Econometrica, Vol. 76, No. 5 (September, 2008), 1103–1142

SEMIPARAMETRIC POWER ENVELOPES FOR TESTS OF THE UNIT ROOT HYPOTHESIS

BY MICHAEL JANSSON[1]

This paper derives asymptotic power envelopes for tests of the unit root hypothesis in a zero-mean AR(1) model. The power envelopes are derived using the limits of experiments approach and are semiparametric in the sense that the underlying error distribution is treated as an unknown infinite-dimensional nuisance parameter. Adaptation is shown to be possible when the error distribution is known to be symmetric and to be impossible when the error distribution is unrestricted. In the latter case, two conceptually distinct approaches to nuisance parameter elimination are employed in the derivation of the semiparametric power bounds. One of these bounds, derived under an invariance restriction, is shown by example to be sharp, while the other, derived under a similarity restriction, is conjectured not to be globally attainable.

KEYWORDS: Unit root testing, semiparametric efficiency.

1. INTRODUCTION

THE UNIT ROOT TESTING PROBLEM is one of the most intensively studied testing problems in econometrics.[2] During the past decade or so, considerable effort has been devoted to the construction of unit root tests enjoying good power properties.[3] Asymptotic power envelopes for unit root tests in the Gaussian AR(1) model were obtained by Elliott, Rothenberg, and Stock (1996; henceforth ERS) and Rothenberg (2000), while Rothenberg and Stock (1997) derived asymptotic power envelopes under rather general distributional assumptions. Rothenberg and Stock (1997) found that significant power gains (relative to the Gaussian case) are available in cases where the underlying distribution is non-Gaussian and known, and pointed out that this finding is in perfect analogy with well known properties of the location model and the stable AR(1) model. The purpose of this paper is to investigate the extent to which departures from normality can be exploited for power purposes also in the (arguably) more realistic case where the error distribution is unknown. To

[1]The author thanks Marcelo Moreira, Whitney Newey, Jack Porter, Jim Powell, Jim Stock, two anonymous referees, and seminar participants at Aarhus, Berkeley, Boston University, Brown, the 2005 CEME conference, Harvard/MIT, Michigan, Penn, UC Davis, UCLA, UCSD, Wisconsin, and the 2005 World Congress of the Econometric Society for helpful comments. Financial support from CREATES (funded by the Danish National Research Foundation) is gratefully acknowledged.

[2]Important early contributions include Dickey and Fuller (1979, 1981), Phillips (1987), and Phillips and Perron (1988). For reviews, see Stock (1994), Phillips and Xiao (1998), and Haldrup and Jansson (2006).

[3]In parallel with the literature exploring power issues, a different branch of the unit root literature has focused on improving the size properties of unit root tests. Noteworthy contributions in that direction include Ng and Perron (1995, 2001), Perron and Ng (1996), Paparoditis and Politis (2003), and Park (2003).

 DOI: 10.3982/ECTA6113

do so, the paper develops asymptotic power envelopes that are semiparametric in the sense that they explicitly account for the fact that the underlying error distribution is known only to belong to some "big" set of error distributions.

An interesting methodological conclusion emerging from the existing literature on optimality theory for unit root testing is that although there is a fundamental sense in which the unit root testing problem is nonstandard, the problem is still amenable to analysis using existing tools (such as those developed for exponential families and elegantly summarized in Lehmann and Romano (2005)). An important methodological motivation for the present work is the general question of whether semiparametric power envelopes for nonstandard testing problems can be obtained by a conceptually straightforward adaptation of semiparametric methods developed for standard problems. For a variety of reasons, the unit root testing problem seems like a natural starting point for such an investigation and although some of the results obtained in this paper are likely to be somewhat specific to unit root testing, it is hoped that interesting general methodological lessons can be learned from studying that particular problem.

Semiparametric testing theory has been developed for models admitting locally asymptotically normal (LAN) likelihood ratios (e.g., Choi, Hall, and Schick (1996)). In those models, testing theory "has little more to offer than the comforting conclusion that tests based on efficient estimators are efficient" (van der Vaart (1998)). On the other hand, little (if any) work has been done for models outside the LAN class, such as the AR(1) model with a root close to, and possibly equal to, unity. The latter model, which is the model under study here, admits likelihood ratios which are locally asymptotically quadratic (LAQ) in the sense of Jeganathan (1995). No universally accepted definition of estimator efficiency exists for LAQ models.[4] Moreover, the duality between point estimation and hypothesis testing typically breaks down in models whose likelihood ratios are LAQ but not LAN.[5] For these reasons, it appears necessary to develop semiparametric envelopes for the unit root testing problem from first principles.

As is the approach to semiparametric efficiency in standard estimation problems (e.g., Begun, Hall, Huang, and Wellner (1983), Bickel, Klaassen, Ritov, and Wellner (1998), Newey (1990)), the approach to optimality theory taken in this paper is based on Stein's (1956) insight that a testing problem is no easier in a semiparametric model than in any parametric submodel of the semi-

[4]Gushchin (1995) established an optimality property of maximum likelihood estimators. See also Jeganathan (1995) and Ling and McAleer (2003).

[5]An exception occurs in models where the limiting experiment becomes a member of a linear exponential family upon conditioning on statistics with certain ancillarity properties. A well known example is models with locally asymptotically mixed normal (LAMN) likelihood ratios, which arise in cointegration analysis (e.g., Phillips (1991), Stock and Watson (1993)). For an example that does not belong to the LAMN class, see Eliasz (2004) and Jansson and Moreira (2006).

parametric model. Consequently, the semiparametric power envelope will be defined as the infimum of the power envelopes associated with smooth parametric submodels embedding the true error density. Although the unit root testing problem differs from standard testing problems in important respects, it turns out that some of the qualitative findings obtained from the least favorable submodel approach bear a noticeable resemblance to the well known results for the location model, a possibly surprising result in light of the fact that the semiparametric properties of the stable AR(1) model are substantially different from those of the location model. Specifically, it is shown in this paper that although the unit root testing problem admits adaptive procedures when the error distribution is known to be symmetric, adaptation is impossible when the error distribution is (essentially) unrestricted. Nevertheless, and in sharp contrast to the location model, the unit root model with an unrestricted error distribution has the property that although adaptation is impossible, departures from normality can be exploited for efficiency purposes. (The magnitude of the available power gains depends on the shape of the error distribution through its Fisher information for location and can be quite substantial when the error distribution has fat tails.)

The paper proceeds as follows. To set the stage, Section 2 introduces the model and the testing problem under consideration, while Section 3 studies unit root testing under the assumption that the error distribution is known. Section 4 extends the results of Section 3 to parametric submodels. Employing those results, Sections 5 and 6 obtain semiparametric power envelopes for the cases of symmetric and (essentially) unrestricted error distributions, respectively. The consequences of accommodating deterministic components and/or serial correlation in the error are briefly explored in Section 7, while Section 8 offers concluding remarks. Proofs of the main results are provided in the Appendix.

2. PRELIMINARIES

Suppose the observed data $y_1, \ldots, y_T$ are generated by the zero-mean AR(1) model

$$y_t = \rho y_{t-1} + \varepsilon_t, \tag{1}$$

where $y_0 = 0$ and the ε_t are unobserved independent and identically distributed (i.i.d.) errors from an unknown continuous distribution with full support, zero mean, and finite variance. Let f denote the unknown error density. Furthermore, and without loss of generality, let the (unknown) error variance be normalized to 1.

The objective of this paper is to develop asymptotic power envelopes for the unit root testing problem in the zero-mean AR(1) model, treating f as an unknown nuisance parameter. In other words, the testing problem under consideration is

$$H_0 : \rho = 1 \quad \text{vs.} \quad H_1 : \rho < 1,$$

and it is assumed to be known that f lies in some set $\mathcal{F}$ of densities. The main goal of the paper is to develop sharp upper bounds on the asymptotic performance of unit root testing procedures in models of this type, with special attention being devoted to semiparametric cases in which $\mathcal{F}$ is infinite-dimensional.

By Donsker's theorem (e.g., Billingsley (1999)), the assumptions $y_0 = 0$ and $\varepsilon_t \sim$ i.i.d.$(0, 1)$ ensure that if $\rho = 1$, then y_t is $I(1)$ in the sense that the weak limit of $T^{-1/2} y_{\lfloor T \cdot \rfloor}$ is a Brownian motion, where $\lfloor \cdot \rfloor$ denotes the integer part of the argument. While Donsker's theorem is valid without the additional assumption that the error distribution is continuous and has full support, most of the statistical analysis conducted in this paper would be invalid without an assumption of this kind.[6] Specifically, the additional assumption on the error distribution implies that the distributions of $\{y_1, \ldots, y_T\}$ induced by different values of ρ are mutually absolutely continuous. Mutual absolute continuity is a finite sample counterpart of the property of (mutual) contiguity, which plays a prominent role in Le Cam's (1972) theory of limits of experiments and will be utilized throughout this paper.[7]

Because the purpose of this paper is to elucidate the role of $\mathcal{F}$ in optimality theory for unit root tests, Sections 3–6 study the zero-mean AR(1) model, which deliberately assumes away deterministic components and serial correlation in the error.[8] Section 7 explores the consequences of relaxing these (implausible) assumptions and finds, in perfect analogy with ERS's results for the Gaussian case, that the results obtained for the zero-mean AR(1) model extend readily to a model with an unknown mean and serial correlation in the error, whereas the presence of a time trend affects the asymptotic power envelope(s).

[6]If the innovation distribution has bounded support, then the conditional distribution of y_t given y_{t-1} has parameter-dependent support, a property which introduces nontrivial complications even in models with i.i.d. data (e.g., Hirano and Porter (2003), Chernozhukov and Hong (2004)).

[7]The property of mutual contiguity is useful in part because it makes it possible to derive conclusions about local asymptotic power functions from assumptions concerning the behavior of certain statistics "under the null," an attractive feature because assumptions of the latter kind tend to be relatively easy to verify. Examples of readily verifiable assumptions required to hold "under the null" are provided by Assumptions LAQ and LAQ* in Sections 3 and 4, respectively, and the condition (15) that underlies the definition of adaptation employed in Section 5.

[8]The model (1) furthermore sets the initial condition y_0 equal to zero and assumes away conditional heteroskedasticity. Proceeding along the lines of Müller and Elliott (2003) and Boswijk (2005), respectively, it may be possible to relax these assumptions, but no attempts to do so will be made in this paper.

3. KNOWN ERROR DISTRIBUTION

In an attempt to further motivate the question addressed by this paper and to facilitate the interpretation of the main results, this section discusses asymptotic optimality theory for the unit root testing problem under the counterfactual assumption that the underlying error distribution is known (i.e., that $\mathcal{F}$ is a singleton). Even if f is known, the unit root testing problem is nonstandard. A well known manifestation of the nonstandard nature of the unit root testing problem is that contiguous alternatives to the unit root null are of the form $\rho = 1 + O(1/T)$. Accordingly, the parameter of interest is henceforth taken to be $c = T(\rho - 1)$, the associated formulation of the unit root testing problem being $H_0 : c = 0$ vs. $H_1 : c < 0$.

Any (possibly randomized) unit root test can be represented by means of a test function $\phi_T : \mathbb{R}^T \to [0, 1]$ such that H_0 is rejected with probability $\phi_T(Y)$ whenever $Y_T := (y_1, \dots, y_T)' = Y$. The power function (with argument c) associated with ϕ_T is given by $E_{\rho_T(c)}\phi_T(Y_T)$, where $\rho_T(c) := 1 + c/T$ and the subscript on E indicates the distribution with respect to which the expectation is taken.

Define the log likelihood ratio function

$$L_T^f(c) := \sum_{t=2}^{T} \log f\left(\Delta y_t - \frac{c}{T} y_{t-1}\right) - \sum_{t=2}^{T} \log f(\Delta y_t).$$

For any $\alpha \in (0, 1)$ and any sample size T, it follows from the Neyman–Pearson lemma that the optimal size α unit root test against the point alternative $c = \bar{c} < 0$ rejects for large values of $L_T^f(\bar{c})$. The power (against the alternative $c = \bar{c}$) of this point optimal test gives the value of the size α power envelope at $c = \bar{c}$.

Under mild assumptions on f, the finite sample power envelope has an asymptotic counterpart which depends on f only through a scalar functional. A sequence of unit root tests ϕ_T is said to have asymptotic size α if

$$\lim_{T\to\infty} E_{\rho_T(0)}\phi_T(Y_T) = \alpha.$$

The asymptotic power envelope for unit root tests of asymptotic size α will be derived under the following high-level assumption on f, in which $o_{p_{0,f}}(1)$ is shorthand for "$o_p(1)$ when H_0 holds and ε has density f" and $\mathcal{L}_f$ denotes the set of functions ℓ_f for which $E[\ell_f(\varepsilon)] = 0$, $E[\varepsilon\ell_f(\varepsilon)] = 1$, and $1 \le E[\ell_f(\varepsilon)^2] < \infty$.

ASSUMPTION LAQ: *If c_T is a bounded sequence, then*

$$L_T^f(c_T) = c_T S_T^f - \frac{1}{2} c_T^2 H_T^{ff} + o_{p_{0,f}}(1),$$

where, for some $\ell_f \in \mathcal{L}_f$,

$$S_T^f := \frac{1}{T}\sum_{t=2}^{T} y_{t-1}\ell_f(\Delta y_t), \quad H_T^{ff} := \frac{\mathcal{I}_{ff}}{T^2}\sum_{t=2}^{T} y_{t-1}^2, \quad \mathcal{I}_{ff} := E[\ell_f(\varepsilon)^2].$$

Assumption LAQ is in the spirit of Jeganathan (1995) and implies that the likelihood ratios are LAQ at $\rho = 1$ in the sense of that paper. In particular, it follows from Donsker's theorem and Chan and Wei (1988, Theorem 2.4) that

$$L_T^f(c) \to_{d_{0,f}} \Lambda_f(c) := c\mathcal{S}_f - \frac{1}{2}c^2\mathcal{H}_{ff} \quad \forall c, \tag{2}$$

where

$$\mathcal{S}_f := \int_0^1 W(r)\,dB_f(r), \quad \mathcal{H}_{ff} := \mathcal{I}_{ff}\int_0^1 W(r)^2\,dr,$$

(W, B_f) is a bivariate Brownian motion with

$$\operatorname{Var}\begin{pmatrix} W(1) \\ B_f(1) \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & \mathcal{I}_{ff} \end{pmatrix},$$

and $\to_{d_{0,f}}$ is shorthand for "$\to_d$ when H_0 holds and ε has density f." Additional discussion of Assumption LAQ, including sufficient conditions for its validity, will be given at the end of this section.

Prohorov's theorem (e.g., Billingsley (1999)) and Le Cam's third lemma (e.g., van der Vaart (2002)) can be used to show that if (2) holds, then every subsequence $\phi_{T'}$ admits a further subsequence $\phi_{T''}$ and a [0, 1]-valued function ψ for which

$$\lim_{T''\to\infty} E_{\rho_{T''}(c)}\phi_{T''}(Y_{T''}) = E\big[\psi(\mathcal{S}_f, \mathcal{H}_{ff})\exp(\Lambda_f(c))\big] \quad \forall c. \tag{3}$$

If ϕ_T has asymptotic size α, then ψ in (3) satisfies $E[\psi(\mathcal{S}_f, \mathcal{H}_{ff})] = \alpha$ and it follows from the Neyman–Pearson lemma that $E[\psi(\mathcal{S}_f, \mathcal{H}_{ff})\exp(\Lambda_f(c))]$ is bounded from above by

$$\Psi_f(c, \alpha) := E\big[\psi_f(\mathcal{S}_f, \mathcal{H}_{ff}|c, \alpha)\exp(\Lambda_f(c))\big],$$

where

$$\psi_f(\mathcal{S}_f, \mathcal{H}_{ff}|c, \alpha) := 1[\Lambda_f(c) > K_\alpha(c; \mathcal{I}_{ff})],$$

$1[\cdot]$ is the indicator function, and $K_\alpha(c;\mathcal{I}_{ff})$ is the $1-\alpha$ quantile of $\Lambda_f(c)$. These facts yield the following theorem, which generalizes a result of ERS to non-Gaussian error distributions.[9]

THEOREM 1: *If Assumption LAQ holds and ϕ_T has asymptotic size α, then*

$$\overline{\lim_{T\to\infty}} E_{\rho_T(c)}\phi_T(Y_T) \le \Psi_f(c,\alpha) \quad \forall c<0. \tag{4}$$

The proof of Theorem 1 given above is based on Le Cam's (1972) theory of limits of experiments. Because f is assumed to be known, a Neyman–Pearson test exists for every T and the use of the theory of limits of experiments can be avoided (e.g., Rothenberg and Stock (1997)). On the other hand, the use of the limits of experiments approach seems unavoidable when studying the models under consideration in the following sections. Specifically, the presence of a nuisance parameter governing distributional shape makes it very difficult (if not impossible) to derive a Neyman–Pearson-type test for any given T. In contrast, the limits of experiments approach is applicable also when f depends on a nuisance parameter, because the limiting experiments associated with such models do admit Neyman–Pearson-type tests.

The asymptotic power bound $\Psi_f(c,\alpha)$ is attainable pointwise (in c) when f is known, $\overline{\lim}_{T\to\infty}$ on the left-hand side of (4) equaling $\lim_{T\to\infty}$ and the inequality being sharp when $\phi_T(Y_T)$ equals

$$\phi_{f,T}(Y_T|c,\alpha) := 1\left[cS_T^f - \frac{1}{2}c^2H_T^{ff} > K_\alpha(c;\mathcal{I}_{ff})\right],$$

the natural finite sample counterpart of $\psi_f(\mathcal{S}_f,\mathcal{H}_{ff}|c,\alpha)$. Moreover, it was found by Rothenberg and Stock (1997) that the local asymptotic power function associated with $\phi_{f,T}(\cdot|\bar{c},\alpha)$ is uniformly (in c) "close" to $\Psi_f(c,\alpha)$ if $\bar{c}$ is chosen appropriately. By implication, Ψ_f is a relevant benchmark.

The envelope Ψ_f depends on f only through $\mathcal{I}_{ff}$. It can be shown that Ψ_f is strictly increasing in $\mathcal{I}_{ff}$ and that $\mathcal{I}_{ff}\ge 1$ with equality if and only if f is the standard normal density. Moreover, ERS's unit root tests, based on the Gaussian (quasi-) likelihood, have local asymptotic power functions that are invariant with respect to f. As a consequence, the Gaussian power envelopes derived by ERS provide a lower bound on maximal attainable local asymptotic power in models with non-Gaussian errors.

Figure 1 plots $\Psi_f(\cdot,0.05)$ for various values of $\mathcal{I}_{ff}$, thereby quantifying the magnitude of the potential power gains, relative to procedures based on a

[9]Theorem 1 is essentially due to Rothenberg and Stock (1997), who obtained a result equivalent to (4) under the (somewhat stronger) assumptions that (i) $E[|\varepsilon|^k + |\ell_f(\varepsilon)|^k] < \infty$ for some $k>2$, where $\ell_f(\varepsilon) := \partial \log f(\varepsilon-\theta)/\partial\theta|_{\theta=0}$, and (ii) ℓ_{ff} satisfies a linear Lipschitz condition, where $\ell_{ff}(\varepsilon) := \partial^2 \log f(\varepsilon-\theta)/\partial\theta^2|_{\theta=0}$.

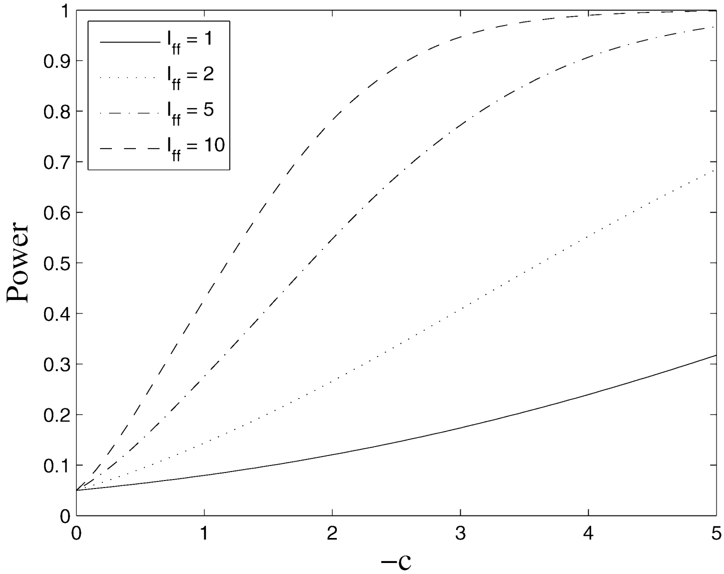

FIGURE 1.—Power envelopes, known error distribution.

Gaussian quasi-likelihood, available in applications with nonnormal errors.[10] Evidently, substantial power gains will be available (for models with $\mathcal{I}_{ff}$ well above unity) if it is possible to construct a unit root test which is computable without knowledge of f and attains Ψ_f for every $f \in \mathcal{F}$. Section 5 shows that this situation occurs when $\mathcal{F}$ consists only of symmetric error densities. More generally, Figure 1 suggests that nontrivial power gains will be available in situations where (attainable) semiparametric power envelopes are qualitatively similar to Ψ_f in the sense that they lie well above the envelope corresponding to the Gaussian distribution. Section 6 shows that this occurs even when $\mathcal{F}$ is unrestricted.

[10]Consider the density given by $f_\lambda(\varepsilon) := C_\lambda^0 \exp(-C_\lambda^1|\varepsilon|^\lambda)$, where $\lambda > 1/2$ and the constants C_λ^0 and C_λ^1 are determined by the requirement

$$\int_{-\infty}^{\infty} f_\lambda(\varepsilon)\,d\varepsilon = \int_{-\infty}^{\infty} \varepsilon^2 f_\lambda(\varepsilon)\,d\varepsilon = 1.$$

The values $\lambda = 2$ and $\lambda = 1$ correspond to the standard normal and rescaled double exponential distributions, respectively, and the associated values of $\mathcal{I}_{ff}$ are 1 and 2, respectively. More generally, it can be shown that the value of $\mathcal{I}_{ff}$ associated with f_λ is given by

$$\mathcal{I}_{ff}(\lambda) := \lambda^2 \frac{[\int_0^\infty r^2 \exp(-r^\lambda)\,dr][\int_0^\infty r^{2(\lambda-1)} \exp(-r^\lambda)\,dr]}{[\int_0^\infty \exp(-r^\lambda)\,dr]^2}.$$

Because $\lim_{\lambda\downarrow 1/2} \mathcal{I}_{ff}(\lambda) = \infty$ and $\mathcal{I}_{ff}(\cdot)$ is continuous, the range of $\mathcal{I}_{ff}(\cdot)$ is $[1, \infty)$. Numerical evaluation shows that $\mathcal{I}_{ff}(0.7709) \approx 5$ and $\mathcal{I}_{ff}(0.6818) \approx 10$, respectively.

Assumption LAQ holds for a wide range of error distributions. For instance, Jeganathan (1995) showed that Assumption LAQ is satisfied (with $\ell_f = -\dot{f}/f$) under the following absolute continuity condition on f.

ASSUMPTION AC: *The density f admits a function $\dot{f}$ such that $f(\varepsilon) = \int_{-\infty}^{\varepsilon} \dot{f}(r)\,dr$ for every $\varepsilon \in \mathbb{R}$ and $\int_{-\infty}^{\infty} [\dot{f}(\varepsilon)^2/f(\varepsilon)]\,d\varepsilon < \infty$.*

Under Assumption AC, ℓ_f is the score function, evaluated at $\theta = 0$, for θ in the location model

$$X_i = \theta + \varepsilon_i, \tag{5}$$

where the ε_i are i.i.d. with density f. Similarly, $\mathcal{I}_{ff}$ is the Fisher information for location associated with f. As a consequence, both ℓ_f and $\mathcal{I}_{ff}$ are readily interpretable. In the location model (5), Assumption AC serves dual purposes: it delivers the LAN property (i.e., a quadratic expansion of the log likelihood ratio function) and enables nonparametric estimation of ℓ_f. Assumption AC will serve similar purposes in Theorems 4 and 6 of this paper.

The following Le Cam (1970) type of assumption is implied by Assumption AC.[11]

ASSUMPTION DQM: *The density f admits a function ℓ_f such that, as $|\theta| \to 0$,*

$$\int_{-\infty}^{\infty} \left(\sqrt{\frac{f(\varepsilon - \theta)}{f(\varepsilon)}} - 1 - \frac{1}{2}\theta \ell_f(\varepsilon) \right)^2 f(\varepsilon)\,d\varepsilon = o(\theta^2).$$

For the purposes of establishing just the LAN property (in the location model), it is well known that differentiability in quadratic mean (Assumption DQM) suffices.[12] It seems natural to ask if the model studied in this paper exhibits a similar feature. An affirmative answer to that question is provided by the following lemma, which therefore shows that the usefulness of Assumption DQM extends beyond the class of models whose likelihood ratios enjoy the LAN property.[13]

LEMMA 2: *Assumption LAQ is implied by Assumption DQM.*

[11] Additional discussion of the relationship between assumptions of the absolute continuity (AC) and differentiability in quadratic mean (DQM) variety can be found in Le Cam (1986, Section 17.3) and Le Cam and Yang (2000, Section 7.3).

[12] Indeed, van der Vaart (2002, p. 676) argued that Assumption DQM is "exactly right for getting the LAN expansion (in the location model)." For an appreciation of differentiability in quadratic mean, see Pollard (1997).

[13] A proof of Lemma 2 can be found in Jansson (2007).

4. PARAMETRIC SUBMODELS

Relaxing the assumption that the error density is known, this section studies unit root testing in parametric submodels. In the present context, a parametric submodel is a model of the form (1) with f embedded in a parametric family $F := \{f(\cdot|\eta) : \eta \in \mathbb{R}\}$ of density functions satisfying

$$\int_{-\infty}^{\infty} \varepsilon f(\varepsilon|\eta)\, d\varepsilon = 0 \quad \text{and} \quad \int_{-\infty}^{\infty} \varepsilon^2 f(\varepsilon|\eta)\, d\varepsilon < \infty$$

for each value of the (nuisance) parameter η.

It is assumed that $f(\cdot) = f(\cdot|0)$; that is, it is assumed that the true (but unknown) value of η is zero. Moreover, the family F is assumed to be "smooth" at $\eta = 0$. Among other things, "smoothness" will imply that the contiguous alternatives to $\eta = 0$ are of the form $\eta = O(1/\sqrt{T})$. In recognition of this fact, all subsequent formulations will employ a local reparameterization of the form $\eta = \eta_T(h) := h/\sqrt{T}$, where the true (but unknown) value of the local parameter h is zero.

Let the log likelihood ratio function associated with F be denoted by

$$L_T^F(c, h) := \sum_{t=2}^{T} \log f\left[\Delta y_t - \frac{c}{T} y_{t-1} \middle| \eta_T(h)\right] - \sum_{t=2}^{T} \log f[\Delta y_t | \eta_T(0)],$$

and let $\mathcal{L}_\eta$ denote the class of functions ℓ_η for which $E[\ell_\eta(\varepsilon)] = 0$, $E[\varepsilon \ell_\eta(\varepsilon)] = 0$, and $E[\ell_\eta(\varepsilon)^2] < \infty$. The degree of smoothness assumed on the part of F is made precise by the following high-level assumption, which generalizes Assumption LAQ to parametric submodels.

ASSUMPTION LAQ*: *If* (c_T, h_T) *is a bounded sequence, then*

$$L_T^F(c_T, h_T) = (c_T, h_T) S_T^F - \frac{1}{2}(c_T, h_T) H_T^F (c_T, h_T)' + o_{p_{0,f}}(1),$$

where, for some $\ell_F := (\ell_f, \ell_\eta)' \in \mathcal{L}_f \times \mathcal{L}_\eta$,

$$S_T^F := \begin{pmatrix} S_T^f \\ S_T^\eta \end{pmatrix} := \begin{pmatrix} \frac{1}{T} \sum_{t=2}^{T} y_{t-1} \ell_f(\Delta y_t) \\ \frac{1}{\sqrt{T}} \sum_{t=2}^{T} \ell_\eta(\Delta y_t) \end{pmatrix},$$

$$H_T^F := \begin{pmatrix} H_T^{ff} & H_T^{f\eta} \\ H_T^{f\eta} & H_T^{\eta\eta} \end{pmatrix} := \begin{pmatrix} \frac{\mathcal{I}_{ff}}{T^2} \sum_{t=2}^{T} y_{t-1}^2 & \frac{\mathcal{I}_{f\eta}}{T^{3/2}} \sum_{t=2}^{T} y_{t-1} \\ \frac{\mathcal{I}_{f\eta}}{T^{3/2}} \sum_{t=2}^{T} y_{t-1} & \mathcal{I}_{\eta\eta} \end{pmatrix},$$

$$\mathcal{I}_F := \begin{pmatrix} \mathcal{I}_{ff} & \mathcal{I}_{f\eta} \\ \mathcal{I}_{f\eta} & \mathcal{I}_{\eta\eta} \end{pmatrix} := E[\ell_F(\varepsilon) \ell_F(\varepsilon)'].$$

The requirement $E[\ell_F(\varepsilon)] = (0,0)'$ of Assumption LAQ* is the familiar zero-mean property of scores, while $E[\varepsilon\ell_F(\varepsilon)] = e_1 := (1,0)'$ will be a consequence of the requirement $\int_{-\infty}^{\infty} \varepsilon f(\varepsilon|\eta)\,d\varepsilon = 0$ under mild smoothness conditions. As is true of Assumption LAQ, Assumption LAQ* is in the spirit of Jeganathan (1995) and implies that the likelihood ratios are LAQ in the sense of that paper. Moreover, proceeding as in the proof of Lemma 2 it can be shown that Assumption LAQ* is implied by the following generalization of Assumption DQM.

ASSUMPTION DQM*: *The family F admits functions ℓ_f and ℓ_η such that, as* $|\theta| + |\eta| \to 0$,

$$\int_{-\infty}^{\infty}\left(\sqrt{\frac{f(\varepsilon-\theta|\eta)}{f(\varepsilon)}} - 1 - \frac{1}{2}[\theta\ell_f(\varepsilon) + \eta\ell_\eta(\varepsilon)]\right)^2 f(\varepsilon)\,d\varepsilon$$
$$= o(\theta^2 + \eta^2).$$

The LAQ property delivered by Assumption LAQ* makes it possible to use the limits of experiments approach to derive asymptotic power envelopes for the unit root testing problem in a model where it is assumed to be known only that $f \in F$. To describe the salient properties of the limiting experiment, let $(W, B_F') := (W, B_f, B_\eta)$ be a trivariate Brownian motion with

$$\text{(6)} \qquad \operatorname{Var}\begin{pmatrix} W(1) \\ B_F(1) \end{pmatrix} = \begin{pmatrix} 1 & e_1' \\ e_1 & \mathcal{I}_F \end{pmatrix}.$$

It follows from standard weak convergence results that

$$(S_T^F, H_T^F) \to_{d_{0,f}} (\mathcal{S}_F, \mathcal{H}_F),$$

where

$$\mathcal{S}_F := \begin{pmatrix} \mathcal{S}_f \\ \mathcal{S}_\eta \end{pmatrix} := \begin{pmatrix} \int_0^1 W(r)\,dB_f(r) \\ B_\eta(1) \end{pmatrix},$$
$$\mathcal{H}_F := \begin{pmatrix} \mathcal{H}_{ff} & \mathcal{H}_{f\eta} \\ \mathcal{H}_{f\eta} & \mathcal{H}_{\eta\eta} \end{pmatrix} := \begin{pmatrix} \mathcal{I}_{ff}\int_0^1 W(r)^2\,dr & \mathcal{I}_{f\eta}\int_0^1 W(r)\,dr \\ \mathcal{I}_{f\eta}\int_0^1 W(r)\,dr & \mathcal{I}_{\eta\eta} \end{pmatrix}.$$

Using Prohorov's theorem, Le Cam's third lemma, and the result

$$L_T^F(c,h) \to_{d_{0,f}} \Lambda_F(c,h) := (c,h)\mathcal{S}_F - \frac{1}{2}(c,h)\mathcal{H}_F(c,h)' \quad \forall (c,h),$$

it can be shown that every subsequence $\phi_{T'}$ admits a further subsequence $\phi_{T''}$ and a $[0,1]$-valued function ψ for which

$$\text{(7)} \qquad \lim_{T''\to\infty} E_{\rho_{T''}(c),\eta_{T''}(h)}\phi_{T''}(Y_{T''}) = E\big[\psi(\mathcal{S}_F, \mathcal{H}_F)\exp(\Lambda_F(c,h))\big] \quad \forall (c,h).$$

Because the true, but unknown, value of h has been normalized to zero, asymptotic power envelopes provide sharp upper bounds on $\overline{\lim}_{T\to\infty} E_{\rho_T(c),\eta_T(0)} \times \phi_T(Y_T)$. In view of (7), these bounds can be obtained by maximizing

$$E\big[\psi(\mathcal{S}_F, \mathcal{H}_F)\exp(\Lambda_F(c,0))\big] = E\big[\psi(\mathcal{S}_F, \mathcal{H}_F)\exp(\Lambda_f(c))\big]$$

with respect to ψ. As in Section 3, the tests under consideration will be assumed to be such that the limiting test functions ψ satisfy

$$E[\psi(\mathcal{S}_F, \mathcal{H}_F)] = \alpha. \tag{8}$$

In an attempt to further ensure that the power envelopes account for the presence of the unknown nuisance parameter h, some additional restrictions will be placed on ψ.

Two classes of test functions, motivated by two conceptually distinct approaches to nuisance parameter elimination in the limiting experiment, will be considered. The first class is motivated by the fact that ψ is α-similar in the limiting experiment if and only if

$$E\big[\psi(\mathcal{S}_F, \mathcal{H}_F)\exp(\Lambda_F(0,h))\big] = \alpha \quad \forall h. \tag{9}$$

Accordingly, a sequence ϕ_T is said to be locally asymptotically α-similar (in F) if any ψ satisfying (7) also satisfies (9). The second class is motivated by a location invariance property enjoyed by testing problems involving c in the limiting experiment. As explained in a remark following the proof of Theorem 3, any (location) invariant test in the limiting experiment admits a representation in which $\psi(\mathcal{S}_F, \mathcal{H}_F)$ depends on $(\mathcal{S}_F, \mathcal{H}_F)$ only through $(\mathcal{S}_{f.\eta}, \mathcal{H}_F)$, where

$$\mathcal{S}_{f.\eta} := \mathcal{S}_f - \frac{\mathcal{H}_{f\eta}}{\mathcal{H}_{\eta\eta}}\mathcal{S}_\eta.$$

Accordingly, a sequence ϕ_T is said to be locally asymptotically α-invariant (in F) if any ψ satisfying (7) also satisfies (8) and can be chosen such that

$$\psi(\mathcal{S}_F, \mathcal{H}_F) = E[\psi(\mathcal{S}_F, \mathcal{H}_F)|\mathcal{S}_{f.\eta}, \mathcal{H}_F]. \tag{10}$$

It is shown in the proof of Theorem 3 that if (10) holds, then

$$E\big[\psi(\mathcal{S}_F, \mathcal{H}_F)\exp(\Lambda_F(c,h))\big] = E\big[\psi(\mathcal{S}_F, \mathcal{H}_F)\exp(\Lambda_{f.\eta}(c))\big], \tag{11}$$

where

$$\Lambda_{f.\eta}(c) := c\mathcal{S}_{f.\eta} - \frac{1}{2}c^2\mathcal{H}_{ff.\eta}, \quad \mathcal{H}_{ff.\eta} := \mathcal{H}_{ff} - \frac{\mathcal{H}_{f\eta}^2}{\mathcal{H}_{\eta\eta}}.$$

Because the right-hand side of (11) does not depend on h, the class of locally asymptotically α-invariant tests is contained in the class of locally asymptotically α-similar tests. Both classes of tests contain most (if not all) existing unit root tests. In particular, it can be shown that both classes of tests contain the point optimal tests of ERS as well as the "robust" unit root tests based on M-estimators and/or ranks proposed by Herce (1996), Hasan and Koenker (1997), Thompson (2004), and Koenker and Xiao (2004). On the other hand, the restrictions imposed are not entirely vacuous, as it follows from Theorem 3 that they are violated by the "oracle" test based on $\phi_{f,T}$ unless the submodel satisfies $\mathcal{I}_{f\eta} \neq 0$.

The next result generalizes Theorem 1 to parametric submodels. It is shown in part (a) that

$$\Psi_F^S(c,\alpha) := E\big[\psi_F^S(\mathcal{S}_F, \mathcal{H}_F | c, \alpha)\exp(\Lambda_f(c))\big]$$

provides an upper bound on local asymptotic power for locally asymptotically α-similar tests, where

$$\psi_F^S(\mathcal{S}_F, \mathcal{H}_F | c, \alpha) := 1[\Lambda_f(c) > K_\alpha^S(\mathcal{S}_\eta, c; \mathcal{I}_F)]$$

and K_α^S is the continuous function satisfying $E[\psi_F^S(\mathcal{S}_F, \mathcal{H}_F | c, \alpha) | \mathcal{S}_\eta] = \alpha$.[14] The envelope for locally asymptotically α-invariant tests is shown in part (b) to be given by

$$\Psi_F^I(c,\alpha) := E\big[\psi_F^I(\mathcal{S}_F, \mathcal{H}_F | c, \alpha)\exp(\Lambda_f(c))\big],$$

where

$$\psi_F^I(\mathcal{S}_F, \mathcal{H}_F | c, \alpha) := 1[\Lambda_{f.\eta}(c) > K_\alpha^I(c; \mathcal{I}_F)]$$

and $K_\alpha^I(c; \mathcal{I}_F)$ is the $1-\alpha$ quantile of $\Lambda_{f.\eta}(c)$.

THEOREM 3: (a) *If Assumption* LAQ* *holds and* ϕ_T *is locally asymptotically* α-*similar*, *then*

$$\overline{\lim_{T\to\infty}} E_{\rho_T(c),\eta_T(0)} \phi_T(Y_T) \le \Psi_F^S(c,\alpha) \quad \forall c < 0. \tag{12}$$

(b) *If*, *moreover*, ϕ_T *is locally asymptotically* α-*invariant*, *then*

$$\overline{\lim_{T\to\infty}} E_{\rho_T(c),\eta_T(0)} \phi_T(Y_T) \le \Psi_F^I(c,\alpha) \quad \forall c < 0. \tag{13}$$

[14] A more explicit characterization of K_α^S is given in the proof of Lemma 7 provided in Jansson (2007).

The bounds derived in Theorem 3 are attainable pointwise if F is known, $\overline{\lim}_{T\to\infty}$ on the left-hand sides of (12) and (13) equaling $\lim_{T\to\infty}$ and the inequalities becoming equalities when $\phi_T(Y_T)$ is given by[15]

$$\phi^S_{F,T}(Y_T|c,\alpha) := 1\left[cS^f_T - \frac{1}{2}c^2 H^{ff}_T > K^S_\alpha(S^\eta_T, c; \mathcal{I}_F)\right]$$

and

$$\phi^I_{F,T}(Y_T|c,\alpha) := 1\left[cS^{f.\eta}_T - \frac{1}{2}c^2 H^{ff.\eta}_T > K^I_\alpha(c; \mathcal{I}_F)\right],$$

respectively, where

$$S^{f.\eta}_T := S^f_T - \frac{H^{f\eta}_T}{\mathcal{I}_{\eta\eta}} S^\eta_T, \quad H^{ff.\eta}_T := H^{ff}_T - \frac{H^{f\eta 2}_T}{\mathcal{I}_{\eta\eta}}.$$

Because the class of locally asymptotically α-invariant tests is contained in the class of locally asymptotically α-similar tests, the power envelopes satisfy $\Psi^I_F \leq \Psi^S_F$ by construction. Moreover, the inequality is strict whenever $\mathcal{I}_{f\eta} \neq 0$, implying that the present model differs in an interesting way from models with LAN likelihood ratios. In a Gaussian shift experiment (the limiting experiment in a model with LAN likelihood ratios) with one element of the mean vector being the parameter of interest and the others being unknown nuisance parameters, the class of α-similar tests contains the class of size α location invariant tests. In other words, the natural counterparts of the restrictions (9) and (10) are nested in the same way as they are here. Unlike the limiting experiment of the model studied here, however, the two classes of restrictions give rise to identical power envelopes in a Gaussian shift experiment (because the best α-similar test is location invariant) and there is no ambiguity about what the "correct" power envelope is.[16] In contrast, because Ψ^I_F and Ψ^S_F differ whenever $\mathcal{I}_{f\eta} \neq 0$, it is unclear which (if any) of these envelopes is the "correct" envelope in the present context.

A potential problem with the power envelope Ψ^S_F is that it is perhaps "too local" in the sense that it fails to adequately reflect the fact that the nuisance parameter h is unknown in the limiting experiment. Specifically, whereas $E[\psi(\mathcal{S}_F, \mathcal{H}_F)\exp(\Lambda_F(c,h))]$ will depend on h in general even if ψ satisfies (9), the object being maximized, $E[\psi(\mathcal{S}_F, \mathcal{H}_F)\exp(\Lambda_f(c))]$, does not depend on h.

[15]The functions $\phi^S_{F,T}$ and $\phi^I_{F,T}$ are the natural finite sample counterparts of ψ^S_F and ψ^I_F, respectively.

[16]This observation combined with the fact that (the counterpart of) local asymptotic α-similarity is a weaker restriction than (the counterpart of) local asymptotic α-invariance in models with LAN likelihood ratios would appear to explain why the latter restriction has received little (if any) attention in the existing literature on semiparametric testing theory.

One way to avoid this potential problem, which further helps clarify the relationship between (9) and (10), is to consider a minimax criterion. Using the Hunt–Stein theorem (e.g., Lehmann and Romano (2005, Theorem 8.5.1)), it can be shown that if ψ satisfies (9), then

$$\inf_{h\in\mathbb{R}} E\big[\psi(\mathcal{S}_F, \mathcal{H}_F)\exp(\Lambda_F(c,h))\big] \le \Psi_F^I(c,\alpha), \tag{14}$$

where the inequality becomes an equality when $\psi = \psi_F^I(\cdot|c,\alpha)$. By implication, Ψ_F^I can be interpreted as a minimax power envelope for locally asymptotically α-similar tests. Indeed, if ϕ_T is locally asymptotically α-similar, then

$$\inf_H \varlimsup_{T\to\infty} \min_{h\in H} E_{\rho_T(c),\eta_T(h)}\phi_T(Y_T) \le \Psi_F^I(c,\alpha) \quad \forall c<0,$$

where the inf is taken over all finite subsets H of $\mathbb{R}$. (Moreover, the inequality becomes an equality and $\varlimsup_{T\to\infty}$ can be replaced by $\lim_{T\to\infty}$ when $\phi_T(\cdot) = \phi_{F,T}^I(\cdot|c,\alpha)$.) This fact, a proof of which can be based on (14) and the methods of van der Vaart (1991), would appear to support the conjecture that Ψ_F^I is the correct power envelope. Additional substantiation of that conjecture is provided by a remark at the end of this section.

Because profile likelihood procedures "work" in conventional (parametric or even semiparametric) problems (e.g., Murphy and van der Vaart (1997, 2000)), it may be worth noting that Ψ_F^I has a profile likelihood interpretation. Specifically, the function $\Lambda_{f.\eta}$ appearing in the definition of Ψ_F^I satisfies

$$\Lambda_{f.\eta}(c) = \max_h \Lambda_F(c,h) - \max_h \Lambda_F(0,h).$$

This is not merely a coincidence, as it follows from Lehmann and Romano (2005, Problem 6.9) that the best location invariant tests are of the profile likelihood variety whenever the log likelihood is quadratic in the location parameter.

Comparing the power envelopes of Theorems 1 and 3, it is seen that $\Psi_F^S \le \Psi_f$, the inequality being strict unless $\mathcal{I}_{f\eta} = 0$. Therefore, the asymptotic power bound(s) for unit root tests in a parametric submodel will be strictly lower than the power bound in the model with a known error density unless the submodel satisfies $\mathcal{I}_{f\eta} = 0$. Because the assumption $\ell_\eta \in \mathcal{L}_\eta$ implies $E[\ell_f(\varepsilon)\ell_\eta(\varepsilon)] = 0$ only when $\ell_f(\varepsilon) = \varepsilon$, the only distribution (with full support and unit variance) for which the condition $\mathcal{I}_{f\eta} = 0$ is satisfied for every smooth submodel is the Gaussian distribution. Therefore, the point optimal tests of ERS are locally asymptotically α-similar/α-invariant in any smooth parametric submodel for which $f(\cdot|0)$ is the Gaussian distribution. Conversely, if f is not Gaussian, then the test $\phi_{f,T}(\cdot|c,\alpha)$ of the previous section violates (9) for some smooth parametric submodel F with $f(\cdot|0)$ equal to the (true) density f. By implication, the concept of point optimality, which has proven successful when dealing with the curvature in the unit root model with a known density, cannot be used to handle the nuisance parameter f. Specifically, a test of the form $\phi_{\bar{f},T}(\cdot|c,\alpha)$ will not be "nearly efficient" even if $\bar{f}$ is chosen carefully.

The results mentioned in the preceding paragraph bear a noticeable resemblance to the point estimation results for the location model, but differ in one important respect from the results for the stable AR(1) model. In the location model, the Cramér–Rao bound is given by $\mathcal{I}_{ff}^{-1}$ when f is known and by $(\mathcal{I}_{ff} - \mathcal{I}_{f\eta}^2/\mathcal{I}_{\eta\eta})^{-1}$ in parametric submodels, implying that the bounds coincide if and only if $\mathcal{I}_{f\eta} = 0$. Moreover, the sample mean, the quasi-maximum likelihood estimator of θ based on the Gaussian distribution, is regular in any submodel, whereas the quasi-maximum likelihood estimator $\hat{\theta}_f := \arg\max_\theta \sum_i \log f(X_i - \theta)$ based on the true density f is regular in a submodel F with $f(\cdot|0) = f$ if and only if the submodel has $\mathcal{I}_{f\eta} = 0$, a condition which is violated by some smooth submodels unless the true distribution happens to be Gaussian.

In the location model, the condition that $\mathcal{I}_{f\eta} = 0$ in all smooth submodels permitted by the set $\mathcal{F}$ of densities to which f is assumed to belong is simply Stein's (1956) necessary condition for adaptation. As is well known, this condition is satisfied when f is assumed to be symmetric, but is violated when f is unrestricted. In fact, adaptive estimation is possible in the symmetric location model under Assumption AC (e.g., Beran (1974, 1978), Stone (1975)), whereas the sample average attains the semiparametric efficiency bound in the location model with an (essentially) unrestricted f (e.g., Levit (1975), Newey (1990)), implying in particular that departures from normality cannot be exploited for efficiency purposes in that model. The latter property is not shared by the stable AR(1) model, which admits adaptive estimators even when f is required only to satisfy Assumption AC (e.g., Kreiss (1987a, 1987b), Drost, Klaassen, and Werker (1997)). Therefore, although the stable AR(1) model and the location model exhibit qualitatively identical behavior when the density is known and/or symmetric, they exhibit drastically different behavior in the semiparametric case where f is treated as an unrestricted nuisance parameter.

Utilizing the results of this section, the following two sections develop power envelopes in the semiparametric cases where f is either assumed to be symmetric or is left unrestricted. It will be shown in Section 5 that the unit root model admits adaptive testing procedures when the errors are assumed to be symmetric and satisfy Assumption AC. Consequently, the analogies pointed out by Rothenberg and Stock (1997, p. 278) extend in a predictable way to the semiparametric model in which only symmetry is assumed on the part of the error distributions. In contrast, it is obvious from the results cited in the previous paragraph that these analogies will not extend to the model in which the error distribution is unrestricted. Studying the unit root model with an (essentially) unrestricted f, Section 6 finds that the semiparametric properties of the unit root model are related to the semiparametric properties of both the location model and the stable AR(1) model. On the one hand, a numerical evaluation of the semiparametric power envelopes will show that these can be well above the power envelope corresponding to the Gaussian distribution. By implication, the unit root model shares some of the semiparametric properties of the stable AR(1) model. On the other hand, the analytical characterization of the semiparametric power envelope for the unit root model turns out to be intimately related to the corresponding characterization of the semiparametric power envelope for the location model (and seemingly unrelated to the characterization of the semiparametric power envelope for the stable AR(1) model).

REMARK: A simple heuristic argument (which can easily be made rigorous) shows that "plug-in" versions of $\phi^S_{F,T}(Y_T|c,\alpha)$ will typically fail to attain Ψ^S_F. Indeed, consider

$$\tilde{\phi}^S_{F,T}(Y_T|c,\alpha) := 1\left[c\tilde{S}_T - \frac{1}{2}c^2 H^{ff}_T > K^S_\alpha(\tilde{S}^\eta_T, c; \mathcal{I}_F)\right],$$

where, for some estimator $\tilde{\eta}_T$ of η (and assuming the derivatives exist),

$$\tilde{S}_T := \frac{1}{T}\sum_{t=2}^{T} y_{t-1}\ell_f(\Delta y_t|\tilde{\eta}_T), \quad \ell_f(\Delta y_t|\eta) := \frac{\partial}{\partial\theta}\log f(\Delta y_t - \theta|\eta)\bigg|_{\theta=0},$$

$$\tilde{S}^\eta_T := \frac{1}{\sqrt{T}}\sum_{t=2}^{T} \ell_\eta(\Delta y_t|\tilde{\eta}_T), \quad \ell_\eta(\Delta y_t|\eta) := \frac{\partial}{\partial\eta}\log f(\Delta y_t|\eta).$$

If $\tilde{\eta}_T$ is asymptotically efficient (i.e., best regular), then $T^{1/2}\tilde{\eta}_T = \mathcal{I}^{-1}_{\eta\eta}S^\eta_T + o_{p_{0,f}}(1)$ and $\ell_f(\cdot|\tilde{\eta}_T)$ should be asymptotically equivalent to

$$\ell_f(\cdot) + \ell_{f\eta}(\cdot)\tilde{\eta}_T, \quad \ell_{f\eta}(\cdot) := \frac{\partial}{\partial\eta}\ell_f(\cdot|\eta)\bigg|_{\eta=0},$$

in the sense that

$$\tilde{S}_T = \frac{1}{T}\sum_{t=2}^{T} y_{t-1}[\ell_f(\Delta y_t) + \ell_{f\eta}(\Delta y_t)\tilde{\eta}_T] + o_{p_{0,f}}(1) = S^{f.\eta}_T + o_{p_{0,f}}(1).$$

Moreover, $\tilde{S}^\eta_T$ should be $o_{p_{0,f}}(1)$, so it should be the case that

$$\tilde{\phi}^S_{F,T}(Y_T|c,\alpha) = 1\left[cS^{f.\eta}_T - \frac{1}{2}c^2 H^{ff}_T > K^S_\alpha(0, c; \mathcal{I}_F)\right] + o_{p_{0,f}}(1).$$

Replacing $\eta = 0$ by an asymptotically efficient estimator is seen to have a nonnegligible impact on the properties of the test. Indeed, the statistics

$\tilde{\phi}^S_{F,T}(Y_T|c,\alpha)$ and $\phi^S_{F,T}(Y_T|c,\alpha)$ are asymptotically equivalent if and only if $\mathcal{I}_{f\eta}=0$. By implication, plug-in versions of $\phi^S_{F,T}(Y_T|c,\alpha)$ generally fail to attain Ψ^S_F. In fact, it follows from the preceding display that $\tilde{\phi}^S_{F,T}(\cdot|c,\alpha)$ is locally asymptotically α-invariant in F and therefore satisfies

$$\overline{\lim_{T\to\infty}}\, E_{\rho_T(c),\eta_T(0)}\tilde{\phi}^S_T(Y_T|c,\alpha;F)\le \Psi^I_F(c,\alpha).$$

The failure of the plug-in approach in this example casts serious doubt on the relevance of the power envelope Ψ^S_F. In contrast, Ψ^I_F is easily shown to be attained by plug-in versions of $\phi^I_{F,T}(Y_T|c,\alpha)$. It will be shown in Section 6 that this property extends to semiparametric models.

5. SYMMETRIC ERROR DISTRIBUTIONS

This section studies unit root testing in the case where f is assumed to belong to $\mathcal{F}^{\mathbb{S}}_{\mathrm{AC}}$, the set of symmetric densities satisfying Assumption AC. As discussed in the previous section, Stein's (1956) necessary condition for adaptation in the location model is also necessary and sufficient for the power envelopes Ψ_f, Ψ^S_F, and Ψ^I_F to coincide for every smooth submodel F permitted by the set $\mathcal{F}$ of densities to which f is assumed to belong. This necessary condition is satisfied when f is assumed to belong to $\mathcal{F}^{\mathbb{S}}_{\mathrm{AC}}$. Theorem 4, the main result of this section, shows that the assumption $f\in\mathcal{F}^{\mathbb{S}}_{\mathrm{AC}}$ is also sufficient for adaptive unit root testing to be possible.

In models with LAN likelihood ratios, the duality between point estimation and hypothesis testing in Gaussian shift experiments (e.g., Choi, Hall, and Schick (1996)) implies that associated with any "reasonable" definition of adaptation for point estimators (e.g., Bickel (1982), Begun et al. (1983)) there is a "reasonable" definition of adaptation for hypothesis tests. On the other hand, because the duality between point estimation and hypothesis testing breaks down in models with LAQ (but not LAN/LAMN) likelihood ratios, some care must be exercised when defining adaptation in the context of the model studied in this paper. In particular, although Ling and McAleer's (2003) definition of adaptation for point estimators generalizes Bickel's (1982) definition to models of the form considered in this paper, it is unclear whether that definition can be translated into a reasonable definition of adaptation for hypothesis tests.

It is by no means difficult to give a reasonable definition of adaptation for tests of the unit root hypothesis. Nevertheless, it seems more attractive to work with a notion of adaptation that depends only on the model under consideration and makes no reference to any particular type of inference (e.g., point estimation or hypothesis testing). Accordingly, the collection $\mathcal{F}^{\mathbb{S}}_{\mathrm{AC}}$ is said to permit adaptive inference if there exists a pair $(\hat{S}_T,\hat{H}_T)$ of statistics such that

$$(\hat{S}_T,\hat{H}_T)=(S^f_T,H^f_T)+o_{P_{0,f}}(1)\quad\forall f\in\mathcal{F}^{\mathbb{S}}_{\mathrm{AC}}. \tag{15}$$

Because (S_T^f, H_T^f) is asymptotically sufficient when f is known and satisfies Assumption LAQ, the present definition is a natural formalization of the requirement that no information is lost (asymptotically) when the density is treated as an unknown nuisance parameter belonging to the set $\mathcal{F}_{AC}^{\mathbb{S}}$.[17]

Theorem 4 shows that $\mathcal{F}_{AC}^{\mathbb{S}}$ permits adaptive inference and uses that result to derive an adaptation result for unit root tests. To describe the latter result, suppose $(\hat{S}_T, \hat{H}_T)$ satisfies (15) and let

$$(16)\qquad \hat{\phi}_T(Y_T|c,\alpha) := 1\left[c\hat{S}_T - \frac{1}{2}c^2\hat{H}_T > K_\alpha(c;\hat{\mathcal{I}}_T)\right], \quad \hat{\mathcal{I}}_T := \frac{\hat{H}_T}{\frac{1}{T^2}\sum_{t=2}^T y_{t-1}^2}.$$

If (15) holds, then a test based on $\hat{\phi}_T(\cdot|c,\alpha)$ will be asymptotically equivalent to the oracle test based on $\phi_{f,T}(\cdot|c,\alpha)$, an adaptation property in view of the fact that $\phi_{f,T}(\cdot|c,\alpha)$ attains $\Psi_f(\cdot)$ and is locally asymptotically α-invariant in $F \subseteq \mathcal{F}_{AC}^{\mathbb{S}}$ whenever F satisfies Assumption DQM*.

As candidate "estimators" of S_T^f and H_T^f, consider

$$(17)\qquad \hat{S}_T := \frac{1}{T}\sum_{t=2}^T y_{t-1}\hat{\ell}_{Tt}^{\mathbb{S}}(\Delta y_t)$$

and

$$(18)\qquad \hat{H}_T := \frac{\hat{\mathcal{I}}_T}{T^2}\sum_{t=2}^T y_{t-1}^2, \quad \hat{\mathcal{I}}_T := \frac{1}{T}\sum_{t=1}^T \hat{\ell}_{Tt}^{\mathbb{S}}(\Delta y_t)^2,$$

where $\{\hat{\ell}_{Tt}^{\mathbb{S}} : 2 \le t \le T\}$ are estimators of ℓ_f. Evidently, $(\hat{S}_T, \hat{H}_T)$ satisfies (15) provided the $\hat{\ell}_{Tt}^{\mathbb{S}}$ are such that $(\hat{S}_T, \hat{\mathcal{I}}_T) = (S_T^f, \mathcal{I}_{ff}) + o_{p_{0,f}}(1)$ for every $f \in \mathcal{F}_{AC}^{\mathbb{S}}$.

This requirement is met by sample splitting estimators of the form

$$(19)\qquad \hat{\ell}_{Tt}^{\mathbb{S}}(\Delta y_t) := \begin{cases} \tilde{\ell}_{T-\tau_T}^{\mathbb{S}}(\Delta y_t|\Delta y_{\tau_T+1},\ldots,\Delta y_T), & t = 1,\ldots,\tau_T, \\ \tilde{\ell}_{\tau_T}^{\mathbb{S}}(\Delta y_t|\Delta y_1,\ldots,\Delta y_{\tau_T}), & t = \tau_T+1,\ldots,T, \end{cases}$$

where τ_T are integers with

$$(20)\qquad 0 < \underline{\lim}_{T\to\infty} \tau_T/T \le \overline{\lim}_{T\to\infty} \tau_T/T < 1$$

[17]Moreover, the definition generalizes in an obvious way to (other classes of densities and) other models with a finite-dimensional asymptotically sufficient statistic and the resulting definition agrees with standard definitions in models where the likelihood ratios happen to be LAN.

and $\tilde{\ell}_T^{\mathbb{S}}$ is a sequence of estimators such that, as $T \to \infty$,

$$\int_{-\infty}^{\infty} \left[\tilde{\ell}_T^{\mathbb{S}}(\varepsilon|\varepsilon_1, \ldots, \varepsilon_T) - \ell_f(\varepsilon)\right]^2 f(\varepsilon)\, d\varepsilon = o_p(1) \tag{21}$$

and

$$\sqrt{T} \int_{-\infty}^{\infty} \tilde{\ell}_T^{\mathbb{S}}(\varepsilon|\varepsilon_1, \ldots, \varepsilon_T) f(\varepsilon)\, d\varepsilon = o_p(1), \tag{22}$$

whenever $\varepsilon_1, \ldots, \varepsilon_T$ are i.i.d. with density $f \in \mathcal{F}_{\mathrm{AC}}^{\mathbb{S}}$.

THEOREM 4: (a) *If* $(\hat{S}_T, \hat{H}_T)$ *is defined as in* (17)–(22), *then* (15) *holds.*
(b) *In particular, if* $f \in \mathcal{F}_{\mathrm{AC}}^{\mathbb{S}}$ *and* $\hat{\phi}_T$ *is defined as in* (16)–(22), *then*

$$\lim_{T\to\infty} E_{\rho_T(c)} \hat{\phi}_T(Y_T|\bar{c}, \alpha) = \lim_{T\to\infty} E_{\rho_T(c)} \phi_{f,T}(Y_T|\bar{c}, \alpha) \quad \forall c \le 0, \bar{c} < 0.$$

REMARK: (i) The main purpose of Theorem 4 is to demonstrate by example that the bound Ψ_f is sharp when the errors are known to be symmetric. Sample splitting estimators of ℓ_f (with the null hypothesis imposed) are employed because such estimators make it possible to give a relatively elementary proof of adaptation under fairly minimal conditions on f. In practice, it may be desirable to use the full sample (along with some estimator of ρ) when estimating ℓ_f. It seems plausible that the methods of Koul and Schick (1997) can be used to justify the use of such an estimator, but an investigation along these lines will not be pursued in this paper. Also left for future work is a numerical investigation of the extent to which the asymptotic power gains documented here are available in small samples when ℓ_f needs to be estimated. A similar remark applies to Theorem 6 of the next section.
(ii) Estimators $\tilde{\ell}_T^{\mathbb{S}}$ that satisfy (21) and (22) can be found in Bickel (1982) and Bickel et al. (1998). For further discussion of these high-level assumptions, see Schick (1986) and Klaassen (1987).

6. UNRESTRICTED ERROR DISTRIBUTIONS

This section obtains semiparametric power envelopes for tests of the unit root hypothesis in the case where f is (essentially) unrestricted in the sense that is assumed to be known only that f belongs to $\mathcal{F}_{\mathrm{DQM}}$, the class of densities satisfying Assumption DQM. In the spirit of Stein (1956), the semiparametric power envelopes will be defined as the infimum of the power envelopes associated with parametric submodels embedding the true error density. In light of the striking similarities between the results derived so far and the corresponding results for the location model, it seems plausible that these asymptotic power envelopes should admit an interpretation analogous to the interpretation of the semiparametric power envelope for tests in the location model.

That conjecture turns out to be correct. Specifically, it turns out that the least favorable submodels in the unit root model coincide with the least favorable submodels in the location model.

In the location model, a least favorable submodel is any submodel for which the associated ℓ_η maximizes the squared correlation $\mathcal{I}_{f\eta}^2\mathcal{I}_{ff}^{-1}\mathcal{I}_{\eta\eta}^{-1}$ of $\ell_f(\varepsilon)$ and $\ell_\eta(\varepsilon)$ subject to the restriction $\ell_\eta \in \mathcal{L}_\eta$. As was seen in Section 4, this property is shared by the unit root model in the Gaussian case, where $\mathcal{I}_{ff} = 1$ and any submodel has $\mathcal{I}_{f\eta} = 0$ (and is least favorable). Presuming that the property is shared by the unit root model also if $\mathcal{I}_{ff} > 1$, it follows that the semiparametric power envelopes for tests of the unit hypothesis should be given by the envelopes Ψ_F^S and Ψ_F^I associated with a submodel F for which $\ell_\eta(\varepsilon) = \ell_f(\varepsilon) - \varepsilon$.

Theorem 5 makes the preceding heuristics precise. Let (W, B_f) and $(\Lambda_f, \mathcal{S}_f, \mathcal{H}_{ff})$ be as in Section 3 and define

$$\Psi_f^S(c,\alpha) := E\big[\psi_f^S(\mathcal{S}_f, \mathcal{H}_{ff}, \mathcal{S}_f^S|c,\alpha)\exp(\Lambda_f(c))\big],$$

$$\Psi_f^I(c,\alpha) := E\big[\psi_f^I(\mathcal{S}_f^I, \mathcal{H}_{ff}^I|c,\alpha)\exp(\Lambda_f(c))\big],$$

where[18]

$$\psi_f^S(\mathcal{S}_f, \mathcal{H}_{ff}, \mathcal{S}_f^S|c,\alpha) := 1[\Lambda_f(c) > K_\alpha^S(\mathcal{S}_f^S, c; \mathcal{J}_f^{LF})],$$

$$\psi_f^I(\mathcal{S}_f^I, \mathcal{H}_{ff}^I|c,\alpha) := 1[\Lambda_f^I(c) > K_\alpha^I(c; \mathcal{J}_f^{LF})],$$

$$\Lambda_f^I(c) := c\mathcal{S}_f^I - \frac{1}{2}c^2\mathcal{H}_{ff}^I,$$

$$\mathcal{S}_f^S := B_f(1) - W(1), \quad \mathcal{J}_f^{LF} := \begin{pmatrix} \mathcal{I}_{ff} & \mathcal{I}_{ff}-1 \\ \mathcal{I}_{ff}-1 & \mathcal{I}_{ff}-1 \end{pmatrix},$$

$$\mathcal{S}_f^I := \mathcal{S}_f - \left(\int_0^1 W(r)\,dr\right)\mathcal{S}_f^S,$$

$$\mathcal{H}_{ff}^I := \mathcal{H}_{ff} - (\mathcal{I}_{ff}-1)\left(\int_0^1 W(r)\,dr\right)^2.$$

Finally, for any $f \in \mathcal{F}_{\mathrm{DQM}}$, let $\mathcal{J}_f$ denote the set of matrices $\mathcal{I}_F$ associated with submodels F satisfying Assumption DQM*.

THEOREM 5: *If $f \in \mathcal{F}_{\mathrm{DQM}}$, then*

$$\inf_{F:\mathcal{I}_F\in\mathcal{J}_f} \Psi_F^S(c,\alpha) = \Psi_f^S(c,\alpha) \quad \forall c < 0, \tag{23}$$

$$\inf_{F:\mathcal{I}_F\in\mathcal{J}_f} \Psi_F^I(c,\alpha) = \Psi_f^I(c,\alpha) \quad \forall c < 0. \tag{24}$$

[18]As defined, ψ_f^S and ψ_f^I are the test functions ψ_F^S and ψ_F^I of Section 4 that correspond to a submodel with $\ell_\eta(\varepsilon) = \ell_f(\varepsilon) - \varepsilon$.

The proof of (23) first shows that

$$\inf_{F:\mathcal{I}_F\in\mathcal{J}_f} \Psi_F^S(c,\alpha) \geq \Psi_f^S(c,\alpha) \quad \forall c<0. \tag{25}$$

The proof of (25) is constructive in the sense that it shows that the test based on

$$\phi_{f,T}^S(Y_T|c,\alpha) := 1\left[cS_T^f - \frac{1}{2}c^2 H_T^{ff} > K_\alpha^S(S_T^{f,S}, c; \mathcal{J}_f^{LF})\right]$$

attains Ψ_f^S and is locally asymptotically α-similar in any smooth submodel, where

$$S_T^{f,S} := T^{-1/2}\sum_{t=2}^{T}[\ell_f(\Delta y_t) - \Delta y_t].$$

Then, using the fact that Ψ_F^S is continuous in $\mathcal{I}_F$, an inequality in the opposite direction is obtained by showing that $\mathcal{J}_f^{LF}$ belongs to the closure of $\mathcal{J}_f$. A similar strategy is used to obtain (24).

Figures 2 and 3 plot $\Psi_f^S(\cdot, 0.05)$ and $\Psi_f^I(\cdot, 0.05)$ for various values of $\mathcal{I}_{ff}$. Comparing Figures 2 and 3 to Figure 1, the semiparametric power envelopes are seen to lie well above the power envelope corresponding to the Gaussian

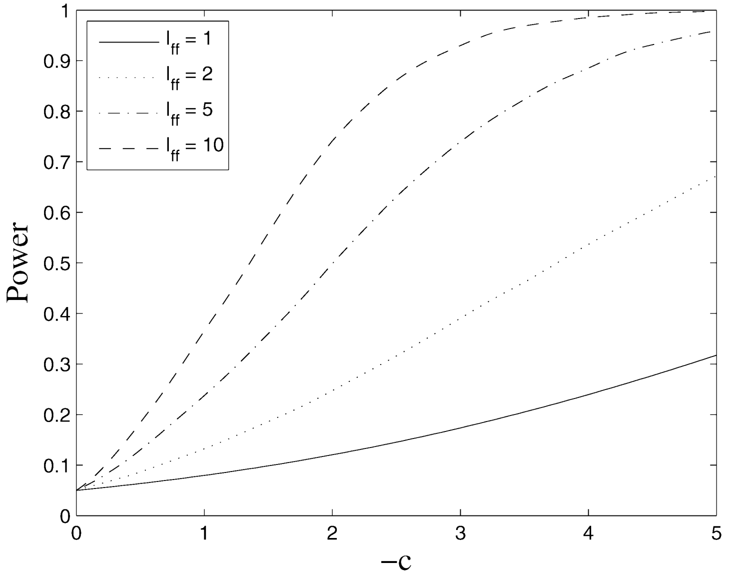

FIGURE 2.—Semiparametric power envelopes, similar tests.

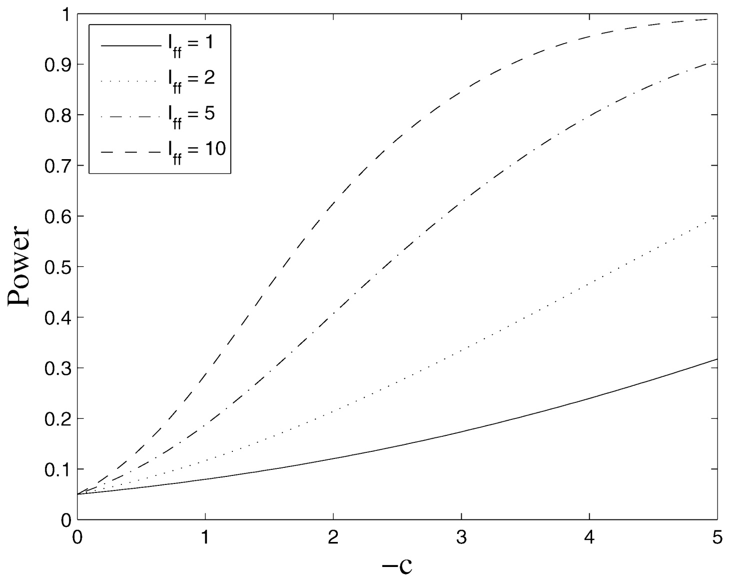

FIGURE 3.—Semiparametric power envelopes, invariant tests.

distribution.[19] In spite of the fact that there is no obvious connection between the analytical semiparametric efficiency results for the unit root model and those for the stable AR(1) model, the numerical results displayed in Figures 2 and 3 are therefore qualitatively similar to the well known results for the stable AR(1) model insofar as Figures 2 and 3 suggest that nonnormality can be a source of potentially substantial power gains in unit root tests even in the absence of knowledge of the error distribution.

It is therefore of significant interest to investigate whether the power bounds reported in Figures 2 and 3 are sharp. The fact that completely consistent (in the terminology of Andrews (1986)) goodness of fit tests exist can be used to show that for any $\bar{f} \in \mathcal{F}_{\mathrm{DQM}}$ it is possible to construct tests that are locally efficient at $\bar{f}$ in the sense that they are locally asymptotically α-similar/α-invariant in any smooth submodel F with $f(\cdot|0) \in \mathcal{F}_{\mathrm{DQM}}$ and attain Ψ_f^S/Ψ_f^I when $f = \bar{f}$. For instance, consider

$$\begin{aligned}\phi^*_{\bar{f},T}(Y_T|c,\alpha) \\ &:= \varphi_T(Y_T|\bar{f})\phi^S_{\bar{f},T}(Y_T|c,\alpha) + [1-\varphi_T(Y_T|\bar{f})]\phi_T^{\mathrm{ERS}}(Y_T|c,\alpha),\end{aligned}$$

[19]The difference between the oracle bounds $\Psi_f(\cdot, 0.05)$, $\Psi_f^S(\cdot, 0.05)$, and $\Psi_f^I(\cdot, 0.05)$ is noticeable in most of the cases considered. Numerical evaluation shows that $\sup_c |\Psi_f(c, 0.05) - \Psi_f^S(c, 0.05)| \approx 0.02$, 0.05, 0.07 and $\sup_c |\Psi_f(c, 0.05) - \Psi_f^I(c, 0.05)| \approx 0.09$, 0.15, 0.16 for $\mathcal{I}_{ff} = 2$, 5, 10.

where $\phi_T^{\mathrm{ERS}}(Y_T|c,\alpha)$ is the test function of ERS's point optimal test and $\varphi_T(\cdot|\bar{f})$ is a (goodness of fit) test function for which

$$\varphi_T(Y_T|\bar{f}) = 1(f=\bar{f}) + o_{p_{0,f}}(1) \quad \forall f \in \mathcal{F}_{\mathrm{DQM}}.$$

This "shrinkage" test is asymptotically equivalent to $\phi_{\bar{f},T}^S(Y_T|c,\alpha)$ when $f=\bar{f}$ and asymptotically equivalent to ERS's test otherwise. In particular, the test is locally asymptotically α-similar in any smooth submodel F with $f(\cdot|0) \in \mathcal{F}_{\mathrm{DQM}}$ and attains Ψ_f^S when $f=\bar{f}$. A similar construction can be used to show that Ψ_f^I provides a (pointwise) sharp upper bound on the local asymptotic power attainable by means of tests that are locally asymptotically α-invariant in any smooth submodel F with $f(\cdot|0) \in \mathcal{F}_{\mathrm{DQM}}$.

The preceding construction is of theoretical interest because it demonstrates by example that the bounds Ψ_f^S and Ψ_f^I are pointwise sharp. (In light of this it seems reasonable to refer to Ψ_f^S and Ψ_f^I as semiparametric power envelopes.) Nevertheless, the shrinkage test based on $\phi_{\bar{f},T}^{S,*}(Y_T|c,\alpha)$ is obviously not recommended for actual use and a more interesting question is therefore whether globally (in f) efficient testing procedures exist. On the one hand, reasoning similar to that of the remark at the end of Section 4 shows that plug-in versions of $\phi_{f,T}^S(Y_T|c,\alpha)$ generally fail to attain Ψ_f^S even if a valid parametric submodel is postulated. In contrast, Theorem 6 will show that the assumption $f \in \mathcal{F}_{\mathrm{AC}}$ is sufficient for the envelope Ψ_f^I to be globally attainable.

Global attainability of Ψ_f^I follows from arguments analogous to those used by Bickel (1982) to show feasibility of adaptive estimation of the slope coefficients in a standard regression model. The proof of (24) uses a finite sample counterpart of $\mathcal{S}_f^I$ given by

$$S_T^{f,I} := S_T^f - \left(\frac{1}{T^{3/2}}\sum_{t=2}^T y_{t-1}\right)\frac{1}{\sqrt{T}}\sum_{t=2}^T[\ell_f(\Delta y_t) - \Delta y_t]. \tag{26}$$

Because

$$S_T^{f,I} = \frac{1}{T}\sum_{t=2}^T y_{t-1}^\mu \ell_f(\Delta y_t) + \left(\frac{1}{T^{3/2}}\sum_{t=2}^T y_{t-1}\right)\frac{1}{\sqrt{T}}(y_T - y_1),$$

$$y_{t-1}^\mu := y_{t-1} - \frac{\sum_{s=2}^T y_{s-1}}{T-1},$$

consistent estimation of $S_T^{f,I}$ turns out to be feasible even though ℓ_f cannot be estimated with small bias. Specifically, the fact that $\sum_{t=2}^T y_{t-1}^\mu = 0$ implies that (the natural counterpart of) the assumption (22) can be avoided when constructing a consistent estimator of $S_T^{f,I}$.[20]

To demonstrate by example that the bound Ψ_f^I is globally (in f) sharp, let

$$\hat{\phi}_T^I(Y_T|c,\alpha) := 1\left[c\hat{S}_T^I - \frac{1}{2}c^2\hat{H}_T^I > K_\alpha^I(c;\hat{\mathcal{J}}_T^{LF})\right], \tag{27}$$

where, for some integers τ_T with

$$\lim_{T\to\infty}\tau_T = \infty \quad \text{and} \quad \lim_{T\to\infty}\tau_T/T = 0, \tag{28}$$

and for some estimator $\hat{\ell}_T$ of ℓ_f,

$$\begin{aligned}\hat{S}_T^I &:= \frac{1}{T}\sum_{t=\tau_T+1}^{T} y_{t-1}\hat{\ell}_T(\Delta y_t) \\ &\quad - \left(\frac{1}{T^{3/2}}\sum_{t=\tau_T+1}^{T} y_{t-1}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=\tau_T+1}^{T}[\hat{\ell}_T(\Delta y_t) - \Delta y_t]\right),\end{aligned} \tag{29}$$

$$\hat{H}_T^I := \frac{\hat{\mathcal{I}}_T}{T^2}\sum_{t=\tau_T+1}^{T} y_{t-1}^2 - (\hat{\mathcal{I}}_T - 1)\left(\frac{1}{T^{3/2}}\sum_{t=\tau_T+1}^{T} y_{t-1}\right)^2, \tag{30}$$

$$\hat{\mathcal{J}}_T^{LF} := \begin{pmatrix}\hat{\mathcal{I}}_T & \hat{\mathcal{I}}_T - 1 \\ \hat{\mathcal{I}}_T - 1 & \hat{\mathcal{I}}_T - 1\end{pmatrix}, \quad \hat{\mathcal{I}}_T := \frac{1}{T}\sum_{t=\tau_T+1}^{T}\hat{\ell}_T(\Delta y_t)^2. \tag{31}$$

As defined, $\hat{\phi}_T^I(Y_T|c,\alpha)$ is a plug-in version of the test $\phi_{f,T}^I(Y_T|c,\alpha)$ used in the proof of Theorem 5(b). In the spirit of Bickel (1982), suppose

$$\hat{\ell}_T(\Delta y_t) := \tilde{\ell}_{\tau_T}(\Delta y_t|\Delta y_1,\ldots,\Delta y_{\tau_T}), \tag{32}$$

where $\tilde{\ell}_T$ is a sequence of estimators such that, as $T\to\infty$,

$$\int_{-\infty}^{\infty}[\tilde{\ell}_T(\varepsilon|\varepsilon_1,\ldots,\varepsilon_T) - \ell_f(\varepsilon)]^2 f(\varepsilon)\,d\varepsilon = o_p(1) \tag{33}$$

whenever $\varepsilon_1,\ldots,\varepsilon_T$ are i.i.d. with density $f\in\mathcal{F}_{\mathrm{AC}}$.

[20]Because adaptive estimation is impossible in the location model with an (essentially) unrestricted f, it follows from Klaassen (1987) that there exists no $\sqrt{T}$-unbiased estimator of ℓ_f when f is (essentially) unrestricted; that is, the natural counterpart of (22) cannot hold when f is (essentially) unrestricted. In contrast, it follows from Bickel (1982) that the natural counterpart of (21) is compatible with the assumption $f\in\mathcal{F}_{\mathrm{AC}}$, so (33) is not void.

THEOREM 6: *If* $f \in \mathcal{F}_{\mathrm{AC}}$ *and* $\hat{\phi}_T^I$ *is defined as in* (27)–(33), *then*

$$\lim_{T\to\infty} E_{\rho_T(c)} \hat{\phi}_T^I(Y_T|\bar{c},\alpha) = E\big[\psi_f^I(\mathcal{S}_f^I, \mathcal{H}_{ff}^I|\bar{c},\alpha)\exp(\Lambda_f(c))\big]$$
$$\forall c \leq 0, \bar{c} < 0.$$

By showing that Ψ_f^I is sharp, Theorem 6 demonstrates in particular that there is a sense in which the tests of ERS are asymptotically inadmissible if the assumption of Gaussian errors is relaxed. This result and the analogous inadmissibility result deducible from Theorem 4 can be viewed as unit root counterparts to the inadmissibility of the least squares estimator of β in the model

$$Y_i = \beta X_i + \varepsilon_i,$$

where the ε_i are i.i.d. with density f and independent of the i.i.d. regressor X_i whose mean is assumed to be different from zero. Specifically, Bickel (1982, Example 2) showed that adaptive estimation of β is possible when f is symmetric, while Schick (1987, Example 2) presented the efficient influence function for β without assuming symmetry of f and showed in particular that departures from normality can be exploited for efficiency purposes also in that case.

7. EXTENSIONS

Sections 3–6 study a model which assumes away the presence of deterministic components and/or serial correlation in the error. In the Gaussian case, the consequences of relaxing these assumptions are well understood from the work of ERS: parameters governing serial correlation in the error can be treated "as if" they are known, as can the value of a constant mean in the observed process, whereas the presence of a time trend affects the asymptotic power envelope. This section briefly explores whether these qualitative conclusions remain valid in models with non-Gaussian errors and finds that they do. In addition, and in perfect analogy with Section 4, it is found that also in models with a time trend, the properties of parametric submodels depend on whether or not Stein's (1956) necessary condition for adaptation in the location model is satisfied.

The consequences of accommodating deterministic components and/or serial correlation in the error will be explored by studying a model in which the observed data $y_1, \ldots, y_T$ are generated as

$$y_t = \mu + \delta t + u_t, \quad (1 - \rho L)\gamma(L)u_t = \varepsilon_t, \tag{34}$$

where μ and δ are the parameters governing the deterministic component, the lag polynomial $\gamma(L) = 1 - \gamma_1 L - \cdots - \gamma_p L^p$ is of (known, finite) order p,[21] the initial conditions are $u_0 = u_{-1} = \cdots = u_{1-p} = 0$, and the ε_t are unobserved i.i.d. errors from a continuous distribution with full support, zero mean, unit variance, and density f. It is assumed that $\min_{|z|\leq 1} |\gamma(z)| > 0$ so that the unit root testing problem is that of testing $H_0 : \rho = 1$ vs. $H_1 : \rho < 1$. As in Section 4, the density f is embedded in a smooth family of densities.

As before, local reparameterizations will be employed in the asymptotic analysis. The appropriate reparameterizations of μ, δ, and $\gamma(L)$ are of the form

$$\mu = \mu_T(m) := \mu_0 + m, \qquad \delta = \delta_T(d) := \delta_0 + \frac{\gamma_0(1)}{\sqrt{T}} d,$$

$$\gamma(L) = \gamma_T(L; g) := \gamma_0(L) + \frac{g(L)}{\sqrt{T}},$$

where μ_0 and δ_0 are known constants, $\gamma_0(L) := 1 - \gamma_{1,0} L - \cdots - \gamma_{p,0} L^p$ is a known lag polynomial with $\min_{|z|\leq 1} |\gamma_0(z)| > 0$, whereas the unknown parameters are m, d, and the coefficients $g := (g_1, \ldots, g_p)'$ of the lag polynomial $g(L) := -g_1 L - \cdots - g_p L^p$.[22] Without loss of generality, it is assumed that μ_0 and δ_0 are equal to zero.

The log likelihood ratio function associated with the chosen reparameterization is of the form

$$\begin{aligned} &L_T^{\mathbf{F}}(c, m, d, g, h) \\ &\quad := L_T^0(c, m, d, g, h) + \sum_{t=p+2}^{T} \log f[\varepsilon_t(c, m, d, g) | \eta_T(h)] \\ &\qquad - \sum_{t=p+2}^{T} \log f[\varepsilon_t(0, 0, 0, 0) | \eta_T(0)], \end{aligned}$$

where $L_T^0(c, h, m, d, g)$ represents the contribution of $y_1, \ldots, y_{p+1}$ and

$$\varepsilon_t(c, m, d, g) := [1 - \rho_T(c) L] \gamma_T(L; g) \left[y_t - m - \frac{d}{\sqrt{T}} t \right] \qquad (t \geq p + 2).$$

[21]Adapting the methods of Jeganathan (1997), it should be possible to allow $\gamma(L)$ to be a smoothly parameterized lag polynomial of infinite order. The qualitative conclusions of this section will not be affected by such an extension, so to conserve space it will not be pursued here.

[22]The term $\gamma_0(1)$ appears in the definition of $\delta_T(d)$ because the resulting definition gives rise to a limiting experiment which depends on d in a particularly simple way.

If y_t is generated by (34), $(c_T, h_T, m_T, d_T, g_T)$ is a bounded sequence, and mild smoothness conditions on F hold, then $L_T^{\mathbf{F}}$ admits an expansion of the form

$$L_T^{\mathbf{F}}(c_T, m_T, d_T, g_T, h_T) = L_T^{f\delta\eta}(c_T, d_T, h_T) + L^{\mu}(m_T) + L_T^{\gamma}(g_T) + o_{p_{0,f}}(1), \tag{35}$$

where $o_{p_{0,f}}(1)$ is shorthand for "$o_p(1)$ when H_0 holds, $(m, d, g) = (0, 0, 0)$, and ε has density f" and the functions $L_T^{f\delta\eta}$, L^{μ}, and L_T^{γ} are given by

$$\begin{aligned} L_T^{f\delta\eta}(c, d, h) &:= cS_T^f - \frac{1}{2}c^2 H_T^{ff} + d[S_T^{\delta}(c) - cH_T^{f\delta}(c)] - \frac{1}{2}d^2\mathcal{H}_{\delta\delta}(c) \\ &\qquad + h[S_T^{\eta} - cH_T^{f\eta} - d\mathcal{H}_{\delta\eta}(c)] - \frac{1}{2}h^2\mathcal{H}_{\eta\eta}, \\ L^{\mu}(m) &:= \sum_{j=0}^{p} \log f(\Delta x_{1+j} + \gamma_{j,0}m) - \sum_{j=0}^{p} \log f(\Delta x_{1+j}) \qquad (\gamma_{0,0} := -1), \\ L_T^{\gamma}(g) &:= g'S_T^{\gamma} - \frac{1}{2}g'\mathcal{H}_{\gamma\gamma}g, \end{aligned}$$

where

$$\begin{aligned} S_T^f &:= \frac{1}{T}\sum_{t=p+2}^{T} x_{t-1}\ell_f(\Delta x_t), \quad H_T^{ff} := \frac{\mathcal{I}_{ff}}{T^2}\sum_{t=p+2}^{T} x_{t-1}^2, \\ S_T^{\delta}(c) &:= \frac{1}{\sqrt{T}}\sum_{t=p+2}^{T} \xi_c\left(\frac{t-1}{T}\right)\ell_f(\Delta x_t), \\ H_T^{f\delta}(c) &:= \frac{\mathcal{I}_{ff}}{T^{3/2}}\sum_{t=p+2}^{T} x_{t-1}\xi_c\left(\frac{t-1}{T}\right), \\ S_T^{\eta} &:= \frac{1}{\sqrt{T}}\sum_{t=p+2}^{T} \ell_{\eta}(\Delta x_t), \quad H_T^{f\eta} := \frac{\mathcal{I}_{f\eta}}{T^{3/2}}\sum_{t=p+2}^{T} x_{t-1}, \\ S_T^{\gamma} &:= \frac{1}{\sqrt{T}}\sum_{t=p+2}^{T} (\Delta y_{t-1}, \ldots, \Delta y_{t-p})'\ell_f(\Delta x_t), \end{aligned}$$

(ℓ_f, ℓ_η) and $(\mathcal{I}_{ff}, \mathcal{I}_{f\eta}, \mathcal{H}_{\eta\eta})$ are as in Section 4, $x_t := \gamma_0(L)y_t$,[23] $\xi_c(r) := 1 - cr$, $\mathcal{H}_{\delta\delta}(c) := \mathcal{I}_{ff}(1 - c + c^2/3)$, $\mathcal{H}_{\delta\eta}(c) := \mathcal{I}_{f\eta}(1 - c/2)$, $\mathcal{H}_{\gamma\gamma} := \Sigma_{\gamma\gamma}\mathcal{I}_{ff}$, and $\Sigma_{\gamma\gamma}$ is a $p \times p$ matrix with element (i, j) given by $E[\gamma_0(L)^{-1}\varepsilon_{t-i}\gamma_0(L)^{-1}\varepsilon_{t-j}]$.

Because neither $L^\mu(\cdot)$ nor $L_T^{f\delta\eta}(\cdot)$ is quadratic, the model with a deterministic component does not admit LAQ likelihood ratios. Nevertheless, the model is well suited for analysis using the limits of experiments approach, as the interesting part of the limiting experiment belongs to a curved exponential family and is amenable to analysis using existing tools. Indeed, for every (c, h, m, d, g),

$$[L_T^{f\delta\eta}(c, d, h), L^\mu(m), L_T^\gamma(g)] \to_{d_{0,f}} [\Lambda_{f\delta\eta}(c, d, h), \Lambda_\mu(m), \Lambda_\gamma(g)],$$

where $\to_{d_{0,f}}$ is shorthand for "$\to_d$ when H_0 holds, $(m, d, g) = (0, 0, 0)$, and ε has density f" and $\Lambda_{f\delta\eta}(c, d, h)$, $\Lambda_\mu(m)$, and $\Lambda_\gamma(g)$ are mutually independent with

$$\Lambda_\gamma(g) := g'\mathcal{S}_\gamma - \frac{1}{2}g'\mathcal{H}_{\gamma\gamma}g, \quad \mathcal{S}_\gamma \sim \mathcal{N}(0, \mathcal{H}_{\gamma\gamma}),$$

$$\Lambda_\mu(m) \sim \sum_{j=0}^{p} \log f(\varepsilon_{1+j} + \gamma_{j,0}m) - \sum_{j=0}^{p} \log f(\varepsilon_{1+j}),$$

and

$$\begin{aligned}\Lambda_{f\delta\eta}(c, d, h) := {}& c\mathcal{S}_f - \frac{1}{2}c^2\mathcal{H}_{ff} + d[\mathcal{S}_\delta(c) - c\mathcal{H}_{f\delta}(c)] - \frac{1}{2}d^2\mathcal{H}_{\delta\delta}(c) \\ & + h[\mathcal{S}_\eta - c\mathcal{H}_{f\eta} - d\mathcal{H}_{\delta\eta}(c)] - \frac{1}{2}h^2\mathcal{H}_{\eta\eta},\end{aligned}$$

where $(\mathcal{S}_f, \mathcal{H}_{ff}, \mathcal{S}_\eta, \mathcal{H}_{f\eta})$ and (W, B_f, B_η) are as in Section 4 and

$$\mathcal{S}_\delta(c) := \int_0^1 \xi_c(r)\, dB_f(r), \quad \mathcal{H}_{f\delta}(c) := \mathcal{I}_{ff}\int_0^1 W(r)\xi_c(r)\, dr.$$

The mutual independence of $\Lambda_{f\delta\eta}(c, d, h)$ and $[\Lambda_\mu(m), \Lambda_\gamma(g)]$ and the additively separable structure of the right-hand side of (35) imply that the derivation of asymptotic power envelopes for tests of the unit root hypothesis can proceed under the "as if" assumption that μ and the coefficients of $\gamma(L)$ are known. Moreover, the distribution of $\Lambda_{f\delta\eta}(c, d, h)$ does not depend on the coefficients of $\gamma_0(L)$, so the power bounds developed in the previous sections

[23]The (presample) values of $y_0, \ldots, y_{1-p}$ are set equal to zero in the definition of $x_1, \ldots, x_p$. Because $x_t = y_t$ when $p = 0$, the present definitions of S_T^f, H_T^{ff}, S_T^η, and $H_T^{f\eta}$ are consistent with those of the previous sections.

(under the assumption that d is known to equal zero) are valid also in the presence of a constant mean and/or serial correlation in the error. Furthermore, the presence of a constant mean and/or serial correlation in the error does not weaken the sense in which the bounds are sharp because

$$\frac{1}{T}\sum_{t=p+2}^{T}\hat{x}_{t-1}\ell_f(\Delta\hat{x}_t)=S_T^f+o_{p_{0,f}}(1),\qquad \frac{\mathcal{I}_{ff}}{T^2}\sum_{t=p+2}^{T}\hat{x}_{t-1}^2=H_T^{ff}+o_{p_{0,f}}(1),$$

and so on, where $\hat{x}_t := \hat{\gamma}(L)(y_t-\hat{\mu})$, $\hat{\mu} := y_1$, and $\hat{\gamma}(L)$ is a discretized, $\sqrt{T}$-consistent estimator of $\gamma(L)$. These qualitative conclusions, which are in perfect agreement with those obtained by ERS in the Gaussian case, show in particular that the inability to do adaptive unit root testing when f is (essentially) unrestricted is not an artifact of the assumption that the deterministic component is known. Nevertheless, it is of some interest to investigate whether the condition $\mathcal{I}_{f\eta}=0$ continues to play an important role also in models with a time trend.

In the case where a time trend is accommodated, the relevant limiting experiment is an extended version of that studied in Section 4. The extended limiting experiment involves the three-dimensional parameter (c,d,h) and is characterized by log likelihood ratios of the form $\Lambda_{f\delta\eta}(c,d,h)$. As in Section 4, a location invariance restriction can be used to remove the nuisance parameter h, the associated log likelihood ratio being given by

$$\begin{aligned}\Lambda_{f\delta.\eta}(c,d) &:= \max_h \Lambda_{f\delta\eta}(c,d,h)-\max_h\Lambda_{f\delta\eta}(0,0,h)\\ &= c\mathcal{S}_{f.\eta}-\frac{1}{2}c^2\mathcal{H}_{ff.\eta}+d[\mathcal{S}_{\delta.\eta}(c)-c\mathcal{H}_{f\delta.\eta}(c)]\\ &\quad -\frac{1}{2}d^2\mathcal{H}_{\delta\delta.\eta}(c),\end{aligned}$$

where $(\mathcal{S}_{f.\eta},\mathcal{H}_{ff.\eta})$ is as in Section 4 and

$$\mathcal{S}_{\delta.\eta}(c):=\mathcal{S}_\delta(c)-\frac{\mathcal{H}_{\delta\eta}(c)}{\mathcal{H}_{\eta\eta}}\mathcal{S}_\eta,\qquad \mathcal{H}_{f\delta.\eta}(c):=\mathcal{H}_{f\delta}(c)-\frac{\mathcal{H}_{f\eta}}{\mathcal{H}_{\eta\eta}}\mathcal{H}_{\delta\eta}(c),$$

$$\mathcal{H}_{\delta\delta.\eta}(c):=\mathcal{H}_{\delta\delta}(c)-\frac{\mathcal{H}_{\delta\eta}(c)^2}{\mathcal{H}_{\eta\eta}}.$$

Similarly, the remaining nuisance parameter d can be removed using the principle of invariance.[24] Indeed, in perfect analogy with ERS's analysis of the

[24]The invariance condition in question is an asymptotic counterpart of the restriction that inference should be invariant under transformations of the form

$$y_t \to y_t + b_\delta t,\qquad b_\delta\in\mathbb{R}.$$

Gaussian case, Lehmann and Romano (2005, Problem 6.9) and the fact that $\Lambda_{f\delta.\eta}(c,d)$ is quadratic in d for any fixed c can be used to show that the power envelope associated with α-invariant tests in the extended limiting experiment is given by

$$\Psi_F^\delta(c,\alpha) := E\big[1(\Lambda_{f.\delta\eta}(c) > K_\alpha^\delta(c;\mathcal{I}_F))\exp(\Lambda_f(c))\big],$$

where $K_\alpha^\delta(c;\mathcal{I}_F)$ is the $1-\alpha$ quantile of

$$\begin{aligned}\Lambda_{f.\delta\eta}(c) &:= \max_d \Lambda_{f\delta.\eta}(c,d) - \max_d \Lambda_{f\delta.\eta}(0,d)\\ &= c\mathcal{S}_{f.\eta} - \frac{1}{2}c^2\mathcal{H}_{ff.\eta} + \frac{1}{2}\frac{[\mathcal{S}_{\delta.\eta}(c) - c\mathcal{H}_{f\delta.\eta}(c)]^2}{\mathcal{H}_{\delta\delta.\eta}(c)} - \frac{1}{2}\frac{\mathcal{S}_{\delta.\eta}(0)^2}{\mathcal{H}_{\delta\delta.\eta}(0)}.\end{aligned}$$

Analogous reasoning shows that if h is assumed to be known to equal zero, then the power envelope associated with α-invariant tests in the relevant limiting experiment is given by

$$\bar{\Psi}_F^\delta(c,\alpha) := E\big[1(\bar{\Lambda}_{f.\delta}(c) > \bar{K}_\alpha^\delta(c;\mathcal{I}_{ff}))\exp(\Lambda_f(c))\big],$$

where $\bar{K}_\alpha^\delta(c;\mathcal{I}_{ff})$ is the $1-\alpha$ quantile of

$$\begin{aligned}\bar{\Lambda}_{f.\delta}(c) &:= \max_d \Lambda_{f\delta\eta}(c,d,0) - \max_d \Lambda_{f\delta\eta}(0,d,0)\\ &= c\mathcal{S}_f - \frac{1}{2}c^2\mathcal{H}_{ff} + \frac{1}{2}\frac{[\mathcal{S}_\delta(c) - c\mathcal{H}_{f\delta}(c)]^2}{\mathcal{H}_{\delta\delta}(c)} - \frac{1}{2}\frac{\mathcal{S}_\delta(0)^2}{\mathcal{H}_{\delta\delta}(0)}.\end{aligned}$$

By inspection, it is seen that $\bar{\Lambda}_{f.\delta}(\cdot)$ and $\Lambda_{f.\delta\eta}(\cdot)$ coincide if and only if $\mathcal{I}_{f\eta} = 0$. In other words, Stein's (1956) necessary condition for adaptation in the location model remains a necessary condition for adaptive unit root testing even when a time trend is included in the deterministic component.

REMARK: Proceeding as in Section 6, it should be possible to give an explicit characterization of the semiparametric power envelope $\Psi_f^\delta(c,\alpha) := \inf_{F:\mathcal{I}_F\in\mathcal{J}_f}\Psi_F^\delta(c,\alpha)$ obtained by minimizing $\Psi_F^\delta(c,\alpha)$ with respect to the submodel F and to demonstrate by example that the envelope is sharp. To conserve space, the details of these extensions are left for future work.

8. CONCLUSION

This paper has derived asymptotic power envelopes for tests of the unit root hypothesis in a zero-mean AR(1) model. The power envelopes have been derived using the limits of experiments approach and are semiparametric in the

(This transformation induces a transformation on the parameter δ of the form $\delta \to \delta + b_\delta$, but leaves all other parameters unchanged.)

sense that the underlying error distribution is treated as an unknown infinite-dimensional nuisance parameter. Adaptation has been shown to be possible when the error distribution is known to be symmetric and to be impossible when the error distribution is (essentially) unrestricted. In the latter case, two conceptually distinct approaches to nuisance parameter elimination were employed in the derivation of the semiparametric power envelopes. One of these power bounds, derived under an invariance restriction, was shown by example to be sharp, while the other, derived under a similarity restriction, was conjectured not to be globally attainable.

Both sets of restrictions imposed when deriving the semiparametric power envelopes have natural counterparts in models with LAN likelihood ratios and give rise to identical power envelopes in such models. The fact that the two sets of restrictions give rise to distinct power envelopes in the present context is perhaps surprising and clearly shows that not all methodological conclusions from the existing literature on semiparametrics will generalize to models not admitting LAN likelihood ratios. On the other hand, it is interesting that one approach to nuisance parameter elimination (albeit one that has not received much attention in the existing literature) "works" both in conventional models and in the model studied herein. It would be of interest to investigate whether this approach to nuisance parameter elimination also "works" in other non-standard hypothesis testing problems involving infinite-dimensional nuisance parameters.

APPENDIX: PROOFS

PROOF OF THEOREM 3: Let $c < 0$ be given.

(a) Because $\Lambda_F(0,h) = h\mathcal{S}_\eta - \frac{1}{2}h^2\mathcal{I}_{\eta\eta}$, it follows from the completeness properties of linear exponential families (e.g., Lehmann and Romano (2005, Theorem 4.3.1)) that ψ satisfies (9) if and only if $E[\psi(\mathcal{S}_F, \mathcal{H}_F)|\mathcal{S}_\eta] = \alpha$. Using this characterization of (9) and the properties of curved exponential families (e.g., Lehmann and Romano (2005, Lemma 2.7.2)), the Neyman–Pearson lemma can be used to show that if ψ satisfies (9), then $E[\psi(\mathcal{S}_F, \mathcal{H}_F) \times \exp(\Lambda_f(c))] \le \Psi_F^S(c, \alpha)$.

(b) By the Neyman–Pearson lemma, the right-hand side in (11) is no greater than $\Psi_F^I(c, \alpha)$ if (8) holds. To complete the proof, it therefore suffices to show that (11) holds whenever (10) does. Now, $\mathcal{S}_\eta \sim \mathcal{N}(0, \mathcal{I}_{\eta\eta})$ is independent of $(\mathcal{S}_{f.\eta}, \mathcal{H}_F)$. Furthermore,

$$\Lambda_F(c,h) = \Lambda_{f.\eta}(c) + \left(c\frac{\mathcal{H}_{f\eta}}{\mathcal{H}_{\eta\eta}} + h\right)\mathcal{S}_\eta - \frac{1}{2}\left(c\frac{\mathcal{H}_{f\eta}}{\mathcal{H}_{\eta\eta}} + h\right)^2 \mathcal{I}_{\eta\eta} \quad \forall (c,h). \tag{36}$$

These facts imply that $E[\exp(\Lambda_F(c,h))|\mathcal{S}_{f.\eta}, \mathcal{H}_F] = \exp(\Lambda_{f.\eta}(c))$ for any (c,h), from which the desired conclusion follows. *Q.E.D.*

REMARK: Using (36) and the fact that $\mathcal{S}_\eta \sim \mathcal{N}(0, \mathcal{I}_{\eta\eta})$ is independent of $(\mathcal{S}_{f.\eta}, \mathcal{H}_F)$, it can be shown that, for any c, h, b_η and for any bounded, measurable function $\varkappa$,

$$\begin{aligned} &E[\varkappa(\mathcal{S}_{f.\eta}, \mathcal{S}_\eta + b_\eta, \mathcal{H}_F)\Lambda_F(c,h)] \\ &\quad = E\left[\varkappa(\mathcal{S}_{f.\eta}, \mathcal{S}_\eta, \mathcal{H}_F)\Lambda_F\left(c, h + \frac{b_\eta}{\mathcal{I}_{\eta\eta}}\right)\right]. \end{aligned}$$

Let $(S_\infty^{f.\eta}, S_\infty^\eta, H_\infty^F)$ denote the weak limit of $(S_T^{f.\eta}, S_T^\eta, H_T^F)$ (under a sequence of parameterizations of the form $(\rho, \eta) = (\rho_T(c), \eta_T(h))$ for some fixed (c, h)).[25] In view of the preceding display, any transformation of the form

$$(S_\infty^{f.\eta}, S_\infty^\eta, H_\infty^F) \to (S_\infty^{f.\eta}, S_\infty^\eta + b_\eta, H_\infty^F), \quad b_\eta \in \mathbb{R}, \tag{37}$$

induces a transformation of the parameter (c, h) given by $(c, h) \to (c, h + b_\eta/\mathcal{I}_{\eta\eta})$. Because h is a nuisance parameter, the testing problem under consideration is invariant with respect to (location) transformations of the form (37). The associated maximal invariant is $(S_\infty^{f.\eta}, H_\infty^F)$. Condition (10) on the test function ψ asserts that the test depends on $(S_\infty^{f.\eta}, S_\infty^\eta, H_\infty^F)$ only through this maximal invariant.

The following simple lemma, a proof of which can be found in Jansson (2007), is used in the proofs of Theorems 4 and 5.

LEMMA 7: *There exists a* (*unique*) *continuous function* K_α^S *such that* ψ_F^S *satisfies* $E[\psi_F^S(\mathcal{S}_F, \mathcal{H}_F | c, \alpha) | \mathcal{S}_\eta] = \alpha$.

PROOF OF THEOREM 4: If (a) holds, then (b) holds because it follows from (a) and the continuity theorem (for convergence in probability) that if $f \in \mathcal{F}_{\mathrm{AC}}^{\mathbb{S}}$, then

$$\hat{\phi}_T(Y_T | \bar{c}, \alpha) = \phi_{f,T}(Y_T | \bar{c}, \alpha) + o_{p_{0,f}}(1).$$

(The continuity theorem is applicable because it follows from Lemma 7 that K_α is continuous.)

To prove (a) it suffices to show that

$$(\hat{S}_T, \hat{\mathcal{I}}_T) = (S_T^f, \mathcal{I}_{ff}) + o_{p_{0,f}}(1) \quad \forall f \in \mathcal{F}_{\mathrm{AC}}^{\mathbb{S}}.$$

Throughout the proof, suppose H_0 holds and let $f \in \mathcal{F}_{\mathrm{AC}}^{\mathbb{S}}$ be given.

The result $\hat{\mathcal{I}}_T = \mathcal{I}_{ff} + o_p(1)$ is (essentially) a special case of Drost, Klaassen, and Werker (1997, Lemma 3.1) and can be proved in exactly the same way.

[25]When $(c, h) = (0, 0)$, $(S_\infty^{f.\eta}, S_\infty^\eta, H_\infty^F) \sim (\mathcal{S}_{f.\eta}, \mathcal{S}_\eta, \mathcal{H}_F)$.

The result $\hat{S}_T = S_T^f + o_p(1)$ will be established by showing that

$$(\hat{S}_{T,\tau}, \hat{S}_T - \hat{S}_{T,\tau}) = (S_{T,\tau}^f, S_T^f - S_{T,\tau}^f) + o_p(1),$$

where $\hat{S}_{T,\tau} := T^{-1}\sum_{t=2}^{\tau_T} y_{t-1}\hat{\ell}_{Tt}^{\mathbb{S}}(\Delta y_t)$ and $S_{T,\tau}^f := T^{-1}\sum_{t=2}^{\tau_T} y_{t-1}\ell_f(\Delta y_t)$. Let $E_{t-1}^{\tau}[\cdot]$ denote conditional expectation given $\{\varepsilon_1, \ldots, \varepsilon_{t-1}\}$ and $\{\varepsilon_{\tau_T+1}, \ldots, \varepsilon_T\}$. By construction, $\sqrt{T}E_{t-1}^{\tau}[\hat{\ell}_{Tt}^{\mathbb{S}}(\Delta y_t)]$ is the same for every $t \le \tau_T$, namely

$$\sqrt{T}E_{t-1}^{\tau}[\hat{\ell}_{Tt}^{\mathbb{S}}(\Delta y_t)] = \sqrt{T}\int_{-\infty}^{\infty} \tilde{\ell}_{T-\tau_T}^{\mathbb{S}}(\varepsilon|\varepsilon_{\tau_T+1}, \ldots, \varepsilon_T) f(\varepsilon)\, d\varepsilon = o_p(1),$$

where the last equality uses (22). Furthermore, $E_{t-1}^{\tau}[\ell_f(\Delta y_t)] = E[\ell_f(\varepsilon)] = 0$ and $\sum_{t=2}^{\tau_T} y_{t-1} = O_p(T^{3/2})$, so

$$\begin{aligned}
&\frac{1}{T}\sum_{t=2}^{\tau_T} E_{t-1}^{\tau}\big[y_{t-1}(\hat{\ell}_{Tt}^{\mathbb{S}}(\Delta y_t) - \ell_f(\Delta y_t))\big] \\
&\quad = \left(\frac{1}{T^{3/2}}\sum_{t=2}^{\tau_T} y_{t-1}\right)\left(\sqrt{T}\int_{-\infty}^{\infty} \tilde{\ell}_{T-\tau_T}^{\mathbb{S}}(\varepsilon|\varepsilon_{\tau_T+1}, \ldots, \varepsilon_T) f(\varepsilon)\, d\varepsilon\right) \\
&\quad = o_p(1).
\end{aligned}$$

It now follows from Drost, Klaassen, and Werker (1997, Lemma 2.2) that $\hat{S}_{T,\tau} = S_{T,\tau}^f + o_p(1)$ because

$$\begin{aligned}
\frac{1}{T^2}\sum_{t=2}^{\tau_T} E_{t-1}^{\tau}\big[y_{t-1}^2(\hat{\ell}_{Tt}^{\mathbb{S}}(\Delta y_t) - \ell_f(\Delta y_t))^2\big] &= \left(\frac{1}{T^2}\sum_{t=2}^{\tau_T} y_{t-1}^2\right) o_p(1) \\
&= o_p(1),
\end{aligned}$$

where the first equality uses the fact that, for every $t \le \tau_T$,

$$\begin{aligned}
&E_{t-1}^{\tau}\big[(\hat{\ell}_{Tt}^{\mathbb{S}}(\Delta y_t) - \ell_f(\Delta y_t))^2\big] \\
&\quad = \int_{-\infty}^{\infty} [\tilde{\ell}_{T-\tau_T}^{\mathbb{S}}(\varepsilon|\varepsilon_{\tau_T+1}, \ldots, \varepsilon_T) - \ell_f(\varepsilon)]^2 f(\varepsilon)\, d\varepsilon \\
&\quad = o_p(1),
\end{aligned}$$

the last equality being a consequence of (21).

Analogous reasoning can be used to show that $\hat{S}_T - \hat{S}_{T,\tau} = S_T^f - S_{T,\tau}^f + o_p(1)$. *Q.E.D.*

PROOF OF THEOREM 5: Let $f \in \mathcal{F}_{\text{DQM}}$ be given.

Equation (23). The inequality (25) follows from the fact that if F satisfies Assumption DQM*, then $\phi^S_{f,T}(\cdot|c,\alpha)$ is locally asymptotically α-similar in F and satisfies $\lim_{T\to\infty} E_{\rho_T(c),\eta_T(0)}\phi^S_T(Y_T|c,\alpha;f) = \Psi^S_f(c,\alpha)$ (for every $c<0$); see Jansson (2007) for details.

Next, because $f \in \mathcal{F}_{\mathrm{DQM}}$, it can be embedded in a family F satisfying Assumption DQM* and it follows from standard spanning arguments (e.g., Newey (1990, Appendix B)) that the collection of functions ℓ_η (defined as in Assumption DQM*) associated with such families F is dense in $\mathcal{L}_\eta$. As a consequence, the set $\mathcal{J}_f$ is dense in the set of symmetric 2×2 matrices $\mathcal{I}_F$ for which the first diagonal element equals $\mathcal{I}_{ff}$ and (6) is positive semidefinite. In particular, the fact that $f\in\mathcal{F}_{\mathrm{DQM}}$ implies that $\mathcal{J}^{LF}_f$ belongs to the closure of $\mathcal{J}_f$.

To complete the proof of the inequality

$$\inf_{F:\mathcal{I}_F\in\mathcal{J}_f} \Psi^S_F(c,\alpha) \le \Psi^S_f(c,\alpha) \quad \forall c<0, \tag{38}$$

it therefore suffices to show that $\Psi^S_F(c,\alpha)$ is a continuous function of $\mathcal{I}_F$. Because K^S_α is continuous, the continuous mapping theorem can be used to show that if the sequence $\mathcal{I}_{F,n}$ is convergent, then $\psi^S_{F,n}(\mathcal{S}_{F,n},\mathcal{H}_{F,n}|c,\alpha)$ (defined from $\mathcal{I}_{F,n}$ in the natural way) converges in distribution. Using this fact and the dominated convergence theorem, it can be shown that $\Psi^S_F(c,\alpha)$ is a continuous function of $\mathcal{I}_F$.

Equation (24). The proof is similar to that of (23) and proceeds by showing that

$$\inf_{F:\mathcal{I}_F\in\mathcal{J}_f} \Psi^I_F(c,\alpha) \ge \Psi^I_f(c,\alpha) \quad \forall c<0 \tag{39}$$

and

$$\inf_{F:\mathcal{I}_F\in\mathcal{J}_f} \Psi^I_F(c,\alpha) \le \Psi^I_f(c,\alpha) \quad \forall c<0. \tag{40}$$

Inequality (40) follows from arguments analogous to those used to prove (38).

To establish (39), let $c<0$ be given and let F be any submodel satisfying assumption LAQ*. Also, let (S^F_T, H^F_T), (W, B_f, B_η) etcetera be as in Section 4 and define

$$\phi^I_{f,T}(Y_T|c,\alpha) := 1\left[cS^{f,I}_T - \frac{1}{2}c^2 H^{ff,I}_T > K^I_\alpha(c;\mathcal{J}^{LF}_f)\right],$$

where $S^{f,I}_T$ is defined in (26) and

$$H^{ff,I}_T := H^{ff}_T - (\mathcal{I}_{ff}-1)\left(T^{-3/2}\sum_{t=2}^T y_{t-1}\right)^2.$$

The statistic $\phi^I_{f,T}(Y_T|c,\alpha)$ satisfies

$$\lim_{T\to\infty} E_{\rho_T(c'),\eta_T(h)}\phi^I_T(Y_T|c,\alpha;f) = E\big[\psi^I_f(\mathcal{S}^I_f,\mathcal{H}^I_{ff}|c,\alpha)\exp(\Lambda_F(c',h))\big]$$

for every (c',h). In particular, $\lim_{T\to\infty} E_{\rho_T(c),\eta_T(0)}\phi^I_T(Y_T|c,\alpha;f) = \Psi^I_f(c,\alpha)$, so the proof of (39) can be completed by showing that $\phi^I_{f,T}(\cdot|c,\alpha)$ is locally asymptotically α-invariant in F.

To do so, it suffices to show that

$$E[\psi^I_f(\mathcal{S}^I_f,\mathcal{H}^I_{ff}|c,\alpha)|\mathcal{S}_F,\mathcal{H}_F] = E[\psi^I_f(\mathcal{S}^I_f,\mathcal{H}^I_{ff}|c,\alpha)|\mathcal{S}_{f.\eta},\mathcal{H}_F].$$

A sufficient condition for this to hold is that $(\mathcal{S}^I_f,\mathcal{H}^I_{ff})$ is independent of $\mathcal{S}_\eta$ conditional on $(\mathcal{S}_{f.\eta},\mathcal{H}_F)$. In turn, this conditional independence property follows from simple algebra and the fact that the conditional distribution of $(\mathcal{S}_f,\mathcal{S}^S_f,\mathcal{S}_\eta)'$ given W is normal with mean $(\int_0^1 W(r)\,dW(r),0,0)'$ and variance

$$\begin{pmatrix} (\mathcal{I}_{ff}-1)\int_0^1 W(r)^2\,dr & (\mathcal{I}_{ff}-1)\int_0^1 W(r)\,dr & \mathcal{I}_{f\eta}\int_0^1 W(r)\,dr \\ (\mathcal{I}_{ff}-1)\int_0^1 W(r)\,dr & \mathcal{I}_{ff}-1 & \mathcal{I}_{f\eta} \\ \mathcal{I}_{f\eta}\int_0^1 W(r)\,dr & \mathcal{I}_{f\eta} & \mathcal{I}_{\eta\eta} \end{pmatrix}.$$

Q.E.D.

PROOF OF THEOREM 6: Suppose H_0 holds and let $f\in\mathcal{F}_{\mathrm{AC}}$ and $\bar{c}<0$ be given.

It suffices to show that

$$\hat{\phi}^I_T(Y_T|\bar{c},\alpha) = \phi^I_{f,T}(Y_T|\bar{c},\alpha) + o_p(1),$$

where $\phi^I_{f,T}(\cdot|\bar{c},\alpha)$ was defined in the proof of Theorem 5. The displayed result will follow from the convergence theorem (for convergence in probability) if it can be shown that

$$\hat{S}^I_T = S^{f,I}_{T,\tau} + o_p(1) = S^{f,I}_T + o_p(1), \tag{41}$$

$$\hat{H}^I_T = H^{ff,I}_{T,\tau} + o_p(1) = H^{ff,I}_T + o_p(1), \tag{42}$$

$$\hat{\mathcal{I}}_T = \mathcal{I}_{ff} + o_p(1), \tag{43}$$

where $S^{f,I}_T$ and $H^{ff,I}_T$ are as in the proof of Theorem 5 and

$$S^{f,I}_{T,\tau} := \frac{1}{T}\sum_{t=\tau_T+1}^{T} y_{t-1}\ell_f(\Delta y_t) - \left(\frac{1}{T^{3/2}}\sum_{t=\tau_T+1}^{T} y_{t-1}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=\tau_T+1}^{T}[\ell_f(\Delta y_t)-\Delta y_t]\right),$$

$$H_{T,\tau}^{ff,I} := \frac{\mathcal{I}_{ff}}{T^2}\sum_{t=\tau_T+1}^{T} y_{t-1}^2 - (\mathcal{I}_{ff}-1)\left(\frac{1}{T^{3/2}}\sum_{t=\tau_T+1}^{T} y_{t-1}\right)^2.$$

The result (43) follows from Bickel (1982, Section 6.2(i)), while (42) follows from (43), $\lim_{T\to\infty}\tau_T/T=0$, and simple algebra. The second equality in (41) follows from $\lim_{T\to\infty}\tau_T/T=0$ and simple algebra. Finally, to establish the first equality in (41), let $\check{\ell}_T$ be a "bias-corrected" version of $\hat{\ell}_T$ given by

$$\check{\ell}_T(\Delta y_t) := \tilde{\ell}_{\tau_T}(\Delta y_t|\Delta y_1,\dots,\Delta y_{\tau_T}) - \int_{-\infty}^{\infty}\tilde{\ell}_{\tau_T}(\varepsilon|\Delta y_1,\dots,\Delta y_{\tau_T})f(\varepsilon)\,d\varepsilon.$$

Reasoning analogous to that of the proof of Theorem 4 can be used to show that

$$\frac{1}{T}\sum_{t=\tau_T+1}^{T} y_{t-1}[\check{\ell}_T(\Delta y_t)-\ell_f(\Delta y_t)] = o_p(1),$$

$$\frac{1}{\sqrt{T}}\sum_{t=\tau_T+1}^{T}[\check{\ell}_T(\Delta y_t)-\ell_f(\Delta y_t)] = o_p(1).$$

The first equality in (41) can be established using these results and the fact that

$$\hat{S}_T^I = \frac{1}{T}\sum_{t=\tau_T+1}^{T} y_{t-1}\check{\ell}_T(\Delta y_t) - \left(\frac{1}{T^{3/2}}\sum_{t=\tau_T+1}^{T} y_{t-1}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=\tau_T+1}^{T}[\check{\ell}_T(\Delta y_t)-\Delta y_t]\right).$$

Q.E.D.

REFERENCES

ANDREWS, D. W. K. (1986): "Complete Consistency: A Testing Analogue of Estimator Consistency," *Review of Economic Studies*, 53, 263–269. [1125]

BEGUN, J. M., W. J. HALL, W.-M. HUANG, AND J. A. WELLNER (1983): "Information and Asymptotic Efficiency in Parametric–Nonparametric Models," *Annals of Statistics*, 11, 432–452. [1104,1120]

BERAN, R. (1974): "Asymptotically Efficient Adaptive Rank Estimates in Location Models," *Annals of Statistics*, 2, 63–74. [1118]

——— (1978): "An Efficient and Robust Adaptive Estimator of Location," *Annals of Statistics*, 6, 292–313. [1118]

BICKEL, P. J. (1982): "On Adaptive Estimation," *Annals of Statistics*, 10, 647–671. [1120,1122, 1126-1128,1139]

BICKEL, P. J., C. A. J. KLAASSEN, Y. RITOV, AND J. A. WELLNER (1998): *Efficient and Adaptive Estimation for Semiparametric Models*. New York: Springer-Verlag. [1104,1122]

BILLINGSLEY, P. (1999): *Convergence of Probability Measures* (Second Ed.). New York: Wiley. [1106,1108]

BOSWIJK, H. P. (2005): "Adaptive Testing for a Unit Root With Nonstationary Volatility," Working Paper, University of Amsterdam. [1106]

CHAN, N. H., AND C. Z. WEI (1988): "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes," *Annals of Statistics*, 16, 367–401. [1108]

CHERNOZHUKOV, V., AND H. HONG (2004): "Likelihood Estimation and Inference in a Class of Nonregular Econometric Models," *Econometrica*, 72, 1445–1480. [1106]

CHOI, S., W. J. HALL, AND A. SCHICK (1996): "Asymptotically Uniformly Most Powerful Tests in Parametric and Semiparametric Models," *Annals of Statistics*, 24, 841–861. [1104,1120]

DICKEY, D. A., AND W. A. FULLER (1979): "Distribution of the Estimators for Autoregressive Time Series With a Unit Root," *Journal of the American Statistical Association*, 74, 427–431. [1103]

——— (1981): "Likelihood Ratio Statistics for Autoregressive Time Series With a Unit Root," *Econometrica*, 49, 1057–1072. [1103]

DROST, F. C., C. A. J. KLAASSEN, AND B. J. M. WERKER (1997): "Adaptive Estimation in Time-Series Models," *Annals of Statistics*, 25, 786–817. [1118,1135,1136]

ELIASZ, P. (2004): "Optimal Median Unbiased Estimation of Coefficients on Highly Persistent Regressors," Working Paper, Princeton University. [1104]

ELLIOTT, G., T. J. ROTHENBERG, AND J. H. STOCK (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, 813–836. [1103]

GUSHCHIN, A. A. (1995): "Asymptotic Optimality of Parameter Estimators Under the LAQ Condition," *Theory of Probability and Its Applications*, 40, 261–272. [1104]

HALDRUP, N., AND M. JANSSON (2006): "Improving Size and Power in Unit Root Testing," in *Palgrave Handbook of Econometrics. Vol. 1, Econometric Theory*, ed. by T. C. Mills and K. Patterson. New York: Palgrave Macmillan, 252–277. [1103]

HASAN, M. N., AND R. W. KOENKER (1997): "Robust Rank Tests of the Unit Root Hypothesis," *Econometrica*, 65, 133–161. [1115]

HERCE, M. A. (1996): "Asymptotic Theory of LAD, Estimation in a Unit Root Process With Finite Variance Errors," *Econometric Theory*, 12, 129–153. [1115]

HIRANO, K., AND J. R. PORTER (2003): "Asymptotic Efficiency in Parametric Structural Models With Parameter-Dependent Support," *Econometrica*, 71, 1307–1338. [1106]

JANSSON, M. (2007): "Supplement to 'Semiparametric Power Envelopes for Tests of the Unit Root Hypothesis'," *Econometrica Supplemental Material*, 76, http://www.econometricsociety.org/ecta/Supmat/6113-proofs.pdf. [1111,1115,1135,1137]

JANSSON, M., AND M. J. MOREIRA (2006): "Optimal Inference in Regression Models With Nearly Integrated Regressors," *Econometrica*, 74, 681–714. [1104]

JEGANATHAN, P. (1995): "Some Aspects of Asymptotic Theory With Applications to Time Series Models," *Econometric Theory*, 11, 818–887. [1104,1108,1111,1113]

——— (1997): "On Asymptotic Inference in Linear Cointegrated Time Series Systems," *Econometric Theory*, 13, 692–745. [1129]

KLAASSEN, C. A. J. (1987): "Consistent Estimation of the Influence Function of Locally Asymptotically Linear Estimators," *Annals of Statistics*, 15, 1548–1562. [1122,1127]

KOENKER, R., AND Z. XIAO (2004): "Unit Root Quantile Autoregression Inference," *Journal of the American Statistical Association*, 99, 775–787. [1115]

KOUL, H. L., AND A. SCHICK (1997): "Efficient Estimation in Nonlinear Autoregressive Time-Series Models," *Bernoulli*, 3, 247–277. [1122]

KREISS, J.-P. (1987a): "On Adaptive Estimation in Autoregressive Models When There Are Nuisance Functions," *Statistics and Decisions*, 5, 59–76. [1118]

——— (1987b): "On Adaptive Estimation in Stationary ARMA Processes," *Annals of Statistics*, 15, 112–133. [1118]

LE CAM, L. (1970): "On the Assumptions Used to Prove Asymptotic Normality of Maximum Likelihood Estimates," *Annals of Mathematical Statistics*, 41, 802–828. [1111]

——— (1972): "Limits of Experiments," in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1. Berkeley, CA: University of California Press, 245–261. [1106,1109]

——— (1986): *Asymptotic Methods in Statistical Decision Theory*. New York: Springer Verlag. [1111]
LE CAM, L., AND G. L. YANG (2000): *Asymptotics in Statistics Some Basic Concepts* (Second Ed.). New York: Springer Verlag. [1111]
LEHMANN, E. L., AND J. P. ROMANO (2005): *Testing Statistical Hypotheses* (Third Ed.). New York: Springer Verlag. [1104,1117,1133,1134]
LEVIT, B. Y. (1975): "On the Efficiency of a Class of Nonparametric Estimates," *Theory of Probability and Its Applications*, 3, 723–740. [1118]
LING, S., AND M. MCALEER (2003): "On Adaptive Estimation in Nonstationary ARMA Models With GARCH Errors," *Annals of Statistics*, 31, 642–674. [1104,1120]
MÜLLER, U. K., AND G. ELLIOTT (2003): "Tests for Unit Root and the Initial Condition," *Econometrica*, 71, 1269–1286. [1106]
MURPHY, S. A., AND A. W. VAN DER VAART (1997): "Semiparametric Likelihood Ratio Inference," *Annals of Statistics*, 25, 1471–1509. [1117]
——— (2000): "On Profile Likelihood," *Journal of the American Statistical Association*, 95, 449–485. [1117]
NEWEY, W. K. (1990): "Semiparametric Efficiency Bounds," *Journal of Applied Econometrics*, 5, 99–135. [1104,1118,1137]
NG, S., AND P. PERRON (1995): "Unit Root Tests Is ARMA Models With Data-Dependent Methods for the Selection of the Truncation Lag," *Journal of the American Statistical Association*, 90, 268–281. [1103]
——— (2001): "Lag Length Selection and the Construction of Unit Root Tests With Good Size and Power," *Econometrica*, 69, 1519–1554. [1103]
PAPARODITIS, E., AND D. N. POLITIS (2003): "Residual-Based Block Bootstrap for Unit Root Testing," *Econometrica*, 71, 813–855. [1103]
PARK, J. Y. (2003): "Bootstrap Unit Root Tests," *Econometrica*, 71, 1845–1895. [1103]
PERRON, P., AND S. NG (1996): "Useful Modifications to Unit Root Tests With Dependent Errors and Their Local Asymptotic Properties," *Review of Economic Studies*, 63, 435–465. [1103]
PHILLIPS, P. C. B. (1987): "Time Series Regression With a Unit Root," *Econometrica*, 55, 277–301. [1103]
——— (1991): "Optimal Inference in Cointegrated Systems," *Econometrica*, 59, 283–306. [1104]
PHILLIPS, P. C. B., AND P. PERRON (1988): "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75, 335–346. [1103]
PHILLIPS, P. C. B., AND Z. XIAO (1998): "A Primer on Unit Root Testing," *Journal of Economic Surveys*, 12, 423–469. [1103]
POLLARD, D. (1997): "Another Look at Differentiability in Quadratic Mean," in *Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics*, ed. by D. Pollar, E. Torgersen, and G. L. Yang. New York: Springer Verlag, 305–314. [1111]
ROTHENBERG, T. J. (2000): "Testing for Unit Roots in AR and MA Models," in *Applications of Differential Geometry to Econometrics*, ed. by P. Marriott and M. Salmon. New York: Cambridge University Press, 281–293. [1103]
ROTHENBERG, T. J., AND J. H. STOCK (1997): "Inference in a Nearly Integrated Autoregressive Model With Nonnormal Innovations," *Journal of Econometrics*, 80, 269–286. [1103,1109,1118]
SCHICK, A. (1986): "On Asymptotically Efficient Estimation in Semiparametric Models," *Annals of Statistics*, 14, 1139–1151. [1122]
——— (1987): "A Note on the Construction of Asymptotically Linear Estimators," *Journal of Statistical Planning and Inference*, 16, 89–105. [1128]
STEIN, C. (1956): "Efficient Nonparametric Estimation and Testing," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1. Berkeley, CA: University of California Press, 187–196. [1104,1118,1120,1122,1128,1133]
STOCK, J. H. (1994): "Unit Roots, Structural Breaks and Trends," in *Handbook of Econometrics*, Vol. 4, ed. by R. F. Engle and D. L. McFadden. New York: North-Holland, 2739–2841. [1103]

STOCK, J. H., AND M. W. WATSON (1993): "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica*, 61, 783–820. [1104]

STONE, C. J. (1975): "Adaptive Maximum Likelihood Estimators of a Location Parameter," *Annals of Statistics*, 3, 267–284. [1118]

THOMPSON, S. B. (2004): "Robust Tests of the Unit Root Hypothesis Should Not Be Modified," *Econometric Theory*, 20, 360–381. [1115]

VAN DER VAART, A. W. (1991): "An Asymptotic Representation Theorem," *International Statistical Review*, 59, 97–121. [1117]

——— (1998): *Asymptotic Statistics*. New York: Cambridge University Press. [1104]

——— (2002): "The Statistical Work of Lucien Le Cam," *Annals of Statistics*, 30, 631–682. [1108,1111]

Dept. of Economics, University of California at Berkeley, 508-1 Evans Hall, 3880, Berkeley, CA 94720-3880, U.S.A. and Center for Research in Econometric Analysis of Time Series (CREATES), School of Economics and Management, University of Aarhus, DK-8000 Aarhus C, Denmark; mjansson@econ.berkeley.edu.

Manuscript received October, 2005; final revision received April, 2007.

Econometrica Supplementary Material

SUPPLEMENT TO "SEMIPARAMETRIC POWER ENVELOPES FOR TESTS OF THE UNIT ROOT HYPOTHESIS"

(*Econometrica*, Vol. 76, No. 5, September 2008, 1103–1142)

BY MICHAEL JANSSON

PROOF OF LEMMA 2: Suppose f satisfies Assumption DQM.

The result $\ell_f \in \mathcal{L}_f$ follows from standard arguments. Specifically, $E[\ell_f(\varepsilon)] = 0$ and $E[\ell_f(\varepsilon)^2] < \infty$ by van der Vaart (2002, Lemma 1.8). Furthermore, using van der Vaart (2002, Example 1.15), the property $E[\varepsilon\ell_f(\varepsilon)] = 1$ can be deduced from the fact that the functional $\int_{-\infty}^{\infty} f(\varepsilon-\theta)\,d\varepsilon = \theta$ is differentiable in the ordinary sense and the sense of van der Vaart (2002, Definition 1.14). Finally, by the Cauchy–Schwarz inequality, $E[\ell_f(\varepsilon)^2] \geq E[\varepsilon^2]/E[\varepsilon\ell_f(\varepsilon)]^2 = 1$.

To establish the locally asymptotically quadratic (LAQ) property, let c_T be a bounded sequence. The log likelihood ratio $L_T^f(c_T)$ admits the expansion

$$\begin{aligned} L_T^f(c_T) = {}& \frac{c_T}{T}\sum_{t=2}^{T} y_{t-1}\ell_f(\Delta y_t) + \sum_{t=2}^{T} R_{Tt} \\ & - \frac{1}{4}\sum_{t=2}^{T}\left[\frac{c_T}{T}y_{t-1}\ell_f(\Delta y_t) + R_{Tt}\right]^2 (1+\beta_{Tt}), \end{aligned}$$

where $R_{Tt} := R_f(\Delta y_t, c_T y_{t-1}/T)$, $\beta_{Tt} := \beta[c_T y_{t-1}\ell_f(\Delta y_t)/T + R_{Tt}]$, and the defining properties of $R_f(\cdot)$ and $\beta(\cdot)$ are

$$\sqrt{\frac{f(\varepsilon-\theta)}{f(\varepsilon)}} = 1 + \frac{1}{2}\theta\ell_f(\varepsilon) + \frac{1}{2}R_f(\varepsilon,\theta),$$

$$\log(1+r) = r - \frac{1}{2}r^2[1+\beta(2r)].$$

The proof of Lemma 2 will be completed by showing that

$$\text{(S1)} \qquad \sum_{t=2}^{T} R_{Tt} = -\frac{1}{4}c_T^2\frac{\mathcal{I}_{ff}}{T^2}\sum_{t=2}^{T} y_{t-1}^2 + o_{p_{0,f}}(1),$$

$$\text{(S2)} \qquad \sum_{t=2}^{T}\left[\frac{c_T}{T}y_{t-1}\ell_f(\Delta y_t) + R_{Tt}\right]^2(1+\beta_{Tt}) = c_T^2\frac{\mathcal{I}_{ff}}{T^2}\sum_{t=2}^{T} y_{t-1}^2 + o_{p_{0,f}}(1).$$

In the rest of the proof, suppose H_0 holds and let ϑ_T be any positive sequence for which $\vartheta_T \to 0$ and $\sqrt{T}\vartheta_T \to \infty$ (as $T \to \infty$).

 DOI: 10.3982/ECTA6113SUPP

Equation (S1). Let $\tilde{R}_{Tt} := 1(|c_T y_{t-1}/T| \leq \vartheta_T)R_{Tt}$ denote a truncated version of R_{Tt}. Because $\max_{2\leq t\leq T}|c_T y_{t-1}/\sqrt{T}| = O_p(1)$ and $\sqrt{T}\vartheta_T \to \infty$, the sequences $\tilde{R}_{Tt}$ and R_{Tt} are asymptotically equivalent in the sense that $\sum_{t=2}^T R_{Tt} = \sum_{t=2}^T \tilde{R}_{Tt} + o_p(1)$.

Now

$$E_{t-1}(\tilde{R}_{Tt}^2) = 1\left(\left|\frac{c_T}{T}y_{t-1}\right| \leq \vartheta_T\right)E_{t-1}\left[R_f\left(\varepsilon_t, \frac{c_T}{T}y_{t-1}\right)^2\right]$$
$$\leq V_f(\vartheta_T)\frac{c_T^2}{T^2}y_{t-1}^2,$$

where $V_f(\vartheta) := \sup_{|\theta|\leq\vartheta,\theta\neq 0}\theta^{-2}E[R_f(\varepsilon,\theta)^2]$ and $E_{t-1}[\cdot]$ denotes conditional expectation given $\{\varepsilon_1,\ldots,\varepsilon_{t-1}\}$. By Assumption DQM, $\lim_{\vartheta\downarrow 0}V_f(\vartheta) = 0$. As a consequence, using $\vartheta_T = o(1)$ and $E(y_{t-1}^2) = t-1$,

$$\sum_{t=2}^T E_{t-1}(\tilde{R}_{Tt}^2) \leq V_f(\vartheta_T)E\left(\frac{c_T^2}{T^2}\sum_{t=2}^T y_{t-1}^2\right) = V_f(\vartheta_T)O(1) = o(1),$$

implying that $\sum_{t=2}^T \tilde{R}_{Tt} = \sum_{t=2}^T E_{t-1}(\tilde{R}_{Tt}) + o_p(1)$. Moreover,

$$\sum_{t=2}^T E_{t-1}(\tilde{R}_{Tt}) = -\frac{1}{4}\mathcal{I}_{ff}\frac{c_T^2}{T^2}\sum_{t=2}^T 1\left(\left|\frac{c_T}{T}y_{t-1}\right| \leq \vartheta_T\right)y_{t-1}^2$$
$$+\sum_{t=2}^T 1\left(\left|\frac{c_T}{T}y_{t-1}\right| \leq \vartheta_T\right)r_f\left(\frac{c_T}{T}y_{t-1}\right),$$

where $r_f(\theta) := \frac{1}{4}\mathcal{I}_{ff}\theta^2 + E[R_f(\varepsilon,\theta)]$ and

$$\frac{1}{T^2}\sum_{t=2}^T 1\left(\left|\frac{c_T}{T}y_{t-1}\right| \leq \vartheta_T\right)y_{t-1}^2 = \frac{1}{T^2}\sum_{t=2}^T y_{t-1}^2 + o_p(1)$$

because $\max_{2\leq t\leq T}|c_T y_{t-1}/\sqrt{T}| = O_p(1)$ and $\sqrt{T}\vartheta_T \to \infty$. The proof of (S1) can therefore be completed by showing that

$$\sum_{t=2}^T 1\left(\left|\frac{c_T}{T}y_{t-1}\right| \leq \vartheta_T\right)r_f\left(\frac{c_T}{T}y_{t-1}\right) = o_p(1).$$

The relationship in the preceding display follows from $\vartheta_T = o(1)$ and the fact that

$$\left|\sum_{t=2}^T 1\left(\left|\frac{c_T}{T}y_{t-1}\right| \leq \vartheta_T\right)r^f\left(\frac{c_T}{T}y_{t-1}\right)\right| \leq v_f(\vartheta_T)\frac{c_T^2}{T^2}\sum_{t=2}^T y_{t-1}^2$$

$$= v_f(\vartheta_T)O_p(1),$$

where $v_f(\vartheta) := \sup_{|\theta|\leq\vartheta,\theta\neq 0} \theta^{-2}|r_f(\theta)| = o(1)$ as $\vartheta \downarrow 0$ (Pollard (1997, Lemma 1)).

Equation (S2). To prove (S2), it suffices to show that

$$\sum_{t=2}^{T}\left[\frac{c_T}{T}y_{t-1}\ell_f(\varepsilon_t)+R_{Tt}\right]^2 = c_T^2\frac{\mathcal{I}_{ff}}{T^2}\sum_{t=2}^{T}y_{t-1}^2+o_p(1)$$

and

$$\max_{2\leq t\leq T}\bigl|\beta[c_T T^{-1}y_{t-1}\ell_f(\varepsilon_t)+R_{Tt}]\bigr| = o_p(1).$$

By Taylor's theorem, $\beta(r)\to 0$ as $|r|\to 0$. Moreover,

$$\max_{2\leq t\leq T}\left|\frac{y_{t-1}}{T}\ell_f(\varepsilon_t)\right| \leq \max_{2\leq t\leq T}\left|\frac{y_{t-1}}{\sqrt{T}}\right|\max_{2\leq t\leq T}\left|\frac{\ell_f(\varepsilon_t)}{\sqrt{T}}\right| = O_p(1)o_p(1) = o_p(1)$$

and $\max_{2\leq t\leq T}|R_{Tt}| \leq \sqrt{\sum_{t=2}^{T}R_{Tt}^2}$. Therefore, the desired result will follow from

$$\text{(S3)}\qquad \frac{1}{T^2}\sum_{t=2}^{T}y_{t-1}^2\ell_f(\varepsilon_t)^2 = \frac{\mathcal{I}_{ff}}{T^2}\sum_{t=2}^{T}y_{t-1}^2+o_p(1)$$

and

$$\text{(S4)}\qquad \sum_{t=2}^{T}R_{Tt}^2 = o_p(1).$$

As noted by Jeganathan (1995, Lemma 24), (S3) can be deduced with the help of the proof of Hall and Heyde (1980, Theorem 2.23) if it can be shown that

$$\frac{1}{T^2}\sum_{t=2}^{T}E_{t-1}\left[y_{t-1}^2\ell_f(\varepsilon_t)^2 1\left(\left|\frac{y_{t-1}}{T}\ell_f(\varepsilon_t)\right|>\varrho\right)\right] = o_p(1)\quad \forall\varrho>0.$$

To do so, let $\varrho>0$ be given and define $Q_f(r) := E[\ell_f(\varepsilon)^2 1(|\ell_f(\varepsilon)|>r)]$. Because Q_f is nonincreasing and $\lim_{r\to\infty}Q_f(r)=0$,

$$\begin{aligned}&\frac{1}{T^2}\sum_{t=2}^{T}E_{t-1}\left[y_{t-1}^2\ell_f(\varepsilon_t)^2 1\left(\left|\frac{y_{t-1}}{T}\ell_f(\varepsilon_t)\right|>\varrho\right)\right]\\ &\quad= \frac{1}{T^2}\sum_{t=2}^{T}y_{t-1}^2 Q_f\left(\frac{\sqrt{T}\varrho}{|y_{t-1}/\sqrt{T}|}\right)\end{aligned}$$

$$\leq \left(\frac{1}{T^2}\sum_{t=2}^{T} y_{t-1}^2\right) \max_{2\leq t\leq T} Q_f\left(\frac{\sqrt{T}\varrho}{|y_{t-1}/\sqrt{T}|}\right)$$
$$= O_p(1)o_p(1) = o_p(1),$$

where the penultimate equality uses $\max_{2\leq t\leq T}|y_{t-1}/\sqrt{T}| = O_p(1)$.

It can be shown that $\sum_{t=2}^{T} R_{Tt}^2 = \sum_{t=2}^{T} \tilde{R}_{Tt}^2 + o_p(1)$. Moreover,

$$\sum_{t=2}^{T} E_{t-1}\left[\tilde{R}_{Tt}^2 1(|\tilde{R}_{Tt}| > \varrho)\right] \leq \sum_{t=2}^{T} E_{t-1}(\tilde{R}_{Tt}^2) = o_p(1) \quad \forall \varrho > 0,$$

where the equality was established in the proof of (S1). A second application of the proof of Hall and Heyde (1980, Theorem 2.23) therefore establishes (S4). *Q.E.D.*

PROOF OF LEMMA 7: For any b, any $c < 0$, any $\alpha \in (0, 1)$, and any symmetric 2×2 matrix $\mathcal{I}_F$ for which

$$\mathrm{Var}\begin{pmatrix} W(1) \\ B_F(1) \end{pmatrix} = \begin{pmatrix} 1 & e_1' \\ e_1 & \mathcal{I}_F \end{pmatrix}$$

is positive semidefinite, let $K_\alpha^S(b, c; \mathcal{I}_F)$ be the $1-\alpha$ quantile of

$$G(W, Z, b, c; \mathcal{I}_F)$$
$$:= c\left[\int_0^1 W(r)\,dW(r) + \frac{\mathcal{H}_{f\eta}}{\mathcal{H}_{\eta\eta}} b + \sqrt{\mathcal{H}_{ff.\eta} - \int_0^1 W(r)^2\,dr}\, Z\right]$$
$$- \frac{1}{2}c^2 \mathcal{H}_{ff},$$

where $Z \sim \mathcal{N}(0, 1)$ is independent of W and $\mathcal{H}_{f\eta}$, $\mathcal{H}_{\eta\eta}$, etc. are as in Section 4.

The function K_α^S satisfies $E[\psi_F^S(\mathcal{S}_F, \mathcal{H}_F|c, \alpha)|\mathcal{S}_\eta] = \alpha$ because it follows from elementary facts about Brownian motions that

$$\frac{\mathcal{S}_{f.\eta} - \int_0^1 W(r)\,dW(r)}{\sqrt{\mathcal{H}_{ff.\eta} - \int_0^1 W(r)^2\,dr}} \sim \mathcal{N}(0, 1)$$

independent of W and $\mathcal{S}_\eta$, where $\mathcal{S}_{f.\eta}$ and $\mathcal{S}_\eta$ are as in Section 4.

Continuity of K_α^S follows from the fact that $G(W, Z, b_n, c_n; \mathcal{I}_{F,n})$ converges in distribution to a continuous random variable whenever the sequence $(b_n, c_n, \mathcal{I}_{F,n})$ is convergent (and $G(W, Z, b_n, c_n; \mathcal{I}_{F,n})$ is well defined for each n). *Q.E.D.*

PROOF OF (27): Let $f \in \mathcal{F}_{\mathrm{DQM}}$ and $c < 0$ be given, suppose F satisfies Assumption DQM*, and let (S_T^F, H_T^F), (W, B_f, B_η), etc. be as in Section 4. Because K_α^S is continuous (Lemma 7) and

$$(S_T^f, H_T^{ff}, S_T^{f,S}, S_T^\eta) \to_{d_{0,f}} (\mathcal{S}_f, \mathcal{H}_{ff}, \mathcal{S}_f^S, \mathcal{S}_\eta),$$

the sequence $\phi_{f,T}^S(\cdot|c, \alpha)$ satisfies

$$\phi_{f,T}^S(Y_T|c, \alpha) \to_{d_{0,f}} \psi_f^S(\mathcal{S}_f, \mathcal{H}_{ff}, \mathcal{S}_f^S|c, \alpha).$$

It follows from these convergence results, Le Cam's third lemma, and the result

$$L_T^F(c, h) \to_{d_{0,f}} \Lambda_F(c, h) := (c, h)\mathcal{S}_F - \frac{1}{2}(c, h)\mathcal{H}_F(c, h)' \quad \forall (c, h)$$

that

$$\begin{aligned}&\lim_{T\to\infty} E_{\rho_T(c'),\eta_T(h)} \phi_T^S(Y_T|c, \alpha; f)\\ &\quad = E\big[\psi_f^S(\mathcal{S}_f, \mathcal{H}_{ff}, \mathcal{S}_f^S|c, \alpha) \exp(\Lambda_F(c', h))\big]\end{aligned}$$

for every (c', h). In particular, $\lim_{T\to\infty} E_{\rho_T(c),\eta_T(0)} \phi_T^S(Y_T|c, \alpha; f) = \Psi_f^S(c, \alpha)$, implying that the proof of (27) can be completed by showing that $\phi_{f,T}^S(\cdot|c, \alpha)$ is locally asymptotically α-similar in F.

To do so, it suffices to show that $E[\psi_f^S(\mathcal{S}_f, \mathcal{H}_{ff}, \mathcal{S}_f^S|c, \alpha)|\mathcal{S}_\eta] = \alpha$. Let

$$\mathcal{S}_\eta^\perp := \mathcal{S}_\eta - \frac{\mathcal{I}_{f\eta}}{\mathcal{I}_{ff} - 1}\mathcal{S}_f^S.$$

Because $B_\eta - \mathcal{I}_{f\eta}(\mathcal{I}_{ff} - 1)^{-1}(B_f - W)$ and (W, B_f) are independent, $\mathcal{S}_\eta^\perp$ is independent of $(\mathcal{S}_f, \mathcal{H}_{ff}, \mathcal{S}_f^S)$ and

$$E[\psi_f^S(\mathcal{S}_f, \mathcal{H}_{ff}, \mathcal{S}_f^S|c, \alpha)|\mathcal{S}_f^S, \mathcal{S}_\eta^\perp] = E[\psi_f^S(\mathcal{S}_f, \mathcal{H}_{ff}, \mathcal{S}_f^S|c, \alpha)|\mathcal{S}_f^S] = \alpha,$$

where the second equality is the defining property of K_α^S. Because $\mathcal{S}_\eta$ is a function of $(\mathcal{S}_f^S, \mathcal{S}_\eta^\perp)$, it therefore follows from the law of iterated expectations that

$$\begin{aligned}E[\psi_f^S(\mathcal{S}_f, \mathcal{H}_{ff}, \mathcal{S}_f^S|c, \alpha)|\mathcal{S}_\eta] &= E\big(E[\psi_f^S(\mathcal{S}_f, \mathcal{H}_{ff}, \mathcal{S}_f^S|c, \alpha)|\mathcal{S}_f^S, \mathcal{S}_\eta^\perp]\big|\mathcal{S}_\eta\big)\\ &= \alpha,\end{aligned}$$

as was to be shown. *Q.E.D.*

REFERENCES

HALL, P., AND C. C. HEYDE (1980): *Martingale Limit Theory and Its Application*. New York: Academic Press. [3,4]

JEGANATHAN, P. (1995): "Some Aspects of Asymptotic Theory With Applications to Time Series Models," *Econometric Theory*, 11, 818–887. [3]

POLLARD, D. (1997): "Another Look at Differentiability in Quadratic Mean," in *Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics*, ed. by D. Pollard, E. Torgersen, and G. L. Yang. New York: Springer Verlag, 305–314. [3]

VAN DER VAART, A. W. (2002): "Semiparametric Statistics," in *Lectures on Probability Theory and Statistics (Saint-Flour, 1999)*, ed. by P. Bernard. New York: Springer Verlag, 341–457. [1]

Dept. of Economics, University of California at Berkeley, 508-1 Evans Hall, 3880, Berkeley, CA 94720-3880, U.S.A. and Center for Research in Econometric Analysis of Time Series (CREATES), School of Economics and Management, University of Aarhus, DK-8000 Aarhus C, Denmark; mjansson@econ.berkeley.edu.

Manuscript received October, 2005; final revision received April, 2007.