The Annals of Statistics
2019, Vol. 47, No. 5, 2601–2638
https://doi.org/10.1214/18-AOS1758
© Institute of Mathematical Statistics, 2019

SEMIPARAMETRICALLY POINT-OPTIMAL HYBRID RANK TESTS FOR UNIT ROOTS

BY BO ZHOU*, RAMON VAN DEN AKKER† AND BAS J. M. WERKER†

Hong Kong University of Science and Technology and Tilburg University†*

We propose a new class of unit root tests that exploits invariance properties in the Locally Asymptotically Brownian Functional limit experiment associated to the 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, that is, have the correct (asymptotic) size, irrespective of the true innovation density. For a correctly specified reference density, our test is point-optimal and nearly efficient. For arbitrary reference densities, we establish a Chernoff–Savage-type result, that is, our test performs as well as commonly used tests under Gaussian innovations but has improved power under other, for example, fat-tailed or skewed, innovation distributions. To avoid nonparametric estimation, we propose a simplified version of our test that exhibits the same asymptotic properties, except for the Chernoff–Savage result that we are only able to demonstrate by means of simulations.

1. Introduction. The 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 earlier 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 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 possibly serially correlated errors. The innovations driving these serially correlated errors are i.i.d., 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

Received March 2017; revised June 2018.
MSC2010 subject classifications. Primary 62G10, 62G20; secondary 62P20, 62M10.
Key words and phrases. Unit root test, semiparametric power envelope, limit experiment, LABF, maximal invariant, rank statistic.

Locally Asymptotically Brownian Functional (LABF) limit experiments (in the Le Cam sense; see Jeganathan (1995)). 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 more difficult 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 nonparametric score function estimator which complicates its implementation. The point-optimal tests proposed in the present paper only require nonparametric estimation of a real-valued cross-information factor and we also provide a simplified version that avoids any nonparametric estimation.

The main contribution of this manuscript is twofold. First, we derive the semiparametric power envelopes of unit root tests with serially correlated errors for two cases: symmetric or possibly nonsymmetric innovation distributions (Section 3). Our method of derivation is novel and exploits the 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 an infinitely-dimensional 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 parameter in this limit experiment. Within the limit experiment, Section 3.3 derives the maximal invariant, that is, a reduction of the data which is invariant with respect to the nuisance parameters (i.e., the unknown drift in the limiting Ornstein–Uhlenbeck experiment). It turns out that this maximal invariant takes a rather simple form: all processes associated to density perturbations have to be replaced by their associated bridges (i.e., consider $W(s) - sW(1)$ for the process $W(s)$ with $s \in [0, 1]$). The power envelopes for invariant tests in the limit experiment then readily follow from the Neyman–Pearson lemma. An application of the Asymptotic Representation Theorem (see, e.g., Theorem 15.1 in van der Vaart (1998)) subsequently yields the local asymptotic power envelope (Theorem 3.3). In case the innovation density is known to be symmetric, the semiparametric power envelope coincides with the parametric power envelope. This implies the existence of an adaptive testing procedure (see also Jansson (2008)). Moreover, 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. Also, the analysis gives an alternative interpretation of the test proposed in Elliott, Rothenberg and Stock (1996)—the ERS test—as this test is also based on an invariant, though not the maximal one (see Remark 3.3).

As a second contribution, we provide two new classes of easy-to-implement unit root tests that are semiparametrically optimal in the sense that their asymptotic power curves are tangent to the associated semiparametric power envelopes (Section 4.1). The form of the maximal invariant developed before suggests how

to construct such tests based on the ranks/signed-ranks (depending on whether the innovation density is known to be symmetric or not) of the increments of the observations, the average of these increments, and an assumed reference density g. These tests are semiparametric in the sense that the reference density need not equal the true innovation density, while they are still valid (i.e., provide the correct asymptotic size). The reference density is not restricted to be Gaussian, which it generally needs to be in more classical QMLE results. When the reference density is correctly specified (i.e., g happens to be equal to the true density f), the asymptotic power curve of our test is tangent to the semiparametric power envelope, and this in turn gives the optimality property. A feasible version of the oracle test using $g = f$ is obtained by using a nonparametrically estimated density $\hat{f}$, of which the corresponding simulation results are provided in Section 5.

In relation to the classical literature on efficient rank-based testing (for instance, Hájek and Šidák (1967), Hallin and Puri (1988) and Hallin, van den Akker and Werker (2011)) our approach can be interpreted as follows. In the aforementioned papers, the invariance arguments (i.e., using the ranks of the innovations) are applied in the sequence of models at hand. We, on the other hand, only apply the invariance arguments in the limit experiment. In this way, we can extend these ideas to non-LAN experiments. For the LAN case, both approaches would effectively lead to the same results. Our tests, despite the absence of a LAN structure, satisfy a Chernoff and Savage (1958)-type result (Corollary 4.1): for any reference density our test outperforms, at any true density, its classical counterpart which, in this case, is the ERS test. We provide, in Section 4.2, even simpler alternative classes of tests that require no nonparametric estimations at all. These (simplified) classes of tests coincide with their corresponding originals for correctly specified reference density and, hence, share the same optimality properties. In case of misspecified reference density, the alternative classes still seem to enjoy the Chernoff–Savage-type property, though only for a Gaussian reference density. This is in line with some 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 envelopes. The classes of hybrid rank tests we propose are 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 Supplementary Material (Zhou, van den Akker and Werker (2019)).

2. The model. Consider observations $Y_1, \dots, Y_T$ generated from the classical component specification, for $t \in \mathbb{Z}_+$,

$$Y_t = \mu + X_t, \tag{2.1}$$

$$X_t = \rho X_{t-1} + v_t, \tag{2.2}$$

$$\Gamma(L) v_t = \varepsilon_t, \tag{2.3}$$

where $v_0 = v_{-1} = \cdots = v_{1-p} = 0$, the innovations $\{\varepsilon_t\}$ form an i.i.d. sequence defined for $t \in \mathbb{Z}$ with density f, and $\Gamma(L)$ is the AR(p) lag polynomial. Moreover, it is assumed that $Y_0 = \mu$. We impose the following assumptions on this innovation density.

ASSUMPTION 1. (a) The density f is absolutely continuous with a.e. derivative f', that is, for all $a < b$ we have

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

(b) $\mathrm{E}_f[\varepsilon_t] = \int e f(e)\,\mathrm{d}e = 0$ 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 positive, that is, $f > 0$.

The imposed smoothness on f is mild and standard (see, e.g., Le Cam (1986), van der Vaart (1998)). 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 intercept 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 J_f in (c) has been standardized by premultiplying with the variance σ_f^2, so that it becomes scale invariant (i.e., invariant with respect to σ_f). In other words, J_f only depends on the shape of the density f and not on its variance σ_f^2. The positivity of the density f in (d) is mainly made for notational convenience.

The assumption on the initial condition, $v_0 = v_{-1} = \cdots = v_{1-p} = 0$, 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 nonnegligible statistical information. Nevertheless, it is still stronger than

[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 (2001) and Callegari, Cappuccio and Lubian (2003).

necessary for the sake of simplicity, and it can be relaxed to the level of generality in Elliott, Rothenberg and Stock (1996).

Let $\mathfrak{F}$ denote the set of densities satisfying Assumption 1. We also investigate in the present paper the special case of symmetric densities. For that purpose, we denote by $\mathfrak{F}_{\mathbb{S}}$ the set of densities which satisfy Assumption 1 and, at the same time, are symmetric about zero. Of course, it follows $\mathfrak{F}_{\mathbb{S}} \subset \mathfrak{F}$.

With respect to the autocorrelation structure $\Gamma(L)$, we impose the following assumption.

ASSUMPTION 2. The lag polynomial $\Gamma(z) := 1 - \Gamma_1 z - \cdots - \Gamma_p z^p$ is of finite order $p \in \mathbb{Z}_+$ and satisfies $\min_{|z| \in \mathbb{C}: |z| \leq 1} |\Gamma(z)| > 0$.

Let $\mathfrak{G} \subset \mathbb{R}^p$ denote the set of $(\Gamma_1, \ldots, \Gamma_p)'$ such that the induced lag polynomial satisfies Assumption 2. For mathematical convenience, we restrict the lag polynomial to be of finite order p. One may expect many of the results in the present paper to extend to the case $p = \infty$ (see, e.g., Jeganathan (1997)).

The main goal of this paper is to develop tests, with optimality features, for the semiparametric unit root hypothesis

$$\begin{aligned} &\mathrm{H}_0 : \rho = 1 \qquad (\mu \in \mathbb{R}, \Gamma \in \mathfrak{G}, f \in \mathfrak{F}) \quad \text{vs.} \\ &\mathrm{H}_a : \rho < 1 \qquad (\mu \in \mathbb{R}, \Gamma \in \mathfrak{G}, f \in \mathfrak{F}), \end{aligned}$$

that is, apart from Assumptions 1, no further structure is imposed on f, the intercept μ and the autocorrelation structure $\Gamma(L)$.

In the following section, we derive the (asymptotic) power envelope of tests that are (locally and asymptotically) invariant with respect to the nuisance parameters μ, Γ and f. We consider both the nonsymmetric ($f \in \mathfrak{F}$) and the symmetric case ($f \in \mathfrak{F}_{\mathbb{S}}$). 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 a Chernoff–Savage type result.

3. The power envelope for invariant tests. This section first introduces some notations and preliminaries (Section 3.1). Afterwards, we will derive the limit experiment (in the Le Cam sense) corresponding to the component unit root model (2.1)–(2.3) 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. At last, in Section 3.4, we exploit the Asymptotic Representation Theorem to translate these results to obtain (asymptotically) optimal invariant test in the sequence of unit root models. Again we consider both the case of unrestricted densities $f \in \mathfrak{F}$ and that of symmetric densities $f \in \mathfrak{F}_{\mathbb{S}}$.

3.1. *Preliminaries.* We first introduce local reparameterizations for the parameter of interest ρ and the nuisance autocorrelation structure $\Gamma(L)$. Then we discuss a convenient parametrization of perturbations to the innovation density f 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 all parameters but ρ, 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.

Local reparameterizations of ρ and $\Gamma(L)$. 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, that is, 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, that is,

$$\rho = \rho_h^{(T)} = 1 + \frac{h}{T}. \tag{3.1}$$

The appropriate local reparameterization for the lag polynomial $\Gamma(L)$ is of a traditional form with rate $\sqrt{T}$, that is,

$$\Gamma_\gamma^{(T)}(L) = \Gamma(L) + \frac{\gamma(L)}{\sqrt{T}}, \tag{3.2}$$

where the local perturbation γ is defined by $\gamma(L) := -\gamma_1 L - \cdots - \gamma_p L^p$ with local parameter $\gamma := (\gamma_1, \dots, \gamma_p)' \in \mathbb{R}^p$. As $\mathfrak{G}$ is open, $\Gamma_\gamma^{(T)} \in \mathfrak{G}$ for T large enough.

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

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

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 $\mathrm{L}_2^{0,f}$ (see, e.g., Rudin (1987), Theorem 3.14). This basis can be chosen such that $b_k \in C_{2,b}(\mathbb{R})$, for all k, that is, each b_k is bounded and two times continuously differentiable with bounded derivatives. Moreover, $\mathrm{E}_f b_k(\varepsilon) = 0$ and $\mathrm{Var}_f\, b_k(\varepsilon) = 1$. 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 $\eta := (\eta_k)_{k\in\mathbb{N}} \in \ell_2 = \{(x_k)_{k\in\mathbb{N}} | \sum_{k=1}^\infty x_k^2 <$

$\infty\}$. 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, that is,

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

Of course, c_{00} is a dense subspace of ℓ_2. Given the orthonormal basis b_k and $\eta \in c_{00}$, we introduce the following perturbation to the density f:

$$f_\eta^{(T)}(e) = f(e)\left(1 + \frac{1}{\sqrt{T}} \sum_{k=1}^{\infty} \eta_k b_k(e)\right), \qquad e \in \mathbb{R}. \tag{3.3}$$

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.

For the symmetric case, we can assume that the perturbations b_k, $k \in \mathbb{N}$, are also symmetric about zero.

The following proposition shows that the perturbations, both for the nonsymmetric and for the symmetric case, are valid in the sense that they satisfy the conditions on the innovation density that we imposed throughout on the model (Assumption 1). The proof is organized in the Supplementary Material (Zhou, van den Akker and Werker (2019)).

PROPOSITION 3.1. *Let $f \in \mathfrak{F}$ and suppose $\eta \in c_{00}$. Then there exists $T' \in \mathbb{N}$ such that for all $T \geq T'$ we have $f_\eta^{(T)} \in \mathfrak{F}$. If we further restrict $f \in \mathfrak{F}_{\mathbb{S}}$ and b_k is chosen symmetric about zero for $k \in \mathbb{N}$, then there exists $T'' \in \mathbb{N}$ such that for all $T \geq T''$ we have $f_\eta^{(T)} \in \mathfrak{F}_{\mathbb{S}}$.*

REMARK 3.1. In semiparametric statistics, one typically parametrizes perturbations to a density by a so-called "nonparametric" score function $h \in \mathrm{L}_2^{0,f}$, that is, the 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 simultaneously 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 (3.3) 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 in the Supplementary Material (Zhou, van den Akker and Werker (2019)).

Define, for $s \in [0, 1]$, the partial sum processes

$$W_{\varepsilon}^{(T)}(s) = \frac{1}{\sqrt{T}} \sum_{t=p+2}^{\lfloor sT \rfloor} \frac{\Gamma(L)\Delta Y_t}{\sigma_f},$$

$$W_{\phi_f}^{(T)}(s) = \frac{1}{\sqrt{T}} \sum_{t=p+2}^{\lfloor sT \rfloor} \sigma_f \phi_f\big(\Gamma(L)\Delta Y_t\big),$$

$$W_{\Gamma}^{(T)}(s) = \frac{1}{\sqrt{T}} \sum_{t=p+2}^{\lfloor sT \rfloor} (\Delta Y_{t-1}, \dots, \Delta Y_{t-p})' \phi_f\big(\Gamma(L)\Delta Y_t\big) \in \mathbb{R}^p,$$

$$W_{b_k}^{(T)}(s) = \frac{1}{\sqrt{T}} \sum_{t=p+2}^{\lfloor sT \rfloor} b_k\big(\Gamma(L)\Delta Y_t\big), \qquad k \in \mathbb{N}.$$

Note that we pick the starting point of the sums at $t = p + 2$, so that these partial sum processes are (maximally) invariant with respect to the intercept μ (otherwise, e.g., for $t = p + 1$, the term $\Gamma(L)\Delta Y_t$ contains $\Delta Y_1 = Y_1 - \mu$).

Using Assumption 1 we find, under the null hypothesis, joint weak convergence of observation processes $W_{\varepsilon}^{(T)}$, $W_{\phi_f}^{(T)}$, $W_{\Gamma}^{(T)}$ and $W_{b_k}^{(T)}$ to Brownian motions that we denote by W_{ε}, W_{ϕ_f}, W_{Γ} and W_{b_k}, respectively.[2] These limiting Brownian motions are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P}_{0,0,0})$. Let us already mention that we will introduce a collection of probability measures $\mathbb{P}_{h,\gamma,\eta}$, on $(\Omega, \mathcal{F})$, representing the limit experiment, in Section 3.2. We use the notational convention that probability measures related to the limit experiment (i.e., to the "W-processes") are denoted by $\mathbb{P}$, while probability measures related to the finite-sample unit root model, that is, observing $Y_1, \dots, Y_T$, will be denoted by $\mathrm{P}^{(T)}$.

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, under the null hypothesis, to the associated stochastic integral with the limiting Brownian motions, that is, 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 can be found in the proof of Proposition 3.2 in the Supplementary Material (Zhou, van den Akker and Werker (2019)).

Behavior of W_{ε}, W_{ϕ_f}, W_{Γ} and W_{b_k} under $\mathbb{P}_{0,0,0}$ when $f \in \mathfrak{F}$. As ε and $b_k(\varepsilon)$ are orthogonal for each k, it holds that W_{ε} and W_{b_k}, $k \in \mathbb{N}$, are all mutually independent. Moreover, we have

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

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

As $\phi_f(\varepsilon)$ is the score of the location model, it is well known (see, e.g., 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, for $f \in \mathfrak{F}$, again because ε and $b_k(\varepsilon)$ are orthogonal for each k, we can decompose

$$\sigma_f \phi_f(\varepsilon) = \sigma_f^{-1}\varepsilon + \sum_{k=1}^{\infty} J_{f,k} b_k(\varepsilon) \tag{3.4}$$

with coefficients $J_{f,k} := \sigma_f \mathrm{E}_f[b_k(\varepsilon)\phi_f(\varepsilon)]$. This establishes

$$W_{\phi_f} = W_\varepsilon + \sum_{k=1}^{\infty} J_{f,k} W_{b_k}. \tag{3.5}$$

Moreover, we have, for $k \in \mathbb{N}$,

$$\mathrm{Cov}_{0,0,0}\big[W_{\phi_f}(1), W_\varepsilon(1)\big] = 1, \qquad \mathrm{Cov}_{0,0,0}\big[W_{\phi_f}(1), W_{b_k}(1)\big] = J_{f,k} \tag{3.6}$$

and

$$\mathrm{Var}_{0,0,0}\big[W_{\phi_f}(1)\big] = J_f = 1 + \sum_{k=1}^{\infty} J_{f,k}^2. \tag{3.7}$$

As for W_Γ: since, under the null, $(\Delta Y_{t-1}, \dots, \Delta Y_{t-p})'$ is independent of the innovation ε_t and $\mathrm{E}_{0,0,0}[\Delta Y_{t-i}] = 0$, $i = 1, \dots, p$, it follows

$$\mathrm{Cov}_{0,0,0}\big[W_\Gamma(1), W_\varepsilon(1)\big] = 0, \qquad \mathrm{Cov}_{0,0,0}\big[W_\Gamma(1), W_{\phi_f}(1)\big] = 0, \tag{3.8}$$

and

$$\mathrm{Cov}_{0,0,0}\big[W_\Gamma(1), W_{b_k}(1)\big] = 0, \qquad k \in \mathbb{N}. \tag{3.9}$$

We define the covariance matrix of $W_\Gamma(1)$ as

$$\mathrm{Var}_{0,0,0}\big[W_\Gamma(1)\big] = \Sigma_\Gamma \in \mathbb{R}^{p\times p} \tag{3.10}$$

with

$$\Sigma_{\Gamma i,j} := J_f \mathrm{E}_f\big[\Gamma(L)^{-1}\varepsilon_{t-i}\Gamma(L)^{-1}\varepsilon_{t-j}\big], \qquad i, j = 1, \dots, p.$$

Behavior of W_ε, W_{ϕ_f}, W_Γ and W_{b_k} under $\mathbb{P}_{0,0,0}$ when $f \in \mathfrak{F}_{\mathbb{S}}$. In this case, the density function $f(\varepsilon)$ is an even function and so are the perturbation functions $b_k(\varepsilon)$, $k \in \mathbb{N}$. The score $\phi_f(\varepsilon)$ is now an odd function. Therefore, $\phi_f(\varepsilon)$ cannot be decomposed by ε and $b_k(\varepsilon)$ anymore as in (3.4) and (3.5). Instead of (3.7) we now have

$$J_{f,k} = \sigma_f \mathrm{E}_f\big[b_k(\varepsilon)\phi_f(\varepsilon)\big] = 0$$

for all $f \in \mathfrak{F}_{\mathbb{S}}$ and $k \in \mathbb{N}$. All the other results mentioned above still hold.

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. Thus, fix $\mu \in \mathbb{R}$, $\Gamma \in \mathfrak{G}$ and $f \in \mathfrak{F}$. Let, for $h \in \mathbb{R}$, $\gamma \in \mathbb{R}^p$ and $\eta \in c_{00}$, $\mathrm{P}^{(T)}_{h,\gamma,\eta;\mu,\Gamma,f}$ denote the law of $Y_1, \dots, Y_T$ under (2.1)–(2.3) with parameter ρ given by (3.1), $\Gamma^{(T)}_\gamma(L)$ given by (3.2), and innovation density (3.3). 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 $\mu \in \mathbb{R}$, $\Gamma \in \mathfrak{G}$, $f \in \mathfrak{F}$, $h \in \mathbb{R}$, $\gamma \in \mathbb{R}^p$ and $\eta \in c_{00}$. Let Δ denote differencing, that is, $\Delta Y_t = Y_t - Y_{t-1}$.*

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

$$\begin{aligned}
\log \frac{\mathrm{dP}^{(T)}_{h,\gamma,\eta;\mu,\Gamma,f}}{\mathrm{dP}^{(T)}_{0,0,0;\mu,\Gamma,f}}
&= \sum_{t=1}^{T} \log \frac{f^{(T)}_\eta\left(\Gamma^{(T)}_\gamma(L)\left(\Delta Y_t - \frac{h}{T}(Y_{t-1}-\mu)\right)\right)}{f(\Gamma(L)\Delta Y_t)} \\
&= h\Delta^{(T)}_f + \gamma'\Delta^{(T)}_\Gamma + \sum_{k=1}^{\infty} \eta_k \Delta^{(T)}_{b_k} - \frac{1}{2}\mathcal{I}^{(T)}(h,\gamma,\eta) + o_P(1),
\end{aligned}$$

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

$$\Delta^{(T)}_f = \frac{1}{T}\sum_{t=p+2}^{T} \Gamma(L)(Y_{t-1} - Y_1)\phi_f\big(\Gamma(L)\Delta Y_t\big),$$

$$\Delta^{(T)}_\Gamma = \frac{1}{\sqrt{T}}\sum_{t=p+2}^{T} (\Delta Y_{t-1}, \dots, \Delta Y_{t-p})'\phi_f\big(\Gamma(L)\Delta Y_t\big),$$

$$\Delta^{(T)}_{b_k} = \frac{1}{\sqrt{T}}\sum_{t=p+2}^{T} b_k\big(\Gamma(L)\Delta Y_t\big), \qquad k \in \mathbb{N}$$

and

$$\begin{aligned}
\mathcal{I}^{(T)}(h,\gamma,\eta) &= h^2 J_f \frac{1}{T^2}\sum_{t=p+2}^{T} \frac{(\Gamma(L)(Y_{t-1}-Y_1))^2}{\sigma_f^2} + \gamma'\Sigma_\Gamma\gamma + \|\eta\|_2^2 \\
&\quad + 2h\frac{1}{T^{3/2}}\sum_{t=p+2}^{T} \frac{\Gamma(L)(Y_{t-1}-Y_1)}{\sigma_f}\sum_{k=1}^{\infty}\eta_k J_{f,k}.
\end{aligned}$$

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

$$\frac{\mathrm{dP}^{(T)}_{h,\gamma,\eta;\mu,\Gamma,f}}{\mathrm{dP}^{(T)}_{0,0,0;\mu,\Gamma,f}} \Rightarrow \exp\left(h\Delta_f + \gamma'\Delta_\Gamma + \sum_{k=1}^{\infty}\eta_k\Delta_{b_k} - \frac{1}{2}\mathcal{I}(h,\gamma,\eta)\right), \tag{3.11}$$

where

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

(iii) *For all* $h \in \mathbb{R}$, $\gamma \in \mathbb{R}^p$ *and* $\eta \in c_{00}$ *the right-hand side of* (3.11) *has unit expectation under* $\mathbb{P}_{0,0,0}$.

Of course, Proposition 3.2 still holds for $f \in \mathfrak{F}_{\mathbb{S}}$; in that case we have $J_{f,k} = 0$. 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. For the sake of completeness, detailed proofs are organized in the Supplementary Material (Zhou, van den Akker and Werker (2019)).

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

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

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

$$\mathcal{E}^{(T)}(\mu,\Gamma,f) = \big(\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T), (\mathrm{P}^{(T)}_{h,\gamma,\eta;\mu,\Gamma,f} | h \in \mathbb{R}, \gamma \in \mathbb{R}^p, \eta \in c_{00})\big)$$

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

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

Note that the latter experiment indeed depends on f as the measure $\mathbb{P}_{h,\gamma,\eta}$ depends on f.

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

The Asymptotic Representation Theorem (see, e.g., Chapter 9 in van der Vaart (1998)) implies that for any statistic A_T which converges in distribution to the law $L_{h,\gamma,\eta}$, under $\mathrm{P}^{(T)}_{h,\gamma,\eta;\mu,\Gamma,f}$, there exists a (possibly randomized) statistic A, defined on $\mathcal{E}(f)$, such that the law of A under $\mathbb{P}_{h,\gamma,\eta}$ is given by $L_{h,\gamma,\eta}$. This allows us to study (asymptotically) optimal inference: the "best" procedure in the limit experiment also 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 statistics attaining it.

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

PROPOSITION 3.3. *Let $h \in \mathbb{R}$, $\gamma \in \mathbb{R}^p$, $\eta \in c_{00}$, and $f \in \mathfrak{F}$. The processes Z_ε, Z_{ϕ_f}, Z_Γ and Z_{b_k}, $k \in \mathbb{N}$, defined by the starting values $Z_\varepsilon(0) = Z_{\phi_f}(0) = Z_{b_k}(0) = 0$, $Z_\Gamma(0) = \mathbf{0}_p$ and the stochastic differential equations, for $s \in [0, 1]$,*

$$\begin{aligned}
\mathrm{d}Z_\varepsilon(s) &= \mathrm{d}W_\varepsilon(s) - hW_\varepsilon(s)\,\mathrm{d}s,\\
\mathrm{d}Z_{\phi_f}(s) &= \mathrm{d}W_{\phi_f}(s) - hJ_f W_\varepsilon(s)\,\mathrm{d}s - \sum_k \eta_k J_{f,k}\,\mathrm{d}s,\\
\mathrm{d}Z_\Gamma(s) &= \mathrm{d}W_\Gamma(s) - \gamma\,\mathrm{d}s,\\
\mathrm{d}Z_{b_k}(s) &= \mathrm{d}W_{b_k}(s) - hJ_{f,k}W_\varepsilon(s)\,\mathrm{d}s - \eta_k\,\mathrm{d}s, \qquad k \in \mathbb{N},
\end{aligned}$$

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

For the case $f \in \mathfrak{F}_{\mathbb{S}}$, Proposition 3.3 still applies with $J_{f,k} = 0$. Moreover, for this case, we denote by $\mathcal{E}^{(T)}_{\mathbb{S}}(f)$ the associated sequence of experiments and by $\mathcal{E}_{\mathbb{S}}(f)$ the associated limit experiment.

REMARK 3.2. Part (i) and (ii) Proposition 3.2 show that the parameter μ vanishes in the limit. More explicitly, in the proof of this proposition, we replace μ in the likelihood ratio term by Y_1 and then show that the difference term is $o_P(1)$. On the other hand, one could also "localize" the parameter μ as $\mu = \mu_d^{(T)} = \mu_0 + d$ (with rate $T^0 = 1$) as in Jansson (2008). As shown in that paper, the term associated to the parameter d does not change with T and is independent of the other terms of the likelihood ratio. By the additively separable structure, we can treat the parameter μ "as if" it is known. In either way, inference for ρ would be invariant with

respect to μ in the limit. Analogously, in the finite-sample experiment $\mathcal{E}^{(T)}(f)$, μ is eliminated (automatically) by using the increments ΔY_t, $t = p+2, \dots, T$, which are (maximally) invariant with respect to μ (see Section 4).

3.3. *The limit experiment*: *Invariance and power envelope*. In this section, we consider the limit experiments $\mathcal{E}(f)$ and $\mathcal{E}_{\mathbb{S}}(f)$. In these experiments, we observe the processes W_ε, W_{ϕ_f}, W_Γ and W_{b_k}, $k \in \mathbb{N}$, continuously on the time interval $[0,1]$ from the model $(\mathbb{P}_{h,\gamma,\eta} | h \in \mathbb{R}, \gamma \in \mathbb{R}^p, \eta \in c_{00})$, and we are interested in the power envelopes for testing the hypothesis

$$\begin{aligned} &\mathrm{H}_0 : h = 0 \qquad \big(\gamma \in \mathbb{R}^p, \eta \in c_{00}\big) \quad \text{versus} \\ &\mathrm{H}_a : h < 0 \qquad \big(\gamma \in \mathbb{R}^p, \eta \in c_{00}\big). \end{aligned} \tag{3.12}$$

To eliminate the nuisance parameters γ and η, we first propose a statistic that is sufficient for the parameter of interest h and does not depend on γ. Afterwards, using Proposition 3.3, we discuss a natural invariance structure with respect to the infinite-dimensional nuisance parameter η. We derive the maximal invariant and apply the Neyman–Pearson lemma to obtain the power envelopes of invariant tests within the experiments $\mathcal{E}(f)$ and $\mathcal{E}_{\mathbb{S}}(f)$, respectively, in Section 3.3.1 and Section 3.3.2.

We begin with the elimination of γ. The statistic $(W_\varepsilon, W_{\phi_f}, (W_{b_k})_{k\in\mathbb{N}})$ serves as a sufficient statistic for the parameter h. This is because, according to the structural version of $\mathcal{E}(f)$ (or $\mathcal{E}_{\mathbb{S}}(f)$) in Proposition 3.3, the distribution of W_Γ is only affected by γ and so is the distribution of the process $(W_\varepsilon, W_{\phi_f}, (W_{b_k})_{k\in\mathbb{N}})$ only affected by h and η. It then follows that the distribution of the statistic $(W_\varepsilon, W_{\phi_f}, W_\Gamma, (W_{b_k})_{k\in\mathbb{N}})$ conditional on the statistic $(W_\varepsilon, W_{\phi_f}, (W_{b_k})_{k\in\mathbb{N}})$ is only a function of W_Γ and γ and, in turn, does not depend on h. Next, observe that the sufficient statistic $(W_\varepsilon, W_{\phi_f}, (W_{b_k})_{k\in\mathbb{N}})$ is independent of the process W_Γ and thus has a distribution that does not depend on γ. This allows us to restrict attention to the sufficient statistic $(W_\varepsilon, W_{\phi_f}, (W_{b_k})_{k\in\mathbb{N}})$ to conduct inference for h, and the nuisance parameter γ disappears from the likelihoods.

3.3.1. *Elimination of η in $\mathcal{E}(f)$ and the associated power envelope*. The elimination of the nuisance parameter η is more involved and different for the limit experiments $\mathcal{E}(f)$ and $\mathcal{E}_{\mathbb{S}}(f)$. We start with $\mathcal{E}(f)$.

In Proposition 3.3, if one applies the decompositions (3.5) and (3.7) to the first equation (of W_ε) and the fourth equation (of W_{b_k}), one retrieves the second equation (of W_{ϕ_f}). This essentially allows us to omit the process W_{ϕ_f} and restrict the observations to the processes W_ε and W_{b_k}, $k \in \mathbb{N}$.

Now we formalize the invariance structure with respect to η. 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 $W \in C[0,1]$,

$$g_{\eta_k} : \big[g_{\eta_k}(W)\big](s) = W(s) - \eta_k s, \qquad s \in [0,1], \tag{3.13}$$

that is, g_{η_k} adds a drift $s \mapsto -\eta_k s$ to its argument process. Proposition 3.3 implies that the law of $(W_\varepsilon, (g_{\eta_k}(W_{b_k}))_{k\in\mathbb{N}})$ under $\mathbb{P}_{h,\gamma,0}$ is the same as the law of $(W_\varepsilon, (W_{b_k})_{k\in\mathbb{N}})$ under $\mathbb{P}_{h,\gamma,\eta}$. Hence our testing problem (3.12) 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, that is, test statistics t that satisfy

$$t\big(W_\varepsilon, \big(g_{\eta_k}(W_{b_k})\big)_{k\in\mathbb{N}}\big) = t\big(W_\varepsilon, (W_{b_k})_{k\in\mathbb{N}}\big) \qquad \text{for all } g_\eta, \eta \in c_{00}. \tag{3.14}$$

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 $k \in \mathbb{N}$,

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

that is, taking the bridge of a process ensures invariance with respect to adding drifts to that process. Define the mapping M by

$$M\big(W_\varepsilon, (W_{b_k})_{k\in\mathbb{N}}\big) := \big(W_\varepsilon, (B_{b_k})_{k\in\mathbb{N}}\big),$$

where $B_{b_k} := B^{W_{b_k}}$. It follows that statistics that are measurable with respect to the σ-field,

$$\mathcal{M} = \sigma\big(M\big(W_\varepsilon, (W_{b_k})_{k\in\mathbb{N}}\big)\big) = \sigma\big(W_\varepsilon, (B_{b_k})_{k\in\mathbb{N}}\big), \tag{3.15}$$

are invariant (with respect to g_η, $\eta \in c_{00}$). It is, however, not (immediately) clear that we did not throw away too much information. 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* (3.15) *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

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

where the conditional expectation indeed does not depend on η precisely because of the invariance. The conditional expectation also does not depend on γ as a result of the arguments stated at the beginning of this section.

To calculate this conditional expectation, we first introduce $B_{\phi_f} = B^{W_{\phi_f}}$, that is, the bridge process associated to W_{ϕ_f}. Follow the decomposition in (3.5), we can decompose $\Delta_f = \int_0^1 W_\varepsilon(s)\,\mathrm{d}W_{\phi_f}(s) = I + II$ with

$$I = \int_0^1 W_\varepsilon(s)\,\mathrm{d}B_{\phi_f}(s) + W_\varepsilon(1)\int_0^1 W_\varepsilon(s)\,\mathrm{d}s,$$

$$II = \left(\sum_{k=1}^{\infty} J_{f,k} W_{b_k}(1)\right)\int_0^1 W_\varepsilon(s)\,\mathrm{d}s.$$

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

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

with

$$\Delta_f^* = \int_0^1 W_\varepsilon(s)\,\mathrm{d}B_{\phi_f}(s) + W_\varepsilon(1)\int_0^1 W_\varepsilon(s)\,\mathrm{d}s, \tag{3.16}$$

$$\begin{aligned}
\mathcal{I}_f^* &= J_f\int_0^1 W_\varepsilon^2(s)\,\mathrm{d}s - \left(\int_0^1 W_\varepsilon(s)\,\mathrm{d}s\right)^2\sum_{k=1}^{\infty} J_{f,k}^2 \\
&= J_f\int_0^1 W_\varepsilon^2(s)\,\mathrm{d}s - \left(\int_0^1 W_\varepsilon(s)\,\mathrm{d}s\right)^2(J_f - 1),
\end{aligned} \tag{3.17}$$

where the last equality follows from (3.7). Note that the distribution of this likelihood ratio indeed does not depend on the nuisance parameters γ or η.

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 $\mathbb{P}_{0,\gamma,\eta}$, which does not depend on either γ or η, by $c(h, J_f; \alpha)$. Define the size-α test $\phi_{f,\alpha}^{\mathfrak{F}*}(\bar{h}) = \mathbb{1}\{\mathrm{d}\mathbb{P}_{\bar{h}}^{\mathcal{M}}/\mathrm{d}\mathbb{P}_0^{\mathcal{M}} \geq c(\bar{h}, J_f; \alpha)\}$, for a fixed value of $\bar{h} < 0$. Note that this is an oracle test in $\mathcal{E}(f)$ that depends on the true value of f. A feasible test is provided in Section 4. The power function of this oracle test is given by

$$h \mapsto \pi_{f,\alpha}^{\mathfrak{F}*}(h; \bar{h}) = \mathbb{E}_0\left[\phi_{f,\alpha}^{\mathfrak{F}*}(\bar{h})\frac{\mathrm{d}\mathbb{P}_h^{\mathcal{M}}}{\mathrm{d}\mathbb{P}_0^{\mathcal{M}}}\right] = \mathbb{E}_0\left[\phi_{f,\alpha}^{\mathfrak{F}*}(\bar{h})\frac{\mathrm{d}\mathbb{P}_{h,\gamma,\eta}}{\mathrm{d}\mathbb{P}_{0,0,0}}\right].$$

An application of the Neyman–Pearson lemma yields the following.

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

$$\pi(\bar{h}) \leq \pi_{f,\alpha}^{\mathfrak{F}*}(\bar{h}; \bar{h}), \qquad \bar{h} < 0.$$

The (oracle) test $\phi_{f,\alpha}^{\mathfrak{F}*}(\bar{h})$ thus is point optimal, that is, its power function is tangent to the (semiparametric) power envelope $h \mapsto \pi_{f,\alpha}^{\mathfrak{F}*}(h; h)$ at $h = \bar{h}$.[3]

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

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

[3]Here and later in this section, the early usage of the concept "power envelope" is due to the fact that it is shown to be the upper bound later in this section and point-wisely attainable by tests in sequence in Section 4.

$$= \exp\left(h \int_0^1 W_\varepsilon(s)\,\mathrm{d}W_\varepsilon(s) - \frac{1}{2}h^2 \int_0^1 W_\varepsilon^2(s)\,\mathrm{d}s\right),$$

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 ERS test statistic equals the likelihood ratio statistic using the (non-maximal) invariant $\mathcal{M}_\varepsilon$. This explains the improved power of our tests within the model we consider. Moreover, for Gaussian f, we have $\mathcal{M}_\varepsilon = \mathcal{M}$ and obtain point-optimality of the ERS test.[4]

REMARK 3.4. The semiparametric power envelope derived above for this case of $f \in \mathfrak{F}$, of course, coincides with the one in Jansson (2008) based on the invariance constraint. This can be seen by rewriting $\Delta_f^* = \int_0^1 W_\varepsilon(s)\,\mathrm{d}W_{\phi_f}(s) - (W_{\phi_f}(1) - W_\varepsilon(1)) \int_0^1 W_\varepsilon(s)\,\mathrm{d}s$. We feel our approach is attractive since, by describing the perturbations on f with an orthonormal basis and a infinite-dimensional parameter instead of one single parameter, there is no need to find the least favorable direction.[5] Moreover, we feel that the use of the invariance principle, rather than the similarity constraint, more naturally suggests (partly) rank-based tests.

3.3.2. *Elimination of η in $\mathcal{E}_{\mathbb{S}}(f)$ and the associated power envelope.* Since in this case we have $J_{f,k} = 0$, the structural representation of experiment $\mathcal{E}_{\mathbb{S}}(f)$ in Proposition 3.3 becomes

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

Note that here the process W_Γ is again removed by sufficiency in order to eliminate the nuisance parameter γ (see the discussion at the beginning of Section 3.3). Following the same argument for the case of $f \in \mathfrak{F}$, statistics that are measurable with respect to the σ-field

$$\mathcal{M}^{\mathbb{S}} = \sigma\big(M\big(W_\varepsilon, W_{\phi_f}, (W_{b_k})_{k\in\mathbb{N}}\big)\big) = \sigma\big(W_\varepsilon, W_{\phi_f}, (B_{b_k})_{k\in\mathbb{N}}\big) \tag{3.18}$$

are invariant with respect to the transformations g_η, $\eta \in c_{00}$. Moreover, in the following theorem, we show that this σ-field is maximally invariant.

THEOREM 3.2. *The σ-field $\mathcal{M}^{\mathbb{S}}$ in (3.15) is maximally invariant for the group of transformations g_η, $\eta \in c_{00}$, in the experiment $\mathcal{E}_{\mathbb{S}}(f)$.*

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

[5]The traditional semiparametric approach, developed for LAN-type experiments, accomplishes this by projecting the score function of the parameter of interest onto the tangent space of nuisance score functions. However, this approach seems not easily generalizable to LABF-type experiments.

The likelihood ratio restricted to $\mathcal{M}^{\mathbb{S}}$ is given by

$$\begin{aligned}\frac{\mathrm{d}\mathbb{P}_h^{\mathcal{M}^{\mathbb{S}}}}{\mathrm{d}\mathbb{P}_0^{\mathcal{M}^{\mathbb{S}}}} &= \mathbb{E}_0\left[\frac{\mathrm{d}\mathbb{P}_{h,\gamma,\eta}}{\mathrm{d}\mathbb{P}_{0,\gamma,\eta}}\Big|\mathcal{M}^{\mathbb{S}}\right] \\ &= \exp\left(h\Delta_f - \frac{1}{2}\mathcal{I}(h,\gamma,\eta)\right) \\ &\quad \times \mathbb{E}_{0,0,0}\left[\exp\left(\sum_{k=1}^{\infty}\eta_k W_{b_k}(1) + \gamma' W_\Gamma(1)\right)\Big|\mathcal{M}^{\mathbb{S}}\right] \\ &= \exp\left(h\Delta_f - \frac{1}{2}\mathcal{I}(h,\gamma,\eta) + \frac{1}{2}\|\eta\|_2^2 + \frac{1}{2}\gamma'\Sigma_\Gamma\gamma\right) \\ &= \exp\left(h\Delta_f - \frac{1}{2}h^2\mathcal{I}_f\right),\end{aligned}$$

where

$$\mathcal{I}_f := J_f \int_0^1 W_\varepsilon^2(s)\,\mathrm{d}s. \tag{3.19}$$

As before, we then formalize the point-optimal invariant tests in $\mathcal{E}_{\mathbb{S}}(f)$. Denote the $(1-\alpha)$-quantile of $\mathrm{d}\mathbb{P}_h^{\mathcal{M}^{\mathbb{S}}}/\mathrm{d}\mathbb{P}_0^{\mathcal{M}^{\mathbb{S}}}$ under $\mathbb{P}_{0,\gamma,\eta}$ by $c_{\mathbb{S}}(h, J_f; \alpha)$. Define the size-$\alpha$ test $\phi_{f,\alpha}^{\mathfrak{F}_{\mathbb{S}}*}(\bar{h}) = \mathbb{1}\{\mathrm{d}\mathbb{P}_{\bar{h}}^{\mathcal{M}^{\mathbb{S}}}/\mathrm{d}\mathbb{P}_0^{\mathcal{M}^{\mathbb{S}}} \geq c_{\mathbb{S}}(\bar{h}, J_f; \alpha)\}$, for a fixed value of $\bar{h} < 0$. The associated power function is

$$h \mapsto \pi_{f,\alpha}^{\mathfrak{F}_{\mathbb{S}}*}(h;\bar{h}) = \mathbb{E}_0\left[\phi_{f,\alpha}^{\mathfrak{F}_{\mathbb{S}}*}(\bar{h})\frac{\mathrm{d}\mathbb{P}_{h,\gamma,\eta}}{\mathrm{d}\mathbb{P}_{0,0,0}}\right].$$

Again, by the Neyman–Pearson lemma, the following corollary holds.

COROLLARY 3.3. *Let $f \in \mathfrak{F}_{\mathbb{S}}$ and $\alpha \in (0,1)$. Let ϕ be a (possibly randomized) test that is $\mathcal{M}^{\mathbb{S}}$-measurable and is of size α, that is, $\mathbb{E}_0\phi \leq \alpha$. Let π denote the power function of this test, that is, $\pi(h) = \mathbb{E}_h\phi$. Then we have*

$$\pi(\bar{h}) \leq \pi_{f,\alpha}^{\mathfrak{F}_{\mathbb{S}}*}(\bar{h};\bar{h}).$$

The likelihood ratio of the maximal invariant $\mathcal{M}^{\mathbb{S}}$, $\mathrm{d}\mathbb{P}_h^{\mathcal{M}^{\mathbb{S}}}/\mathrm{d}\mathbb{P}_0^{\mathcal{M}^{\mathbb{S}}}$, equals the likelihood ratio of the full observation $(W_\varepsilon, W_{\phi_f}, W_\Gamma, (W_{b_k})_{k\in\mathbb{N}})'$ with known $\eta = 0$ (i.e., known f) and $\gamma = 0$, $\mathrm{d}\mathbb{P}_{h,0,0}/\mathrm{d}\mathbb{P}_{0,0,0}$. This shows that the "semi-parametric" power envelope $\pi_{f,\alpha}^{\mathfrak{F}_{\mathbb{S}}*}$ actually coincides with the parametric power envelope, which is defined based on the likelihood ratio $\mathrm{d}\mathbb{P}_{h,0,0}/\mathrm{d}\mathbb{P}_{0,0,0}$. This verifies again the adaptivity result in Jansson (2008) under the same conditions for this

unit root testing problem.[6] We provide, in Sections 4.1 and 4.2, a class of adaptive unit root tests based on signed-rank statistics for this setting.

REMARK 3.5. The semiparametric power envelopes $\pi_{f,\alpha}^{\mathfrak{F}*}$ and $\pi_{f,\alpha}^{\mathfrak{F}_{\mathbb{S}}*}$ are scale invariant, that is, 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 3.6. The problem of eliminating the nuisance parameter η in the symmetric density case ($f \in \mathfrak{F}_{\mathbb{S}}$) is actually the same as that of eliminating the autocorrelation parameter γ in the beginning of this section: the nuisance parameter appears only in a process which is independent of all the other processes. To be specific, η only affects the distribution of W_{b_k}, which is independent of W_{ϕ_f} as well as W_ε. Therefore, the distribution of $(W_\varepsilon, W_{\phi_f}, (W_{b_k})_{k\in\mathbb{N}})$ conditional on the statistic $(W_\varepsilon, W_{\phi_f})$ does not depend on the parameter of interest h. Hence, the statistic $(W_\varepsilon, W_{\phi_f})$ serves as a sufficient statistic for h which is also invariant with respect to η.

3.4. *The asymptotic power envelope for asymptotically invariant tests.* Now we translate the results for the limit LABF experiment to the unit root model of interest. To mimic the invariance in the limit experiment we introduce the following definition.

DEFINITION 1. Let $\mu \in \mathbb{R}$, $\Gamma \in \mathfrak{G}$, $f \in \mathfrak{F}$. A sequence of test statistics $\psi^{(T)}$ is said to be asymptotically invariant if the distribution of $\psi^{(T)}$ weakly converges, under $\mathrm{P}_{h,\gamma,\eta;\mu,\Gamma,f}^{(T)}$ for all $h \le 0$, $\gamma \in \mathbb{R}^p$, and $\eta \in c_{00}$, to the distribution of an invariant test in the limit experiment $\mathcal{E}(f)/\mathcal{E}_{\mathbb{S}}(f)$, under $\mathbb{P}_{h,\gamma,\eta}$.

The Asymptotic Representation Theorem (see, e.g., van der Vaart (1998) Chapter 9) now yields the following main result on the asymptotic power envelope.

THEOREM 3.3. *Let $\mu \in \mathbb{R}$, $\Gamma \in \mathfrak{G}$, $f \in \mathfrak{F}$, and $\alpha \in (0,1)$. Let $\phi_T(Y_1, \dots, Y_T)$, $T \in \mathbb{N}$, be an asymptotically invariant test of size α, that is, $\limsup_{T\to\infty} \mathrm{E}_{0,\gamma,\eta} \times \phi_T \le \alpha$ for all $\gamma \in \mathbb{R}^p$ and $\eta \in c_{00}$. Let π_T denote the power function of ϕ_T, that is, $\pi_T(h,\gamma,\eta) = \mathrm{E}_{h,\gamma,\eta;\mu,\Gamma,f}\phi_T$. Then we have*

$$\limsup_{T\to\infty} \pi_T(h,\gamma,\eta) \le \pi_{f,\alpha}^{\mathfrak{F}*}(h;h), \qquad h < 0, \gamma \in \mathbb{R}^p, \text{ and } \eta \in c_{00}.$$

For the case of $f \in \mathfrak{F}_{\mathbb{S}}$, we have

$$\limsup_{T\to\infty} \pi_T(h,\gamma,\eta) \le \pi_{f,\alpha}^{\mathfrak{F}_{\mathbb{S}}*}(h;h), \qquad h < 0, \gamma \in \mathbb{R}^p, \text{ and } \eta \in c_{00}.$$

[6] A discussion about the notion of "adaptiveness" in this nonstandard testing problem can be found in Section 5 of Jansson (2008).

These power envelopes for invariant tests in the limit experiments $\mathcal{E}(f)$ and $\mathcal{E}_{\mathbb{S}}(f)$ thus provide upper bounds to the asymptotic powers of invariant tests for the unit root hypotheses in $\mathcal{E}^{(T)}(f)$ and $\mathcal{E}_{\mathbb{S}}^{(T)}(f)$, respectively. The next section introduces two classes of tests (based on rank and signed-rank statistics) that (in a point-wise sense) attain these power envelopes and, thereby, demonstrates that these bounds are sharp. We additionally provide a Chernoff–Savage-type result for these classes 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 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 $\Gamma(L)\Delta Y_t$ among $\Gamma(L)\Delta Y_{p+2}, \dots, \Gamma(L)\Delta Y_T$. 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 remains valid, that is, 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 3. The density $g \in \mathfrak{F}$, with finite variance σ_g^2, satisfies

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

with location score function $\phi_g(\varepsilon) := -(g'/g)(\varepsilon)$, where J_g is the standardized Fisher information for location of g.[7]

Now we can formulate the following direct extension of Lemma A.1 in Hallin, van den Akker and Werker (2011) and its signed-rank counterpart. The proof for the weak convergence of the stochastic integrals in (4.3) and (4.7) is provided in the Supplementary Material (Zhou, van den Akker and Werker (2019)).

LEMMA 4.1. *Let $\mu \in \mathbb{R}$, $\Gamma \in \mathfrak{G}$ and g satisfy Assumption 3.*

[7]Similar to the standardized Fisher information J_f of f, the Fisher information J_g of g is standardized by the variance σ_g^2. As a result, it is scale invariant.

(i) *For the case* $f \in \mathfrak{F}$, *consider the partial sum process*

$$B_{\phi_g}^{(T)}(s) = \frac{1}{\sqrt{T}} \sum_{t=p+2}^{\lfloor sT \rfloor} \sigma_g \left(\phi_g \left(G^{-1} \left(\frac{R_t}{T-p} \right) \right) - \bar{\phi}_g^{(T)} \right) \tag{4.1}$$

for $s \in [0,1]$, *where* $\bar{\phi}_g^{(T)} := T^{-1} \sum_{i=1}^{T-p-1} \phi_g(G^{-1}(i/(T-p)))$, *and* R_t *denotes the rank of* $\Gamma(L)\Delta Y_t$, $t = p+2, \dots, T$. *Then, under* $P_{0,0,0;\mu,\Gamma,f}^{(T)}$ *and as* $T \to \infty$, *we have*

$$[W_\varepsilon^{(T)}, W_{\phi_f}^{(T)}, B_{\phi_g}^{(T)}]' \quad \Rightarrow \quad [W_\varepsilon, W_{\phi_f}, B_{\phi_g}]', \tag{4.2}$$

and

$$\int_0^1 W_\varepsilon^{(T)}(s-)\,\mathrm{d}B_{\phi_g}^{(T)}(s) \quad \Rightarrow \quad \int_0^1 W_\varepsilon(s)\,\mathrm{d}B_{\phi_g}(s). \tag{4.3}$$

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* $(\Omega, \mathcal{F}, \mathbb{P}_{0,0,0})$ *as* W_ε *and* W_{ϕ_f}, *with covariance matrix*

$$\mathrm{Cov}_{0,0,0} \begin{bmatrix} W_\varepsilon(1) \\ W_{\phi_f}(1) \\ W_{\phi_g}(1) \end{bmatrix} = \begin{pmatrix} 1 & 1 & \sigma_{\varepsilon\phi_g} \\ & J_f & J_{fg} \\ & & J_g \end{pmatrix}, \tag{4.4}$$

where

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

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

(ii) *For the case* $f \in \mathfrak{F}_{\mathbb{S}}$, *consider the partial sum process*

$$W_{\phi_g}^{(T)}(s) = \frac{1}{\sqrt{T}} \sum_{t=p+2}^{\lfloor sT \rfloor} s_t \sigma_g \left(\phi_g \left(G^{-1} \left(\frac{1}{2} + \frac{R_t^+}{2(T-p)} \right) \right) \right) \tag{4.5}$$

for $s \in [0,1]$, *where* s_t *and* R_t^+ *denote the sign of* $\Gamma(L)\Delta Y_t$ *and the rank of its absolute value, respectively, for* $t = p+2, \dots, T$. *Then, under* $P_{0,0,0;\mu,\Gamma,f}^{(T)}$ *and as* $T \to \infty$, *we have*

$$[W_\varepsilon^{(T)}, W_{\phi_f}^{(T)}, W_{\phi_g}^{(T)}]' \quad \Rightarrow \quad [W_\varepsilon, W_{\phi_f}, W_{\phi_g}]', \tag{4.6}$$

and

$$\int_0^1 W_\varepsilon^{(T)}(s-)\,\mathrm{d}B_{\phi_g}^{(T)}(s) \quad \Rightarrow \quad \int_0^1 W_\varepsilon(s)\,\mathrm{d}B_{\phi_g}(s), \tag{4.7}$$

where the law of W_{ϕ_g} *is given in* (4.4).

In practice, the autoregressive structure $\Gamma(L)$ will not be known. In that case, one must rely on ranks (and signs) based on innovations calculated using an *estimated* autoregressive structure, that is, using residuals. This is usually referred to as inference based on *aligned ranks*. For this purpose, we first introduce the following assumption on the estimation of $\Gamma(L)$ (see Hallin and Puri (1994)).

ASSUMPTION 4. (i) There exists, under the null hypothesis, a $\sqrt{T}$-consistent estimator $(\widehat{\Gamma}_1, \dots, \widehat{\Gamma}_p)'$ of $(\Gamma_1, \dots, \Gamma_p)'$. That is, for all $\mu \in \mathbb{R}$, $\Gamma \in \mathfrak{G}$, $f \in \mathfrak{F}$ and all $\epsilon > 0$, there exists b and T_b such that

$$P^{(T)}_{0,0,0;\mu,\Gamma,f}\{\|\sqrt{T}(\widehat{\Gamma} - \Gamma)\| > b\} < \epsilon \qquad \forall t \geq T_b. \tag{4.8}$$

(ii) The estimator $(\widehat{\Gamma}_1, \dots, \widehat{\Gamma}_p)'$ is discretized on grids of mesh width $T^{-1/2}$. That is, the number of possible values of $(\widehat{\Gamma}_1, \dots, \widehat{\Gamma}_p)'$ in balls of the form $\{\Upsilon \in \mathbb{R}^p \| T^{-1/2}(\Upsilon - \Upsilon_0) \leq c\}$ remains bounded, as $T \to \infty$, for all $\Upsilon_0 \in \mathbb{R}^p$ and all $c > 0$.

Part (i) of Assumption 4 is mild as many known estimators of $(\Gamma_1, \dots, \Gamma_p)'$ exist for the AR(p) model that can be applied to the increments ΔY_t; see, for example, Brockwell and Davis (2016). Part (ii) is standard in the semiparametric literature and can easily be met by transforming the estimator in Part (i); see, for example, Bickel (1982) and Kreiss (1987). Such an estimator is often called *locally asymptotically discrete*. The use of aligned ranks does not invalidate the conclusion of Lemma 4.1. This is the content of the following result, of which the proof is provided in the Supplementary Material (Zhou, van den Akker and Werker (2019)).

LEMMA 4.2. *Under Assumption 4, Lemma 4.1 remains valid in case aligned signs and ranks are used, that is, signs and ranks of innovations calculated using the estimated autoregressive coefficients* $(\widehat{\Gamma}_1, \dots, \widehat{\Gamma}_p)$.

We denote by $\widehat{B}^{(T)}_{\phi_g}$ and $\widehat{W}^{(T)}_{\phi_g}$ the aligned-rank-based counterparts of the rank-based processes $B^{(T)}_{\phi_g}$ and $W^{(T)}_{\phi_g}$, respectively.

4.1. *Hybrid rank tests based on a reference density.* In this section, we propose our unit root tests, for both the nonsymmetric case ($f \in \mathfrak{F}$) and the symmetric case ($f \in \mathfrak{F}_{\mathbb{S}}$). Section 3.3 provides the basis for optimal invariant tests in the limit experiment $\mathcal{E}(f)$ and $\mathcal{E}_{\mathbb{S}}(f)$. Lemmas 4.1 and 4.2 can subsequently be used to approximate, in the sequences of unit root experiments $\mathcal{E}^{(T)}(f)$ and $\mathcal{E}^{(T)}_{\mathbb{S}}(f)$, the observable processes in these limit experiments. More specifically, these two lemmas indicate that constructing the rank-based score partial sum processes $B^{(T)}_{\phi_g}$ and $W^{(T)}_{\phi_g}$ as in (4.1) and (4.5), together with $W^{(T)}_{\varepsilon}$, intuitively corresponds to observing the σ-fields $\mathcal{M}_g := \sigma(W_\varepsilon, B_{\phi_g})$ and $\mathcal{M}^{\mathbb{S}}_g := \sigma(W_\varepsilon, W_{\phi_g})$ in the limit experiments

$\mathcal{E}(f)$ and $\mathcal{E}_{\mathbb{S}}(f)$, respectively. This then leads to our asymptotically invariant tests based on $\mathcal{M}_g$ and $\mathcal{M}_g^{\mathbb{S}}$.

The following proposition establishes the likelihood ratio restricted to the information $\mathcal{M}_g$ and $\mathcal{M}_g^{\mathbb{S}}$. The Neyman–Pearson lemma implies that tests based on these likelihood ratios are point optimal amongst the class of invariant tests in the limit experiments.

PROPOSITION 4.1. *Define $W_\perp$ implicitly via the decomposition*

$$W_{\phi_g} = \sigma_{\varepsilon\phi_g} W_\varepsilon + \sqrt{J_g - \sigma_{\varepsilon\phi_g}^2}\, W_\perp, \tag{4.9}$$

which is a standard Brownian motion under $\mathbb{P}_{0,0,0}$ and denote the associated bridge by $B_\perp$.[8]

(i) *The likelihood ratio $\mathrm{d}\mathbb{P}_h^{\mathcal{M}}/\mathrm{d}\mathbb{P}_0^{\mathcal{M}}$ restricted to the outcome space $\mathcal{M}_g$ is given by*

$$\frac{\mathrm{d}\mathbb{P}_h^{\mathcal{M}_g}}{\mathrm{d}\mathbb{P}_0^{\mathcal{M}_g}} = \mathbb{E}_0\left[\frac{\mathrm{d}\mathbb{P}_h^{\mathcal{M}}}{\mathrm{d}\mathbb{P}_0^{\mathcal{M}}}\Big|\mathcal{M}_g\right] = \exp\left(h\Delta_g - \frac{1}{2}h^2\mathcal{I}_g\right), \tag{4.10}$$

where

$$\Delta_g = \Delta_\varepsilon + \lambda\Delta_\perp,$$
$$\mathcal{I}_g = \int_0^1 W_\varepsilon^2(s)\,\mathrm{d}s + \lambda^2\left(\frac{J_g}{\sigma_{\varepsilon\phi_g}^2} - 1\right)\left[\int_0^1 W_\varepsilon(s)^2\,\mathrm{d}s - \left(\int_0^1 W_\varepsilon(s)\,\mathrm{d}s\right)^2\right],$$

with $\Delta_\varepsilon = \int_0^1 W_\varepsilon(s)\,\mathrm{d}W_\varepsilon(s)$, $\Delta_\perp = \sqrt{J_g/\sigma_{\varepsilon\phi_g}^2 - 1}\int_0^1 W_\varepsilon(s)\,\mathrm{d}B_\perp(s)$, and

$$\lambda = (J_{fg}\sigma_{\varepsilon\phi_g} - \sigma_{\varepsilon\phi_g}^2)/(J_g - \sigma_{\varepsilon\phi_g}^2).$$

(ii) *The likelihood ratio $\mathrm{d}\mathbb{P}_h^{\mathcal{M}_g^{\mathbb{S}}}/\mathrm{d}\mathbb{P}_0^{\mathcal{M}_g^{\mathbb{S}}}$ restricted to the outcome space $\mathcal{M}_g^{\mathbb{S}}$ is given by*

$$\frac{\mathrm{d}\mathbb{P}_h^{\mathcal{M}_g^{\mathbb{S}}}}{\mathrm{d}\mathbb{P}_0^{\mathcal{M}_g^{\mathbb{S}}}} = \mathbb{E}_0\left[\frac{\mathrm{d}\mathbb{P}_h^{\mathcal{M}^{\mathbb{S}}}}{\mathrm{d}\mathbb{P}_0^{\mathcal{M}^{\mathbb{S}}}}\Big|\mathcal{M}_g\right] = \exp\left(h\Delta_g^{\mathbb{S}} - \frac{1}{2}h^2\mathcal{I}_g^{\mathbb{S}}\right), \tag{4.11}$$

[8]The implicit requirement $J_g \geq \sigma_{\varepsilon\phi_g}^2$ in the decomposition (4.9) is directly guaranteed the Cauchy–Schwarz inequality, $\sigma_g^2\int_0^1|\phi_g(G^{-1}(u))|^2\,\mathrm{d}u \cdot \sigma_f^{-2}\int_0^1|F^{-1}(u)|^2\,\mathrm{d}u \geq |\sigma_f^{-1}\sigma_g\int_0^1 F^{-1}(u)\phi_g(G^{-1}(u))\,\mathrm{d}u|^2$, and the fact that $\sigma_f^{-2}\int_0^1[F^{-1}(u)]^2\,\mathrm{d}u = 1$.

where

$$\Delta_g^{\mathbb{S}} = \Delta_\varepsilon + \lambda \Delta_\perp^{\mathbb{S}},$$
$$\mathcal{I}_g^{\mathbb{S}} = \left(1 + \lambda^2 \frac{J_g}{\sigma_{\varepsilon\phi_g}^2} - \lambda^2\right)\left[\int_0^1 W_\varepsilon(s)^2 \,\mathrm{d}s\right],$$

with $\Delta_\perp^{\mathbb{S}} = \sqrt{J_g/\sigma_{\varepsilon\phi_g}^2 - 1} \int_0^1 W_\varepsilon(s)\,\mathrm{d}W_\perp(s)$.

REMARK 4.1. To construct the test statistic in the limit experiment $\mathcal{E}(f)$, instead of simply replacing B_{ϕ_f} by B_{ϕ_g}, we rely on the likelihood ratio of $\mathcal{M}_g$. This, by the Neyman–Pearson lemma, is the optimal way regardless of complexity. Clearly, it has $\mathcal{M}_g \subseteq \mathcal{M}$ so that $\mathcal{M}_g$ is invariant for the group of transformations g_η.[9] When $g = f$, $\mathcal{M}_g = \mathcal{M}$ so that it is maximally invariant, which means that we capture all available information about h, and this in turn gives the asymptotic optimality for the finite-sample counterpart below. The same argument also holds for the limit experiment $\mathcal{E}_{\mathbb{S}}(f)$.

REMARK 4.2. The result of Proposition 4.1 can also be achieved by first applying Girsanov's theorem to the experiment associated to observing W_ε and W_{ϕ_g} defined via

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

to get the likelihood ratio of $\mathcal{M}_g^{\mathbb{S}}$. This experiment is obtained by combining the limit experiment in Proposition 3.3 and the covariance matrix in (4.4). In other words, this provides the limit experiment associated to $\mathcal{M}_g^{\mathbb{S}}$. Subsequently, one can take the expectation of the likelihood ratio of $\mathcal{M}_g^{\mathbb{S}}$ obtained above conditional on $\mathcal{M}_g$ to get the likelihood ratio of $\mathcal{M}_g$. The associated limit experiment is given by

$$\mathrm{d}W_\varepsilon(s) = hW_\varepsilon(s)\,\mathrm{d}s + \mathrm{d}Z_\varepsilon(s),$$
$$\mathrm{d}B_{\phi_g}(s) = hJ_{fg}\big[W_\varepsilon(s) - \overline{W}_\varepsilon\big]\,\mathrm{d}s + \mathrm{d}\big[Z_{\phi_g}(s) - sZ_{\phi_g}(1)\big],$$

where $\overline{W}_\varepsilon = \int_0^1 W_\varepsilon(r)\,\mathrm{d}r$.

Observe that $W_\perp$ is a standard Brownian motion under $\mathbb{P}_{0,0,0}$ and 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, we have $\Delta_g = \Delta_f^*$ and $\mathcal{I}_g = \mathcal{I}_f^*$, and $\Delta_g^{\mathbb{S}} = \Delta_f$ and $\mathcal{I}_g^{\mathbb{S}} = \mathcal{I}_f$.

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

The central idea to construct a hybrid rank test is to use a (quasi)-likelihood ratio test based on $L_{\mathcal{M}_g}(h,\lambda) := h\Delta_g - \frac{1}{2}h^2\mathcal{I}_g$ from (4.10) for the case of $f \in \mathfrak{F}$, and $L_{\mathcal{M}_g^{\mathbb{S}}}(h,\lambda) := h\Delta_g^{\mathbb{S}} - \frac{1}{2}h^2\mathcal{I}_g^{\mathbb{S}}$ from (4.11) for the case of $f \in \mathfrak{F}_{\mathbb{S}}$. In both cases, we then replace W_ε and B_{ϕ_g} by their finite-sample counterparts in Lemma 4.2 (or those in Lemma 4.1 in case $\Gamma(L)$ would be known, for example, when $p=0$). The remaining unknown finite-sample parameters σ_f^2 and λ are replaced by estimates that need to satisfy the following condition.

ASSUMPTION 5. 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 $\mu \in \mathbb{R}$, $\Gamma \in \mathfrak{G}$, $f \in \mathfrak{F}$, 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 $\mathrm{P}^{(T)}_{0,0,0;\mu,\Gamma,f}$ 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, for example, 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 and Paindaveine (2010). It is also worth noting that the consistency automatically also holds under local alternatives due to Le Cam's third lemma.

Based on a chosen reference density g satisfying Assumption 3 and estimators $\hat{\sigma}_f$, $\hat{\sigma}_{\varepsilon\phi_g}$ and $\hat{J}_{fg}$ satisfying Assumption 5, we introduce the following partial sum processes:

$$\widehat{W}_\varepsilon^{(T)}(s) = \frac{1}{\sqrt{T}} \sum_{t=p+2}^{\lfloor sT \rfloor} \frac{\widehat{\Gamma}(L)\Delta Y_t}{\hat{\sigma}_f},$$

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

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

where $\widehat{B}_{\phi_g}^{(T)}(s)$ and $\widehat{W}_{\phi_g}^{(T)}(s)$ are defined by Lemma 4.2. Note that in the case of known $\Gamma(L)$, for example, the i.i.d. case, one can simply use $\widehat{\Gamma}(L) = \Gamma(L)$ (where $\widehat{B}_{\phi_g}^{(T)}(s) = B_{\phi_g}^{(T)}(s)$ and $\widehat{W}_{\phi_g}^{(T)}(s) = W_{\phi_g}^{(T)}(s)$). Now, given a fixed alternative $\bar{h} < 0$, we define

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

with

$$\widehat{\Delta}_g = \widehat{\Delta}_\varepsilon + \hat{\lambda}\widehat{\Delta}_\perp,$$
$$\widehat{\mathcal{I}}_g = \int_0^1 \big(\widehat{W}_\varepsilon^{(T)}(s-)\big)^2\,\mathrm{d}s + \hat{\lambda}^2\Big(\frac{J_g}{\hat{\sigma}^2_{\varepsilon\phi_g}} - 1\Big) \times \Big[\int_0^1 \big(\widehat{W}_\varepsilon^{(T)}(s-)\big)^2\,\mathrm{d}s - \Big(\int_0^1 \widehat{W}_\varepsilon^{(T)}(s-)\,\mathrm{d}s\Big)^2\Big],$$

where $\widehat{\Delta}_\varepsilon = \int_0^1 \widehat{W}_\varepsilon^{(T)}(s-)\,\mathrm{d}\widehat{W}_\varepsilon^{(T)}(s)$, $\widehat{\Delta}_\perp = \sqrt{J_g/\hat{\sigma}^2_{\varepsilon\phi_g} - 1}\int_0^1 \widehat{W}_\varepsilon^{(T)}(s-)\,\mathrm{d}\widehat{B}_\perp^{(T)}(s)$ and $\hat{\lambda} = (\hat{J}_{fg}\hat{\sigma}_{\varepsilon\phi_g} - \hat{\sigma}^2_{\varepsilon\phi_g})/(J_g - \hat{\sigma}^2_{\varepsilon\phi_g})$. Define also for the case of $f \in \mathfrak{F}_{\mathbb{S}}$,

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

with

$$\widehat{\Delta}_g^{\mathbb{S}} = \widehat{\Delta}_\varepsilon + \hat{\lambda}\widehat{\Delta}_\perp^{\mathbb{S}},$$
$$\widehat{\mathcal{I}}_g^{\mathbb{S}} = \Big(1 + \hat{\lambda}^2\frac{J_g}{\hat{\sigma}^2_{\varepsilon\phi_g}} - \hat{\lambda}^2\Big)\Big[\int_0^1 \big(\widehat{W}_\varepsilon^{(T)}(s-)\big)^2\,\mathrm{d}s\Big],$$

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

By Slutsky's theorem, we have the convergences $(\widehat{W}_\varepsilon^{(T)}, \widehat{B}_\perp^{(T)})' \Rightarrow (W_\varepsilon, B_\perp)'$ and $(\widehat{W}_\varepsilon^{(T)}, \widehat{W}_\perp^{(T)})' \Rightarrow (W_\varepsilon, W_\perp)'$, and the convergences of stochastic integrals $\int_0^1 \widehat{W}_\varepsilon^{(T)}(s-)\,\mathrm{d}\widehat{B}_\perp^{(T)}(s) \Rightarrow \int_0^1 W_\varepsilon(s)\,\mathrm{d}B_\perp(s)$ and $\int_0^1 \widehat{W}_\varepsilon^{(T)}(s-)\,\mathrm{d}\widehat{W}_\perp^{(T)}(s) \Rightarrow \int_0^1 W_\varepsilon(s)\,\mathrm{d}W_\perp(s)$ (see the proof of Lemma 4.1). We thus also obtain $\widehat{L}_{\mathcal{M}_g}(\bar{h}, \hat{\lambda}) \Rightarrow L_{\mathcal{M}_g}(\bar{h}, \lambda)$ and $\widehat{L}_{\mathcal{M}_g^{\mathbb{S}}}(\bar{h}, \hat{\lambda}) \Rightarrow L_{\mathcal{M}_g^{\mathbb{S}}}(\bar{h}, \lambda)$ under $P^{(T)}_{0,0,0;\mu,\Gamma,f}$. Define the critical values $c_{\mathcal{M}_g}(\bar{h}, \sigma_{\varepsilon\phi_g}, \lambda, J_g; \alpha)$ and $c_{\mathcal{M}_g^{\mathbb{S}}}(\bar{h}, \sigma_{\varepsilon\phi_g}, \lambda, J_g; \alpha)$ by the $(1-\alpha)$-quantiles of $L_{\mathcal{M}_g}(\bar{h}, \lambda)$ and $L_{\mathcal{M}_g^{\mathbb{S}}}(\bar{h}, \lambda)$, respectively. This leads to the (feasible) tests

$$\phi_{\mathcal{M}_g}(\bar{h}, \alpha) := \mathbb{1}\big\{\widehat{L}_{\mathcal{M}_g}(\bar{h}, \hat{\lambda}) \geq c_{\mathcal{M}_g}(\bar{h}, \hat{\sigma}_{\varepsilon\phi_g}, \hat{\lambda}, J_g; \alpha)\big\}$$

and

$$\phi_{\mathcal{M}_g^{\mathbb{S}}}(\bar{h}, \alpha) := \mathbb{1}\big\{\widehat{L}_{\mathcal{M}_g^{\mathbb{S}}}(\bar{h}, \hat{\lambda}) \geq c_{\mathcal{M}_g^{\mathbb{S}}}(\bar{h}, \hat{\sigma}_{\varepsilon\phi_g}, \hat{\lambda}, J_g; \alpha)\big\}.$$

These tests are not only based on the ranks of ΔY_t but also their average, therefore, we name them Hybrid Rank Tests (HRTs).

We can now state our main theoretical result.

THEOREM 4.1. (i) *Under the Assumptions* 1–5, *for each chosen* $\alpha \in (0, 1)$ *and* $\bar{h} \in (-\infty, 0)$, *we have*:

(1) *The HRT* $\phi_{\mathcal{M}_g}(\bar{h}, \alpha)$ *is asymptotically of size* α.
(2) *The HRT* $\phi_{\mathcal{M}_g}(\bar{h}, \alpha)$ *is asymptotically invariant.*
(3) *The HRT* $\phi_{\mathcal{M}_g}(\bar{h}, \alpha)$ *is point-optimal, at* $h = \bar{h}$, *if* $g = f$.

(ii) *Under the Assumptions 1–5 and* $g \in \mathfrak{F}_{\mathbb{S}}$, *for each chosen* $\alpha \in (0, 1)$ *and* $\bar{h} \in (-\infty, 0)$, *we have*:

(1) *The HRT* $\phi_{\mathcal{M}_g^{\mathbb{S}}}(\bar{h}, \alpha)$ *is asymptotically of size* α.
(2) *The HRT* $\phi_{\mathcal{M}_g^{\mathbb{S}}}(\bar{h}, \alpha)$ *is asymptotically invariant.*
(3) *The HRT* $\phi_{\mathcal{M}_g^{\mathbb{S}}}(\bar{h}, \alpha)$ *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. Moreover, in the corollary below, we state that the HRTs enjoy a Chernoff–Savage-type result.

COROLLARY 4.1. *Fix* $\alpha \in (0, 1)$ *and* $\bar{h} < 0$. *The HRT* $\phi_{\mathcal{M}_g}(\bar{h}, \alpha)$ *is, for any reference density* g *satisfying Assumption 3, more powerful, at* $h = \bar{h}$ *and for* $\mu \in \mathbb{R}$, $\Gamma \in \mathfrak{G}$ *and* $f \in \mathfrak{F}$, *than the ERS test except when* f *is Gaussian where they have equal powers. The same argument holds for the HRT* $\phi_{\mathcal{M}_g^{\mathbb{S}}}(\bar{h}, \alpha)$ *with* $g \in \mathfrak{F}_{\mathbb{S}}$ *for any* $f \in \mathfrak{F}_{\mathbb{S}}$.

Corollary 4.1 is a particularly useful result for applied work. The HRT dominates its classical canonical Gaussian counterpart, that is, the ERS test in the present model, for any reference density g. Traditionally, this claim can only be made for Gaussian reference densities, but the 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 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 4.3. It is worth noting that the invariance constraint is only imposed in the limit and, therefore, the maximal invariant needs only to be derived in the limit experiment. In other words, in the finite-sample unit root testing experiment $\mathcal{E}^{(T)}(f)$ (or $\mathcal{E}_{\mathbb{S}}^{(T)}(f)$ for $f \in \mathfrak{F}_{\mathbb{S}}$), we actually use statistics that are only asymptotically invariant (i.e., their limiting equivalents are measurable with respect to the maximally invariant sigma-field $\mathcal{M}$ (or $\mathcal{M}^{\mathbb{S}}$ for $f \in \mathfrak{F}_{\mathbb{S}}$), while, for finite T, they are not necessarily (maximally) invariant with respect to some transformation

on the density f. In fact, a maximal invariant may very well not even exist in the finite-sample experiments. Specifically, $W_\varepsilon^{(T)}$ approximates W_ε in the limit, whose distribution does not change with the density f, while the distribution of $W_\varepsilon^{(T)}$ does depend on the density f. Instead, the statistic $B_{\phi_g}^{(T)}$ (or $W_{\phi_g}^{(T)}$ for $f \in \mathfrak{F}_{\mathbb{S}}$) is distribution-free, that is, its distribution is not affected by any transformation on the density f. In Section 5 below, we will show that the asymptotic approximations work well even in smaller samples.

REMARK 4.4. The additional power of the HRT compared to the ERS test is not free due to the stronger weak convergence assumption employed. Consequently, the class of models for which the HRTs are valid forms a subclass of the class where the ERS tests are valid. In this subclass, the HRT dominates the ERS 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 ERS tests. Again, within the class of models where the ERS 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. *Approximate hybrid rank tests*. A somewhat inconvenient aspect of the hybrid rank tests is that we need to estimate "the real-valued parameter" J_{fg}. As mentioned before, this is (much) less complicated than estimating the score function ϕ_f (as needed in Jansson (2008)), but might still be considered cumbersome, despite all references mentioned below Assumption 5. Moreover, the critical value $c_{\mathcal{M}_g}(\bar{h}, \hat{\sigma}_{\varepsilon\phi_g}, \hat{\lambda}, J_g; \alpha)$ depends on the estimates $\hat{\sigma}_{\varepsilon\phi_g}$ and $\hat{\lambda}$ (henceforth $\hat{J}_{fg}$). Of course, this introduces no difficulty to implementing the test for a single dataset (though one would need to simulate a critical value), however, when it comes to a Monte Carlo study to access the performances of the HRTs, the computational effort will be significant. Therefore, we introduce a simplified version of the hybrid rank test. This simplified test is obtain by invoking $\lambda = 1$, which holds in case $g = f$.

To be precise, define

$$\widehat{L}_g(\bar{h}) := \widehat{L}_{\mathcal{M}_g}(\bar{h}, 1) = \bar{h}\widehat{\Delta}_g - \frac{1}{2}\bar{h}^2\widehat{\mathcal{I}}_g, \tag{4.12}$$

where

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

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

and

$$\widehat{L}^{\mathbb{S}}_g(\bar{h}) := \widehat{L}_{\mathcal{M}^{\mathbb{S}}_g}(\bar{h}, 1) = \bar{h}\widehat{\Delta}^{\mathbb{S}}_g - \frac{1}{2}\bar{h}^2\widehat{\mathcal{I}}^{\mathbb{S}}_g, \tag{4.13}$$

where

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

$$\widehat{\mathcal{I}}^{\mathbb{S}}_g = \frac{J_g}{\hat{\sigma}^2_{\varepsilon\phi_g}} \int_0^1 \big(\widehat{W}^{(T)}_\varepsilon(s-)\big)^2 \,\mathrm{d}s.$$

Also define $L_g(\bar{h}) := L_{\mathcal{M}_g}(\bar{h}, 1)$ and $L^{\mathbb{S}}_g(\bar{h}) := L_{\mathcal{M}^{\mathbb{S}}_g}(\bar{h}, 1)$, then we have $\widehat{L}_g(\bar{h}) \Rightarrow L_g(\bar{h})$ and $\widehat{L}^{\mathbb{S}}_g(\bar{h}) \Rightarrow L^{\mathbb{S}}_g(\bar{h})$ under $\mathrm{P}^{(T)}_{0,0,0;\mu,\Gamma,f}$. Denoting the $(1-\alpha)$-quantiles of $L_g(\bar{h})$ and $L^{\mathbb{S}}_g(\bar{h})$ by $c_g(\bar{h}, \sigma_{\varepsilon\phi_g}, J_g; \alpha)$ and $c^{\mathbb{S}}_g(\bar{h}, \sigma_{\varepsilon\phi_g}, J_g; \alpha)$, respectively. These lead to the feasible tests

$$\phi_g(\bar{h}, \alpha) := \mathbb{1}\big\{L_g(\bar{h}) \geq c_g(\bar{h}, \hat{\sigma}_{\varepsilon\phi_g}, J_g; \alpha)\big\}$$

and

$$\phi^{\mathbb{S}}_g(\bar{h}, \alpha) := \mathbb{1}\big\{L^{\mathbb{S}}_g(\bar{h}) \geq c^{\mathbb{S}}_g(\bar{h}, \hat{\sigma}_{\varepsilon\phi_g}, J_g; \alpha)\big\}.$$

Since $\phi_g(\bar{h}, \alpha)$ and $\phi^{\mathbb{S}}_g(\bar{h}, \alpha)$ are approximate versions of the Hybrid Rank Tests $\phi_{\mathcal{M}_g}(\bar{h}, \alpha)$ and $\phi_{\mathcal{M}^{\mathbb{S}}_g}(\bar{h}, \alpha)$, we refer to them as Approximate Hybrid Rank Tests (AHRTs).

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 convergences $\widehat{L}_g(\bar{h}) \Rightarrow L_g(\bar{h})$ and $\widehat{L}^{\mathbb{S}}_g(\bar{h}) \Rightarrow L^{\mathbb{S}}_g(\bar{h})$. The simulation results in Section 5 show that these asymptotic properties carry over to finite samples.

REMARK 4.5. 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 AHRTs, 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.

TABLE 1

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 on the interval $[0, \sqrt{J_g}]$ with a grid where adjacent points are 0.01 apart

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

From a computational point of view, the AHRTs have 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 $c_g(\bar{h}, \sigma_{\varepsilon\phi_g}, J_g; \alpha)$ and $c_g^{\mathbb{S}}(\bar{h}, \sigma_{\varepsilon\phi_g}, J_g; \alpha)$ are 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 $\sqrt{J_g}$. For the chosen three reference densities, the critical value functions are listed in Table 1. These are obtained by fitting a fourth-order polynomial to the exact critical values. In the Monte Carlo study (Section 5), we use these approximating critical value functions for computational speed.

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 estimators are based on the order statistics of the residuals $\hat{\varepsilon}_t$. Under a suitable consistency condition, the HRT based on $\hat{f}$ asymptotically is conjectured to behave 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.

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 performance 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 Elliott, Rothenberg and Stock (1996). Second, as $\hat{\sigma}_{\varepsilon\phi_g}$ appears in the denominator in the AHRT statistic (4.12), the statistic becomes better behaved for values of $\hat{\sigma}_{\varepsilon\phi_g}$ close to zero. The corresponding critical value functions for various reference densities are provided in Table 1. The estimators for σ_f^2 and $\sigma_{\varepsilon\phi_g}$ we use are

$$\hat{\sigma}_f^2 = \frac{1}{T-p-1}\sum_{t=p+2}^{T}\left(\widehat{\Gamma}(L)\left(\Delta Y_t - \frac{1}{T}\sum_{t=p+2}^{T}\Delta Y_t\right)\right)^2,$$

$$\hat{\sigma}_{\varepsilon\phi_g} = \frac{1}{T-p-1}\sum_{t=p+2}^{T}\frac{\widehat{\Gamma}(L)\Delta Y_t}{\hat{\sigma}_f}\sigma_g\phi_g\left(G^{-1}\left(\frac{\hat{R}_t}{T-p}\right)\right).$$

Moreover, to simplify the notation, we denote the Approximate Hybrid Rank Test with reference density g by AHRT-g and, in particular, by AHRT-ϕ for Gaussian reference density and by AHRT-$\hat{f}$ for a (nonparametrically) estimated reference density. To be specific, for the nonparametric estimation of f, we employ a kernel-based method $\hat{f}(x) = (T\mathfrak{h})^{-1}\sum_{t=1}^{T}K((x-\hat{\varepsilon}_t)/\mathfrak{h})$, where the kernel K is chosen to be Gaussian, $\mathfrak{h}$ is the bandwidth chosen by the rule $(4/(3T))^{\frac{1}{5}}\hat{\sigma}_f$, and $\hat{\varepsilon}_t$ are the residuals from the regression (5.1) below. Throughout, we use significance level $\alpha = 5\%$ and all results are based on 20,000 Monte-Carlo replications.

We compare the performances of the propsed AHRT tests 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

$$Y_t = \mu + \rho Y_{t-1} + \sum_{i=1}^{p}\Gamma_i\Delta Y_{t-i} + \varepsilon_t. \tag{5.1}$$

The critical values for this test are -13.52 for $T = 100$ and -14.05 for $T = 2500$. The second competitor is the ERS 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

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

where $\hat{\beta}$ is estimated by regressing $Y_{\bar{\alpha}}$ on $Z_{\bar{\alpha}}$. The long-run variance estimator (for ERS test), $\hat{\omega}^2$, is chosen to be $\hat{\omega}^2_{AR(p)} = \hat{\sigma}_e^2/(1-\sum_{i=1}^{p}\widehat{\Gamma}_i)^2$ with $\hat{\sigma}_e^2 = \sum_{t=1}^{T}\hat{\varepsilon}_t^2/T$, where the residuals $\hat{\varepsilon}_t$ and coefficient estimates $\widehat{\Gamma}_i$ are from the regression in (5.1). 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 DF-GLS test proposed in Elliott, Rothenberg and

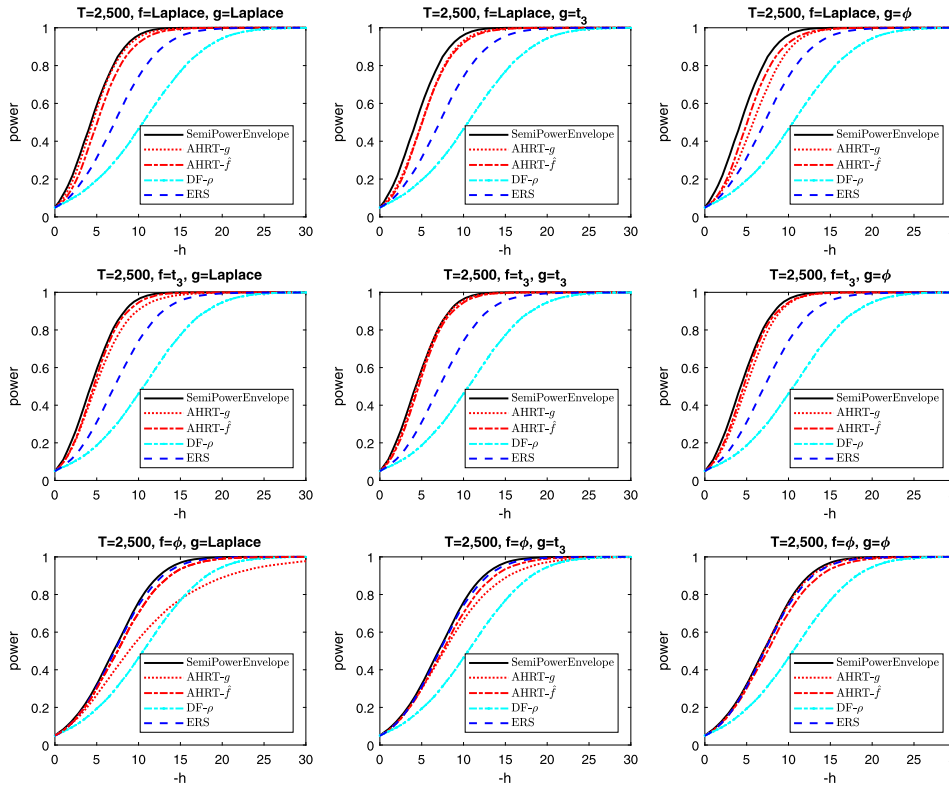

FIG. 1. *IID errors case*: *large-sample power functions of the AHRT-g with various reference densities g, the AHRT-$\hat{f}$ and other selected unit root tests under the true innovation densities f*: *Gaussian*, *Laplace*, *Student's* t_3.

Stock (1996) is also omitted as it behaves asymptotically the same as the ERS test, but can be oversized in smaller samples.

Below, we start with the Monte Carlo study with i.i.d. innovations, that is, the data is generated by the model in (2.1)–(2.3) with $\Gamma(L) = 1$. Therefore, for the newly-proposed AHRT test and its competitors introduced above, we have $\widehat{\Gamma}(L) = 1$ (or, equivalently, $\widehat{\Gamma}_1 = \cdots = \widehat{\Gamma}_p = 0$). The Monte Carlo study with i.i.d. innovations and $f \in \mathfrak{F}_{\mathbb{S}}$, the one with ARMA innovations, and the one illustrating how the AHRT-f test's power function converges to the power envelope as the sample size T increases are provided in the Supplementary Material (Zhou, van den Akker and Werker (2019)).

Large-sample performance. We first use the large-sample performances to illustrate the asymptotic properties. In particular, the chosen sample size T is 2500.

Figure 1 shows the power curves for 9 combinations of 3 innovation densities f and 3 reference densities g (for AHRT-g): f and g are chosen to be

Laplace, Student t_3 or Gaussian.[10] In line with our theoretical results, we find that the AHRT-g test 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 $-h = 7$. This verifies the point-optimal result of the AHRT-f test in Theorem 4.2. The AHRT-$\hat{f}$ test has a similar behavior as the AHRT-f but with a slightly lower power due to the efficiency loss in estimating the density f. This small amount of power loss is also different from case to case, for example, for $f = t_3$, this power loss is almost indistinguishable; and the amount decreases to zero as T goes to infinity. Moreover, when the reference density g is Gaussian (the three right-most graphs), the AHRT-ϕ outperforms its competitors for non-Gaussian f; while for Gaussian f, the AHRT-ϕ test and the ERS test have indistinguishable power (and they outperform the Dickey–Fuller-ρ test). This corroborates the Chernoff–Savage property of the AHRT-ϕ test mentioned in Remark 4.5.

In order to investigate the Chernoff–Savage result for the AHRT-ϕ test even further, we consider in Figure 2 the AHRT-ϕ test for i.i.d errors generated with nine different 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 parameter values 0.5 for stability, 0 for skewness, 1 for scale and 0 location. As these densities do not all satisfy our maintained assumptions, these graphs exclude power envelopes and the AHRT-$\hat{f}$ power functions. The top three graphs in Figure 2 show that the AHRT-ϕ 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-ϕ's power is higher when f is skewed-normal (with skewness 0.8145) than that when f is normal (in Figure 1). This indicates that the AHRT-ϕ can acquire power from skewness. The same conclusion can be drawn from the comparison of the AHRT-ϕ 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 power functions, it validates again 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 tests. In all cases where the true density f satisfies our maintained assumption, that is, $f \in \mathfrak{F}$ (that is all cases in Figure 1 and the skewnormal, t_4, Pearson-I, Pearson-II and Pearson-III in Figure 2), the simulated sizes are between 4.9% and 5.1%. This verifies the validity of the AHRTs

[10]The semiparametric power envelopes are based on 40,000 Monte Carlo replications where the W-processes are approximated by a simple Euler approximation using 2500 grid points.

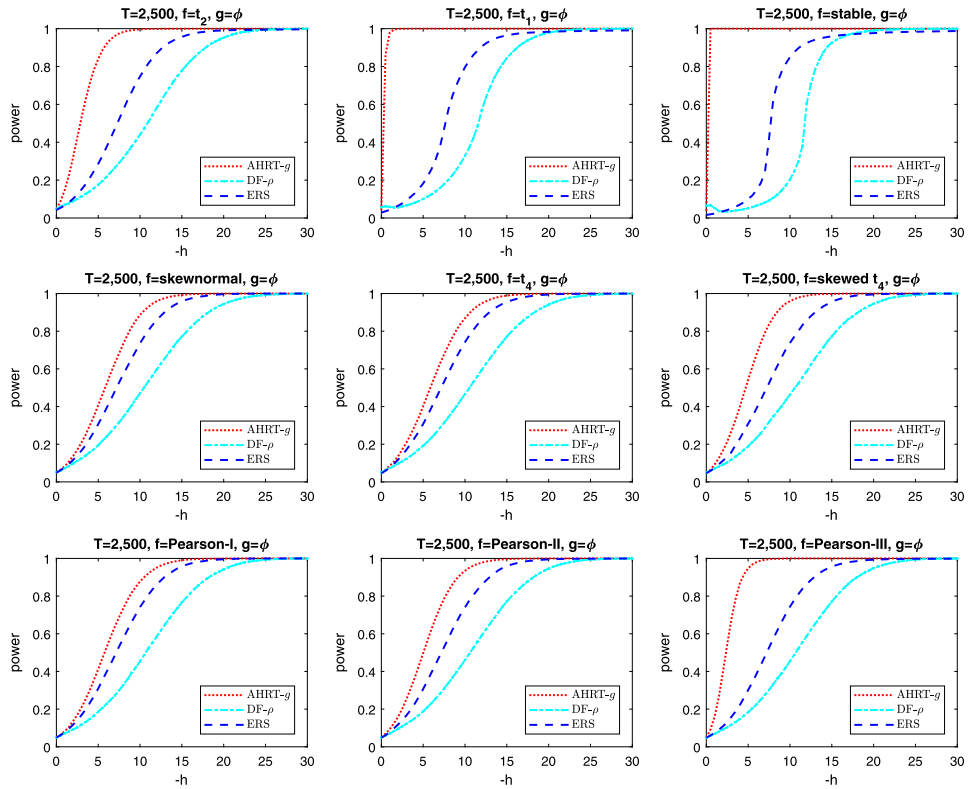

FIG. 2. *IID errors case*: *illustration of the Chernoff–Savage result. The figure shows large-sample power functions of the AHRT-ϕ and other selected unit root tests under various true innovation densities f.*

claimed in Theorem 4.2. In the other cases, that is, $f \notin \mathfrak{F}$, the AHRT is somewhat conservative. More precisely, the simulated sizes of the AHRT-ϕ are 4.8%, 4.1%, 3.7% and 4.7% for the Student t_2, t_1, stable and skew-t_4 distribution, respectively. This result seems consistent over all simulations.

Small-sample performance. We also report the performance of the AHRTs and the two competitors introduced above for smaller samples. Figures 3 and 4 are the small-sample versions, with $T = 100$, of Figures 1 and 2, 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 this small-sample case. For larger values of h, the DF-ρ test sometimes dominates the other two tests. This is due to the fact that the DF-ρ in this Monte Carlo setting appears to have a superior convergence speed (towards its asymptotic power as sample size T increases) to those of the AHRTs and the ERS test. This phenomenon appears only in (ultra)-small-sample cases, and disappears when T gets larger (e.g., $T = 200$).

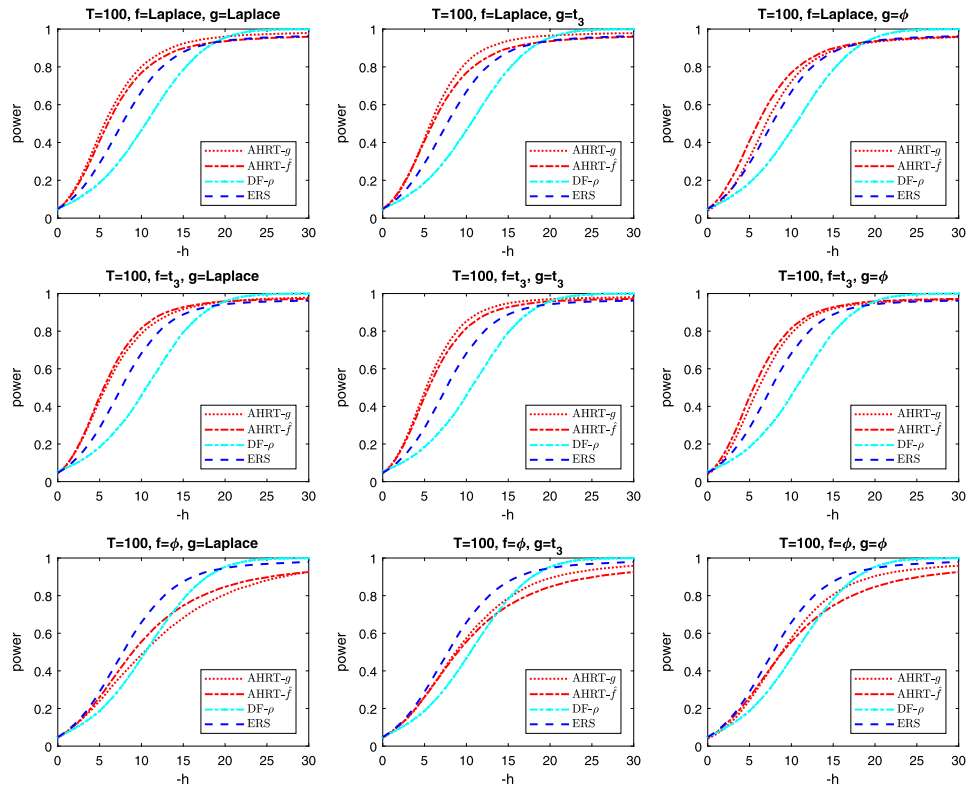

FIG. 3. *IID errors case*: *small-sample power functions of the AHRT-g with various reference densities g*, *the AHRT-* $\hat{f}$ *and other selected unit root tests under various true innovation densities*: *Gaussian*, *Laplace*, *Student* t_3.

In cases with enough samples (say, $T \geq 100$) and when f is significantly away from the Gaussian density, irrespective of the choice of g, the AHRTs performs favorably.

Concerning the small-sample size, we find it to range from about 4.0% to 4.5% for the cases where $f \in \mathfrak{F}$. Again, when f does not satisfy our maintained assumptions ($f \notin \mathfrak{F}$) 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.

6. Conclusion. 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.

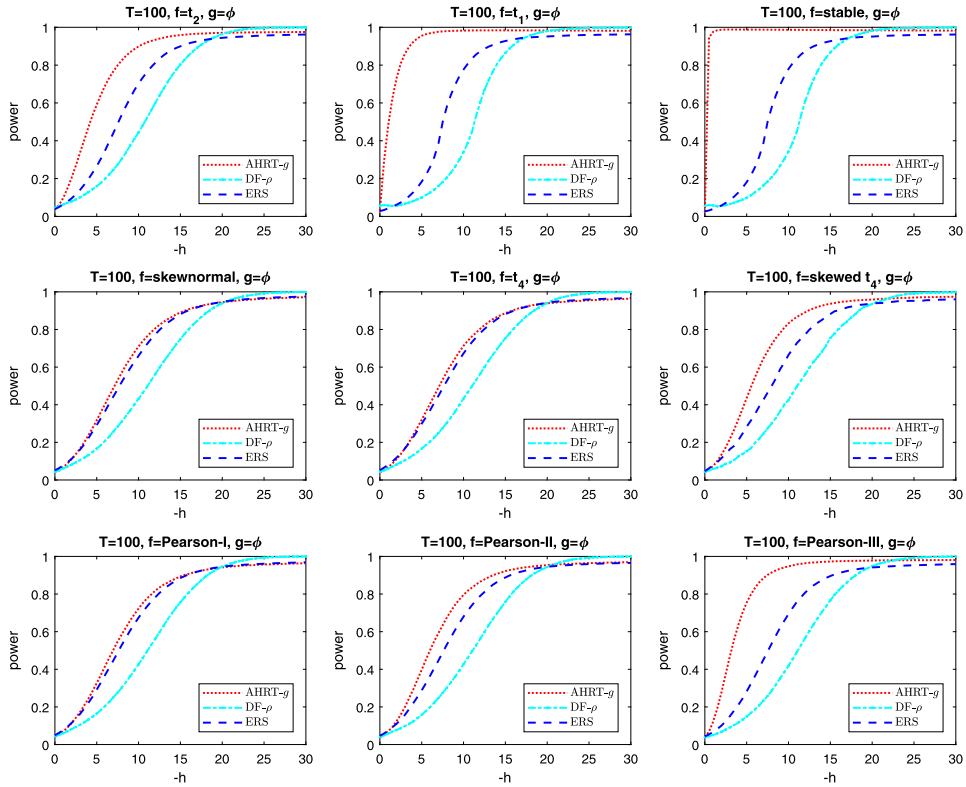

FIG. 4. *IID errors case*: *small-sample power functions of the Gaussian AHRT and other selected unit root tests under some more true innovation densities*.

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 ERS test. Finally, we introduced a simplified version of our test and show, in a Monte Carlo study, that our theoretical results carry over to finite samples.

As potential future work we mention the use of similar ideas to construct hybrid rank-based tests in more general time-series models with, for instance, a deterministic time trend term, or stochastic volatility. Also, the structural representation of the limit experiment and its invariance properties could be applied to other nonstationary time-series models, for instance, cointegration and predictive regression models.

SUPPLEMENTARY MATERIAL

Supplement to "Semiparametrically optimal hybrid rank tests for unit roots" (DOI: 10.1214/18-AOS1758SUPP; .pdf). This supplemental file contains technical proofs for propositions and theorems in the main context.

REFERENCES

AHN, S. K., FOTOPOULOS, S. B. and HE, L. (2001). Unit root tests with infinite variance errors. *Econometric Rev.* **20** 461–483. MR1868343

BICKEL, P. J. (1982). On adaptive estimation. *Ann. Statist.* **10** 647–671. MR0663424

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1998). *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York. Reprint of the 1993 original. MR1623559

BROCKWELL, P. J. and DAVIS, R. A. (2016). *Introduction to Time Series and Forecasting*, 3rd ed. *Springer Texts in Statistics*. Springer, Cham. MR3526245

CALLEGARI, F., CAPPUCCIO, N. and LUBIAN, D. (2003). Asymptotic inference in time series regressions with a unit root and infinite variance errors. *J. Statist. Plann. Inference* **116** 277–303. MR1997142

CASSART, D., HALLIN, M. and PAINDAVEINE, D. (2010). On the estimation of cross-information quantities in rank-based inference. In *Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis*: *A Festschrift in Honor of Professor Jana Jurečková*. *Inst. Math. Stat.* (*IMS*) *Collect.* **7** 35–45. IMS, Beachwood, OH. MR2808364

CHAN, N. H. and WEI, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist.* **16** 367–401. MR0924877

CHERNOFF, H. and SAVAGE, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *Ann. Math. Stat.* **29** 972–994. MR0100322

CHOI, I. (2015). *Almost All About Unit Roots*. *Themes in Modern Econometrics*. Cambridge Univ. Press, New York. MR3558861

DICKEY, D. A. and FULLER, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Amer. Statist. Assoc.* **74** 427–431. MR0548036

DICKEY, D. A. and FULLER, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* **49** 1057–1072. MR0625773

DUFOUR, J.-M. and KING, M. L. (1991). Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary or nonstationary AR(1) errors. *J. Econometrics* **47** 115–143. MR1087209

ELLIOTT, G. and MÜLLER, U. K. (2006). Minimizing the impact of the initial condition on testing for unit roots. *J. Econometrics* **135** 285–310. MR2328403

ELLIOTT, G., ROTHENBERG, T. J. and STOCK, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica* **64** 813–836. MR1399219

HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests*. Academic Press, New York. MR0229351

HALLIN, M. and PURI, M. L. (1988). Optimal rank-based procedures for time series analysis: Testing an ARMA model against other ARMA models. *Ann. Statist.* **16** 402–432. MR0924878

HALLIN, M. and PURI, M. L. (1994). Aligned rank tests for linear models with autocorrelated error terms. *J. Multivariate Anal.* **50** 175–237. MR1293044

HALLIN, M., VAN DEN AKKER, R. and WERKER, B. J. M. (2011). A class of simple distribution-free rank-based unit root tests. *J. Econometrics* **163** 200–214. MR2812866

HALLIN, M., VAN DEN AKKER, R. and WERKER, B. J. M. (2015). On quadratic expansions of log-likelihoods and a general asymptotic linearity result. In *Mathematical Statistics and Limit Theorems* 147–165. Springer, Cham. MR3380735

HASAN, M. N. (2001). Rank tests of unit root hypothesis with infinite variance errors. *J. Econometrics* **104** 49–65. MR1862029

JANSSON, M. (2008). Semiparametric power envelopes for tests of the unit root hypothesis. *Econometrica* **76** 1103–1142. MR2455121

JEGANATHAN, P. (1995). Some aspects of asymptotic theory with applications to time series models. *Econometric Theory* **11** 818–887. MR1458943

JEGANATHAN, P. (1997). On asymptotic inference in linear cointegrated time series systems. *Econometric Theory* **13** 692–745. MR1491257

KLAASSEN, C. A. J. (1987). Consistent estimation of the influence function of locally asymptotically linear estimators. *Ann. Statist.* **15** 1548–1562. MR0913573

KREISS, J.-P. (1987). On adaptive estimation in stationary ARMA processes. *Ann. Statist.* **15** 112–133. MR0885727

LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. *Springer Series in Statistics*. Springer, New York. MR0856411

MÜLLER, U. K. (2011). Efficient tests under a weak convergence assumption. *Econometrica* **79** 395–435. MR2809375

MÜLLER, U. K. and ELLIOTT, G. (2003). Tests for unit roots and the initial condition. *Econometrica* **71** 1269–1286. MR1995831

MÜLLER, U. K. and WATSON, M. W. (2008). Testing models of low-frequency variability. *Econometrica* **76** 979–1016. MR2455119

PATTERSON, K. (2011). *Unit Root Tests in Time Series*. *Vol.* 1. *Palgrave Texts in Econometrics*. Palgrave Macmillan, New York. MR3184412

PATTERSON, K. (2012). *Unit Root Tests in Time Series*. *Vol.* 2. *Palgrave Texts in Econometrics*. Palgrave Macmillan, New York. MR3237671

PHILLIPS, P. C. B. (1987). Time series regression with a unit root. *Econometrica* **55** 277–301. MR0882096

PHILLIPS, P. C. B. and PERRON, P. (1988). Testing for a unit root in time series regression. *Biometrika* **75** 335–346. MR0946054

ROTHENBERG, T. J. and STOCK, J. H. (1997). Inference in a nearly integrated autoregressive model with nonnormal innovations. *J. Econometrics* **80** 269–286. MR1483180

RUDIN, W. (1987). *Real and Complex Analysis*, 3rd ed. McGraw-Hill, New York. MR0924157

SAIKKONEN, P. and LUUKKONEN, R. (1993). Point optimal tests for testing the order of differencing in ARIMA models. *Econometric Theory* **9** 343–362. MR1241978

SCHICK, A. (1986). On asymptotically efficient estimation in semiparametric models. *Ann. Statist.* **14** 1139–1151. MR0856811

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247

WHITE, J. S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *Ann. Math. Stat.* **29** 1188–1197. MR0100952

ZHOU, B., VAN DEN AKKER, R. and WERKER, B. J. M. (2019). Supplement to "Semiparametrically point-optimal hybrid rank tests for unit roots." DOI:10.1214/18-AOS1758SUPP.

B. ZHOU
DEPARTMENT OF INFORMATION SYSTEMS,
BUSINESS STATISTICS
AND OPERATIONS MANAGEMENT
HONG KONG UNIVERSITY
OF SCIENCE AND TECHNOLOGY
CLEAR WATER BAY
KOWLOON
HONG KONG
E-MAIL: bozhouhkust@ust.hk

R. VAN DEN AKKER
B. J. M. WERKER
ECONOMETRICS GROUP AND FINANCE GROUP
TILBURG UNIVERSITY
PO BOX 90153
5000 LE TILBURG
THE NETHERLANDS
E-MAIL: r.vdnakker@tilburguniversity.edu
b.j.m.werker@tilburguniversity.edu