SEMIPARAMETRICALLY OPTIMAL HYBRID RANK TESTS FOR UNIT ROOTS

Bo Zhou[1], Ramon van den Akker[2] and Bas J.M. Werker[3]

We propose a new class of unit root tests that exploits invariance properties in the Locally Asymptotically Brownian Functional limit experiment associated to the standard unit root model. The invariance structures naturally suggest tests that are based on the ranks of the increments of the observations, their average, and an assumed reference density for the innovations. The tests are semiparametric in the sense that they are valid, i.e., have the correct (asymptotic) size, irrespective of the true innovation density. For correctly specified reference density, our test is point-optimal and nearly efficient. For arbitrary reference density, we establish a Chernoff-Savage type result, i.e., our test performs as well as commonly used tests under Gaussian innovations but has improved power under other, e.g., fat-tailed or skewed, innovation distributions. We also propose a simplified version of our test that exhibits the same properties, however the Chernoff-Savage type result is restricted to Gaussian reference densities and can only be demonstrated by simulations.

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

1. INTRODUCTION

The recent monographs of Patterson (2011, 2012) and Choi (2015) provide an overview of the literature on unit roots tests. This literature traces back to White (1958) and includes seminal papers as Dickey and Fuller (1979, 1981), Phillips (1987), Phillips and Perron (1988), and Elliott, Rothenberg, and Stock (1996). The present paper fits into the stream of literature that focuses on "optimal" testing for unit roots. Important early contributions here are Dufour and King (1991), Saikkonen and Luukkonen (1993), and Elliott, Rothenberg, and Stock (1996). The latter paper derives the asymptotic power envelope for unit root testing in models with Gaussian innovations. Rothenberg and Stock (1997) and Jansson (2008) consider subsequently the non-Gaussian case.

The present paper considers testing for unit roots in a semiparametric setting. Following earlier literature, we focus on a simple AR(1) model driven by i.i.d. in-

[1]B.Zhou@tilburguniversity.edu. Econometrics Group, Tilburg University, the Netherlands.
[2]R.vdnAkker@tilburguniversity.edu. Econometrics Group, Tilburg University, the Netherlands.
[3]B.J.M.Werker@tilburguniversity.edu. Econometrics and Finance Group, Tilburg University, the Netherlands.

novations whose distribution is considered a nuisance parameter. Apart from some smoothness and the existence of relevant moments, no assumptions are imposed on this distribution. From earlier work it is known that the unit root model leads to Locally Asymptotically Brownian Functional (LABF) limit experiments (in the Hájek-Le Cam sense). As a consequence, no uniformly most powerful test exists (even in case the innovation distribution would be known) – see also Elliott, Rothenberg, and Stock (1996). In the semiparametric case the limit experiment becomes even more difficult, precisely due to the infinite-dimensional nuisance parameter. Jansson (2008) derives the semiparametric power envelope by mimicking ideas that hold for Locally Asymptotically Normal (LAN) models. However, the proposed test needs a full nonparametric score function estimator which complicates its implementation. Our optimal test only requires a nonparametric estimation of a real-valued cross-information factor.

The main contribution of this manuscript is twofold. First, we provide a new derivation of the semiparametric asymptotic power envelope for unit root tests (Section 3). This derivation is build upon invariance structures embedded in the semiparametric unit root model. To be precise, we use a "structural" description of the LABF limit experiment (Section 3.2), obtained from Girsanov's theorem. This limit experiment corresponds to observing a multivariate Ornstein-Uhlenbeck process (on the time interval $[0, 1]$). The unknown innovation density in the semiparametric unit root model, takes the form of an unknown drift in the limit experiment. Within this limit experiment, we subsequently (Section 3.3) derive the maximal invariant, i.e., a reduction of the data which is invariant with respect to the nuisance parameters (that is, the unknown drift in the limiting Ornstein-Uhlenbeck experiment). It turns out that this maximal invariant takes a rather simple form (all components, but one, of the multivariate process have to be replaced by their associated bridges). The power envelope for invariant tests in the limit experiment then follows from the Neyman-Pearson lemma. An application of the Asymptotic Representation Theorem subsequently yields the local asymptotic power envelope (Theorem 3.2). We note that our analysis of invariance structures in the LABF experiment is also of independent interest and could, for example, be exploited in the analysis of optimal inference for cointegration or predictive regression models. Moreover, it also gives an alternative interpretation of the test proposed in Elliott, Rothenberg, and Stock (1996) as it is also based on an invariant, though not the maximal one.

As a second contribution, we provide a new class of easy-to-implement unit root tests that are semiparametrically optimal in the sense that their asymptotic power curve is tangent to the semiparametric power envelope (Section 4.1). The form of the maximal invariant developed before suggests how to construct such tests based on the ranks of the increments of the observations, the average of these increments, and an assumed reference density. These tests are semiparametric in the sense that the reference density need not equal the true innovation density, while they still provide the correct asymptotic size. This reference density is not restricted to be Gaussian, which it generally is in more classical QMLE results. When the reference density is correctly specified, 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 a fixed alternative that yields a "nearly efficient" test, i.e., one for which the asymptotic local power function is uniformly close to the semiparametric power envelope. Our tests, despite the absence of a LAN structure, satisfy a Chernoff and Savage (1958) type result (Corollary 4.1): with any reference density our test outperforms, at any true density, its classical counterpart which in this case, is the Elliott, Rothenberg, and Stock (1996) test. We provide, in Section 4.2, an even simpler alternative class of tests. Both classes of tests coincide for correctly specified reference density and, thus, share the same optimality properties. In case of misspecified reference density, the alternative class still seems to enjoy the Chernoff-Savage type property, though only for Gaussian reference density. This is in line with with the traditional Chernoff-Savage results for Locally Asymptotically Normal models.

The remainder of this paper is organized as follows. Section 2 introduces the model assumptions and some notation. Next, Section 3 contains the analysis of the limit experiment. In particular we study invariance properties in the limit experiment leading to our new derivation of the semiparametric power envelope. The class of hybrid rank tests we propose is introduced in Section 4. Section 5 provides 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 $Y_1, \dots, Y_T$ generated from the classical component specification

$$Y_t = \mu + X_t, \quad t \in \mathbb{N}, \tag{1}$$

$$X_t = \rho X_{t-1} + \varepsilon_t, \quad t \in \mathbb{N}, \tag{2}$$

where $X_0 = 0$ and the innovations $\{\varepsilon_t\}$ form an i.i.d. sequence with density f. We impose the following assumptions on this innovation density.

Assumption 1

(a) The density f is absolutely continuous 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) $\mathrm{E}_f[\varepsilon_t] = \int ef(e)\mathrm{d}e = 0$ and $\sigma_f^2 = \mathrm{Var}_f[\varepsilon_t] < \infty$.

(c) The standardized Fisher-information for location,

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

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

(d) The density f is strictly positive, i.e., $f > 0$.

Let $\mathcal{F}$ denote the set of densities satisfying Assumption 1.

The imposed smoothness assumptions (a) on f are mild and standard. The finite variance assumption (b) 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 zero mean assumption in (b) excludes a deterministic trend in the model. Such a trend leads to an entirely different asymptotic analysis, see Hallin, Van den Akker, and Werker (2011). The Fisher information in (c) is standardized by the variance σ_f^2 so that it becomes scale invariant. The strict positivity of the density f in (d) is mainly made for notational convenience. The assumption on the initial condition, $X_0 = 0$, is less innocent then it may appear. Indeed, it is 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 tests, with optimality features, for the semiparametric unit root hypothesis

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

[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 (Section 5) 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).

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), Chan and Wei (1988) 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.,

$$\rho = \rho_h^{(T)} = 1 + \frac{h}{T}, \tag{3}$$

and we can rewrite our hypothesis of interest as

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

In the following section, we derive the (asymptotic) power envelope of tests that are (asymptotically) invariant with respect to the nuisance parameters μ and f. Section 4 is subsequently devoted to tests, depending on a reference density g that can be freely chosen, that are point optimal with respect to this power envelope and proves the Chernoff-Savage result.

3. THE POWER ENVELOPE FOR INVARIANT TESTS

This section first introduces some notations and preliminaries (Section 3.1). Next, we will derive the limit experiment (in Hájek-Le Cam sense) corresponding to the component unit root model (1)-(2) and provide a “structural” representation of this limit experiment (Section 3.2). In Section 3.3 we discuss, exploiting this structural representation, a natural invariance restriction, to be imposed on tests for the unit root hypothesis with respect to the infinite-dimensional nuisance parameter associated to the innovation density. We derive the maximal invariant and obtain from this the power envelope for invariant tests in the limit experiment.

3.1. *Preliminaries*

We first discuss a convenient parametrization of perturbations to the innovation density which we use to deal with the semiparametric nature of the testing problem. These perturbations follow the standard approach of local alternatives in (semiparametric) models commonly used in experiments that are Locally Asymptotically Normal

(LAN). We will see that, with respect to the innovation density f alone, the model is actually LAN; compare also Remark 3.1 below. Moreover, we introduce some partial sum processes that we need in the sequel, as well as their Brownian limits.

Perturbations to the innovation density

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

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

where $\mathrm{L}_2^{f}(\mathbb{R},\mathcal{B})$ denotes, the space of Lebesgue-measurable functions $b : \mathbb{R} \to \mathbb{R}$ satisfying $\int b^2(e)f(e)\mathrm{d}e < \infty$. Because of the separability, there exists a countable orthonormal basis b_k, $k \in \mathbb{N}$, of $\mathrm{L}_2^{0,f}$. This basis can be chosen 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 $b_k \in L_2^{0,f}$ with $\mathrm{Var}_f\, b_k(\varepsilon) = 1$, $\eta \in c_{00}$ we now introduce the following perturbation to the density f:

$$f_\eta^{(T)}(e) = f(e)\left(1 + \frac{1}{\sqrt{T}} \sum_{k=1}^{\infty} \eta_k b_k(e)\right), \quad e \in \mathbb{R}. \tag{4}$$

The rate $T^{-1/2}$ is already indicative of the standard LAN behavior of the nuisance parameter f as will formally follow from Proposition 3.2 below. The following proposition shows that these perturbations are valid in the sense that they satisfy the conditions on the innovation density that we imposed throughout on the model (Assumption 1).

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

REMARK 3.1 In semiparametric statistics one typically parametrizes perturbations (paths in semiparametric parlor) to a density by a so-called "non-parametric" score $h \in \mathrm{L}_2^{0,f}$, i.e., a perturbation takes the form $f(e)k(T^{-1/2}h(e))) \approx f(e)(1+T^{-1/2}h(e))$ for a suitable function k; see, for example, Bickel et al. (1998) for details. 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 instead restrict to c_{00} which ensures (4) to be a density (for large T). For our purposes this restriction will be without cost. Intuitively, this is 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 extends to the closure).

Partial sum processes

To describe the limit experiment in Section 3.2, we introduce some partial sum processes and their limits. These results are fairly classical but, for completeness, precise statements are organized in Lemma A.1.

As usual, Δ denotes differencing, i.e., $\Delta Y_t = Y_t - Y_{t-1}$. Define, for $s \in [0,1]$,

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

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

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

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. Also note that the sums start at $t = 2$, so the partial sum processes are (maximally) invariant with respect to the intercept μ. Using Assumption 1 we find, under the null hypothesis, weak convergence[2] of $W_\varepsilon^{(T)}$, $W_{\phi_f}^{(T)}$, and $W_{b_k}^{(T)}$ to Brownian motions that we denote by W_ε, W_{ϕ_f}, and W_{b_k}, respectively. These limiting Brownian motions W_ε, W_{ϕ_f}, and W_{b_k} are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P}_{0,0})$. Let us already mention that we will introduce a collection of probability measures $\mathbb{P}_{h,\eta}$ representing the limit experiment, in Section 3.2. We use the notational convention that probability measures related to the limit experiment (i.e.,

[2]All weak convergences in this paper are in product spaces of $D[0,1]$ with the uniform topology.

to the Brownian motions) are denoted by $\mathbb{P}$, while probability measures related to the finite-sample unit root model will be denoted by $\mathrm{P}^{(T)}$.

As ε and $b_k(\varepsilon)$ are orthogonal for each k, we find that W_ε and W_{b_k}, $k \in \mathbb{N}$, are all mutually independent. Moreover,

$$\mathrm{Var}_{0,0}[W_\varepsilon(1)] = 1 \text{ and } \mathrm{Var}_{0,0}[W_{b_k}(1)] = 1.$$

As $\phi_f(\varepsilon)$ is the score of the location model, it is well known (see, for example, Bickel et al. (1998)) that we have (under Assumption 1) $\mathrm{E}_f[\phi_f(\varepsilon)] = 0$ and $\mathrm{E}_f[\varepsilon\phi_f(\varepsilon)] = 1$. Consequently, again because ε and $b_k(\varepsilon)$ are orthogonal for each k, we can decompose $\sigma_f\phi_f(\varepsilon) = \sigma_f^{-1}\varepsilon + \sum_{k=1}^\infty J_{f,k}b_k(\varepsilon)$, with coefficients $J_{f,k} = \sigma_f \mathrm{E}_f[b_k(\varepsilon)\phi_f(\varepsilon)]$. This establishes, for $f \in \mathcal{F}$,

$$W_{\phi_f} = W_\varepsilon + \sum_{k=1}^\infty J_{f,k} W_{b_k}. \tag{5}$$

Moreover, we have

$$\mathrm{Cov}_{0,0}(W_{\phi_f}(1), W_\varepsilon(1)) = 1, \quad \mathrm{Cov}_{0,0}(W_{\phi_f}(1), W_{b_k}(1)) = J_{f,k},\ \ k \in \mathbb{N}, \tag{6}$$

and

$$\mathrm{Var}_{0,0}[W_{\phi_f}(1)] = J_f = 1 + \sum_{k=1}^\infty J_{f,k}^2. \tag{7}$$

We remark that integrals like $\int_0^1 W_\varepsilon^{(T)}(s-)\mathrm{d}W_{\phi_f}^{(T)}(s)$ can be shown to converge weakly to the associated stochastic integral with the limiting Brownian motions, i.e., to $\int_0^1 W_\varepsilon(s)\mathrm{d}W_{\phi_f}(s)$. Weak convergence of integrals like $\int_0^1 (W_\varepsilon^{(T)}(s-))^2\mathrm{d}s$ follows from an application of the continuous mapping theorem. Again, details are provided in Appendix A.

3.2. *A structural representation of the limit experiment*

The results in the previous section are needed to study the asymptotic behavior of log-likelihood ratios. These in turn determine the limit experiment, which we use to study asymptotically optimal procedures invariant with respect to the nuisance parameters f and μ. Thus, fix $f \in \mathcal{F}$ and $\mu \in \mathbb{R}$. Let, for $h \in \mathbb{R}$ and $\eta \in c_{00}$, $\mathrm{P}^{(T)}_{h,\eta;\mu,f}$ denote the law of $Y_1, \ldots, Y_T$ under (1)-(2) with autoregression parameter ρ given by (3) and innovation density (4). The following proposition shows that the semiparametric unit root model is of the Locally Asymptotically Brownian Functional (LABF) type introduced in Jeganathan (1995).

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

(i) Then we have, under $\mathrm{P}^{(T)}_{0,0;\mu,f}$,

$$\log \frac{\mathrm{dP}^{(T)}_{h,\eta;\mu,f}}{\mathrm{dP}^{(T)}_{0,0;\mu,f}} = \log \frac{f^{(T)}_{\eta}\left(Y_1-\mu\right)}{f(Y_1-\mu)} + \sum_{t=2}^{T} \log \frac{f^{(T)}_{\eta}\left(\Delta Y_t - \frac{h}{T}(Y_{t-1}-\mu)\right)}{f(\Delta Y_t)} \tag{8}$$

$$= h\Delta^{(T)}_{f} + \sum_{k=1}^{\infty} \eta_k \Delta^{(T)}_{b_k} - \frac{1}{2}\mathcal{I}^{(T)}_{f}(h,\eta) + o_P(1),$$

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

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

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

and

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

$$= h^2 J_f \frac{1}{T^2}\sum_{t=2}^{T}\frac{(Y_{t-1}-Y_1)^2}{\sigma_f^2} + \|\eta\|_2^2 + 2h\frac{1}{T^{3/2}}\sum_{t=2}^{T}\frac{(Y_{t-1}-Y_1)}{\sigma_f}\sum_{k=1}^{\infty}\eta_k J_{f,k}.$$

(ii) Moreover, with $\Delta_f = \int_0^1 W_{\varepsilon}(s)\mathrm{d}W_{\phi_f}(s)$ and $\Delta_{b_k} = W_{b_k}(1)$, $k \in \mathbb{N}$, we have, still under $\mathrm{P}^{(T)}_{0,0;\mu,f}$ and as $T \to \infty$,

$$\frac{\mathrm{dP}^{(T)}_{h,\eta;\mu,f}}{\mathrm{dP}^{(T)}_{0,0;\mu,f}} \Rightarrow \exp\left(h\Delta_f + \sum_{k=1}^{\infty}\eta_k\Delta_{b_k} - \frac{1}{2}\mathcal{I}_f(h,\eta)\right), \tag{9}$$

where

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

(iii) For all $h \in \mathbb{R}$ and $\eta \in c_{00}$ the right-hand side of (9) has unit expectation under $\mathbb{P}_{0,0}$.

The proof of (i) follows by an application of Proposition 1 in Hallin, Van den Akker, and Werker (2015) which provides generally applicable 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) discussed above. Finally, Part (iii) follows by verifying the Novikov condition. All proofs are organized in Appendix B.

Part (iii) of the proposition implies that we can introduce, for $h \in \mathbb{R}$ and $\eta \in c_{00}$, new probability measures $\mathbb{P}_{h,\eta}$ on the measurable space $(\Omega, \mathcal{F})$ (on which the processes W_ε, W_{ϕ_f}, and W_{b_k} were defined) by their Radon-Nikodym derivatives with respect to $\mathbb{P}_{0,0}$:

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

Proposition 3.2 then implies that the sequence of unit root experiments (each $T \in \mathbb{N}$ yields an experiment) weakly converges (in the Hájek-Le Cam sense) to the experiment described by the probability measures $\mathbb{P}_{h,\eta}$. Formally, we define the sequence of experiments of interest by

$$\mathcal{E}^{(T)}(\mu, f) = \left(\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T), (\mathrm{P}^{(T)}_{h,\eta;\mu,f} \,|\, h \in \mathbb{R},\, \eta \in c_{00})\right),\ T \in \mathbb{N}, \tag{10}$$

and the limit experiment by, with $\mathcal{B}_C$ the Borel σ-field on $C[0,1]$,

$$\mathcal{E}(f) = \Big(C[0,1] \times C^{\mathbb{N}}[0,1], \mathcal{B}_{\mathcal{C}} \otimes (\otimes_{k=1}^{\infty} \mathcal{B}_{\mathcal{C}}), (\mathbb{P}_{h,\eta} \,|\, h \in \mathbb{R},\, \eta \in c_{00})\Big). \tag{11}$$

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 $\mathcal{E}(f)$.

The Asymptotic Representation Theorem (see, e.g., Chapter 9 in Van der Vaart (2000)) implies that for any statistic A_T which converges in distribution to the law $L_{h,\eta}$, under $\mathrm{P}^{(T)}_{h,\eta;\mu,f}$, there exists a (randomized) statistic A, defined on $\mathcal{E}(f)$, such that the law of A under $\mathbb{P}_{h,\eta}$ is given by $L_{h,\eta}$. This allows us to study (asymptotically) optimal inference: 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 this bound, it follows that the bound is sharp and the statistic is called (asymptotically) optimal. This is precisely what we do: Section 3.3 establishes the bound and in Section 4 we introduce a statistic attaining it.

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

PROPOSITION 3.3 Let $f \in \mathcal{F}$, $\eta \in c_{00}$, and $h \in \mathbb{R}$. Then the processes Z_ε and Z_{b_k}, $k \in \mathbb{N}$, defined by the starting values $Z_\varepsilon(0) = Z_{b_k}(0) = 0$ and the stochastic differential equations, for $s \in [0,1]$,

$$\begin{aligned} \mathrm{d}Z_\varepsilon(s) &= \mathrm{d}W_\varepsilon(s) - hW_\varepsilon(s)\mathrm{d}s, \\ \mathrm{d}Z_{b_k}(s) &= \mathrm{d}W_{b_k}(s) - hJ_{f,k}W_\varepsilon(s)\mathrm{d}s - \eta_k \mathrm{d}s, \quad k \in \mathbb{N}, \end{aligned}$$

are Brownian motions under $\mathbb{P}_{h,\eta}$. Their joint law is that of $(W_\varepsilon, (W_{b_k})_{k\in\mathbb{N}})$ under $\mathbb{P}_{0,0}$.

3.3. *The limit experiment: invariance and power envelope*

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. In Section 3.4 we then exploit the Asymptotic Representation Theorem to translate these results to obtain (asymptotically) optimal invariant test in the sequence of unit root models.

Consider the testing problem for the limit experiment $\mathcal{E}(f)$. We thus observe the processes W_ε and W_{b_k}, $k \in \mathbb{N}$, (continuously) 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

$$\text{(12)} \qquad \mathrm{H}_0 : \, h = 0, \, (\eta \in c_{00}) \text{ versus } \mathrm{H}_a : \, h < 0, \, (\eta \in c_{00}).$$

We focus on test statistics that are invariant with respect to the value of the nuisance parameter η, i.e., these test statistics take the same value irrespective of the value of η. We now formalize this invariance structure.

Introduce, for $\eta \in c_{00}$, the transformation $g_\eta = (g_{\eta_k})_{k\in\mathbb{N}} : C^{\mathbb{N}}[0,1] \to C^{\mathbb{N}}[0,1]$ defined by, for $W \in C[0,1]$,

$$\text{(13)} \qquad g_{\eta_k} : \, [g_{\eta_k}(W)]\,(s) = W(s) - \eta_k s, \quad s \in [0,1],$$

i.e., g_{η_k} adds a drift $s \mapsto -\eta_k s$ to W. Proposition 3.3 implies that the law of $(W_\varepsilon, (g_{\eta_k}(W_{b_k}))_{k\in\mathbb{N}})$ under $\mathbb{P}_{h,0}$ is the same as the law of $(W_\varepsilon, (W_{b_k})_{k\in\mathbb{N}})$ under $\mathbb{P}_{h,\eta}$. Hence our testing problem (12) is invariant with respect to the transformations g_η. Therefore, following the invariance principle, it is natural to restrict attention to test statistics that are invariant with respect to these transformations as well, i.e., test statistics t that satisfy

$$\text{(14)} \qquad t(W_\varepsilon, (g_{\eta_k}(W_{b_k}))_{k\in\mathbb{N}}) = t(W_\varepsilon, (W_{b_k})_{k\in\mathbb{N}}) \text{ for all } g_\eta,\ \eta \in c_{00}.$$

Given a process W let us define the associated *bridge process* by $B^W(s) = W(s) - sW(1)$. Now note that we have, for all $s \in [0,1]$ and $k \in \mathbb{N}$,

$$\begin{aligned} B^{g_{\eta_k}(W)}(s) &= [g_{\eta_k}(W)](s) - s[g_{\eta_k}(W)](1) \\ &= W(s) - s\eta_k - s(W(1) - 1 \times \eta_k) \\ &= W(s) - sW(1) \\ &= B^W(s), \end{aligned}$$

i.e., taking the bridge of a process ensures invariance with respect to adding drifts to that process. Define the mapping M by $M(W_\varepsilon, (W_{b_k})_{k\in\mathbb{N}}) := (W_\varepsilon, (B_{b_k})_{k\in\mathbb{N}})$, with $B_{b_k} = B^{W_{b_k}}$. It follows that statistics that are measurable with respect to the σ-field,

$$\mathcal{M} = \sigma\left(M(W_\varepsilon, (W_{b_k})_{k\in\mathbb{N}})\right) = \sigma(W_\varepsilon, (B_{b_k})_{k\in\mathbb{N}}), \tag{15}$$

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 The σ-field $\mathcal{M}$ in (15) is maximally invariant for the group of transformations g_η, $\eta \in c_{00}$, in the experiment $\mathcal{E}(f)$.

The above theorem implies that invariant inference must 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. To be precise, consider the likelihood ratios restricted to $\mathcal{M}$, 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 precisely because of the invariance. To calculate this conditional expectation we first introduce $B_{\phi_f} = B^{W_{\phi_f}}$, i.e., the bridge process associated to W_{ϕ_f} defined in (5). Now we can decompose $\Delta_f = \int_0^1 W_\varepsilon(s)\mathrm{d}W_{\phi_f}(s) = I + II$ with

$$\begin{aligned} I &= \int_0^1 W_\varepsilon(s)\mathrm{d}B_{\phi_f}(s) + W_\varepsilon(1)\int_0^1 W_\varepsilon(s)\mathrm{d}s, \\ II &= \left(\sum_{k=1}^\infty J_{f,k} W_{b_k}(1)\right)\int_0^1 W_\varepsilon(s)\mathrm{d}s. \end{aligned}$$

Note that part I is $\mathcal{M}$-measurable. Under $\mathbb{P}_{0,0}$ the random variables $W_{b_k}(1)$, $k \in \mathbb{N}$, are independent to W_ε and B_{b_k}, $k \in \mathbb{N}$. Indeed, the independence to W_ε holds by construction and the independence to B_{b_k} is a well-known, and easy to verify, property of Brownian bridges. We thus obtain, since $\mathcal{I}_f(h,\eta)$ is $\mathcal{M}$-measurable as well,

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

This yields

$$\frac{\mathrm{d}\mathbb{P}_h^{\mathcal{M}}}{\mathrm{d}\mathbb{P}_0^{\mathcal{M}}} = \exp\left(h\Delta_f^* - \frac{1}{2}h^2\mathcal{I}_f^*\right) \text{ with}$$

$$\Delta_f^* = \int_0^1 W_\varepsilon(s)\mathrm{d}B_{\phi_f}(s) + W_\varepsilon(1)\int_0^1 W_\varepsilon(s)\mathrm{d}s, \tag{16}$$

$$\begin{aligned}\mathcal{I}_f^* &= J_f\int_0^1 W_\varepsilon^2(s)\mathrm{d}s - \left(\int_0^1 W_\varepsilon(s)\mathrm{d}s\right)^2\sum_{k=1}^{\infty}J_{f,k}^2 \\ &= J_f\int_0^1 W_\varepsilon^2(s)\mathrm{d}s - \left(\int_0^1 W_\varepsilon(s)\mathrm{d}s\right)^2(J_f - 1),\end{aligned} \tag{17}$$

where the last equality follows from (7). Note that this likelihood ratio is indeed invariant with respect to η and one can also verify directly that $\mathcal{I}_f^*$ is the quadratic counterpart of Δ_f^*.

We can now formalize the notion of point-optimal invariant tests in the limit experiment. To that end, 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 $c(h,f;\alpha)$. Define the size-α test $\phi_{f,\alpha}^*(\bar{h}) = I\left\{\mathrm{d}\mathbb{P}_{\bar{h}}^{\mathcal{M}}/\mathrm{d}\mathbb{P}_0^{\mathcal{M}} \geq c(\bar{h},f;\alpha)\right\}$, for a fixed value of $\bar{h} < 0$. Note that this is an oracle test depending on f, and the feasible version will be provided in Section 4. The power function of this oracle test is given by

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

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

$$\pi(\bar{h}) \le \pi^*_{f,\alpha}(\bar{h};\bar{h}).$$

The test $\phi^*_{f,\alpha}(\bar{h})$ thus is point optimal, i.e., its power function is tangent to the power envelope[3] $h \mapsto \pi^*_{f,\alpha}(h;h)$ at $h = \bar{h}$.

REMARK 3.2 The semiparametric power envelope $\pi^*_{f,\alpha}$ of the limit experiment in Proposition 3.3 is scale invariant, i.e., invariant with respect to the value of $\sigma_f > 0$. This is easily seen from the fact that W_ε, W_{ϕ_f} and J_f are all scale invariant.

REMARK 3.3 The notion of invariance in the limit experiment leads to another interpretation of the Elliott, Rothenberg, and Stock (1996) statistic. Note that $\mathcal{M}_\varepsilon = \sigma\left(W_\varepsilon(s);\, s \in [0,1]\right)$ is also invariant, though not maximally so. We now calculate the likelihood ratio conditional on observing $\mathcal{M}_\varepsilon$ only, by further projecting the likelihood ratio on $\mathcal{M}_\varepsilon$:

$$\begin{aligned}
\frac{\mathrm{d}\mathbb{P}_h^{\mathcal{M}_\varepsilon}}{\mathrm{d}\mathbb{P}_0^{\mathcal{M}_\varepsilon}} &= \mathbb{E}_0\left[\frac{\mathrm{d}\mathbb{P}_h^{\mathcal{M}}}{\mathrm{d}\mathbb{P}_0^{\mathcal{M}}} \mid \mathcal{M}_\varepsilon\right] \\
&= \exp\left(h\int_0^1 W_\varepsilon(s)\mathrm{d}B_\varepsilon(s) + hW_\varepsilon(1)\int_0^1 W_\varepsilon(s)\mathrm{d}s - \frac{1}{2}h^2\mathcal{I}_f^*\right) \\
&\quad \times \mathbb{E}_0\left[\exp\left(h\int_0^1 W_\varepsilon(s)\mathrm{d}B_b(s)\right) \mid \mathcal{M}_\varepsilon\right] \\
&= \exp\left(h\int_0^1 W_\varepsilon(s)\mathrm{d}W_\varepsilon(s) - \frac{1}{2}h^2\mathcal{I}_f^*\right) \\
&\quad \times \exp\left(\frac{1}{2}h^2\left[\int_0^1 W_\varepsilon^2(s)\mathrm{d}s - \left(\int_0^1 W_\varepsilon(s)\mathrm{d}s\right)^2\right](J_f - 1)\right) \\
&= \exp\left(h\int_0^1 W_\varepsilon(s)\mathrm{d}W_\varepsilon(s) - \frac{1}{2}h^2\int_0^1 W_\varepsilon^2(s)\mathrm{d}s\right),
\end{aligned}$$

where $W_b(s) = \sum_{k=1}^\infty J_{f,k}W_{b_k}(s)$ and $B_b(s) = B^{W_b}(s)$ for notational simplicity. As a result, the Elliott, Rothenberg, and Stock (1996) test statistic equals the likelihood ratio statistic from using the (non-maximal) invariant $\mathcal{M}_\varepsilon$. This is an alternative explanation for the improved power of our tests. Moreover, for Gaussian f, we have

[3] Here and later in this section, the early usage of the concept "power envelope" (instead of "upper bound") is due to the fact that it is shown to be pointwise attainable in Section 4.

$\mathcal{M}_\varepsilon = \mathcal{M}$ and obtain point-optimality of the Elliott, Rothenberg, and Stock (1996) test for this case.[4]

3.4. *The asymptotic power envelope for asymptotically invariant 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 $\psi^{(T)}$ is said to be asymptotically invariant if the distribution of $\psi^{(T)}$ weakly converges, under $\mathrm{P}^{(T)}_{h,\eta;\mu,f}$ for all $h \leq 0$ and $\eta \in c_{00}$, to the distribution of an invariant test in the limit experiment $\mathcal{E}(f)$, under $\mathbb{P}_{h,\eta}$.

The Asymptotic Representation Theorem (see, e.g., Chapter 9 in Van der Vaart (2000)) now yields the following main result on the asymptotic power envelope.

THEOREM 3.2 Let $f \in \mathcal{F}$, $\mu \in \mathbb{R}$, and $\alpha \in (0,1)$. Let $\phi_T(Y_1, ..., Y_T)$, $T \in \mathbb{N}$, be an asymptotically 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

$$\limsup_{T\to\infty} \pi_T(h,\eta) \leq \pi^*_{f,\alpha}(h;h), \quad \eta \in c_{00} \text{ and } h < 0. \qquad \square$$

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, thereby, demonstrates that the bound indeed constitutes the asymptotic power envelope for invariant unit root tests. We also provide a Chernoff-Savage type result for this class of tests.

4. A CLASS OF SEMIPARAMETRICALLY OPTIMAL HYBRID RANK TESTS

The appearance of the bridge process B_{ϕ_f} in the "efficient central sequence" Δ^*_f naturally suggests the (partial) use of ranks in the construction of feasible test statistics. Indeed, we can construct an empirical analogue of B_{ϕ_f} by considering a partial-sum process which only depends on the observations via the ranks R_t of ΔY_t amongst

[4]Similarly, one could try to derive the statistic resulting from using $\mathcal{M}_B = \sigma\left(B_{b_k}(s);\, s \in [0,1]\right)$ as an invariant. However, that does not seem to lead to an insightful result.

$\Delta Y_2, \ldots, \Delta Y_T$. We allow for the use of a *reference density* g that may or may not be equal to the true underlying innovation density f. Our findings compare to Quasi-ML methods: if the true innovation density happens to be the same as the selected reference density the inference procedure is point-optimal. At the same time, the procedure is valid, i.e., has proper asymptotic size, even in case the true innovation density does not coincide with the reference density. Note that these results also hold in case the reference density is non-Gaussian, while Quasi-ML results are generally restricted to Gaussian reference densities.

We need the following mild assumption on the reference density.

ASSUMPTION 2 The density $g \in \mathcal{F}$, with finite variance σ_g^2, satisfies

$$\lim_{T\to\infty} \frac{\sigma_g^2}{T} \sum_{i=1}^{T} \phi_g^2 \left(G^{-1} \left(\frac{i}{T+1} \right) \right) = J_g,$$

with location score function $\phi_g(e) := -(g'/g)(e)$, where J_g is the standardized Fisher information[5] for location of g.

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

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

$$B_{\phi_g}^{(T)}(s) = \frac{1}{\sqrt{T}} \sum_{t=2}^{[sT]} \sigma_g \phi_g \left(G^{-1} \left(\frac{R_t}{T+1} \right) \right), \tag{18}$$

where R_t denotes the rank of ΔY_t, $t = 2, \ldots, T$. Then, under $P_{0,0;\mu,f}^{(T)}$ and as $T \to \infty$, we have[6]

$$\begin{bmatrix} W_\varepsilon^{(T)} \\ W_{\phi_f}^{(T)} \\ B_{\phi_g}^{(T)} \end{bmatrix} \Rightarrow \begin{bmatrix} W_\varepsilon \\ W_{\phi_f} \\ B_{\phi_g} \end{bmatrix} \quad \text{and} \tag{19}$$

$$\int_0^1 W_\varepsilon^{(T)}(s-) \mathrm{d}B_{\phi_g}^{(T)}(s) \Rightarrow \int_0^1 W_\varepsilon(s) \mathrm{d}B_{\phi_g}(s). \tag{20}$$

[5]Similarly as the standardized Fisher information J_f of f, Fisher information J_g of g is also standardized, by the variance σ_g^2, so that it is scale invariant.

[6]Equation (20) holds because of the Theorem 2.1 in Hansen (1992).

Here, B_{ϕ_g} is the associated Brownian bridge of W_{ϕ_g}, which itself is a Brownian motion defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P}_{0,0})$ as W_ε and W_{ϕ_f}, with covariance matrix

$$\text{Cov}_{0,0}\begin{pmatrix} W_\varepsilon(1) \\ W_{\phi_f}(1) \\ W_{\phi_g}(1) \end{pmatrix} = \begin{pmatrix} 1 & 1 & \sigma_{\varepsilon\phi_g} \\ & J_f & J_{fg} \\ & & J_g \end{pmatrix}, \tag{21}$$

where

$$\sigma_{\varepsilon\phi_g} = \sigma_f^{-1}\sigma_g \int_0^1 F^{-1}(u)\phi_g(G^{-1}(u))\mathrm{d}u, \tag{22}$$

$$J_{fg} = \sigma_f\sigma_g \int_0^1 \phi_f(F^{-1}(u))\phi_g(G^{-1}(u))\mathrm{d}u. \tag{23}$$

4.1. *The hybird rank tests based on a reference density*

The weak convergence in Lemma 4.1 indicates that constructing the partial sum processes as described above (with rank statistics and a reference density g for $B_{\phi_g}^{(T)}$), as T approaches infinity, corresponds to observing the σ-field $\mathcal{M}_g = \sigma(W_\varepsilon(s), B_{\phi_g}(s); s \in [0,1])$ in the limit. Clearly, $\mathcal{M}_g \subseteq \mathcal{M}$[7] so that $\mathcal{M}_g$ is invariant for the group of transformations g_η. When $g = f$, $\mathcal{M}_g = \mathcal{M}$ so that it is maximally invariant, which means that we capture all available information about h.

The following proposition establishes the likelihood ratio restricted to the information $\mathcal{M}_g$.

Proposition 4.1 Define the standard Brownian motion $W_\perp$, under $\mathbb{P}_{0,0}$, via the decomposition

$$\frac{W_{\phi_g}}{\sigma_{\varepsilon\phi_g}} = W_\varepsilon + \sqrt{\frac{J_g}{\sigma^2_{\varepsilon\phi_g}} - 1}W_\perp, \tag{24}$$

and denote associated Brownian bridge by $B_\perp$. The likelihood ratio $\mathrm{d}\mathbb{P}_h/\mathrm{d}\mathbb{P}_0$ restricted to the outcome space $\mathcal{M}_g$ is given by

$$\frac{\mathrm{d}\mathbb{P}_h^{\mathcal{M}_g}}{\mathrm{d}\mathbb{P}_0^{\mathcal{M}_g}} = \mathbb{E}_0\left[\frac{\mathrm{d}\mathbb{P}_h^{\mathcal{M}}}{\mathrm{d}\mathbb{P}_0^{\mathcal{M}}} \mid \mathcal{M}_g\right] = \exp\left(h\Delta_g - \frac{1}{2}h^2\mathcal{I}_g\right), \tag{25}$$

[7] This is due to the decomposition $B_{\phi_g} = \sigma_{\varepsilon\phi_g}B_\varepsilon + \sum_{k=1}^\infty J_{g,k}B_{b_k}$.

with

$$(26) \qquad \Delta_g = \Delta_\varepsilon + \lambda \Delta_\perp, \quad \text{and}$$

$$(27) \qquad \mathcal{I}_g = \int_0^1 W_\varepsilon^2(s)\mathrm{d}s + \lambda^2 \left(\frac{J_g}{\sigma^2_{\varepsilon\phi_g}} - 1 \right) \left[\int_0^1 W_\varepsilon(s)^2 \mathrm{d}s - \left(\int_0^1 W_\varepsilon(s)\mathrm{d}s \right)^2 \right],$$

where $\Delta_\varepsilon = \int_0^1 W_\varepsilon(s)\mathrm{d}W_\varepsilon(s)$, $\Delta_\perp = \sqrt{J_g/\sigma^2_{\varepsilon\phi_g} - 1} \int_0^1 W_\varepsilon(s)\mathrm{d}B_\perp(s)$, and $\lambda = (J_{fg}\sigma_{\varepsilon\phi_g} - \sigma^2_{\varepsilon\phi_g})/(J_g - \sigma^2_{\varepsilon\phi_g})$.

Remark 4.1 The result of Proposition 4.1 can also be achieved by firstly applying Girsanov's Theorem to the following experiment

$$\mathrm{d}W_\varepsilon(s) = hW_\varepsilon(s)\mathrm{d}s + \mathrm{d}Z_\varepsilon(s),$$
$$\mathrm{d}W_{\phi_g}(s) = hJ_{fg}W_\varepsilon(s)\mathrm{d}s + \eta_g \mathrm{d}s + \mathrm{d}Z_{\phi_g}(s),$$

to get the likelihood ratio of $\sigma(W_\varepsilon(s), W_{\phi_g}(s), s \in [0,1])$ and, subsequently, taking the expectation of it conditional on $\mathcal{M}_g$. The experiment above is obtained by combining the limit experiment in Proposition 3.3 and the covariance matrix in (21). Here $\eta_g = \sum_k \eta_k J_{g,k}$ with $J_{g,k} = \mathrm{Cov}_{0,0}(W_{\phi_g}(1), W_{b_k}(1))$.

Observe that $W_\perp$ is a standard Brownian motion under $P^{(T)}_{0,0;\mu,f}$ which is independent of W_ε. When $g = f$, we have $J_{fg} = J_f = J_g$ and $\sigma_{\varepsilon\phi_g} = 1$, so that $\lambda = 1$ and $B_{\phi_g} = B_{\phi_f}$. As a result $\Delta_g = \Delta_f^*$ and $\mathcal{I}_g = \mathcal{I}_f^*$.

The central idea to construct a hybrid rank test is to use a (quasi)-log-likelihood ratio test based on $L_{\mathcal{M}_g}(h, \lambda) = h\Delta_g - \frac{1}{2}h^2\mathcal{I}_g$ from (25), where we replace W_ε and B_{ϕ_g} by their finite-sample counterparts from Lemma 4.1 and the unknown parameters σ_f^2 and λ by estimates. Therefore, we impose the following condition.

Assumption 3 There exist consistent, under the null hypothesis, estimators $\hat{\sigma}_f^2 > 0$ *a.s.*, $\hat{\sigma}_{\varepsilon\phi_g}$, and $\hat{J}_{fg}$ of σ_f^2, $\sigma_{\varepsilon\phi_g}$, and J_{fg}, respectively. More precisely, for all $f \in \mathcal{F}$, we have $\hat{\sigma}_f^2 \xrightarrow{p} \sigma_f^2$, $\hat{\sigma}_{\varepsilon\phi_g} \xrightarrow{p} \sigma_{\varepsilon\phi_g}$, and $\hat{J}_{fg} \xrightarrow{p} J_{fg}$, under $\mathrm{P}^{(T)}_{0,0;\mu,f}$ as $T \to \infty$.

Such estimators are easily constructed, although $\hat{J}_{fg}$ is somewhat more involved. Estimating the *real-valued* cross-information J_{fg} requires nonparametric techniques, but is considerably simpler than a full *nonparametric* estimation of ϕ_f. Estimating J_{fg} can be done along similar lines as estimating the Fisher information J_f, see, e.g., Bickel (1982), Bickel et al. (1998), Schick (1986), and Klaassen (1987). A direct rank-based estimator of J_{fg} has been proposed in Cassart, Hallin, Paindaveine (2010).

Next, based on a chosen reference density g satisfying Assumption 2 and some estimators $\hat{\sigma}_f$, $\hat{\sigma}_{\varepsilon\phi_g}$ and $\hat{J}_{fg}$ satisfying Assumption 3, we introduce the following two partial sum processes:

$$\widehat{W}_{\varepsilon}^{(T)}(s) = \frac{1}{\sqrt{T}} \sum_{t=2}^{[sT]} \frac{\Delta Y_t}{\hat{\sigma}_f}, \tag{28}$$

$$\widehat{B}_{\perp}^{(T)}(s) = \left(\frac{J_g}{\hat{\sigma}^2_{\varepsilon\phi_g}} - 1\right)^{-\frac{1}{2}} \left[\frac{B_{\phi_g}^{(T)}(s)}{\hat{\sigma}_{\varepsilon\phi_g}} - \left(\widehat{W}_{\varepsilon}^{(T)}(s) - \widehat{W}_{\varepsilon}^{(T)}(1)[sT]\right)\right], \tag{29}$$

where $B_{\phi_g}^{(T)}(s)$ is defined in (18). Then, given a fixed alternative $\bar{h} < 0$, we define

$$\widehat{L}_{\mathcal{M}_g}^{(T)}(\bar{h}, \hat{\lambda}) := \bar{h}\widehat{\Delta}_g^{(T)} - \frac{1}{2}\bar{h}^2\widehat{\mathcal{I}}_g^{(T)}, \tag{30}$$

with

$$\begin{aligned} \widehat{\Delta}_g^{(T)} &= \widehat{\Delta}_{\varepsilon}^{(T)} + \hat{\lambda}\widehat{\Delta}_{\perp}^{(T)}, \\ \widehat{\mathcal{I}}_g^{(T)} &= \int_0^1 \left(\widehat{W}_{\varepsilon}^{(T)}(s-)\right)^2 \mathrm{d}s \\ &\quad + \hat{\lambda}^2 \left(\frac{J_g}{\hat{\sigma}^2_{\varepsilon\phi_g}} - 1\right) \left[\int_0^1 \left(\widehat{W}_{\varepsilon}^{(T)}(s-)\right)^2 \mathrm{d}s - \left(\int_0^1 \widehat{W}_{\varepsilon}^{(T)}(s-)\mathrm{d}s\right)^2\right], \end{aligned}$$

where $\widehat{\Delta}_{\varepsilon}^{(T)} = \int_0^1 \widehat{W}_{\varepsilon}^{(T)}(s-)\mathrm{d}\widehat{W}_{\varepsilon}^{(T)}(s)$, $\widehat{\Delta}_{\perp}^{(T)} = \sqrt{J_g/\hat{\sigma}^2_{\varepsilon\phi_g} - 1}\int_0^1 \widehat{W}_{\varepsilon}^{(T)}(s-)\mathrm{d}\widehat{B}_{\perp}^{(T)}(s)$ and $\hat{\lambda} = (\hat{J}_{fg}\hat{\sigma}_{\varepsilon\phi_g} - \hat{\sigma}^2_{\varepsilon\phi_g})/(J_g - \hat{\sigma}^2_{\varepsilon\phi_g})$.

By Slutsky's theorem and (19), we have $\left[\widehat{W}_{\varepsilon}^{(T)} \quad \widehat{B}_{\phi_g}^{(T)}\right]' \Rightarrow \left[W_{\varepsilon} \quad B_{\phi_g}\right]'$, and thus $\left[\widehat{W}_{\varepsilon}^{(T)} \quad \widehat{B}_{\perp}^{(T)}\right]' \Rightarrow \left[W_{\varepsilon} \quad B_{\perp}\right]'$. It follows $\widehat{L}_{\mathcal{M}_g}^{(T)}(\bar{h}, \hat{\lambda}) \Rightarrow L_{\mathcal{M}_g}(\bar{h}, \lambda)$. Define the critical value $c_{\mathcal{M}_g}(\bar{h}, \sigma_{\varepsilon\phi_g}, \lambda, J_g; \alpha)$ by the $(1-\alpha)$-quantile of $L_{\mathcal{M}_g}(\bar{h}, \lambda)$. This leads to the feasible test

$$\phi_{\mathcal{M}_g}^{(T)}(\bar{h}, \alpha) := I\left\{\widehat{L}_{\mathcal{M}_g}^{(T)}(\bar{h}, \hat{\lambda}) \geq c_{\mathcal{M}_g}(\bar{h}, \hat{\sigma}_{\varepsilon\phi_g}, \hat{\lambda}, J_g; \alpha)\right\}.$$

Since these tests are based on the ranks of ΔY_t, but also their average, we name them Hybrid Rank Tests (HRTs). We can now state our main theoretical result.

THEOREM 4.1 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 Hybrid Rank Test $\phi^{(T)}_{\mathcal{M}_g}(\bar{h}, \alpha)$ is asymptotically of size α.
(ii) The Hybrid Rank Test $\phi^{(T)}_{\mathcal{M}_g}(\bar{h}, \alpha)$ is asymptotically invariant.
(iii) The Hybrid Rank Test $\phi^{(T)}_{\mathcal{M}_g}(\bar{h}, \alpha)$ is point-optimal, at $h = \bar{h}$, if $g = f$.

Theorem 4.1 shows the HRTs are valid irrespective of the choice of the reference density and point-optimal for a correctly specified reference density. The proof of this theorem is based on weak convergence of the test statistic $\widehat{L}^{(T)}_{\mathcal{M}_g}(\bar{h}, \hat{\lambda})$ to its limit $L_{\mathcal{M}_g}(\bar{h}, \lambda)$ as shown above. Then, (i) is derived directly by the design of the test; (ii) is proved by the fact that $L_{\mathcal{M}_g}(\bar{h}, \lambda)$ is $\mathcal{M}$-measurable. The proof of (iii) comes from last part of Proposition 4.1: when $g = f$, $L_{\mathcal{M}_g}(h, \lambda) = \log\left(\mathrm{d}\mathbb{P}^{\mathcal{M}}_h / \mathrm{d}\mathbb{P}^{\mathcal{M}}_0\right)$.

Corollary 4.1 (Chernoff-Savage type result) Fix $\alpha \in (0, 1)$ and $\bar{h} < 0$. The Hybrid Rank Test $\phi^{(T)}_{\mathcal{M}_g}(\bar{h}, \alpha)$ is, for any reference density g satisfying Assumption 2, more powerful, at $h = \bar{h}$ and for $\mu \in \mathbb{R}$ and $f \in \mathcal{F}$, than the Elliott, Rothenberg, and Stock (1996) test. Both tests have equal power if f is Gaussian.

Proof: Recall once more that $\widehat{L}^{(T)}_{\mathcal{M}_g}(\bar{h}, \hat{\lambda})$ weakly converges to $L_{\mathcal{M}_g}(\bar{h}, \lambda)$, which, in the limit, is the likelihood ratio restricted to the σ-field $\mathcal{M}_g$. Then, by the Neyman-Pearson Lemma, we conclude from $\mathcal{M}_\varepsilon \subseteq \mathcal{M}_g$ that the HRT is more powerful than the Elliott, Rothenberg, and Stock (1996) test at $h = \bar{h}$. Recalling the decomposition (24), write the limit experiment in Remark 4.1 as

$$
\begin{aligned}
\mathrm{d}W_\varepsilon(s) &= hW_\varepsilon(s)\mathrm{d}s + \mathrm{d}Z_\varepsilon(s), \\
\mathrm{d}W_\perp(s) &= h\frac{J_{fg} - \sigma_{\varepsilon\phi_g}}{\sqrt{J_g - \sigma^2_{\varepsilon\phi_g}}} W_\varepsilon(s)\mathrm{d}s + \frac{\eta_g}{\sqrt{J_g - \sigma^2_{\varepsilon\phi_g}}}\mathrm{d}s + \mathrm{d}Z_\perp(s).
\end{aligned}
$$

When f is Gaussian, $W_\perp$ (or W_{ϕ_g}) provides no more information about h than W_ε since $J_{fg} = \sigma_{\varepsilon\phi_g}$. In that case the HRT and the Elliott, Rothenberg, and Stock (1996) test are asymptotically equivalent. *Q.E.D.*

Corollary 4.1 is a particularly useful result for applied work. The HRT dominates its classical canonical Gaussian counterpart, i.e., the Elliott, Rothenberg, and Stock (1996) test in the present model, for any reference density g. Traditionally, this claim can only be made for Gaussian reference densities, but the non-LAN framework here even allows for a stronger result. Our formulation of the testing problem using invariance arguments is convenient in this respect: the larger the invariant σ-field that is used, the more powerful the test.

The situation can be compared to Quasi Maximum Likelihood methods. However, again, in classical situations these methods are generally restricted to Gaussian reference densities. In the present setup, any reference density g (subject to the regularity conditions imposed) can be used. The resulting test will always be valid, but more powerful in case the reference density chosen is closer to the true underlying density f.

REMARK 4.2 The additional power of the HRT compared to the Elliott, Rothenberg, and Stock (1996) test is not free of charge due to the stronger weak convergence assumption employed. Consequently, the class of models for which the HRTs are valid forms a sub-class of the class where the Elliott, Rothenberg, and Stock (1996) tests are valid. In this sub-class, the HRT dominates the Elliott, Rothenberg, and Stock (1996) test, but outside they may even loose validity. In the opposite direction, the Müller and Watson (2008) low-frequency unit root test can be applied in a even larger class of models than the Elliott, Rothenberg, and Stock (1996) tests. Again, within the class of models where the Elliott, Rothenberg, and Stock (1996) test is valid, it has lower power. A more general and detailed discussion in this direction can be found in Müller (2011).

Our test will still be relevant in many applications, notably those where policy implications are derived under an i.i.d. assumption on the innovations. Also, our approach can most likely be extended to situations where the innovations are described by some explicit dynamic location-scale model. We come back to this point in Section 6.

4.2. *The approximate hybrid rank tests*

A somewhat inconvenient aspect of the hybrid rank tests is that we need to estimate J_{fg}. As mentioned before, this is (much) less complicated than estimating the score function ϕ_f, but might still be considered cumbersome, despite all references mentioned below Assumption 3. Moreover, the critical value $c_{\mathcal{M}_g}(\bar{h}, \hat{\sigma}_{\varepsilon\phi_g}, \hat{\lambda}, J_g; \alpha)$ depends on estimates $\hat{\sigma}_{\varepsilon\phi_g}$ and $\hat{\lambda}$ (henceforth $\hat{J}_{fg}$). This introduces no difficulty to implementing the test, however, when it comes to simulations, the computational effort will be significant. Therefore, we introduce additionally a simplified version of the hybrid rank test. This simplified test is obtain by setting $\lambda = 1$, which holds in case $g = f$.

To be precise, define

$$\widehat{L}_g^{(T)}(\bar{h}) := \widehat{L}_{\mathcal{M}_g}^{(T)}(\bar{h}, 1) = \bar{h}\widehat{\Delta}_g^{(T)} - \frac{1}{2}\bar{h}^2\widehat{\mathcal{I}}_g^{(T)}, \tag{31}$$

where

$$\widehat{\Delta}_g^{(T)} = \frac{1}{\hat{\sigma}_{\varepsilon\phi_g}}\int_0^1 \widehat{W}_\varepsilon^{(T)}(s-)\mathrm{d}\widehat{B}_{\phi_g}^{(T)}(s) + \widehat{W}_\varepsilon^{(T)}(1)\int_0^1 \widehat{W}_\varepsilon^{(T)}(s-)\mathrm{d}s, \tag{32}$$

$$\widehat{\mathcal{I}}_g^{(T)} = \frac{J_g}{\hat{\sigma}_{\varepsilon\phi_g}^2}\int_0^1 \left(\widehat{W}_\varepsilon^{(T)}(s-)\right)^2 \mathrm{d}s - \left(\int_0^1 \widehat{W}_\varepsilon^{(T)}(s-)\mathrm{d}s\right)^2 \left(\frac{J_g}{\hat{\sigma}_{\varepsilon\phi_g}^2} - 1\right), \tag{33}$$

and $L_g(\bar{h}) := L_{\mathcal{M}_g}(\bar{h}, 1)$. By the same arguments as before, we have $\widehat{L}_g^{(T)}(\bar{h}) \Rightarrow Ł_g(\bar{h})$. Denoting the $(1-\alpha)$-quantile of $L_g(\bar{h})$ by $c_g(\bar{h}, \sigma_{\varepsilon\phi_g}, J_g; \alpha)$, this leads to the feasible test

$$\phi_g^{(T)}(\bar{h}, \alpha) := I\left\{L_g^{(T)}(\bar{h}) \geq c_g(\bar{h}, \hat{\sigma}_{\varepsilon\phi_g}, J_g; \alpha)\right\}.$$

Since $\phi_g^{(T)}(\bar{h}, \alpha)$ is an approximate version of the Hybrid Rank Test $\phi_{\mathcal{M}_g}^{(T)}(\bar{h}, \alpha)$, we refer to it as Approximate Hybrid Rank Test (AHRT).

THEOREM 4.2 Under the same conditions as Theorem 4.1, the asymptotic properties of the Hybrid Rank Tests — validity, invariance, and point-optimality when $g = f$ — also hold for the Approximate Hybrid Rank Tests.

The proof of Theorem 4.2 follows along the same lines as that of Theorem 4.1 but using the weak convergence $\widehat{L}_g^{(T)}(\bar{h}) \Rightarrow L_g(\bar{h})$. The simulation results in Section 5 show that these asymptotic properties carry over to finite samples.

REMARK 4.3 (Chernoff-Savage result for the AHRTs) Although we are not able to provide a rigorous mathematical proof, the Monte-Carlo study indicates that the Chernoff-Savage property is also preserved for the AHRT, at least in case the reference density g is chosen to be Gaussian. Such a result would be more in line with applications of the Chernoff-Savage result in classical LAN situations.

From a computational point of view, the AHRT has the advantage that nonparametric estimation of J_{fg} is no longer needed. This significantly reduces the computational effort in the Monte-Carlo study. Indeed, even though the critical value $c_g(\bar{h}, \sigma_{\varepsilon\phi_g}, J_g; \alpha)$ is still data dependent, it is, for given α, $\bar{h}$, and reference density g, a function of only one argument — the parameter $\sigma_{\varepsilon\phi_g}$. Observe, by Cauchy-Schwarz, that $\sigma_{\varepsilon\phi_g}$ is bounded by $\sqrt{J_g}$. Table I and Figure 1 show that, for three canonical reference densities, the critical values are well approximated by a polynomial regression of order 5. In the Section 5, we use these estimated critical value functions for computational speed.

TABLE I

THIS TABLE PROVIDES ESTIMATED CRITICAL VALUE FUNCTIONS FOR THREE REFERENCE DENSITIES: GAUSSIAN ($J_g = 1$), LAPLACE ($J_g = 2$), AND STUDENT t_3 ($J_g = 2$) AT $\alpha = 5\%$ AND $\bar{h} = -7\sigma_{\varepsilon\phi_g}$. FOR EACH CASE, THE CRITICAL VALUE FUNCTION IS ESTIMATED BY OLS USING SIMULATED CRITICAL VALUES, SEE FIGURE 1, ON THE INTERVAL $[0, \sqrt{J_g}]$ WITH A GRID WHERE ADJACENT POINTS ARE 0.01 APART.

g	
Gaussian	$c_g(-7\sigma_{\varepsilon\phi_g}, \sigma_{\varepsilon\phi_g}, 1; 5\%) = 0.96 + 1.88\sigma_{\varepsilon\phi_g} - 3.98\sigma^2_{\varepsilon\phi_g} + 6.74\sigma^3_{\varepsilon\phi_g} - 5.45\sigma^4_{\varepsilon\phi_g} + 1.69\sigma^5_{\varepsilon\phi_g}$
Laplace	$c_g(-7\sigma_{\varepsilon\phi_g}, \sigma_{\varepsilon\phi_g}, 2; 5\%) = 0.25 + 2.30\sigma_{\varepsilon\phi_g} - 3.58\sigma^2_{\varepsilon\phi_g} + 4.30\sigma^3_{\varepsilon\phi_g} - 2.45\sigma^4_{\varepsilon\phi_g} + 0.54\sigma^5_{\varepsilon\phi_g}$
Student t_3	$c_g(-7\sigma_{\varepsilon\phi_g}, \sigma_{\varepsilon\phi_g}, 2; 5\%) = 0.25 + 2.30\sigma_{\varepsilon\phi_g} - 3.58\sigma^2_{\varepsilon\phi_g} + 4.30\sigma^3_{\varepsilon\phi_g} - 2.45\sigma^4_{\varepsilon\phi_g} + 0.54\sigma^5_{\varepsilon\phi_g}$

4.3. *Possible extensions*

We discuss three possible extensions of HRTs and AHRTs: a version based on signed rank statistics in case the error distribution is known to be symmetric, a version based on aligned ranks for the case where the innovations ε_t are serially correlated, and a version based on a nonparametrically estimated reference density which addresses globally optimality in $\mathcal{F}$. Note that these cases may be concurrent and we can combine these extensions accordingly. Details of proofs and associated Monte-Carlo study are left for future work.

REMARK 4.4 (Symmetric error distributions) In this remark we consider the case that $f \in \mathcal{F}$ is known to be symmetric. The density f is modeled nonparametrically as in equation (4) but with all perturbation scores $b_k(e)$ being even functions. We have $J_{f,k} = \sigma_f \mathrm{E}_f[b_k(\varepsilon)\phi_f(\varepsilon)] = 0$ since the score function ϕ_f is odd. As a consequence, Δ_f is independent of Δ_{b_k} for all k. This gives adaptivity[8] for testing the parameter of interest h in the presence of the nuisance parameter η: applying Girsanov's Theorem

[8]A discussion about definition of "adaptive" in this nonstandard unit root testing problem can be found in Section 5 of Jansson (2008).

gives the limit experiment structurally as in Proposition 3.3

$$
\begin{aligned}
\mathrm{d}Z_{\varepsilon}(s) &= \mathrm{d}W_{\varepsilon}(s) - hW_{\varepsilon}(s)\mathrm{d}s, \\
\mathrm{d}Z_{\phi_f}(s) &= \mathrm{d}W_{\phi_f}(s) - hJ_f W_{\varepsilon}(s)\mathrm{d}s, \\
\mathrm{d}Z_{b_k}(s) &= \mathrm{d}W_{b_k}(s) - \eta_k \mathrm{d}s \quad k \in \mathbb{N}.
\end{aligned}
$$

The equation for W_{ϕ_f} is not omitted since (5) does not hold anymore. By the same method used to prove Theorem 3.1, we can show that $\sigma(W_\varepsilon, W_{\phi_f}, B_{b_k})$ is maximal invariant. Subsequently, after some simple algebra, we find the semiparametric power envelope based on this maximal invariant coincides with parametric power envelope where f is known. This verifies again the adaptation result from Jansson (2008) under the same condition with the new approach. To demonstrate that the semiparametric power envelope is sharp, we propose a test based on signed-rank statistics. This is a natural counterpart of the maximal invariant in the sequence

$$
\begin{aligned}
\widehat{W}_{\varepsilon}^{(T)}(s) &= \frac{1}{\sqrt{T}} \sum_{t=2}^{[sT]} \frac{\Delta Y_t}{\hat{\sigma}_f}, \\
\widehat{W}_{\phi_g}^{(T)}(s) &= \frac{1}{\sqrt{T}} \sum_{t=2}^{[sT]} s_t \sigma_g \phi_g \left(G^{-1} \left(\frac{T+1+R_t^+}{2(T+1)} \right) \right)
\end{aligned}
$$

where $(R_1^+, ..., R_T^+)$ are the ranks of absolute values of $(\hat{\varepsilon}_1, ..., \hat{\varepsilon}_T)$, and $(s_1, ..., s_T)$ are signs of $(\hat{\varepsilon}_1, ..., \hat{\varepsilon}_T)$. The symmetric reference density g is assumed to be symmetric with variance σ_g, score function ϕ_g and quantile function G^{-1}. Under the symmetric density condition, $\widehat{W}_{\varepsilon}^{(T)}$ and $\widehat{W}_{\phi_f}^{(T)}$ weakly converges to W_ε and W_{ϕ_f}, respectively.

Remark 4.5 (Serial-correlated errors) In this remark, we discuss possible extension for the case where errors are possibly serially correlated. To be specific, in model equation (2), let v_t denote the innovation at time t instead of ε_t, and model it as $v_t = \gamma_1 v_{t-1} \cdots - \gamma_p v_{t-p} + \varepsilon_t$. We assume the same assumptions on ε_t as above. The inference for ρ is adaptive to the present of γ in the sense that their corresponding score functions are asymptotically independent (see Section 7 of Jansson (2008)). Therefore, replacing γ by some consistent estimator $\hat{\gamma}$ will not affect the result of testing ρ (asymptotically). Recall the i.i.d error case considered above, we use ΔY_t, which actually plays the role of estimates $\hat{\varepsilon}_t$ for ε_t under the null hypothesis, and R_t, which is the rank of $\hat{\varepsilon}_t$, to build the HRT and the AHRT statistics. In this case, estimates for ε_t becomes $\hat{\varepsilon}_t = \Delta Y_t - \hat{\gamma}_1 \Delta Y_{t-1} \cdots - \hat{\gamma}_p \Delta Y_{t-p}$, and subsequentially, the rank of the new estimates $\hat{\varepsilon}_t$, R_t, becomes aligned ranks. Consistency of $\hat{\gamma}$ gives the consistency of $\hat{\varepsilon}_t$. Then, with these consistent estimates of errors and associated aligned ranks, the convergences in (19) and (20) are preserved and hence the properties of HRTs and AHRTs are also preserved for the aligned-rank-based versions.

Remark 4.6 (Nonparametrically estimated reference density) The Hybrid Rank Test and the Approximate Hybrid Rank Test are optimal when the reference density g coincides with the actual innovation density f. It is therefore reasonable to consider these test using a nonparametric estimate of f, say $\hat{f}$, as reference density. Commonly such estimate is based on the order statistics of the innovations ε_t and thus independent of the ranks in the HRT. Under a suitable consistency condition, the HRT based on $\hat{f}$ asymptotically behaves as the HRT based on the true innovation density f. Thus, such test achieves the optimality properties of Theorem 4.1 and Theorem 4.2 globally. Notably, even if there exists relatively large bias in the estimation of f, the usage of rank statistics ensures zero expectation of the feasible score function $\phi_{\hat{f}}[\widehat{F}^{-1}(R_t/(T+1))]$, which furthermore ensures the validity of the HRTs and the AHRTs. This argument can also be showed with the fact that rank-based score converges to Brownian bridge as $T \to \infty$, and $\int_0^1 W_\varepsilon(s)\mathrm{d}B_{\phi_f}(s) = \int_0^1 \overline{W}_\varepsilon(s)\mathrm{d}W_{\phi_f}(s)$, where $\overline{W}_\varepsilon(s) = W_\varepsilon(s) - \int_0^1 W_\varepsilon(s)\mathrm{d}s$. Thus, a drift in W_{ϕ_f} caused by estimation bias will be canceled out.

5. MONTE CARLO STUDY

This section reports the results of a Monte Carlo study to corroborate our asymptotic results, and to analyze the small-sample performances of the Approximate Hybrid Rank Tests. As mentioned earlier, we use the Approximate Hybrid Rank Tests in this simulation to avoid having to simulate the critical value for each individual replication. For the fixed alternative, we choose $\bar{h} = -7\sigma_{\varepsilon\phi_g}$ for two reasons. First, when $g = f$, we have $\sigma_{\varepsilon\phi_g} = 1$ and hence $\bar{h} = -7$, which is in line with the Elliott, Rothenberg, and Stock (1996) paper. Second, if we choose $\bar{h} = -7$, the critical value will approach $-\infty$ when $\sigma_{\varepsilon\phi_g} \to 0$. The corresponding critical value functions for various reference densities are provided in Table I. The estimators for σ_f^2 and $\sigma_{\varepsilon\phi_g}$ we use are

$$\hat{\sigma}_f^2 = \frac{1}{T-1}\sum_{t=2}^{T}(\Delta Y_t - \frac{1}{T}\sum_{t=2}^{T}\Delta Y_t)^2, \tag{34}$$

$$\hat{\sigma}_{\varepsilon\phi_g} = \frac{1}{T-1}\sum_{t=2}^{T}\frac{\Delta Y_t}{\hat{\sigma}_f}\sigma_g\phi_g\left(G^{-1}\left(\frac{R_t}{T+1}\right)\right). \tag{35}$$

Moreover, to simplify the notations, we denote the Approximate Hybrid Rank test with reference density g by AHRT^g and, in particular, by AHRT$^\phi$ for Gaussian reference density. Throughout we use the significance level $\alpha = 5\%$ and all simulations are based on 20,000 Monte-Carlo repetitions.

We compare our AHRT with two alternatives. First we consider the Dickey-Fuller test (denoted by DF-ρ) from Dickey and Fuller (1979). This test is based on the statistic $T(\hat{\rho}-1)$ where $\hat{\rho}$ is the least-squares estimator in the regression $Y_t = \mu + \rho Y_{t-1} + \varepsilon_t$. The critical values for this test are -13.52 for $T = 100$ and -14.05 for $T = 2500$. The second competitor is the Elliott, Rothenberg, and Stock (1996) test with $\bar{h} = -7$. This test is based on the statistic $[S(\bar{\alpha}) - \bar{\alpha}S(1)]/\hat{\omega}^2$ with $\bar{\alpha} = 1 + T^{-1}\bar{h}$ and $S(a) = (Y_a - Z_a\hat{\beta})'(Y_a - Z_a\hat{\beta})$, with Y_a and Z_a defined as

$$Y_a = (Y_1, Y_2 - aY_1, ..., Y_T - aY_{T-1})',$$
$$Z_a = (1, 1-a, ..., 1-a)',$$

where $\hat{\beta}$ is estimated by regressing $Y_{\bar{\alpha}}$ on $Z_{\bar{\alpha}}$. Since in the present model we employ the i.i.d. assumption on the innovations, the long-run variance estimator $\hat{\omega}^2$ is chosen to be $\hat{e}'\hat{e}/T$, where $\hat{e}$ is the residual vector from the regression $\Delta Y_t = \mu + \delta Y_{t-1} + \varepsilon_t$. The critical values for this test are 3.11 for $T = 100$ and 3.26 for $T = 2500$. We do not consider the Dickey-Fuller t-test as it is dominated by the DF-ρ test in the current model. Similarly, the Elliott, Rothenberg, and Stock (1996) DF-GLS test is also omitted as it behaves asymptotically the same as the Elliott, Rothenberg, and Stock (1996) test, but can be oversized in small samples.

5.1. *Large-sample performance*

In order to check the large-sample performance, we compare the three tests mentioned above in a setting of $T = 2,500$.

Figure 2 shows the power curves for 9 combinations of 3 innovation densities f and 3 reference densities g. Each are chosen to be Laplace, Student t_3, or Gaussian. In line with our theoretical results, we find that the AHRT outperforms the two competitors in most cases. More specifically, when $g = f$ (the graphs on the diagonal), the AHRT^f has power very close to the semiparametric power envelope and it is tangent to it at the point $h = -7$. Moreover, when the reference density g is Gaussian (the three right-most graphs), the AHRT$^\phi$ outperforms the competitors for all three true densities f. This corroborates the Chernoff-Savage property of the AHRT$^\phi$ test mentioned in Remark 4.3. When both g and f are Gaussian, the AHRT test and the Elliott, Rothenberg, and Stock (1996) test have indistinguishable power.

In order to investigate the Chernoff-Savage result even further, we consider in Figure 3 the AHRT$^\phi$ test for nine true innovation densities f. These include innovation densities f that are extremely heavy-tailed, skewed, or both. The first row of graphs shows three extremely heavy-tailed distributions: Student t_2, Student t_1, and a stable distribution with stability parameter $\alpha = 0.5$, skewness parameter $\beta = 0$, scale parameter $c = 1$, and location parameter $\mu = 0$. As these densities do not all satisfy our maintained assumptions, these graphs do not show power envelopes.

The top three graphs in Figure 3 show that the AHRT$^\phi$ is much more powerful than its competitors and that its power increases with the heaviness of the tail. The second and third row show the effect of skewness in f. Specifically, the AHRT$^\phi$ power function when f is skewed-normal (with skewness 0.8145) is higher than that when f is normal (in Figure 2). This indicates that the AHRT$^\phi$ can acquire power from skewness. The same conclusion can be drawn from the comparison of the AHRT$^\phi$ power function for t_4 and that of a skewed t_4 with skewness ≈ 2.7. To further remove the effects of the other moments, in the third row, we also employ the Pearson distributions with identical mean, variance and kurtosis, but different skewness — skewness = 1 for Pearson-I, skewness = 3 for Pearson-II, and skewness = 6 for Pearson-III. Comparing the corresponding three AHRT$^\phi$ power functions, it seems that the larger the skewness of the true distribution f is, the more powerful the AHRT$^\phi$ becomes.

A final remark on the size of the AHRT. In all cases where the true density f satisfies our maintained assumption, i.e., $f \in \mathcal{F}$ (that is all cases in Figure 2 and the skewnormal, t_4, Pearson-I, Pearson-II, and Pearson-III in Figure 3), the simulated sizes are between 4.9% and 5.1%. This verifies the validity of the AHRTs claimed in Theorem 4.2. In the other cases, i.e., $f \notin \mathcal{F}$, the AHRT is somewhat conservative. More precisely, the simulated sizes of the AHRT$^\phi$ are 4.845%, 4.085%, 3.725%, and 4.735% for the t_2, t_1, stable, and skew-t_4 distribution, respectively. This result seems consistent over all simulations, but we have not been able to provide formal proof.

5.2. *Small-sample performance*

We finally report the performance of the AHRTs and the two competitors described above for samples of size $T = 100$. Figures 4 and 5 are the small-sample versions, with $T = 100$, of Figures 2 and 3, respectively. We observe that, even with a slight downward shift of the power functions for all three tests considered, the findings of the large-sample case remain valid in the small-sample case. For larger values of h, the DF-ρ test sometimes dominates the other two tests. Again, when f is significantly away from the Gaussian density, irrespective of the choice of g, the AHRT dominates the other two tests.

Concerning the small-sample size, we find it to range from about 4.0% to 4.5% for the cases where $f \in \mathcal{F}$. Again, when f does not satisfy our maintained assumptions ($f \notin \mathcal{F}$) the AHRT turns out to be conservative. More precisely, we find a size of 3.7%, 3.1%, 2.4%, and 4.1% for the t_2, t_1, stable, and skew-t_4 distribution, respectively. This makes the improved power even more remarkable.

It may also be useful to illustrate the convergence of the power function of the AHRT^f to the semiparametric power envelope as sample size T increases. This is the purpose of Figure 6. For three cases: Gaussian, Laplace, and Student t_3, we find that the convergence indeed occurs already at relatively small samples, which is not always the case for alternative unit root tests.

6. CONCLUSION AND DISCUSSION

This paper has provided a structural representation of the limit experiment of the standard unit root model in a univariate but semiparametric setting. Using invariance arguments, we have derived the semiparametric power envelope. These invariance structures also lead, using the Neyman-Pearson lemma, to point-optimal semiparametric tests. The analysis naturally leads to the use of rank-based statistics.

Our tests are asymptotically valid, invariant, and (with a correctly chosen reference density) point-optimal. Moreover, we establish a Chernoff-Savage type property of our test: irrespective of the reference density chosen, our test outperforms its classical competitor which in this case is the Elliott, Rothenberg, and Stock (1996) test. Finally, we introduced a simplified version of our test and show, in a small Monte-Carlo study, that our theoretical results carry over to small samples.

As potential future work we mention the use of similar ideas to construct hybrid rank-based tests in more general time-series models which, for instance, allow for serial correlation in the error terms, a deterministic time trend term, or stochastic volatility. Also, the structural representation of the limit experiment and its invariance properties can be applied to other non-stationary time-series models (for instance cointegration or predictive regression models).

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. (1982), On Adaptive Estimation. *Annals of Statistics*, 10, 647–671.

Bickel, P.J., C.A.J. Klaassen, Y. Ritov, and J.A. Wellner (1998), *Efficient and Adaptive Estimation for Semiparametric Models.* Amsterdam: Springer-Verlag.

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.

Cassart, D., Hallin, M., and Paindaveine, D. (2010). On the estimation of cross-information quantities in rank-based inference. *Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in honor of Professor Jana Jureckova*, Vol.7, pages 35-45.

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.

Choi, I. (2015). *Almost All About Unit Roots*. Cambridge University Press.

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.

Hall, P. and C.C. Heyde (1980). Martingale limit theory and its application. Academic Press, New York.

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.

Hansen, B.E. (1992). Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory*, 8, 489–500.

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.

Jacod, J. and A.N. Shiryaev (2002). Limit Theorems for Stochastic Processes. Springer, Berlin.

Jansson, M. (2008). Semiparametric power envelopes for tests of the unit root hypothesis. *Econometrica*, 76, 1103–1142.

Jansson, M. and Nielsen, M. (2012). Nearly efficient likelihood ratio tests of the unit root hypothesis. *Econometrica*, 80, 2321–2332.

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

Klaassen, C. A. J. (1987). Consistent Estimation of the Influence Function of Locally Asymptotically Linear Estimators. *Annals of Statistics*, 15, 1548–1562.

Karatzas, I. and S.E. Shreve (1991). *Brownian motion and stochastic calculus.* (second edition) Springer-Verlag.

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

Lehmann, Erich L., Romano, Joseph P. (2005). *Testing Statistical Hypothesis* (third edition). Springer, New York.

Müller, U.K. (2011). Efficient tests under a weak convergence assumption. *Econometrica*, 79, 395–435.

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

Müller, U.K., and M. W. Watson (2008). Testing Models of Low-Frequency Variability. *Econometrica*, 76, 979-1016.

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.

Schick, A. (1986). On Asymptotically Efficient Estimation in Semiparametric Models. *Annals of Statistics*, 14, 11391151.

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: WEAK CONVERGENCE OF PARTIAL SUM PROCESSES

Introduce the filtrations $\mathbb{F}^{(T)} := \big(\mathcal{F}_u^{(T)},\ u \in [0,1]\big)$, $T \in \mathbb{N}$, defined by $\mathcal{F}_u^{(T)} := \sigma\big(Y_t,\ t \in \mathbb{N} : t \leq [uT]\big)$, $u \in [0,1]$. The angle-bracket process $\big\langle A_i^{(T)}, A_j^{(T)}\big\rangle(u)$ and the straight-bracket process $\big[A_i^{(T)}, A_j^{(T)}\big](u)$ are now well-defined for all $\mathbb{F}^{(T)}$-adapted locally square-integrable martingales and semimartingales $A_i^{(T)}$, respectively (see, e.g., Jacod and Shiryaev (2002)). If $A_i^{(T)}$, $i = 1, 2$, are square-integrable martingales of the form $A_i^{(T)}(u) = \sum_{t=1}^{[uT]} I_{Tt}^{(i)}$ with $I_{Tt}^{(i)}$ $\mathcal{F}_t$-measurable, we have $\big[A_1^{(T)}, A_2^{(T)}\big](u) = \sum_{t=1}^{[uT]} I_{Tt}^{(1)} I_{Tt}^{(2)\prime}$ and $\big\langle A_1^{(T)}, A_2^{(T)}\big\rangle(u) = \sum_{t=1}^{[uT]} \mathrm{E}\big[I_{Tt}^{(1)} I_{Tt}^{(2)\prime} \mid \mathcal{F}_{t-1}\big]$. Recall that for a square-integrable martingale with continuous sample paths the angle-brackets and straight-brackets coincide.

The lemma below shows that the partial sum processes introduced in Section 3.1 weakly converges to the associated Brownian motions. Due to the i.i.d.-ness of the innovations, the lemma is a direct corollary to the functional central limit theorem VIII.3.33 in Jacod and Shiryaev (2002).

Lemma A.1 Let $f \in \mathcal{F}$ and let, with $m \geq 2$, $k_1, \dots, k_{m-2} \in \mathbb{N}$. Define, with the notation of Section 3.1,

$$\begin{aligned}\mathcal{W}^{(T)} &= (W^{(T)}_{\varepsilon}, W^{(T)}_{\phi_f}, W^{(T)}_{b_1}, \ldots, W^{(T)}_{b_{m-2}})' \text{ and}\\ \mathcal{W} &= (W_{\varepsilon}, W_{\phi_f}, W_{b_1}, \ldots, W_{b_{m-2}})'.\end{aligned}$$

Then we have, in $D_{\mathbb{R}^m}[0,1]$ and under $\mathrm{P}^{(T)}_{0,0;\mu,f}$,

$$\mathcal{W}^{(T)} \Rightarrow \mathcal{W}, \tag{36}$$

and, still under $\mathrm{P}^{(T)}_{0,0;\mu,f}$,

$$\langle \mathcal{W}^{(T)}, \mathcal{W}^{(T)} \rangle(1) = \left[\mathcal{W}^{(T)}, \mathcal{W}^{(T)}\right](1) + o_{\mathrm{P}}(1) = \mathrm{Var}_f\left(\mathcal{W}(1)\right) + o_{\mathrm{P}}(1). \tag{37}$$

APPENDIX B: MAIN PROOFS

PROOF OF PROPOSITION 3.1:
For notational convenience we drop the superscript "(T)" in the following and thus write f_η instead of $f^{(T)}_\eta$. It is clear, since η has finite support, that we have $f_\eta > 0$ 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_η integrates to 1. Similarly, $\int b_k(e) e f(e) \mathrm{d}e = 0$ implies $\mathrm{E}_{f_\eta}[\varepsilon_t] = 0$. Of course, absolute continuity of f_η follows from $f \in \mathcal{F}$ 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 $\mathrm{Var}_{f_\eta}[\varepsilon_t] < \infty$. Only $J_{f_\eta} < \infty$ requires a bit of straightforward calculus. 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) $\left\|1 + T^{-1/2} \sum_{k=1}^{\infty} \eta_k b_k\right\|_\infty \leq C_1$ and $\left\|T^{-1/2} \sum_{k=1}^{\infty} \eta_k b'_k\right\|^2_\infty \leq C_2$. Moreover, there exists $C_3 > 0$ such that, for all $T \geq T'$, $\|(1 + T^{-1/2} \sum_{k=1}^{\infty} \eta_k b_k)^{-1}\|^2_\infty \leq C_3$. Using these observations we immediately obtain, for $T \geq T'$,

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

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

PROOF OF PROPOSITION 3.2:
We first note that Part (iii) directly follows from an application of Corollary 3.5.16 in Karatzas and Shreve (1991). In the following we evaluate, unless mentioned otherwise, expectations, O_{P}'s, and o_{P}'s under $\mathrm{P}^{(T)}_{0,0;\mu,f}$. We first establish Part (ii) and

prove the quadratic expansion (i) afterwards. Let $k_1, \dots, k_m$ denote the elements for which η does not vanish, i.e. $\eta_{k_j} \neq 0$ and $\eta_i = 0$ for $i \notin \{k_1, \dots, k_m\}$. And let $a_T = (h_T, \eta_{k_1}, \dots, \eta_{k_m})'$ and $a = (h, \eta_{k_1}, \dots, \eta_{k_m})'$.

Proof of Part (ii): First we introduce auxiliary processes $\tilde{\Delta}^{(T)}$, $T \in \mathbb{N}$, by

$$\tilde{\Delta}^{(T)}(r) = \left(\int_0^r W_\varepsilon^{(T)}(s-) \mathrm{d} W_{\phi_f}^{(T)}(s), W_{b_{k_1}}^{(T)}(r), \dots, W_{b_{k_m}}^{(T)}(r) \right), \; r \in [0,1].$$

A combination of Lemma A.1 with Theorem 2.1 in Hansen (1992) (the conditions are trivially met) yields $\tilde{\Delta}^{(T)} \Rightarrow \tilde{\Delta}$ in $D_{\mathbb{R}^{m+1}}[0,1]$, where $\tilde{\Delta}$ is given by

$$\tilde{\Delta}(r) = \left(\int_0^r W_\varepsilon(s) \mathrm{d} W_{\phi_f}(s), W_{b_{k_1}}(r), \dots, W_{b_{k_m}}(r) \right)',$$

$r \in [0,1]$ (which we evaluate under $\mathbb{P}_{0,0}$). Using this weak convergence, the identity $[A, B](r) = A(r)B(r) - A(0)B(0) - \int_0^r A(s-)\mathrm{d}B(s) - \int_0^r B(s-)\mathrm{d}A(s)$ and the continuous mapping theorem in combination with Theorem 2.1 in Hansen (1992) (the condition to this theorem is met as $\tilde{\Delta}^{(T)}$ is a martingale with respect to $\mathbb{F}^{(T)}$ and as we have $\sum_{t=1}^T \mathrm{E}|\tilde{\Delta}^{(T)}(t/T) - \tilde{\Delta}^{(T)}((t-1)/T)|^2 = O(1)$) yields,

$$\left(\tilde{\Delta}^{(T)}(1), \left[\tilde{\Delta}^{(T)}, \tilde{\Delta}^{(T)} \right](1) \right) \Rightarrow \left(\tilde{\Delta}(1), \left\langle \tilde{\Delta}, \tilde{\Delta} \right\rangle (1) \right). \tag{38}$$

The quadratic variation at time 1, $\left\langle \tilde{\Delta}, \tilde{\Delta} \right\rangle_1$, is given by

$$\left\langle \tilde{\Delta}_1, \tilde{\Delta}_1 \right\rangle_1 = J_f \int_0^1 W_\varepsilon^2(s) \mathrm{d}s, \quad \left\langle \tilde{\Delta}_{j+1}, \tilde{\Delta}_{j+1} \right\rangle_1 = 1,$$
$$\left\langle \tilde{\Delta}_1, \tilde{\Delta}_{j+1} \right\rangle_1 = J_{k_j,f} \int_0^1 W_\varepsilon(s) \mathrm{d}s, \text{ and } \left\langle \tilde{\Delta}_{1+i}, \tilde{\Delta}_{1+j} \right\rangle_1 = 0,$$

or $i \neq j \in \{1, \dots, m\}$. The angle brackets of $\tilde{\Delta}^{(T)}$ at time 1, $\left\langle \tilde{\Delta}^{(T)}, \tilde{\Delta}^{(T)} \right\rangle_1$, take a similar form (just replace the limiting Brownian motions by their empirical analogues). On noting that we have

$$\mathcal{I}_f^{(T)}(h, \eta) = a' \left\langle \tilde{\Delta}^{(T)}, \tilde{\Delta}^{(T)} \right\rangle_1 a \quad \text{and} \quad \mathcal{I}_f(h, \eta) = a' \left\langle \tilde{\Delta}, \tilde{\Delta} \right\rangle_1 a, \tag{39}$$

the proof of Part (ii) follows if we show

$$\left\langle \tilde{\Delta}^{(T)}, \tilde{\Delta}^{(T)} \right\rangle_1 - \left[\tilde{\Delta}^{(T)}, \tilde{\Delta}^{(T)} \right]_1 = o_{\mathrm{P}}(1). \tag{40}$$

Indeed, a combination of (38)-(40) with the continuous mapping theorem yields (ii). To demonstrate (40) we first note that Lemma A.1 yields, for $i,j=1,\dots,m$, $\left\langle \tilde{\Delta}_{1+i}^{(T)}, \tilde{\Delta}_{1+j}^{(T)} \right\rangle(1) - \left[\tilde{\Delta}_{1+i}^{(T)}, \tilde{\Delta}_{1+j}^{(T)}\right](1) = o_{\mathrm{P}}(1)$. So we only need to consider

$$r_1^{(T)} = \left[\tilde{\Delta}_1^{(T)}, \tilde{\Delta}_1^{(T)}\right]_1 - \left\langle \tilde{\Delta}_1^{(T)}, \tilde{\Delta}_1^{(T)} \right\rangle_1 = \frac{1}{T^2}\sum_{t=2}^{T} \sigma_f^{-2}(Y_{t-1}-Y_1)^2\left(\sigma_f^2\phi_f^2(\varepsilon_t) - J_f\right) \text{ and}$$
$$r_{2,j}^{(T)} = \left[\tilde{\Delta}_1^{(T)}, \tilde{\Delta}_{1+j}^{(T)}\right]_1 - \left\langle \tilde{\Delta}_1^{(T)}, \tilde{\Delta}_{1+j}^{(T)} \right\rangle_1 = \frac{1}{T^{3/2}}\sum_{t=2}^{T} \sigma_f^{-1}(Y_{t-1}-Y_1)\left(\sigma_f\phi_f(\varepsilon_t)b_{k_j}(\varepsilon_t) - J_{f,k_j}\right),$$

for $j=1,\dots,m$. We have

$$\mathbb{E}(r_{2,j}^{(T)})^2 = \frac{1}{T}\,\mathrm{Var}_f\left[\sigma_f\phi_f(\varepsilon_1)b_{k_1}(\varepsilon_t)\right]\int_0^1 \mathbb{E}(W_\varepsilon^{(T)}(u-))^2 \mathrm{d}u = o(1).$$

For $r_1^{(T)}$ the same line of reasoning can be followed in case $\phi_f(\varepsilon_1)$ has a finite fourth moment. This, however, does not need to be the case under Assumption 2. Therefore we resort to an application of Theorem 2.23 in Hall and Heyde (1980) which shows that $r_1^{(T)} = o_{\mathrm{P}}(1)$ if, for all $\delta > 0$,

$$\sum_{t=2}^{T} \frac{1}{T^2}\mathrm{E}\left[(Y_{t-1}-Y_1)^2\phi_f^2(\varepsilon_t)1_{\{|(Y_{t-1}-Y_1)\phi_f(\varepsilon_t)|>\delta T\}} \mid \mathcal{F}_{t-1}\right] = o_{\mathrm{P}}(1). \tag{41}$$

Using the notation $\zeta(M) = \mathrm{E}\left[\sigma_f^2\phi_f^2(\varepsilon_1)1_{\{|\sigma_f\phi_f(\varepsilon_1)|\geq M\}}\right]$, we see that the left-hand-side of the previous display is bounded by

$$\zeta\left(\frac{\delta\sqrt{T}}{\|W_\varepsilon^{(T)}\|_\infty}\right)\int_0^1 \left(W_\varepsilon^{(T)}(u-)\right)^2 \mathrm{d}u = o_{\mathrm{P}}(1),$$

by a combination of Lemma A.1, the continuous mapping theorem, and $\zeta(M) \to 0$ as $M \to \infty$ (dominated convergence). This concludes the proof of Part (ii).

Proof of Part (i): We use Proposition 1 in Hallin, Van den Akker, and Werker (2015) to prove the expansion. To this end we set $\tilde{P}_T = \mathrm{P}_{h_T,\eta;\mu,f}^{(T)}$, $P_T = \mathrm{P}_{0,0;\mu,f}^{(T)}$, and $\mathcal{F}_{Tt} = \sigma(Y_1,\dots,Y_t)$. And we introduce

$$S_{Tt} = \left(T^{-1}(Y_{t-1}-Y_1)\phi_f(\Delta Y_t), T^{-1/2}b_{k_1}(\Delta Y_t),\dots,T^{-1/2}b_{k_m}(\Delta Y_t)\right)'$$

for $t = 2, \ldots, T$ and $T \in \mathbb{N}$. Notice that (see the proof of Part (ii) above) $\tilde{\Delta}^{(T)}(1) = \sum_{t=1}^{T} S_{Tt}$.

In the notation of Proposition 1 in Hallin, Van den Akker, and Werker (2015) we have, for $t \geq 2$,

(42)
$$LR_{Tt} = \frac{f\left(\Delta Y_t - w_{Tt}\right)\left(1 + \frac{1}{\sqrt{T}}\sum_{j=1}^{m}\eta_{k_j} b_{k_j}\left(\Delta Y_t - w_{Tt}\right)\right)}{f(\Delta Y_t)} \text{ with } w_{Tt} = \frac{h_T}{T}(Y_{t-1} - \mu),$$

Assumption 1 implies (see, e.g., Le Cam (1986, Section 17.3) and Le Cam and Yang (2000, Section 7.3)) that the mapping $e \mapsto f^{1/2}(e)$ is differentiable in quadratic mean:

$$\frac{\sqrt{f}(e-w)}{\sqrt{f}(e)} = 1 + \frac{1}{2}\left[\phi_f(e)w + r(e,w)\right],$$

where

(43) $$\mathrm{E}r^2(\varepsilon_1, w) = o_{\mathrm{P}}(w^2),$$

which implies, by Cauchy-Schwarz inequality,

(44) $$\mathrm{E}r(\varepsilon_1, w) = o_{\mathrm{P}}(w).$$

Let $B_{Tt} = T^{-1/2}\sum_{j=1}^{m}\eta_{k_j} b_{k_j}\left(\Delta Y_t - w_{Tt}\right)$, $B_{Tt}^0 = T^{-1/2}\sum_{j=1}^{m}\eta_{k_j} b_{k_j}\left(\Delta Y_t\right)$, and introduce

(45) $$R_{Tt}^b = 2\left(\sqrt{1 + B_{Tt}} - 1 - \frac{1}{2}B_{Tt}^0\right),$$

where, by Taylor's theorem (twice) and the assumption that the continuous derivatives of b_{k_j} are bounded, we have

(46) $$\max_{2 \leq t \leq T} |R_{Tt}^b| = o_{\mathrm{P}}(\frac{1}{T}).$$

Recall that $a_T = (h_T, \eta_{k_1}, \ldots, \eta_{k_m})'$ and $a = (h, \eta_{k_1}, \ldots \eta_{k_m})'$. We have, for $t \geq 2$,

(47)
$$\sqrt{LR_{Tt}} = \left(1 + \frac{1}{2}w_{Tt}\phi_f(\varepsilon_t) + \frac{1}{2}r(\varepsilon_t, w_{Tt})\right)\left(1 + \frac{1}{2}B_{Tt}^0 + \frac{1}{2}R_{Tt}^b\right) = 1 + \frac{1}{2}a_T' S_{Tt} + \frac{1}{2}R_{Tt},$$

with

(48) $$R_{Tt} = r(\varepsilon_t, w_{Tt}) + R_{Tt}^b + \frac{1}{2}(B_{Tt}^0 + R_{Tt}^b)\left(\phi_f(\varepsilon_t)w_{Tt} + r(\varepsilon_t, w_{Tt})\right).$$

So we can conclude that expansion (i) holds once we verify the conditions in Proposition 1 of Hallin, Van den Akker, and Werker (2015).

Condition (a). This is immediate as a_T converges by assumption.

Condition (b). Square-integrability follows from our assumption $f \in \mathcal{F}$. Display (2) in Condition (b) of Hallin, Van den Akker, and Werker (2015) follows immediately from the independence of ε_t and $\mathcal{F}_{T,t-1}$, $\mathrm{E}\phi_f(\varepsilon_t) = 0$, and $\mathrm{E}b_{k_j}(\varepsilon_t) = 0$, $j = 1, ..., m$. The second equation in Display (3) in Condition (b) is immediate as $J_T = \sum_{t=1}^T \mathrm{E}\left[S_{Tt}S'_{Tt}|\mathcal{F}_{t-1}\right] = \langle\tilde{\Delta}^{(T)}, \tilde{\Delta}^{(T)}\rangle_1 = O_{\mathrm{P}}(1)$ (see (38)). Next we verify the conditional Lindeberg condition (the first equation in Display (3)), which is, for all $\delta > 0$,

$$\sum_{t=2}^T \mathrm{E}\left[(a'_T S_{Tt})^2 1_{\left\{|a'_T S_{Tt}|>\delta\right\}}|\mathcal{F}_{t-1}\right] = o_{\mathrm{P}}(1). \tag{49}$$

Observe

$$\begin{aligned}
&\sum_{t=2}^T \mathrm{E}\left[(a'_T S_{Tt})^2 1_{\left\{|a'_T S_{Tt}|>\delta\right\}}|\mathcal{F}_{t-1}\right]\\
=\;& \sum_{t=2}^T \mathrm{E}\left[\left(\frac{h_T}{T}(Y_{t-1}-Y_1)\phi_f(\Delta Y_t) + \sum_{j=1}^m \frac{\eta_{k_j}}{\sqrt{T}} b_{k_j}(\Delta Y_t)\right)^2 1_{\left\{\left(a'_T S_{Tt}\right)^2>\delta^2\right\}}|\mathcal{F}_{t-1}\right]\\
\leq\;& (m+1)^2 \sum_{t=2}^T \mathrm{E}\left[\left(\frac{h_T^2}{T^2}(Y_{t-1}-Y_1)^2\phi_f^2(\Delta Y_t) 1_{\left\{(m+1)^2 h_T^2 (Y_{t-1}-Y_1)^2 \phi_f^2(\Delta Y_t)>\delta^2 T^2\right\}}\right)|\mathcal{F}_{t-1}\right]\\
&+ \sum_{j=1}^m (m+1)^2 \sum_{t=2}^T \mathrm{E}\left[\frac{\eta_{k_j}^2}{T} b_{k_j}^2(\Delta Y_t) 1_{\left\{(m+1)^2\eta_{k_j}^2 b_{k_j}^2(\Delta Y_t)>\delta^2 T\right\}}|\mathcal{F}_{t-1}\right].
\end{aligned}$$

To complete the proof, we just need to show separately that, for any given $\delta > 0$,

$$\sum_{t=2}^T \mathrm{E}\left[\left(\frac{h_T^2}{T^2}(Y_{t-1}-Y_1)^2\phi_f^2(\Delta Y_t) 1_{\left\{|(m+1)h_T(Y_{t-1}-Y_1)\phi_f(\Delta Y_t)|>\delta T\right\}}\right)|\mathcal{F}_{t-1}\right] = o_{\mathrm{P}}(1), \tag{50}$$

$$\sum_{t=2}^T \mathrm{E}\left[\frac{\eta_{k_j}^2}{T} b_{k_j}^2(\Delta Y_t) 1_{\left\{|(m+1)\eta_{k_j} b_{k_j}(\Delta Y_t)|>\delta\sqrt{T}\right\}}|\mathcal{F}_{t-1}\right] = o_{\mathrm{P}}(1). \tag{51}$$

Here, (50) and (51) can be shown in the same way as (41) is proved.

Condition (c). By (43) and (44), we have $\mathrm{E}[r^2(\varepsilon_t, w_{Tt})|\mathcal{F}_{t-1}] = o_{\mathrm{P}}(T^{-2})$ and $\mathrm{E}[r(\varepsilon_t, w_{Tt})|\mathcal{F}_{t-1}] = o_{\mathrm{P}}(T^{-1})$. Moreover, since all b_{k_j}s are bounded, we have $\max_{2\leq t\leq T}|B^0_{Tt}| = O_{\mathrm{P}}(T^{-1/2})$

and $\sum_{t=1}^{T} E[B_{Tt}^0|\mathcal{F}_{t-1}] = O_{\mathrm{P}}(1)$. Together with (46), this yields

$$\sum_{t=2}^{T} \mathrm{E}\left[(r(\varepsilon_t, w_{Tt}) + R_{Tt}^b)^2|\mathcal{F}_{t-1}\right] = o_{\mathrm{P}}(1),$$
$$\sum_{t=2}^{T} \mathrm{E}\left[(r(\varepsilon_t, w_{Tt}) + R_{Tt}^b)(B_{Tt}^0 + R_{Tt}^b)\left(\phi_f(\varepsilon_t)w_{Tt} + r(\varepsilon_t, w_{Tt})\right)|\mathcal{F}_{t-1}\right] = o_{\mathrm{P}}(1),$$
$$\sum_{t=2}^{T} \mathrm{E}\left[\left((B_{Tt}^0 + R_{Tt}^b)\left(\phi_f(\varepsilon_t)w_{Tt} + r(\varepsilon_t, w_{Tt})\right)\right)^2|\mathcal{F}_{t-1}\right] = o_{\mathrm{P}}(1),$$

and, thus, by (48), we have

$$\sum_{t=2}^{T} \mathrm{E}\left[R_{Tt}^2|\mathcal{F}_{t-1}\right] = o_{\mathrm{P}}(1).$$

This establishes Display (4). As we assumed the density f to be strictly positive, Display (5) is immediate by plugging in (42) to its left-hand side.

Condition (d). This follows easily from

$$\log \frac{f_{\eta}^{(T)}(Y_1 - \mu)}{f(Y_1 - \mu)} = \log\left[1 + \frac{1}{\sqrt{T}}\sum_{j=1}^{m} \eta_{k_j} b_{k_j}(\varepsilon_1)\right] = o_{\mathrm{P}}(1).$$

This completes the proof. *Q.E.D.*

Proof of Theorem 3.1:

Let G be the group of translations g_η with $\eta \in c_{00}$ defined in (13). Invariance of M has been shown in Section 3.3. In order to prove that M is maximal invariant, following the idea in Section 6.2 of Lehmann and Romano (2005), we establish that $M((W_\varepsilon(s), (B_{b_k}(s))_{k\in\mathbb{N}})', s \in [0,1]) = M((\widetilde{W}_\varepsilon(s), (\widetilde{B}_{b_k}(s))_{k\in\mathbb{N}})', s \in [0,1])$ implies $\widetilde{W}_\varepsilon(s) = W_\varepsilon(s)$ and $\widetilde{W}_{b_k}(s) = g_{\eta_k}(W_{b_k}(s))$ with $s \in [0,1]$, for some $g_\eta = (g_{\eta_k})_{k\in\mathbb{N}} \in G$.

Suppose $M((W_\varepsilon(s), (B_{b_k}(s))_{k\in\mathbb{N}})', s \in [0,1]) = M((\widetilde{W}_\varepsilon(s), (\widetilde{B}_{b_k}(s))_{k\in\mathbb{N}})', s \in [0,1])$, that is

$$W_\varepsilon(s) = \widetilde{W}_\varepsilon(s),$$
$$B_{b_k}(s) = \widetilde{B}_{b_k}(s), \quad k \in \mathbb{N}.$$

This implies, for $\eta_k = W_{b_k}(1) - \widetilde{W}_{b_k}(1)$,

$$W_\varepsilon(s) - \widetilde{W}_\varepsilon(s) = 0,$$
$$W_{b_k}(s) - \widetilde{W}_{b_k}(s) = \eta_k s, \quad k \in \mathbb{N}.$$

Hence $\widetilde{W}_\varepsilon(s) = W_\varepsilon(s)$ and $\widetilde{W}_{b_k}(s) = g_{\eta_k}(W_{b_k}(s))$ with $s \in [0,1]$, which completes the proof. *Q.E.D.*

Proof of Proposition 4.1:

Recall $\mathcal{M}_g = \sigma(W_\varepsilon, B_{\phi_g}) \subseteq \mathcal{M} = \sigma(W_\varepsilon, B_{b_k}; k \in \mathbb{N})$, the likelihood ratio of $\mathcal{M}_g$ can be derived by taking the expectation of the likelihood ratio of $\mathcal{M}$ conditional on the information $\mathcal{M}_g$. We find

$$\begin{aligned}
\frac{\mathrm{d}\mathbb{P}_h^{\mathcal{M}_g}}{\mathrm{d}\mathbb{P}_0^{\mathcal{M}_g}} &= \mathbb{E}_0\left[\frac{\mathrm{d}\mathbb{P}_h^{\mathcal{M}}}{\mathrm{d}\mathbb{P}_0^{\mathcal{M}}}\Big|\mathcal{M}_g\right] \\
&= \mathbb{E}_0\Bigg[\exp\Bigg(h\left\{\int_0^1 W_\varepsilon(s)\mathrm{d}B_{\phi_f}(s) + W_\varepsilon(1)\int_0^1 W_\varepsilon(s)\mathrm{d}s\right\} \\
&\qquad - \frac{1}{2}h^2\left\{J_f\int_0^1 W_\varepsilon^2(s)\mathrm{d}s - \left(\int_0^1 W_\varepsilon(s)\mathrm{d}s\right)^2(J_f-1)\right\}\Bigg)\Big|\mathcal{M}_g\Bigg].
\end{aligned}$$

Based on the covariance matrix (21) and using $\lambda = (J_{fg}\sigma_{\varepsilon\phi_g} - \sigma^2_{\varepsilon\phi_g})/(J_g - \sigma^2_{\varepsilon\phi_g})$, we have the decomposition $W_{\phi_f} = (1-\lambda)W_\varepsilon + \lambda W_{\phi_g}/\sigma_{\varepsilon\phi_g} + W_\ddagger$, where $W_\ddagger$ is a Brownian motion (not necessarily standard) independent of both W_ε and W_{ϕ_g}. Together with the decomposition $W_{\phi_g}/\sigma_{\varepsilon\phi_g} = W_\varepsilon + \sqrt{J_g/\sigma^2_{\varepsilon\phi_g} - 1}W_\perp$, we have

$$W_{\phi_f} = W_\varepsilon + \lambda\sqrt{\frac{J_g}{\sigma^2_{\varepsilon\phi_g}} - 1}W_\perp + W_\ddagger.$$

Define $B_\varepsilon = B^{W_\varepsilon}$, $B_\perp = B^{W_\perp}$, and $B_\ddagger = B^{W_\ddagger}$. It follows

$$B_{\phi_f} = B_\varepsilon + \lambda\sqrt{\frac{J_g}{\sigma^2_{\varepsilon\phi_g}} - 1}B_\perp + B_\ddagger.$$

Plugging this into the previous equation leads to

$$\begin{aligned}
&\frac{\mathrm{d}\mathbb{P}_h^{\mathcal{M}_g}}{\mathrm{d}\mathbb{P}_0^{\mathcal{M}_g}}\\
=&\mathbb{E}_0\Bigg[\exp\Bigg(h\Bigg\{\int_0^1 W_\varepsilon(s)\Bigg(\mathrm{d}B_\varepsilon(s)+\lambda\sqrt{\frac{J_g}{\sigma^2_{\varepsilon\phi_g}}-1}\mathrm{d}B_\perp(s)+\mathrm{d}B_\ddagger(s)\Bigg)\\
&+W_\varepsilon(1)\int_0^1 W_\varepsilon(s)\mathrm{d}s\Bigg\}-\frac{1}{2}h^2\left\{J_f\int_0^1 W_\varepsilon^2(s)\mathrm{d}s-\left(\int_0^1 W_\varepsilon(s)\mathrm{d}s\right)^2(J_f-1)\right\}\Bigg)\Bigg|\mathcal{M}_g\Bigg]\\
=&\mathbb{E}_0\Bigg[\exp\Bigg(h\left\{\int_0^1 W_\varepsilon(s)\mathrm{d}W_\varepsilon(s)+\lambda\sqrt{\frac{J_g}{\sigma^2_{\varepsilon\phi_g}}-1}\int_0^1 W_\varepsilon(s)\mathrm{d}B_\perp(s)+\int_0^1 W_\varepsilon(s)\mathrm{d}B_\ddagger(s)\right\}\\
&-\frac{1}{2}h^2\left\{J_f\int_0^1 W_\varepsilon^2(s)\mathrm{d}s-\left(\int_0^1 W_\varepsilon(s)\mathrm{d}s\right)^2(J_f-1)\right\}\Bigg)\Bigg|\mathcal{M}_g\Bigg]\\
=&\mathbb{E}_0\Bigg[\exp\Bigg(h\left\{\Delta_\varepsilon+\lambda\Delta_\perp+\Delta_\ddagger\right\}-\frac{1}{2}h^2\left\{\langle\Delta_\varepsilon+\lambda\Delta_\perp+\Delta_\ddagger,\Delta_\varepsilon+\lambda\Delta_\perp+\Delta_\ddagger\rangle\right\}\Bigg)\,|\mathcal{M}_g\Bigg]
\end{aligned}$$

where Δ_ε and $\Delta_\perp$ are defined in the present proposition, and $\Delta_\ddagger := \int_0^1 W_\varepsilon(s)\mathrm{d}B_\ddagger(s)$. Under $\mathbb{P}_{0,0}$ the process $B_\ddagger$ is independent of W_ε and B_{ϕ_g} (henceforth $B_\perp$). Consequently, $\langle\Delta_\varepsilon,\Delta_\ddagger\rangle=0$ and $\langle\Delta_\perp,\Delta_\ddagger\rangle=0$. Noting that Δ_ε, $\Delta_\perp$, $\langle\Delta_\varepsilon\rangle$, $\langle\Delta_\perp\rangle$ and $\langle\Delta_\ddagger\rangle$ are all $\mathcal{M}_g$-measurable, we thus obtain,

$$\begin{aligned}
\frac{\mathrm{d}\mathbb{P}_h^{\mathcal{M}_g}}{\mathrm{d}\mathbb{P}_0^{\mathcal{M}_g}}&=\mathbb{E}_0\Bigg[\exp\Bigg(h\left\{\Delta_\varepsilon+\lambda\Delta_\perp+\Delta_\ddagger\right\}-\frac{1}{2}h^2\left\{\langle\Delta_\varepsilon\rangle+\lambda^2\langle\Delta_\perp\rangle+\langle\Delta_\ddagger\rangle\right\}\Bigg)\,|\mathcal{M}_g\Bigg]\\
&=\exp\Bigg(h\left\{\Delta_\varepsilon+\lambda\Delta_\perp\right\}-\frac{1}{2}h^2\left\{\langle\Delta_\varepsilon\rangle+\lambda^2\langle\Delta_\perp\rangle+\langle\Delta_\ddagger\rangle\right\}\Bigg)\mathbb{E}_0\Big[\exp\left(h\Delta_\ddagger\right)|\mathcal{M}_g\Big]\\
&=\exp\Bigg(h\left\{\Delta_\varepsilon+\lambda\Delta_\perp\right\}-\frac{1}{2}h^2\left\{\langle\Delta_\varepsilon\rangle+\lambda^2\langle\Delta_\perp\rangle\right\}\Bigg).
\end{aligned}$$

The last equality holds since $\mathbb{E}_0\Big[\exp\left(h\Delta_\ddagger\right)|\mathcal{M}_g\Big]=\exp\left(\frac{1}{2}h^2\langle\Delta_\ddagger\rangle\right)$.

Q.E.D.

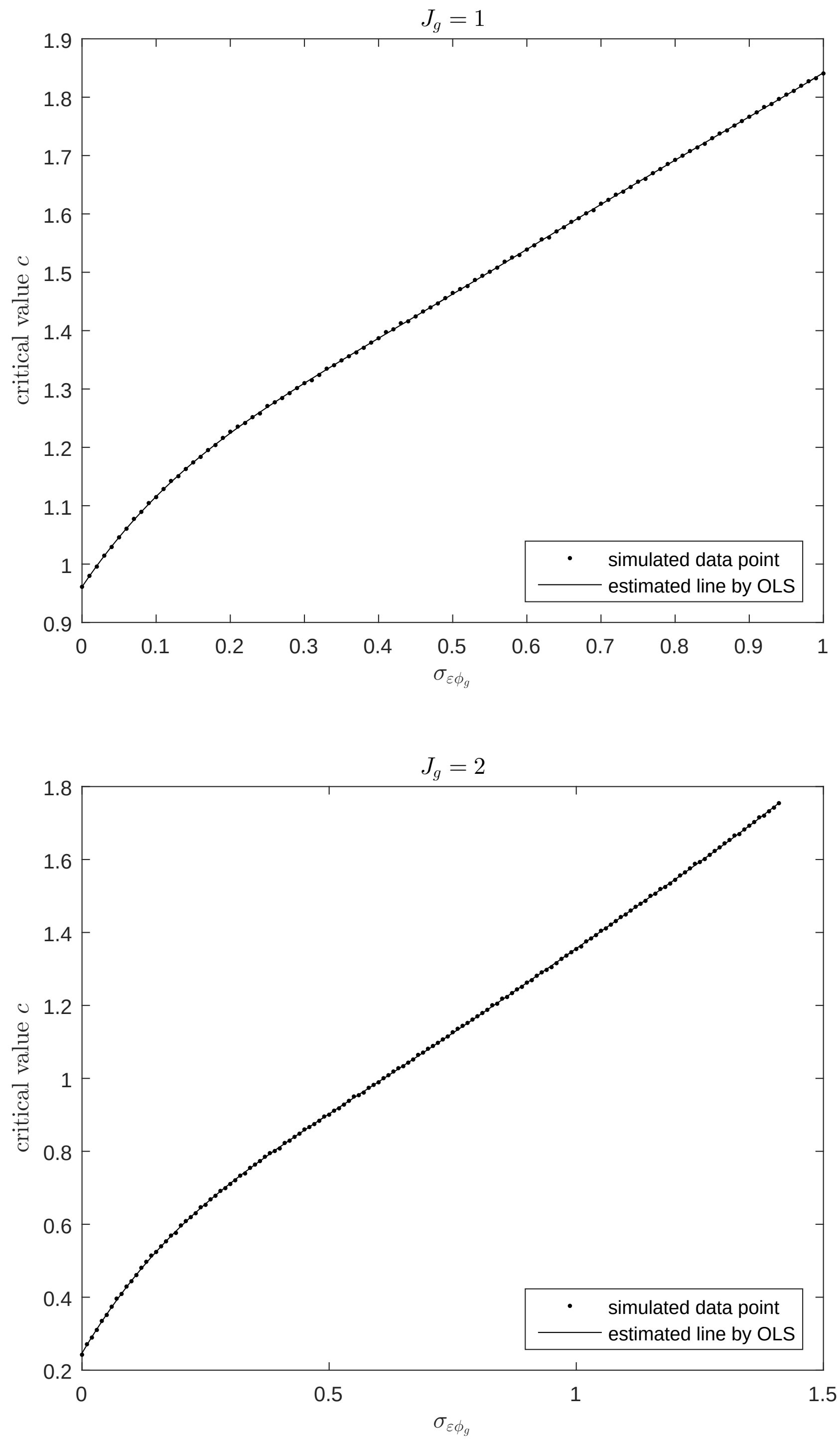

FIGURE 1.— Fitted regression lines for the critical value functions in Table I. Each critical value is simulated using Brownian motions with a time step 0.001 over the interval $[0, 1]$ and 10,000,000 replications.

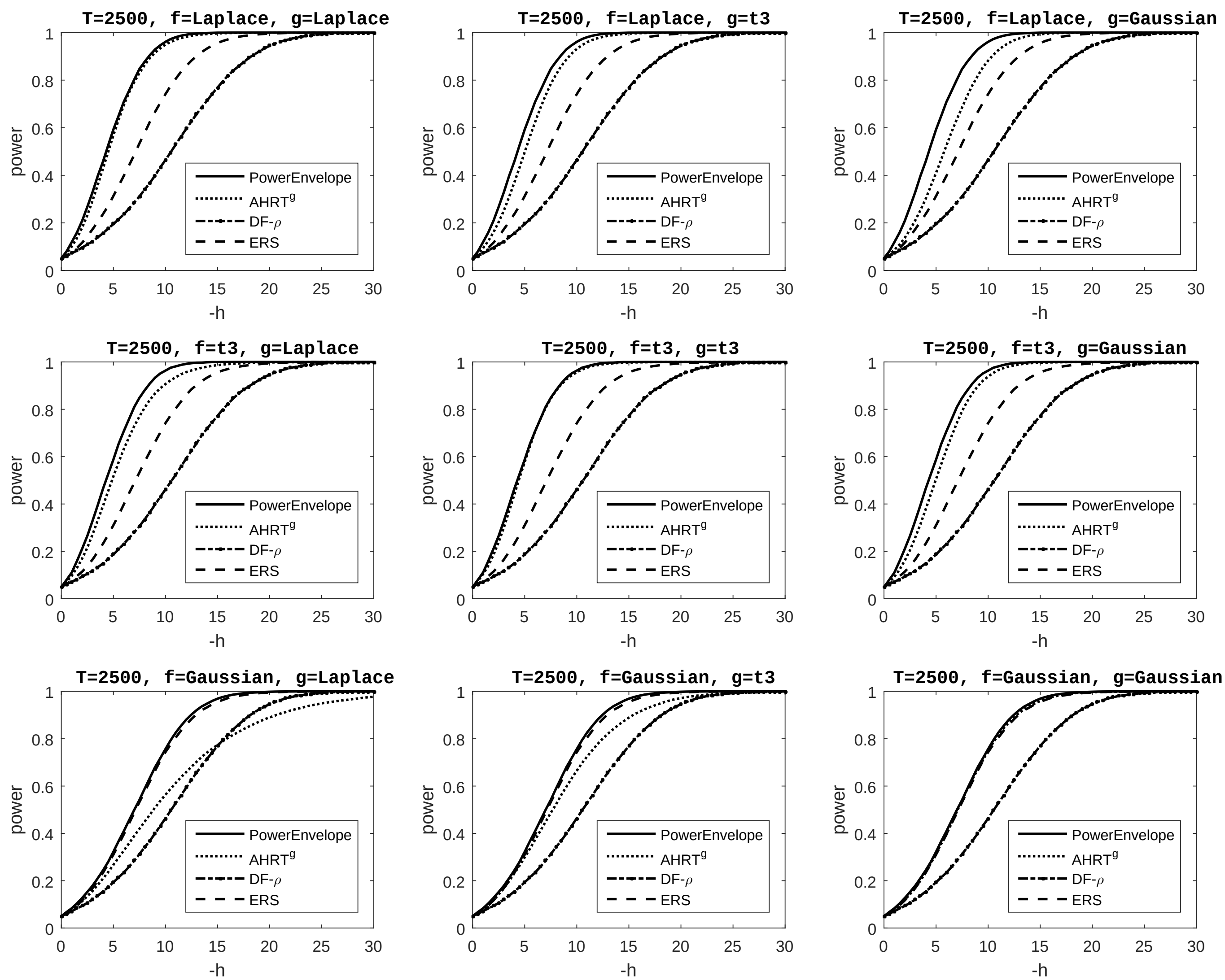

FIGURE 2.— Asymptotic power functions of the Hybrid Rank Test for various reference densities g and other selected unit root tests under the true innovation densities f: Gaussian, Laplace, Student's t_3.

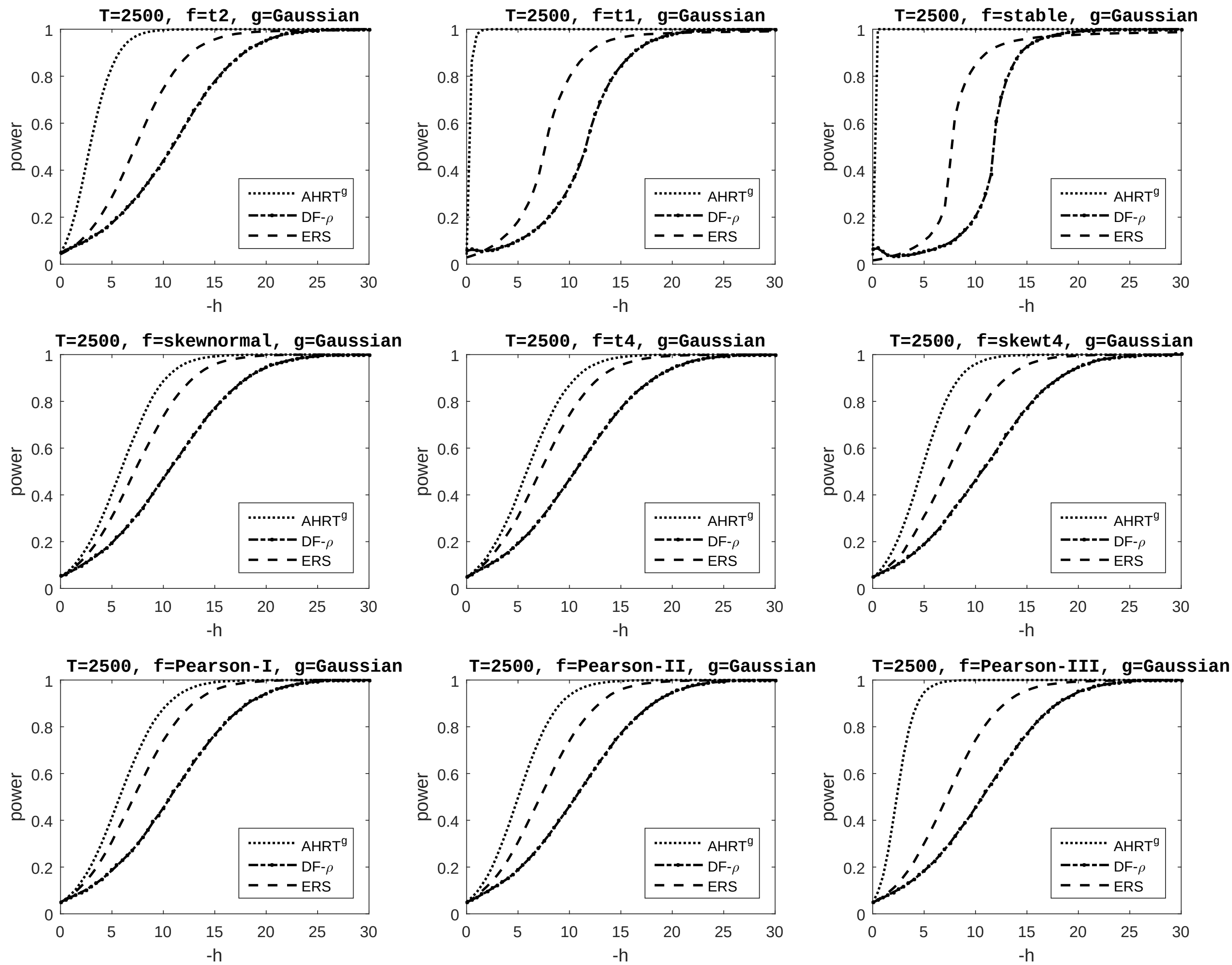

FIGURE 3.— Illustration of the Chernoff-Savage result. The figure shows asymptotic power functions for the Gaussian Hybrid Rank Test and selected unit root tests under various true innovation densities f.

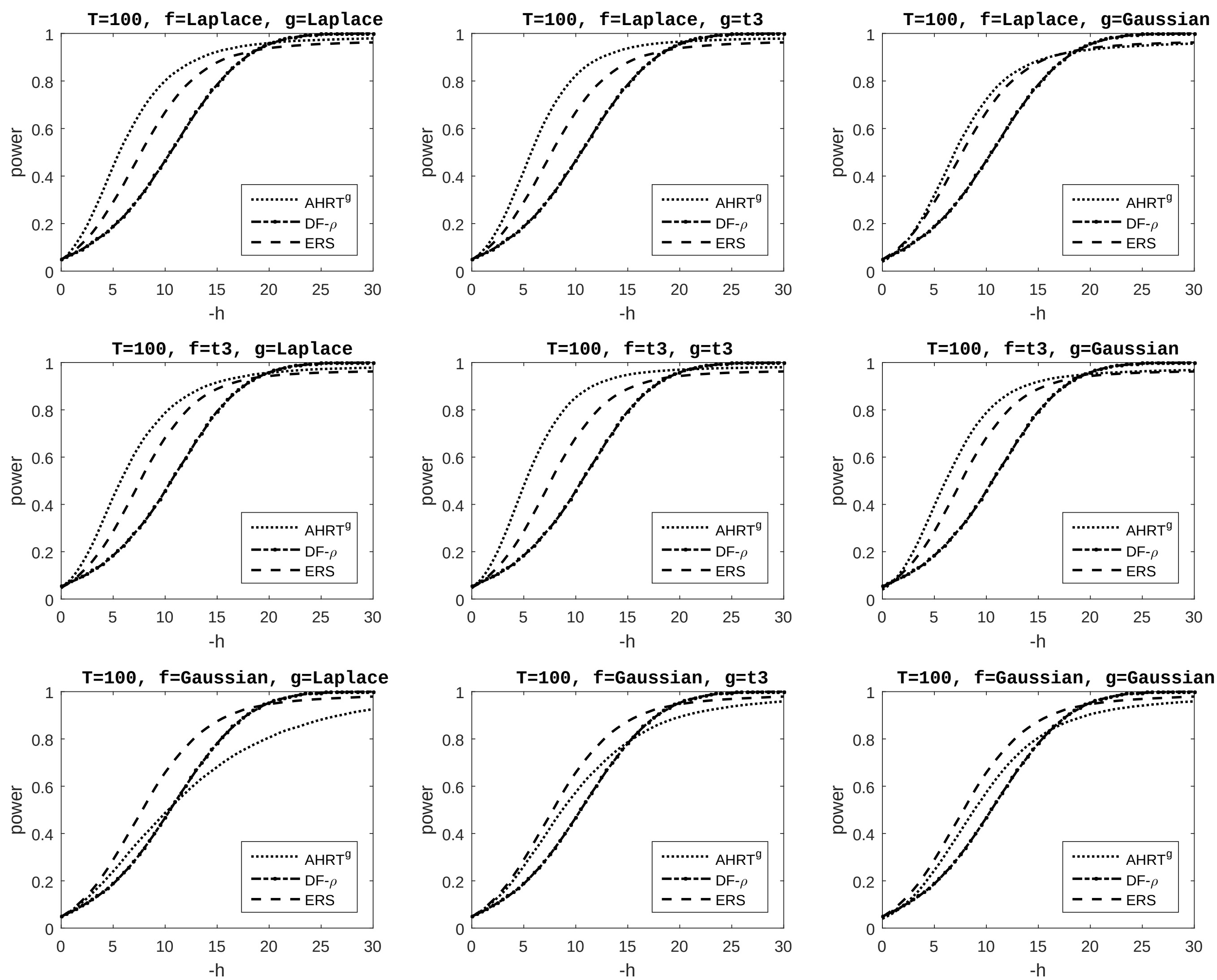

FIGURE 4.— Small-sample ($T = 100$) power functions of selected unit root tests and various true innovation densities: Gaussian, Laplace, Student t_3.

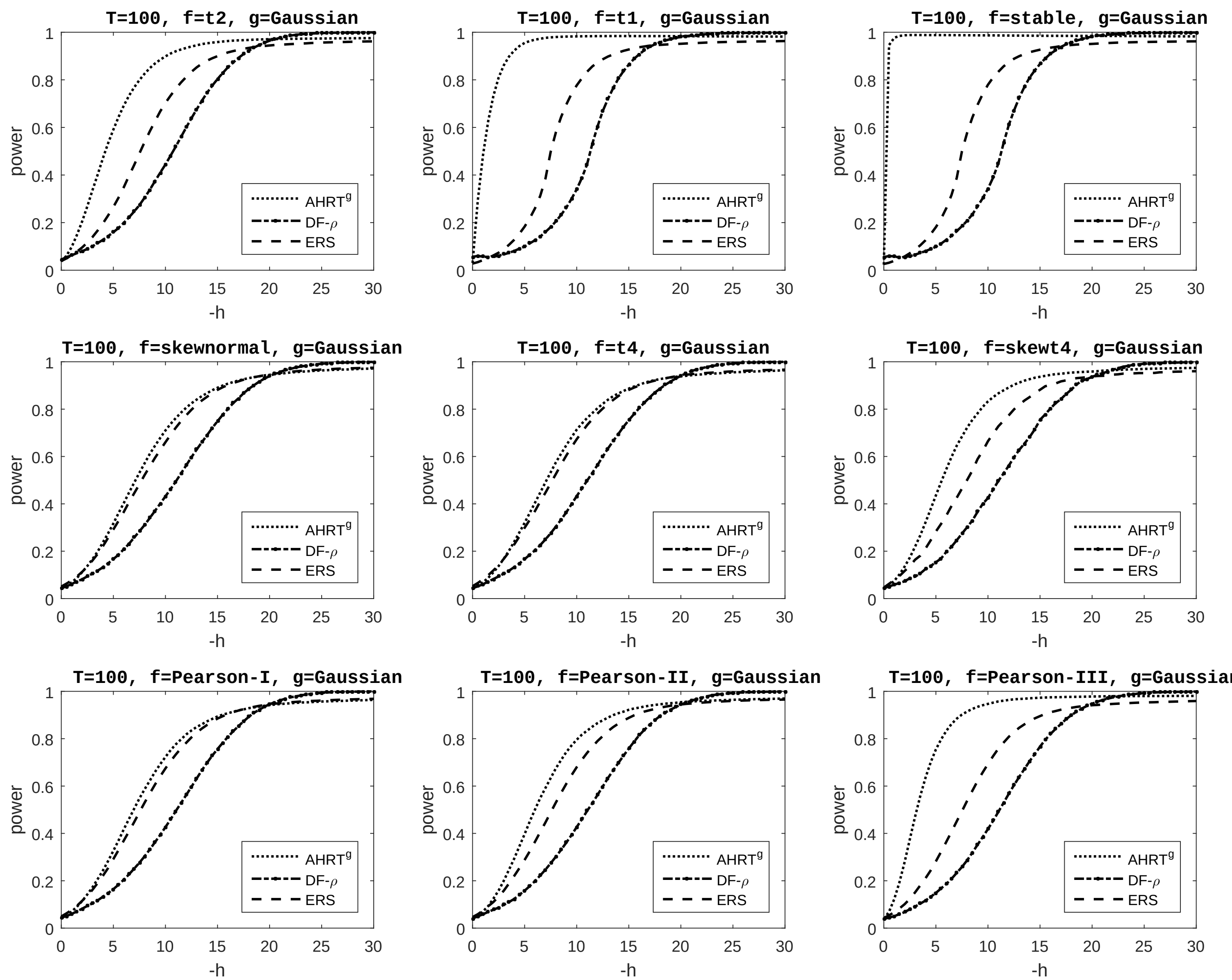

FIGURE 5.— Small-sample ($T = 100$) power functions of selected unit root tests and various true innovation densities.

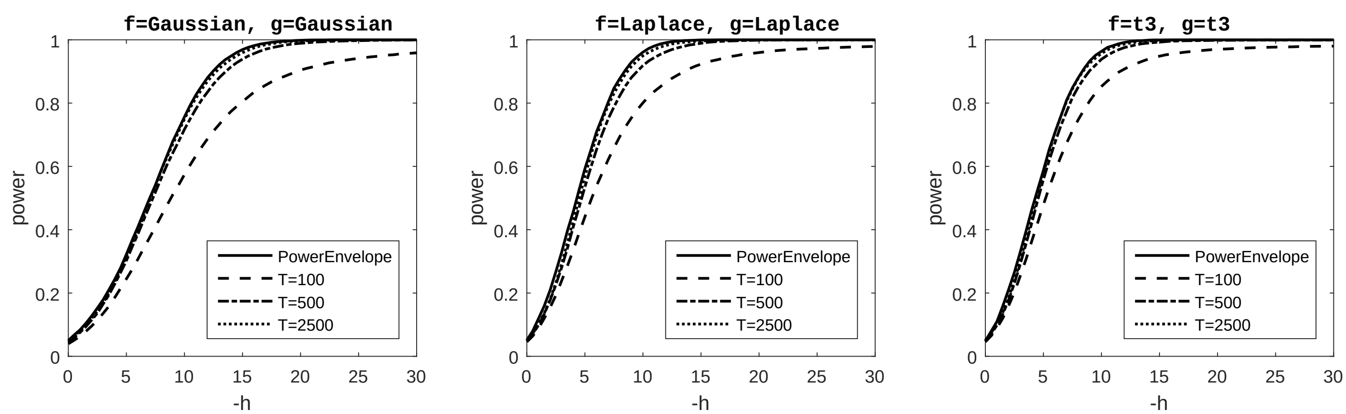

FIGURE 6.— Powers of HRT^g when $g = f$ with different sample sizes.