EFWW
7/9/08
11:00 a.m.

An Alternative Sense of Asymptotic Efficiency*

Ulrich K. Müller
Princeton University
Economics Department
Princeton, NJ, 08544

Abstract

The paper studies the asymptotic efficiency and robustness of hypothesis tests when models of interest are defined in terms of a weak convergence property. The null and local alternatives induce different limiting distributions for a random element, and a test is considered robust if it controls asymptotic size for all data generating processes for which the random element has the null limiting distribution. Under weak regularity conditions, asymptotically robust and efficient tests are then simply given by efficient tests of the limiting problem—that is, with the limiting random element assumed observed—evaluated at sample analogues. These tests typically coincide with suitably robustified versions of optimal tests in canonical parametric versions of the model. This paper thus establishes an alternative and broader sense of asymptotic efficiency for many previously derived tests in econometrics, such as tests for unit roots, parameter stability tests and tests about regression coefficients under weak instruments, and it provides a concrete limit on the potential for more powerful tests in less parametric set-ups.

JEL classification: C12, C14

Keywords: hypothesis tests, optimality, robustness, weak convergence

*The author would like to thank Mark Watson and seminar participants at Berkeley, Brown, Penn State and Stanford for helpful comments and discussions, as well as Jia Li for excellent research assistance. Financial support by the NSF through grant SES-0518036 is gratefully acknowledged.

1 Introduction

A continued focus of recent the econometrics literature has been the development of asymptotically optimal inference procedures for nonstandard problems: For instance, Elliott, Rothenberg, and Stock (1996) (abbreviated ERS in the following), Elliott (1999), Müller and Elliott (2003) and Müller (2007) derive optimal tests for an autoregressive unit root in a univariate framework; Elliott and Jansson (2003) derive optimal tests for a unit root with stationary covariates; Elliott, Jansson, and Pesavento (2005) derive optimal tests for the null hypothesis of no cointegration with known cointegrating vector; Jansson (2005) derives optimal tests for the null hypothesis of cointegration; Nyblom (1989), Andrews and Ploberger (1994) and Elliott and Müller (2006) derive optimal tests of parameter stability; Stock and Watson (1996) and Jansson and Moreira (2006) derive optimal inference in regression models with nearly integrated regressors; and Andrews, Moreira and Stock (2006, 2007) derive optimal tests for regression coefficients in the presence of weak instruments. By construction, these tests are optimal for a specific parametric version of the model, usually assuming i.i.d. Gaussian disturbances, in the sense of maximizing local asymptotic power. Furthermore, appropriate versions of these tests are robust in the sense that they yield the same asymptotic rejection probability under the null hypothesis (and local alternatives) for a wide range of data generating processes. The assumption of i.i.d. Gaussian disturbances is thus a natural starting point for the development of asymptotically efficient and robust tests.

Nevertheless, with a focus on efficiency, it is natural to ask whether there exist tests that are as good in the Gaussian case, but have higher local asymptotic power for non-Gaussian versions of the models. And indeed, Jansson (2007) draws on and extends the theory of semi-parametrically efficient tests to derive such tests for the unit root null hypothesis in the AR(1) model with i.i.d. driving errors of unknown distribution.[1] Also, as for the tests derived under Gaussianity, Jansson (2007) shows that suitably modified versions continue to have correct asymptotic rejection probability under the null hypothesis for a range of serial correlation structures.

This paper also considers the construction of asymptotically efficient tests for nonstandard problems, but with a stronger focus on robustness. For many models and hypothesis

[1] Also see Rothenberg and Stock (1997) on tests of the unit root hypothesis under non-Gaussian disturbances.

tests of interest, typical data generating processes imply the weak convergence of some random element to a limiting random element, whose distribution is different under the null and local alternative. For instance, under the null hypothesis of a unit root, the data (suitably scaled) converges to a Wiener process, and it converges to an Ornstein-Uhlenbeck process under the usual local-to-unity alternative. Typically, it is relatively straightforward to think about efficient tests in the 'limiting problem', where the limiting random element is directly observed. For instance, by the Neyman Pearson Lemma, the Radon-Nikodym derivative of the distribution of an Ornstein-Uhlenbeck process (for some fixed mean reversion parameter) and the distribution of a Wiener process is the best point-optimal test statistics in the limiting version of the unit root testing problem. Now suppose one is sufficiently unsure about the nature of the short run dynamics that one would like the test not to overreject whenever the data converges to a Wiener process—or, more generally, whenever the random element converges to its null limiting distribution. If one restricts attention to tests that are robust in this sense, then it is shown that (under mild regularity conditions), the best test statistic is simply given by the best test in the limiting problem, evaluated at sample analogues. In the unit root testing example, this test is asymptotically equivalent to the best unit root test under Gaussian i.i.d. disturbances, so that the test derived by ERS is this best robust test. Any test that has higher asymptotic power than this test for some non-Gaussian version of the model (such as Jansson's (2007) test) necessarily lacks robustness: its asymptotic rejection probability is larger than the nominal level for some model whose suitably scaled data converges weakly to a Wiener process.

The upshot of this analysis is straightforward: to determine the asymptotically efficient robust test in the sense described above, one only needs to consider efficient tests in the limiting problem, where the limiting random element is assumed observed. The potentially complicated small sample testing problem is thus replaced by a (typically) much simpler one. This aspect of the approach makes it somewhat akin to LeCam's Limits of Experiments—see van der Vaart (1998) for an introduction. The arguments, however, have distinct starting points: The Limit of Experiments approach considers a sequence of fully specified parametric models, and derives implications from the limiting behavior of the (small sample efficient) likelihood ratio statistics; the approach here, in contrast, defines models in terms of their weak convergence properties, and studies efficiency by considering the implied asymptotic properties of tests.

The basic result holds quite generally, including for cases where tests are restricted to satisfy some asymptotic unbiasedness or similarity constraint, or to be invariant. The results here may thus be applied to argue for a broader asymptotic efficiency of the test statistics derived in the 14 papers cited in the first paragraph of this introduction. In addition, they provide a precise sense in which Sowell's (1996) GMM parameter stability tests are asymptotically efficient. Finally, the results of this paper also imply efficiency of some recent nonstandard methods that take a weak convergence assumption as their starting point, such as those suggested in Müller and Watson (2007, 2008) and Ibragimov and Müller (2007).

The remainder of the paper is organized as follows. Section 2 introduces the formal framework and contains the main result. Section 3 discusses extensions regarding consistently estimable nuisance parameters, invariance restrictions and uniformity issues. A running example throughout Sections 2 and 3 is the problem of testing for an autoregressive unit root in a univariate time series. Section 4 discusses the application of the approach to three additional testing problems: Elliott and Jansson's (2003) point-optimal tests for unit roots with stationary covariates; Andrews, Moreira and Stock's (2006) optimal tests statistics for linear instrumental variable regressions; and Sowell's (1996) tests for GMM parameter stability. Section 4 concludes. All proofs are collected in an appendix.

2 Efficiency and Robustness under a Weak Convergence Assumption

2.1 Set-up

The following notation and conventions are used throughout the paper: All limits are taken as $T \to \infty$. If S_1 is a metric space with metric d_{S_1}, then $\mathfrak{B}(S_1)$ is its Borel σ-algebra. If μ is a probability measure on $\mathfrak{B}(S_1)$, then its image measure under the $\mathfrak{B}(S_1)\backslash\mathfrak{B}(S_2)$ measurable mapping $f : S_1 \mapsto S_2$, where S_1 and S_2 are space with metrics d_{S_1} and d_{S_2}, is denoted $f\mu$. If no ambiguity arises, we suppress the dummy variable of integration, that is we write $\int f d\mu$ for $\int f(x)d\mu(x)$. By default, the product space $S_1 \times S_2$ is equipped with the metric $d_{S_1} + d_{S_2}$. We write $\mu_T \rightsquigarrow \mu_0$ or $X_T \rightsquigarrow X_0$ for the weak convergence of the random elements $X_0, X_1, \ldots$ with probability measures $\mu_0, \mu_1, \ldots$ on $\mathfrak{B}(S_1)$, and $X_T \xrightarrow{p} X$ for convergence in probability. The $\mathbb{R} \mapsto \mathbb{R}$ function $x \mapsto \lfloor x \rfloor$ is the integer part of x.

In a sample of size T, suppose we observe data $Y_T \in \mathbb{R}^{nT}$, which is the Tth row of a double-array of random variables. The distribution of Y_T depends on the statistical model m with parameter $\theta \in \Theta$, where Θ is a metric space, so that the distribution $F_T(m, \theta)$ of Y_T is a probability kernel. The hypotheses of interest are

$$H_0 : \theta \in \Theta_0 \qquad \text{against} \qquad H_1 : \theta \in \Theta_1 \tag{1}$$

where $\Theta = \Theta_0 \cup \Theta_1$.

Let h_T be a sequence of measurable functions $h_T : \mathbb{R}^{nT} \mapsto S$, where S is a complete and separable metric space. Denote by $P_T(m, \theta)$ the distribution of $X_T = h_T(Y_T)$ in model m with parameter θ, that is $P_T(m, \theta) = h_T F_T(m, \theta)$. Suppose the typical model m satisfies the following weak convergences under the null and alternative hypothesis

$$P_T(m, \theta) \rightsquigarrow P(\theta) \quad \text{pointwise for all } \theta \in \Theta_0 \tag{2}$$

$$P_T(m, \theta) \rightsquigarrow P(\theta) \quad \text{pointwise for all } \theta \in \Theta_1 \tag{3}$$

to some statistical model $P(\theta)$, where, for each $\theta_1, \theta_2 \in \Theta$, the probability measures $P(\theta_1)$ and $P(\theta_2)$ on $\mathfrak{B}(S)$ are absolutely continuous. The parameter θ should be thought of as describing *local* alternatives, such as the magnitude of the Pitman drift.

Unit Root Test Example: Consider testing for a unit root in a model with no deterministics against the local-to-unity alternative: We observe data $Y_T = (u_{T,1}, \cdots, u_{T,T})'$ from the model $u_{T,t} = \rho_T u_{T,t-1} + \nu_{T,t}$ and $u_{T,0} = 0$ for all T, where $\rho_T = 1 - c/T$ for some fixed $c \geq 0$, and the hypotheses are $H_0 : c = 0$ against $H_1 : c > 0$ (so that $\theta = c$, $\Theta_0 = \{0\}$ and $\Theta_1 = (0, \infty)$). Let $\hat{\omega}_T^2$ be a specific, "reasonable" long-run variance estimator. With $S = D_{[0,1]}$ the space of cadlag functions on the unit interval, equipped with the Billingsley (1968) metric, a typical model m for the disturbances $\nu_{T,t}$ satisfies $T^{-1/2}\hat{\omega}_T^{-1} u_{T,\lfloor \cdot T \rfloor} = h_T(Y_T) = \hat{J}_T(\cdot) \rightsquigarrow J_c(\cdot)$ on $D_{[0,1]}$, where J_c is an Ornstein-Uhlenbeck process $J_c(s) = \int_0^s e^{-c(s-r)} dW(r)$ with W a standard Wiener process. It is well known that the measure of J_c is absolutely continuous of with respect to the measure of $J_0 = W$. ▲

2.2 Limiting Problem

In this set-up, in the typical model m, the random element X_T converges weakly to X with distribution $P(\theta)$. It will be useful in the sequel to first consider in detail the "limiting problem", where X is directly observed:

$$H_0^{lp}: X \sim P(\theta), \quad \theta \in \Theta_0 \quad \text{against} \quad H_1^{lp}: X \sim P(\theta), \quad \theta \in \Theta_1. \tag{4}$$

Possibly randomized tests of H_0^{lp} in (4) are measurable functions $\varphi_S : S \mapsto [0,1]$, where $\varphi_S(x)$ indicates the probability of rejection conditional on observing $X = x$, so that the overall rejection probability of the test φ_S when $X \sim P(\theta)$ is given by $\int \varphi_S dP(\theta)$. The test φ_S is thus of level α when $\sup_{\theta \in \Theta_0} \int \varphi_S dP(\theta) \leq \alpha$. In many nonstandard problems, no uniformly most powerful test exists, so consider tests that maximize a weighted average power criterion

$$\text{WAP}(\varphi_S) = \int \left(\int \varphi_S dP(\theta) \right) dw(\theta), \tag{5}$$

where w is a probability measure on Θ_1. In general, the weighting function w describes the importance a researcher attaches to the ability of the test to reject for certain alternatives. A point-optimal test is a special case of a weighted average power maximizing test for a degenerate weighting function w that puts all mass at one point. Also, if a uniformly most powerful test exists, then it maximizes WAP for all choices for w. The WAP criterion is statistically convenient, since by standard arguments, the WAP maximizing test equivalently is the best test of H_0^{lp} in (4) against the single alternative $H_{1,w}^{lp} : X \sim \int P(\theta) dw(\theta)$.

With the WAP criterion as efficiency measure, efficient level-α tests φ_S^* in the limiting problem (4) thus maximize WAP subject to $\sup_{\theta \in \Theta_0} \int \varphi_S dP(\theta) \leq \alpha$.

Unit Root Test Example, ctd: The weak convergence $\hat{J}_T(\cdot) \rightsquigarrow J_c(\cdot)$ leads to the limiting problem where we directly observe the continuous time process X, and $H_0^{lp} : X \sim J_0(\cdot)$ against $H_1^{lp} : X \sim J_c(\cdot)$ with $c > 0$. As a weighting function in the WAP criterion, consider a degenerate distribution with all mass at c_1, so that we consider a point-optimal test, just as ERS. By Girsanov's theorem, the Radon-Nikodym derivative of the distribution of J_{c_1} with respect to the distribution of J_0, evaluated at X, is given by $L(X) = \exp[-\frac{1}{2}c_1(X(1)^2 - 1) - \frac{1}{2}c_1^2 \int_0^1 X(s)^2 ds]$. Thus, by the Neyman Pearson Lemma, the point-optimal test in the limiting problem is of the form $\varphi_S^*(X) = \mathbf{1}[L(X) > \text{cv}]$, where cv solves $P(L(J_0(\cdot)) > \text{cv}) = \alpha$. ▲

When Θ_0 is not a singleton, that is if H_0^{lp} is composite, the derivation of a WAP maximizing test is typically much more involved. The weighted average power maximizing test under a composite null hypothesis is typically given by the Neyman-Pearson test of $H_{0,\Lambda}^{lp} : X \sim \int P(\theta) d\Lambda(\theta)$ against $H_{1,w}^{lp} : X \sim \int P(\theta) dw(\theta)$, where Λ is the least favorable distribution for θ—see chapter 3.8 of Lehmann (1986) for discussion. For many problems, however, it is difficult to identify the least favorable distribution Λ. To make further progress, researchers therefore often restrict the class of tests under consideration by additional constraints, and derive the best test in the restricted class. Sometimes, the WAP maximizing test in the restricted class turns out to be uniformly most powerful (that is, maximizes WAP for all weighting functions), so that the issue of how to choose an appropriate weighting function is also avoided by imposing additional constraints.

We discuss invariance as a restriction in Section 3.2 below, and focus here on two other constraints on the tests φ_S. First, consider

$$\int \varphi_S dP(\theta) \geq \pi_0(\theta) \quad \text{for all } \theta \in \Theta_1 \tag{6}$$

for some function $\pi_0 : \Theta_1 \mapsto \mathbb{R}$. The formulation (6) allows for a range of cases: with $\pi_0 = 0$, (6) never binds; with $\pi_0 = \alpha$, (6) imposes unbiasedness; with π_0 equal to the power of the locally best test for some arbitrarily small neighborhood of Θ_0, (6) effectively selects φ_S to be the locally best test.

Second, consider a (conditional) similarity constraint of the form

$$\int (\varphi_S - \alpha) f_S dP(\theta) = 0 \quad \text{for all } \theta \in \bar{\Theta}_0 \text{ and } f_S \in \mathcal{F}_S \tag{7}$$

for some $\bar{\Theta}_0 \subset \Theta_0$, which would typically be the intersection of Θ_0 with the closure of Θ_1, and $\mathcal{F}_S$ some set of measurable and bounded functions $f_S : S \mapsto \mathbb{R}$. With $\mathcal{F}_S$ only containing the zero function, (7) never binds. With $\mathcal{F}_S$ only containing the function that is equal to one, (7) imposes similarity. Finally, suppose $\vartheta : S \mapsto U$ is a measurable function, and f_S contains all $S \mapsto \mathbb{R}$ functions of the form $f_U \circ \vartheta$, where $f_U : U \mapsto \mathbb{R}$ is continuous and bounded. Then (7) amounts to the restriction that the rejection probability of φ_S for θ on the boundary between the null and alternative hypothesis, conditional on $\vartheta(X)$, is equal to α, so that φ_S is a conditionally similar test.

To sum up, we will refer to level-α tests φ_S^* in the limiting problem (4) as efficient when φ_S^* maximizes weighted average power (5), subject to (6) and (7).

2.3 Asymptotically Efficient and Robust Tests

In the original hypothesis testing problem (1) with Y_T observed, tests are measurable functions $\varphi_T : \mathbb{R}^{nT} \mapsto [0, 1]$, where $\varphi_T(y_T)$ indicates the probability of rejection conditional on observing $Y_T = y_T$. The overall rejection probability of the test φ_T in model m is thus given by $\int \varphi_T dF_T(m, \theta)$.

As in the discussion of the limiting problem, we consider weighted average power as the criterion to measure the efficiency of tests φ_T. In particular, define

$$\text{WAP}_T(\varphi_T, m) = \int \int \varphi_T dF_T(m, \theta) dw(\theta),$$

where the probability measure w on Θ_1 has the same interpretation as discussed below (5). Also, define the asymptotic null rejection probability of test φ_T in model m as

$$\text{ARP}_0(\varphi_T, m) = \sup_{\theta \in \Theta_0} \limsup_{T \to \infty} \int \varphi_T dF_T(m, \theta).$$

With these definitions, an asymptotically powerful level-α test φ_T has large $\lim_{T\to\infty} \text{WAP}_T(\varphi_T, m)$, while $\text{ARP}_0(\varphi_T, m) \leq \alpha$. A reasonable definition of an asymptotically robust test is to impose that $\text{ARP}_0(\varphi_T, m) \leq \alpha$ for a large class of models m. Let $\mathcal{M}_0$ be the set of models satisfying (2), i.e. $\mathcal{M}_0$ collects all data generating processes for Y_T such that $P_T(m, \theta) = h_T F_T(m, \theta) \rightsquigarrow P(\theta)$ for all $\theta \in \Theta_0$. In this paper, the notion of robustness is interpreted as imposing that a test has asymptotic null rejection probability no larger than the nominal level for all models $m \in \mathcal{M}_0$, that is formally if

$$\sup_{\theta \in \Theta_0} \limsup_{T \to \infty} \int \varphi_T dF_T(m, \theta) \leq \alpha \quad \text{for all } m \in \mathcal{M}_0. \tag{8}$$

Analogously, define $\mathcal{M}_1$ as the set of models m that satisfy (3).

Unit Root Test Example, ctd: The literature has developed a large number of sufficient conditions on the disturbances $\nu_{T,t}$ that imply $\hat{J}_T(\cdot) \rightsquigarrow J_c(\cdot)$—see, for instance, McLeish (1974) for a martingale difference sequence framework, Wooldridge and White (1988) for mixing conditions, Phillips and Solo (1992) for linear process assumptions, Davidson (2002) for near-epoch dependence, and Stock (1994b) for general discussion. Arguably, when invoking such assumptions, researchers do not typically have a specific data generating process in mind that is known to satisfy the conditions; rather there is great uncertainty about the true data generating process, and the hope is that by deriving tests that are valid for a large class of data generating processes, the true model is also covered. The primitive conditions are therefore quite possibly not a reflection of what researchers are sure is true about the data generating process, but rather an attempt to assume little in order to gain robustness. In that perspective, it seems quite natural to further strengthen the robustness requirement and to impose that the asymptotic rejection probability is no bigger than the nominal level for all models that satisfy $\hat{J}_T(\cdot) \rightsquigarrow W(\cdot)$. In fact, Stock (1994a), White (2001, p. 179), Breitung (2002), Davidson (2002, 2007) and Müller (2006) *define* the unit root null hypothesis in terms of the convergence $T^{-1/2}u_{T,\lfloor \cdot T \rfloor} \rightsquigarrow \omega W(\cdot)$, making the requirement (8) quite natural for a unit root test. ▲

In addition to the robustness constraint (8), we allow for the possibility that tests φ_T are restricted to possess further asymptotic properties. In particular, we consider

$$\liminf_{T\to\infty} \int \varphi_T dF_T(m,\theta) \quad \geq \quad \pi_0(\theta) \text{ for all } m \in \mathcal{M}_1, \theta \in \Theta_1 \tag{9}$$

$$\lim_{T\to\infty} \int (\varphi_T - \alpha)(f_S \circ h_T) dF_T(m,\theta) \quad = \quad 0 \text{ for all } m \in \mathcal{M}_0, \theta \in \bar{\Theta}_0 \text{ and } f_S \in \mathcal{F}_S. \tag{10}$$

The constraints (9) and (10) are asymptotic analogues of the constraints (6) and (7) in the limiting problem introduced above. So setting $\pi_0 = \alpha$, for instance, imposes asymptotic unbiasedness of the test φ_T in the sense that for all models $m \in \mathcal{M}_1$, the asymptotic rejection probability of φ_T under the alternative is not smaller than the nominal level. The formulation (10) of "asymptotic conditional similarity" is convenient, as it avoids explicit limits of conditional distributions; see Jansson and Moreira (2006), page 694 for discussion and references. Also, without loss of generality, we can always impose (9) and (10), since with $\pi_0 = 0$ and $\mathcal{F}_S = \{0\}$, they do not constrain the tests φ_T in any way.

The main result of this paper is that under the robustness constraint (8), efficient tests in the limiting problem φ_S^*, evaluated at sample analogues with X replaced by $X_T = h_T(Y_T)$, yield asymptotically efficient tests in the original problem involving the observations Y_T.

Theorem 1 *Let $\varphi_S^* : S \mapsto [0,1]$ be a level-α test in the limiting problem (4) that maximizes weighted average power (5) subject to (6) and (7). Suppose φ_S^* is $P(\theta_0)$-almost everywhere continuous for some $\theta_0 \in \Theta_0$, and define $\hat{\varphi}_T^* : \mathbb{R}^{nT} \mapsto [0,1]$ as $\hat{\varphi}_T^* = \varphi_S^* \circ h_T$. Then*

(i) $\lim_{T\to\infty} \int \hat{\varphi}_T^* dF(m,\theta) \leq \alpha$ for all $m \in \mathcal{M}_0$ and $\theta \in \Theta_0$, and $\lim_{T\to\infty} \int\int \hat{\varphi}_T^* dF_T(m,\theta) dw(\theta) = \int\int \varphi_S^* dP(\theta) dw(\theta)$ for all $m \in \mathcal{M}_1$.

(ii) For any test $\varphi_T : \mathbb{R}^{nT} \mapsto [0,1]$ satisfying (8), (9) and (10), $\limsup_{T\to\infty} \int\int \varphi_T dF_T(m,\theta) dw(\theta) \leq \int\int \varphi_S^* dP(\theta) dw(\theta)$ for all $m \in \mathcal{M}_1$.

Unit Root Test Example, ctd: The function $\varphi_S^* : D_{[0,1]} \mapsto [0,1]$ is continuous at almost all realizations of W, so that part (i) of Theorem 1 shows that the test $\hat{\varphi}_T^*(Y_T) = \varphi_S^*(\hat{J}_T(\cdot)) = \mathbf{1}[\exp[-\frac{1}{2}c_1(\hat{J}_T(1)^2 - 1) - \frac{1}{2}c_1^2 \int \hat{J}_T(s)^2 ds] > \text{cv}]$ has asymptotic null rejection probability equal to the nominal level and asymptotic weighted average power equal to $P(L(J_{c_1}) > \text{cv})$ for all models in $\mathcal{M}_0$ and $\mathcal{M}_1$, respectively, that is models that satisfy $\hat{J}_T(\cdot) = T^{-1/2}\hat{\omega}_T^{-1}u_{T,\lfloor \cdot T \rfloor} \rightsquigarrow W(\cdot)$ and $\hat{J}_T(\cdot) \rightsquigarrow J_c(\cdot)$. Note that $\hat{\varphi}_T^*(Y_T)$ is asymptotically equivalent to the efficient unit root test statistic derived by ERS, so the contribution of part (i) of Theorem 1 for the unit root testing example is only to point out that $\hat{\varphi}_T^*$ has the same asymptotic properties under the null and alternative hypothesis for all models in $\mathcal{M}_0$ and $\mathcal{M}_1$, respectively.

The more interesting finding is part (ii) of Theorem 1: For any unit root test that has higher asymptotic power than $\hat{\varphi}_T^*$ for *any* model satisfying $\hat{J}_T(\cdot) \rightsquigarrow J_{c_1}(\cdot)$, there exists a model m where $\hat{J}_T(\cdot) \rightsquigarrow W(\cdot)$ for which the test has asymptotic null rejection probability larger than the nominal level. Any adaption to a non-Gaussian error distribution that leads to higher asymptotic power than ERS's test necessarily implies violation of the robustness condition (8). In other words, ERS's test is point-optimal in the class of all robust tests, i.e. test with asymptotic null rejection probability of at most α for all models that satisfy $\hat{J}_T(\cdot) \rightsquigarrow W(\cdot)$. ▲

The proof of part (i) of Theorem 1 follows from the definition of weak convergence, the continuous mapping theorem and dominated convergence. To gain some intuition for part (ii), consider the case where the hypotheses are simple, $\Theta_0 = \{\theta_0\}$ and $\Theta_1 = \{\theta_1\}$, and (9) and (10) do not bind. Let $L : S \mapsto \mathbb{R}$ be the Radon-Nikodym derivative of $P(\theta_1)$ with respect to $P(\theta_0)$, so that by the Neyman-Pearson Lemma, φ_S^* rejects for large values of L. For simplicity, assume that $L^i = 1/L$ is continuous and bounded. The central idea is to take the model $m \in \mathcal{M}_1$, and to reweigh the probabilities according to $L^i \circ h_T$ to construct a corresponding model in $\mathcal{M}_0$. This reweighed probability distribution needs to integrate to one, so let $\kappa_T = \int (L^i \circ h_T) dF_T(m, \theta_1) = \int L^i dP_T(m, \theta_1)$, and define the measure G_T on $\mathbb{R}^{nT}$ via $\int_A dG_T = \kappa_T^{-1} \int_A (L^i \circ h_T) dF_T(m, \theta_1)$ for all $A \in \mathfrak{B}(\mathbb{R}^{nT})$. By construction, under G_T, the function h_T induces the measure Q_T on S, where Q_T satisfies $\int \vartheta dQ_T = \kappa_T^{-1} \int \vartheta L^i dP_T(m, \theta_1)$ for any bounded and continuous function $\vartheta : S \mapsto \mathbb{R}$. Further, the $S \mapsto \mathbb{R}$ functions ϑL^i and L^i are bounded and continuous, so that $P_T(m, \theta_1) \rightsquigarrow P(\theta_1)$ implies $\kappa_T \to \int L^i dP(\theta_1) = \int L^i L dP(\theta_0) = \int dP(\theta_0) = 1$ and $\int \vartheta L^i dP_T(m, \theta_1) \to \int \vartheta L^i dP(\theta_1) = \int \vartheta dP(\theta_0)$, so that $h_T G_T \rightsquigarrow P(\theta_0)$. Thus, by (8), $\limsup_{T \to \infty} \int \varphi_T dG_T \leq \alpha$. Furthermore, by construction, the Radon-Nikodym derivative between G_T and $F_T(m, \theta_1)$ is given by $\kappa_T(L \circ h_T)$. Therefore, by the Neyman-Pearson Lemma, the best test of $\tilde{H}_0 : Y_T \sim G_T$ against $\tilde{H}_1 : Y_T \sim F_T(m, \theta_1)$ rejects for large values of $L \circ h_T$, and no test can have a better asymptotic level and power trade-off than this sequence of optimal tests. But $\hat{\varphi}_T^*$ also rejects for large values of $L \circ h_T$ and has the same asymptotic null rejection probability, and the result follows. The proof of Theorem 1 (ii) in the appendix deviates somewhat from this construction in order accommodate the additional constraints (9) and (10).

Comments

1. As already mentioned in the introduction, recall the standard approach for the derivation of asymptotically efficient and robust hypothesis tests in econometrics: Initially, restrict attention to the canonical parametric version of the model of interest, usually with Gaussian i.i.d. disturbances. Call this model m^*, so that $Y_T \sim F_T(m^*, \theta)$, and for simplicity, consider the problem of testing the simple hypotheses $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$. In this parametric model, $F_T(m^*, \theta_1)$ is absolutely continuous with respect to $F_T(m^*, \theta_0)$, and the small sample Likelihood Ratio statistic LR_T can be derived. The small sample optimal test in model m^* thus rejects for large values of LR_T. Express LR_T (up to asymptotically negligible terms) as a continuous function $L : S \mapsto \mathbb{R}$ of a random element $X_T^* = h_T^*(Y_T)$ that converges weakly under the null and contiguous alternative: $\mathrm{LR}_T = L(X_T^*) + o_p(1)$, where $X_T^* \rightsquigarrow X$ with $X \sim P(\theta)$. Thus, by the continuous mapping theorem, also the likelihood ratio statistic converges weakly under the null and alternative, $\mathrm{LR}_T \rightsquigarrow \mathrm{LR} \sim LP(\theta)$, and the asymptotic critical value is computed from the distribution $LP(\theta_0)$. Furthermore, an asymptotically robust test statistic is given by $L(X_T)$, where $X_T = h_T(Y_T)$ is a "robustified" version of X_T^* such that $X_T \rightsquigarrow X \sim P(\theta)$ in many models m of interest, including m^* (whereas typically, $X_T^* = h_T^*(Y_T) \not\rightsquigarrow X$ for some plausible models).

Unit Root Test Example, ctd: Under i.i.d. standard normal driving disturbances, the small sample efficient unit root test rejects for large values of $\mathrm{LR}_T = \exp[-\frac{1}{2}c_1 T^{-1}(u_{T,T}^2 - \sum_{t=1}^T (u_{T,t} - u_{T,t-1})^2) - \frac{1}{2}c_1^2 T^{-2} \sum_{t=1}^T u_{T,t-1}^2]$. With $h_T^*(Y_T) = X_T^* = T^{-1/2} u_{T,\lfloor \cdot T \rfloor} \rightsquigarrow J_c(\cdot)$, we thus have $\mathrm{LR}_T = L(X_T^*) + o_p(1)$ with $L(x) = \exp[-\frac{1}{2}c_1(x(1)^2 - 1) - \frac{1}{2}c_1^2 \int_0^1 x(s)^2 ds]$. The asymptotically robustified test that allows for serially correlated and non-Gaussian disturbances is based on $L(X_T)$, where $X_T = \hat{J}_T(\cdot) = T^{-1/2}\hat{\omega}_T^{-1} u_{T,\lfloor \cdot T \rfloor}$. ▲

The end product of this standard approach is a test based on the statistic $L(X_T)$, with critical value computed from the distribution $LP(\theta_0)$. Now generically, this test is identical to the test $\hat{\varphi}_T^*$ of Theorem 1. This follows from a general version of LeCam's Third Lemma (see, for instance, Lemma 27 of Pollard (2001)): If the models $Y_T \sim F_T(m^*, \theta_0)$ and $Y_T \sim F_T(m^*, \theta_1)$ are contiguous with likelihood ratio statistic LR_T, and under $Y_T \sim F_T(m^*, \theta_0)$,

$(\mathrm{LR}_T, X_T^*) \rightsquigarrow (L(X), X)$ with $X \sim P(\theta_0)$ and some function $L : S \mapsto \mathbb{R}$, then under $Y_T \sim F_T(m^*, \theta_1)$, $X_T^* \rightsquigarrow X \sim Q$, and the Radon-Nikodym derivative of Q with respect to $P(\theta_0)$ is equal to L. So if it is known that under $Y_T \sim F_T(m^*, \theta_1)$, $X_T^* \rightsquigarrow X \sim P(\theta_1)$, then it must be the case that $Q = P(\theta_1)$, and $L(X)$ is recognized as the Neyman-Pearson test statistic of the limiting problem $H_0^{lp} : \theta = \theta_0$ against $H_1^{lp} : \theta = \theta_1$ with $X \sim P(\theta)$ observed. The test that rejects for large values of $L(X_T)$ is thus simply the efficient test of this limiting problem, evaluated at sample analogues. This explains why in the unit root example the test $\hat{\varphi}_T^*$ of Theorem 1 had to be asymptotically equivalent to ERS's statistic.

2. This standard construction of tests starting from the canonical parametric model m^*, i.e. rejecting for large values of $L(X_T)$, is by construction asymptotically efficient in model m^*, and by part (i), it has the same asymptotic local power for all models $m \in \mathcal{M}_1$ for which $X_T \rightsquigarrow X \sim P(\theta_1)$. This does not, however, make the test $L(X_T)$ necessarily overall asymptotically efficient: It might be that there exists another test with the same asymptotic power in model m^*, and higher asymptotic power for at least some other models m for which $X_T \rightsquigarrow X \sim P(\theta_1)$. The semi-parametrically efficient unit root test by Jansson (2007) is an example of a test with the same asymptotic power as ERS's test for Gaussian i.i.d. disturbances, and higher asymptotic power for some other driving disturbances.

Now part (ii) of Theorem 1 shows that whenever a test has higher asymptotic power than $\hat{\varphi}_T^*$ for some alternative model $m \in \mathcal{M}_1$, then it cannot satisfy the robustness constraint (8). So any partial adaption to models $m \neq m^*$, if successful, necessarily implies the existence of a model $m \in \mathcal{M}_0$ for which the test has asymptotic rejection probability larger than the nominal level. So in particular, Theorem 1 implies the existence of a double array process $(u_{T,1}, \cdots, u_{T,T})'$ satisfying $T^{-1/2}\hat{\omega}_T^{-1}u_{T,\lfloor \cdot T \rfloor} \rightsquigarrow W(\cdot)$ for which Jansson's (2007) test has asymptotic rejection probability larger than the nominal level.

In other words, under the robustness constraint (8), Theorem 1 shows $\hat{\varphi}_T^*$ to be an overall asymptotically efficient test, because no test can exist with higher asymptotic (weighted average) power for *any* model $m \in \mathcal{M}_1$.

3. In this sense, Theorem 1 implies a particular version of an asymptotic essentially complete class result for the hypothesis test (1): Set π_0 in (9) equal to the power function of an admissible test in the limiting problem (4), so that (9) effectively determines φ_S^*. The theorem then shows that no test φ_T of (1) can be robust in the sense of (8) and have higher asymptotic local power than $\hat{\varphi}_T^*$ uniformly over $\theta \in \Theta_1$. As long as all admissible φ_S^*

are $P(\theta_0)$-almost everywhere continuous, the resulting tests $\hat{\varphi}_T^*$ thus form an "essentially complete class of asymptotically admissible robust tests" of the original problem (1).

In particular, if a uniformly most powerful test exists in the limiting problem, then φ_S^* is this test for any weighting function w, and repeated application of Theorem 1 with w having point mass for any $\theta \in \Theta_1$ then shows that the test $\hat{\varphi}_T^*$ is correspondingly asymptotically uniformly most powerful.

4. The appeal of the efficiency property of $\hat{\varphi}_T^*$ depends crucially on the appropriateness and desirability of the robustness constraint (8). One might think about the relative gain in robustness of tests satisfying (8) rather than the more standard "correct asymptotic null rejection probability for a wide range of primitive assumptions about disturbances that all imply (2)" in two ways.

On the one hand, one might genuinely worry that the true data generating process happens to be in the set of models that satisfy (2), but the disturbances do not satisfy the primitive conditions. Whenever tests with higher power exist under the more standard assumption, this set cannot be empty. This line of argument then faces the question whether such non-standard data generating processes are plausible. Especially in a time series context, primitive conditions are often quite opaque (could it be that interest rate are not mixing?), so it is not clear how and with what arguments one would discuss such a possibility. It is probably fair to say, however, that very general forms of sufficient primitive conditions for Central Limit Theorems and alike were derived precisely because researchers felt uncomfortable assuming more restricted (but still quite general) conditions, so one might say that imposing (8) constitutes only one more natural step in this progression of generality.

On the other hand, one might argue that the only purpose of an asymptotic analysis is to generate approximations for the small sample under study. In that perspective, it is irrelevant whether interest rates are indeed mixing or not, and the only interesting question becomes whether asymptotic properties derived under an assumption of mixing are useful approximations for the small sample under study. So even in an i.i.d. setting, one might be reluctant to rely on an adaptive test—not because it wouldn't be true that with a very large data set, the adaptive test would be excellent, but because asymptotics might be a poor guide to the behavior of the test in the sample under study. The robustness constraint (8) is then motivated by a concern that additional asymptotic implications of the primitive conditions beyond (2) are potentially poor approximations for the sample under study, and attempts to exploit them may lead to non-trivial size distortions.

Low Frequency Unit Root Test Example: Müller and Watson (2007) argue that in a macroeconomic context, it makes sense to take asymptotic implications of standard models of low frequency variability seriously only over frequencies below the business cycle. So in particular, when $u_{T,t}$ is modelled as local-to-unity, then the usual asymptotic implication is the functional convergence $T^{-1/2}u_{T,\lfloor \cdot T \rfloor} \rightsquigarrow \omega J_c(\cdot)$. Müller and Watson (2007) instead derive a scale invariant (we discuss invariance in Section 3.2) point-optimal unit root test that only assumes a subset of this convergence, that is

$$\left\{ T^{-3/2} \sum_{t=1}^{T} \psi_l(t/T) u_{T,t} \right\}_{l=1}^{q} \rightsquigarrow \left\{ \omega \int_0^1 \psi_l(s) J_\theta(s) ds \right\}_{l=1}^{q}, \tag{11}$$

where $\psi_l(s) = \sqrt{2}\cos(\pi l s)$ and q is chosen so that the frequency of the weight functions ψ_l, $l = 1, \cdots, q$ are below business cycle frequency for the span of the sample under study. The rationale is that picking q larger would implicitly imply a flat spectrum for $u_{T,t} - u_{T,t-1}$ in the I(1) model over business cycle frequencies, which is not an attractive assumption for macroeconomic data. So even if one were certain that for a long enough span of observations, the functional convergence $T^{-1/2}u_{T,\lfloor \cdot T \rfloor} \rightsquigarrow \omega J_c(\cdot)$ becomes a good approximation eventually, it does not seem well-advised to exploit its implications beyond (11) for the sample under study. ▲

5. Weak convergence statements of the form (2) and (3) can be viewed as a way of expressing regularity one is willing to impose on some inference problem. Implicitly, this is standard practice: invoking standard normal asymptotics for the OLS estimator of the largest autoregressive root ρ is formally justified for any value of $|\rho| < 1$, but effectively amounts to the assumption that the true parameter in the sample under study is not close to the local-to-unity region. Similarly, a choice of weak vs strong instrument asymptotics or local vs non-local time varying parameter asymptotics expresses knowledge of regularity in terms of weak convergences.

In some instances, it might be natural to express all regularity that one is willing to impose in this form, and Theorem 1 then shows that $\hat{\varphi}_T^*$ efficiently exploits this information. The i.i.d. nature of (standard) cross sectional data is not easily embedded in a weak convergence statement, so that such a starting point is much more convincing in a time series context. Also, interesting high level weak convergence assumptions are certainly not entirely arbitrary, but derive their plausibility from the knowledge that there exists a range of underlying primitive conditions that would imply them.

If one expresses regularity of a problem in terms of weak convergences, one faces a choice of what to assume. But not all weak convergences are relevant for deciding between H_0 and H_1: Whenever the Radon-Nikodym derivative in the limiting problem remains the same, then $\hat{\varphi}_T^*$ of Theorem 1 retains its optimality. This holds, for instance, for any additional convergence in probability to a constant limiting element (whose value does not depend on θ), but is more generally true if the conditional distribution of the additional limiting element is the same for all $\theta \in \Theta$. In the unit root test example, for instance, ERS's test remains asymptotically efficient in the sense of Theorem 1 if in addition to $\hat{J}_T(\cdot) \rightsquigarrow J_c(\cdot)$, the average sample kurtosis of $\Delta u_{T,t} = u_{T,t} - u_{T,t-1}$, $T^{-1} \sum_{t=1}^{T} (\Delta u_{T,t})^3$, is assumed to converge to zero in probability for any $c \geq 0$, or that $T^{-1/2} \sum_{t=1}^{\lfloor \cdot T \rfloor} (\Delta u_{T,t})^3 \rightsquigarrow W_\Delta(\cdot)$ for any $\theta \in \Theta$, where $W_\Delta(\cdot)$ is a Wiener process independent of J_c. Additional weak convergence restrictions of this type strengthen the efficiency claim by weakening the robustness requirement (8) of tests, since $\mathcal{M}_0$ becomes a smaller set, but, of course, the resulting efficiency correspondingly only holds for the smaller alternative set $\mathcal{M}_1$.

At the same time, one might also be reluctant to impose the full extent of the 'usual' weak convergence assumption, and in general, this leads to less powerful inference. The efficiency claim of Theorem 1 then shows that it is impossible to use data-dependent methods to improve inference for more regular data while still remaining robust in the sense of (8).

Low Frequency Unit Root Test Example, ctd: Since (11) is strictly weaker than the standard assumption $T^{-1/2} u_{T,\lfloor \cdot T \rfloor} \rightsquigarrow \omega J_c(\cdot)$, Müller and Watson's (2007) low-frequency unit root test is less powerful than a standard ERS test. It is nevertheless point-optimal in the sense of efficiently extracting all regularity contained in the weaker statement (11): Theorem 1 implies that it is impossible to let the data decide whether (11) holds for q larger than assumed (that is, whether the local-to-unity model provides good approximations also over business cycle frequencies), and to conduct more powerful inference if it is, without inducing size distortions for some model satisfying (11) for $c = 0$. ▲

6. No matter how one views the desirability of the robustness constraint (8), one appeal of Theorem 1 is that it suggests a general method for constructing reasonable tests. In nonlinear and/or dynamic models, it might be difficult to derive the small sample likelihood ratio statistic, even under strong parametric assumptions, while high level weak convergence properties might be easier to think about. The problem of testing for parameter instability in

a general GMM framework, as considered by Sowell (1996) and discussed in detail in Section 4.3 below, or the weak instrument problem in a general GMM framework, as considered by Stock and Wright (2000), arguably fall into this class. For such problems, efficient tests of the limiting problem, evaluated at sample analogues, are a natural starting point for sensible tests in the original problem.

3 Extensions

3.1 Consistently Estimable Nuisance Parameters

Suppose the testing problem (1) involves an additional nuisance parameter $\gamma \in \Gamma$, where Γ is a metric space, so that now $Y_T \sim F_T(m, \theta, \gamma)$, and the null and alternative hypotheses become

$$H_0 : (\theta, \gamma) \in (\Theta_0, \Gamma) \qquad \text{against} \qquad H_1 : (\theta, \gamma) \in (\Theta_1, \Gamma). \tag{12}$$

Suppose γ can be consistently estimated by the estimator $\hat{\gamma}_T$ under the null and alternative hypothesis. For a fixed value of $\gamma = \gamma_0$, that is when Γ is a singleton, this is a special case of what is covered by Theorem 1, as discussed in comment 5. But when Γ is not a singleton, the analysis above is not immediately applicable, because the limiting measures of X were assumed to be mutually absolutely continuous for all parameter values.

Thus denote by $P_T^e(m, \theta, \gamma)$ ('e' for extended) the distribution of $(\hat{\gamma}_T, X_T) = h_T^e(Y_T)$ when $Y_T \sim F_T(m, \theta, \gamma)$, so that the weak convergence assumption analogous to (2) and (3) now becomes

$$P_T^e(m, \theta, \gamma) \rightsquigarrow P^e(\theta, \gamma) \text{ pointwise for all } (\theta, \gamma) \in (\Theta \times \Gamma) \tag{13}$$

where $P^e(\theta, \gamma)$ is the product measure between the measure $P(\theta, \gamma)$ of X (which might depend on γ) on $\mathfrak{B}(S)$ and the degenerate probability measure on $\mathfrak{B}(\Gamma)$ that puts all mass on the point γ. The limiting problem now becomes testing (12) with $X \sim P(\theta, \gamma)$ observed and γ known, and optimal tests φ_S^{e*} in the limiting problem are indexed by $\gamma \in \Gamma$, $\varphi_S^{e*} : \Gamma \times S \mapsto [0, 1]$, so that for each $\gamma_0 \in \Gamma$, $\varphi_S^{e*}(\gamma_0, \cdot)$ is the weighted average power maximizing test of (12) for $\gamma = \gamma_0$, possibly subject to constraints of the form (6) and (7).

Now as long as the test φ_S^{e*} is, for each $\gamma \in \Gamma$, $P^e(\gamma, \theta_0)$-almost everywhere continuous, the same arguments as employed in the proof of Theorem 1 part (i) imply that under (13) and for each $\gamma \in \Gamma$, the test $\hat{\varphi}_T^{e*} = \varphi_S^{e*}(\hat{\gamma}_T, X_T)$ has the same asymptotic size and power as φ_S^{e*} does. Furthermore, for any value of $\gamma = \gamma_0$, one can invoke Theorem 1 (ii) as above to conclude that no test $\hat{\varphi}_T$ can exist with better asymptotic weighted average power, for any model satisfying $P_T^e(m, \theta, \gamma_0) \rightsquigarrow P^e(\theta, \gamma_0)$. Problems that involve consistently estimable parameters γ with an impact on the efficient limiting tests are thus covered by the results of Theorem 1 under an additional assumption of the family of efficient limiting tests, indexed by γ, to depend on γ sufficiently smoothly.

Unit Root Test Example, ctd: Instead of $T^{-1/2}\hat{\omega}_T^{-1}u_{T,\lfloor \cdot T\rfloor} = \hat{J}_T(\cdot) \rightsquigarrow J_c(\cdot)$, consider the weak convergences $h_T^e(Y_T) = (\hat{\omega}_T^2, T^{-1/2}u_{T,\lfloor \cdot T\rfloor}) \rightsquigarrow (\omega^2, \omega J_c(\cdot))$ as a starting point. In the limiting problem, $X = \omega J_c(\cdot)$ is observed with ω^2 known, and the point-optimal test is of the form $\varphi_S^{e*}(\omega^2, X) = \mathbf{1}[\exp[-\frac{1}{2}c_1(\omega^{-2}X(1)^2 - 1) - \frac{1}{2}c_1^2\omega^{-2}\int_0^1 X(s)^2 ds] > \text{cv}]$. A calculation shows this to be a continuous function $(0, \infty) \times D_{[0,1]} \mapsto \mathbb{R}$ for almost all realizations of (ω^2, J_0), so the test $\varphi_S^{e*}(\hat{\omega}_T^2, T^{-1/2}u_{T,\lfloor \cdot T\rfloor})$ is asymptotically efficient among all unit root tests with correct asymptotic null rejection probability whenever $(\hat{\omega}_T^2, T^{-1/2}u_{T,\lfloor \cdot T\rfloor}) \rightsquigarrow (\omega^2, \omega J_0(\cdot))$ against all models satisfying $(\hat{\omega}_T^2, T^{-1/2}u_{T,\lfloor \cdot T\rfloor}) \rightsquigarrow (\omega^2, \omega J_{c_1}(\cdot))$. ▲

3.2 Invariance

The majority of efficient tests for nonstandard problems cited in the introduction rely on invariance considerations. In the framework here, invariance may be invoked at two levels: On the one hand, one might consider a weak convergence as a starting point that is a function of a small sample maximal invariant. On the other hand, invariance might instead be employed in the limiting problem as a way of dealing with nuisance parameters. This subsection discusses the link between these two notions, and the interaction of the concept of invariance with the efficiency statements of Theorem 1.

The first case is entirely straightforward: suppose $\phi_T(Y_T)$ with $\phi_T : \mathbb{R}^{nT} \mapsto \mathbb{R}^{nT}$ is a maximal invariant to some group of transformations. By Theorem 1 on page 285 of Lehmann (1968), all invariant tests can be written as functions of a maximal invariant. So if h_T of Section 2.1 is of the form $h_T = h_T^{\phi} \circ \phi_T$, then Theorem 1 applies and yields an asymptotic efficiency statement among all invariant tests relative to a robustness constraint (8) in terms of the weak convergence of a function of the small sample maximal invariant.

Unit Root Test Example, ctd: Consider the problem of testing for a unit root in a model with unknown mean, and suppose $\omega = 1$ is known for simplicity. A maximal invariant is given by the demeaned data $\{\hat{u}_{T,t}\}_{t=1}^T$ with $\hat{u}_{T,t} = y_{T,t} - \bar{y}_T$, where $\bar{y}_T = T^{-1}\sum_{t=1}^T y_{T,t}$ and $Y_T = (y_{T,1}, \cdots, y_{T,T})'$. The typical model satisfies $T^{-1/2}\hat{u}_{T,\lfloor \cdot T\rfloor} \rightsquigarrow J_c^\mu(\cdot)$, where $J_c^\mu(s) = J_c(s) - \int_0^1 J_c(l)dl$. Theorem 1 now shows that rejecting for large values of $L^\mu(T^{-1/2}\hat{u}_{T,\lfloor \cdot T\rfloor})$ is asymptotically point-optimal among all tests whose asymptotic null rejection probability is at most α whenever $T^{-1/2}\hat{u}_{T,\lfloor \cdot T\rfloor} \rightsquigarrow J_c^\mu(\cdot)$, where L^μ is the Radon-Nikodym derivative of the probability measure of $J_{c_1}^\mu$ with respect to the measure of J_0^μ.[2] We conclude that rejecting for large values of $L^\mu(T^{-1/2}\hat{u}_{T,\lfloor \cdot T\rfloor})$ is the asymptotically point-optimal translation invariant test among all tests that do not overreject asymptotically whenever $T^{-1/2}\hat{u}_{T,\lfloor \cdot T\rfloor} \rightsquigarrow J_0^\mu(\cdot)$. ▲

In the second case, one considers the typical weak convergence in a model with nuisance parameters, and applies invariance only in the limiting problem. Formally, let $\tilde{g} : R \times S \mapsto S$ be such that the $S \mapsto S$ functions $x \mapsto \tilde{g}(r, x)$, indexed by $r \in R$, form the group $\tilde{\mathcal{G}}$. Further, suppose $\tilde{\phi} : S \mapsto S$ is a maximal invariant to $\tilde{\mathcal{G}}$, so that the efficient invariant test in the limiting problem $\varphi_S^{\phi *}$ is the efficient test of $H_0^{\phi,lp} : \Xi \sim \tilde{\phi}P(\theta)$, $\theta \in \Theta_0$ against $H_1^{\phi,lp} : \Xi \sim \tilde{\phi}P(\theta)$, $\theta \in \Theta_1$, where $\Xi = \tilde{\phi}(X)$. It is not clear whether or in which sense this test, evaluated at sample analogues would be asymptotically efficient.

Unit Root Test Example, ctd: In the parametrization $y_{T,t} = u_{T,t} + T^{1/2}\alpha_y$, we obtain with $\theta = (c, \alpha_y)'$ that $T^{-1/2}y_{T,\lfloor \cdot T\rfloor} \rightsquigarrow J_\theta^y(\cdot)$, where $J_\theta^y(s) = J_c(s) + \alpha_y$. The parameter α_y is a nuisance parameters in the limiting problem. Define the transformations $\tilde{g} : \mathbb{R} \times D_{[0,1]} \mapsto D_{[0,1]}$ as $\tilde{g}(r, x) = x(\cdot) + r$, with $r \in R = \mathbb{R}$. The limiting problem is seen to be invariant to these transformations, and $\tilde{\phi} : D_{[0,1]} \mapsto D_{[0,1]}$ with $\tilde{\phi}(x) = x(\cdot) - \int_0^1 x(l)dl$, is a maximal invariant. Since $\tilde{\phi}(J_\theta^y) \sim J_c^\mu(\cdot)$, the point-optimal invariant test in the limiting problem rejects for large values of $L^\mu(J_\theta^y)$. This test, evaluated at sample analogues, yields $L^\mu(T^{-1/2}\hat{u}_{T,\lfloor \cdot T\rfloor})$, just as above.

Even though in this example, the efficient invariant test in the limiting problem, evaluated at sample analogues, is small sample invariant, one still cannot claim this test to be the asymptotically point-optimal test among all small sample invariant tests that are robust whenever $T^{-1/2}y_{T,\lfloor \cdot T\rfloor} \rightsquigarrow J_c(\cdot) + \alpha_y$ with $\alpha_y \in \mathbb{R}$. The reason is that the set of

[2] Under the assumption of $u_{T,0} = 0$ for the initial condition, $L^\mu(x) = L(x(\cdot) - x(0))$, where L is the Radon-Nikodym derivative of the probability measure of J_{θ_1} with respect to the measure of J_0, so that rejecting for large values of $L^\mu(T^{-1/2}\hat{u}_{T,\lfloor \cdot T\rfloor})$ leads to the same asymptotic power as without translation invariance. This equivalence does not hold, however, when the initial condition $u_{T,0}$ is of the same order of magnitude as $u_{T,\lfloor sT\rfloor}$ for $s > 0$. See Müller and Elliott (2003) for discussion.

models satisfying $T^{-1/2}y_{T,\lfloor \cdot T\rfloor} \rightsquigarrow J_c(\cdot)+\alpha_y$ is a proper subset of the set of models satisfying $T^{-1/2}\hat{u}_{T,\lfloor \cdot T\rfloor} \rightsquigarrow J_\theta^\mu(\cdot)$. The efficiency of $L^\mu(T^{-1/2}\hat{u}_{T,\lfloor \cdot T\rfloor})$ noted above is thus relative to a more stringent robustness constraint, and it remains unclear whether an efficiency claim can also be made relative to the weaker constraint of asymptotic size control whenever $T^{-1/2}y_{T,\lfloor \cdot T\rfloor} \rightsquigarrow J_0(\cdot)+\alpha_y$. ▲

We now show how one can make an asymptotic efficiency claim when invariance is employed in the limiting problem by relating the limiting group of transformations to a sequence of small sample groups. So for each T, suppose the measurable function $g_T : R\times\mathbb{R}^{nT} \mapsto \mathbb{R}^{nT}$ is such that the $\mathbb{R}^{nT}\mapsto\mathbb{R}^{nT}$ functions $y\mapsto g_T(r,y)$, indexed by $r\in R$, form the group $\mathcal{G}_T$. Let $\phi_T:\mathbb{R}^{nT}\mapsto\mathbb{R}^{nT}$ be a maximal invariant of $\mathcal{G}_T$. Assume that the small sample and limiting invariance correspond in the sense that the small sample maximal invariant converges weakly to the maximal invariant of the limiting problem, i.e.

$$(h_T\circ\phi_T)F_T(m,\theta) \quad\rightsquigarrow\quad \tilde{\phi}P(\theta) \quad \text{pointwise for } \theta\in\Theta_0 \tag{14}$$

$$(h_T\circ\phi_T)F_T(m,\theta) \quad\rightsquigarrow\quad \tilde{\phi}P(\theta) \quad \text{pointwise for } \theta\in\Theta_1. \tag{15}$$

Let $\mathcal{M}_0^\phi$ and $\mathcal{M}_1^\phi$ the set of models m satisfying (14) and (15), respectively. Since $\varphi_S^{\phi *}$ was assumed to be the efficient test of $H_0^{\phi,lp}:\Xi\sim\tilde{\phi}P(\theta)$, $\theta\in\Theta_0$ against $H_1^{\phi,lp}:\Xi\sim\tilde{\phi}P(\theta)$, $\theta\in\Theta_1$, one can apply Theorem 1 to conclude that $\varphi_S^{\phi *}\circ h_T\circ\phi_T$ is the asymptotically efficient small sample invariant (with respect to $\mathcal{G}_T$) robust test *relative to the weak convergence assumptions described by* $\mathcal{M}_0^\phi$ *and* $\mathcal{M}_1^\phi$. As noted in the unit root example above, however, one cannot conclude that $\varphi_S^{\phi *}\circ h_T\circ\phi_T$ is also the asymptotically efficient small sample invariant robust test relative to the weak convergence $h_TF_T(m,\theta)\rightsquigarrow P(\theta)$, i.e. relative to the (typically strictly smaller) sets $\mathcal{M}_0$ and $\mathcal{M}_1$.

It would be possible to draw this additional conclusion if for small sample invariant tests, constraints with respect to models in $\mathcal{M}_j$ are implied by the analogous constraints with respect to $\mathcal{M}_j^\phi$, $j=0,1$: For each model $m\in\mathcal{M}_j^\phi$, there must exist a corresponding model in $\mathcal{M}_j$ that is identical up to a transformation in $\mathcal{G}_T$. The following Theorem provides conditions under which this is the case.

Theorem 2 *Suppose (i) R is a separable and complete metric space; (ii) the mapping $x\mapsto\tilde{g}(r,x)$ is continuous for all $r\in R$; (iii) there exists a measurable function $\tilde{\rho}:S\mapsto R$ such that $x=\tilde{g}(\tilde{\rho}(x),\tilde{\phi}(x))$ for all $x\in S$; (iv)* $\sup_{r\in R,y\in\mathbb{R}^{nT}} d_S(h_T(g_T(r,y)),\tilde{g}(r,h_T(y)))\to 0$, *where d_S is the metric on S; (v) $\tilde{\phi}$ and ϕ_T select specific orbits, i.e. for all $y\in\mathbb{R}^{nT}$ and*

$x \in S$, there exist $r_y, r_x \in R$ so that $\phi_T(y) = g_T(r_y, y)$ and $\tilde{\phi}(x) = \tilde{g}(r_x, x)$. Then for any $m \in \mathcal{M}_j^\phi$ and $\theta \in \Theta_j$, $j = 0, 1$, there exists a sequence of measures G_T on $\mathfrak{B}(\mathbb{R}^{nT})$ such that $\phi_T F_T(m, \theta) = \phi_T G_T$ and $h_T G_T \rightsquigarrow P(\theta)$.

Unit Root Test Example, ctd: With $g_T(r, Y_T) = (y_{T,1}+rT^{1/2}, \cdots, y_{T,T}+rT^{1/2})'$, $\phi_T(Y_T) = g_T(-\bar{y}_{T,}Y_T)$, and $\tilde{\rho}(x) = \int_0^1 x(s)ds$, we find $(h_T \circ \phi_T)(Y_T) = T^{-1/2}\hat{u}_{T,\lfloor \cdot T \rfloor} \rightsquigarrow J_c^\mu \sim \tilde{\phi}(J_c)$, $\sup_{y \in \mathbb{R}^T} d_{D_{[0,1]}}(T^{-1/2}(y_{\lfloor sT \rfloor} + T^{1/2}r) - (T^{-1/2}y_{\lfloor sT \rfloor} + r)) = 0$, and $x = \tilde{\phi}(x) + \tilde{\rho}(x) = \tilde{g}(\tilde{\rho}(x), \tilde{\phi}(x))$ for all $x \in D_{[0,1]}$, so that the assumptions of Theorem 2 hold. We can therefore conclude that rejecting for large values of $L^\mu(T^{-1/2}\hat{u}_{T,\lfloor \cdot T \rfloor})$ is also the asymptotically point-optimal test among all translation invariant tests with asymptotic null rejection probability of at most α whenever $T^{1/2}y_{T,\lfloor \cdot T \rfloor} \rightsquigarrow J_0(\cdot) + \alpha_y$, $\alpha_y \in \mathbb{R}$. ▲

For the proof of Theorem 2, note that with $x = \tilde{g}(\tilde{\rho}(x), \tilde{\phi}(x))$ for all $x \in S$ and $\theta \in \Theta$, one can construct the distribution $P(\theta)$ by applying an appropriate *random* transformation $\tilde{g}$ to each x drawn under $\tilde{\phi}P(\theta)$—in the unit root example, the appropriate r is distributed as $J_c^\mu(0)$, since $J_c(\cdot) = J_c^\mu(\cdot) - J_c^\mu(0)$ a.s. The assumptions of Theorem 2 are now sufficient to ensure a tight enough link between this construction for $P(\theta)$ and the limit of the small sample analogously transformed $\mathcal{M}_j^\phi$ models. For each model in $\mathcal{M}_j^\phi$, one can thus construct a corresponding model in $\mathcal{M}_j$ by applying an appropriate random transformation $g_T \in \mathcal{G}_T$ for each T.

While asymptotic efficiency statements about invariant tests based on Theorems 1 and 2 require a tight link between the limiting group $\tilde{\mathcal{G}}$ and the small sample groups $\mathcal{G}_T$, the link does not need to be perfect: Even if the distribution of the limiting maximal invariant $\tilde{\phi}P(\theta)$ does not depend a subset of the parameter θ (so that a nuisance parameter is eliminated by invariance), it is not assumed that the small sample counterpart $\phi_T F_T(m, \theta)$ shares this feature. Also assumption (iv) does not require the small sample and limiting group actions to exactly coincide, which is important for, say, arguing for the asymptotic efficiency of translation and trend invariant unit root tests with respect to the weak convergence $T^{-1/2}y_{T,\lfloor \cdot T \rfloor} \rightsquigarrow J_c(\cdot) + \alpha_y + \cdot\beta_y$.

3.3 Uniformity

The discussion so far concerned the pointwise asymptotic properties of tests φ_T, i.e. the rejection probability as $T \to \infty$ for a fixed model m and parameter value θ. This is standard practice in much of econometric theory, including in the literature on semiparametrically efficient tests, such as Jansson (2007).

When one insists on uniform results in the set-up here, it is natural to ask whether the pointwise robustness constraint (8) can be replaced by a uniform constraint over models, that is to demand that for large enough T, the null rejection probability $\int \varphi_T dF_T(m,\theta)$ for $\theta \in \Theta_0$ to be close to α for all data generating processes under consideration. It is not hard to see, however, that with $\mathcal{M}_0$ the set of all models satisfying (2), such a uniformity cannot hold for non-trivial tests: for any T, the distribution $F_T(m,\theta)$ of Y_T is entirely unrestricted, as the convergence $P_T(m,\theta) = h_T F_T(m,\theta) \rightsquigarrow P(\theta)$ can occur 'later'.

To generate uniform results, one must therefore reduce the set of models $\mathcal{M}_0$ and impose a lower limit on the speed of convergence. For two probability measures μ_1 and μ_2 on $\mathfrak{B}(S)$, define Δ_{BL} as $\Delta_{BL}(\mu_1,\mu_2) = \sup_{||f||_{BL}\leq 1} |\int f d\mu_1 - \int f d\mu_2|$, where $f : S \mapsto \mathbb{R}$ are $\mathfrak{B}(S)\backslash\mathfrak{B}(\mathbb{R})$ measurable and $||f||_{BL} = \sup_{x\in S}|f(x)| + \sup_{x,y\in S} \frac{|f(x)-f(y)|}{d_U(x,y)}$. It is known that Δ_{BL} metrizes weak convergence on separable metric spaces (Dudley (2002, p. 395)). Also, let the real sequence δ_T be such that $\delta_T \to 0$. Now define $\mathcal{M}_0^u(\delta)$ ('u' for uniform) as the set of models m satisfying

$$\sup_{\theta\in\Theta_0} \Delta_{BL}(P_T(m,\theta), P(\theta)) \leq \delta_T,$$

that is $\mathcal{M}_0^u(\delta)$ is the collection of models m for which the distribution $P_T(m,\theta) = h_T F_T(m,\theta)$ of $h_T(Y_T)$ differs by at most δ_T from its limit $P(\theta)$ as measured by Δ_{BL}, uniformly over Θ_0. It then makes sense to ask whether the rejection probability of a test φ_T converges to the nominal level uniformly over $\theta \in \Theta_0$ and $\mathcal{M}_0^u(\delta)$, that is if

$$\limsup_{T\to\infty} \sup_{\theta\in\Theta_0, m\in\mathcal{M}_0^u(\delta)} \int \varphi_T dF_T(m,\theta) \leq \alpha. \tag{16}$$

By the continuity of φ_S^*, (16) holds for the test $\varphi_T = \hat{\varphi}_T^*$ in Theorem 1, i.e. for large enough T, the rejection probability of $\hat{\varphi}_T^*$ is close to α for all models in $\mathcal{M}_0^u(\delta)$, uniformly over Θ_0.[3] Similar restrictions and arguments could be made regarding the constraints (9) and (10).

It is not clear, however, whether all tests that satisfy the point-wise robustness (8) also satisfy (16), or vice versa. Theorem 1 therefore does not imply that $\hat{\varphi}_T^*$ also maximizes

[3] Construct a decreasing sequence of functions ℓ_T that converge to φ_S^* pointwise $P(\theta_0)$-almost everywhere with $\delta_T||\ell_T||_{BL} \to 0$, as in chapter 7.1 of Pollard (2002). Then $\sup_{\theta\in\Theta_0, m\in\mathcal{M}_0^u(\delta)} \int \hat{\varphi}_T^* dF_T(m,\theta) - \alpha \leq \sup_{\theta\in\Theta_0, m\in\mathcal{M}_0^u(\delta)} \int \ell_T(dP_T(m,\theta) - dP(\theta)) \leq ||\ell_T||_{BL}\Delta_{BL}(P_T(m,\theta), P(\theta)) \to 0$, where the last inequality uses $||f_1 \cdot f_2||_{BL} \leq ||f_1||_{BL} \cdot ||f_2||_{BL}$.

asymptotic weighted average power in the class of all tests that satisfy (16). A partial result in that regard is provided by the following Theorem for the case of a single null hypothesis and without the constraints (9) and (10).

Theorem 3 *Suppose* $\Theta_0 = \{\theta_0\}$*, let* $L : \Theta \times S \mapsto \mathbb{R}$ *be the Radon-Nikodym derivative of* $P(\theta)$ *with respect to* $P(\theta_0)$*, define* $\bar{L}(x) = \int L(\theta, x) dw(\theta)$*, let* $\varphi_S^* : S \mapsto [0, 1]$ *be the level* α *Neyman-Pearson test that rejects for large values of* $\bar{L}$*, and define* $\hat{\varphi}_T^* = \varphi_S^* \circ h_T$*. Suppose that for all* $\varepsilon > 0$ *there exists an open set* $D_\varepsilon \in \mathfrak{B}(S)$ *with* $\int_{D_\varepsilon} dP(\theta_0) > 1 - \varepsilon$ *so that the* $D_\varepsilon \mapsto \mathbb{R}$ *function* $x \mapsto \bar{L}(x)$ *is Lipschitz, and assume that the models* m_0 *and* m_1 *are such that* $\Delta_{BL}(P_T(m_0, \theta_0), P(\theta_0))/\delta_T \to 0$ *and* $\Delta_{BL}(\int P_T(m_1, \theta) dw(\theta), \int P(\theta) dw(\theta))/\delta_T \to 0$*. Then for any test* φ_T *that satisfies (16),* $\limsup_{T \to \infty} \int \int \varphi_T dF_T(m_1, \theta) dw(\theta) \leq \lim_{T \to \infty} \int \int \hat{\varphi}_T^* dF_T(m_1, \theta) dw(\theta) = \int \int \varphi_S^* dP(\theta) dw(\theta)$*.*

Under a stronger continuity assumption on the limiting problem, Theorem 3 shows that no test can satisfy (16) and have higher asymptotic weighted average power than $\hat{\varphi}_T^*$ for alternative models whose (average) weak convergence is faster than the lower bound δ_T. In other words, $\hat{\varphi}_T^*$ is the efficient test in the class of test satisfying (16) against any set of alternative models in which $\int P_T(m_1, \theta) dw(\theta) \rightsquigarrow \int P(\theta) dw(\theta)$ converges faster than the slowest convergence $P_T(m_1, \theta_0) \rightsquigarrow P(\theta_0)$ for which the test controls asymptotic size uniformly.

The proof of Theorem 3 follows closely the heuristic sketch of the proof Theorem 1 outlined above and exploits the linearity in both the definition of Δ_{BL} and the reweighed probability assignments. Dudley (2002, p. 411) shows that the Prohorov metric Δ_P (which also metrizes weak convergence) satisfies $\Delta_P \leq 2\Delta_{BL}^{1/2}$ and $\Delta_{BL} \leq 2\Delta_P$, so that Theorem 3 could equivalently be formulated in terms of Δ_P.

4 Applications

4.1 Unit Root Tests with Stationary Covariates

Elliott and Jansson (2003) consider the model

$$\begin{pmatrix} y_{T,t} \\ x_{T,t} \end{pmatrix} = \begin{pmatrix} \alpha_y + \beta_y t \\ \alpha_x + \beta_x t \end{pmatrix} + \begin{pmatrix} u_{T,t} \\ \nu_{T,t}^x \end{pmatrix} \tag{17}$$

where $Y_T = ((y_{T,1}, x'_{T,1})', \cdots, (y_{T,T}, x'_{T,T}))' \in \mathbb{R}^{nT}$ is observed, α_y, β_y and $u_{T,t} = \rho_T u_{T,t-1} + \nu^y_{T,t}$ are scalars, $u_{T,0} = O_p(1)$, $\rho_T = 1 - c/T$ for some fixed $c \geq 0$, and $x_{T,t}$, α_x, β_x and $\nu^x_{T,t}$ are $n-1$ dimensional vectors. The objective is to efficiently exploit the stationary covariates $x_{T,t}$ in the construction of a test of the null hypothesis of a unit root in $y_{T,t}$, $H_0 : c = 0$ against the alternative $H_1 : c > 0$. Consider first the case with $\alpha_y = \alpha_x = \beta_y = \beta_x = 0$ known. The approach of Elliott and Jansson (2003) is to first apply the Neyman-Pearson Lemma to determine, for each T, the point-optimal test against $c = c_1$ when $\nu_{T,t} = (\nu^y_{T,t}, \nu^{x\prime}_{T,t})' \sim$i.i.d.$\mathcal{N}(0, \Omega)$ for known Ω. In a second step, they construct a feasible test that is (i) asymptotically equivalent the point-optimal test when $\nu_{T,t} \sim$i.i.d.$\mathcal{N}(0, \Omega)$ and (ii) that is robust to a range of autocorrelation structures and error distributions. So by construction, their test can only claim efficiency for the special case of i.i.d. Gaussian disturbances.

In order to apply the results in Sections 2 and 3 of this paper, we consider the typical weak convergence properties of model (17). Standard weak dependence assumptions on $\nu_{T,t}$ imply for some suitable long-run covariance matrix estimator $\hat{\Omega}_T$

$$\hat{\Omega}_T \xrightarrow{p} \Omega \text{ and } G_T(\cdot) = \begin{pmatrix} T^{-1/2} u_{T,\lfloor \cdot T \rfloor} \\ T^{-1/2} \sum_{t=1}^{\lfloor \cdot T \rfloor} \nu^x_{T,t} \end{pmatrix} \rightsquigarrow G(\cdot) \tag{18}$$

where Ω is positive definite, $G(s) = \int_0^s \operatorname{diag}(e^{-c(s-r)}, 1, \cdots, 1)\Omega^{1/2} dW(r)$, and W is a $n \times 1$ standard Wiener process. By Girsanov's Theorem, the Radon-Nikodym derivative of the distribution of G with $c = c_1$ with respect to the distribution of G with $c = 0$, evaluated at $G = (G_y, G'_x)'$, is given by

$$\begin{aligned} L(\Omega, G) &= \exp\left[-c_1 \int_0^1 G(s)' S_1 \Omega^{-1} dG(s) - \tfrac{1}{2} c_1^2 \int_0^1 G(s)' S_1 \Omega^{-1} S_1 G(s) ds\right] \\ &= \exp\left[-\tfrac{1}{2} c_1 (\omega_{yy}^{-2} G_y(1)^2 - 1) - c_1 \omega_{yx} \int_0^1 G_y(s) dG_x(s) - \tfrac{1}{2} c_1^2 \omega_{yy}^{-2} \int_0^1 G_y(s)^2 ds\right] \end{aligned} \tag{19}$$

where S_1 is the $n \times n$ matrix $S_1 = \operatorname{diag}(1, 0, \cdots, 0)$ and the first row of Ω^{-1} is $(\omega^2_{yy}, \omega_{yx})$. By the Neyman-Pearson Lemma, the point-optimal test in the limiting problem rejects for large values of $L(\Omega, G)$. Since $\int_0^1 G_y(s) dG_x(s)$ is not a continuous mapping, we cannot directly apply Theorem 1. However, typical weak dependence assumptions on $\nu_{T,t}$ also imply (see, for instance, Phillips (1988), Hansen (1990) and de Jong and Davidson (2000)) that

$$\Upsilon_T = T^{-1} \sum_{t=2}^{T} u_{T,t-1} \nu^x_{T,t} - \hat{\Sigma}_T \rightsquigarrow \Upsilon = \int_0^1 G_y(s) dG_x(s) \tag{20}$$

for a suitably defined $(n-1)\times 1$ vector $\hat{\Sigma}_T \xrightarrow{p} \Sigma$ (which equals $\sum_{s=1}^{\infty} E[\nu^x_{T,t}\nu^y_{T,t+s}]$ when $\nu_{T,t}$ is covariance stationary) jointly with (18). Clearly, the Radon-Nikodym derivative of the measure of (G,Υ) for $c=c_1$ with respect to the measure of (G,Υ) with $c=0$, evaluated at G, is also given by $L(\Omega,G)$ in (19), and one can write $L(\Omega,G) = L^{\Upsilon}(\Omega,G,\Upsilon)$ for a continuous function L^{Υ}. The discussion of Section 3.1 thus applies, and Theorem 1 shows that rejecting for large values of $L^{\Upsilon}(\hat{\Omega}_T,G_T,\Upsilon_T)$ is the point-optimal unit root test for the alternative $c=c_1$ among all tests that have correct asymptotic null rejection probabilities whenever (18) and (20) hold.

Since the model with $\nu_{T,t}$ ∼i.i.d.$\mathcal{N}(0,\Omega)$ satisfies (18) and (20), the test derived by Elliott and Jansson (2003) is by construction—as explained in Comment 1 of Theorem 1—asymptotically equivalent to a test that rejects for large values of $L^{\Upsilon}(\hat{\Omega}_T,G_T,\Upsilon_T)$. The derivation here, which starts with the Radon-Nikodym derivative directly, is arguably a more straightforward way of determining a test in this equivalence class. But the more important insight concerns the optimality properties of this test: While Elliott and Jansson (2003) could only claim optimality for the model with i.i.d. Gaussian disturbances, Theorem 1 shows that the test to be efficient against *all* alternatives satisfying (18) and (20) with $c=c_1$ if one imposes size control for all models satisfying (18) and (20) with $c=0$. In other words, under this robustness constraint, no test exists with higher asymptotic power for any disturbance distribution or autocorrelation structure satisfying (18) and (20) with $c=c_1$.

When the deterministic terms are not fully known, i.e. the parameters α_y, α_x, β_y, and/or β_x are not known, it is natural to impose an appropriate invariance requirement. Specifically, considering the case where α_y and α_x are unconstrained and $\beta_y=\beta_x=0$, one might impose invariance to the transformations

$$\{(y_{T,t},x'_{T,t})'\}_{t=1}^T \to \{(y_{T,t}+a_y,x'_{T,t}+a'_x)'\}_{t=1}^T \quad a_y\in\mathbb{R},\, a_x\in\mathbb{R}^{n-1}. \tag{21}$$

A maximal invariant of this group of transformations is given by the demeaned data $\{(\hat{y}_{T,t},\hat{x}'_{T,t})'\}_{t=1}^T$, where $\hat{y}_{T,t}=y_{T,t}-\bar{y}_T$, $\hat{x}_{T,t}=x_{T,t}-\bar{x}_T$, $\bar{y}_T=T^{-1}\sum_{t=1}^T y_{T,t}$ and $\bar{x}_T=T^{-1}\sum_{t=1}^T x_{T,t}$. Elliott and Jansson (2003) derive the limiting behavior of the likelihood ratio statistics of this maximal invariant when $\nu_{T,t}$ ∼i.i.d.$\mathcal{N}(0,\Omega)$, and thus obtain the asymptotically point-optimal invariant unit root test under that assumption. Considering again the weak convergence properties of a typical model, we obtain

$$\hat{\Omega}_T \xrightarrow{p} \Omega \quad \text{and} \quad \begin{pmatrix} T^{-1/2}\hat{y}_{T,\lfloor \cdot T \rfloor} \\ T^{-1/2}\sum_{t=1}^{\lfloor \cdot T \rfloor} \hat{x}_{T,t} \\ T^{-1/2}\sum_{t=2}^{T} \hat{y}_{T,t-1}\hat{x}_{T,t} - \hat{\Sigma}_T \end{pmatrix} \rightsquigarrow \begin{pmatrix} \hat{G}_y(\cdot) \\ \hat{G}_x(\cdot) \\ \int_0^1 \hat{G}_y(s)d\hat{G}_x(s) \end{pmatrix} \tag{22}$$

where $\hat{G}_y(s) = G_y(s) - \int_0^1 G_y(l)dl$ and $\hat{G}_x(s) = G_x(s) - sG_x(1)$. For brevity, we omit an explicit expression for the Radon-Nikodym derivative $L^{\hat{G}}$ of the measure of $(\hat{G}_y, \hat{G}_x)$ with $c = c_1$ with respect to the measure of $(\hat{G}_y, \hat{G}_x)$ when $c = 0$. By the Neyman-Pearson Lemma and Theorem 1, rejecting for large values of $L^{\hat{G}}$, evaluated at sample analogues, is the asymptotically point-optimal robust test for models defined via (22). Furthermore, in the notation of Section 3.2, with $h_T(Y_T) = (T^{-1/2}y_{T,\lfloor \cdot T \rfloor}, T^{-1/2}\sum_{t=1}^{\lfloor \cdot T \rfloor} x_{T,t}, T^{-1}\sum_{t=2}^{T} y_{T,t-1}x_{T,t} - \hat{\Sigma}_T) \in D_{[0,1]} \times D_{[0,1]}^{n-1} \times \mathbb{R}^{n-1}$, $r = (r_y, r_x')' \in \mathbb{R} \times \mathbb{R}^{n-1}$, $g_T(r, \{(y_{T,t}, x_{T,t}')'\}_{t=1}^T) = \{(y_{T,t} + T^{1/2}r_y, x_{T,t}' + T^{-1/2}r_x')'\}_{t=1}^T$, $\tilde{g}(r, (y, x, z)) = (y(\cdot) + r_y, x(\cdot) + r_x\cdot, z + r_y(x(1) - x(0)) + r_x\int_0^1 y(s)ds)$, $\phi_T(\{(y_{T,t}, x_{T,t}')'\}_{t=1}^T) = \{(\hat{y}_{T,t}, \hat{x}_{T,t}')'\}_{t=1}^T$ and $\tilde{\phi}(y, x, z) = (y(\cdot) - \int_0^1 y(s)ds, x(\cdot) - \cdot x(1), z - (x(1) - x(0))\int_0^1 y(s)ds)$, we find that Theorem 2 is applicable, and that rejecting for large values of $L^{\hat{G}}$, evaluated at sample analogues, is also the asymptotically point-optimal invariant unit root test among all tests with correct asymptotic null rejection probability for all models satisfying (18) and (20) with $c = 0$.

4.2 Linear Regression with Weak Instruments

As Andrews, Moreira, and Stock (2006) (abbreviated AMS in the following), consider the problem of inference about the coefficient of a scalar endogenous variable in the presence of weak instruments. The reduced form equations are given by (cf. equation (2.4) of AMS)

$$\begin{aligned} y_{1,t} &= z_t'\pi\beta + x_t'\zeta_1 + v_{t,1} \\ y_{2,t} &= z_t'\pi + x_t'\zeta_2 + v_{t,2} \end{aligned} \tag{23}$$

for $t = 1, \cdots, T$, where $y_{1,t}$ and $y_{2,t}$ are scalars, z_t is $k \times 1$ and x_t is $p \times 1$, and z_t are the residuals of a linear regression of the original instruments $\tilde{z}_t$ on x_t. For further reference, define $\bar{z}_t' = (\tilde{z}_t', x_t')$ and $v_t' = (v_{t,1}, v_{t,2}')$. AMS initially consider small sample efficient tests of

$$H_0 : \beta = \beta_0 \tag{24}$$

for nonstochastic $\bar{z}_t$ and $v_t = (v_{1,t}, v_{2,t})' \sim$i.i.d.$\mathcal{N}(0, \Omega)$ with Ω known. A sufficiency argument shows that tests may be restricted to functions of the $2(k+p) \times 1$ multivariate normal statistic

$$\sum_{t=1}^{T} \begin{pmatrix} z_t y_{1,t} \\ z_t y_{2,t} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} S_z \pi \beta \\ S_z \pi \end{pmatrix}, \Omega \otimes S_z \right) \tag{25}$$

where $S_z = \sum_{t=1}^{T} z_t z_t'$, and AMS derive weighted average power maximizing similar tests that are invariant to the group of transformations

$$\{z_t\}_{t=1}^{T} \to \{O z_t\}_{t=1}^{T} \text{ for any orthogonal matrix } O. \tag{26}$$

For their asymptotic analysis, AMS employ Staiger and Stock (1997) weak instrument asymptotics, where $\pi = T^{-1/2} C$ for some fixed matrix C. AMS then exploit their small sample efficiency results to construct tests (i) that maximize weighted average asymptotic power among all asymptotically invariant and asymptotically similar test when $v_t \sim$i.i.d.$\mathcal{N}(0, \Omega)$ independent of $\{(x_t', z_t')\}_{t=1}^{T}$; and (ii) that yield correct null rejection probability under much broader conditions (the working paper of AMS contains the details of the construction of heteroskedasticity, and of heteroskedasticity and autocorrelation robust tests).

To apply the results from Sections 2 and 3 to this problem, consider the set of weak convergences in the double array version of model (23)

$$\begin{aligned} &\hat{D}_Z = T^{-1} \textstyle\sum_{t=1}^{T} z_{T,t} z_{T,t}' \xrightarrow{p} D_z, \ \hat{\Sigma} \xrightarrow{p} \Sigma, \\ &X_T = T^{-1/2} \textstyle\sum_{t=1}^{T} \begin{pmatrix} z_{T,t} y_{1,T,t} \\ z_{T,t} y_{2,T,t} \end{pmatrix} \rightsquigarrow X \sim \mathcal{N}\left(\begin{pmatrix} D_z C \beta \\ D_z C \end{pmatrix}, \Sigma \right) \end{aligned} \tag{27}$$

where $\hat{\Sigma}$ is some standard estimator of the long run variance of $\text{vec}(z_t v_t')$, and D_z and Σ have full rank. The limiting problem in the sense of Section 2.2 above is thus the test of (24) based on observing the random variable X distributed as in (27), with D_z and Σ known.

For $k = 1$, i.e. in the just-identified case, this problem is exactly equivalent to the small sample problem (25) considered by AMS. In fact, already Moreira (2001) has shown that for $k = 1$, the Anderson and Rubin (1949) statistic $AR_0 = (b_0' X)^2 / b_0' \Sigma b_0$ with $b_0 = (1, -\beta_0)'$ yields the uniformly most powerful unbiased test φ_S^*. The discussion of Section 3.1 and Theorem 1 thus imply that for $k = 1$, rejecting for large values of $(b_0' X_T)^2 / b_0' \hat{\Sigma} b_0$ maximizes asymptotic power uniformly in all models that satisfy (27) with $\beta \neq \beta_0$ among all tests that are asymptotically unbiased with at most nominal asymptotic null rejection probability for all models that satisfy (27) with $\beta = \beta_0$.

The robustness constraint here—correct asymptotic null rejection probability for all models that satisfy (27) with $\beta = \beta_0$—is quite stringent, since there are many ways

$T^{-1/2}\sum_{t=1}^{T} z_{T,t}y_{T,t}$ can converge to a normal vector (for instance, one could set $z_{T,t}y_{T,t} = 0$ for all $t < T$ and $z_{T,t}y_{T,t} = X$). To the extent that one would be prepared to rule out such models a priori, this decreases the appeal of the efficiency result. But as discussed in comment 5 in Section 2.3 above, one can impose additional weak convergences without necessarily affecting φ_S^*: Supplementing (27) by the FCLT type convergence

$$T^{-1/2}\sum_{t=1}^{\lfloor \cdot T \rfloor} \begin{pmatrix} z_{T,t}y_{1,T,t} \\ z_{T,t}y_{2,T,t} \end{pmatrix} \rightsquigarrow G(\cdot) \text{ with } G(s) = s \begin{pmatrix} D_z C\beta \\ D_z C \end{pmatrix} + \Sigma^{1/2}W(s),$$

with $W(s)$ is a 2×1 standard Wiener process, for instance, rules out such pathological cases, and yet still yields φ_S^* to be the efficient test, since $G(1)$ is sufficient for the unknown parameters C and β.

For $k > 1$, Theorems 1 and 2 can again be invoked to yield analogous asymptotic efficiency statements for the statistics developed in AMS under the assumption that Σ is of the Kronecker form $\Sigma = \Omega \otimes D_z$, as in (25). But this form naturally arises only in the context of a serially uncorrelated homoskedastic model, so the resulting efficiency statements are of limited appeal. The approach here thus points to a general solution of the limiting problem without the constraint $\Sigma = \Omega \otimes D_z$ as an interesting missing piece in the literature on efficient inference in linear regressions with weak instruments.

4.3 GMM Parameter Stability Tests

Following Sowell (1996), suppose we are interested in testing the null hypothesis that a parameter $\beta \in \mathbb{R}^k$ in a GMM framework is constant through time. Parametrizing $\beta_{T,t} = \beta_0 + T^{-1/2}\theta(t/T)$, where $\theta \in D_{[0,1]}^k$ and $\theta(0)$ is normalized to zero $\theta(0) = 0$, this is equivalent to the hypothesis test

$$H_0 : \theta = 0 \quad \text{against} \quad H_1 : \theta \neq 0. \tag{28}$$

Denote by $g_{T,t}(\beta) \in \mathbb{R}^p$ with $p \geq k$ the sample moment condition for $y_{T,t}$ evaluated at β, so that under the usual assumptions, the moment condition evaluated at the true parameter value satisfies a central limit theorem, that is $T^{-1/2}\sum_{t=1}^{T} g_{T,t}(\beta_{T,t}) \rightsquigarrow \mathcal{N}(0, V)$ for some positive definite $p \times p$ matrix V. Furthermore, with $\hat{\beta}_T$ the usual full sample GMM estimator of β with optimal weighting matrix converging to V^{-1}, we obtain under typical assumptions that for some suitable estimators $\hat{H}_T$ and $\hat{V}_T$ (cf. Theorem 1 of Sowell (1996))

$$G_T(\cdot) = T^{-1/2} \sum_{t=1}^{\lfloor \cdot T \rfloor} g_{T,t}(\hat{\beta}_T) \rightsquigarrow G(\cdot) \quad \text{and} \quad \hat{H}_T \xrightarrow{p} H, \, \hat{V}_T \xrightarrow{p} V \tag{29}$$

where the convergence to G is on $D^p_{[0,1]}$, $G(s) = V^{1/2}W(s) - sH(H'V^{-1}H)^{-1}H'V^{-1/2}W(1) + H\left(\int_0^s \theta(l)dl - s\int_0^1 \theta(l)dl\right)$ with W a $p \times 1$ standard Wiener process and H some $p \times k$ matrix full column rank matrix (which is the probability limit of the average of the partial derivatives of $g_{T,t}$). Andrews (1993), Sowell (1996) and Li and Müller (2007) discuss primitive conditions for these convergences. Sowell (1996) goes on to derive weighted average power maximizing tests of (28) as a function of G (that is, he computes φ^*_S in the notation of Theorem 1), and he denotes the resulting test evaluated at $G_T(\cdot)$, $\hat{H}_T$ and $\hat{V}_T$ (that is, $\hat{\varphi}^*_T$ in the notation of Theorem 1), an "optimal" test for structural change.

Without further restrictions, however, such tests cannot claim to be efficient: As a simple example, consider the scalar model with $y_{T,t} = \beta + \theta(t/T) + \varepsilon_t$, where ε_t is i.i.d. with $P(\varepsilon_t = -1) = P(\varepsilon_t = 1) = 1/2$. This model is a standard time varying parameter GMM model with $g_{T,t}(\beta) = y_{T,t} - \beta = \theta(t/T) + \varepsilon_t$ satisfying (29), yet in this model, the test φ^{**}_T that rejects whenever any one of $\{y_{T,t} - y_{T,t-1}\}_{t=1}^T$ is not -2, 0 or 2 has level zero for any $T \geq 2$ and has asymptotic power equal to one against any local alternative.

Theorem 1 provides a sense in which the tests derived by Sowell (1996) are asymptotically optimal: they maximize asymptotic weighted average power among all tests that have correct asymptotic null rejection probability whenever (29) holds with $\theta = 0$. Tests that exploit specificities of the error distribution, such as φ^{**}_T, to gain higher power necessarily do not lead to correct asymptotic null rejection probability for all stable models satisfying (29).

5 Conclusion

This paper analyzes a new notion of asymptotic efficiency. The starting point is the idea that models of interest can be delineated in terms of a weak convergence property, so that under the null and alternative hypothesis, an observed random element converges weakly to a limiting random element. It is shown that if one restricts attention to tests that have nominal asymptotic rejection probability for all models that satisfy the weak convergence under the null hypothesis, then efficient tests in the original problem are simply given by efficient tests in the limiting problem (that is, with the limiting random element observed), evaluated at sample analogues. These efficient tests generically coincide with robustified versions of efficient tests that are derived as the limit of small sample efficient tests in canonical parametric versions of the model. The results of this paper thus provide an alternative and broader sense of asymptotic efficiency for many previously derived tests in econometrics, and provides a concrete limit how far these tests can be improved upon in less parametric set-ups.

It is a severe restriction to force tests to have correct asymptotic null rejection probability for all models that satisfy the weak convergence, because there are typically many such models. Relative to the semi-parametrically efficient testing literature, the issue of asymptotic efficiency is thus approached here with a stronger focus on robustness of inference. This allows for the application of a different set of arguments, and we accordingly obtain a different (negative) result: it is impossible to even partially adapt to non-canonical versions of the model. Given this focus on the robustness of inference, the results derived here are most naturally to applied in time series econometrics, where often there is substantial uncertainty over the joint distribution of the data (and where semiparametric considerations face serious challenges).

Given the difference in focus, the approach considered here complements rather than substitutes the approach based on semiparametric efficiency. It seems an interesting topic for future research to more deeply explore the relationship between the two approaches.

6 Appendix

Proof of Theorem 1:

(i) We prove the second claim, the first is proved analogously.

Since φ_S^* is $P(\theta_0)$-almost everywhere continuous, it is also $P(\theta)$-almost everywhere continuous for any $\theta \in \Theta_1$, and by the Continuous Mapping Theorem, $P_T(m, \theta) \rightsquigarrow P(\theta)$ implies $\int \varphi_S^* dP_T(m, \theta) \to \int \varphi_S^* dP(\theta)$ for all $\theta \in \Theta_1$. Since $0 \leq \varphi_S^* \leq 1$, the result follows by dominated convergence.

(ii) Pick any $m \in \mathcal{M}_1$. Define $\bar{F}_T = \int F_T(m, \theta) dw(\theta)$, $\bar{P}_T = \int P_T(m, \theta) dw(\theta)$ and $\bar{P} = \int P(\theta) dw(\theta)$. For any bounded and continuous function $\vartheta : S \mapsto \mathbb{R}$, $\int \vartheta dP_T(m, \theta) \to \int \vartheta dP(\theta)$ for all $\theta \in \Theta_1$ by the continuous mapping theorem, so that by dominated convergence, also $\int \vartheta d\bar{P}_T \to \int \vartheta d\bar{P}$. Thus, $h_T \bar{F}_T = \bar{P}_T \rightsquigarrow \bar{P}$. Let $D^n_{[0,1]}$ be the space of n-valued cadlag functions on the unit interval, equipped with the Billingsley (1968) metric, and define the mapping $\chi_T : \mathbb{R}^{nT} \mapsto D^n_{[0,1]}$ as $\{y_t\}_{t=1}^T \mapsto T^{-1} \Phi_Z(y_{\lfloor \cdot T \rfloor})$, where Φ_Z is the c.d.f. of a standard normal applied element by element. Note that χ_T is injective, and denote by χ_T^{-1} a $D^n_{[0,1]} \mapsto \mathbb{R}^{nT}$ function such that $\chi_T^{-1}(\chi_T(y)) = y$ for all $y \in \mathbb{R}^{nT}$. Since $\sup_{s \in [0,1]} ||\chi_T(s)|| \leq 1/T \to 0$, the probability measures $(h_T, \chi_T)\bar{F}_T$ on the complete and separable space $S \times D^n_{[0,1]}$ converge weakly to the product measure $\bar{P} \times \delta_0$, where δ_0 puts all mass at the zero function in $D^n_{[0,1]}$. Let $T_1 \to \infty$ be any subsequence of T such that $\lim_{T_1 \to \infty} \int \varphi_{T_1} d\bar{F}_{T_1} = \limsup_{T \to \infty} \int \varphi_T d\bar{F}_T$. Since the probability measures $(h_{T_1}, \chi_{T_1}, \varphi_{T_1})\bar{F}_{T_1}$ on the complete and separable space $S \times D^n_{[0,1]} \times [0, 1]$ are tight, by Prohorov's Theorem (see, for instance, Theorem 36 on p. 185 of Pollard (2002)), there exists a subsequence T_2 of T_1 such that $(h_{T_2}, \chi_{T_2}, \varphi_{T_2})\bar{F}_{T_2} \rightsquigarrow \bar{\nu}$ as $T_2 \to \infty$, where $(\pi_X, \pi_Y)\bar{\nu} = \bar{P} \times \delta_0$, and π_X, π_Y and π_φ are the projections of $S \times D^n_{[0,1]} \times [0, 1]$ on S, $D^n_{[0,1]}$ and $[0, 1]$, respectively. For notational convenience, write T for T_2 in the following. By Theorem 11.7.2 of Dudley (2002), there exists a probability space $(\Omega^*, \mathcal{F}^*, P^*)$ and functions $\eta_T : \Omega^* \mapsto S \times D^n_{[0,1]} \times [0, 1]$ such that $\eta_T P^* = (h_T, \chi_T, \varphi_T)\bar{F}_T$, $\eta_0 P^* = \bar{\nu}$ and $\eta_T(\omega^*) \to \eta_0(\omega^*)$ for P^*-almost all $\omega^* \in \Omega^*$, and by Theorem 11.7.3 of Dudley (2002), we may assume Ω^* to be complete and separable. In this construction, note that for P^*-almost all ω^*, $h_T \circ \chi_T^{-1} \circ \pi_Y \circ \eta_T(\omega^*) = \pi_X \circ \eta_T(\omega^*)$ and $\varphi_T \circ \chi_T^{-1} \circ \pi_Y \circ \eta_T(\omega^*) = \pi_\varphi \circ \eta_T(\omega^*)$. Furthermore, $(\pi_X \circ \eta_T)P^* = \bar{P}_T$, $(\chi_T^{-1} \circ \pi_Y \circ \eta_T)P^* = \bar{F}_T$ and $\int \varphi_T d\bar{F}_T = \int (\pi_\varphi \circ \eta_T) dP^* \to \int (\pi_\varphi \circ \eta_0) dP^*$, where the convergence follows from the dominated convergence theorem, since $\pi_\varphi \circ \eta_T(\omega^*) \in [0, 1]$ for all $\omega^* \in \Omega^*$. We need to show that $\int (\pi_\varphi \circ \eta_0) dP^* \leq \int \varphi_S^* d\bar{P}$.

Note that $P(\theta)$ is absolutely continuous with respect to $\bar{P}$ for any $\theta \in \Theta$, and denote by $L(\theta)$ the Radon-Nikodym derivative of $P(\theta)$ with respect to $\bar{P}$, so that for all $A \in \mathfrak{B}(S)$, $\int_A dP(\theta) = \int_A L(\theta)d\bar{P}$ (existence of L is ensured by the Radon-Nikodym Theorem; see, for instance, page 56 of Pollard (2002)). Define $Q^*(\theta)$ to be the probability measure on $\mathcal{F}^*$, indexed by $\theta \in \Theta$, as $\int_A dQ^*(\theta) = \int_A (L(\theta) \circ \pi_X \circ \eta_0)dP^*$ for all $A \in \mathcal{F}^*$. By construction, $(\pi_X \circ \eta_0)Q^*(\theta) = P(\theta)$, since for all $A \in \mathfrak{B}(S)$, $\int_A (\pi_X \circ \eta_0)dQ^*(\theta) = \int_A L(\theta)d\bar{P} = \int_A dP(\theta)$. Consider the hypothesis test

$$H_0^* : \omega^* \sim Q^*(\theta), \theta \in \Theta_0 \text{ vs } H_1^* : \omega^* \sim Q^*(\theta), \theta \in \Theta_1. \tag{30}$$

Because the Radon-Nikodym derivative between $Q^*(\theta_1)$ and $Q^*(\theta_2)$ is given by $(L(\theta_1)/L(\theta_2)) \circ \pi_X \circ \eta_0$, the statistic $\pi_X \circ \eta_0 : \Omega^* \mapsto S$ is sufficient for θ by the Factorization Theorem (see, for instance, Theorem 2.21 of Schervish (1995)). Thus, for any test $\varphi_\Omega : \Omega^* \mapsto [0,1]$, one can define a corresponding test $\varphi_S : S \mapsto [0,1]$ via $\varphi_S(x) = E[\varphi_\Omega|(\pi_X \circ \eta_0)(\omega^*) = x]$, which satisfies $\int \varphi_\Omega dQ^*(\theta) = \int(\varphi_S \circ \pi_X \circ \eta_0)dQ^*(\theta) = \int \varphi_S dP(\theta)$ for all $\theta \in \Theta$ (cf. Theorem 3.18 of Schervish (1995)), and also $\int (f_S \circ \pi_X \circ \eta_0)(\varphi_\Omega - \alpha)dQ^*(\theta) = \int f_S(\varphi_S - \alpha)dP(\theta)$ for any $f_S \in \mathcal{F}_S$. Since the level α test $\varphi_S^* : S \mapsto [0,1]$ of $H_0 : X \sim P(\theta)$, $\theta \in \Theta_0$ against $H_1 : X \sim P(\theta)$, $\theta \in \Theta_1$ maximizes weighted average power subject to (6) and (7), the level α test $\varphi_S^* \circ \pi_X \circ \eta_0 : \Omega^* \mapsto [0,1]$ of (30) maximizes weighted average power $\int\int \varphi_\Omega dQ^*(\theta)dw(\theta)$ among all level α tests $\varphi_\Omega : \Omega^* \mapsto [0,1]$ of (30) that satisfy

$$\int \varphi_\Omega dQ^*(\theta) \;\geq\; \pi_0(\theta) \quad \text{for all } \theta \in \Theta_1 \tag{31}$$

$$\int (f_S \circ \pi_X \circ \eta_0)(\varphi_\Omega - \alpha)dQ^*(\theta) \;=\; 0 \text{ for all } \theta \in \bar{\Theta}_0 \text{ and } f_S \in \mathcal{F}_S, \tag{32}$$

and it achieves the same weighted average power $\int\int(\varphi_S^* \circ \pi_X \circ \eta_0)dQ^*(\theta)dw(\theta) = \int\int \varphi_S^* dP(\theta)dw(\theta)$.

Now define the sequence of measures $G_T(\theta)$ on $\mathfrak{B}(\mathbb{R}^{nT})$, indexed by $\theta \in \Theta$, via $G_T(\theta) = (\chi_T^{-1} \circ \pi_Y \circ \eta_T)Q^*(\theta)$, which induce the measures $h_T G_T(\theta) = (h_T \circ \chi_T^{-1} \circ \pi_Y \circ \eta_T)Q^*(\theta) = (\pi_X \circ \eta_T)Q^*(\theta)$ on $\mathfrak{B}(S)$. By construction of η_T and absolute continuity of $Q^*(\theta)$ with respect to P^*, we have $\eta_T(\omega^*) \to \eta_0(\omega^*)$ for $Q^*(\theta)$-almost all ω^*, and since almost sure convergence implies weak convergence, $h_T G_T(\theta) \rightsquigarrow (\pi_X \circ \eta_0)Q^*(\theta) = P(\theta)$ pointwise in $\theta \in \Theta$. Thus, for any test that satisfies (8), (9) and (10), we have $\limsup_{T\to\infty} \int \varphi_T dG_T(\theta) \leq \alpha$ for all $\theta \in \Theta_0$, $\liminf_{T\to\infty} \int \varphi_T dG_T(\theta) \geq \pi_0(\theta)$ for all $\theta \in \Theta_1$ and $\lim_{T\to\infty} \int(\varphi_T - \alpha)(f_S \circ h_T)dG_T(\theta) = 0$ for all $\theta \in \bar{\Theta}_0$ and $f_S \in \mathcal{F}_S$. By the dominated convergence theorem and the construction of

G_T, $\int \varphi_T dG_T(\theta) = \int(\varphi_T \circ \chi_T^{-1} \circ \pi_Y \circ \eta_T)dQ^*(\theta) = \int(\pi_\varphi \circ \eta_T)dQ^*(\theta) \to \int(\pi_\varphi \circ \eta_0)dQ^*(\theta)$ for all $\theta \in \Theta$ and $\int(\varphi_T - \alpha)(f_S \circ h_T)dG_T(\theta) = \int((\pi_\varphi \circ \eta_T) - \alpha)(f_S \circ \pi_X \circ \eta_T)dQ^*(\theta) \to \int((\pi_\varphi \circ \eta_0) - \alpha)(f_S \circ \pi_X \circ \eta_0)dQ^*(\theta)$ for all $\theta \in \bar{\Theta}_0$ and $f_S \in \mathcal{F}_S$. We can conclude that the test $\pi_\varphi \circ \eta_0 : \Omega^* \mapsto [0,1]$ of (30) is of level α and satisfies (31) and (32). Its weighted average power $\int \int(\pi_\varphi \circ \eta_0)dQ^*(\theta)dw(\theta)$ is therefore smaller or equal than the weighted average power $\int \int \varphi_S^* dP(\theta)dw(\theta)$ of the test $\varphi_S^* \circ \pi_X \circ \eta_0 : \Omega^* \mapsto [0,1]$. The result now follows from noting that $\int \int(\pi_\varphi \circ \eta_0)dQ^*(\theta)dw(\theta) = \int \int(L(\theta) \circ \pi_X \circ \eta_0)(\pi_\varphi \circ \eta_0)dP^* dw(\theta) = \int(\pi_\varphi \circ \eta_0)dP^*$, since $\int L(\theta)dw(\theta) = 1$, and the change of the order of integration is allowed by Fubini's Theorem.

Proof of Theorem 2:

For notational convenience, write $F_{0,T} = F_T(m,\theta)$, $P_T = h_T F_{0,T}$, $P_{0,T}^\phi = (h_T \circ \phi_T)F_{0,T}$, $P_0 = P(\theta)$ and $P_0^\phi = \tilde{\phi}P_0$. Proceed in analogy to the proof of Theorem 1 part (ii) to argue for the existence of a probability space $(\Omega^*, \mathcal{F}^*, P^*)$ with complete and separable Ω^* and functions $\eta_T : \Omega^* \mapsto S \times D_{[0,1]}^n$ such that $\eta_T P^* = (h_T \circ \phi_T, \chi_T)F_{0,T}$, $\eta_0 P^* = P_0^\phi \times \delta_0$, where the probability measure δ_0 puts all mass on the zero function $D_{[0,1]}^n \mapsto \mathbb{R}$, and $\eta_T(\omega^*) \to \eta_0(\omega^*)$ for P^*-almost all $\omega^* \in \Omega^*$. In particular, $(\pi_X \circ \eta_T)P^* = P_{0,T}^\phi$ and $(\chi_T^{-1} \circ \pi_Y \circ \eta_T)P^* = F_{0,T}$, where π_X and π_Y are the usual projections of $S \times D_{[0,1]}^n$ on S and $D_{[0,1]}^n$, respectively. Also, for almost all ω^*, $h_T \circ \phi_T \circ \chi_T^{-1} \circ \pi_Y \circ \eta_T(\omega^*) = \pi_X \circ \eta_T(\omega^*)$.

Let ν be the probability measure on $\mathfrak{B}(R \times S)$ induced by $(\tilde{\rho}, \tilde{\phi}) : S \mapsto R \times S$ under P_0, i.e. $\nu = (\tilde{\rho}, \tilde{\phi})P_0$. Since $x = \tilde{g}(\tilde{\rho}(x), \tilde{\phi}(x))$ for all $x \in S$, $P_0 = \tilde{g}\nu$. By Proposition 10.2.8 of Dudley (2002) there exists a probability kernel ν_x from $(S, \mathfrak{B}(S))$ to $(R, \mathfrak{B}(R))$ such that for each $A \in \mathfrak{B}(S)$ and $B \in \mathfrak{B}(R)$, $\nu(A \times B) = \int_A \nu_x(B)dP_0^\phi(x)$. Note that the $\Omega^* \times \mathfrak{B}(S) \mapsto [0,1]$ mapping defined via $(\omega^*, B) \mapsto \nu_{\pi_X \circ \eta_0(\omega^*)}(B)$ is a probability kernel from $(\Omega^*, \mathcal{F}^*)$ to $(R, \mathfrak{B}(R))$. We can thus construct the probability measure μ^* on $(\Omega^* \times R)$, $(\mathcal{F}^* \otimes \mathfrak{B}(R))$ via $\mu^*(C \times B) = \int_C \nu_{\pi_X \circ \eta_0(\omega^*)}(B)dP^*(\omega^*)$, and by construction, the mapping $(\omega^*, r) \mapsto \tilde{g}(r, \pi_X \circ \eta_0(\omega^*))$ induces the measure P_0 under μ^*.

Now from assumptions (iv), for μ^*-almost all (ω^*, r), $h_T \circ g_T(r, \phi_T \circ \chi_T^{-1} \circ \pi_Y \circ \eta_T(\omega^*)) = \tilde{g}(r, h_T \circ \phi_T \circ \chi_T^{-1} \circ \pi_Y \circ \eta_T(\omega^*)) + o(1) = \tilde{g}(r, \pi_X \circ \eta_T(\omega^*)) + o(1) \to \tilde{g}(r, \pi_X \circ \eta_0(\omega^*))$ where the convergence follows from the continuity of $\tilde{g}$. But almost sure convergence implies weak convergence, so that the measures G_T on $\mathfrak{B}(\mathbb{R}^{nT})$ induced by the mapping $(\omega^*, r) \mapsto g_T(r, \phi_T \circ \chi_T^{-1} \circ \pi_Y \circ \eta_T(\omega^*))$ under μ^* satisfy $h_T G_T \rightsquigarrow P_0$. Finally, from $\phi_T \circ g_T(r, \phi_T(y)) = \phi_T(y)$ for all $r \in R$ and $y \in \mathbb{R}^{nT}$, it follows that $\phi_T G_T = (\phi_T \circ \chi_T^{-1} \circ \pi_Y \circ \eta_T)P^* = \phi_T F_{0,T}$.

Proof of Theorem 3:

The equality $\lim_{T\to\infty}\int\int\hat{\varphi}_T^* dF_T(m_1,\theta)dw(\theta)=\int\int\varphi_S^* dP(\theta)dw(\theta)$ follows as in the proof of Theorem 1 part (i).

For the inequality, write $P_0=P(\theta_0)$, $\bar{P}=\bar{L}P_0$, $F_{0,T}=F_T(m_0,\theta_0)$, $P_{0,T}=h_T F_{0,T}$, $\bar{F}_T=\int F_T(m_1,\theta)dw(\theta)$ and $\bar{P}_T=h_T\bar{F}_T$, so that $P_{0,T}\rightsquigarrow P_0$ and $\bar{P}_T\rightsquigarrow\bar{P}$ by assumption. Pick $1/2>\epsilon>0$ such that $P_0(\bar{L}=\epsilon)=0$ and define $B_\epsilon=\{x\in S:\bar{L}>\epsilon\}$. Note that $\int_{S\backslash B_\epsilon}d\bar{P}=\int_{S\backslash B_\epsilon}\bar{L}dP_0\leq\epsilon\int_{S\backslash B_\epsilon}dP_0\leq\epsilon$, so that $\int_{B_\epsilon}d\bar{P}\geq 1-\epsilon$. The assumption about $\bar{L}$ in the statement of the theorem also implies the existence of an open set $\bar{B}_\epsilon$ such that $\int_{\bar{B}_\epsilon}d\bar{P}>1-\epsilon$ and $\bar{L}:\bar{B}_\epsilon\mapsto\mathbb{R}$ is Lipschitz, since $\bar{P}$ is absolutely continuous with respect to P_0 (cf. Example 3 on page 55 of Pollard (2002)). With $B=B_\epsilon\cap\bar{B}_\epsilon$, $\bar{L}^i:B\mapsto\mathbb{R}$ with $\bar{L}^i=1/\bar{L}$ is thus bounded and Lipschitz. Furthermore, since $S\backslash B$ is closed, there exists a Lipschitz function $\ell:S\mapsto[0,1]$ that is zero on $S\backslash B$ and for which $\int\ell d\bar{P}\geq 1-3\epsilon$ (see Pollard (2002), p. 172-173 for an explicit construction). For future reference, define $\mathcal{B}$ and $\mathcal{B}_\epsilon$ to be the indicator functions of B and B_ϵ, respectively, and note that $\mathcal{B}\ell=\mathcal{B}_\epsilon\ell=\ell$. Define the scalar sequence

$$\begin{aligned}\kappa_T &= \int(\ell\circ h_T)dF_{0,T}\Big/\int(\ell\circ h_T)(\bar{L}^i\circ h_T)d\bar{F}_T\\ &= \int\ell dP_{0,T}\Big/\int\ell\bar{L}^i d\bar{P}_T\end{aligned}$$

and note that $\kappa_T\to 1$ because $\int\ell dP_{0,T}\to\int\ell dP_0$ and $\int\ell\bar{L}^i d\bar{P}_T\to\int\ell\bar{L}^i d\bar{P}=\int\mathcal{B}\ell\bar{L}^i\bar{L}dP_0=\int\ell dP_0$ by the Continuous Mapping Theorem. Further, define the probability distribution G_T on $\mathfrak{B}(\mathbb{R}^{nT})$ via

$$\int_A dG_T=\kappa_T\int_A(\ell\circ h_T)(\bar{L}^i\circ h_T)d\bar{F}_T+\int_A((1-\ell)\circ h_T)dF_{0,T}$$

for any $A\in\mathfrak{B}(\mathbb{R}^{nT})$. Then by construction, h_T induces the probability distribution Q_T under G_T, where Q_T satisfies

$$\int_A dQ_T=\kappa_T\int_A\ell\bar{L}^i d\bar{P}_T+\int_A(1-\ell)dP_{0,T}$$

for any $A\in\mathfrak{B}(S)$. Now

$$\begin{aligned}\Delta_{BL}(Q_T,P_0) &= \sup_{||\vartheta||_{BL}\leq 1}\left|\int\vartheta(dQ_T-dP_0)\right|\\ &\leq \sup_{||\vartheta||_{BL}\leq 1}\left|\int\vartheta\ell\bar{L}^i(\kappa_T d\bar{P}_T-d\bar{P})\right|+\sup_{||\vartheta||_{BL}\leq 1}\left|\int\vartheta(1-\ell)(dP_{0,T}-dP_0)\right|\\ &\leq ||\ell\bar{L}^i||_{BL}(\Delta_{BL}(\bar{P}_T,\bar{P})+|\kappa_T-1|)+||1-\ell||_{BL}\Delta_{BL}(P_{0,T},P_0)\end{aligned}$$

where the manipulations after the first inequality use $\int \vartheta \ell dP_0 = \int \vartheta \ell \bar{L}^i d\bar{P}$ and the second inequality exploits that $||\cdot||_{BL}$ is a submultiplicative norm on the set of bounded Lipschitz functions $S \mapsto \mathbb{R}$ (cf. Proposition 11.2.1 of Dudley (2002)). Also,

$$\begin{aligned} |\kappa_T - 1| &= \left|\frac{\int \ell dP_{0,T}}{\int \ell \bar{L}^i d\bar{P}_T} - \frac{\int \ell dP_0}{\int \ell \bar{L}^i d\bar{P}}\right| \\ &\leq \frac{||\ell \bar{L}^i||_{BL} \Delta_{BL}(\bar{P}_T, \bar{P}) + ||\ell||_{BL} \Delta_{BL}(P_{0,T}, P_0)}{\int \ell \bar{L}^i d\bar{P}_T}. \end{aligned}$$

Thus, $\lim_{T\to\infty} \Delta_{BL}(Q_T, P_0)/\delta_T = 0$, so that for large enough T, (16) implies $\limsup_{T\to\infty} \int \varphi_T dG_T \leq \alpha$.

Now define the probability measures $\tilde{F}_T$ via

$$\begin{aligned} \int_A d\tilde{F}_T &= \tilde{\kappa}_T \int_A (\mathcal{B}_\epsilon \circ h_T)(\bar{L} \circ h_T) dG_T \\ &= \tilde{\kappa}_T \kappa_T \int_A (\ell \circ h_T) d\bar{F}_T + \tilde{\kappa}_T \int_A (\mathcal{B}_\epsilon (1-\ell) \bar{L} \circ h_T) dF_{0,T} \end{aligned}$$

for any $A \in \mathfrak{B}(\mathbb{R}^{nT})$, where $\tilde{\kappa}_T = 1/(\kappa_T \int (\ell \circ h_T) d\bar{F}_T + \int (\mathcal{B}_\epsilon (1-\ell)\bar{L} \circ h_T) dF_{0,T}) \to \tilde{\kappa} = 1/\int (\ell + \mathcal{B}_\epsilon - \mathcal{B}_\epsilon \ell) d\bar{P} = 1/\int \mathcal{B}_\epsilon d\bar{P}$ and $1 \leq 1/\int \mathcal{B}_\epsilon d\bar{P} \leq 1 + 2\epsilon$. By the Neyman-Pearson Lemma, the best test of $\tilde{H}_0 : Y_T \sim G_T$ against $\tilde{H}_1 : Y_T \sim \tilde{F}_T$ thus rejects for large values of $(\mathcal{B}_\epsilon \bar{L}) \circ h_T$, i.e. $\bar{L} \circ h_T$.

For any T, denote by $\tilde{\varphi}_T^* : \mathbb{R}^{nT} \mapsto [0,1]$ the test that rejects for large values of $\bar{L} \circ h_T$ of level $\int \tilde{\varphi}_T^* dG_T = \max(\int \varphi_T dG_T, \alpha)$, so that $\int (\tilde{\varphi}_T^* - \varphi_T) d\tilde{F}_T \geq 0$ for all T. By LeCam's First Lemma (Lemma 6.4 in van der Vaart (1998)), $\tilde{F}_T$ is contiguous to G_T, since under G_T, the Radon-Nikodym derivative $\tilde{\kappa}_T (\mathcal{B}_\epsilon \bar{L}) \circ h_T$ converges weakly to the distribution $\tilde{\kappa} \mathcal{B}_\epsilon \bar{L} P_0$ by the Continuous Mapping Theorem, and $\int \tilde{\kappa} \mathcal{B}_\epsilon \bar{L} dP_0 = 1$. Since both $\tilde{\varphi}_T^*$ and $\hat{\varphi}_T^*$ reject for large values of $\bar{L} \circ h_T$ and are of asymptotic level α, we have $\int |\tilde{\varphi}_T^* - \hat{\varphi}_T^*| dG_T \to 0$, so that by contiguity, also $\int |\tilde{\varphi}_T^* - \hat{\varphi}_T^*| d\tilde{F}_T \to 0$. Thus $\limsup_{T\to\infty} \int (\varphi_T - \hat{\varphi}_T^*) d\tilde{F}_T \leq 0$. To complete the proof, note that the total variation distance between $\tilde{F}_T$ and $\bar{F}_T$ is bounded above by

$$\begin{aligned} \int |1 - \tilde{\kappa}_T \kappa_T (\ell \circ h_T)| d\bar{F}_T &\leq |1 - \tilde{\kappa}_T \kappa_T| + |\tilde{\kappa}_T \kappa_T \int (1-\ell) d\bar{P}_T| \\ &\to 1/\int \mathcal{B}_\epsilon d\bar{P} - 1 + (1 - \int \ell d\bar{P})/\int \mathcal{B}_\epsilon d\bar{P} \leq 8\epsilon \end{aligned}$$

so that $\limsup_{T\to\infty} \int (\varphi_T - \hat{\varphi}_T^*) d\bar{F}_T \leq 8\epsilon$, and the result follows, since the choice of $1/2 > \epsilon > 0$ was arbitrary.

References

ANDERSON, T. W., AND H. RUBIN (1949): "Estimators of the Parameters of a Single Equation in a Complete Set of Stochastic Equations," *The Annals of Mathematical Statistics*, 21, 570–582.

ANDREWS, D. W. K. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856.

ANDREWS, D. W. K., M. J. MOREIRA, AND J. H. STOCK (2006): "Optimal Two-Sided Invariant Similar Tests for Instrumental Variabes Regression," *Econometrica*, 74, 715–752.

——— (2007): "Efficient Two-Sided Nonsimilar Invariant Tests in IV Regression with Weak Instruments," *Working paper, Harvard University.*

ANDREWS, D. W. K., AND W. PLOBERGER (1994): "Optimal Tests When a Nuisance Parameter Is Present Only under the Alternative," *Econometrica*, 62, 1383–1414.

BILLINGSLEY, P. (1968): *Convergence of Probability Measure*. Wiley, New York.

BREITUNG, J. (2002): "Nonparametric Tests for Unit Roots and Cointegration," *Journal of Econometrics*, 108, 343–363.

DAVIDSON, J. (2002): "Establishing Conditions for the Functional Central Limit Theorem in Nonlinear and Semiparametric Time Series Processes," *Journal of Econometrics*, 106, 243–269.

——— (2007): "When is a Time Series I(0)?," *University of Exeter Working Paper.*

DE JONG, R. M., AND J. DAVIDSON (2000): "The Functional Central Limit Theorem and Weak Convergence to Stochastic Integrals I: Weakly Dependent Processes," *Econometric Theory*, 16, 621–642.

DUDLEY, R. M. (2002): *Real Analysis and Probability*. Cambridge University Press, Camebridge, UK.

ELLIOTT, G. (1999): "Efficient Tests for a Unit Root When the Initial Observation is Drawn From its Unconditional Distribution," *International Economic Review*, 40, 767–783.

Elliott, G., and M. Jansson (2003): "Testing for Unit Roots with Stationary Covariates," *Journal of Econometrics*, 115, 75–89.

Elliott, G., M. Jansson, and E. Pesavento (2005): "Optimal Power for Testing Potential Cointegrating Vectors with Known Parameters for Nonstationarity," *Journal of Business and Economics Statistics*, 23, 34–48.

Elliott, G., and U. K. Müller (2006): "Efficient Tests for General Persistent Time Variation in Regression Coefficients," *Review of Economic Studies*, 73, 907–940.

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

Hansen, B. E. (1990): "Convergence to a Stochastic Integrals for Dependent Heterogeneous Processes," *Econometric Theory*, 8, 489–500.

Ibragimov, R., and U. K. Müller (2007): "T-Statistic Based Correlation and Heterogeneity Robust Inference," *Harvard Institute of Economic Research Discussion Paper No. 2129.*

Jansson, M. (2005): "Point Optimal Tests of the Null Hypothesis of Cointegration," *Journal of Econometrics*, 124, 187–201.

——— (2007): "Semiparametric Power Envelopes for Tests of the Unit Root Hypothesis," *forthcoming in Econometrica.*

Jansson, M., and M. J. Moreira (2006): "Optimal Inference in Regression Models with Nearly Integrated Regressors," *Econometrica*, 74, 681–714.

Li, H., and U. K. Müller (2007): "Valid Inference in Partially Unstable GMM Models," *Working paper, Princeton University.*

McLeish, D. L. (1974): "Dependent Central Limit Theorems and Invariance Principles," *The Annals of Probability*, 2, 620–628.

Moreira, M. J. (2001): "Tests with Correct Size When Instruments Can Be Arbitrarily Weak," *Working Paper, Berkeley.*

Müller, U. K. (2006): "The Impossibility of Consistent Discrimination Between I(0) and I(1) Processes," *forthcoming in Econometric Theory.*

——— (2007): "Comment on: Unit Root Testing in Practice Dealing with Uncertainty Over the Trend and Initial Condition," *forthcoming in Econometric Theory.*

Müller, U. K., and G. Elliott (2003): "Tests for Unit Roots and the Initial Condition," *Econometrica*, 71, 1269–1286.

Müller, U. K., and M. W. Watson (2007): "Testing Models of Low-Frequency Variability," *forthcoming in Econometrica.*

——— (2008): "Low-Frequency Robust Cointegration Testing," *Working paper, Princeton University.*

Nyblom, J. (1989): "Testing for the Constancy of Parameters Over Time," *Journal of the American Statistical Association*, 84, 223–230.

Phillips, P. C. B. (1988): "Weak Convergence of Sample Covariance Matrices to Stochastic Integrals Via Martingale Approximations," *Econometric Theory*, 4, 528–533.

Phillips, P. C. B., and V. Solo (1992): "Asymptotics for Linear Processes," *Annals of Statistics*, 20, 971–1001.

Pollard, D. (2001): "Contiguity," Working paper, Yale University.

——— (2002): *A User's Guide to Measure Theoretic Probability.* Cambridge University Press, Cambridge, UK.

Rothenberg, T. J., and J. H. Stock (1997): "Inference in a Nearly Integrated Autoregressive Model with Nonnormal Innovations," *Journal of Econometrics*, 80, 269–286.

Schervish, M. J. (1995): *Theory of Statistics.* Springer, New York.

Sowell, F. (1996): "Optimal Tests for Parameter Instability in the Generalized Method of Moments Framework," *Econometrica*, 64, 1085–1107.

Staiger, D., and J. H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

STOCK, J. H. (1994a): "Deciding Between I(1) and I(0)," *Journal of Econometrics*, 63, 105–131.

——— (1994b): "Unit Roots, Structural Breaks and Trends," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. McFadden, vol. 4, pp. 2740–2841. North Holland, New York.

STOCK, J. H., AND M. W. WATSON (1996): "Confidence Sets in Regressions with Highly Serially Correlated Regressors," *Working Paper, Harvard University.*

STOCK, J. H., AND J. H. WRIGHT (2000): "GMM with Weak Identification," *Econometrica*, 68, 1055–1096.

VAN DER VAART, A. W. (1998): *Asymptotic Statistics.* Cambridge University Press, Cambridge, UK.

WHITE, H. (2001): *Asymptotic Theory for Econometricians.* Academic Press, San Diego, revised edn.

WOOLDRIDGE, J. M., AND H. WHITE (1988): "Some Invariance Principles and Central Limit Theorems for Dependent Heterogeneous Processes," *Econometric Theory*, 4, 210–230.