewed. Monte Carlo evidence shows that our test also behaves well in small samples.

Keywords: unit root test, semiparametric power envelope, limit experiment, LABF, maximal invariant, rank statistic.

1. INTRODUCTION

The recent monographs Patterson (2011, 2012) provide a summary of the literature on unit roots which traces back to White (1958) and which got an enormous boost after the seminal papers Dickey and Fuller (1979, 1981), Phillips (1987), Phillips and Perron (1988), and Elliott, Rothenberg, and Stock (1996). This paper fits into the stream of literature that focuses on "optimal" testing for unit roots. Important early contributions are those by Dufour and King (1991), Saikkonen and Luukkonen (1993), and Elliott, Rothenberg, and Stock (1996), which derived the asymptotic power envelope for unit root testing in settings where the underlying innovations of the time series model are Gaussian, and Rothenberg and Stock (1997) which considered the non-Gaussian case.

This paper considers semiparametric optimal testing for unit roots. Following earlier literature, we focus on a simple AR(1) model driven by i.i.d. innovations. Apart from some smoothness and existence of relevant moments, no assumptions are imposed on the distribution of the innovations. From earlier work it is already known that the unit root model leads to Locally Asymptotically Brownian Functional (LABF) limit experiments (in the Le Cam sense). As a consequence, no uniformly most powerful test exists (even if the innovation distribution is known) – see also Elliott, Rothenberg, and Stock (1996). In the semiparametric case the limit experiment becomes even more difficult, because then one also has to deal with the infinite-dimensional nuisance parameter. Jansson (2008) managed to derive the semiparametric power envelope by mimicking ideas that hold for Locally Asymptotically Normal (LAN) models. However, no (feasible) test attaining the power envelope was provided. This paper aims to fill this gap. The main contribution of this manuscript is two-fold.

First, we provide a new derivation of the semiparametric asymptotic power envelope for unit root tests. We focus on unit root tests that are (locally) invariant with respect to the distribution of the innovations. Using the Asymptotic Representation Theorem we can obtain the asymptotic power envelope for invariant tests by studying an associated inference problem in the LABF limit experiment. Girsanov's theorem provides a "structural" description of the LABF limiting structure, which corresponds to observing a countable collection of Ornstein-Uhlenbeck processes (on the time interval $[0, 1]$). We exploit the structural description to derive the maximal invariant, i.e. the minimal reduction of the data which is invariant with respect to the nuisance parameters. It turns out that this maximal invariant takes a rather simple form (all but one processes have to be replaced by the associated "bridge processes"). Now the power envelope for the LABF limit experiment easily follows by an application of the Neyman-Pearson lemma. An application of the Asymptotic Representation Theorem now, almost directly, yields the local asymptotic power envelope. We note that our invariance analysis of the LABF experiment is of independent interest and could, for example, also be exploited for the analysis of optimal inference for cointegration models or predictive regression models with nearly unstable predictors.

Second, we provide a new class of easy-to-implement unit root tests that are semiparametrically optimal in the sense that they are tangent to the semiparametric power envelope. The form of the maximal invariant naturally

Department of Econometrics and OR, Tilburg University, P.O. Box 90153, NL-5000LE Tilburg, the Netherlands
B.Zhou@TilburgUniversity.edu; R.vdnAkker@gmail.com; Werker@TilburgUniversity.edu
[1]Econometrics Group, CentER, Tilburg University
[2]Finance Group, CentER, Tilburg University

suggests tests based on the ranks of the increments of the observations, their mean, and an assumed reference density for the innovations. The tests are semiparametric in the sense that the reference density need not equal the true innovation density. For correctly specified reference density and a chosen alternative, the asymptotic power curve of our test is tangent to the semiparametric power envelope. Following Elliott, Rothenberg, and Stock (1996) we also discuss the selection of an alternative that yields a "nearly" optimal tests, i.e. whose asymptotic local power function is very close to the semiparametric asymptotic power envelope. Monte Carlo results show that when using a Gaussian reference density, our test performs as well as commonly used tests under true Gaussian innovations and better under other distributions, e.g., fat-tailed or skewed.

The remainder of this paper is organized as follows. Section 2 introduces the model assumptions and some notation. Next, Section 3 contains the new derivation of the semiparametric power envelope for unit root tests. The new class of hybrid rank tests is introduced in Section 4. Section 5 contains the results of a Monte Carlo study and Section 6 contains a discussion of possible extensions of our results. All proofs are organized in the appendix.

2. THE MODEL

We consider observations $Y_1, \ldots, Y_T$ generated from the component specification

$$Y_t = \mu + Z_t, \quad t \in \mathbb{N}, \tag{2.1}$$

$$Z_t = \rho Z_{t-1} + \varepsilon_t, \quad t \in \mathbb{N}, \tag{2.2}$$

where $Z_0 = 0$ and the innovations $\{\varepsilon_t\}$ form an i.i.d. sequence with density f. We impose the following assumptions on the innovation density.

Assumption 1

(a) The density f is absolutely continuous, $f \in \mathcal{F}_{ac}$, with a.e. derivative f', i.e. for all $a < b$ we have

$$f(b) - f(a) = \int_a^b f'(e)\mathrm{d}e.$$

(b) The Fisher-information for location,

$$J_f = \int \phi_f^2(e) f(e) \mathrm{d}x, \tag{2.3}$$

where $\phi_f(e) = -(f'/f)(e)1\{f(e) > 0\}$ is the *location score*, is finite.

(c) We have $\mathbb{E}_f \varepsilon_t = \int e f(e)\mathrm{d}e = 0$ and $\sigma_f^2 = \mathrm{var}_f(\varepsilon_t) < \infty$. □

Let $\mathcal{F} \subset \mathcal{F}_{ac}$ denote the set of densities satisfying this assumption.

The imposed smoothness assumptions (a) and (b) on f are mild and standard. The finite variance assumption is important to our asymptotic results as it is essential to the weak convergence, to a Brownian motion, of the partial-sum process generated by the innovations.[1] The assumption on the initial condition, Z_0, is less innocent then it may appear. Indeed, it is well-known, see Müller and Elliott (2003) and Elliott and Müller (2006), that, even asymptotically, the initial condition can contain non-negligible statistical information.

The main goal of this paper is to develop a semiparametrically optimal test for the unit root hypothesis

$$\mathrm{H}_0: \rho = 1, \, (\mu \in \mathbb{R}, \, f \in \mathcal{F}) \text{ versus } \mathrm{H}_a: \rho < 1, \, (\mu \in \mathbb{R}, \, f \in \mathcal{F}),$$

i.e. apart from Assumption 1 no further structure is imposed on f and the intercept μ is also treated as a nuisance parameter. It is well-known, and goes back to Phillips (1987) and Phillips and Perron (1988), that the contiguity rate for the unit root testing problem, i.e. the fastest convergence rate at which it is possible to distinguish (with non-trivial power) the unit root $\rho = 1$ from a stationary alternative $\rho < 1$, is given by T^{-1}. Therefore, in order to compare performances of tests with this proper rate of convergence, we reparametrize the autoregression parameter ρ into its local-to-unity form, i.e.

$$\rho = \rho_h^{(T)} = 1 + \frac{h}{T}, \tag{2.4}$$

and we can rewrite our hypothesis of interest as

$$\mathrm{H}_0: h = 0, \; (\mu \in \mathbb{R}, \, f \in \mathcal{F}) \text{ versus } \mathrm{H}_a: h < 0, \; (\mu \in \mathbb{R}, \, f \in \mathcal{F}).$$

[1] Let us already mention that, although not allowed for in our theoretical results, we will also assess the finite-sample performances of the proposed tests for innovation distributions with infinite variance. For tests specifically developed for such cases we refer to Hasan (2001), Ahn, Fotopoulos, and He (2003), and Callegari, Cappuccio, and Lubian (2003).

3. THE ASYMPTOTIC POWER ENVELOPE FOR INVARIANT TESTS

In this section we derive the asymptotic power envelope for invariant tests. We will first discuss some notation and preliminaries in Section 3.1. Next, we will derive the LABF limit experiment corresponding to the unit root model and provide a "structural" description of this limit experiment (Section 3.2). In Section 3.3 we discuss, exploiting the structural representation of the limit experiment, that there is a natural invariance restriction with respect to the infinite-dimensional nuisance parameter associated to the innovation density. We derive the maximal invariant for the limit experiment and obtain from this the power envelope for invariant tests in the limit experiment. Using the Asymptotic Representation Theorem we translate these results back to the unit root model of interest.

3.1. *Preliminaries*

This section discusses a convenient parametrization of perturbations to the innovation density which we will need to deal with the semiparametric nature of the testing problem. Moreover, we introduce some partial-sum processes and their Brownian limits which we will use later on in the analysis of the LABF limit experiment.

Perturbations to the innovation density

To describe the perturbations to the density f, we need the separable Hilbert space

$$\mathrm{L}_2^{0,f} = \mathrm{L}_2^{0,f}(\mathbb{R},\mathcal{B},f) = \left\{ b \in \mathrm{L}_2(\mathbb{R},\mathcal{B},f) \;\middle|\; \int b(e)f(e)\mathrm{d}e = 0,\ \int b(e)ef(e)\mathrm{d}e = 0 \right\},$$

where $\mathrm{L}_2(\mathbb{R},\mathcal{B},f)$ denotes, as usual, the space of Borel-measurable functions $b:\ \mathbb{R}\to\mathbb{R}$ satisfying $\int b^2(e)f(e)\mathrm{d}e < \infty$. There exists a countable basis b_k, $k\in\mathbb{N}$, of $\mathrm{L}_2^{0,f}$ such that $b_k \in C_{2,b}(\mathbb{R})$, for all k, i.e. each b_k is bounded and two times continuously differentiable with bounded derivatives. Hence each function $b\in\mathrm{L}_2^{0,f}$ can be written as $b=\sum_{k=1}^\infty \eta_k b_k$, for some $(\eta_k)_{k\in\mathbb{N}} \in \ell_2 = \{(x_k)_{k\in\mathbb{N}} \mid \sum_{k=1}^\infty x_k^2 < \infty\}$. Besides the sequence space ℓ_2 we also need the sequence space c_{00} which is defined as the set of sequences with finite support, i.e.

$$c_{00} = \left\{ (x_k)_{k\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \;\middle|\; \sum_{k=1}^\infty 1\{x_k \neq 0\} < \infty \right\}.$$

Of course, c_{00} is a dense subspace of ℓ_2. For $\eta\in c_{00}$ we now introduce the following perturbation to the density f:

$$f_\eta^{(T)}(e) = f(e)\left(1 + \frac{1}{\sqrt{T}}\sum_{k=1}^\infty \eta_k b_k(e)\right), \quad e\in\mathbb{R}. \tag{3.1}$$

The following proposition shows that these perturbations are valid in the sense that they belong to the model for the innovation density.

Proposition 3.1 Let Assumption 1 hold and $\eta \in c_{00}$. Then there exists $T'\in\mathbb{N}$ such that for all $T\geq T'$ we have $f_\eta^{(T)} \in \mathcal{F}$. □

It is clear, since η has finite support, that we have $f_\eta^{(T)} \geq 0$ for large enough T. The mean restrictions $\int b_k(e)f(e)\mathrm{d}e = 0$, together with the finite support of η, guarantee that $f_\eta^{(T)}$ integrates to 1. Similarly, $\int b_k(e)ef(e)\mathrm{d}e = 0$ implies $\mathbb{E}_{f_\eta^{(T)}}[\varepsilon_t] = 0$. Of course, absolute continuity of $f_\eta^{(T)}$ follows from $f\in\mathcal{F}_{ac}$ and, again because η has finite support, $\sum_{k=1}^\infty \eta_k b \in C_{2,b}(\mathbb{R})$. These properties also easily yield $\mathrm{var}_{f_\eta^{(T)}}[\varepsilon_t] < \infty$. Only $J_{f_\eta^{(T)}} < \infty$ requires a bit of straightforward calculus. For the sake of completeness, this calculation is organized in Appendix A.

Remark 3.1 Typically (see, for example, Bickel, Klaassen, Ritov, and Wellner (1993)) one parametrizes perturbations to a density by a so-called "non-parametric" score $h\in\mathrm{L}_2^{0,f}$, i.e. a perturbation takes the form $f(e)(1+T^{-1/2}h(e))$. By using the basis b_k, $k\in\mathbb{N}$, we instead tackle all such perturbations via the infinite-dimensional nuisance parameter η. Of course, one would need to use ℓ_2 as parameter space to "generate" all score functions h. We can, however, restrict to c_{00} without cost. Intuively, this is clear since c_{00} is a dense subspace of ℓ_2 (so if a property is "sufficiently continuous" one only needs to establish it on c_{00} because it automatically extends to the closure). □

Partial sum processes

To describe the limit experiment in Section 3.2 we need to introduce some partial sum processes and their limits. As usual, Δ denotes differencing, i.e. $\Delta X_t = X_t - X_{t-1}$. Define, for $u \in [0,1]$,

$$W_{\varepsilon}^{(T)}(u) = \frac{1}{\sqrt{T}} \sum_{t=2}^{[uT]} \Delta Y_t,$$

$$W_{\phi_f}^{(T)}(u) = \frac{1}{\sqrt{T}} \sum_{t=2}^{[uT]} \phi_f(\Delta Y_t), \quad f \in \mathcal{F},$$

$$W_{k}^{(T)}(u) = \frac{1}{\sqrt{T}} \sum_{t=2}^{[uT]} b_k(\Delta Y_t), \quad k \in \mathbb{N}.$$

The rationale of our notation is that we have $\Delta Y_t = \varepsilon_t$, for $t \geq 2$, under the null hypothesis of a unit root. Together with Assumption 1 this yields, still under the null hypothesis, weak convergence to Brownian motions. Note that the sums start at $t = 2$, so the partial sum processes are invariant with respect to the intercept μ. This property will facilitate the construction of tests that are invariant with respect to the intercept μ.

To introduce the limiting Brownian motions, we first note that there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P}_{0,0})$ supporting mutually independent Brownian motions W_ε and W_k, $k \in \mathbb{N}$, with

$$\operatorname{var}[W_\varepsilon(1)] = \sigma_f^2 \text{ and } \operatorname{var}[W_k(1)] = 1.$$

As $\phi_f(\varepsilon_1)$ is the score of the location model, it is well known (see, for example, Van der Vaart (2000)) that we have (under Assumption 1) $\mathbb{E}_f \phi_f(\varepsilon_1) = 0$ and $\mathbb{E}_f \phi_f(\varepsilon_1)\varepsilon_1 = 1$. Consequently, because ε_1 and $b_k(\varepsilon_1)$ are orthogonal for each k, we can decompose $\phi_f(\varepsilon_1) = \sigma_f^{-2}\varepsilon_1 + \sum_{k=1}^{\infty} \mathcal{J}_{f,k} b_k(\varepsilon_1)$, with coefficients $\mathcal{J}_{f,k} = \mathbb{E}_f b_k(\varepsilon_1)\phi_f(\varepsilon_1)$. This motivates, for $f \in \mathcal{F}$, the introduction of

$$W_{\phi_f} = \frac{1}{\sigma_f^2} W_\varepsilon + \sum_{k=1}^{\infty} \mathcal{J}_{f,k} W_k. \tag{3.2}$$

It is easy to verify that this indeed defines a Brownian motion and that we have

$$\operatorname{cov}(W_{\phi_f}(1), W_\varepsilon(1)) = 1, \quad \operatorname{cov}(W_{\phi_f}(1), W_k(1)) = \mathcal{J}_{f,k}, \; k \in \mathbb{N}, \text{ and } \operatorname{var}[W_{\phi_f}(1)] = J_f = \frac{1}{\sigma_f^2} + \sum_{k=1}^{\infty} \mathcal{J}_{f,k}^2. \tag{3.3}$$

Using the functional central limit theorem it follows that, under the null hypothesis, the partial sum processes weakly converge to the associated Brownian motions. And integrals like $\int_0^1 W_\varepsilon^{(T)}(u-) \mathrm{d}W_{\phi_f}^{(T)}(u)$ weakly converge to the associated stochastic integral with the limiting Brownian motions, i.e. $\int_0^1 W_\varepsilon(u) \mathrm{d}W_{\phi_f}(u)$. The precise statements are organized in Lemma **??** in the Appendix.

3.2. *Weak convergence of experiments and a structural representation of the limit experiment*

Fix $f \in \mathcal{F}$ and $\mu \in \mathbb{R}$. Let, for $\eta \in c_{00}$, $\mathrm{P}_{\mu,h,\eta}^{(T)}$ denote the law of $Y_1, \ldots, Y_T$ under (2.1)-(2.2) with autoregression parameter ρ given by (2.4) and innovation density (3.1). The following proposition shows that the semiparametric unit root model is of the Locally Asymptotically Brownian Functional (LABF) type; see Jeganathan (1995).

Proposition 3.2 Let $f \in \mathcal{F}$, $\eta \in c_{00}$, and $h_T \to h \in \mathbb{R}$.

(i) Then we have, under $\mathrm{P}_{\mu,0,0}^{(T)}$ and as $T \to \infty$,

$$\begin{aligned} \log \frac{\mathrm{dP}_{\mu,h_T,\eta}^{(T)}}{\mathrm{dP}_{\mu,0,0}^{(T)}} &= \log \frac{f_\eta^{(T)}(Y_1 - \mu)}{f(Y_1 - \mu)} + \sum_{t=2}^{T} \log \frac{f_\eta^{(T)}\left(\Delta Y_t - \frac{h}{T}(Y_{t-1} - \mu)\right)}{f(\Delta Y_t)} \\ &= h\Delta_{\rho;f}^{(T)} + \sum_{k=1}^{\infty} \eta_k \Delta_{b_k}^{(T)} - \frac{1}{2}\mathcal{I}_f^{(T)}(h,\eta) + o_P(1), \end{aligned} \tag{3.4}$$

where the central-sequence $\Delta^{(T)} = (\Delta^{(T)}_{\rho;f}, \Delta^{(T)}_{b})$, with $\Delta^{(T)}_{b} = (\Delta^{(T)}_{b_k})_{k\in\mathbb{N}}$, is given by

$$\Delta^{(T)}_{\rho;f} = \int_0^1 W^{(T)}_{\varepsilon}(u-)\mathrm{d}W^{(T)}_{\phi_f}(u) = \frac{1}{T}\sum_{t=2}^{T}(Y_{t-1}-Y_1)\phi_f(\Delta Y_t),$$

$$\Delta^{(T)}_{b_k} = W^{(T)}_{k}(1) = \frac{1}{\sqrt{T}}\sum_{t=2}^{T} b_k(\Delta Y_t), \quad k\in\mathbb{N},$$

and

$$\mathcal{I}^{(T)}_{f}(h,\eta) = h^2 J_f \int_0^1 (W^{(T)}_{\varepsilon}(u-))^2\mathrm{d}u + \|\eta\|_2^2 + 2h\int_0^1 W^{(T)}_{\varepsilon}(u-)\mathrm{d}u \sum_{k=1}^{\infty}\eta_k \mathcal{J}_{f,k}.$$

(ii) Moreover, with $\Delta_{\rho;f} = \int_0^1 W_\varepsilon(u)\mathrm{d}W_{\phi_f}(u)$ and $\Delta_{b_k} = W_k(1)$, we have, still under $\mathrm{P}^{(T)}_{\mu,0,0}$,

$$\frac{\mathrm{dP}^{(T)}_{\mu,h_T,\eta}}{\mathrm{dP}^{(T)}_{\mu,0,0}} \xrightarrow{d} \exp\left(h\Delta_{\rho;f} + \sum_{k=1}^{\infty}\eta_k\Delta_{b_k} - \frac{1}{2}\mathcal{I}_f(h,\eta)\right), \tag{3.5}$$

with

$$\mathcal{I}_{f}(h,\eta) = h^2 J_f \int_0^1 (W_{\varepsilon}(u))^2\mathrm{d}u + \|\eta\|_2^2 + 2h\int_0^1 W_{\varepsilon}(u)\mathrm{d}u \sum_{k=1}^{\infty}\eta_k \mathcal{J}_{f,k}.$$

(iii) For all $h\in\mathbb{R}$ and $\eta\in\ell_2$ the right-hand-side of (3.5) has expectation 1 under $\mathbb{P}_{0,0}$. □

The proof of (i) follows by an application of Proposition 1 Hallin, Van den Akker, and Werker (2015) which provides sufficient conditions for the quadratic expansion of log likelihood ratios. Of course, Part (ii) is not surprising and follows using the weak convergence of the partial-sum processes to Brownian motions (and integrals involving the partial-sum processes to stochastic integrals). And Part (iii) follows by verifying the Novikov condition. All these proofs are organized in the appendix.

Part (iii) of the proposition implies that we can introduce, for $h\in\mathbb{R}$ and $\eta\in\ell_2$, new probability measures $\mathbb{P}_{h,\eta}$ on the measurable space $(\Omega,\mathcal{F})$ (on which the Brownian motions W_ε, W_k, etc. were defined) by their Radon-Nikodym derivatives with respect to $\mathbb{P}_{0,0}$:

$$\frac{\mathrm{d}\mathbb{P}_{h,\eta}}{\mathrm{d}\mathbb{P}_{0,0}} = \exp\left(h\Delta_{\rho;f} + \sum_{k=1}^{\infty}\eta_k\Delta_{b_k} - \frac{1}{2}\mathcal{I}_f(h,\eta)\right).$$

Proposition 3.2 implies that the sequence of unit root experiments (each $T\in\mathbb{N}$ yields an experiment) weakly converges (in the Le Cam sense) to the experiment described by the probability measures $\mathbb{P}_{h,\eta}$. To formulate this formally, we define the sequence of experiments by $\mathcal{E}^{(T)}(\mu,f) = \left(\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T), (\mathrm{P}^{(T)}_{\mu,h,\eta}\,|\, h\in\mathbb{R},\, \eta\in c_{00})\right)$, $T\in\mathbb{N}$, and the limit experiment by, with $\mathcal{B}_C$ the Borel σ-field on $C[0,1]$, $\mathcal{E} = (C[0,1]\times C^{\mathbb{N}}[0,1], \mathcal{B}_{\mathcal{C}}\otimes(\otimes_{k=1}^{\infty}B_C), (\mathbb{P}_{h,\eta}\,|\, h\in\mathbb{R},\, \eta\in c_{00}))$.

Corollary 3.1 Let $\mu\in\mathbb{R}$ and $f\in\mathcal{F}$. Then the sequence of experiments $\mathcal{E}^{(T)}(\mu,f)$, $T\in\mathbb{N}$, converges (as $T\to\infty$) to the experiment $\mathcal{E}$. □

The Asymptotic Representation Theorem (see, for example, Chapter 9 in Van der Vaart (2000)) now shows us that for any statistic A_T which converges in distribution to the law $L_{h,\eta}$ under $\mathrm{P}^{(T)}_{\mu,h,\eta}$ there exists a (randomized) statistic A, defined on $\mathcal{E}$, such that the law of A under $\mathbb{P}_{h,\eta}$ is given by $L_{h,\eta}$. This allows us to study optimal inference in the limit experiment: the "best" procedure in the limit experiment yields a bound for the sequence of experiments. If one is able to construct a statistic (for the sequence) that attains the bound, it follows that the bound is sharp and the statistic is (asymptotically) optimal.

To obtain more insight in the limit experiment $\mathcal{E}$ the following proposition, which follows by an application of Girsanov's theorem, provides a "structural" description of the limit experiment.

PROPOSITION 3.3 Let $f \in \mathcal{F}$, $\eta \in \ell_2$, and $h \in \mathbb{R}$. Then the processes Z_ε and Z_k, $k \in \mathbb{N}$, defined by the starting values $Z_\varepsilon(0) = Z_k(0) = 0$ and the stochastic differential equations, for $u \in [0,1]$,

$$\begin{aligned} \mathrm{d}Z_\varepsilon(u) &= \mathrm{d}W_\varepsilon(u) - hW_\varepsilon(u)\mathrm{d}u, \\ \mathrm{d}Z_k(u) &= \mathrm{d}W_k - h\mathcal{J}_{f,k}W_\varepsilon(u)\mathrm{d}u - \eta_k \mathrm{d}u, \quad k \in \mathbb{N}, \end{aligned}$$

are Brownian motions under $\mathbb{P}_{h,\eta}$: their law is the same as the law of $(W_\varepsilon, (W_k)_{k\in\mathbb{N}})$ under $\mathbb{P}_{0,0}$.

3.3. *Invariance and power envelopes*

Using Proposition 3.3 we first discuss a natural invariance structure, with respect to the infinite-dimensional nuisance parameter η, for the limit experiment. We derive the maximal invariant and apply the Neyman-Pearson lemma to obtain the power envelope for invariant tests in the limit experiment. Next we exploit the Asymptotic Representation Theorem to translate the results to the sequence of unit root models.

The limit experiment

We first discuss the testing problem for the limit experiment $\mathcal{E}$. We thus observe the processes W_ε and W_k, $k \in \mathbb{N}$, on the time interval $[0,1]$ from the model $(\mathbb{P}_{h,\eta} \,|\, h \in \mathbb{R},\, \eta \in c_{00})$. We are interested in the power envelope for testing the hypothesis

$$\text{(3.6)} \qquad \mathrm{H}_0 : \; h = 0, \; (\eta \in c_{00}) \text{ versus } \mathrm{H}_a : \; h < 0, \; (\eta \in c_{00}).$$

We will focus on test statistics whose distribution, under $\mathbb{P}_{h,\eta}$, does not depend on η (for all h). Their law is thus invariant with respect to the nuisance parameter.

To see how such statistics should look like, we introduce, for $\eta \in \ell_2$, the transformations $g^\eta = (g_k)_{k\in\mathbb{N}} : C^{\mathbb{N}}[0,1] \to C^{\mathbb{N}}[0,1]$ defined by

$$g_k : \; C[0,1] \ni (W_k(u))_{u\in[0,1]} \mapsto (W(u) - u\eta_k)_{u\in[0,1]} \in C[0,1],$$

i.e. g_k adds a drift $u \mapsto -\eta_k u$ to W_k. Proposition 3.3 implies that the law of $(W_\varepsilon, (g_k(W_k))_{k\in\mathbb{N}})$ under $\mathbb{P}_{h,0}$ is the same as the law of $(W_\varepsilon, (W_k)_{k\in\mathbb{N}})$ under $\mathbb{P}_{h,\eta}$. Hence our testing problem (3.5) remains invariant with respect to the transformation g^η. Therefore, following the invariance principle, it is natural to focus on test statistics which are invariant with respect to these transformations, i.e.

$$\text{(3.7)} \qquad t(W_\varepsilon, (g_k^\eta(W_k))_{k\in\mathbb{N}}) = t(W_\varepsilon, W_k) \text{ for all } g^\eta, \; \eta \in c_{00}.$$

Given a process W let us define the associated *bridge process* by $B^W(u) = W(u) - uW(1)$. Now note that we have, for all $u \in [0,1]$,

$$\begin{aligned} B^{g_k(W_k)}(u) &= [g_k(W_k)](u) - u[g_k(W_k)](1) = W_k(u) - u\eta_k - u(W_k(1) - 1 \times \eta_k) = W_k(u) - uW_k(1) \\ &= B^{W_k}(u), \end{aligned}$$

i.e. taking the bridge of the observed processes ensures invariance with respect to adding drifts. This shows that statistics that are measurable with respect to the σ-field, with $B_k = B^{W_k}$,

$$\mathcal{M} = \sigma\left(W_\varepsilon(u),\, B_k(u),\, u \in [0,1]\right)$$

are invariant (with respect to g^η, $\eta \in c_{00}$). It is, however, not clear that we did not throw away too much data. Formally, we need $\mathcal{M}$ to be *maximally invariant* which means that each invariant statistic is $\mathcal{M}$-measurable. The following theorem, which once more exploits the structural description of the limit experiment, shows that this indeed is the case.

THEOREM 3.1 Let $f \in \mathcal{F}$. The σ-field $\mathcal{M}$ is maximally invariant. □

The theorem implies that invariant inference should be based on $\mathcal{M}$. An application of the Neyman-Pearson lemma, using $\mathcal{M}$ as observation, yields the power envelope for the class of invariant tests. We thus need to consider the likelihood ratios of $\mathcal{M}$, which are given by

$$\frac{\mathrm{d}\mathbb{P}_h^{\mathcal{M}}}{\mathrm{d}\mathbb{P}_0^{\mathcal{M}}} = \mathbb{E}_0\left[\frac{\mathrm{d}\mathbb{P}_{h,\eta}}{\mathrm{d}\mathbb{P}_{0,\eta}} \mid \mathcal{M}\right],$$

where the conditional expectation indeed does not depend on η because of the invariance. To calculate the conditional expectation we first introduce $B_{\phi_f} = B^{W_{\phi_f}}$, i.e. the bridge process associated to W_{ϕ_f} (see (3.2)). Now we can decompose $\Delta_{\rho;f} = I + II$ with

$$I = \int_0^1 W_\varepsilon(u)\mathrm{d}B_{\phi_f}(u) + \frac{1}{\sigma_f^2}W_\varepsilon(1)\int_0^1 W_\varepsilon(u)\mathrm{d}u \text{ and } II = \left(\sum_{k=1}^\infty \mathcal{J}_{f,k}W_k(1)\right)\int_0^1 W_\varepsilon(u)\mathrm{d}u.$$

Note that part I is $\mathcal{M}$-measurable. Under $\mathbb{P}_{0,0}$ the random variables $W_k(1)$, $k \in \mathbb{N}$, are independent to W_ε and B_k, $k \in \mathbb{N}$. Indeed, the independence to W_ε holds by construction and the independence to B_k follows from the Gaussianity and $\mathrm{cov}_{0,0}(B_k(u), W_k(1)) = \mathrm{cov}_{0,0}(W_k(u), W_k(1)) - u\,\mathrm{cov}_{0,0}(W_k(1), W_k(1)) = 0$. We thus obtain

$$\begin{aligned}\mathbb{E}_0\left[\frac{\mathrm{d}\mathbb{P}_{h,\eta}}{\mathrm{d}\mathbb{P}_{0,\eta}} \mid \mathcal{M}\right] &= \exp\left(h \times I - \frac{1}{2}\mathcal{I}_f(h,\eta)\right)\mathbb{E}_{0,0}\left[\exp\left(\sum_{k=1}^\infty (h\mathcal{J}_{f,k}\int_0^1 W_\varepsilon(u)\mathrm{d}u + \eta_k)W_k(1)\right) \mid \mathcal{M}\right] \\ &= \exp\left(h \times I - \frac{1}{2}\mathcal{I}_f(h,\eta)\right)\exp\left(\frac{1}{2}\sum_{k=1}^\infty (h\mathcal{J}_{f,k}\int_0^1 W_\varepsilon(u)\mathrm{d}u + \eta_k)^2\right).\end{aligned}$$

This yields

$$\begin{aligned}\frac{\mathrm{d}\mathbb{P}_h^{\mathcal{M}}}{\mathrm{d}\mathbb{P}_0^{\mathcal{M}}} &= \exp\left(h\Delta_f^\star - \frac{1}{2}h^2\mathcal{I}_f^*\right) \text{ with} \\ \Delta_f^\star &= \int_0^1 W_\varepsilon(u)\mathrm{d}B_{\phi_f}(u) + \frac{1}{\sigma_f^2}W_\varepsilon(1)\int_0^1 W_\varepsilon(u)\mathrm{d}u, \\ \mathcal{I}_f^* &= J_f\int_0^1 W_\varepsilon^2(u)\mathrm{d}u - \left(\int_0^1 W_\varepsilon(u)\mathrm{d}u\right)^2\sum_{k=1}^\infty \mathcal{J}_{f,k}^2 = J_f\int_0^1 W_\varepsilon^2(u)\mathrm{d}u - \left(\int_0^1 W_\varepsilon(u)\mathrm{d}u\right)^2\left(J_f - \frac{1}{\sigma_f^2}\right).\end{aligned}$$

Let us denote the $(1-\alpha)$-quantile of $\mathrm{d}\mathbb{P}_h^{\mathcal{M}}/\mathrm{d}\mathbb{P}_0^{\mathcal{M}}$ under $\mathbb{P}_{0,\eta}$, which does not depend on η, by $c(h, f; \alpha)$. Let $\phi_{f,\alpha}^\star(\bar{h}) = 1\{\mathrm{d}\mathbb{P}_h^{\mathcal{M}}/\mathrm{d}\mathbb{P}_0^{\mathcal{M}} \geq c(\bar{h}, f; \alpha)\}$, which of course is a test of size α. The power function of this test is given by

$$h \mapsto \pi_{f,\alpha}^*(h; \bar{h}) = \mathbb{E}_0\left[\phi_{f,\alpha}^\star(\bar{h})\frac{\mathrm{d}\mathbb{P}_h^{\mathcal{M}}}{\mathrm{d}\mathbb{P}_0^{\mathcal{M}}}.\right]$$

An application of the Neyman-Pearson lemma yields the following corollary.

COROLLARY 3.2 Let $f \in \mathcal{F}$ and $\alpha \in (0,1)$. Let ϕ be a (possibly randomized) test that is $\mathcal{M}$-measurable and is of size α, i.e. $\mathbb{E}_0\phi \leq \alpha$. Let π denote the power function of this test, i.e. $\pi(h) = \mathbb{E}_h\phi$. Then we have

$$\pi(h) \leq \pi_{f,\alpha}^*(h; h).$$

□

The test $\phi_{f,\alpha}^\star(\bar{h})$ thus is point-optimal, i.e. it is tangent to the power envelope $h \mapsto \pi_{f,\alpha}^*(h; h)$ at $h = \bar{h}$.

The asymptotic power envelope for asymptotically invariant unit root tests

Now we translate the results for the limiting LABF experiment to the unit root model of interest. To mimick the invariance in the limit experiment we introduce the following definition.

DEFINITION 1 A sequence of test statistics ψ_T is said to be (asymptotically) invariant if the distribution of ψ_T weakly converges under $\mathrm{P}^{(T)}_{[}h, \eta]$ for all $h \leq 0$ and $\eta \in c_{00}$ to the distribution of an invariant test in the limit experiment $\mathcal{E}$. □

The Asymptotic Representation Theorem now yields the following corollary.

COROLLARY 3.3 Let $f \in \mathcal{F}$, $\mu \in \mathbb{R}$, and $\alpha \in (0,1)$. Let ϕ_T, $T \in \mathbb{N}$, an invariant test of size α, i.e. $\limsup_{T\to\infty}\mathbb{E}_{0,\eta}\phi_T \leq \alpha$ for all $\eta \in c_{00}$. Let π_T denote the power function of ϕ_T, i.e. $\pi_T(h,\eta) = \mathbb{E}_{h,\eta}\phi_T$. Then we have

$$\limsup_{T\to\infty}\pi_T(h,\eta) \leq \pi_{f,\alpha}^*(h; h), \quad \eta \in c_{00}.$$

□

The power envelope for invariant tests in the limit experiment thus provides an upper bound to the asymptotic power of invariant tests for the unit root hypothesis. The next section introduces a class of tests that attains this bound (point-wise) and therefore demonstrates that the bound indeed constitutes the asymptotic power envelope for invariant unit root tests.

4. A CLASS OF SEMIPARAMETRICALLY OPTIMAL HYBRID RANK TESTS

The appearance of the bridge process B_{ϕ_f} in the "efficient central-sequence" $\Delta_f^\star$ naturally suggests the (partial) use of ranks in the construction of feasible test statistics: we can construct an empirical anologue of B_{ϕ_f} by considering a partial-sum process which only depends on the observations $Y_1, \dots, Y_T$ via the ranks, R_t, of ΔY_t, $t = 2, \dots, T$.

This rank process requires the choice of a *reference density*. This should be compared to Quasi-ML methods: if the true innovation density happens to be the same as the selected reference density the inference procedure is optimal, while the procedure is valid, i.e. has the proper size, in case the true innovation density does not coincide with the reference density. We need the following mild (and standard) assumption on the reference density.

ASSUMPTION 2 The density $g \in \mathcal{F}$ satisfies

$$\lim_{T\to\infty} \frac{1}{T}\sum_{i=1}^{T}\left\{\frac{g'}{g}\left(G^{-1}\Big(\frac{i}{T+1}\Big)\right)\right\}^2 = I_g$$

□

Now we can formulate the following extension of Lemma A.1 in Hallin, Van den Akker, and Werker (2011).

LEMMA 4.1 Let $f \in F$, $\mu \in \mathbb{R}$, and g satisfy Assumption 2. Consider the partial sum process, defined on $[0, 1]$,

$$B_g^{(T)}(u) = \frac{1}{\sqrt{T}}\sum_{t=2}^{[uT]}\phi_g\left(G^{-1}\left(\frac{R_t}{T+1}\right)\right). \tag{4.1}$$

We have, under $P_{\mu,0,f}$ and as $T \to \infty$,

$$\begin{bmatrix} W_\varepsilon^{(T)} \\ W_{\phi_f}^{(T)} \\ B_g^{(T)} \\ \int_0^1 W_\varepsilon^{(T)}(u-)\mathrm{d}B_g^{(T)}(u) \end{bmatrix} \Rightarrow \begin{bmatrix} W_\varepsilon \\ W_{\phi_f} \\ B_{\phi_g} \\ \int_0^1 W_\varepsilon(u)\mathrm{d}B_{\phi_g}(u) \end{bmatrix}, \tag{4.2}$$

where B_{ϕ_g} is the Brownian bridge associated to a Brownian motion W_{ϕ_g} and the covariance per unit of time of this process is given by

$$\mathrm{cov}\begin{pmatrix} W_\varepsilon(1) \\ W_{\phi_f}(1) \\ W_{\phi_g}(1)\end{pmatrix} = \begin{pmatrix} \sigma_f^2 & 1 & \int_0^1 F^{-1}(u)\phi_g(G^{-1}(u))\mathrm{d}u \\ & J_f & I_{fg} \\ & & J_g \end{pmatrix},$$

where

$$I_{fg} = \int_0^1 \phi_f(F^{-1}(u))\phi_g(G^{-1}(u))\mathrm{d}u.$$

The weak convergence in (4.2) is on $D^3[0, 1] \times \mathbb{R}$ equipped with the uniform topology. □

ASSUMPTION 3 Let $\hat\sigma_T^2$ a consistent estimator of σ_f^2 under the null hypothesis, i.e. for all $f \in \mathcal{F}$ $\hat\sigma_T^2 \xrightarrow{p} \sigma_f^2$ under $\mathrm{P}_{0,\eta}^{(T)}$, for all $\eta \in c_{00}$ and as $T \to \infty$. □

The lemma motivates to consider statistics of the type, for $\bar h < 0$ fixed,

$$\hat L_T^g(\bar h) := \bar h\tilde\Delta_g^{(T)} - \frac{1}{2}\bar h^2\tilde{\mathcal{I}}_g, \tag{4.3}$$

with

$$\tilde{\Delta}_g^{(T)} = \int_0^1 W_\varepsilon^{(T)}(u-)\mathrm{d}B_{\phi_g}^{(T)}(u) + \frac{1}{\hat{\sigma}_T^2} W_\varepsilon^{(T)}(1) \int_0^1 W_\varepsilon^{(T)}(u-)\mathrm{d}u$$

and

$$\tilde{\mathcal{I}}_g^{(T)} = J_g \int_0^1 W_\varepsilon^{(T)^2}(u-)\mathrm{d}u - \left(\int_0^1 W_\varepsilon^{(T)}(u-)\mathrm{d}u\right)^2 \left(J_g - \frac{1}{\sigma_g^2}\right).$$

Lemma 4.1 implies that we have, under $P_{\mu,0,f}$,

$$\tilde{\Delta}_g^{(T)} \xrightarrow{d} \tilde{\Delta}_g = \int_0^1 W_\varepsilon(u)\mathrm{d}B_{\phi_g}(u) + \frac{1}{\sigma^2} W_\varepsilon(1) \int_0^1 W_\varepsilon(u)\mathrm{d}u \text{ and}$$
$$\tilde{\mathcal{I}}_g^{(T)} \xrightarrow{d} J_g \int_0^1 W_\varepsilon^2(u)\mathrm{d}u - \left(\int_0^1 W_\varepsilon(u)\mathrm{d}u\right)^2 \left(J_g - \frac{1}{\sigma_g^2}\right).$$

Unfortunately, the limiting null distribution $(\tilde{\Delta}_g, \tilde{\mathcal{I}}_g)$ is not distribution-free, because it depends on σ_f^2 and $\rho_{\varepsilon,g} = \mathrm{cov}(W_\varepsilon(1), W_{\phi_g}(1))$ which does not vanish in general. Denote the $(1-\alpha)$-quantile of $\bar{h}\tilde{\Delta}_g - 0.5\bar{h}^2\tilde{\mathcal{I}}_g$ by $c(\bar{h}, \sigma_f^2, \rho_{\varepsilon,g}, J_g; \alpha)$. The parameter $\rho_{\varepsilon,g}$ can be estimated consistently from the data by

$$\hat{\rho}_{\varepsilon,g}^{(T)} = \frac{1}{T}\sum_{t=2}^{T} \Delta Y_t \phi_g(G^{-1}(R_t/(T+1))).$$

This leads to the test

$$\phi_T^g(\bar{h}, \alpha) := \mathbb{1}\left\{\hat{L}_T^g(\bar{c}) \geq c(\bar{h}, \hat{\sigma}_f^2, \hat{\rho}_{\varepsilon,g}^{(T)}, J_g; \alpha)\right\}. \tag{4.4}$$

Since the test is based on the ranks and the levels of ΔY_t, we call these tests Hybrid Rank Tests (HRT).

Proposition 4.2 Let $\mu \in \mathbb{R}$, $f \in \mathcal{F}$, $\alpha \in (0,1)$, $\bar{h} < 0$, and g satisfy Assumption 2. Then we have:
(i) The HRT $\phi_T^g(\bar{h}, \alpha)$ is asymptotically of size α.
(ii) The HRT $\phi_T^g(\bar{h}, \alpha)$ is asymptotically invariant.
(iii) If $f = g$, the HRT $\phi_T^g(\bar{h}, \alpha)$ is point-optimal at $h = \bar{h}$.
The HRTs thus are valid irrespective of the choice of the reference density and are (point) optimal for a correctly specified reference density.

5. MONTE CARLO STUDY

This section reports the results of a Monte Carlo study to assess the quality of the asymptotic approximations and to analyze the finite-sample performances of the proposed Hybrid Rank Tests (HRTs). We compare the performances of the HRTs, for a selection of reference densities g (denoted HRT^g), to those of Dickey-Fuller t-test (denoted DF-t), Dickey Fuller estimator test (denoted DF-ρ) from Dickey and Fuller (1979) and the family of point-optimal tests (denoted ERS), modified DF test (denoted DF-GLS) from Elliott, Rothenberg, and Stock (1996).

Since the model 2.1 contains an unknown constant in the deterministic term, we follow ERS(1996) and choose -7 to be the fixed alternative for ERS test. Similar as the family of ERS point optimal tests, HRT^g for given g is also a family of point-optimal tests associated with a fixed alternative $\bar{h}$. In this versions we fix $\bar{h} = -7$ for HRT^g in the simulation study.

In section 5.1 we provide an analysis of sizes and a table of critical values for the HRT^g test with three reference density functions: Gaussian, Laplace and Student's t_3. In section 5.2, we study the large sample performance of the tests we listed above with various innovation densities. And finally we provide the finite sample performance in section 5.3. Throughout we report nominal rejection frequencies.

5.1. *Sizes*

The asymptotic critical values of the HRTs are simulation based. In this version we throughout use, for computational reasons, $\rho_{\varepsilon,g} = 1$ (and thus do not estimate this quantity). Table I presents the simulated values for various values α of the tests sizes and various reference densities g (Gaussian, Laplace and Student's t_3).

TABLE I

CRITICAL VALUES FOR $\bar{h} = -7$

g		$\alpha = 1\%$	$\alpha = 2.5\%$	$\alpha = 5\%$	$\alpha = 10\%$
Gaussian	T=25	1.97	1.64	1.28	0.68
	T=50	2.21	1.89	1.52	0.90
	T=100	2.35	2.03	1.66	1.04
	T=250	2.44	2.13	1.76	1.14
	T=2500	2.50	2.21	1.83	1.21
Laplace	T=25	2.48	1.81	1.14	0.15
	T=50	2.76	2.01	1.27	0.19
	T=100	2.87	2.10	1.32	0.20
	T=250	2.94	2.14	1.34	0.21
	T=2500	2.97	2.16	1.35	0.21
t_3	T=25	2.59	2.09	1.59	0.85
	T=50	2.77	2.20	1.64	0.79
	T=100	2.86	2.25	1.62	0.68
	T=250	2.93	2.25	1.56	0.54
	T=2500	2.95	2.25	1.45	0.31

Simulated critical values (based on 100,000 replications) for various reference densities g (Gaussian, Laplace and Student's t_3) and significant levels α.

5.2. *Large sample performance*

In this section we evaluate the (local) powers of the HRTs for large samples together with those of the competing tests. The results are based on 20,000 replications, 2500 sample size and 5% significance level.

Figure 1 graphs the power functions for large samples of selected tests along with the semiparametric power envelope when the true innovation density f is Laplace, Student t_3 and Gaussian, respectively. The results suggest two conclusions. First, when the chosen reference density g happens to be the true density f, the large sample power of HRT$^{g=f}$ test is very close to the semiparametric power envelope and tangent with it at a certain point (the chosen alternative point -7). This is indeed why we call HRT$^{g=f}$ point-optimal. Especially when f is not Gaussian, the power of HRT$^{g=f}$ is much higher than those of the other tests.

Second, if we keep choosing g to be Gaussian, say HRT$^{g=\phi}$, when f is Gaussian, it works as well as the other tests for large sample size; while when f is not Gaussian, HRT$^{g=\phi}$ is still of more power than the other tests (This property corresponse to the Charnoff-Savage results).

In figure 2, we tried some other innovation densities: Student's t_1, t_2 and skewed normal distribution (with skewness 0.8145). It shows that our hybrid rank test with Gaussian reference density (HRT$^{g=\phi}$) always dominates the other tests when the true innovation density is not normally distributed.

5.3. *Finite sample performance*

The convergence behaviour of HRT$^{g=f}$ is shown in figure 3: with the increase of sample size, the power of HRT$^{g=f}$ converges to the corresponding power envelope.

Figure 4 is the finite-sample version of figure 1: it graphs the finite-sample power functions of selected tests along with the semiparametric power envelope when the true innovation density f is Gaussian, Laplace and Student t_3, respectively. And it leads to the following conclusions. First, when f is not Gaussian, the power of HRT$^{g=f}$ is larger than that of HRT$^{g=\phi}$, while the power of the latter is still larger than that of ERS. Only when f is Gaussian, the power of HRT$^{g\neq f}$ is lower than that of ERS, and the power of HRT$^{g=\phi=f}$ is very close to that of ERS.

Comparing with DF-ρ becomes a bit more complicated for the finite-sample case than for the asymptotic case: for all three different true innovation densities: both HRT$^{g=f}$ and HRT$^{g=\phi}$ has apparent larger power than DF-ρ for small alternatives ($-c < 15$); while for large alternatives ($-c > 20$), the powers of HRT$^{g=f}$ and HRT$^{g=\phi}$ is slightly lower than the powers of DF-ρ but never far away.

In figure 5 we tried more types of f but with only HRT$^{g=\phi}$ (Gaussian reference density). The set of f contains: Student's t_1, t_2, skewed-normal distribution (with skewness 0.8145), skewed-t_4 (with skewness approximates 3.7), stable distribution (with stability parameter 0.75, skewness parameter 0, scale parameter 1 and location parameter 0) and Pesrson distribution (with location parameter 0, scale parameter 1, skewness parameter 9 and kurtosis parameter 1296). The graph shows that HRT$^{g=\phi}$ dominates ERS test by gaining plenty of power from non-normality in all these cases.

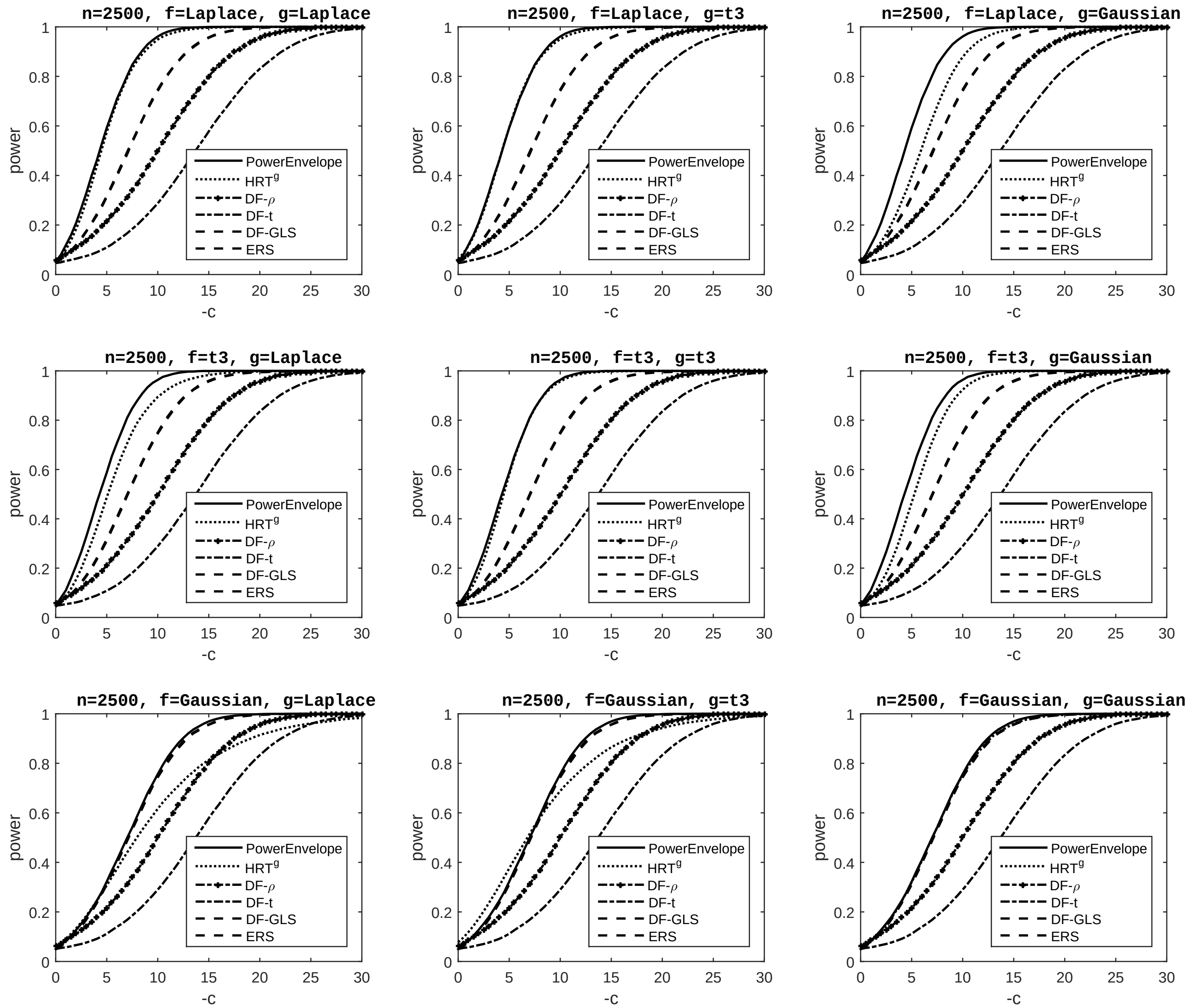

FIGURE 1.— Asymptotic power functions of selected unit root tests and various true innovation densities: Gaussian, Laplace, Student's t_3.

6. CONCLUSION AND DISCUSSION

To be completed

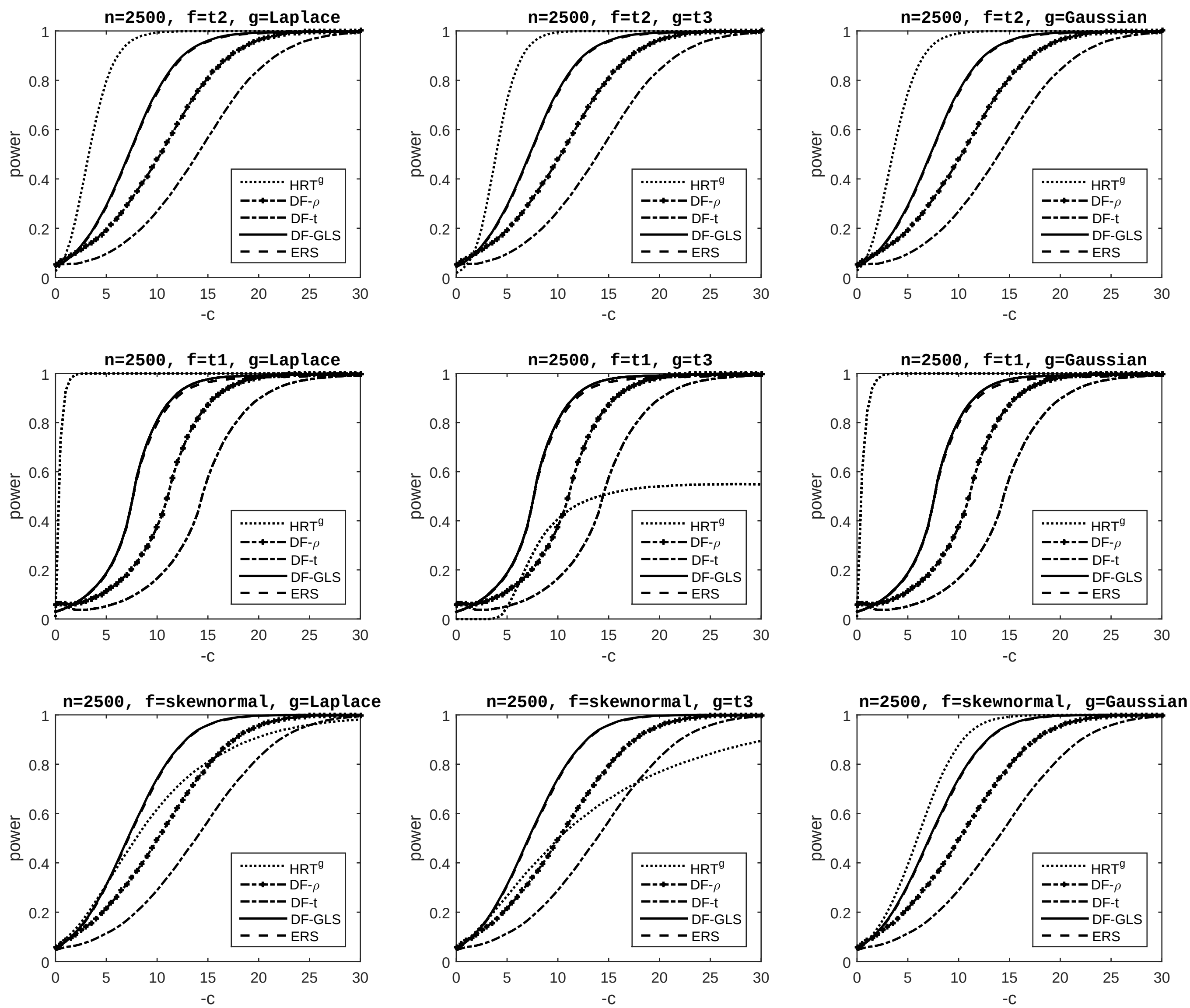

FIGURE 2.— Asymptotic power functions of selected unit root tests and various true innovation densities: Student's t_1, Student's t_2, skewed-normal.

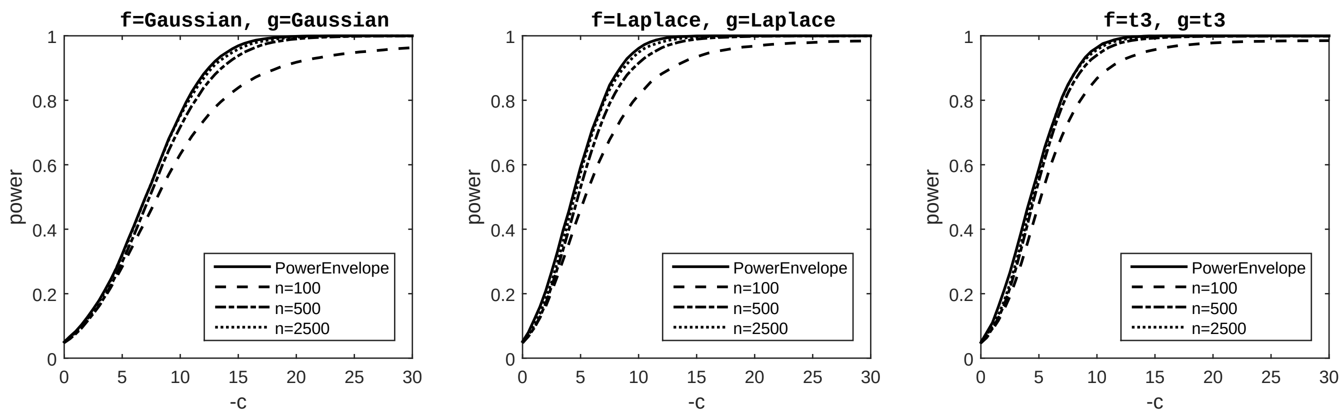

FIGURE 3.— Powers of HRT^g when $g = f$ with different sample sizes.

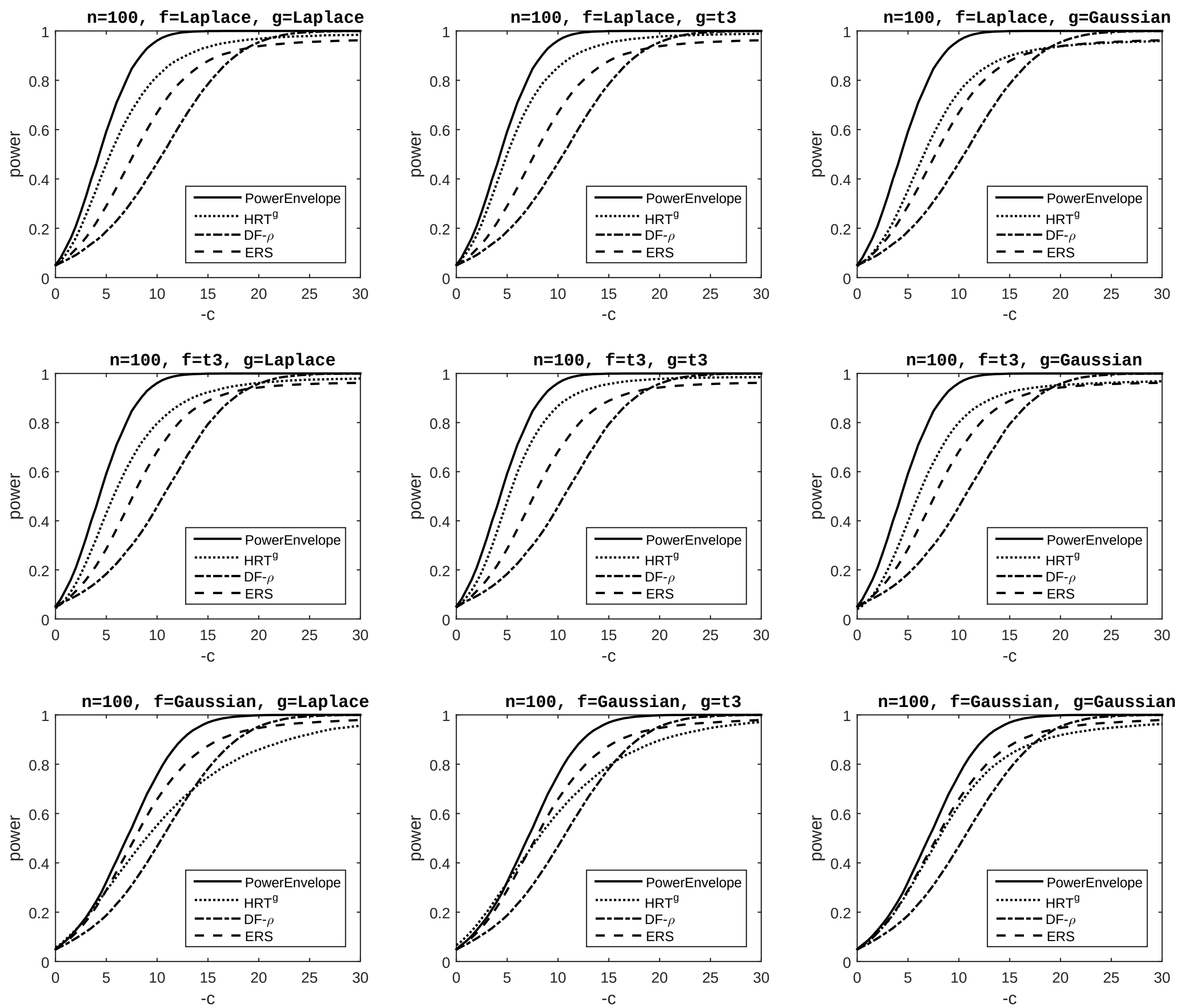

FIGURE 4.— Finite-sample power functions of selected unit root tests and various true innovation densities: Gaussian, Laplace, Student's t_3.

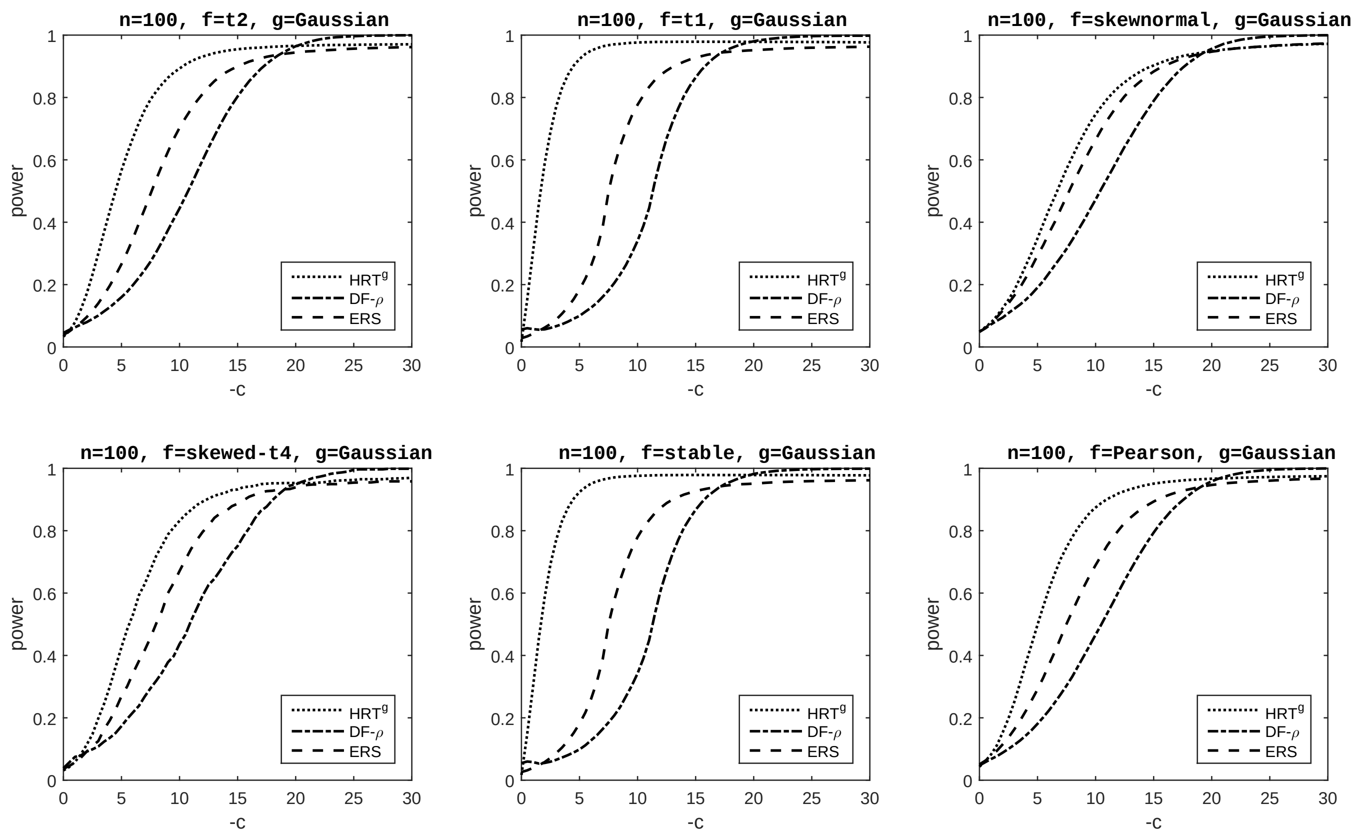

FIGURE 5.— Finite-sample power functions of selected unit root tests and various true innovation densities: Student's t_1, Student's t_2, skewed-normal, skewed-t_4, stable distribution, Pearson distribution.

REFERENCES

Ahn, S.A., S.B. Fotopoulos, and L. He (2003). Unit root tests with infinite variance errors. *Econometric Reviews* 20, 461–483.

Bhargava, A. (1986). On the theory of testing for unit roots in observed time series. *The Review of Economic Studies* 53, 369–384.

Bickel, P.J., C.A.J. Klaassen, Y. Ritov, and J.A. Wellner (1993), *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press.

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

Callegari, F., N. Cappuccio, and D. Lubian (2003). Asymptotic inference in time series regressions with a unit root and infinite variance errors. *Journal of Statistical Planning and Inference* 116, 277–303.

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.

Coudin, E. and J.M. Dufour (2009). Finite-sample distribution-free inference in linear median regression under heteroskedasticity and nonlinear dependence of unknown form. *Econometrics Journal* 12, 19–49.

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.

Dufour, J.-M. and M.L. 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. and U.K. Müller (2006). Minimizing the impact of the initial condition on testing for unit roots. *Journal of Econometrics* 135, 285–310.

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.

Hallin, M., R. van den Akker, and B.J.M. Werker (2011). A class of simple distribution-free rank-based unit root tests. *Journal of Econometrics* 163, 200–214.

Hallin, M., R. van den Akker, and B.J.M. Werker (2015). On quadratic expansions of log-likelihoods and a general asymptotic linearity result. *Mathematical Statistics and Limit Theorems : Festschrift in honour of Paul Deheuvels*, 12 p.

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.

Jeganathan, P. (1995). Some aspects of asymptotic theory with applications to time series models. *Econometric Theory* 11, 818–887.

Le Cam, L.M. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer-Verlag, New York.

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

Patterson, K. (2011). Unit root tests in time series, Volume 1: key concepts and problems. Palgrave Macmillan.

Patterson, K. (2012). Unit root tests in time series, Volume 2: extensions and developments. Palgrave Macmillan.

Phillips, P.C.B. (1987). Time series regression with a unit root. *Econometrica* 55, 277-301.

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.

Saikkonen, P. and R. Luukkonen (1993). Point optimal tests for testing the order of differencing in ARIMA models. *Econometric Theory* 9, 343–362.

Van der Vaart, A. (2000). *Asymptotic Statistics* (first edition). Cambridge University Press.

White, J.S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *The Annals of Mathematical Statistics* 29, 1188–1197.

APPENDIX A: PROOFS

Proof of Proposition 3.1:
For notational convenience we drop the superscript "(T)" in the following and thus write f_η instead of $f_\eta^{(T)}$. Moreover, we consider T' such that f_η is nonnegative. We have

$$f_\eta'(e) = f'(e)\left(1+\frac{1}{\sqrt{T}}\sum_{k=1}^\infty \eta_k b_k(e)\right) + f(e)\frac{1}{\sqrt{T}}\sum_{k=1}^\infty \eta_k b_k'(e), \quad \text{a.e.}.$$

There exist $C_1, C_2 < \infty$ such that we have, for all $T \geq T'$, $\left\|1+T^{-1/2}\sum_{k=1}^\infty \eta_k b_k\right\|_\infty \leq C_1$ and $\left\|T^{-1/2}\sum_{k=1}^\infty \eta_k b_k'\right\|_\infty^2 \leq C_2$. Moreover, there exists $C_3 > 0$ such that, again for all $T \geq T'$, $\|(1+T^{-1/2}\sum_{k=1}^\infty \eta_k b_k)^{-1}\|_\infty^2 \leq C_3$. Using these observations we immediately obtain

$$\int \left(-\frac{f_\eta'(e)}{f_\eta(e)}\right)^2 f_\eta(e)\mathrm{d}e \leq 2C_1 J_f + 2\frac{C_2}{C_3}\int f_\eta(e)\mathrm{d}e < \infty,$$

which concludes the proof. *Q.E.D.*

PROOF OF PROPOSITION 3.2: In this proof, all limits, o_p, and O_p quantities are to be understood as $T \to \infty$ and under the measure in which H_0 holds and the true density function of ε is f. The log-likelihood ratio is

$$L_T^{f,h}(c,\eta) = \sum_{t=2}^{T} \log \frac{f(\Delta y_t - \frac{c}{T} y_{t-1})}{f(\Delta y_t)} + \sum_{t=2}^{T} \log[1 + \frac{1}{\sqrt{T}} \sum_k \eta_k b_k (\Delta y_t - \frac{c}{T} y_{t-1})].$$

For the first term, by the proof is given by Jansson(2008) we have

$$\sum_{t=2}^{T} \log \frac{f(\Delta y_t - \frac{c}{T} y_{t-1})}{f(\Delta y_t)} = c \left[\frac{1}{T} \sum_{t=2}^{T} \phi_f(\Delta y_t) y_{t-1} \right] - \frac{1}{2} c^2 \left[\frac{\mathcal{I}_{ff}}{T^2} \sum_{t=2}^{T} y_{t-1}^2 \right] + o_p(1).$$

For the second term,

$$\begin{aligned}
&\sum_{t=2}^{T} \log \left[1 + \frac{1}{\sqrt{T}} \sum_k \eta_k b_k (\Delta y_t - \frac{c}{T} y_{t-1}) \right] \\
=&\sum_{t=2}^{T} \log \left[1 + \frac{1}{\sqrt{T}} \sum_k \eta_k \left(b_k(\Delta y_t) - \frac{c}{T} y_{t-1} b_k'(\Delta y_t) - \frac{c}{T} y_{t-1} r_k (\Delta y_t - \frac{c}{T} y_{t-1}) \right) \right] \\
=&\sum_{t=2}^{T} \log \left[1 + \sum_k \left(\frac{\eta_k}{\sqrt{T}} b_k(\Delta y_t) - \frac{c\eta_k}{T^{3/2}} y_{t-1} b_k'(\Delta y_t) + R_k^{Tt} \right) \right] \\
=&\sum_{t=2}^{T} \left[\sum_k \left(\frac{\eta_k}{\sqrt{T}} b_k(\Delta y_t) - \frac{c\eta_k}{T^{3/2}} y_{t-1} b_k'(\Delta y_t) + R_k^{Tt} \right) \right] \\
&- \frac{1}{2} \sum_{t=2}^{T} \left[\sum_k \left(\frac{\eta_k}{\sqrt{T}} b_k(\Delta y_t) - \frac{c\eta_k}{T^{3/2}} y_{t-1} b_k'(\Delta y_t) + R_k^{Tt} \right) \right]^2 (1 + \beta^{Tt})
\end{aligned}$$

where $R_k^{Tt} = \frac{c^2 \eta_k}{T^{5/2}} y_{t-1}^2 b_k''(a)$ for some value a between $\min\{\Delta y_t - \frac{c}{T} y_{t-1}, \Delta y_t\}$ and $\max\{\Delta y_t - \frac{c}{T} y_{t-1}, \Delta y_t\}$; and $\beta^{Tt} = \beta \left[\sum_k \left(\frac{\eta_k}{\sqrt{T}} b_k(\Delta y_t) - \frac{c\eta_k}{T^{3/2}} y_{t-1} \right. \right.$ $\beta(\cdot)$ is defined by

$$\log(1+x) = x - \frac{1}{2} x^2 [1 + \beta(x)], \qquad \lim_{x \to 0} \beta(x) = 0.$$

To complete the proof, it is sufficient to show that

$$\left| \sum_{t=2}^{T} R_k^{Tt} \right| = o_p(1), \quad \text{for each } k, \tag{A.1}$$

$$\max_{2 \le t \le T} \left| \beta^{Tt} \right| = o_p(1), \tag{A.2}$$

$$\sum_{t=2}^{T} \left[\sum_k \left(\frac{\eta_k}{\sqrt{T}} b_k(\Delta y_t) - \frac{c\eta_k}{T^{3/2}} y_{t-1} b_k'(\Delta y_t) + R_k^{Tt} \right) \right]^2 = \sum_k \eta_k^2 \mathcal{I}_{h_k h_k} + o_p(1), \tag{A.3}$$

$$\frac{c\eta_k}{T^{3/2}} \sum_{t=2}^{T} b_k'(\Delta y_t) y_{t-1} = \frac{c\eta_k}{T^{3/2}} \mathcal{I}_{f h_k} \sum_{t=2}^{T} y_{t-1} + o_p(1), \qquad \text{for each } k. \tag{A.4}$$

Equation (A.1). Since c and η_k are bounded sequences, (A.1) can be proved by

$$\left| \sum_{t=2}^{T} R_k^{Tt} \right| = \left| \frac{c^2 \eta_k}{T^{5/2}} \sum_{t=2}^{T} y_{t-1}^2 b_k''(a) \right| \le \left| \frac{c^2 \eta_k}{\sqrt{T}} M \right| \left| \frac{1}{T^2} \sum_{t=2}^{T} y_{t-1}^2 \right| = o_p(1) O_p(1) = o_p(1).$$

Equation (A.2). Since $\lim_{x \to 0} \beta(x) = 0$, it is sufficient to show

$$\max_{2 \le t \le T} \left| \frac{1}{T^{3/2}} b_k'(\Delta y_t) y_{t-1} \right| = o_p(1), \quad \max_{2 \le t \le T} \left| \frac{1}{\sqrt{T}} b_k(\Delta y_t) \right| = o_p(1), \quad \max_{2 \le t \le T} \left| R_k^{Tt} \right| = o_p(1).$$

The proof of the first one is given by

$$\max_{2 \le t \le T} \left| \frac{1}{T^{3/2}} b_k'(\Delta y_t) y_{t-1} \right| = \max_{2 \le t \le T} \left| \frac{1}{\sqrt{T}} y_{t-1} \right| \max_{2 \le t \le T} \left| \frac{1}{T} b_k'(\Delta y_t) \right| = O_p(1) o_p(1) = o_p(1),$$

the second one is obvious since by Assumption 2 function b_k is bounded, the proof of last one is given by

$$\max_{2 \le t \le T} \left| \frac{c^2 \eta_k}{T^{5/2}} y_{t-1}^2 b_k''(a) \right| \le \max_{2 \le t \le T} \left| \frac{c^2 \eta_k}{T^{5/2}} y_{t-1}^2 \right| M = o_p(1).$$

Equation (A.3). To prove (A.3), it is sufficient to show that

$$\sum_{t=2}^{T}\left[\sum_{k}\left(\frac{\eta_k}{\sqrt{T}}b_k(\Delta y_t)-\frac{c\eta_k}{T^{3/2}}y_{t-1}b_k'(\Delta y_t)+R_k^{Tt}\right)\right]^2=\sum_k\frac{\eta_k^2}{T}\sum_{t=2}^{T}b_k^2(\Delta y_t)+o_p(1), \tag{A.5}$$

and

$$\frac{\eta_k^2}{T}\sum_{t=2}^{T}b_k^2(\Delta y_t)=\eta_k^2\mathcal{I}_{b_kb_k}+o_p(1),\quad \forall k. \tag{A.6}$$

Equation (A.6) directly follows from the Law of Large Number (LLN). For equation (A.5), we are going to show all terms of quadratic part of $L_T^{f,h}(c,\eta)$ except $\sum_k\frac{\eta_k^2}{T}\sum_{t=2}^{T}b_k^2(\Delta y_t)$ are $o_p(1)$:

I.

$$\left|\sum_{t=2}^{T}(R_k^{Tt})^2\right|=\left|\frac{c^4\eta_k^2}{T^5}\sum_{t=2}^{T}y_{t-1}^4b_k''(a)^2\right|\le c^4\eta_k^2\left|\frac{1}{T^3}\sum_{t=2}^{T}y_{t-1}^4\right|\frac{M^2}{T^2}=O_p(1)o_p(1)=o_p(1),$$

II.

$$\left|\frac{1}{T^2}\sum_{t=2}^{T}b_k(\Delta y_t)b_s'(\Delta y_t)y_{t-1}\right|^2\le\left|\frac{1}{T^2}\sum_{t=2}^{T}b_k^2(\Delta y_t)b_s'^2(\Delta y_t)\right|\left|\frac{1}{T^2}\sum_{t=2}^{T}y_{t-1}^2\right|$$
$$\le\frac{M^2}{T}\left|\frac{1}{T}\sum_{t=2}^{T}b_s'^2(\Delta y_t)\right|\left|\frac{1}{T^2}\sum_{t=2}^{T}y_{t-1}^2\right|=o_p(1)O_p(1)O_p(1)=o_P(1),$$

here, $\left|\frac{1}{T}\sum_{t=2}^{T}b_s'^2(\Delta y_t)\right|=O_p(1)$ is because of $b_s'\in\mathcal{L}^2$.

III.

$$\left|\frac{1}{T^3}\sum_{t=2}^{T}b_k'(\Delta y_t)b_s'(\Delta y_t)y_{t-1}^2\right|^2\le\left|\frac{1}{T^3}\sum_{t=2}^{T}b_k'^2(\Delta y_t)b_s'^2(\Delta y_t)\right|\left|\frac{1}{T^3}\sum_{t=2}^{T}y_{t-1}^4\right|$$
$$\le\frac{1}{T}\left|\frac{1}{T}\sum_{t=2}^{T}b_k'^2(\Delta y_t)\right|\left|\frac{1}{T}\sum_{t=2}^{T}b_s'^2(\Delta y_t)\right|\left|\frac{1}{T^3}\sum_{t=2}^{T}y_{t-1}^4\right|=o_p(1)O_p(1)O_p(1)O_p(1)=o_p(1)$$

IV.

$$\left|\frac{1}{\sqrt{T}}\sum_{t=2}^{T}b_k(\Delta y_t)R_s^{Tt}\right|^2\le\left|\frac{1}{T}\sum_{t=2}^{T}b_k^2(\Delta y_t)\right|\left|\sum_{t=2}^{T}(R_s^{Tt})^2\right|\le M^2\left|\sum_{t=2}^{T}(R_s^{Tt})^2\right|=o_p(1),$$

V.

$$\left|\frac{1}{T^{3/2}}\sum_{t=2}^{T}b_k'(\Delta y_t)y_{t-1}R_s^{Tt}\right|^2\le\left|\frac{1}{T}\sum_{t=2}^{T}b_k'(\Delta y_t)^2\right|\left|\frac{1}{T}\sum_{t=2}^{T}y_{t-1}^2\right|\left|\frac{1}{T}\sum_{t=2}^{T}(R_s^{Tt})^2\right|=O_p(1)O_p(1)o_p(1)=o_p(1),$$

VI.

$$\left|\sum_{t=2}^{T}R_k^{Tt}R_s^{Tt}\right|\le\frac{1}{2}\left|\sum_{t=2}^{T}(R_k^{Tt})^2\right|+\frac{1}{2}\left|\sum_{t=2}^{T}(R_k^{Tt})^2\right|=o_p(1),$$

VII.

$$\left|\frac{1}{T}\sum_{t=2}^{T}b_k(\Delta y_t)b_s(\Delta y_t)\right|=o_p(1).$$

Equation (A.4.). Now prove $\sum_k\frac{c\eta_k}{T^{3/2}}\sum_{t=2}^{T}b_k'(\Delta y_t)y_{t-1}=\sum_k\frac{c\eta_k}{T^{3/2}}\mathcal{I}_{fh_k}\sum_{t=2}^{T}y_{t-1}+o_p(1)$. Firstly, $\frac{1}{T^{3/2}}b_k'(\Delta y_t)y_{t-1}$ is square-integrable and adapted to the filtration $(\mathbb{F}_{T,t})_{0\le t\le T}$, and the following condition holds:

$$\frac{1}{T^3}\sum_{t=2}^{T}E\left[b_k'(\Delta y_t)^2y_{t-1}^2\middle|\mathbb{F}_{T,t-1}\right]=\frac{1}{T^3}\sum_{t=2}^{T}y_{t-1}^2E\left[b_k'(\Delta y_t)^2\middle|\mathbb{F}_{T,t-1}\right]$$
$$=\frac{1}{T}\left(\frac{1}{T^2}\sum_{t=2}^{T}y_{t-1}^2\right)E\left[b_k'(\Delta y_t)^2\right]=o_p(1).$$

We then apply Hallin, van den Akker and Werker (2015, Lemma 2),

$$
\begin{aligned}
&\sum_k \frac{c\eta_k}{T^{3/2}} \sum_{t=2}^{T} b_k'(\Delta y_t) y_{t-1} \\
&= \sum_k \frac{c\eta_k}{T^{3/2}} \sum_{t=2}^{T} E\left[b_k'(\Delta y_t) y_{t-1} \middle| \mathbb{F}_{T,t-1} \right] + o_p(1) \\
&= \sum_k \frac{c\eta_k}{T^{3/2}} \sum_{t=2}^{T} y_{t-1} E\left[b_k'(\Delta y_t) \right] + o_p(1) \\
&= \sum_k \frac{c\eta_k}{T^{3/2}} \mathcal{I}_{c\eta_k} \sum_{t=2}^{T} y_{t-1} + o_p(1),
\end{aligned}
$$

the last equality comes from $E\left[b_k'(\Delta y_t) \middle| \mathbb{F}_{T,t-1} \right] = E[b_k'(\varepsilon)] = \int_{-\infty}^{\infty} b_k'(\varepsilon) f(\varepsilon) \mathrm{d}\varepsilon = b_k(\varepsilon) f(\varepsilon) \big|_{-\infty}^{+\infty} - \int_{-\infty}^{\infty} b_k(\varepsilon) f'(\varepsilon) \mathrm{d}\varepsilon = \int_{-\infty}^{\infty} b_k(\varepsilon)(-\frac{f'}{f}(\varepsilon)) f(\varepsilon) \mathrm{d}\varepsilon = E[b_k(\varepsilon)(-\frac{f'}{f}(\varepsilon))] = \mathcal{I}_{f b_k}$.

Q.E.D.

PROOF OF THEOREM 3.1: Let $\bar{G}$ be the group of translations $\bar{g}\eta = \eta + C$, where $C := (C_1, C_2, \cdots)'$ with $-\infty < C_1, C_2, \cdots < \infty$. Then it the same to show that statistic $\mathcal{M}$ is maximal invariant under $\bar{G}$. Let G be a group of transformations of the sample space, defined by

$$
gW(u) := \begin{pmatrix} W_\varepsilon(u) \\ W_1(u) + C_1 u \\ W_2(u) + C_2 u \\ \vdots \end{pmatrix}.
$$

It is easy to see the fact that G is a homomorphism of $\bar{G}$ (see, e.g., section 6.1 of Lehmann and Romano (2005)). Then it is same to show that $\mathcal{M}$ is a maximal invariant under G. Suppose $\mathcal{M} = \tilde{\mathcal{M}}$, which explicitly is

$$
\begin{aligned}
&W_\varepsilon(u) = \tilde{W}_\varepsilon(u), \\
&B_k(u) = \tilde{B}_k(u), \quad k \in \mathbb{N}.
\end{aligned}
$$

It implies that, for $C_k := (W_k(1) - \tilde{W}_k(1))$, $k \in \mathbb{N}$,

$$
\begin{aligned}
&W_\varepsilon(u) - \tilde{W}_\varepsilon(u) = 0, \\
&W_k(u) - \tilde{W}_k(u) = C_k u,
\end{aligned}
$$

which is $gW(t) = \tilde{W}(t)$. Thus by definition (see, e.g., section 6.2 of Lehmann and Romano (2005)), $\mathcal{M}$ is maximal invariant. *Q.E.D.*

TO BE COMPLETED