An Alternative Sense of Asymptotic Efficiency

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

Preliminary and Incomplete
October 2007

Abstract

This paper studies the asymptotic efficiency and robustness of tests when models of interest are defined in terms of a weak convergence property. The null and local alternatives induce different distributions for a limiting random element, and a test is deemed robust if it controls asymptotic size for all data generating processes for which the random element has the null limiting distribution. It is found that under weak regularity conditions, asymptotically robust and efficient tests in the original problem are then simply given by efficient tests of the limiting problem (that is, with the limiting random element observed), evaluated at sample analogues. These tests often coincide with suitably robustified versions of optimal tests for i.i.d. Gaussian disturbances in the original problem. The result therefore implies that many standard methods cannot be improved upon without losing robustness, and thus limits the scope for successful adaption.

JEL classification: C32

Keywords: hypothesis tests, optimality, robustness, weak convergence, adaption

1 Introduction

A continued focus of recent econometrics 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) and Elliott and Müller (2003) 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, Nyblom (1989), Andrews and Ploberger (1994) and Elliott and Müller (2006) derive optimal tests of parameter stability, Jansson (2005) derives optimal tests for the null hypothesis of cointegration, 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) derive optimal tests for regression coefficients in the presence of potentially weak instruments, just to name a few. 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, with suitable modifications, 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 a natural question 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, Rothenberg and Stock (1997) and Jansson (2007) derive such tests for the unit root null hypothesis by drawing on and extending the theory of semi-parametrically efficient tests. Also, as for the tests derived under Gaussianity, Rothenberg and Stock (1997) and Jansson (2007) show 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 tests of interest, a wide range of data generating processes imply the weak convergence of some random elements 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. Suppose one is sufficiently unsure about the nature of the short run dynamics that one would like the test not to reject 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 given by the Neyman-Pearson test of the limiting random element, evaluated at sample analogues. In the unit root testing example, the best test to discriminate between a Wiener process and an Ornstein-Uhlenbeck process, evaluated at sample analogues, 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 best robust test for some non-Gaussian disturbances necessarily lacks robustness: its asymptotic rejection probability is larger than the nominal level for some model whose data converges 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 derive the Neyman-Pearson test for the limiting random element. The potentially complicated small sample testing problem it 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.

Beyond the unit root testing problem, the results in this paper can be applied to yield stronger efficiency implications for a number of standard tests that are asymptotically efficient by construction for i.i.d. Gaussian disturbances. In particular, Elliott and Jansson's (2003) and Jansson's (2004, 2005) point-optimal tests for unit roots with stationary covariates, stationarity tests with covariates and tests for the null hypothesis of cointegration, respectively, are more generally asymptotically point-optimal in the sense described above: if one imposes robustness in the sense of correct asymptotic null rejection probability for all models that satisfy the usual I(0)/I(1) convergences (that is, I(1) processes and partial sums of I(0) processes are defined in terms of convergence to a Wiener process), then these tests are asymptotically point-optimal against all models that satisfy the usual convergence under the local alternative. In other words, any test with higher asymptotic local power against *any* usual local alternative (with non-Gaussian and correlated disturbances) necessarily lacks robustness. The same holds for the usual Generalized Method of Moments (GMM) Wald test (where the GMM estimator converges weakly to a multivariate normal under the null and local alternatives) and Sowell's (1996) GMM parameter stability tests. Furthermore, the results of this paper also imply efficiency of nonstandard methods that take a weak convergence assumption as their starting point, such as those suggested in Müller and Watson (2006) and Ibragimov and Müller (2007).

The remainder of the paper is organized as follows. Section 2 contains the main result, and provides a discussion of the issue of uniformity over alternative models with the same weak convergence property and discusses the implications of an invariance requirement on tests in this context. A running example throughout this Section is the problem of testing for an autoregressive unit root in a univariate time series. Section 3 provides details on the application of the main result for some standard tests. Section 4 concludes. All proofs are collected in an appendix.

2 Efficiency and Robustness under a Weak Convergence Assumption

2.1 Set-up and Main Result

This subsection formally introduces the set-up and contains the main result of this paper. The following notation and conventions are used throughout the paper: All limits are taken as $T \to \infty$. If S is a metric space with metric d_S, then $\mathfrak{B}(S)$ is its Borel σ-algebra. If μ is a probability measure on $\mathfrak{B}(S)$, then its image measure under the $\mathfrak{B}(S)\backslash\mathfrak{B}(U)$ measurable mapping $f : S \mapsto U$, where U is another space with metric d_U, 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 $S \times U$ is equipped with the metric $d_U + d_S$. 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 $\mathfrak{B}(S)$. 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 the 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 parameters $\theta \in \{\theta_0\} \cup \Theta$ and $\gamma \in \Gamma$, where Θ and Γ are metric spaces, so that the distribution $F_T(m, \gamma, \theta)$ of Y_T is a probability kernel. The set Γ may consist of a singleton. The hypotheses of interest are

$$H_0 : \theta = \theta_0 \qquad \text{against} \qquad H_1 : \theta \in \Theta. \tag{1}$$

Let ϕ_T be a sequence of Borel measurable functions $\phi_T : \mathbb{R}^{nT} \mapsto S$, where S is a metric space. Denote by $P_T(m, \gamma, \theta)$ the distribution of $X_T = \phi_T(Y_T)$ in model m with parameters γ and θ, that is $P_T(m, \gamma, \theta) = \phi_T F_T(m, \gamma, \theta)$. Suppose the typical model m satisfies the following convergences in distribution

$$P_T(m, \gamma, \theta) \rightsquigarrow P(\gamma, \theta) \;\; \text{pointwise for all } \theta \in \{\theta_0\} \cup \Theta_1 \text{ and } \gamma \in \Gamma$$

where, for each $\gamma \in \Gamma$, the probability measure $P(\gamma, \theta)$ on S is absolutely continuous with respect to $P(\gamma, \theta_0)$, with Radon-Nikodym derivative $\mathrm{LR} : \Gamma \times \Theta \times S \mapsto \mathbb{R}$, that is for each $A \in \mathfrak{B}(S)$, $\int_A dP(\gamma, \theta) = \int_A \mathrm{LR}(\gamma, \theta, x) dP(\gamma, \theta_0)(x)$. In addition, assume there exists an estimator $\hat{\gamma}_T : \mathbb{R}^{nT} \mapsto \Gamma$ of γ that is consistent for all values of $\gamma \in \Gamma$, irrespective of the value of $\theta \in \Theta$. Denoting the distribution of $(\hat{\gamma}_T, X_T) = \phi_T^e(Y_T)$ by $P_T^e(m, \gamma, \theta) = \phi_T^e F_T(m, \gamma, \theta)$ ('e' for extended), we might thus summarize the behavior of the typical model m under the null and alternative hypothesis as

$$H_0 \quad : \quad P_T^e(m, \gamma, \theta_0) \rightsquigarrow P^e(\gamma, \theta_0) \quad \text{pointwise for all } \gamma \in \Gamma \tag{2}$$

$$H_1 \quad : \quad P_T^e(m, \gamma, \theta) \rightsquigarrow P^e(\gamma, \theta) \quad \text{pointwise for all } \gamma \in \Gamma,\ \theta \in \Theta \tag{3}$$

where $P^e(\gamma, \theta)$ is the product measure between the degenerate probability measure on $\mathfrak{B}(\Gamma)$ that puts all mass on the point γ and the measure $P(\gamma, \theta)$ on $\mathfrak{B}(S)$.

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 $\rho_T = 1 - \theta/T$ for some fixed $\theta \geq 0$, and the hypotheses are $H_0 : \theta = 0$ against $H_1 : \theta > 0$. With $S = D_{[0,1]}$ the space of cadlag functions on the unit interval, equipped with the Billingsley metric, a typical model m for the disturbances $\nu_{T,t}$ satisfies $T^{-1/2} u_{T,\lfloor \cdot T \rfloor} \rightsquigarrow \omega J_\theta(\cdot)$ on $D_{[0,1]}$, where J_θ is an Ornstein-Uhlenbeck process $J_\theta(s) = \int_0^s e^{-\theta(s-r)} dW(r)$ with W a standard Wiener process, and $\gamma = \omega^2$ is the positive long-run variance of $\nu_{T,t}$. The probability measure of the Gaussian process ωJ_θ is absolutely continuous with respect to the measure of $\omega J_0 = \omega W$, and by Girsanov's Theorem, the Radon-Nikodym derivative equals $\mathrm{LR}(\omega^2, \theta, x) = \exp[-\frac{1}{2}\omega^2\theta(x(1)^2 - 1) - \frac{1}{2}\omega^2\theta^2 \int_0^1 x(s)^2 ds]$. Furthermore, let $\hat{\omega}_T^2$ be a specific, "reasonable" long-run variance estimator, which is consistent for ω^2 in the typical model m. The standard asymptotic implications of model m are thus summarized by $(T^{-1/2} u_{T,\lfloor \cdot T \rfloor}, \hat{\omega}_T^2) \rightsquigarrow (\omega J_\theta(\cdot), \omega^2)$. ▲

Possibly randomized tests of H_0 in (1) are measurable functions $\varphi_T : \mathbb{R}^{nT} \mapsto [0, 1]$, where $\varphi_T(Z_T)$ indicates the probability of rejection conditional on observing $Y_T = Z_T$, so that the overall rejection probability of the test φ_T in model m is given by $\int \varphi_T dF_T(m, \gamma, \theta)$. In many nonstandard problems, no uniformly most powerful test exists, so consider tests that maximize a weighted average power criterion

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

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_T for all choices for w.

Furthermore, define asymptotic null rejection probability of test φ_T in model m as

$$\text{ARP}_0(\varphi_T, m, \gamma) = \limsup_{T\to\infty} \int \varphi_T dF_T(m, \gamma, \theta_0).$$

With these definitions, asymptotically efficient level-α tests φ_T maximize $\lim_{T\to\infty}\text{WAP}_T(\varphi_T, m, \gamma)$ subject to $\text{ARP}_0(\varphi_T, m, \gamma) \leq \alpha$, for all γ. A reasonable definition of an asymptotically robust test is to impose that $\text{ARP}_0(\varphi_T, m, \gamma) \leq \alpha$ 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 $P_T^e(m, \gamma, \theta_0) = \phi_T^e F_T(m, \gamma, \theta_0) \rightsquigarrow P^e(\gamma, \theta_0)$. In this paper, we refer to a test as robust if it has asymptotic null rejection probability no larger than the nominal level for all models $m \in \mathcal{M}_0$, that is formally if

$$\text{ARP}_0(\varphi_T, m, \gamma) \leq \alpha \quad \text{for all } m \in \mathcal{M}_0 \text{ and } \gamma \in \Gamma. \tag{4}$$

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 $(T^{-1/2}u_{T,\lfloor \cdot T\rfloor}, \hat{\omega}_T^2) \rightsquigarrow (\omega J_\theta(\cdot), \omega^2)$—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

[1] For maximal generality, one could additionally index the weighting function w by γ, at the cost of more cumbersome notation.

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 $(T^{-1/2}u_{T,\lfloor \cdot T\rfloor}, \hat{\omega}_T^2) \rightsquigarrow (\omega W(\cdot), \omega^2)$. In fact, Stock (1994a), White (2001, p. 179), Breitung (2002), Davidson (2002, 2007) and Müller (2004) *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 (4) entirely natural for a unit root test. ▲

The main result of this paper is that imposing (4) renders the question about asymptotically efficient tests straightforward to answer:

Theorem 1 *For all* $\gamma \in \Gamma$, *let* $\overline{\mathrm{LR}}(\gamma, \cdot) = \int \mathrm{LR}(\gamma, \theta, \cdot) dw(\theta)$, *and suppose that (i)* $\overline{\mathrm{LR}}(\gamma, \cdot)$ *is* $P(\gamma, \theta_0)$*-almost everywhere continuous; (ii) for some functions* $\mathrm{cv} : \Gamma \mapsto \mathbb{R}$ *and* $p : \Gamma \mapsto [0,1]$, *the* $\Gamma \times S \mapsto [0,1]$ *function*

$$\varphi^*(\gamma_1, x) = \begin{cases} 1 \text{ if } \overline{\mathrm{LR}}(\gamma_1, x) > \mathrm{cv}(\gamma_1) \\ p(\gamma_1) \text{ if } \overline{\mathrm{LR}}(\gamma_1, x) = \mathrm{cv}(\gamma_1) \\ 0 \text{ if } \overline{\mathrm{LR}}(\gamma_1, x) < \mathrm{cv}(\gamma_1) \end{cases}$$

is $P^e(\gamma, \theta_0)$*-almost everywhere continuous and satisfies* $\int \varphi^*(\gamma, x) dP(\gamma, \theta_0)(x) = \alpha > 0$. *Then*

(i) $ARP_0(\hat{\varphi}_T^*, m) = \alpha$ *for all* $m \in \mathcal{M}_0$ *and* $\gamma \in \Gamma$, *and* $\lim_{T\to\infty} WAP_T(\hat{\varphi}_T^*, m, \gamma) = \int \varphi^* dP(\theta, \gamma) dw(\theta) = \beta(\gamma)$ *for all* $m \in \mathcal{M}_1$, *where* $\hat{\varphi}_T^* : \mathbb{R}^{nT} \mapsto [0,1]$ *is defined as* $\hat{\varphi}_T^*(y) = \varphi^*(\hat{\gamma}(y), \phi_T(y))$.

(ii) For any test φ_T *that satisfies (4),* $\limsup_{T\to\infty} WAP_T(\varphi_T, m, \gamma) \leq \beta(\gamma)$ *for all* $m \in \mathcal{M}_1$ *and* $\gamma \in \Gamma$.

Unit Root Test Example, ctd: Consider the construction of a point-optimal unit root test, so that w puts all mass at θ_1. Then $\overline{\mathrm{LR}}(\omega^2, x) = \exp[\frac{1}{2}\omega^2\theta_1(x(1)^2 - 1) - \frac{1}{2}\omega^2\theta_1^2 \int_0^1 x(s)^2 ds]$, which is continuous for all $x \in D_{[0,1]}$, and $\varphi^*(\omega^2, x) = \mathbf{1}[\overline{\mathrm{LR}}(\omega^2, x) > e^{\omega^2 c}]$, where c is chosen to satisfy $\int \varphi^*(\gamma, x) dP(\gamma, \theta_0)(x) = P(\overline{\mathrm{LR}}(1, W) > e^c) = \alpha$. Then $\varphi^* : (0, \infty) \times D_{[0,1]} \mapsto [0,1]$ is seen to be continuous at almost all realizations of (ω^2, W), and part (i) of Theorem 1 shows that the test $\hat{\varphi}_T^*(Y_T) = \varphi^*(\hat{\omega}_T^2, T^{-1/2}u_{T,\lfloor \cdot T\rfloor}) =$

$\mathbf{1}[\exp[-\frac{1}{2}\hat{\omega}_T^2\theta_1(T^{-1}u_{T,T}^2-1)-\frac{1}{2}\hat{\omega}_T^2\theta_1^2T^{-1}\int_0^1 u_{T,\lfloor sT\rfloor}^2 ds]>e^{\hat{\omega}_T^2 c}]$ has asymptotic null rejection probability equal to the nominal level and asymptotic weighted average power equal to $\beta=P(\overline{\mathrm{LR}}(1,J_{\theta_1})>e^c)$ for all models in $\mathcal{M}_0$ and $\mathcal{M}_1$, respectively, that is models that satisfy $(T^{-1/2}u_{T,\lfloor\cdot T\rfloor},\hat{\omega}_T^2)\rightsquigarrow(\omega W(\cdot),\omega^2)$ and $(T^{-1/2}u_{T,\lfloor\cdot T\rfloor},\hat{\omega}_T^2)\rightsquigarrow(\omega J_\theta(\cdot),\omega^2)$. 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 $(T^{-1/2}u_{T,\lfloor\cdot T\rfloor},\hat{\omega}_T^2)\rightsquigarrow(\omega J_{\theta_1}(\cdot),\omega^2)$, there exists a model m satisfying $(T^{-1/2}u_{T,\lfloor\cdot T\rfloor},\hat{\omega}_T^2)\rightsquigarrow(\omega W(\cdot),\omega^2)$ 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 the statistic by ERS necessarily implies violation of the robustness condition (4). 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 $\mathcal{M}_0$. ▲

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 the proof of part (ii), consider the case where w is degenerate with all weight on θ_1, Γ is a singleton $\Gamma=\{\gamma\}$, and $\overline{\mathrm{LR}}^i(x)=1/\overline{\mathrm{LR}}(\gamma,x)=1/\mathrm{LR}(\gamma,\theta_1,x)$ is continuous and bounded on S. The idea is to take the model $m\in\mathcal{M}_1$, and to reweigh the probabilities according to the Radon-Nikodym derivative of the limiting random elements, i.e. according to $\overline{\mathrm{LR}}^i\circ\phi$, to construct a corresponding model in $\mathcal{M}_0$. This reweighed probability distribution needs to integrate to one, so let $\kappa_T=\int(\overline{\mathrm{LR}}^i\circ\phi)dF_T(m,\gamma,\theta_1)=\int\overline{\mathrm{LR}}^i dP_T(m,\gamma,\theta_1)$, and define the measure G_T on $\mathbb{R}^{nT}$ via $\int_A dG_T=\kappa_T^{-1}\int_A(\overline{\mathrm{LR}}^i\circ\phi)dF_T(m,\gamma,\theta_1)$ for all $A\in\mathfrak{B}(\mathbb{R}^{nT})$. By construction, G_T induces the measure Q_T on S, where Q_T satisfies $\int\vartheta dQ_T=\kappa_T^{-1}\int\vartheta\overline{\mathrm{LR}}^i dP_T(m,\gamma,\theta_1)$ for any continuous function $\vartheta:S\mapsto\mathbb{R}$. Further, the $S\mapsto\mathbb{R}$ functions $\vartheta\overline{\mathrm{LR}}^i$ and $\overline{\mathrm{LR}}^i$ are bounded and continuous, so that $P_T(m,\gamma,\theta_1)\rightsquigarrow P(\gamma,\theta_1)$ implies $\kappa_T\to\int\overline{\mathrm{LR}}^i dP(\gamma,\theta_1)=\int\overline{\mathrm{LR}}^i\cdot\overline{\mathrm{LR}}(\gamma,x)dP(\gamma,\theta_0)(x)=\int dP(\gamma,\theta_0)=1$ and $\int\vartheta\overline{\mathrm{LR}}^i dP_T(m,\gamma,\theta_1)\to\int\vartheta\overline{\mathrm{LR}}^i dP(\gamma,\theta_1)=\int\vartheta dP(\gamma,\theta_0)$, so that

$\phi G_T \rightsquigarrow P(\gamma, \theta_0)$, and G_T is in $\mathcal{M}_0$. Thus, by (4), $\limsup_{T\to\infty} \int \varphi_T dG_T \leq \alpha$. Furthermore, by construction, the Radon-Nikodym derivative between G_T and $F_T(m, \gamma, \theta_1)$ is given by $\kappa_T \overline{\mathrm{LR}}(\gamma, \phi_T(Y_T))$. Therefore, by the Neyman-Pearson Lemma, the best test of $\tilde{H}_0 : Y_T \sim G_T$ against $\tilde{H}_1 : Y_T \sim F_T(m, \gamma, \theta_1)$ rejects for large values of $\overline{\mathrm{LR}}(\gamma, \phi_T(Y_T))$, 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 $\overline{\mathrm{LR}}(\gamma, \phi_T(Y_T))$ and has the same asymptotic rejection probability, and the result follows.

Discussion of Theorem 1

1. Applications of the Central Limit Theorem and the Functional Central Limit Theorem lead, of course, to Gaussian limiting distributions. So typically, the limiting measures $P(\gamma, \theta_0)$ and $P(\gamma, \theta)$ of X_T are Gaussian, and the Radon-Nikodym derivative LR is the limit of the small sample likelihood ratio statistics in the Gaussian version of the model, i.e. when X_T is normally distributed for each T. This arises, for instance, when ϕ_T is linear and Y_T is normally distributed, as in the unit root example with $\nu_{T,t}$ i.i.d. mean zero normal. In this case, $\hat{\varphi}_T^*$ is asymptotically equivalent to the optimal test that is derived under a Gaussianity assumption, and part (ii) of Theorem 1 becomes a statement about the limits of adaption: Any attempt to adapt to the potentially non-Gaussian nature of the underlying disturbances, if successful, necessarily means that there is a model $m \in \mathcal{M}_0$ for which the test has asymptotic rejection probability larger than the nominal level. In this sense, adaption is always costly in terms of the robustness properties of the resulting test.

2. No matter how one views the desirability of the robustness constraint (4), one appeal of the framework here 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), or the weak instrument problem in a general GMM framework, as considered by Stock and Wright (2000), arguably fall into this class. The Radon-Nikodym derivative between the limiting measures under the local alternative and the null hypothesis,

evaluated at the sample analogues, then is a natural starting point for a reasonable test statistic.

3. The test $\varphi^*(\gamma, \cdot)$ is the Neyman-Pearson test of $H_0 : X \sim P(\gamma, \theta_0)$ against $H_1 : X \sim \int P(\gamma, \theta)dw(\theta)$, and Theorem 1 shows a tight link between this limiting problem and the achievable asymptotic weighted average power in the hypothesis testing problem (1) concerning Y_T. These results assume $P(\gamma, \theta)$ to be absolutely continuous with respect to $P(\gamma, \theta_0)$. If instead $P(\gamma, \theta)$ is singular with respect to $P(\gamma, \theta_0)$, then the liming problem becomes trivial, and one might imagine that this must imply the existence of a consistent test φ_T. But this is not always the case: Consider the problem of discriminating between an I(0) series and an I(1) series with a scale invariant procedure, that is in the same set-up as in the unit root example, assume that $\mathcal{M}_0$ and $\mathcal{M}_1$ contain all models satisfying $\sum_{t=1}^{\lfloor \cdot T \rfloor} u_{T,t} / \sum_{t=1}^{T} u_{T,t} \rightsquigarrow W(\cdot)/W(1)$ and $\sum_{t=1}^{\lfloor \cdot T \rfloor} u_{T,t} / \sum_{t=1}^{T} u_{T,t} \rightsquigarrow \int_0^{\cdot} W(s)ds / \int_0^1 W(s)ds$, respectively. The implied limiting measures are singular ($W(\cdot)/W(1)$ is not differentiable with probability one, while $\int_0^{\cdot} W(s)ds / \int_0^1 W(s)ds$ is), yet Müller (2004) shows that there does not exist a consistent test satisfying the robustness requirement (4). At the same time, there does exist a consistent test with the null and alternative hypothesis reversed, so one cannot draw any general conclusions about the existence of a consistent test from the observation that the limiting problem is trivial.

4. The importance of the efficiency property of $\hat{\varphi}_T^*$ depends crucially on the appropriateness and desirability of the robustness constraint (4). One might think about the relative gain in robustness of tests satisfying (4) 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 (4) 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 billion data points, 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 (4) 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.

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. Of course, interesting high level weak convergence assumptions are not entirely arbitrary, but derive their plausibility from the knowledge that there exists a range of underlying primitive conditions that would imply them. In general, weaker regularity conditions (that is fewer weak convergence assumptions) lead to less powerful inference, and Theorem 1 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 (4).

Low Frequency Unit Root Test Example: Müller and Watson (2006) 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 $T^{-1/2}u_{T,\lfloor\cdot T\rfloor} \rightsquigarrow \omega J_\theta(\cdot)$, but Müller and Watson (2006) derive a scale invariant point-optimal unit root test that only assumes a finite subset of this convergence, that is

$$\left\{T^{-3/2}\sum_{t=1}^{T}\psi_l(t/T)u_{T,t}\right\}_{l=1}^{q} \rightsquigarrow \left\{\omega\int_0^1 \psi_l(s)J_\theta(s)ds\right\}_{l=1}^{q}, \tag{5}$$

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. But (5) is strictly weaker than the standard assumption$T^{-1/2}u_{T,\lfloor\cdot T\rfloor} \rightsquigarrow \omega J_\theta(\cdot)$, and accordingly, Müller and Watson's (2006) 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 (5): Theorem 1 implies that it is impossible to let the data decide whether (5) 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 I(1) model satisfying (5). ▲

2.2 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. While this is standard practice in much of econometric theory, it does not ensure that there for large enough T, the null rejection probability $\int \varphi_T dF_T(m, \gamma, \theta_0)$ is close to α for all data generating processes under consideration. In fact, it is not hard to see 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 $F_T(m, \gamma, \theta_0)$ of Y_T is entirely unrestricted, as the convergence $P_T^e(m, \gamma, \theta_0) = \phi_T^e F_T(m, \gamma, \theta_0) \rightsquigarrow P^e(\gamma, \theta_0)$ can occur 'later'.

To generate uniform results, we thus reduce the set of models $\mathcal{M}_0$ and impose a lower limit on the speed of convergence. Let Δ be a metric that metrizes weak convergence of probability measures on $\mathfrak{B}(\Gamma \times S)$, and let $\delta : \Gamma \times \mathbb{N} \mapsto \mathbb{R}$ be such that for all $\gamma \in \Gamma$, $\lim_{T\to\infty} \delta(\gamma, T) = 0$. Now define $\mathcal{M}_0^u(\delta)$ ('u' for uniform) as the set of models m satisfying

$$\Delta(\phi^e F_T(m, \gamma, \theta_0), P^e(\gamma, \theta_0)) \leq \delta(\gamma, T),$$

that is $\mathcal{M}_0^u(\delta)$ is the collection of models m for which the distribution $P_T^e(m, \gamma, \theta_0)$ of $\phi_T^e(Y_T) = (\hat{\gamma}_T, X_T)$ differs by at most $\delta(\gamma, T)$ from its limit $P^e(\gamma, \theta_0)$ as measured by Δ. It then makes sense to ask whether the rejection probability of a test φ_T converges to the nominal level uniformly over $\mathcal{M}_0^u(\delta)$, that is if[2]

$$\lim_{T\to\infty} \sup_{m\in\mathcal{M}_0^u(\delta)} \int \varphi_T dF_T(m, \gamma, \theta_0) \leq \alpha \quad \text{ for all } \gamma \in \Gamma. \tag{6}$$

By the definition of weak convergence and the continuity of φ^*, (6) holds for the test $\varphi_T = \hat{\varphi}_T^*$ under the conditions of Theorem 1, i.e. for large enough T, the rejection probability of $\hat{\varphi}_T^*$ is close to α for all models in $\mathcal{M}_0^u(\delta)$.

It is not clear whether all tests that satisfy the point-wise robustness (4) also satisfy (6), or vice versa. Theorem 1 therefore does not imply that $\hat{\varphi}_T^*$ also maximizes asymptotic weighted average power in the class of all tests that satisfy (6). To make further progress, we restrict attention to the specific metric $\Delta = \Delta_{BL}$: For a separable metric space U with metric d_U and two probability measures μ_1 and μ_2 on $\mathfrak{B}(U)$, Δ_{BL} is defined as $\Delta_{BL}(\mu_1, \mu_2) = \sup_{||f||_{BL}\leq 1} |\int f d\mu_1 - \int f d\mu_2|$, where $f : U \mapsto \mathbb{R}$ are $\mathfrak{B}(U)\backslash\mathfrak{B}(\mathbb{R})$ measurable and $||f||_{BL} = \sup_{x\in U} |f(x)| + \sup_{x,y\in U} \frac{|f(x)-f(y)|}{d_U(x,y)}$. It is known that Δ_{BL}

[2] The uniformity here is over the models $m \in \mathcal{M}_0^u(\delta)$, but it is still a pointwise statement in Γ, since $\lim_{n\to\infty} \delta(\gamma, n) = 0$ for all $\gamma \in \Gamma$ does, of course, not imply that $\lim_{n\to\infty} \sup_{\gamma\in\Gamma} \delta(\gamma, n) = 0$. Uniformity over parameters is a well-studied problem, and in order to isolate the less standard issue of uniformity over models defined in terms of a weak convergence property, we do not impose an additional unformity over Γ.

metrizes weak convergence on separable metric spaces (Dudley (2002, p. 395)). A partial answer to the question of efficiency of $\hat{\varphi}_T^*$ in the class of tests satisfying (6) is provided by the following Theorem.

Theorem 2 *Suppose that S and Γ are separable. Under the assumptions of Theorem 1, pick $\gamma \in \Gamma$, and suppose that for all $\varepsilon > 0$ there exists an open set $D \in \mathfrak{B}(S)$ with $P(\gamma, \theta_0)(D) > 1 - \varepsilon$ so that the $D \mapsto \mathbb{R}$ function $x \mapsto \overline{\mathrm{LR}}(\gamma, x)$ is Lipschitz, and assume that the models m_0 and m_1 are such that $\Delta_{BL}(\phi_T^e F_T(m_0, \gamma, \theta_0), P^e(\gamma, \theta_0))/\delta(\gamma, T) \to 0$ and $\Delta_{BL}(\int \phi_T^e F_T(m_1, \gamma, \theta) dw(\theta), \int P^e(\gamma, \theta) dw(\theta))/\delta(\gamma, T) \to 0$. Then for any test φ_T that satisfies (6) with $\Delta = \Delta_{BL}$, $\limsup_{T\to\infty} WAP_T(\varphi_T, m_1, \gamma) \leq \lim_{T\to\infty} WAP_T(\hat{\varphi}_T^*, m_1, \gamma) = \beta(\gamma)$.*

Under a stronger continuity assumption for $\overline{\mathrm{LR}}$, Theorem 2 shows that no test can satisfy (6) and have higher asymptotic weighted average power than $\hat{\varphi}_T^*$ for alternative models whose weak convergence is faster than the lower bound $\delta(\gamma, T)$, as measured by Δ_{BL}. In other words, $\hat{\varphi}_T^*$ is the efficient test in the class of test satisfying (6) against any set of alternative models for which the mixture $\int \phi_T^e F_T(m_1, \gamma, \theta) dw(\theta) \rightsquigarrow \int P^e(\gamma, \theta) dw(\theta)$ becomes a good approximation faster than the slowest convergence $\phi_T^e F_T(m_1, \gamma, \theta_0) \rightsquigarrow P^e(\gamma, \theta_0)$ for which the test controls asymptotic size uniformly. The proof of Theorem 2 is similar to the proof of Theorem 1 and exploits the linearity in both the definition of Δ_{BL} and the construction of the reweighed alternative model G_T. 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 2 also holds for Δ_P when the models m_0 and m_1 are such that $\Delta_P(\phi_T^e F_T(m_0, \gamma, \theta_0), P^e(\gamma, \theta_0))/\delta(\gamma, T)^2 \to 0$ and $\Delta_P(\int \phi_T^e F_T(m_1, \gamma, \theta) dw(\theta), \int P^e(\gamma, \theta) dw(\theta))/\delta(\gamma, T)^2 \to 0$ for all $\theta \in \Theta$.

2.3 Invariance

In many models, it is natural to assume a weak convergence property on some model component that is not directly observed because it depends on additional parameters $\xi \in \Xi$. But in the set-up of Section 2.1, $X_T = \phi_T(Y_T)$ is a function of observables, so that natural choices for X_T are estimates of the unobserved component. The efficiency statement of Theorem 1 is then relative to all tests that are robust over all models whose estimated unobserved component satisfies the weak convergence property.

Unit Root Test Example, ctd: Consider the problem of testing for a unit root in a model with an unknown constant ξ, so that we observe $Y_T = (y_{T,1}, \cdots, y_{T,T})'$ with $y_{T,t} = u_{T,t} + \xi$. Assume for simplicity that $\omega = 1$ is known. As in the case without deterministics, it would be natural to assume $T^{-1/2}u_{T,\lfloor \cdot T \rfloor} \rightsquigarrow J_\theta(\cdot)$. With ξ unknown, however, $T^{-1/2}u_{T,\lfloor \cdot T \rfloor}$ is unobserved. Replacing ξ with the estimator $\hat{\xi}_T = T^{-1}\sum_{t=1}^T y_{T,t}$, we obtain $\phi_T^\mu(Y_T) = T^{-1/2}(y_{T,\lfloor \cdot T \rfloor} - \hat{\xi}_T) \rightsquigarrow J_\theta^\mu(\cdot)$, where $J_\theta^\mu(s) = J_\theta(s) - \int_0^1 J_\theta(r)dr$. Defining $\mathcal{M}_0$ to be the set of models m satisfying $T^{-1/2}(y_{T,\lfloor \cdot T \rfloor} - \hat{\xi}_T) \rightsquigarrow J_0^\mu(\cdot)$, Theorem 1 shows that rejecting for large values of $\mathrm{LR}^\mu(\theta_1, \phi_T^\mu(Y_T))$ is asymptotically efficient among all tests whose asymptotic null rejection probability is at most α for all models $m \in \mathcal{M}_0$, where $\mathrm{LR}^\mu(\theta_1, \cdot)$ is the Radon-Nikodym derivative of the probability measure of $J_{\theta_1}^\mu$ with respect to the measure of J_0^μ.[3] ▲

While this approach is quite general and leads to a well-defined efficiency statement, it is interesting to explore whether an efficiency statement can be made when the weak convergence property is assumed for the unobservable component directly. This section shows that this is sometimes possible by drawing on the theory of invariance in hypothesis testing.

Specifically, with the additional nuisance parameter $\xi \in \Xi$, the distribution of Y_T is now given by $F_T(m, \gamma, \theta, \xi)$. Assume that for the typical model m, for any $\gamma \in \Gamma$ there exists a sequence $\xi_T \in \Xi$ such that $\phi_T F_T(m, \gamma, \theta_0, \xi_T) \rightsquigarrow P^i(\gamma, \theta_0)$ ('i' for ideal). Denote by $\mathcal{M}_0^i$ the set of all such models, and consider tests that satisfy

$$\limsup_{T\to\infty} \int \varphi_T dF_T(m, \gamma, \theta_0, \xi_T) \le \alpha \text{ for all } m \in \mathcal{M}_0^i \text{ and } \gamma \in \Gamma, \quad (7)$$

and define the set of alternative models $\mathcal{M}_1^i$ analogously. Note that in general, the limiting distributions $P^i(\gamma, \theta_0)$ and $P^i(\gamma, \theta)$ cannot directly serve as the basis for the

[3] Under the assumption of $u_{T,0} = 0$ for the initial condition, $\mathrm{LR}^\mu(\theta_1, \phi_T^\mu(Y_T)) = \mathrm{LR}(\theta_1, \varsigma(\phi_T^\mu(Y_T)))$, where $\varsigma : D_{[0,1]} \mapsto D_{[0,1]}$ is defined as $\varsigma(x) = x(\cdot) - x(0)$ and $\mathrm{LR}(\theta_1, \cdot)$ is the Randon-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 $\mathrm{LR}^\mu(\theta_1, \phi_T^\mu(Y_T))$ leads to the same asymptotic power than in the model with $\xi = 0$ known. This equivalence ceases to hold, however, when the initial condition $u_{T,0}$ is of the same order of magnitude than $u_{T,[sT]}$ for $s > 0$. See Müller and Elliott (2003) for discussion.

construction of tests, since the distribution of $\phi_T(Y_T)$ depends on the unknown parameter ξ.

For each T, suppose there is a group of functions $\mathcal{G}_T$ indexed by $r \in R$, such that for each $r \in R$, $g_r \in \mathcal{G}_T$ maps $\mathbb{R}^{nT}$ to $\mathbb{R}^{nT}$. Let $h_T : \mathbb{R}^{nT} \mapsto \mathbb{R}^{nT}$ be a maximal invariant of $\mathcal{G}_T$ which selects a specific orbit, i.e. for all $y \in \mathbb{R}^{nT}$ there exists $g_r \in \mathcal{G}_T$ so that $h_T(y) = g_r(y)$, and assume that the distribution $h_T F_T(m, \gamma, \theta, \xi)$ of $h_T(Y_T)$ does not depend on ξ for all $\gamma \in \Gamma$ and $\theta \in \{\theta_0\} \cup \Theta$—by Theorem 3 on p. 292 of Lehmann (1986), this is always the case when $\mathcal{G}_T$ induces a transitive group on the parameter space Ξ, i.e. if $g_r F_T(m, \gamma, \theta_0, \xi) = F_T(m, \gamma, \theta_0, \bar{g}_r(\xi))$ for some $\bar{g}_r : \Xi \mapsto \Xi$ and for all $\xi_1, \xi_2 \in \Xi$, there is an $r \in R$ such that $\xi_1 = \bar{g}_r(\xi_2)$. Denote by $P_T^h(m, \gamma, \theta)$ the distribution of $\phi_T(h_T(Y_T))$, and assume that in the typical model m, $P_T^h(m, \gamma, \theta) \rightsquigarrow P^h(\gamma, \theta)$ for all $\gamma \in \Gamma$, $\theta \in \Theta \cup \{\theta_0\}$. Denote by $\mathcal{M}_0^h$ and $\mathcal{M}_1^h$ the set of all models where, for all $\gamma \in \Gamma$ and $\xi \in \Xi$, $P_T^h(m, \gamma, \theta_0, \xi) \rightsquigarrow P^h(\gamma, \theta_0)$ and $P_T^h(m, \gamma, \theta, \xi) \rightsquigarrow P^h(\gamma, \theta)$ for all $\theta \in \Theta$, respectively. Assuming that the Radon-Nikodym derivative of $P^h(\gamma, \theta)$ with respect to $P^h(\gamma, \theta_0)$ satisfies the required assumption, Theorem 1 is applicable and provides for a given weighting function the asymptotic weighted average power maximizing test among all tests that have correct asymptotic null rejection probability for all models $\mathcal{M}_0^h$. Note that this test—say, $\hat{\varphi}_T^h$—is a function of $\phi_T(h_T(Y_T))$, and thus invariant with respect to $\mathcal{G}_T$, so that $\hat{\varphi}_T^h$ is trivially also the best invariant test among such tests. Furthermore, all models in $\mathcal{M}_j^i$ are also in $\mathcal{M}_j^h$ for $j = 0, 1$, so that $\hat{\varphi}_T^h$ also satisfies (7). It is tempting to conclude that $\hat{\varphi}_T^h$ also maximizes weighted average power in the class of all invariant tests that satisfy (7). This is not obvious, though, since not all models in $\mathcal{M}_0^h$ are necessarily also in $\mathcal{M}_0^i$, so that $\hat{\varphi}_T^h$ is optimal with respect to an asymptotic size constraint over a larger set of null models.

Unit Root Test Example, ctd: With $\xi = \xi_T = 0$ and $\phi_T : \mathbb{R}^T \mapsto D_{[0,1]}$ defined as $\phi_T(y) = T^{-1/2} y_{\lfloor \cdot T \rfloor}$, we have $\phi_T(Y_T) = T^{-1/2} u_{T, \lfloor \cdot T \rfloor}$, so $\mathcal{M}_0^i$ contains all models with $T^{-1/2} u_{T, \lfloor \cdot T \rfloor} \rightsquigarrow J_0(\cdot)$. Let $R = \mathbb{R}$, define $g_r \in \mathcal{G}_T$ as $g_r(y) = y + rT^{1/2}$ and let $h_T(Y_T) = (y_{T,1} - \bar{y}_T, \cdots, y_{T,T} - \bar{y}_T)' = g_{-\bar{y}_T}(Y_T)$ with $\bar{y}_T = T^{-1} \sum_{t=1}^T y_{T,t}$. Then $\mathcal{M}_0^h$ is the set of all models satisfying $\phi_T(h_T(Y_T)) = \phi_T^\mu(Y_T) = T^{-1/2}(u_{T, \lfloor \cdot T \rfloor} - T^{-1} \sum_{t=1}^T u_{T,t}) \rightsquigarrow J_0^\mu(\cdot)$. Note that this is a strictly larger set than the set of models for which $T^{-1/2} u_{T, \lfloor \cdot T \rfloor} \rightsquigarrow J_0(\cdot)$. The efficiency of $\mathrm{LR}^\mu(\theta_1, \phi_T^\mu(Y_T))$ noted above is thus not sufficient to conclude that rejecting for large values of $\mathrm{LR}^\mu(\theta_1, \phi_T^\mu(Y_T))$ is the asymptotically point-optimal translation invariant test with correct asymptotic rejection probability whenever $T^{-1/2}u_{T,\lfloor \cdot T\rfloor} \rightsquigarrow J_0(\cdot)$. ▲

It would be possible to draw the desired conclusion if any invariant test that satisfies (7) necessarily also has correct asymptotic null rejection probability for all models in $\mathcal{M}_0^h$, i.e. if for invariant tests, (7) implies $\limsup_{T\to\infty} \int \varphi_T dF_T(m, \gamma, \theta_0, \xi) \leq \alpha$ for all $m \in \mathcal{M}_0^h$ and $\gamma \in \Gamma$. The following Theorem provides general conditions under which this is the case.

Theorem 3 *Suppose (i) R and S are complete and separable metric spaces; (ii) there exists a Borel-measurable function $\tilde{h} : S \mapsto S$ such that $P^h(\gamma, \theta_0) = \tilde{h}P^i(\gamma, \theta_0)$; (iii) for all $r \in R$, there exists a continuous function $\tilde{g}_r : S \mapsto S$ such that $\sup_{y\in\mathbb{R}^{nT}} d_S(\phi_T(g_r(y)), \tilde{g}_r(\phi_T(y))) \to 0$, where d_S is the metric on S; (iv) there exists a Borel-measurable function $\rho : S \mapsto R$ such that $x = \tilde{g}_{\rho(x)}(\tilde{h}(x))$ for all $x \in S$. If a test φ_T is invariant with respect to $\mathcal{G}_T$ and satisfies (7), then* $\limsup_{T\to\infty} \int \varphi_T dF_T(m, \gamma, \theta_0, \xi) \leq$ *α for all $m \in \mathcal{M}_0^h$, $\gamma \in \Gamma$ and $\xi \in \Xi$.*

Unit Root Test Example, ctd: With $\tilde{h} : D_{[0,1]} \mapsto D_{[0,1]}$ defined as $\tilde{h}(x) = x(\cdot) - \int_0^1 x(s)ds$, $\tilde{g}_r : D_{[0,1]} \mapsto D_{[0,1]}$ defined as $\tilde{g}_r(x) = x(\cdot) + r$ and $\rho : D_{[0,1]} \mapsto \mathbb{R}$ defined as $\rho(x) = \int_0^1 x(s)ds$, we find $J_0^\mu = \tilde{h}(J_0)$ a.s., $\sup_{y\in\mathbb{R}^T} \sup_{s\in[0,1]} |T^{-1/2}(y_{\lfloor sT\rfloor} + T^{1/2}r) - (T^{-1/2}y_{\lfloor sT\rfloor} + r)| = 0$, and $x = \tilde{h}(x) + \rho(x) = \tilde{g}_{\rho(x)}(\tilde{h}(x))$ for all $x \in D_{[0,1]}$, so that the assumptions of Theorem 3 hold. We can therefore conclude that for any translation invariant unit root test φ_T with asymptotic null rejection probability of at most α whenever $T^{1/2}u_{T,\lfloor \cdot T\rfloor} \rightsquigarrow J_0(\cdot)$ also has asymptotic null rejection probability of at most α whenever $\phi_T^\mu(Y_T) \rightsquigarrow J_0^\mu$. As noted above, the best test in the latter class rejects for large values of $\mathrm{LR}^\mu(\theta_1, \phi_T^\mu(Y_T))$, so that this test is also the asymptotically point-optimal translation invariant test. ▲

For the proof of Theorem 3, note that with $x = \tilde{g}_{\rho(x)}(\tilde{h}(x))$ for all $x \in S$, one can construct the distribution $P^i(\gamma, \theta_0)$ by applying an appropriate *random* transformation $\tilde{g}_r$ to each x drawn under $P^h(\gamma, \theta_0)$—in the unit root example, the appropriate r is distributed as $J_0^\mu(0)$, since $J_0(s) = J_0^\mu(s) - J_0^\mu(0)$ a.s. The assumptions of Theorem 3 are now sufficient to ensure a tight enough link between this construction for $P^i(\gamma, \theta_0)$ and the limit of the small sample analogously transformed $\mathcal{M}_0^h$ models. For each model in $\mathcal{M}_0^h$, one can thus construct a corresponding model in $\mathcal{M}_0^i$ by applying an appropriate random transformation $g_r \in \mathcal{G}_T$ for each T. But the rejection probability of invariant tests, by definition, do not change by such transformation, and the result follows.

A related, but slightly different application of Theorem 3 concerns the elimination of nuisance parameters from the limiting problem. In general, the limiting measures of $\phi_T(Y_T)$ under the null and local alternatives may depend on nuisance parameters that cannot be consistently estimated. If these parameters can be eliminated by an appeal to invariance with maximal invariant $\tilde{h} : S \mapsto S$, then one can straightforwardly invoke Theorem 1 to $\phi_T^h(Y_T) = \tilde{h}(\phi_T(Y_T))$, whose limiting distribution, by construction, does not depend on such nuisance parameters. But as above, this way of proceeding does not guarantee that the resulting test is also the best invariant test in the original problem, and Theorem 3 provides conditions under which it is.

Low Frequency Unit Root Test Example, ctd: The weak convergence (5) to the q dimensional multivariate limiting random variable X in $X_T = \phi_T(Y_T) = (T^{-3/2}\sum_{t=1}^T \psi_1(t/T)u_{T,t}, \cdots, T^{-3/2}\sum_{t=1}^T \psi_q(t/T)u_{T,t})' \rightsquigarrow X = (\omega\int_0^1 \psi_1(s)J_\theta(s)ds, \cdots, \omega\int_0^1 \psi_q(s)J_\theta(s)ds)'$ cannot directly serve as a basis for the construction of a point-optimal unit root test, since ω is unknown and cannot be consistently estimated without imposing additional regularity. However, with $\tilde{h}(X) = X/\sqrt{X'X}$ recognized as a maximal invariant to the scale transformations $X \to cX$ for $c \neq 0$, the point-optimal scale invariant test in the limiting problem simply rejects for large values of the Radon-Nikodym derivative LR^h of the measures of $\tilde{h}(X)$ under the local-to-unity alternative with $\theta = \theta_1$ and the null with $\theta = 0$. By Theorem 1, this test is efficient in the class of unit root tests with correct asymptotic null rejection probability whenever $\tilde{h}(\phi_T(Y_T)) \rightsquigarrow \tilde{h}(X)$ with $\theta = 0$. In addition, by Theorem 3, this test is also the point-optimal scale invariant unit root test with correct asymptotic rejection probability whenever $\phi_T(Y_T) \rightsquigarrow X$ with $\theta = 0$. ▲

3 Applications

The results discussed in the last section imply that a number of standard inference procedures are efficient in the class of all tests with correct null asymptotic rejection probability for all models that satisfy a certain weak convergence property. This section provides the details of this claim for three problems other than the univariate unit root testing problem discussed above[4]; specifically, we consider Elliott and Jansson's (2003) point-optimal tests for unit roots with stationary covariates, the standard Generalized Method of Moments (GMM) Wald tests and tests for parameter stability in a GMM framework as considered by Sowell (1996). It is not hard to see that the results of this paper also strengthen the optimality of Jansson's (2004, 2005) point-optimal stationarity tests and tests for the null hypothesis of cointegration, respectively, which are optimal by construction for i.i.d. Gaussian disturbances. We omit details for brevity. In addition, the results of this paper are useful for making efficiency claims about non-standard econometric methods that have a high-level weak convergence assumption as their starting point; see Müller and Watson (2006) and Ibragimov and Müller (2007).

3.1 Efficient Unit Root Tests with Stationary Covariates

Elliott and Jansson (2003) consider the model

$$\begin{pmatrix} y_{T,t} \\ x_{T,t} \end{pmatrix} = \begin{pmatrix} \alpha_y + \beta_y t \\ \alpha_x + \beta_x t \end{pmatrix} + \begin{pmatrix} u_{T,t} \\ \nu^x_{T,t} \end{pmatrix} \tag{8}$$

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, $u_{T,0} = O_p(1)$ for all T, $\rho_T = 1 - \theta/T$ for some fixed $\theta \geq 0$, 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 $y_{T,t}$, $H_0 : \theta = 0$ against the alternative $H_1 : \theta > 0$. Consider first the case with $\alpha_y = \alpha_x = \beta_y = \beta_x = 0$ known. The approach of Elliott and Jansson (2003) is to

[4] The efficiency claim about ERS's point-optimal statistic in the case with an unknown mean is easily generalized to the case with unknown time trend, and to alternative assumptions about the initial condition, as considered by Elliott (1999) and Elliott and Müller (2003).

first apply the Neyman-Pearson Lemma to determine, for each T, the point-optimal test against $\theta = \theta_1$ 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 Section 2 of this paper, we need to consider the typical weak convergence properties of model (8). Standard weak dependence assumptions on $\nu_{T,t}$ imply for some suitable long-run covariance matrix estimator $\hat{\Omega}_T$

$$\hat{\Omega}_T \xrightarrow{p} \Omega \text{ and } G_T(\cdot) = \begin{pmatrix} T^{-1/2} u_{T,\lfloor \cdot T \rfloor} \\ T^{-1/2} \sum_{t=1}^{\lfloor \cdot T \rfloor} \nu^x_{T,t} \end{pmatrix} \rightsquigarrow G^\theta(\cdot) \tag{9}$$

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

$$\begin{aligned} \mathrm{LR}(\Omega, \theta, G) &= \exp\left[-\theta \int_0^1 G(s)' S_1 \Omega^{-1} dG(s) - \tfrac{1}{2}\theta^2 \int_0^1 G(s)' S_1 \Omega^{-1} S_1 G(s) ds\right] \qquad (10) \\ &= \exp\left[-\tfrac{1}{2}\frac{\theta}{\omega_{yy}}(G_y(1)^2 - 1) - \theta\omega_{yx} \int_0^1 G_y(s) dG_x(s) - \tfrac{1}{2}\frac{\theta^2}{\omega_{yy}^2} \int_0^1 G_y(s)^2 ds\right] \end{aligned}$$

where S_1 is the $n \times n$ matrix $S_1 = \mathrm{diag}(1, 0, \cdots, 0)$ and the first row of Ω^{-1} is $(\omega_{yy}, \omega_{yx})$. 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

$$U_T = T^{-1} \sum_{t=2}^{T} u_{T,t-1} \nu^x_{T,t} - \hat{\Sigma}_T \rightsquigarrow U^\theta = \int_0^1 G^\theta_y(s) dG^\theta_x(s) \tag{11}$$

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 (9). Clearly, the Radon-Nikodym derivative of the measure of (G^θ, U^θ) with respect to the measure of (G^0, U^0), evaluated at G^0, is also given by $\mathrm{LR}(\Omega, \theta, G^0)$ in (10), and one can write $\mathrm{LR}(\Omega, \theta, G^0) = \mathrm{LR}^U(\Omega, \theta, G^0, U^0)$

for a continuous function LR^U. Theorem 1 thus applies and shows that rejecting for large values of $\mathrm{LR}^U(\hat{\Omega}_T, \theta_1, G_T, U_T)$ is the point-optimal unit root test for the alternative $\theta = \theta_1 > 0$ among all tests that have correct asymptotic null rejection probabilities whenever (9) and (11) hold.

In fact, the feasible tests suggested by Elliott and Jansson (2003) are asymptotically equivalent to a test that rejects for large values of $\mathrm{LR}^U(\hat{\Omega}_T, \theta_1, G_T, U_T)$.[5] This equivalence is no coincidence: i.i.d. Gaussian $\nu_{T,t}$ obviously satisfy $G_T \rightsquigarrow G^\theta$ for $\theta \in \{0, \theta_1\}$, and if the small sample likelihood ratio statistics converge jointly to some limiting random variable with unit expectation for $\theta = 0$, then contiguity implies that this limiting random variable is the Radon-Nikodym derivative of the distribution of G^{θ_1} with respect to the distribution of G^0 (see, for instance, Lemma 27 of Pollard (2001)). The method above, which starts with the Radon-Nikodym derivative directly, is arguably a more straightforward way of determining a test in this equivalence class. But the much more important insight, of course, concerns the optimality properties of this test: While Elliott and Jansson (2003) could only claim optimality for data with i.i.d. Gaussian disturbances, Theorem 1 shows that the test is efficient against *all* local alternatives satisfying (9) and (11) with $\theta = \theta_1$ if one imposes size control for all models satisfying (9) and (11) with $\theta = 0$. In other words, under this size constraint, no test exists with higher asymptotic power for any disturbance distribution or autocorrelation structure satisfying (9) and (11) with $\theta = \theta_1$.

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

$$\{(y_{T,t}, x'_{T,t})'\}_{t=1}^T \to \{(y_{T,t} + a_y, x'_{T,t} + a'_x)'\}_{t=1}^T \quad a_y \in \mathbb{R},\, a_x \in \mathbb{R}^{n-1}. \tag{12}$$

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

[5] Elliott and Jansson (2003) do not rely on (11), though, but exploit the assumption of a finite lag VAR autocorrelation structure in $\nu_{T,t}$ to implicitly construct an appropriate U_T.

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

$$\hat{\Omega}_T \xrightarrow{p} \Omega \quad \text{and} \quad \begin{pmatrix} T^{-1/2}\hat{y}_{T,\lfloor \cdot T \rfloor} \\ T^{-1/2}\sum_{t=1}^{\lfloor \cdot T \rfloor} \hat{x}_{T,t} \\ T^{-1/2}\sum_{t=2}^{T} \hat{y}_{T,t-1}\hat{x}_{T,t} - \hat{\Sigma}_T \end{pmatrix} \rightsquigarrow \begin{pmatrix} \hat{G}_y^{\theta}(\cdot) \\ \hat{G}_x^{\theta}(\cdot) \\ \int_0^1 \hat{G}_y^{\theta}(s) d\hat{G}_x^{\theta}(s) \end{pmatrix} \tag{13}$$

where $\hat{G}_y^{\theta}(s) = G_y^{\theta}(s) - \int_0^1 G_y^{\theta}(s)ds$ and $\hat{G}_x^{\theta}(s) = G_x^{\theta}(s) - sG_x^{\theta}(1)$. For brevity, we omit an explicit expression for the Radon-Nikodym derivative $\mathrm{LR}^{\hat{G}}$ of the measure of $(\hat{G}_y^{\theta}, \hat{G}_x^{\theta})$ with respect to the measure of $(\hat{G}_y^{0}, \hat{G}_x^{0})$, but note that it could be deduced from the limiting result in Elliott and Jansson (2003) by again invoking the change in asymptotic measure implied by contiguity, just as in the argument above. By Theorem 1, rejecting for large values of $\mathrm{LR}^{\hat{G}}$, evaluated at sample analogoues, is the asymptotically efficient robust test for models defined via (13). Furthermore, in the notation of Section 2.3, with $\phi_T(Y_T) = (T^{-1/2}y_{T,\lfloor \cdot T \rfloor}, T^{-1/2}\sum_{t=1}^{\lfloor \cdot T \rfloor} x_{T,t}, T^{-1}\sum_{t=2}^{T} y_{T,t-1}x_{T,t}) \in D_{[0,1]} \times D_{[0,1]}^{n-1} \times \mathbb{R}^{n-1}$, $r = (r_y, r_x')' \in \mathbb{R} \times \mathbb{R}^{n-1}$, $g_r(\{(y_{T,t}, x_{T,t}')'\}_{t=1}^{T}) = \{(y_{T,t} + T^{1/2}r_y, x_{T,t}' + T^{-1/2}r_x')'\}_{t=1}^{T}$, $\tilde{g}_r(y,x,z) = (y(\cdot)+r_y, x(\cdot)+r_x\cdot, z+r_y(x(1)-x(0))+r_x\int_0^1 y(s)ds)$, $h_T(\{(y_{T,t}, x_{T,t}')'\}_{t=1}^{T}) = \{(\hat{y}_{T,t}, \hat{x}_{T,t}')'\}_{t=1}^{T}$ and $\tilde{h}(y,x,z) = (y(\cdot) - \int_0^1 y(s)ds, x(\cdot) - \cdot x(1), z - (x(1)-x(0))\int_0^1 y(s)ds)$, we find that Theorem 3 is applicable, and that rejecting for large values of $\mathrm{LR}^{\hat{G}}$, evaluated at sample analogues, is also the asymptotically point-optimal invariant unit root test among all tests with correct asymptotic null rejection probability all models satisfying (9) and (11) with $\theta = 0$.

3.2 GMM Wald Tests

Consider a standard GMM set-up with parameter of interest $\eta \in \mathbb{R}^k$ (and possibly some additional finite dimensional nuisance parameter) in a sample of size T. Parametrize the true parameter η as $\eta = \eta_0 + T^{-1/2}\theta$ and suppose we are interested in testing $H_0 : \theta = 0$ against $H_1 : \theta \neq 0$. For a large number of primitive conditions, the GMM estimator $\hat{\eta}$ satisfies $\sqrt{T}(\hat{\eta} - \eta_0) \rightsquigarrow X \sim \mathcal{N}(\theta, V)$, and there exists an estimator $\hat{V}_T \xrightarrow{p} V$. If, for a given sample of size T, the usual Wald statistic $T(\hat{\eta} - \eta_0)'\hat{V}_T^{-1}(\hat{\eta} - \eta_0)$ is compared to the

$1-\alpha$ quantile of a Chi-squared distribution with k degrees of freedom, at a minimum, one must deem this asymptotic approximation reasonably accurate for $\theta = 0$.

The limiting random variable X is a multivariate normal with mean θ and known covariance matrix V. It is well known that for this limiting problem, the test that rejects for large values of $X'V^{-1}X$ maximizes weighted average power with respect to any weighting function $w(\theta)$ that depends on θ through $\theta' V^{-1}\theta$—such weighting functions put equal weight on alternatives that are equally difficult to distinguish from the null hypothesis (cf. Wald (1943)). By Theorem 1, rejecting for large values of $T\hat{\eta}'\hat{V}_T^{-1}\hat{\eta}$, that is employing the usual GMM Wald test, therefore maximizes this asymptotic weighted average power among all tests that have correct asymptotic null rejection probability for all models in which $\sqrt{T}(\hat{\eta}-\eta_0) \rightsquigarrow \mathcal{N}(0,V)$ and $\hat{V}_T \xrightarrow{p} V$. Thus, without additional regularity assumptions on the GMM set-up, the standard test is efficient. As argued in Section 2.1, a desire to derive tests that are robust for all such models might be motivated either because assuming more would require to take a specific stand on, say, the uncertain time series properties of the underlying data Y_T, or because one worries about the approximation accuracy of additional asymptotic implications of, say, the knowledge that the underlying data is i.i.d.

3.3 GMM Parameter Stability Tests

Following Sowell (1996), suppose we are interested in testing the null hypothesis that a parameter η in a GMM framework is constant through time. Parametrizing $\eta_{T,t} = \eta_0 + T^{-1/2}\theta(t/T)$, where $\theta \in D^k_{[0,1]}$, this is equivalent to the hypothesis test

$$H_0 : \theta = 0 \quad \text{against} \quad H_1 : \theta \text{ is not constant.} \tag{14}$$

With $Y_T = (y_{T,1}, \cdots, y_{T,T})'$, denote by $g_{T,t}(\eta) \in \mathbb{R}^\ell$ with $\ell \geq k$ the sample moment condition for $y_{T,t}$ evaluated at η, so that under the usual assumptions, the moment condition evaluated at the true parameter value satisfies a central limit theorem, that is $T^{-1/2}\sum_{t=1}^{T} g_{T,t}(\eta_{T,t}) \rightsquigarrow \mathcal{N}(0,V)$ for some positive definite $\ell \times \ell$ matrix V. Furthermore, with $\hat{\eta}_T$ the usual full sample GMM estimator of η 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))

$$G_T(\cdot) = T^{-1/2} \sum_{t=1}^{\lfloor \cdot T \rfloor} g_{T,t}(\hat{\eta}) \rightsquigarrow G^\theta(\cdot) \quad \text{and} \quad \hat{H}_T \xrightarrow{p} H, \, \hat{V}_T \xrightarrow{p} V \tag{15}$$

where the convergence to G^θ is on $D^\ell_{[0,1]}$, $G^\theta(s) = V^{1/2}W(s) - sH(H'V^{-1}H)^{-1}H'V^{-1/2}W(1) + H\left(\int_0^s \theta(l)dl - s\int_0^1 \theta(l)dl\right)$ with W a $\ell \times 1$ standard Wiener process and H some $\ell \times k$ 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 (2006) discuss primitive conditions for these convergences. Sowell (1996) goes on to derive weighted average power maximizing tests of (14) as a function of G^θ (that is, he computes φ^* 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 $y_{T,t} = \eta + \theta(t/T) + \varepsilon_t$, 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 $g_{T,t}(\eta) = y_{T,t} - \eta = \theta(t/T) + \varepsilon_t$ satisfying (15), 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 maximizes asymptotic weighted average power among all tests that have correct asymptotic null rejection probability whenever (15) holds with $\theta = 0$. Tests that exploit specificities of the error distribution, such as φ^{**}_T, to gain higher power necessarily do not control size for all stable models satisfying (15).

4 Conclusion

This paper suggests a new notion of asymptotic efficiency by imposing size control over a set of models that satisfy a specific weak convergence. Under relatively weak regularity conditions on the Radon-Nikodym derivative of the distribution of the limiting random element under the null and alternative hypothesis, it is found that rejecting for large values of this derivative, evaluated at sample analogues, yields an efficient test in this sense. Since this test typically coincides with the asymptotically efficient test under the assumption of Gaussian i.i.d. disturbances, this result may be interpreted as providing a bound on the possibility of adaption. A number of standard tests in the literature are shown to be efficient in this new sense.

5 Appendix

Proof of Theorem 1:

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

Pick any γ. Since φ^* is $P^e(\gamma,\theta_0)$-almost everywhere continuous, it is also $P^e(\gamma,\theta)$-almost everywhere continuous, and $P_T^e(m,\gamma,\theta) \rightsquigarrow P^e(\gamma,\theta)$ implies $\int \varphi^*(\gamma_1,x)dP_T^e(m,\gamma,\theta)(\gamma_1,x) \to \int \varphi^*(\gamma_1,x)dP^e(\gamma,\theta)(\gamma_1,x) = \int \varphi^*(\gamma,x)dP(\gamma,\theta)(x)$ for all $\theta \in \Theta_1$. Since $0 \leq \varphi^* \leq 1$, the result follows by dominated convergence.

(ii) Take any $m \in \mathcal{M}_1$ and $\gamma \in \Gamma$. Denote by $\bar{F}_T$ the distribution $\int F_T(m,\gamma,\theta)dw(\theta)$. By assumption, the measure $P_T(m,\gamma,\theta)$ of $X_T = \phi_T(Y_T)$ induced by $F_T(m,\gamma,\theta)$ satisfies (3). By definition of weak convergence, this implies that for any bounded and continuous function $\vartheta : S \mapsto \mathbb{R}$, $\int \vartheta dP_T(m,\gamma,\theta) \to \int \vartheta dP(\gamma,\theta)$ for any $\theta \in \Theta_1$, so that by dominated convergence $\int \vartheta d\bar{P}_T = \int\int \vartheta dP_T(m,\gamma,\theta)dw(\theta) \to \int\int \vartheta dP(\gamma,\theta)dw(\theta) = \int \vartheta d\bar{P}$. For notational brevity, write P_0 for the measure $P(\gamma,\theta_0)$ on S, and also $\overline{\mathrm{LR}}$ for the $S \mapsto \mathbb{R}$ function $\overline{\mathrm{LR}}(\gamma,\cdot)$. Note that $\bar{P}$ is absolutely continuous with respect to P_0, with Radon-Nikodym derivative $\overline{\mathrm{LR}}$. Pick $\varepsilon > 0$ for which $\int \mathbf{1}[\overline{\mathrm{LR}} = \varepsilon]dP_0 = 0$, and note that $\int \mathbf{1}[\overline{\mathrm{LR}} \leq \varepsilon]d\bar{P} \leq \varepsilon \int \mathbf{1}[\overline{\mathrm{LR}} \leq \varepsilon]dP_0 \leq \varepsilon$. Let B be the indicator function of the set $A = \{x \in S : \overline{\mathrm{LR}} > \varepsilon\}$. Denote by $\overline{\mathrm{LR}}^i$ the $S \mapsto \mathbb{R}$ function $1/\overline{\mathrm{LR}}$, and pick some $m_0 \in \mathcal{M}_0$ to define the real number $\kappa_T = \int(\mathrm{B}\circ\phi_T)dF_T(m_0,\gamma,\theta_0)/\int(\mathrm{B}\circ\phi_T)(\overline{\mathrm{LR}}^i \circ \phi_T)d\bar{F}_T = \int \mathrm{B}dP_T(m_0,\gamma,\theta_0)/\int \mathrm{B}\overline{\mathrm{LR}}^i d\bar{P}_T$ and the probability distribution G_T of Y_T via

$$\int_A dG_T = \kappa_T \int_A (\mathcal{B}\circ\phi_T)(\overline{\mathrm{LR}}^i \circ \phi_T)d\bar{F}_T + \int_A (1-(\mathcal{B}\circ\phi_T))dF_T(m_0,\gamma,\theta_0)$$

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

$$\int \vartheta dQ_T = \kappa_T \int \vartheta \mathcal{B}\overline{\mathrm{LR}}^i d\bar{P}_T + \int \vartheta(1-\mathcal{B})dP_T(m_0,\gamma,\theta_0).$$

Since B, ϑ and $\mathrm{B}\overline{\mathrm{LR}}^i$ are bounded and $P(\gamma,\theta_0)$-almost everywhere continuous (and thus also $\bar{P}$-almost everywhere continuous) functions $S \mapsto \mathbb{R}$, it follows that $\int \mathrm{B}dP_T(m_0,\gamma,\theta_0) \to \int \mathrm{B}dP_0$, $\int \vartheta(1-\mathrm{B})dP_T(m_0,\gamma,\theta_0) \to \int \vartheta(1-\mathrm{B})dP_0$, $\int \mathrm{B}\overline{\mathrm{LR}}^i d\bar{P}_T \to \int \mathrm{B}\overline{\mathrm{LR}}^i d\bar{P} = \int \mathrm{B}dP_0$ and $\int \vartheta \mathrm{B}\overline{\mathrm{LR}}^i d\bar{P}_T \to \int \vartheta \mathrm{B}\overline{\mathrm{LR}}^i d\bar{P} = \int \vartheta \mathrm{B}dP_0$, so that $\int \vartheta dQ_T \to \int \vartheta dP_0$. Thus, $Q_T \rightsquigarrow P_0$, and (4) implies that $\limsup_{T\to\infty} \int \varphi_T dG_T \leq \alpha$.

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

$$\begin{aligned}\int_A d\tilde{F}_T &= \tilde{\kappa}_T \int_A (\mathcal{B} \circ \phi_T)(\overline{\text{LR}} \circ \phi_T) dG_T \\ &= \tilde{\kappa}_T \kappa_T \int_A (\mathcal{B} \circ \phi_T) d\bar{F}_T\end{aligned}$$

for any $\in \mathfrak{B}(\mathbb{R}^{nT})$, where $\tilde{\kappa}_T = 1/(\kappa_T \int(\text{B}\circ\phi_T)d\bar{F}_T) \to \tilde{\kappa} = 1/\int \mathbf{1}[\overline{\text{LR}} > \varepsilon]d\bar{P}$. By the Neyman-Pearson Lemma, the best test of $\tilde{H}_0 : Y_T \sim G_T$ against $\tilde{H}_1 : Y_T \sim \tilde{F}_T$ rejects for large values of $(\text{B}\overline{\text{LR}}) \circ \phi_T = f \circ \overline{\text{LR}} \circ \phi_T$, where $f(x) = x\mathbf{1}[x > \varepsilon]$. Thus, the best test for large values of $\overline{\text{LR}} \circ \phi_T$. For any T, denote by $\tilde{\varphi}^*_T : \mathbb{R}^{nT} \mapsto [0,1]$ the (possibly randomized) test that rejects for large values of $y \mapsto \overline{\text{LR}}(\gamma, \phi_T(y))$ of level $\int \tilde{\varphi}^*_T dG_T = \max(\int \varphi_T dG_T, \alpha)$, so that $\int(\tilde{\varphi}^*_T - \varphi_T)d\tilde{F}_T \geq 0$. Note that $\tilde{F}_T$ is contiguous to G_T, since under G_T, the Radon-Nikodym derivative $f \circ \overline{\text{LR}} \circ \phi_T$ converges weakly to the distribution of $\tilde{\kappa}\text{B}\overline{\text{LR}}$ under $\bar{P}$ by the Continuous Mapping Theorem, and $\int \tilde{\kappa}\text{B}\overline{\text{LR}}d\bar{P} = 1$. Define $\varphi^*_T : \mathbb{R}^{nT} \mapsto [0,1]$ as $\varphi^*_T(y) = \varphi^*(\gamma, \phi_T(y))$. Proceeding as in part (i) of the proof shows that $\text{ARP}_0(\varphi^*_T, m) = \alpha$ and $\lim_{T\to\infty}\text{WAP}_T(\varphi^*_T, m, \gamma) = \beta(\gamma)$. Since both $\tilde{\varphi}^*_T$ and φ^*_T reject for large values of $\overline{\text{LR}}$ and are of asymptotic level α, we have $\int |\tilde{\varphi}^*_T - \varphi^*_T| dG_T \to 0$, so that by contiguity, also $\int |\tilde{\varphi}^*_T - \varphi^*_T| d\tilde{F}_T \to 0$. Thus $\limsup_{T\to\infty} \int(\varphi^*_T - \varphi_T)d\tilde{F}_T \geq 0$. To complete the proof, note that the total variation distance between $\tilde{F}_T$ and $\bar{F}_T$ is bounded above by $\int((1-\text{B})\circ\phi_T)d\bar{F}_T \to \int \mathbf{1}[\overline{\text{LR}} \leq \varepsilon]d\bar{P} \leq \varepsilon$, so that $\limsup_{T\to\infty}\text{WAP}_T(\varphi_T, m, \gamma) \leq \beta(\gamma) + \varepsilon$ and the result follows, since ε was arbitrary.

Proof of Theorem 2:

Similar to proof of Theorem 1 above. Pick $1/2 > \varepsilon > 0$ such that $P_0(\overline{\text{LR}} = \varepsilon) = 0$. Define $A_{\overline{\text{LR}}} = \{x \in S : \overline{\text{LR}} > \varepsilon\}$. Since $\bar{P}$ is absolutely continuous with respect to P_0, there exists an open set A_L such that $\bar{P}(A_L) > 1 - \varepsilon$ and $\overline{\text{LR}} : A_L \mapsto \mathbb{R}$ is Lipschitz, so that with $A = A_L \cap A_{\overline{\text{LR}}}$, $\overline{\text{LR}}^i : A \mapsto \mathbb{R}$ is bounded and Lipschitz. Furthermore, since $S\backslash A$ is closed, there exists a Lipschitz function $\mathcal{C} : S \mapsto [0,1]$ that is zero on $S\backslash A$ and for which $\int \mathcal{C}d\bar{P} \geq 1 - 3\varepsilon$ (see Pollard (2002), p. 172-173 for an explicit construction). For future reference, define $\text{B}_{\overline{\text{LR}}}$ to be the indicator function of $A_{\overline{\text{LR}}}$, and note that $\text{B}_{\overline{\text{LR}}}\mathcal{C} = \mathcal{C}$.

With $\bar{F}_T = \int F_T(m_1,\gamma,\theta)dw(\theta)$ and $\bar{P}_T = \phi_T \bar{F}_T$, define

$$\begin{aligned}
\kappa_T &= \int (\mathcal{C}\circ\phi_T)dF_T(m_0,\gamma,\theta_0) / \int (\mathcal{C}\circ\phi_T)(\overline{\mathrm{LR}}^i\circ\phi_T)d\bar{F}_T \\
&= \int \mathcal{C}dP_T(m_0,\gamma,\theta_0) / \int \mathcal{C}\overline{\mathrm{LR}}^i d\bar{P}_T \to 1
\end{aligned}$$

and the probability distribution G_T of Y_T via

$$\int_A dG_T = \kappa_T \int_A (\mathcal{C}\circ\phi_T)(\overline{\mathrm{LR}}^i\circ\phi_T)d\bar{F}_T + \int_A (1-(\mathcal{C}\circ\phi_T))dF_T(m_0,\gamma,\theta_0)$$

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

$$\int \vartheta dQ_T = \kappa_T \int \vartheta\mathcal{C}\overline{\mathrm{LR}}^i d\bar{P}_T + \int \vartheta(1-\mathcal{C})dP_T(m_0,\gamma,\theta_0).$$

Now

$$\begin{aligned}
\Delta(Q_T,P_0) &= \sup_{||\vartheta||_{BL}\le 1} |\int \vartheta(dQ_T - dP_0)| \\
&= \sup_{||\vartheta||_{BL}\le 1} |\int (\vartheta\mathcal{C}(\kappa_T\overline{\mathrm{LR}}^i d\bar{P}_T - dP_0) + \vartheta(1-\mathcal{C})(dP_T(m_0,\gamma,\theta_0) - dP_0)| \\
&\le \sup_{||\vartheta||_{BL}\le 1} |\int \vartheta\mathcal{C}\overline{\mathrm{LR}}^i(\kappa_T d\bar{P}_T - d\bar{P})| + \sup_{||\vartheta||_{BL}\le 1} |\int \vartheta(1-\mathcal{C})(dP_T(m_0,\gamma,\theta_0) - dP_0)| \\
&\le ||\mathcal{C}\overline{\mathrm{LR}}^i||_{BL}(\Delta_{BL}(\bar{P}_T,\bar{P}) + |1-\kappa_T|) + ||1-\mathcal{C}||_{BL}\Delta_{BL}(P_T(m_0,\gamma,\theta_0),P_0)
\end{aligned}$$

and also

$$\begin{aligned}
|\kappa_T - 1| &= |\frac{\int \mathcal{C}dP_T(m_0,\gamma,\theta_0)}{\int \mathcal{C}\overline{\mathrm{LR}}^i d\bar{P}_T} - \frac{\int \mathcal{C}dP_0}{\int \mathcal{C}\overline{\mathrm{LR}}^i d\bar{P}}| \\
&\le \frac{||\mathcal{C}\overline{\mathrm{LR}}^i||_{BL}\Delta_{BL}(\bar{P}_T,\bar{P}) + ||\mathcal{C}||_{BL}\Delta_{BL}(P_T(m_0,\gamma,\theta_0),P_0)}{\int \mathcal{C}\overline{\mathrm{LR}}^i d\bar{P}_T}.
\end{aligned}$$

Thus, $\limsup_{T\to\infty}\Delta_{BL}(Q_T,P_0)/\delta(\gamma,T) = 0$, so that for large enough T, $G_T \in \mathcal{M}_0^u(\delta)$, and (6) implies $\limsup_{T\to\infty}\int \varphi_T dG_T \le \alpha$.

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

$$\begin{aligned}
\int_D d\tilde{F}_T &= \tilde{\kappa}_T \int_D (\mathcal{B}_{\overline{\mathrm{LR}}}\circ\phi_T)(\overline{\mathrm{LR}}\circ\phi_T)dG_T \\
&= \tilde{\kappa}_T\kappa_T \int_D (\mathcal{C}\circ\phi_T)d\bar{F}_T + \tilde{\kappa}_T \int_D (\mathcal{B}_{\overline{\mathrm{LR}}}(1-\mathcal{C})\overline{\mathrm{LR}}\circ\phi_T)dF_T(m_0,\gamma,\theta_0)
\end{aligned}$$

for $D \in \mathfrak{B}(\mathbb{R}^{nT})$, where $\tilde{\kappa}_T = 1/(\kappa_T \int (\mathcal{C} \circ \phi_T) d\bar{F}_T + \int (\mathrm{B}_{\overline{\mathrm{LR}}}(1 - \mathcal{C})\overline{\mathrm{LR}} \circ \phi_T) dF_T(m_0, \gamma, \theta_0)) \to 1/\int (\mathcal{C} + \mathrm{B}_{\overline{\mathrm{LR}}} - \mathrm{B}_{\overline{\mathrm{LR}}}\mathcal{C}) d\bar{P} = 1/\int \mathrm{B}_{\overline{\mathrm{LR}}} d\bar{P}$ and $1 \leq 1/\int \mathrm{B}_{\overline{\mathrm{LR}}} d\bar{P} \leq 1 + 2\varepsilon$. 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 $(\mathrm{B}_{\overline{\mathrm{LR}}}\overline{\mathrm{LR}}) \circ \phi_T$, i.e. $\overline{\mathrm{LR}} \circ \phi_T$. Noting that the total variation distance between $\tilde{F}_T$ and $\bar{F}_T$ is bounded above by

$$\begin{aligned} \int |1 - \tilde{\kappa}_T \kappa_T (\mathcal{C} \circ \phi_T)| d\bar{F}_T &\leq |1 - \tilde{\kappa}_T \kappa_T| + \int (1 - \mathcal{C}) d\bar{P}_T \\ &\to |\int \mathcal{B}_{\overline{\mathrm{LR}}} d\bar{P}|^{-1} + 1 - \int \mathcal{C} d\bar{P} \leq 4\varepsilon \end{aligned}$$

the remainder of the proof is just as the proof of Theorem 1 above.

Proof of Theorem 3:

Pick any $\gamma \in \Gamma$, $\xi \in \Xi$ and $m \in \mathcal{M}_0^h$. Denote the distribution of Y_T by $F_T = F_T(m, \gamma, \theta_0, \xi)$, and write $P_T = \phi_T F_T$, $P_T^h = (\phi_T \circ h_T) F_T$, $P_0 = P^i(\gamma, \theta_0)$ and $P^h = \tilde{h} P_0$. Let $D_{[0,1]}^n$ the space of n-valued cadlag functions on the unit interval, equipped with the Billingsley metric, and define the mapping $\chi_T : \mathbb{R}^{nT} \mapsto D_{[0,1]}^n$ as $\{y_t\}_{t=1}^T \mapsto T^{-1}\Phi(y_{\lfloor \cdot T \rfloor})$, where Φ is the cdf of a standard normal applied element by element. Note that χ_T is injective, and denote by χ_T^{-1} the $D_{[0,1]}^n \mapsto \mathbb{R}^{nT}$ 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 $(\phi_T \circ h_T, \chi_T) F_T$ on the separable space $S \times D_{[0,1]}^n$ converge weakly to the product measure $P^h \times \delta_0$, where δ_0 puts all mass at the zero function in $D_{[0,1]}^n$. By Theorem 11.7.2 of Dudley (2002), there exists a probability space $(\Omega^*, \mathcal{F}^*, P^*)$ and functions $\eta_T : \Omega^* \mapsto S \times D_{[0,1]}^n$ such that $\eta_T P^* = (\phi_T \circ h_T, \chi_T) F_T$, $\eta P^* = P^h \times \delta_0$ and $\eta_T(\omega^*) \to \eta(\omega^*)$ for P^*-almost all $\omega^* \in \Omega^*$. In particular, $(\sigma \circ \eta_T) P^* = P_T^h$ and $(\chi_T^{-1} \circ d \circ \eta_T) P^* = F_T$, where $\sigma : S \times D_{[0,1]}^n \mapsto S$ and $d : S \times D_{[0,1]}^n \mapsto D_{[0,1]}^n$ are the usual projections of $S \times D_{[0,1]}^n$ on S and $D_{[0,1]}^n$, respectively. Also, for almost all ω^*, $\phi_T \circ h_T \circ \chi_T^{-1} \circ d \circ \eta_T(\omega^*) = \sigma \circ \eta_T(\omega^*)$. Let ν be the probability measure on $\mathfrak{B}(R \times S)$ induced by $(\rho, \tilde{h}) : S \mapsto R \times S$ under P_0, i.e. $\nu = (\rho, \tilde{h}) P_0$. Since $x = \tilde{g}_{\rho(x)}(\tilde{h}(x))$ for all $x \in S$, $P_0 = \tilde{g}\nu$, where $\tilde{g} : R \times S \mapsto S$ is defined as $\tilde{g}(r, x) = \tilde{g}_r(x)$. 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 $B \in \mathfrak{B}(R)$, $\nu(A \times B) = \int_A \nu_x(B) dP^h(x)$. Note that the mapping $\Omega^* \times \mathfrak{B}(S) \mapsto [0, 1]$ defined via $(\omega^*, A) \mapsto \nu_{\sigma \circ \eta(\omega^*)}(A)$ is a probability kernel from $(\Omega^*, \mathcal{F}^*)$ to $(R, \mathfrak{B}(R))$. We can thus construct the product probability measure μ

on $(\Omega^* \times R)$, $(\mathcal{F}^* \otimes \mathfrak{B}(R))$ via $\mu(C \times A) = \int_C \nu_{\sigma \circ \eta(\omega^*)}(A) dP^*(\omega^*)$, and by construction, the mapping $(\omega^*, r) \mapsto \tilde{g}_r \circ \sigma \circ \eta(\omega^*)$ induces the measure P_0 under μ. Furthermore, note that for μ-almost all (ω^*, r), $\phi_T \circ g_r \circ h_T \circ \chi_T^{-1} \circ d \circ \eta_T(\omega^*) = \tilde{g}_r \circ \phi_T \circ h_T \circ \chi_T^{-1} \circ d \circ \eta_T(\omega^*) + o(T) = \tilde{g}_r \circ \sigma \circ \eta_T(\omega^*) + o(T) \to \tilde{g}_r \circ \sigma \circ \eta(\omega^*)$. But almost sure convergence implies weak convergence, so that the measures G_T on $\mathbb{R}^{nT}$ induced by the mapping $(\omega^*, r) \mapsto g_r \circ h_T \circ \chi_T^{-1} \circ d \circ \eta_T(\omega^*)$ under μ satisfy $\phi_T G_T \rightsquigarrow P_0$. Thus, by assumption, $\limsup_{T \to \infty} \int \varphi_T dG_T \leq \alpha$. Since φ_T is invariant and h_T is a maximal invariant, there is no loss in generality to assume that $\varphi_T(y_T)$ is a function of $h_T(y_T)$ by Theorem 1 on p. 285 of Lehmann (1986). From $h_T \circ g_r \circ h_T = h_T$ we have $h_T G_T = h_T F_T$, and thus also $\limsup_{T \to \infty} \int \varphi_T dF_T \leq \alpha$.

References

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.

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

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., and U. K. Müller (2003): "Optimally Testing General Breaking Processes in Linear Time Series Models," *UCSD Working Paper 2003-07.*

——— (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. (2004): "Stationarity Testing with Covariates," *Economtric Theory*, 20, 56–94.

——— (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," *Working paper, UC Berkeley.*

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

Lehmann, E. L. (1986): *Testing Statistical Hypotheses.* Wiley, New York, second edn.

Li, H., and U. K. Müller (2006): "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.

Müller, U. K. (2004): "The Impossibility of Consistent Discrimination Between I(0) and I(1) Processes," *Working paper, Princeton University.*

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 (2006): "Testing Models of Low-Frequency Variability," *NBER Working Paper W12671.*

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.

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

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.

WALD, A. (1943): "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large," *Transactions of the American Mathematical Society*, 54, 426–482.

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.