Efficient Tests for Autoregressive Unit Roots in Panel Data

David Bowman*

August 1999

Abstract

In this paper the class of admissable tests for unit roots in panel data sets of autoregressive, Gaussian time series will be partially characterized. Using this characterization, several recently suggested tests are shown to be inadmissable. Since the sufficient statistic for this testing problem is multidimensional, there is no uniformly most powerful test, however, in light of the inadmissability result, a new test is proposed that appears to do well relative to existing tests. The test is parameterized in a way that allows the choice of different directional deviations from the null hypothesis over which power is to be maximized, giving added flexibility to researchers.

JEL Classification Codes: C12, C15, C22, C23.

Keywords: Admissable tests, nonstationarity, panel data.

---

*Board of Governors of the Federal Reserve System. Mail Stop 24. Washington, D.C. 20551. Tel:(202) 452-2334. Email: david.h.bowman@frb.gov. I wish to thank Andrew Levin for introducing me to this topic and Jon Faust for his continuing willingness to look at graphs of acceptance regions and power functions. Gauss programs implementing the tests described in this paper are available from the author on request. This paper represents views of the author and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or other members of its staff.

1 Introduction

Following the large body of literature on testing a null hypothesis that a univariate time series is integrated of order one against the alternative hypothesis that it is integrated of order zero, a number of unit root tests have recently been developed for use with panel data (Quah(1994); Levin and Lin(1993); Im, Pesaran, and Shin(1997); and Maddala and Wu (1996)). These tests are being used with increasing frequency in applied work.[1] While results have been obtained concerning test optimally in the univariate context (see Elliot, Stock and Rothenberg (1996) and the references therein), there are as of yet no similar results for tests applied to cross sections of time series. This is of some importance because the tests that have been proposed for use with panels differ markedly in form, and there is ample room for confusion without some guidance as to the contexts in which a test will or will not be optimal.

In this paper the class of admissable tests for unit roots in multivariate data sets of autoregressive, Gaussian time series will be partially characterized. Using this characterization, several recently suggested tests are shown to be inadmissable. Since the sufficient statistic for this testing problem is multidimensional, there is no uniformly most powerful test, however, in light of the inadmissability result, a new class of tests is proposed that appear to do well relative to existing tests. The class is parameterized in a way that allows choice of different directional deviations from the null hypothesis over which power is to be maximized, giving added flexibility to researchers.

Given a set of realizations $y = (\ldots, y_{nt}, \ldots)$ from N stochastic processes, indexed by $n = 1 \ldots N$, observed over time periods $t = 0 \ldots T$, existing tests assume that each individual process has an autoregressive form

$$\triangle y_{nt} = \theta_n y_{nt-1} + \bar{\beta}_n' \left(\triangle \bar{z}_t - \theta_n \bar{z}_{t-1}\right) + \beta_n' z_{nt} + e_{nt} \tag{1}$$

where $\triangle y_{nt} \equiv y_{nt} - y_{nt-1}$, $\bar{z}_t$ is a set of nonstochastic regressors,

$$z_{nt} = \left[\begin{array}{ccc} \triangle y_{nt-1} & \ldots & \triangle y_{nt-P} \end{array} \right]',$$

e_{nt} is i.i.d. and has mean zero and variance σ_n^2, and the first $P+1$ observations are treated as fixed. The paper will employ generalizations of this model. In contrast to most of the literature, it will not be assumed that the e_{nt} are necessarily independent across groups, however it will be assumed in line with

[1]This work includes Bernard and Jones (1996); Canzoneri, Cumby and Diba(1999); Evans and Karras (1996); Frankel and Rose (1995); O'Connell (1998); Oh (1996); Papell (1997); Wei and Parsley (1995); and Wu (1996). These tests have also been used in constructing tests of cointegration in panel data (see Pedroni (1997)).

existing literature that the first $P+1$ observation are uninformative about the parameters of the model.[2]

A panel unit root test will be defined as a test of the joint hypothesis $\theta_n = 0$ for $n = 1 \ldots N$. Assuming that one is uninterested in explosive cases, there is a multivariate set of one-sided alternatives: $\theta_n \leq 0$ for all n and $\theta_n < 0$ for at least one n. There are two aspects that make the construction of optimal tests more complicated than in either the case of a univariate unit root test or a multivariate one-sided test with I(0) data: (1) there is a multidimensional sufficient statistic for each group even as T grows large, (2) we wish to combine sufficient statistics across groups to make a joint inference concerning the θ_n. It is the combination of these two aspects that lead to the inadmissability of several of the tests proposed in the existing literature. In more standard problems either there would be no problem of joint inference or each group's parameter of interest would have a one-dimensional sufficient statistic (at least asymptotically). The analysis requires added statistical theory that has not been used in the univariate case, and leads to a somewhat richer set of results. For instance, there will be an asymptotically uniformly most powerful test for certain directional deviations from the null, something that is not true in the univariate case.

Initial tests in this literature (Quah (1994) and Levin and Lin (1993)) either implicitly or explicitly assumed that the θ_n were homogenous, so that either all groups were $I(1)$ or all groups were $I(0)$ with $\theta_n = \bar{\theta} < 0$. For small T this hypothesis may be a practical necessity since power against heterogeneous alternatives is likely to be low. However for larger T, for example sample sizes on the order of those typically found in macro data sets, the assumption of homogeneity can be relaxed while maintaining good power against a wide set of alternatives. Because theoretical arguments for homogeneity are often weak, the primary focus of this paper will be on optimality results that can be established when allowing the alternative set to include heterogeneous θ_n's, however some further optimality results will also be provided for the more restrictive case of homogeneity.

Readers at this stage might expected to wonder in what sense a test which does not constrain the θ_n to be homogeneous can be thought of as a "panel" test or why, if the assumption of homogeneity is dropped, univariate unit root tests should not be calculated for each group individually. This will be briefly addressed before proceeding. On a practical level, the approach taken in this

[2]Extending the paper's results to cases in which the first observations are informative about θ_n (as in Elliot (1993)) or d_n (Elliot, Stock and Rothenberg (1996)) can be accomplished in a straightforward manner, but is at odds with the tests that have been considered in the existing literature and so will not be discussed in the interest of brevity.

paper will be shown to lead to tests that can offer some attractive trade-offs relative to tests that assume homogeneity or that calculate individual test statistics for each group. On a more abstract level, any problem of joint inference will involve pooling data across groups and in this sense can be thought of in terms of a panel. In estimation problems with a cross section of time series observations will be pooled across groups if they are modeled as sharing some set of parameters that determine each group's distribution. In the notation of Liu and Tiao (1980), if the probability distribution of y_n is $p(y_n; \theta_n)$, where θ_n is drawn from a distribution $\pi(\theta; \alpha)$, then

$$p(\theta_n; y) = \int \pi(\theta_n; \alpha) p(\alpha; y) d\alpha,$$

and information will be pooled across groups to estimate each individual group's parameters. In problems of testing a joint hypothesis involving parameters from different groups observations will *automatically* be pooled because they must be used to form a single test statistic. However, unless there is a uniformly most powerful test, different admissable tests will have maximize power against different regions of the space of alternatives. In a rough sense a choice of an admissable test will amount to the choice of a weighting function $\pi(\theta; \alpha)$ of alternatives against which the test will have maximal power. Thus, in estimation problems pooling observations is a modelling choice; in multivariate testing problems involving parameters from different groups pooling is automatic, and the choice of test represents a choice of alternatives against which power is to be maximized.

The rest of the paper will proceed as follows. Section 2 will provide a brief outline of existing tests and will present some definitions required in the following sections. Section 3 will analyze the case in which all nuisance parameters are known and will discuss the relevance of these results to existing tests. Section 4 will introduce a new class of tests within the context of known nuisance parameters. Section 5 will then drop the assumption that nuisance parameters are known and will show how the results of Section 3 can be generalized. Section 6 will adjust the tests introduced in Section 4 for unknown nuisance parameters and conclude.

2 Existing Tests and Required Definitions

2.1 Existing Tests

Existing tests assume that the data is generated by (1) and deal with three standard cases concerning the elements included in the deterministic component $\bar{z}_{nt}$: (1) no deterministic components; (2) a constant mean; and (3) a constant and time trend. Time effects are sometimes allowed and are treated by removing time averages from each variable before testing (assuming that the panel is balanced).

The simplest and most frequently used method of testing the hypothesis $\theta_n = 0$ for $n = 1 \dots N$ is to conduct a univariate unit root test for each group. To ensure that these N tests have an overall size of α, the Bonferroni bound can be used, this bound implying that if each test is conducted individually at the α/N level then the set of N tests will have joint size no greater than α, where the joint hypothesis is rejected if any of the individual tests exceed the α/N critical value. While the other tests that will be described typically assume that ε_{nt} is independent across groups, the Bonferroni bounds will produce a test with correct size even when the error term is correlated across groups.[3] Throughout the paper will refer to the "Bonferroni test" as a test that uses the Bonferroni bounds in conjunction with individual augmented Dickey Fuller (ADF) t-tests. This is a useful but somewhat distortionary piece of terminology since the Bonferroni bounds can be used with any set of tests.

Building upon work by Quah (1994), Levin and Lin (1993) propose a t-test based upon a GLS regression estimate of θ under the assumption that $\theta_n = \bar{\theta}$ in equation (1) while β_n, $\bar{\beta}_n$, and σ_n^2 are allowed to be heterogenous in the regression. The GLS estimator is used to correct for heteroscedasticity ($\sigma_i^2 \neq \sigma_j^2$). When a deterministic component is included (other than time effects) Levin and Lin propose a nonparametric correction for the bias induced by this inclusion.[4]

Im, Pesaran, and Shin (1997) propose two sets of tests, one based upon the average of individual ADF t-tests of $\theta_n = 0$, the other based on the average of individual Lagrange multiplier tests of the same hypothesis. Specifically, if τ_n is either the ADF t-test or LM test for group n, then Im, Pesaran, and Shin's test has the form

$$N^{-\frac{1}{2}} \sum_{n=1}^{N} \frac{\tau_n - a}{b}$$

where a and b are the mean and standard deviation of an ADF t-test or LM test (which are calculated by Monte Carlo methods and depend upon T, the

[3]If independence across groups is a maintained hypothesis then the Bonferroni bound can be tightened so that the test rejects if any individual test execeeds the $\frac{\alpha}{N} + R(\alpha, N)$ critical level, where $R(\alpha, N) = 1 - \frac{\alpha}{N} - (1-\alpha)^{\frac{1}{N}} > 0$. For standard values of α or moderate N this correction will be neglible — for instance, $R(0.05, 2) = 0.00032$.

[4]Levin and Lin state conditions under which they claim their proposed test statistic will have a standard normal distribution as both N and T grow large, $\frac{N}{T} \rightarrow 0$, however as will be discussed in more detail in section 6, the conditions they provide are too weak to establish the claim.

order of the autoregression, and the form of the deterministic component). Im, Pesaran, and Shin state that, assuming normally distributed errors, their test based on LM statistics will have a standard normal distribution as both N and T grow large, $\frac{N}{T} \to \kappa$, $0 < \kappa < \infty$. While they are unable to prove that their test based on averages of ADF t-tests has the same property, their small sample Monte Carlo results indicate that their ADF-based test has somewhat better power than the LM-based test.

Another test has been proposed by Maddala and Wu (1996). Based upon theory developed by Fisher (1932), Maddala and Wu propose combining p-values of individual unit root tests. If τ_n is a test of the univariate hypothesis $\theta_n = 0$ and $P(\ \tau_n)$ is its significance level, then Maddala and Wu propose the statistic

$$-2 \sum_{n=1}^{N} \log\left(P(\tau_n)\right),$$

which will be distributed as a χ^2 with $2N$ degrees of freedom. In theory, any univariate unit root test could be used in this procedure, however Maddala and Wu use ADF t-tests.

In this paper we will concentrate upon the Bonferroni, Levin and Lin (LL), and Im, Pesaran, and Shin (IPS) tests, since these are the most commonly used. However, some of the theory developed below will also apply to the test proposed by Maddala and Wu.

2.2 Defining Required Terms

Defining a parameter space Θ and (strict) subsets $\Theta_0 \subset \Theta$, $\Theta_1 = \Theta - \Theta_0$, we are interested in testing a hypothesis of the form $H_0 : \theta \in \Theta_0$ against the alternative $\theta \in \Theta_1$. Considering the probability density function $p(y; \theta)$ parameterized by $\theta \in \Theta$, a test of H_0 is a measurable function $\tau = \tau(y)$, $0 \leq \tau(y) \leq 1$, where H_0 is rejected with probability $\tau(y)$ when sample y is observed. A test is nonrandomized if $\tau \in \{0, 1\}$. Defining Y as the support of y and $\beta_\tau(\theta) = 1 - \int_Y p(y; \theta)\tau(y)dy$ as the probability that the test τ will accept H_0 when θ is the parameter point, the risk function of a test τ is defined as:

$$r_\tau(\theta) = \begin{cases} \beta_\tau(\theta) & \theta \in \Theta_1 \\ 1 - \beta_\tau(\theta) & \theta \in \Theta_0 \end{cases}$$

A test τ' is *uniformly as good as* another test τ if $r_{\tau'}(\theta) \geq r_\tau(\theta)$ for all θ. A test τ' is *uniformly better* than another test τ if it is uniformly as good as τ and $r_{\tau'}(\theta) > r_\tau(\theta)$ for some θ. A test is *admissable* if there is no other test which is uniformly better. A class of tests is called *complete* if for every test outside the class there is a test within the class which is uniformly better. A class of tests is called *essentially complete* if for every test outside the class there is a test which is uniformly as good as it within the class.

A probability measure π in Θ is an *a priori distribution in* Θ. The *average risk function with respect to* π is defined as $r^*(\pi, \tau) = \int_\Theta r_\tau(\Theta) d\pi(\Theta)$. A test τ_0 is a *Bayes solution relative to* π if it minimizes $r^*(\pi, \tau)$. A test τ is a *Bayes solution in the strict sense* if there is an a priori distribution, π, such that τ is a Bayes solution relative to π. A test τ is a Bayes solution in the *wide sense* if there is exists a sequence of a priori distributions, $\{\pi_i\}$, such that τ is a Bayes solution relative to $\{\pi_i\}$.

3 Known Nuisance Parameters

3.1 Characterization of Admissability

In this section it will be assumed that all nuisance parameters are known, so that the only uncertainty concerns the vector θ. We begin with this case not because it is realistic, but because the basic mechanisms are most clear in this context.

The presentation requires first establishing some notation. If x_{nt} is a variable defined over $n = 1 \ldots N, t = 0 \ldots T$ then let x_n and x_{n-} respectively be the $T \times 1$ vectors $[x_{n1} \ldots x_{nT}]'$ and $[x_{n0} \ldots x_{nT-1}]'$, and let x_t be the $N \times 1$ vector $[x_{1t} \ldots x_{Nt}]'$. If X is a square matrix, then let $vech(X)$ be the linear operator vectorizing the lower triangular portion of X and let $avech(X)$ be the same as $vech(X)$ after multiplying the off diagonal elements by 2. Thus if

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$$

then

$$\begin{aligned} vech(X) &= \begin{bmatrix} x_{11} & x_{21} & x_{22} & x_{31} & x_{32} & x_{33} \end{bmatrix}' \\ avech(X) &= \begin{bmatrix} x_{11} & 2x_{21} & x_{22} & 2x_{31} & 2x_{32} & x_{33} \end{bmatrix}'. \end{aligned}$$

This section will employ the following generalization of (1). It will be assumed that the observations are generated by

$$\begin{aligned} y_{nt} &= d_{nt} + u_{nt} && (2) \\ \triangle u_{nt} &= \theta_n u_{nt-1} + v_{nt} && (3) \end{aligned}$$

where d_{nt} is a deterministic component and v_{nt} is an unobserved stationary mean-zero Gaussian process with a positive spectral density function at frequency zero. The assumptions made in this section are set out in Assumptions 1 and 2:

Assumption 1 *The sequence $\{v_t\}$ is stationary and has a positive definite spectral density matrix; it has a moving average representation $v_t = \sum_{k=0}^{\infty} \Psi_k e_{t-k}$ where the e_t are independent $N(0, \Omega_e)$ random variables and $\sum_{k=0}^{\infty} k \mid \psi_{kij} \mid < \infty$ for each i, $j = 1 \ldots N$, ψ_{kij} denoting the row i column j element of Ψ_k. The distribution of the initial observation u_0 independent of θ.*

Assumption 2 *The parameters $\Omega_v \equiv Evv'$ and d are known.*

Stack equations (2) and (3) across both groups and time:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} d_1 \\ \vdots \\ d_N \end{bmatrix} + \begin{bmatrix} u_1 \\ \vdots \\ u_N \end{bmatrix}$$

$$\begin{bmatrix} \triangle u_1 \\ \vdots \\ \triangle u_N \end{bmatrix} = \begin{bmatrix} u_{1-} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & u_{N-} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_N \end{bmatrix} + \begin{bmatrix} v_1 \\ \vdots \\ v_N \end{bmatrix}$$

or, more compactly,

$$y = d + u \tag{4}$$

$$\triangle u = u_- \theta + v \tag{5}$$

We are interested the problem of testing $\theta = 0$ against the alternative hypothesis $\theta \leq 0, \theta \neq 0$. Formally, we consider the parameter set $\Theta \subset \{\theta \in R^N \mid \theta \leq 0\}$ and study tests with (simple) null $\theta = 0$ and (possibly composite) alternative set $\Theta_1 = \Theta \setminus 0$. Without loss of generality it will be assumed that for each $n = 1 \ldots N$ there is a $\theta \in \Theta_1$ such that $\theta_n < 0$, since any group known to be $I(1)$ can be excluded from the test.

Under assumptions 1 and 2 the probability density for y is

$$\begin{aligned} p(y;\theta) &= h(y) \exp\left(\theta' \left[u_-' \Omega_v^{-1} \triangle u\right] - \frac{1}{2}\theta' \left[u_-' \Omega_v^{-1} u_-\right] \theta\right) \\ &= h(y) \exp\left(\theta' \left[u_-' \Omega_v^{-1} \triangle u\right] - \ avech(\theta\theta')' [vech(u_-' \Omega_v^{-1} u_-)]\right) \\ &= h(y) \exp\left(\gamma(\theta)' s\right) \end{aligned}$$

where

$$\gamma(\theta) = [\theta' \quad - avech(\theta\theta')']',$$

and

$$s = [u'_{-}\Omega_v^{-1} \triangle u \quad vech(u'_{-}\Omega_v^{-1}u_{-})']'.$$

This describes a curved exponential family (Effron 1975,1978) with a $\frac{N(N+3)}{2}$ dimensional sufficient statistic for θ. While $vech(u'_{-}\Omega_v^{-1}u_{-})$ would have a degenerate distribution for large T if one was testing a stationary null against stationary alternatives (for instance testing $\theta = -1$ against $\theta \neq -1$), it has a nondegenerate distribution for large T under the null $\theta = 0$, so that the dimension of the sufficient statistic remains $\frac{N(N+3)}{2}$ even as $T \to \infty$ in the current problem.[5] In considering the set of admissable tests attention can be restricted to tests based upon the sufficient statistic. The problem is then one of optimally combining $\frac{N(N+3)}{2}$ statistics into a single, one-dimensional, test statistic.

When the alternative set Θ_1 has a larger dimension than one there can be no uniformly most powerful test.[6] To obtain some sense of what properties an optimal test might have it is helpful to consider a test of the null against a simple fixed alternative, $\theta^a = (\ldots, \theta_n^a, \ldots)$. By the Neyman-Pearson lemma, the most powerful test of the null $\theta = 0$ against the alternative $\theta = \theta^a$ has the

[5]Under Assumption 1 the probability density for y_n is

$$p(y_n \mid \theta_n) = h_n(y_n) \exp\left(\theta_n \left[u'_{-}\Omega_v^{-1} \triangle u\right] - \frac{1}{2}\theta_n^2 \left[u'_{n-}\Omega_{v_n}^{-1}u_{n-}\right]\right).$$

This describes a curved exponential family with a two-dimensional sufficient statistic for θ_n. In contrast to linear exponential families, there is no uniformly most powerful test of the one-sided hypothesis $\theta_n = 0$ versus $\theta_n < 0$. If $\Omega_{v_n} = I$ then Effon's measure of statistical curvature when evaluated at the null is $\gamma = \frac{2}{3}\left(\frac{T^2+5T-5}{T(T-1)}\right)^{\frac{1}{2}}$, where Effon's measure of curvature is defined as $\gamma = (\det(M)/M_{11})^{\frac{1}{2}}$

$$M = \begin{bmatrix} var\left(\frac{\partial \log(p(y_n;\theta_n))}{\partial\theta_n}\right) & cov\left(\frac{\partial \log(p(y_n;\theta_n))}{\partial\theta_n}, \frac{\partial^2 \log(p(y_n;\theta_n))}{(\partial\theta_n)^2}\right) \\ cov\left(\frac{\partial \log(p(y_n;\theta_n))}{\partial\theta_n}, \frac{\partial^2 \log(p(y_n;\theta_n))}{(\partial\theta_n)^2}\right) & var\left(\frac{\partial^2 \log(p(y_n;\theta_n))}{(\partial\theta_n)^2}\right) \end{bmatrix}.$$

For any positive definite Ω_{v_n} the curvature of this family remains positive as $T \to \infty$ ($\lim_{T\to\infty} \gamma = \frac{2}{3}$), indicating that both elements of the sufficient statistic have nondegenerate distributions for large T, so that even asymptotically the sufficient statistic remains two-dimensional.

[6]If this is not immediately obvious, consider that a size α test of $\theta = 0$ that attempts to maximize power against alternatives of the form $\theta_i < 0$ can only have power α against alternatives of the form $\theta_j < 0$ if $i \neq j$.

form

$$\tau = \begin{cases} 0 & \text{if} \quad \gamma(\theta^a)'s < c \\ 1 & \text{otherwise} \end{cases}$$

Two properties are evident. First, the test is less likely to reject as any element of s increases, and second, the set of values of s for which the test accepts the null hypothesis is convex. Both properties are intuitive. Assuming for a moment that $\Omega_v = I$ and $d = 0$, then $s = [\ldots \triangle\, y_n' y_{n-} \ldots y_{n-}' y_{n-} \ldots]'$. If y_n is a random walk, then $\triangle y_n' y_{n-}$ should be close to zero, while if it is stationary then $\triangle y_n' y_{n-}$ should be negative; likewise $y_{n-}' y_{n-}$ should be larger if y_n is a random walk than if it is stationary. The second property, a convex acceptance set, is one shared by most tests in econometrics.[7]

So far we have only shown that an optimal test has these properties when a single alternative point is considered. The following lemma states that these two properties are shared by any Bayes solution with respect to a discrete set of alternative points.

Lemma 1 *Let τ be a Bayes solution in the strict sense with respect to a discrete probability measure π over the parameter space Θ. Then τ is a.e. equivalent to a nonincreasing and quasiconvex function of the sufficient statistic s.*

Proof: Pick $J+1$ discrete elements of Θ and let π be a probability measure over these points, $\pi(\theta^j)$ representing the probability of $\theta^j \in \Theta$, $j = 0 \ldots J$, where $\theta^0 = 0$ and $\theta^j \in \Theta_1$ for $j > 0$. By the Neyman-Pearson lemma, τ is a Bayes solution relative to π if

$$\tau = \begin{cases} 0 & \text{if} \quad p(y;0)\pi(0) > \sum_{j=1}^{J} p(y;\theta^j)\pi(\theta^j) \\ 1 & \text{if} \quad p(y;0)\pi(0) < \sum_{j=1}^{J} p(y;\theta^j)\pi(\theta^j) \\ 0 \le \bar{\delta} \le 1 & \text{if} \quad p(y;0)\pi(0) = \sum_{j=1}^{J} p(y;\theta^j)\pi(\theta^j) \end{cases}$$

We have seen that

$$p(y;\theta^j) = h(y)\exp\left(\gamma(\theta^j)'s\right).$$

[7]Within the context of *linear* exponential families Stein (1956) provides a general sufficient condition for admissibility of any test with a closed convex acceptance set in the sufficient statistics.

Define $\lambda(s;\theta^j) = \pi(\theta^j)\exp(\gamma(\theta^j)'s) > 0$. Given the absolute continuity of $p(y;\theta)$, τ is a.e. equivalent to

$$\tau = \begin{cases} 0 & \text{if } \sum_{j=1}^{J} \lambda(s;\theta^j) < \pi(0) \\ 1 & \text{otherwise} \end{cases}$$

Defining $\tau(s) = \sum_{j=1}^{J} \lambda(s;\theta^j)$, we can see that $\tau(s)$ is analytic and that

$$\begin{aligned} D_s\tau(s) &= \sum_{j=1}^{J} \lambda(s;\theta^j)\gamma(\theta^j)' \leq 0 \\ D_s^2\tau(s) &= \sum_{j=1}^{J} \lambda(s;\theta^j)\gamma(\theta^j)\gamma(\theta^j)' \geq 0 \end{aligned}$$

which implies that τ is nonincreasing and convex, therefore quasiconvex, function of the sufficient statistic s. QED.

While the set of Bayes solutions in the strict sense with respect to a discrete measure do not form an essentially complete class, Wald (1950) establishes that the set of Bayes solutions in the wide sense with respect to discrete measures do form an essentially complete set.[8] The content of the next theorem is that limits of Bayes solutions with respect to discrete probability measures share the two properties under discussion.

Theorem 2 *The set of nonrandomized tests that are nonincreasing, quasiconvex functions of the sufficient statistics is essentially complete.*

Theorem 2 states that one can always do at least as well by using a test that is a nonincreasing, quasiconvex function of the sufficient statistics. However, it does not state that one can always do better. Although it seems plausible that quasiconvexity is necessary for admissibility, at the moment only the weaker claim that a test must be a nonincreasing function of the sufficient statistics in order to be admissable can be proved.

Theorem 3 *Any test that is not almost everywhere equivalent to a nonrandomized test that is a nonincreasing function of the sufficient statistics is inadmissable.*

[8] If we were willing to assume that the intersection of the closure of Θ_0 and the closure of Θ_1 were empty, then the methodology of lemma 1 could be used to show that the set of nonrandomized tests that are nonincreasing, quasiconvex functions of the sufficient statistics is complete. However, this type of assumption is uncommon in econometrics.

3.2 Homogenous Alternatives

Theorems 2 and 3 characterize the set of admissable tests against any general alternative set $\Theta_1 = \Theta \setminus 0$. As noted in the introduction, some existing tests either explicitly or implicitly assume that the alternative set is one dimensional and that the parameter of interest is homogeneous across groups. If it is assumed that the θ_n are homogeneous across groups, so that $\theta_n = \bar{\theta}$ for $n = 1 \ldots N$, then there is a two-dimensional sufficient statistic for $\bar{\theta}$.

$$p(y;\bar{\theta}) = h(y) \exp\left(\bar{\theta}\left[i'_{NT} u'_{-} \Omega_v^{-1} \bigtriangleup u\right] - \frac{1}{2}\bar{\theta}^2 \left[i'_{NT} u'_{-} \Omega_v^{-1} u_{-} i\right]\right) \tag{6}$$

where i_{NT} is an $NT \times 1$ column vector of ones. As before, $p(y;\bar{\theta})$ forms a curved exponential family, however the curvature vanishes as $N \to \infty$ under the conditions of Assumption 1. Assuming that $\Omega_v = I$ (if this is not the case, a rotation of the variables that satisfies this condition can always be chosen when any nuisance parameters are known) equation (6) becomes

$$p(y;\bar{\theta}) = h(y) \exp\left(\bar{\theta}\left[\sum_{N=1}^{N} u'_{n-} \bigtriangleup u_n\right] - \frac{1}{2}\bar{\theta}^2 \left[\sum_{N=1}^{N} u'_{n-} u_{n-}\right]\right)$$

While $N^{-\frac{1}{2}} \sum_{N=1}^{N} u'_{n-} \bigtriangleup u_n$ has a nondegenerate distribution for large N, $N^{-1} \sum_{N=1}^{N} u'_{n-} u_{n-}$ converges weakly to a constant, so that for $\bar{\theta}$ within an $N^{-\frac{1}{2}}$ neighborhood of the null the density becomes

$$p(y;\bar{\theta}) = c(\bar{\theta}) h(y) \exp\left(\bar{\theta}\left[\sum_{N=1}^{N} u'_{n-} \bigtriangleup u_n\right]\right)$$

as $N \to \infty$. Standard results then imply the existence on a uniformly most powerful one-sided test. This is the logic behind the following theorem.

Theorem 4 *Let $\Theta \subset \{\theta \in R^N \mid \theta_n = \bar{\theta} \leq 0\}$. Under Assumption 1 and 2, for $N \to \infty$ there is a uniformly most powerful test of the null hypothesis $H_0 : \theta = 0$ against the alternative $H_1 : \theta \in \Theta \setminus 0$.*

The phenomenon discussed in this theorem can be confirmed in Figure 1[9], which shows the power envelope and a locally most powerful test of $\bar{\theta} = 0$ versus $\bar{\theta} < 0$ when $T = 10$ (all tests have 5% size). If a uniformly most powerful test exists, it must be equivalent to the locally most powerful test. As shown in the figure, for $N = 1$ the power of the local test is well below the power envelope for nonlocal alternatives. However by the time $N = 25$ the power of the local test is virtually indistinguishable from the power envelope.

[9] Details of all monte carlo experiments are described in the second appendix.

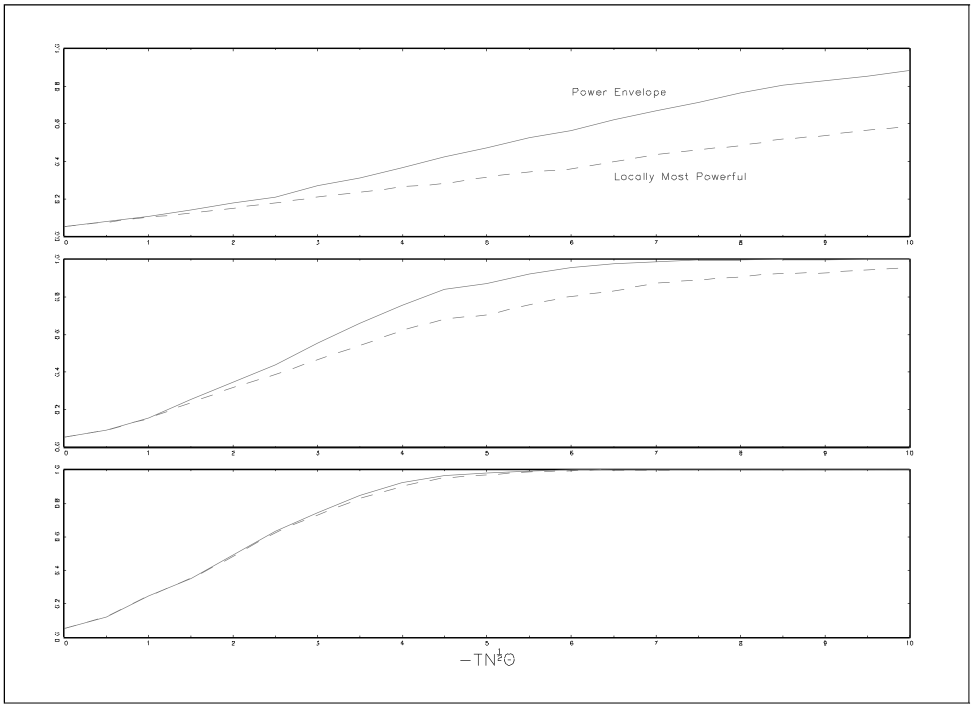

Figure 1: Power function of locally most powerful test of $\bar{\theta} = 0$ compared to the power envelope for N = 1 (top panel), N = 4 (middle panel), and N = 25 (bottom panel); 5% size, T = 10.

3.3 Relevance To Existing Tests

We will examine the Bonferroni, LL, and IPS tests in light of the previous section. The presentation will be made most clear by assuming that there are no nuisance parameters, so that $\Omega_v = I$, $d = 0$, and $s = [\ldots \triangle y_n' y_{n-} \ldots y_{n-}' y_{n-} \ldots]'$, however any statements made for this case will also be true when there are known nuisance parameters. Under these conditions the Levin and Lin test is equivalent to

$$\tau_{LL} = \frac{\sum_{n=1}^{N} \triangle y_n' y_{n-}}{\sqrt{\sum_{n=1}^{N} y_{n-}' y_{n-}}}$$

Since we have seen from Theorem 4 that the denominator will converge in probability to a constant, τ_{LL} will be uniformly most powerful against alternatives of the form $\theta_n = \bar{\theta} < 0$ for large N.

When there are no nuisance parameters the ADF-based IPS test is equivalent to

$$\tau_{IPS(ADF)} = \sum_{n=1}^{N} \frac{\triangle y_n' y_{n-}}{\sqrt{y_{n-}' y_{n-}}}$$

and the LM-based IPS test is equivalent to

$$\tau_{IPS(LM)} = \sum_{n=1}^{N} \frac{(\triangle y_n' y_{n-})^2}{y_{n-}' y_{n-}}.$$

Neither statistic is a non-increasing or quasiconvex function of s, therefore by Theorem 3 they are inadmissable.

The Maddala and Wu proposal of summing p-values across individual unit root tests would, if augmented Dickey-Fuller tests were used for each group, have the form

$$\tau_{MW} = \sum_{n=1}^{N} \log\left(P\left(\frac{\triangle y_n' y_{n-}}{\sqrt{y_{n-}' y_{n-}}}\right)\right)$$

and this is likewise neither non-increasing nor quasiconvex.

If unknown nuisance parameters are introduced then Theorems 2 and 3 must be extended, however it is straightforward to show that neither the IPS nor Maddala-Wu test (if it uses augmented Dickey-Fuller tests) will be nonincreasing or quasiconvex in the relevant subvector of sufficient statistics discussed in the extensions presented in the next section.

To obtain a sense of how the Bonferroni, LL and IPS tests perform in practice it is useful to introduce a benchmark that they can be compared to. The right panel of Figure 2 presents the point by point maximal power obtainable against possible alternatives by a symmetric test when there are two groups. It is the envelope of the set of uniformly most powerful invariant tests (where invariance is with respect to permutations of the groups) against each of the simple alternatives represented. With two groups ($N = 2$) and one hundred time series observations per group ($T = 100$) the figure graphs the envelope of a collection of Neyman-Pearson tests of the simple null $(\theta_1, \theta_2) = (0, 0)$ against simple alternatives $(\theta_1, \theta_2) = (\theta_1^a, \theta_2^a)$, where $(\theta_1^a, \theta_2^a) \in [-.01, -.20] \times [-.01, -.20]$. Each point on the z-axis represents the maximal power by a permutation-invariant test against the specific alternative represented by the x- and y-axes. The right panel of Figure 2 displays a segment of the symmetric power envelope for $N = 10$ (for $N > 2$ the full envelope cannot be shown). It graphs the maximal power against point alternatives in which $K \leq N$ groups are stationary with $\theta_n = \bar{\theta} < 0$ and the remaining $N - K$ groups have $\theta_n = 0$ as

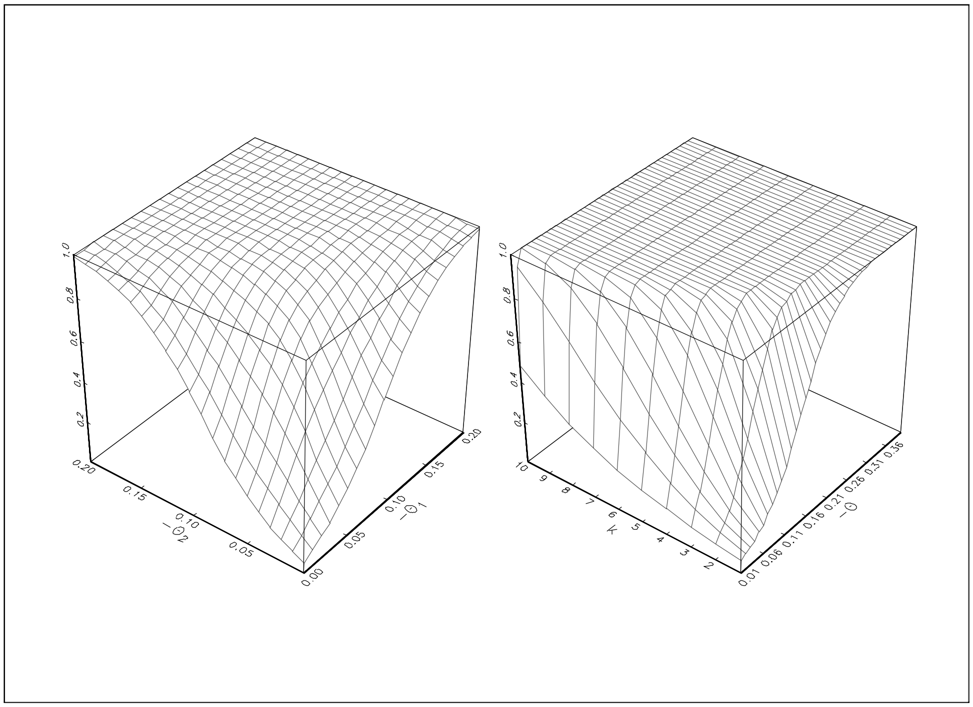

Figure 2: Full symmetric power envelope for N = 2 (left panel) and segment of symmetric power envelope for N = 10 (right panel); 5% size, T = 100.

K and $\bar{\theta}$ are varied over $K = 1 \ldots N$, $\bar{\theta} = -0.01 \ldots - 0.40$.[10] These envelopes provide a benchmark, however it is important to keep in mind that since there is no uniformly most powerful test for this multidimensional set of alternatives, it would be unrealistic to expect any test to be everywhere close to the power envelope.

Figure 3 compares the powers of the Bonferroni, LL and IPS tests to the symmetric power envelopes shown in Figure 2. The Bonferroni bound does well when there are a only a few outlier groups with nonzero θ_n and for those groups θ_n is well below zero. The LL test is quite near the power envelope when θ_n are close to being equal for all groups, but is well below the power envelope when most groups are I(1) and only a few are I(0). Since the results of Im, Pesaran, and Shin indicate that their ADF-based test has better power

[10]To generate this envelope each possible value of $\bar{\theta}$ requires consideration of $\sum_{K=1}^{N} \frac{N!}{K!(N-K)!}$ alternative points per monte carlo replication, so that in practice generating this type of envelope for N much larger than 10 quickly becomes infeasible.

properties than their LM-based test in small samples, we focus on the ADF-based test in Monte Carlo experiments. The IPS ADF-based test is seen to be fairly close to the LL test, doing a bit less well when the θ_n are nearly equal for all groups, but better than the LL test against other alternatives.

4 A New Class of Tests

The Im, Pesaran, and Shin tests have been shown to be inadmissable in the absence of unknown nuisance parameters. In the absence of unknown nuisance parameters the Levin and Lin test meets the admissability criteria presented in section 3, and if N is of reasonable size will be close to uniformly most powerful when the parameters of interest are homogeneous, however it does poorly against other alternatives. The Bonferroni bound does well when there are a few stationary outliers mixed in with groups that are I(1), but does not do well when all or most groups are stationary and highly persistent. Given this, it may be desirable to look for a new form of test.

The theory of the previous section demonstrated that Bayes tests will automatically be nonincreasing and quasiconvex, therefore it is natural to explore tests from that set. In doing so the emphasis will be on choosing a weighting function over alternatives that yields a convenient test with generally good properties, rather than on choosing a function based upon actual a priori beliefs. For instance, Wald tests correspond to Bayes tests with weighting functions that are constant on certain ellipses, and locally most powerful tests correspond to Bayes tests that put weight only on alternatives very near the null. These weighting functions are used not because they are plausible, but because they yield easily calculable tests that in certain circumstances have good properties.

If $\pi(\theta_1, \ldots, \theta_N)$ is a probability measure over $\theta \in R^N_-$, then a test that is a Bayes solution with respect to π will have the form

$$\int_{-\infty}^{0} \cdots \int_{-\infty}^{0} \frac{p(y; \theta_1, \ldots, \theta_N)}{p(y; 0, \ldots, 0)} \pi(\theta_1, \ldots, \theta_N) d\theta_1 \cdots d\theta_N \tag{7}$$

For the moment assume that the error term, v_n, is independent across groups — correlated errors will be discussed later. In the current context one flexible class of tests that does well is the following. If we assume that the measure π is such that each θ_n is independent with identical marginal probability densities then (with some abuse of notation) $\pi(\theta_1, \ldots, \theta_N) = \prod_{n=1}^{N} \pi(\theta_n)$. Defining

$$\lambda(y_n; \theta_n) = \frac{p(y_n; \theta_n)}{p(y_n; 0)},$$

independence of v_n across groups implies that $\frac{p(y;\theta_1,\ldots,\theta_N)}{p(y;0,\ldots,0)} = \prod_{n=1}^{N} \lambda(y_n;\theta_n)$ and (7) becomes

$$\int_{-\infty}^{0} \lambda(y_1;\theta_1)\pi(\theta_1)d\theta_1 \cdots \int_{-\infty}^{0} \lambda(y_N;\theta_N)\pi(\theta_N)d\theta_N. \tag{8}$$

Taking the log yields the test statistic

$$\tau_B = \sum_{n=1}^{N} \log\left(\pi(0) + \int_{\theta<0} \lambda(y_n;\theta)\pi(\theta)d\theta\right)$$

Like the tests proposed by Levin and Lin; Im, Pesaran, and Shin; and Maddala and Wu, this test sums individual test statistics across groups, however it has a clear weighting function against the multivariate set of alternatives. A test that rejects for large values of τ_B is equivalent to a Bayes solution in the strict sense with respect to a probability measure that assigns probability $\pi(\theta)$ to the event $\theta_n = \theta$ independent of the values of the other θ_n.

The nature of the test as $\pi(0)$ approaches the extremes 0 or 1 can be understood by rewriting (8)

$$\prod_{n=1}^{N} \left(\pi(0) + (1-\pi(0))\int_{\theta<0} \lambda(y_n;\theta)\pi(\theta;\theta<0)d\theta\right)$$

and recognizing that the term $\pi(0)^N$ in the resulting polynomial is a constant, and therefore immaterial to size or power of the test since it can be absorbed in the critical value. Holding the conditional probabilities $\pi(\theta;\theta<0)$ constant and taking the limit as $\pi(0) \to 1$ yields a test statistic equivalent to

$$\sum_{n=1}^{N} \int_{\theta<0} \lambda(y_n;\theta)\pi(\theta;\theta<0)d\theta \tag{9}$$

Taking the limit as $\pi(0) \to 0$ yields a test statistic equivalent to

$$\prod_{n=1}^{N} \int_{\theta<0} \lambda(y_n;\theta)\pi(\theta;\theta<0)d\theta \tag{10}$$

A test that rejects for large values of (9) is a Bayes solution with respect to a probability measure that assigns probability one to alternatives in which there is a single group that is $I(0)$, each group having equal weight.[11] Equation

[11] The critical value for the Bonferroni bounds is calculated by assuming that with probability one there is at most one group that will exceed the critical value under the null, each group having equal probability of being the one to reject, which is suggestive of (9).

(10) is a Bayes solution with respect to a probability measure that assigns probability one to alternatives in which all N groups are $I(0)$. Further, if $\pi(\theta; \theta < 0)$ is chosen to place probability one on a single point θ^a then (10) becomes $\prod_{n=1}^{N} \lambda(y_n; \theta^a)$, and as $\theta^a \to 0$ this statistic is equivalent to a locally most powerful test of the null hypothesis when the θ_n are constrained to be homogenous — a test that the previous section has shown will be arbitrarily close to the uniformly most powerful test (and the Levin and Lin test) for large N in the absence of nuisance parameters.

For purposes of computation the integral can be approximated by approximating $\pi()$ with a discrete probability measure; in practice it appears that a probability measure over at most four or five discrete points is sufficiently flexible. In univariate curved exponential families it is frequently the case that Neyman-Pearson tests of the null against a fixed alternative have good power, so long as the alternative is in a range for which the test has 50% to 80% power (see Effron (1975) and Davies (1969); Elliot, Rothenberg, and Stock (1996) is an example in the case of a univariate unit root test), so testing against a fixed discrete distribution over alternatives in this range has some appeal in the multivariate context. In this case the test statistic becomes

$$\tau_B = \sum_{n=1}^{N} \log \left(\pi_0 + \sum_{j=1}^{J} \lambda(y_n; \theta^j) \pi_j \right).$$

By varying π one can produce tests which have maximal power against clear regions of the alternative space. This feature is shown in Figure 4, where $\pi(\theta; \theta < 0)$ is chosen to place probability one on a single point, θ^a, and the parameters (π_0, θ^a) are varied.

Figure 4 demonstrated that Bayes tests can closely match both the Levin and Lin (when $\pi_0 = 0$) and Bonferroni (when $\pi_0 = 1$) bound. This suggests that Bayes tests that move π_0 away from these two limiting values could potentially do about as well in the regions where the Levin and Lin or Bonferroni bound maximize power, while perhaps doing better elsewhere. As the top and middle panels of Figure 5 show, Bayes tests can both mimic the LL and Bonferroni tests and do much better in areas where the tests are weak without giving up significant power in areas of the alternative space where the tests are good (for brevity only the more relevant case of $N = 10$ will be shown from this point on). The previous section proved that the IPS test was inadmissable, implying that there is at least one test that does uniformly better. While it isn't possible to prove that a test of the chosen form will uniformly dominate the IPS test, the bottom panel of Figure 5 shows that the Bayes tests have the ability to significantly outperform the IPS test over a fairly representative area of the alternative space.[12]

Allowing for correlation across groups introduces further considerations. Both Levin and Lin and Im, Pesaran, and Shin recommend subtracting time averages to control for time effects — one standard method for allowing correlation across groups in panel data. If $\bar{y}_t = \frac{1}{N}\sum_{n=1}^{N} y_{nt}$ then (1) implies that

$$(\triangle y_{nt} - \triangle \bar{y}_t) = \theta_n(y_{nt-1} - \bar{y}_{t-1}) + v_{nt} \tag{11}$$

where v_n is an $I(0)$ process. There are several reasons why subtracting time averages in this manner is unattractive. First, the time effects model is convenient when there is a single common factor affecting each group in the same manner and the principle desire is to answer questions concerning the data *conditional on the common factor*. However in testing for unit roots the time series properties of such a factor is of direct interest — if there is a common factor and it is $I(1)$ then the null hypothesis is correct — and so conditioning on any common factor will yield a test with incorrect size if the null hypothesis is that the unconditional series are $I(1)$. Second, while it is true that removing time effects is appropriate if there is a common factor that is known to be $I(0)$ and only homogenous alternatives are considered, the same procedure will yield an inconsistent test if heterogeneous alternatives are considered. If group n is $I(0)$ while some other groups in the panel are $I(1)$ then $(y_{nt} - \bar{y}_t)$ will contain a unit root even if y_{nt} does not, and a test based on (11) will tend to find that $\theta_n = 0$ when it is not. The top panels of Figure 6 displays the power functions of the Levin and Lin and Im, Pesaran and Shin ADF tests when the error term has an idiosyncratic and common component, $e_{nt} = \varepsilon_{nt} + c_t$, where both ε and c are i.i.d. $N(0,1)$ variables and time effects are used to remove the component c_t from the data. As can be seen, removing time effects reduces the power of both tests against heterogeneous alternatives — most markedly for the IPS test since the LL test already did poorly against such alternatives.

Rather than time effects, error correlation will be modelled here by supposing that (1) holds with $\Omega = Ee_te_t'$ nondiagonal. This usefully circumvents the first objection to the time effects model raised above.[13] Because it is a Bayesian test with an explicit prior, the test can be adjusted to allow for correlation across groups in a conceptually straightforward manner based on (7).

[12] While the previous section proved the existence of a test that is uniformly better than the IPS test, there are no practical methods of finding such a test. Since it is infeasible to examine the entire (ten-dimensional) space of alternatives, it cannot be and is not claimed that the test shown is uniformly better. The only claim is that it presents an attractive alternative.

[13] This model is also used by Abuaf and Jorion (1990) and O'Connell (1998). Given a consistent estimate $\hat{\Omega} = PP'$, both employ a t-test based on the pooled GLS estimator of θ

The test statistic will no longer have the form of a simple sum across groups since covariance terms will also be included in τ_B, but the analogue can be calculated. The middle and bottom panels of Figure 6 graph the relative power of the Bayes tests shown in Figure 4, using the Monte Carlo experiment described above. As can be seen, in contrast to the IPS and LL tests the power of the tests (with the exception of $\pi_0 = 0$) against heterogeneous alternatives actually *rises*.

For the purposes of comparison the power functions in Figure 6 are shown relative to the power envelope used in the previous figures, where the error term was independent across groups, explaining why the Bayes tests are actually above the envelope in the figure. Comparing the power functions to the power envelope generated with the true Ω would make the performance of LL and IPS tests appear worse and would make the powers of the Bayes tests appear fairly close to those shown in Figure 4. When the covariance matrix $\Omega = Ee_te_t'$ is known the maximal power possible against any given alternative point is higher for heterogeneous alternatives (it is unchanged for homogeneous alternatives), and this is the source of the increased power of the Bayes tests. (The performance of the Bayes tests can be improved by adjusting its parameters to reflect this change in the power envelope, but this is not done in Figure 6). As noted in a previous footnote, versions of the IPS and LL test that make use of the known Ω matrix have results that are quite close to the time effects model, so the better relative performance of the Bayes tests is not due to their use of the known covariance matrix for this particular experiment. Of course, when Ω is unknown it must be estimated and this will diminish power for small T.

This performance is not completely costless. The critical values for the Bayes tests will in general depend on parameters of the covariance matrix when error correlation is present and they become computationally more bur-

obtained by fitting form (1) to the orthogonalized data:

$$P^{-1} \triangle y_t = \theta P^{-1} y_{t-1} + \bar{\beta}'(\triangle \bar{z}_t - \theta \bar{z}_{t-1}) + \beta' P^{-1} z_t + P^{-1} e_t$$

This procedure corresponds to a variant of the LL test. As with time effects, this would be valid for homogeneous alternatives (where θ is scalar), however if heterogeneous alternatives are considered it will yield an inconsistent test since if *any* of the N groups in y_t is $I(1)$ then in general *all* of the N data series in $P^{-1}y_t$ will be $I(1)$ (unless θ is scalar or Ω is diagonal then in general $P^{-1}\theta \neq \theta P^{-1}$). A monte carlo experiment simliar to the one used to generate figure 6 found that the power functions of the Abuaf and Jorion/O'Conell test and also the IPS-ADF test when the data is first orthogonalized using the true Ω are quite close to the corresponding power functions (shown in the top panels of figure 6) when time effects are removed, although the IPS test does somewhat better for alternatives near the null by orthogalizing than by removing time effects.

densome. However, the dependence of the critical value on the parameters of the covariance matrix may be simply unavoidable if one wishes to have nontrivial power against heterogenous alternatives in this context, and computational costs can be cut by using approximations to the Bayes test that are computationally simpler (if less well theoretically based) and appear to preserve power.

5 Unknown Nuisance Parameters

In this section the restriction that nuisance parameters are known will be relaxed. In order to parameterize the model it will be assumed that d_{nt} is a linear combination of a set of D nonstochastic regressors, $\bar{z}_t$, and that the process generating the data can be written in the form:

$$\triangle y_t = \theta y_{t-1} + \bar{\beta}' \left(\triangle \bar{z}_t - \theta \bar{z}_{t-1} \right) + \beta' z_t + e_{nt} \tag{12}$$

where

$$\theta = \begin{bmatrix} \theta_1 & & \\ & \ddots & \\ & & \theta_n \end{bmatrix}$$

is an $N \times N$ diagonal matrix[14], and

$$z_t = \begin{bmatrix} \triangle y'_{t-1} & \dots & \triangle y'_{t-P} \end{bmatrix}'$$

is a $PN \times 1$ vector of observed data. The first $P+1$ data points, $y_{1-P}, \dots, y_0$, will be considered fixed and e_t will be taken to be a sequence of independent N-dimensional Gaussian variables with mean zero and variance Ω_e.

Equation (12) is more restrictive than (2) - (3), however it continues to encompass (1) and generalizes the assumption made in Levin and Lin and Im, Pesaran, and Shin, that the error term is independent across groups.

If x_t is a given $K \times 1$ vector then let x, $\triangle x$ and x_- respectively be the $K \times T$ matrices $[x_{kt}]$, $[\triangle x_{kt}]$ and $[x_{kt-1}]$, where $k = 1 \dots K$, $t = 1 \dots T$. We can then write:

$$\begin{aligned} e &= \triangle y' - \theta y_- - \bar{\beta}' \left(\triangle \bar{z}' - \theta \bar{z}'_- \right) - \beta' z \\ &= Q' M \end{aligned}$$

where $Q' = \begin{bmatrix} I_N & -\theta & -\bar{\beta}' & -\theta\bar{\beta}' & -\beta' \end{bmatrix}$ and $M' = \begin{bmatrix} \triangle y' & y'_- & \triangle \bar{z}' & \bar{z}'_- & z' \end{bmatrix}$.

[14] We will switch between defining θ as an $N \times 1$ vector to defining it as an $N \times N$ diagonal matrix with elements θ_n on the diagonal; the use should be clear from context.

The probability density for y is

$$\begin{aligned} p(y;\theta,\xi) &= (\det(2\pi\Omega_e))^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}tr\left(e'\Omega_e^{-1}e\right)\right\} \\ &= c(\xi)\exp\left\{-\frac{1}{2}tr\left(MM'Q\Omega_e^{-1}Q'\right)\right\} \\ &= c(\xi)\exp\left\{-\frac{1}{2}avech\left(Q\Omega_e^{-1}Q'\right)' vech\left(MM'\right)\right\} \end{aligned}$$

As in the previous section, this describes a curved exponential family, now with sufficient statistic $vech\,(MM')$. Partition the sufficient statistic into two subvectors, $vech\,(MM') = [s'\ r']'$, where

$$s' = [\ldots \triangle\, y_n' y_{n-} \ldots y_{n-}' y_{n-} \ldots]$$

is a $2N$ vector and r contains the remaining elements of $vech\,(MM')$. The only restrictions placed upon the parameters are that $\theta \leq 0$ and that Ω_e is a positive definite and symmetric matrix. As a result most of the vector $avech\,(Q\Omega_e^{-1}Q')$ is unrestricted, however the restrictions that have been placed are sufficient to guarantee knowledge of the sign of the components of $avech\,(Q\Omega_e^{-1}Q')$ associated with s. This knowledge allows Theorems 2 and 3 to be generalized with respect to the subvector s holding the remaining elements of the sufficient statistic constant — that is, to consider the behavior of any test $\tau(s,r)$ as r is held fixed. In this sense the presence of unknown nuisance parameters forces Theorems 2 and 3 to be weakened, however the set s considered in this section is directly comparable to the set of sufficient statistics considered in the previous section when examining the tests suggested by Levin and Lin, Im, Pesaran and Shin, and Maddala and Wu, so our generalizations will be sufficiently strong to be of use in understanding current panel unit root tests.

Defining the $N \times N$ matrices $\beta_1 \cdots \beta_P$, where $\beta' = \left[\begin{array}{ccc} \beta_1 & \ldots & \beta_P \end{array}\right]$, the vector $\xi \equiv (vec(\bar{\beta})', vec(\beta)', vech(\Omega_e)')$ and $v_t \equiv (I - \beta_1 L - \cdots - \beta_P L^P)^{-1} e_t$, we suppose that $\xi \in \Xi$, where Ξ is a linear subspace of $R^{(D+PN+\frac{N+1}{2})N}$ containing the point $\xi_0 = (0,\ldots,0,vech(I)')$ such that v_t satisfies the conditions of Assumption 1 for all elements ξ and consider the parameter set $\Theta \times \Xi$. We study tests with composite null $\{(\theta,\xi) \in \Theta \times \Xi \mid \theta = 0\}$ and alternative set $\{(\theta,\xi) \in \Theta \times \Xi \mid \theta \neq 0\}$.

The following theorem is counterpart to Theorems 2 and 3:

Theorem 5 *(1)The set of nonrandomized tests that are nonincreasing, quasiconvex functions of s conditional on r is essentially complete. (2) Any test that is not almost everywhere equivalent to a nonrandomized test that is a nonincreasing function of s conditional on r is inadmissable.*

5.1 Homogenous Alternatives with Nuisance Parameters

The results of Section 4 clearly lessen the practical importance Theorem 4. Although a uniformly most powerful test exists for known nuisance parameters when alternatives are restricted to be homogenous in the parameter of interest, Bayes tests exist that do virtually as well against such alternatives and substantially better over others. Nonetheless, extensions of Theorem 4 to cases with unknown nuisance parameters has theoretical interest, if only in pointing out certain difficulties that have been missed in earlier literature on panel unit root tests.

In this subsection it will be assumed, as in Levin and Lin, that the data is generated by (1) and e_{nt} is independent across groups. The full set of assumptions are:[15]

Assumption 3 $\{y_{nt}\}_{n=1\ldots N,t=1\ldots T}$ *is generated by (1) where the* e_{nt} *are independent* $N(0,\sigma_n^2)$*;* $y_{nt}=0$ *for* $t\leq 0$*. If* $v_{nt}=\left(1-\beta_{n1}L-\cdots\beta_{nP}L^P\right)^{-1}e_{nt}$*, then* v_n *is stationary with strictly positive spectral density and a moving average representation* $v_{nt}=\sum_{j=0}^{\infty}\psi_{nj}e_{t-j}$ *that for* $n=1\ldots N$ *satisfies the conditions* $\sum_{j=0}^{\infty}j\mid\psi_{nj}\mid<\infty$*,*

$$\lim_{N\to\infty}N^{-1}\sum_{n=1}^{N}\left(\sum_{j=0}^{\infty}j\mid\psi_{nj}\mid\right)^2<\infty$$

$$\lim_{N\to\infty}N^{-2}\sum_{n=1}^{N}\left(\sum_{j=0}^{\infty}j\mid\psi_{nj}\mid\right)^4=0.$$

Since the work of Levin and Lin emphasis has largely been on models that allow β_n and σ_n to differ across groups as $N\to\infty$. Tests with an infinite number of nuisance parameters are inherently difficult to deal with, and here the asymptotics will have to be altered to let $T\to\infty$ at a fast enough rate to obtain consistent estimates of the heterogeneous nuisance parameters. The primary difficulty with allowing nuisance parameters to differ across groups as both $N\to\infty$ and $T\to\infty$ is that tools such as the continuous mapping

[15] As is standard within the literature on univariate unit root tests, Levin and Lin assume that the number of lagged differences of the dependent variable in each regression may grow at rate $o(T^{1/4})$. However Faust (1998) has shown that uniform rather than pointwise convergence is required to obtain tests with nontrivial size, and that uniform convergence requires stronger condtions on the moving average representation of the process than is implied by allowing the number of lagged differences to grow at such a rate. Whatever conditions are to be assumed, it is clearly necessary that there be some uniform bound on the number of lags in each equation across groups in the current context.

theorem cannot be automatically applied to triangular arrays, as demonstrated in the following claim:

Claim 6 *$x_{nT} = o_p(1)$ for $n = 1 \ldots N_T$ does not imply $N_T^{-1} \sum_{n=1}^{N_T} x_{nT} = o_p(1)$.*

Proof: Examples are simple to create. Suppose for instance that $N_T = kT^{\gamma}$ for some $\infty > \gamma$, $k > 0$. In this case consider the indexed array of independent random variables x_{nT}, where $x_{nT} = 0$ with probability $1 - T^{-\gamma}$ and $x_{nT} = T^{\gamma}$ with probability $T^{-\gamma}$. Then $\lim_{T \to \infty} pr(x_{nT} = 0) = 1$ for $n = 1 \ldots N_T$, but

$$\lim_{T \to \infty} pr(N_T^{-1} \sum_{n=1}^{N_T} x_{nT} = \frac{1}{k}) = k \exp(-k) > 0$$

where $\frac{1}{k} > 0$.

The example demonstrates that even if certain terms weakly converge to zero for each group as $T \to \infty$, the average across groups need not converge to zero, or converge at all, if both $N \to \infty$ and $T \to \infty$, even if T grows at a faster rate than N.

To see understand the complications this presents in the current context, assume for a moment that there are no deterministic components and define the matrices $Q_n = z_n(z_n' z_n)^{-} z_n'$, $M_n = I_T - Q_n$ and the scalar $\psi_n(1) = \sum_{j=0}^{\infty} \psi_{nj}$. Levin and Lin assert that the pooled least squares estimator of θ has a limiting normal distribution:

$$\hat{\theta} = N^{-\frac{1}{2}} T^{-1} \sum_{n=1}^{N} \hat{\sigma}_n^{-2} y_{n-}' M_n e_n \Longrightarrow N(0, \frac{1}{2} V) \tag{13}$$

where $\hat{\sigma}_n^2$ is a consistent estimate of σ_n^2 under the null and

$$V \equiv \lim_{N \to \infty} N^{-1} \sum_{n=1}^{N} \psi_n(1)^2.$$

Under the assumptions of this section it is relatively straightforward to establish that

$$N^{-\frac{1}{2}} T^{-1} \sum_{n=1}^{N} \sigma_n^{-2} y_{n-}' e_n \Longrightarrow N(0, \frac{1}{2} V).$$

Therefore if (13) is true then it must imply

$$\phi \equiv N^{-\frac{1}{2}} T^{-1} \sum_{n=1}^{N} \left(\left(\hat{\sigma}_n^{-2} - \sigma_n^{-2} \right) y_{n-}' M_n e_n - \sigma_n^{-2} y_{n-}' Q_n e_n \right) = o_p(1).$$

It is in fact the case that (by the continuous mapping theorem) ϕ will converge to zero as $T \to \infty$ *holding N fixed*, but the example above shows that this is insufficient. Instead, a law of large numbers must be applied directly to ϕ. Billingsly's weak law of large numbers for triangular arrays states that if $Var(\phi) \to 0$ then $p\lim \phi = \lim E\phi$, requiring statements about the limit of the first two moments of ϕ. This is nontrivial because the moments of ϕ involves moments of inverted quadratic forms in the error term e_{nt}, and in the case of Q_n it involves a rather complicated function of such forms. The conditions of Levin and Lin are too weak to establish that ϕ even has moments, let alone that they converge to zero.[16]

The most general counterpart to Theorem 4 that can be constructed at this point is accordingly fairly limited and given by the following, where invariance refers to invariance to location (more formally, invariance to the parameters comprising $\lim_{T\to\infty} E_0 y_T$, which refers to both invariance to $\bar{\beta}$ under either the null or alternatives and to invariance to the initial value, y_0, under the null):

Theorem 7 *Let $\Theta \subset \{\theta \in R^N \mid \theta_n = \bar{\theta} \le 0\}$. (1) Under Assumption 3, for $N \to \infty$, $T \to \infty$, $NT^{-1} \to 0$ there is a uniformly most powerful invariant test of the null hypothesis $H_0 : (\theta, \xi) \in 0 \times \Xi$ against the alternative $H_1 : (\theta, \xi) \in (\Theta \setminus 0) \times \Xi$ if $N^{-\frac{1}{2}} T^{-1} \sum_{n=1}^{N} \sigma_n^{-2} y_{n-}' Q_n e_n = o_p(1)$. (2) If in addition $P \le 1$ and $N^{-1} \sum_{n=1}^{N} \psi_n(1)^2 \gamma_{n0}^{\frac{1}{2}} < \infty$, $N^{-2} \sum_{n=1}^{N} \psi_n(1)^4 \gamma_{n0} < \infty$, where $\gamma_{n0} = \lim_{T\to\infty} E(\triangle y_{nt}^2)$, then $N^{-\frac{1}{2}} T^{-1} \sum_{n=1}^{N} \sigma_n^{-2} y_{n-}' Q_n e_n = o_p(1)$.*

The first part of Theorem 7 states a general sufficient condition for existence of a uniformly most powerful invariant test. The proof establishes that under Assumption 3

$$N^{-\frac{1}{2}} T^{-1} \sum_{n=1}^{N} \left(\hat{\sigma}_n^{-2} - \sigma_n^{-2} \right) y_{n-}' M_n e_n = o_p(1),$$

it is the second term in ϕ that presents more difficulty. This difficulty is demonstrated in the second part of the theorem, where it can only be established by assuming that the first-differenced data is generated by at most an AR(1) under the null. Whether this can be generalized to higher order autoregressive

[16]Establishing that the needed moments are finite for finite T is relatively straightforward, but this is not sufficient. Levin and Lin implicitly assume that if x_T is some random variable such that Ex_T is finite for finite T and $E\lim_{T\to\infty} x_T$ is finite, then $\lim_{T\to\infty} Ex_T = E\lim_{T\to\infty} x_T$. This does not follow without stronger restrictions (for instance, in the example to the claim in the text $Ex_{nT} = 1$ for all T, but $E\lim_{T\to\infty} x_{nT} = 0$.) In fact, even if it is true that $\lim_{T\to\infty} Ex_{nT} = E\lim_{T\to\infty} x_{nT}$, this would not be of direct use unless it could also be established that $\lim_{T\to\infty} N_T \sum_{n=1}^{N_T} Ex_{nT} = E\lim_{T\to\infty} N_T \sum_{n=1}^{N_T} x_{nT}$

models is unclear. There is a fairly wide literature on the exact moments of least squares estimators for autoregressive models with normal errors in the AR(1) case, including Sawa (1978) and Nankervis and Savin (1988). There is no corresponding literature for AR(P) models, indicating that the task may be difficult, since the analytics that would be used to calculate the moments of the estimators for a higher order autoregressive model are similar to what would be required to establish the desired generalization. Without machinery that can either establish the existence of the required limiting moments or a law of large numbers that does not require their existence, analysis of panel tests in which both N and T grow large appears difficult in the present context of autoregressive models in which the coefficients on lagged differences are group-specific nuisance parameters.

6 Adjusting the Bayes Tests for Nuisance Parameters

Unknown nuisance parameters force certain practical considerations in finite samples. Typical sizes of N and T in panel data will preclude fitting the data as a general vector autoregression since doing so would necessitate the estimation of an unreasonably large number of parameters. Accordingly, the focus of this section will be on constructing tests when each series can be written as a linear function of its own lags, as in (1). Since standard tests of homogeneity of the nuisance parameters, $\beta_n = \beta$ and $\sigma_n = \sigma$, will have chi-squared distributions under the usual regularity conditions regardless of whether the data contain I(1) components, it may make sense to test this hypothesis and impose it if it cannot be rejected before moving on to a unit root test. However, in most macroeconomic data sets T will be large enough relative to N for consistent estimators of group specific nuisance parameters to yield reasonably precise estimates and here it will simply be assumed that — as seems most relevant — the nuisance parameters are heterogeneous. In light of the previous section and because with macroeconomic data sets it seems most compelling to consider N fixed, it will be assumed that N is fixed and $T \to \infty$ (or more informally we will think of T being large relative to N).

Assuming independence across groups, but allowing for the other nuisance parameters, the proposed test can be adjusted to one having a form similar to (8). When there known nuisance parameters and (1) holds, the Bayes tests of section 4 are defined in terms of

$$\lambda(y_n;\theta_n) = \frac{\theta_n}{\sigma_n^2}(\triangle y_n - \bar{\beta}' \triangle \bar{z}_n - \beta z_n)'(y_{n-} - \bar{\beta}'\bar{z}_n) - \frac{\theta_n^2}{2\sigma_n^2}(y_{n-} - \bar{\beta}'\bar{z}_n)(y_{n-} - \bar{\beta}'\bar{z}_n).$$

Defining (as in the previous section, but now with deterministic components) the least squares projection matrix

$$M_n = I - \begin{bmatrix} \bar{z}_n & z_n \end{bmatrix} \left(\begin{bmatrix} \bar{z}_n & z_n \end{bmatrix}' \begin{bmatrix} \bar{z}_n & z_n \end{bmatrix} \right)^{-} \begin{bmatrix} \bar{z}_n & z_n \end{bmatrix}'$$

the LL, IPS, and Bonferroni tests are functions of the statistics $\hat{\sigma}_n^{-2} \triangle y_n' M_n y_{n-}$ and $\hat{\sigma}_n^{-2} y_{n-}' M_n y_{n-}$, where $\hat{\sigma}_n$ is a consistent estimator of σ_n. For fixed N and $T \to \infty$ the statistic

$$\theta_n(\hat{\sigma}_n^{-2} \triangle y_n' M_n y_{n-}) - \frac{\theta_n^2}{2}(\hat{\sigma}_n^{-2} y_{n-}' M_n y_{n-})$$

is asymptotically equivalent to $\lambda(y_n; \theta_n)$ under the null and will inherit the same invariance properties that the other tests have.

The distribution of $\lambda(y_n; \theta_n)$ under the null will depend on the nuisance parameter $\psi_n(1)$ (see, for instance, Hamilton (1994)). The test statistic is easily adjusted to be asymptotically invariant to the full set of nuisance parameters by using a consistent estimator $\hat{\psi}_n(1)$ of $\psi_n(1)$ and defining

$$\lambda(y_n; \theta, \hat{\psi}_n(1)) = \frac{\theta}{\hat{\psi}_n(1)}(\hat{\sigma}_n^{-2} \triangle y_n' M_n y_{n-}) - \frac{1}{2}\left(\frac{\theta}{\hat{\psi}_n(1)}\right)^2 (\hat{\sigma}_n^{-2} y_{n-}' M_n y_{n-})$$

The proposed test statistic is then

$$\tau_B = \sum_{n=1}^{N} \log \left(\pi_0 + \sum_{j=1}^{J} \lambda(y_n; \theta^j, \hat{\psi}_n(1)) \pi_j \right)$$

where $\hat{\beta}_n$, $\hat{\sigma}_n$ are estimated from (1) using OLS and the estimate of $\psi_n(1)$ is

$$\hat{\psi}_n(1) = \left(1 - \sum_{p=1}^{P} \hat{\beta}_{nj}\right)^{-1}$$

As discussed in section 4, when $\Omega = E e_t e_t'$ is nondiagonal then in general Bayes tests will have increased power against heterogeneous alternatives but will not be invariant to the full set of nuisance parameters, and critical values must be calculated on a case by case basis, for instance through Monte Carlo methods.[17]

[17]Defining $\tilde{y}_n = M_n y_n$ and $\hat{\psi}(1)$ as the diagonal matrix with elements $\hat{\psi}_1(1) \cdots \hat{\psi}_N(1)$, then for nondiagonal Ω the general form of $\lambda(y; \theta, \hat{\psi}(1))$ is

$$tr(\theta \hat{\psi}(1)^{-1} \hat{\Omega}^{-1} \sum_{t=1}^{T} \triangle \tilde{y}_t \tilde{y}_{t-1}') - \frac{1}{2} tr(\theta \hat{\psi}(1)^{-1} \hat{\Omega}^{-1} \hat{\psi}(1)^{-1} \theta \sum_{t=1}^{T} \tilde{y}_{t-1t} \tilde{y}_{t-1}')$$

Three Monte Carlo experiments were run in order to study the properties of the Bayes test in the presence of unknown nuisance parameters. In the first it was assumed that groups were independent with $\bar{z}_{nt} = 0$ and an AR(2) was estimated for each group to provide estimates of $\psi_n(1)$ and σ_n^2 (the data were generated with $\beta_n = 0$ and $\sigma_n^2 = 1$ for each n and $T = 100$). Table 1 displays the powers of the LL, IPS, Bonferroni, and two Bayes tests for alternatives in which there are K groups that are stationary with $\theta = \bar{\theta}$. The fact β_n and σ_n must be estimated results in relatively small losses in power that are roughly similar across the tests, and so the conclusions made in the case with known nuisance parameters remain. While the LL test tends to do somewhat better when $N = K$, the IPS test does nearly as well and does better against other alternatives, and the Bonferroni test does best when only a small proportion of the groups are stationary. The parameters of the test Bayes1 where chosen so that it was roughly similar to the IPS test for homogeneous alternatives and better for heterogeneous alternatives. The test Bayes2 was chosen to do roughly as well as the Bonferroni for heterogeneous alternatives and better for homogenous alternatives. While the Bayes test and IPS test have quite similar power when K is close to N, the Bayes test does better than the IPS test when K is small, having up to 60% better power against IPS and up to 85% better power against LL. Likewise, the power of the Bayes2 test is quite close to the Bonferroni for $K = 1$ or 3, while having up to 60% better power for $N = K$.

Tables 2 and 3 repeat the exercise with a constant (Table 2) and a constant and time trend (Table 3) included in each regression. Because estimation of either a group specific constant or group-specific constant and time trend noticeably reduces the power of all tests, the *relative* power of the Bayes tests tends to be diminished (since there is a lower bound on the power of any test). However appropriately chosen forms of the Bayes tests (again labelled Bayes1) continue to have as much as 30% better power in Table 2 and 25-30% better power in Table 3 than the IPS test and as much as 70-80% better power than LL in either table when K is small, while having nearly identical power to the IPS test when K is close to N. And forms of the Bayes tests (again labelled Bayes2) continue to have up to 70% better power in Table 2 and 50% better power in Table 3 than the Bonferroni test when $K = N$, while having nearly identical power for smaller K.

These results argue that Bayes tests can be a feasible and attractive alternative to the existing tests. While existing tests may be simpler to calculate and can (in the cases of the IPS and Bonferroni tests) be based on output from standard statistical packages, Bayes tests offer a far wider range of flexibility concerning where power is to be concentrated, and can offer substantial increases in power against many ranges of the alternative set. While for comparability this paper has dealt only with the case in which initial observations are treated as fixed, it is straightforward to extend most of the results to other assumptions concerning the initial observation, for instance that of Elliot, Rothenberg, and Stock (1996), and one would expect increases in power similar to those they report for the univariate case when their assumptions are met by all groups within the panel .

7 Appendix-Proofs

Proof of Theorem 2: By Wald (1950) Theorem 18, the closure (in the regular sense) of set of all Bayes solution in the strict sense with respect to a discrete probability measure π over the parameter space Θ is essentially complete. In the current framework, Wald's definition of regular convergence corresponds to weak* convergence, therefore by Lemma 2 we must show that the set tests that are nonincreasing, quasiconvex functions of the sufficient statistics are closed in the sense of weak* convergence.

Let Φ be the set of nonempty, closed, convex subsets of R^{2N}. If V is a convex cone then let $\Phi(V)$ be the elements of Φ such that if $C \in \Phi$ and c_0 is a boundary point of C then $V \subseteq C - c_0$. A test τ is said to have an acceptance region C if $\tau = 0$ on the interior of C and $\tau = 1$ on the complement of C. Let $D(V)$ be the set of tests which have acceptance regions C such that $C \in \Phi(V)$, where $\tau \equiv 1$ is taken to be an element of $D(V)$. Consider a sequence of tests $\{\tau_i\}$, $\tau_i \in D(V)$. The set $D(V)$ meets the conditions under which Eaton (1970) and Matthes and Truax (1967) prove closure in the sense of weak* convergence, so that any limit point τ must itself be an element of $D(V)$. By Lemma 2 each Bayes solution in the strict sense with respect to discrete support has an acceptance set of the form $C = \{s \mid \tau(s) \leq \pi(\theta^0)\}$. The proof of Lemma 2 implies that $C \in \Phi(R_+^{2N})$, so that Bayes solutions in the strict sense with respect to a discrete support is an element of $D(R_+^{2N})$. Each element of $D(R_+^{2N})$ will be a nonincreasing and quasiconvex function of the sufficient statistics, and therefore since this set is closed in the sense of weak* convergence each limit point will have the same property. Absolute continuity of the probability measure $p(\cdot)$ allows us to restrict attention to nonrandomized tests, i.e. the set of $\tau \in D(R_+^{2N})$ for which there is a $C \in \Phi(R_+^{2N})$ and $\tau = 0$ on the interior of C, $\tau = 1$ everywhere else. QED.

Proof of Theorem 3: Since $p(y;\theta)$ is absolutely continuous with respect to Lebesgue measure, we can restrict ourselves to nonrandomized tests that are functions of the sufficient statistics. Without loss of generality let τ be a test with the form

$$\tau = \begin{cases} 0 & \text{if} \quad \tau(s) < c \\ 1 & \quad \text{otherwise} \end{cases}$$

for some critical value c. Assume that τ is *not* almost everywhere equivalent to a nonincreasing function of the sufficient statistics. Then by assumption there exist measurable sets $U \subset \{s \mid \tau(s) < c\}$ and $B \subset \{s' \mid s' \geq s \ \forall s \in U\}$ such that $B \cap \{s \mid \tau(s) < c\}$ is of measure zero. By the absolute continuity of p we can choose measurable subsets $U' \subset U$ and $B' \subset B$ such that $p(s \in U'; 0) = p(s \in B'; 0)$. But $p(s \in B'; \theta) \geq p(s \in U'; \theta)$ for all $\theta \in \Theta_1$, with strict inequality for at least one $\theta \in \Theta_1$, therefore the test with acceptance region $\{B' \cup \{s \mid \tau(s) < c, s \notin U'\}\}$ is uniformly better. QED.

Proof of Theorem 4: The conditions of Assumptions 1 and 2 are sufficient to ensure that if $\lim_{N\to\infty} N^{-1}\frac{\partial^2 \log(p(y;\bar{\theta}))}{(\partial\theta)^2}$ converges in probability to a constant $C > 0$ under the null then the log-likelihood will have the form

$$\log(p(y;\hat{\theta})) = \log(p(y;\bar{\theta})) + \frac{C}{2}\left(\hat{\theta} - \bar{\theta}\right)^2$$

for $\hat{\theta} - \bar{\theta} = O(N^{-\frac{1}{2}})$, where $\hat{\theta}$ is the maximum likelihood estimate of $\bar{\theta}$. Since this is a linear exponential family with univariate sufficient statistic $\hat{\theta}$, there is a uniformly most powerful test of hypotheses of the form $\bar{\theta} = 0$ versus alternatives $\bar{\theta} < 0$ within an $N^{-\frac{1}{2}}$ neighborhood of the null.

We need now only verify that $\lim_{N\to\infty} N^{-1}\frac{\partial^2 \log(p(y;\bar{\theta}))}{(\partial\theta)^2}$ converges in probability to some constant $C > 0$. For fixed N, Assumption 1 implies that $v_t = \sum_{k=0}^{\infty} \Psi_k \varepsilon_{t-k}$ is invertible and that there is a nonsingular matrix W such that $\Omega_\varepsilon = WW'$. Writing $\Psi(L) = \Psi_0 + \Psi_1 L + \Psi_2 L^2 + \cdots$ and imposing the constraint $\theta_n = \bar{\theta}$ for $n = 1 \ldots N$, equation (3) can be rewritten:

$$\triangle u_t = \bar{\theta} u_{t-1} + \Psi(L) e_t$$

Multiplying by $W^{-1}\Psi(L)^{-1}$ and defining $\omega_t = W^{-1}\Psi(L)^{-1} u_t$, we have

$$\triangle \omega_t = \bar{\theta}\omega_{t-1} + \varepsilon_t$$

where $\varepsilon_t = W^{-1} e_t$ is a sequence of independent $N(0, I)$ variables. The likelihood therefore takes the form

$$p(y;\bar{\theta}) = h(y)\exp\left(\bar{\theta}\left[\sum_{t=1}^{T} \triangle\omega_t'\omega_{t-1}\right] - \frac{1}{2}\bar{\theta}^2\left[\sum_{t=1}^{T}\omega_{t-1}'\omega_{t-1}\right]\right),$$

implying

$$N^{-1}\frac{\partial^2 \log(p(y;\bar{\theta}))}{(\partial\bar{\theta})^2} = N^{-1}\sum_{t=1}^{T}\omega'_{t-1}\omega_{t-1}$$

We can write $N^{-1}\sum_{t=1}^{T}\omega'_{t-1}\omega_{t-1} = \sum_{n=1}^{N}\chi_{Nn}$, where $\chi_{Nn} = N^{-1}\sum_{t=1}^{T}\omega^2_{nt-1}$ and $(\chi_{N1},\ldots\chi_{NN})$ is a set of N independent random variables. Well known results (see for instance Fuller (1976)) imply that under the null $E(\sum_{n=1}^{N}\chi_{Nn}) = \frac{T(T-1)}{2}$ and $V(\sum_{n=1}^{N}\chi_{Nn}) = \frac{T(T-1)(T^2-T+1)}{3N}$. Since $\lim_{N\to\infty}V(\sum_{n=1}^{N}\chi_{Nn})^{-\frac{1}{2}} = 0$ Theorem 6.2 of Billingsley (1986) establishes that $\sum_{n=1}^{N}\chi_{Nn}$ will converge in probability to $C = E(\sum_{n=1}^{N}\chi_{Nn}) > 0$. QED.

Proof of Theorem 5

Step (1): Deriving the conditional density.

If X is a square matrix, then let $vec(X)$ vectorize the elements of X, $vecl(X)$ vectorize the elements of X below the diagonal, and let $vecd(X)$ vectorize the diagonal of X. Thus if

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$$

then

$$\begin{aligned} vec(X) &= \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{21} & x_{22} & x_{23} & x_{31} & x_{32} & x_{33} \end{bmatrix}' \\ vecl(X) &= \begin{bmatrix} x_{21} & x_{31} & x_{32} \end{bmatrix}' \\ vecd(X) &= \begin{bmatrix} x_{11} & x_{22} & x_{33} \end{bmatrix}'. \end{aligned}$$

Let $\tilde{z}' \equiv \begin{bmatrix} \triangle\bar{z}' & \bar{z}'_- & z' \end{bmatrix}$ and $\tilde{\beta}' \equiv \begin{bmatrix} \bar{\beta}' & \theta\bar{\beta}' & \beta' \end{bmatrix}$. Define

$$\begin{aligned} s &= \begin{bmatrix} vecd(\triangle yy'_-) \\ vecd(y_-y'_-) \end{bmatrix} \\ \gamma(\theta,\xi) &= \begin{bmatrix} \frac{1}{2}vecd(\theta\Omega_e^{-1}) \\ -\frac{1}{2}vecd(\theta\Omega_e^{-1}\theta') \end{bmatrix} \\ r &= \begin{bmatrix} vech\left(\begin{bmatrix} \triangle y \\ \tilde{z} \end{bmatrix}\begin{bmatrix} \triangle y' & \tilde{z}' \end{bmatrix}\right) \\ vec(y'_-\tilde{z}) \\ vecl(\triangle yy'_-) \\ vecl(y_-y'_-) \end{bmatrix} \end{aligned}$$

$$\omega(\theta,\xi) = \begin{bmatrix} -\frac{1}{2}vech\left(\begin{bmatrix} I_N \\ -\tilde{\beta} \end{bmatrix}\Omega_e^{-1}\begin{bmatrix} I_N & -\tilde{\beta}' \end{bmatrix}\right) \\ -vec(\theta\Omega_e^{-1}\tilde{\beta}') \\ vecl(\theta\Omega_e^{-1}) \\ -vecl(\theta\Omega_e^{-1}\theta')' \end{bmatrix}$$

Then

$$\begin{aligned} p(y;\theta,\xi) &= c(\xi)\exp\left\{-\frac{1}{2}vech\left(Q\Omega_e^{-1}Q'\right)' vech\left(MM'\right)\right\} \\ &= c(\xi)\exp\left\{\gamma(\theta,\xi)'s + \omega(\theta,\xi)'r\right\} \end{aligned}$$

Since Ω_e is positive definite, the diagonal elements of Ω_e^{-1} must be strictly positive, and therefore $\gamma(\theta,\xi) \le 0$ for all $(\theta,\xi) \in \Theta \times \Xi$, $\gamma(\theta,\xi) = 0$ if and only if $\theta = 0$. Since the signs of the elements of $\tilde{\beta}$ and the off-diagonal elements of Ω_ε are unconstrained, the sign of any element of $\omega(\theta,\xi)$ is unrestricted without further assumptions.

Let $S \times R$ be the support of (s,r) and $p(s;r,\theta,\xi)$ denote the probability density of s conditional on r given parameter (θ,ξ). By Lemma 2.8 of Lehmann (1959) $p(s;r,\theta,\xi)$ can be represented in the form

$$p(s;r,\theta,\xi) = c_r(\theta,\xi)\exp\left\{\gamma(\theta,\xi)'s\right\} p(s;r,0,\xi_0)$$

where in this instance

$$c_r(\theta,\xi) = \left[\int_S \exp\left\{\gamma(\theta,\xi)'s\right\} p(s;r,0,\xi_0)ds\right]^{-1}$$

and straightforward calculations shows that $c_r(\theta,\xi) > 0$, $c_r(0,\xi) = 1$.

Step (2): Establishing a counterpart to Lemma 1.

Suppose that the conditional density of s given $r \in B \subset R$ has the form

$$\mu(s;r \in B,\theta,\xi) = \tilde{c}(B,\theta,\xi)\exp\left\{\gamma(\theta,\xi)'s\right\}\mu(s;r \in B,0,\xi_0)$$

where $\tilde{c}(B,\theta,\xi) > 0$ and the measure $\mu(s;r \in B,0,\xi_0)$ is absolutely continuous. (Note that $p(s;r,\theta,\xi)$ has this form.)

Pick $J = J_0 + J_1$ discrete elements of $\Theta \times \Xi$ and let π be a probability measure over these points, $\pi(\theta^j,\xi^j)$ representing the probability of $(\theta^j,\xi^j) \in \Theta \times \Xi$, $j = 0 \ldots J$, where $\theta^j = 0$ for $1 \le j \le J_0$ and $\theta^j \in \Theta_1$ for $J_0 < j \le J$. By the Neyman-Pearson lemma and the absolute continuity of μ, τ is a Bayes solution relative to π if it is a.e. equivalent to

$$\tau = \begin{cases} 0 & \text{if } \sum_{j=1}^{J_0}\mu(s;r \in B,\theta^j,\xi^j)\pi(\theta^j,\xi^j) > \sum_{j=J_0+1}^{J}\mu(s;r \in B,\theta^j,\xi^j)\pi(\theta^j,\xi^j) \\ 1 & \text{otherwise} \end{cases}$$

Define $\lambda(s;B,\theta,\xi)=\pi(\theta,\xi)\tilde{c}(B,\theta,\xi)\exp(\gamma(\theta,\xi)'s)>0$. From the expression for $\mu(s;r\in B,\theta,\xi)$, τ is equivalent to

$$\tau=\begin{cases}0 & \text{if} \quad \sum\limits_{j=J_0+1}^{J}\lambda(s;B,\theta^j,\xi^j)<c \\ 1 & \qquad\qquad \text{otherwise}\end{cases}$$

where $c=\sum\limits_{j=1}^{J_0}\tilde{c}(B,0,\xi^j)\pi(0,\xi^j)$ is a positive constant. Defining $\tau(s;B)=\sum\limits_{j=J_0+1}^{J}\lambda(s;B,\theta^j,\xi^j)$, we can see that τ is analytic and that

$$\begin{aligned}D_s\tau &= \sum_{j=J_0+1}^{J}\lambda(s;B,\theta^j,\xi^j)\gamma(\theta^j,\xi^j)'\leq 0\\ D_s^2\tau &= \sum_{j=J_0+1}^{J}\lambda(s;B,\theta^j,\xi^j)\gamma(\theta^j,\xi^j)\gamma(\theta^j,\xi^j)'\geq 0\end{aligned}$$

which implies that $\tau(s;B)$ is nonincreasing and convex function of the sufficient statistic s. Given density μ and any $B\subset R$ the arguments of Theorems 2 and 3 can be applied to the set of tests conditioned on the event $r\in B$.

Step (3): Establishing the measurability of the proposed test statistic.

This step follows an argument in Matthes and Truax (1966). Partition R into a countable number of Borel sets, let B_n be the σ-field generated by this partition, and let F_n be the associated σ-field in $S\times R$. Take a sequence of such partitions and σ-fields, each a refinement of the previous one, such that the smallest σ-field B_∞ (F_∞) containing them is the class of all Borel sets in R ($S\times R$). In terms of F_n define (up to a constant of integration) the conditional density

$$\mu_n(s;r,\theta,\xi)\equiv\exp(\gamma(\theta,\xi)'s)p(s;r\in B,0,\xi_0)$$

if r is in an atom B of B_n. Let $\tau(s,r)$ be any test of the null hypothesis $\{(\theta,\xi)\in\Theta\times\Xi\mid\theta=0\}$. Define the sequence of tests

$$\begin{aligned}\tau_n(s,r) &= \int_{r\in B}\tau(s,r)p(r;0,\xi_0)dr\\ &= E\left[\tau(s,r)\mid F_n,s\right]\end{aligned}$$

if r is an atom B of B_n.

For each atom B of B_n we can use step 2 to prove the existence of a non-randomized test $\tau_n'(s,B)$ that is a non-increasing and quasiconvex function of s such that τ_n' is uniformly as good as τ_n and if τ_n is increasing then τ_n'

is uniformly better with respect to μ_n. Since there are a denumerable set of atoms in B_n the resulting function $\tau_n'(s,r)=\tau_n'(s,B)$ if $r\in B$ is a test. Since τ_n' can be chosen to have the same size as τ, we have just established that

$$\int_S [\tau_n'(s,r)-\tau_n(s,r)]\exp(\gamma(\theta,\xi)'s)\mu_n(s;r,0,\xi_0)ds\geq 0 \tag{14}$$

with equality for all $\theta\in\Theta_0$ and strict inequality for at least one $\theta\in\Theta_1$ if τ is increasing in s conditional on r. The argument of Matthes and Truax's Theorem 3.1 establishes that

$$\int_S \tau_n'(s,r)\exp(\gamma(\theta,\xi)'s)\mu_n(s;r,0,\xi_0)ds\rightarrow\int_S \tau'(s,r)\exp(\gamma(\theta,\xi)'s)p(s;r,0,\xi_0)ds$$

Application of the martingale convergence theorem implies that

$$\int_S \tau_n(s,r)\exp(\gamma(\theta,\xi)'s)\mu_n(s;r,0,\xi_0)ds\rightarrow\int_S \tau(s,r)\exp(\gamma(\theta,\xi)'s)p(s;r,0,\xi_0)ds$$

Theorem 4.2 of Rao (1962) implies that if $\tau_n'\rightarrow\tau'$ then τ' must itself be nonincreasing and quasiconvex in s for given r. Taking the limits of both sides of (14) and multiplying both sides by the positive constant $c_r(\theta,\xi)$ establishes for any test τ the existence of a test τ' which is a nonincreasing and quasiconvex function of s conditional on r such that

$$\int_S \tau'(s,r)p(s;r,\theta,\xi)ds\geq\int_S \tau(s,r)p(s;r,\theta,\xi)ds \tag{15}$$

with equality for all $(\theta,\xi)\in 0\times\Xi$ and strict inequality for at least one $(\theta,\xi)\in(\Theta\setminus 0)\times\Xi$ if τ is increasing in s conditional on r. Since (15) represents the conditional power of the two tests, the claim of the theorem is established. QED.

Prelude to the proof of Theorem 7: Stacking across $t=1\dots T$ for some group n gives

$$\triangle y_n=\bar{\theta}y_{n-}+\left(\triangle\bar{z}_n-\bar{\theta}\bar{z}_{n-}\right)\xi_{n0}+z_n\beta_n+e_n$$

where under the conditions of Theorem 6 $\bar{z}_n=[\bar{z}_{nt}]_{t=1\dots T}$, $z_n=[\triangle y_{nt-1}\cdots\triangle y_{nt-P}]_{t=1\dots T}$, and $e_n=[e_{nt}]_{t=1\dots T}$ are $T\times D$, $T\times P$ and $T\times 1$ matrices respectively; $\bar{z}_{nt}\in\{\{0\},\{1\},\{1,t\}\}$; $e_{nt}\sim N\left(0,\sigma_n^2\right)$ is independent across n and t; and $\beta_n=[\xi_{n1}\cdots\xi_{nP}]'$ is a $P\times 1$ vector.

Define the $T\times T$ matrices $\bar{Q}_n=\bar{z}_n(\bar{z}_n'\bar{z}_n)^{-}\bar{z}_n'$, $\bar{M}_n=I-\bar{Q}_n$, and $M_n=\bar{M}_n-\bar{M}_nz_n\left(z_n'\bar{M}_nz_n\right)^{-}z_n'\bar{M}_n$. All three matrices are symmetric, idempotent, positive semidefinite. Since $\triangle\bar{z}_n$ and $\bar{z}_{n-}$ both lie in the space spanned by $\bar{z}_n$ we have

$$\bar{M}_n \triangle y_n = \bar{\theta}\bar{M}_n y_{n-} + \bar{M}_n z_n \beta_n + \bar{M}_n e_n$$

The proof will frequently employ the Beveridge-Nelson decomposition under the null hypothesis

$$\sigma_n^{-1}\bar{M}_n y_{n-} = \psi_n(1)\bar{M}_n H\varepsilon_n + \bar{M}_n u_{n-} \tag{16}$$

where $\psi_n(1) = \sum_{j=0}^{\infty}\psi_{nj}$, $\varepsilon_n = [\varepsilon_{nt}]_{t=1\ldots T}$, the ε_{nt} are i.i.d. standard normal variables, $u_{n-} = [\eta_{nt-1} - \eta_{n0}]_{t=1\ldots T}$, $\eta_t = \sum_{j=0}^{\infty}\alpha_{nj}\varepsilon_{t-j}$, $\sum_{j=0}^{\infty} \mid \alpha_{nj} \mid = \sum_{j=0}^{\infty} j \mid \psi_{nj} \mid$, and

$$H = \begin{bmatrix} 0 & 1 & 1 & \cdots & 1 \\ 0 & 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

The proof of Theorem 7 will employ the following lemmas

Lemma 8 *Let $\varepsilon = [\varepsilon_1 \ldots \varepsilon_k]'$ be a vector of k i.i.d. N(0,1) variables. If A is a $k \times k$ nonstochastic, real symmetric and positive semidefinite matrix then $E(\varepsilon' A\varepsilon)^q \leq \frac{(2q!)}{2^q(q!)}[tr(A)]^q$.*

Proof: Since A is real symmetric, $A = SDS'$ where S and D are $k \times k$ matrices $S'S = I$, and D is a diagonal matrix whose diagonal elements are the eigenvalues of A. If $\varepsilon \sim N(0, I_k)$ then $S'\varepsilon \sim N(0, I_k)$ since S is orthogonal. Therefore

$$\varepsilon' A\varepsilon = \sum_{i=1}^{r}\delta_i\epsilon_i^2$$

where the ϵ_i are i.i.d. N(0,1) variables, the δ_i are the nonzero eigenvalues of A, and $r = rank(A)$. By independence and the formula for the moments of a standard normal variable

$$\begin{aligned} E\epsilon_1^{2q_1}\epsilon_2^{2q_2}\cdots\epsilon_r^{2q_r} &= \frac{(2q_1!)}{2^{q_1}(q_1!)}\frac{(2q_2!)}{2^{q_2}(q_2!)}\cdots\frac{(2q_r!)}{2^{q_r}(q_r!)} \\ &\leq \frac{(2q!)}{2^q(q!)} \end{aligned}$$

if $q = q_1 + q_2 + \cdots + q_r$. If $\delta_i \geq 0$ for $i = 1 \ldots r$ then by expanding terms in the polynomial $(\varepsilon' A \varepsilon)^q$ this implies

$$E\left(\varepsilon' A \varepsilon\right)^q \leq \frac{(2q!)}{2^q (q!)} \left(\sum_{i=1}^{r} \delta_i\right)^q$$

since $tr(A) = \sum_{i=1}^{r} \delta_i$ this completes the proof. QED.

Lemma 9 *The following moment bounds hold under the null hypothesis, where* $T = 1 \ldots \infty$, $0 \leq q < \infty$, $0 \leq v < \infty$*:*

(a) $E\left(T^{-2}\psi_n(1)^2 \varepsilon_n' H' \bar{M}_n H \varepsilon_n\right)^q \leq \frac{(2q!)}{2^q(q!)} \left[\frac{1}{2}(1 - \frac{1}{T})\right]^q \psi_n(1)^{2q}$.

(b) $E\left(T^{-2} u_{n-}' \bar{M}_n u_{n-}\right)^q \leq \frac{(2q!)}{2^q(q!)} \left(\frac{4}{T}\right)^q \left[\sum_{j=0}^{\infty} \alpha_{nj}^2\right]^q$.

(c) If i_T is a $T \times 1$ column vector of ones, then
$E \mid T^{-2}\psi_n(1) \varepsilon_n' H' \bar{M}_n u_{n-} \mid^q \leq \frac{(2q!)}{2^q(q!)} \left[\frac{1}{2}(1 - \frac{1}{T})\right]^{\frac{q}{2}} \left(\frac{4}{T}\right)^{\frac{q}{2}} \left[\sum_{j=0}^{\infty} \alpha_{nj}^2\right]^{\frac{q}{2}} \psi_n(1)^q$

(d) $E\left(T^{-1}\sigma_n^{-2} \bigtriangleup y_n' M_n \bigtriangleup y_n\right)^{-q} \leq \frac{T^q}{\prod_{i=1}^{q}(T-P-2-2i)}$ for $T - P - 2 > 2q$.

(e) If $Q_n = \bar{M}_n z_n \left(z_n' \bar{M}_n z_n\right)^{-} z_n' \bar{M}_n$, then

$$E \mid T^{-1}\psi_n(1) \varepsilon_n' H' Q_n \varepsilon_n \mid^q \leq c_{e1}(T, q) \psi_n(1)^q$$
$$E \mid T^{-1} u_{n-}' Q_n \varepsilon_n \mid^q \leq c_{e2}(T, q) \left[\sum_{j=0}^{\infty} \alpha_{nj}^2\right]^{\frac{q}{2}}$$

where $c_{e1}(T, q)$ is $O(1)$ and $c_{e2}(T, q)$ is $O(T^{-\frac{q}{2}})$.

(f) If $P = 1$ then

$$E(T^{-1}\sigma_n^{-2} z_n' \bar{M}_n z_n)^{-q} \leq \frac{(4T)^q}{\prod_{i=1}^{q}(T - D - 2i)} \text{ for } T - D > 2q$$
$$E\left(T^{-2}\sigma_n^{-2} z_n' \bar{M}_n H H' \bar{M}_n z_n\right)^q \leq \frac{2^q(2q!)}{(q!)} \psi_n(1)^{2q} \gamma_0$$

where $\gamma_{n0} \equiv \lim_{T \to \infty} Var(\triangle y_{nt})$.

Proof of (a): Since

$$\varepsilon_n' H' H \varepsilon_n = \varepsilon_n' H' \bar{Q}_n H \varepsilon_n + \varepsilon_n' H' \bar{M}_n H \varepsilon_n$$

the following bound

$$0 \leq \left(\varepsilon_n' H' \bar{M}_n H \varepsilon_n\right)^q \leq (\varepsilon_n' H' H \varepsilon_n)^q$$

holds for $q \geq 0$. $T^{-2}\varepsilon_n' H' H \varepsilon_n$ is a quadratic form in independent standard normal variables and $T^{-2}H'H$ is positive definite, $tr(T^{-2}H'H) = \frac{1}{2}(1 - \frac{1}{T})$, therefore by Lemma 8

$$E\left(T^{-2}\varepsilon_n' H' H \varepsilon_n\right)^q < \frac{(2q!)}{2^q(q!)}\left[\frac{1}{2}(1 - \frac{1}{T})\right]^q < \infty$$

for $T = 1 \ldots \infty$, $0 \leq q < \infty$. Since $\mid \psi_n(1) \mid < \infty$ we have the stated result. Note for later reference that the expression holds with equality for $q = 1$.

Proof of (b): This can be established using the same method as in (a) since $T^{-2}u_{n-}' u_{n-}$ can be written as a positive semidefinite form in i.i.d. N(0,1) variables, the sum of the eigenvalues of the quadratic form bounded by $\frac{4}{T}\sum_{j=0}^{\infty}\alpha_{nj}^2$, which is finite by the absolute summability of the α_{nj}.

Proof of (c): By Hölder's inequality

$$\begin{aligned} E \quad \mid \quad & T^{-2}\psi_n(1)\varepsilon_n' H' \bar{M}_n u_{n-} \mid^q \leq \\ & \left[E\left(T^{-2}\psi_n(1)^2\varepsilon_n' H' \bar{M}_n H \varepsilon_n\right)^q E\left(T^{-2}u_{n-}' \bar{M}_n u_{n-}\right)^q\right]^{\frac{1}{2}} \end{aligned}$$

and the bound follows from (a),(b).

Proof of (d): Under the null $\sigma_n^{-2} \triangle y_n' M_n \triangle y_n = \varepsilon' M_n \varepsilon$. Since M_n is idempotent

$$\varepsilon' M_n \varepsilon = \sum_{i=1}^{r}\epsilon_i^2$$

where $r = rank(M_n)$, $T \geq r \geq T - P - 2$, and the ϵ_i are independent $N(0,1)$ variables. Obviously

$$E\left(T^{-1}\sum_{i=1}^{r}\epsilon_i^2\right)^{-q} \leq E\left(T^{-1}\sum_{i=1}^{T-P-2}\epsilon_i^2\right)^{-q}$$

Using the density for a χ_k^2 variable

$$E\left(\sum_{i=1}^{k}\epsilon_i^2\right)^{-q} = \prod_{i=1}^{q}(k - 2i)^{-1}$$

for $k > 2q$. Combining gives the stated result.

Proof of (e): Since $\varepsilon_n' H' Q_n \varepsilon_n = \varepsilon_n' \left(\frac{H'Q_n + Q_n H}{2}\right) \varepsilon_n$ one can write

$$\varepsilon_n' H' Q_n \varepsilon_n = \sum_{i=1}^{r} \delta_i \epsilon_i^2$$

where ϵ_i are i.i.d. N(0,1) variables, the δ_i are the nonzero eigenvalues of $\left(\frac{H'Q_n + Q_n H}{2}\right)$ and r is its rank. If the δ_i are arranged in ascending order then by Fan and Hoffman (1955) $\mid \delta_i \mid \leq (\mu_i)^{\frac{1}{2}}$, where the μ_i are the eigenvalues of $(Q_n H' H Q_n)$, also placed in ascending order. By the Poincaré Separation Theorem there are P nonzero eigenvalues of $(Q_n H' H Q_n)$ and $\lambda_i \leq \mu_i \leq \lambda_{T-P+i}$, $i = 1 \ldots P$, where the λ_i are the eigenvalues of $H'H$. But the eigenvalues of $H'H$ are all positive and the largest eigenvalue is less than T^2. Therefore

$$\mid T^{-1} \varepsilon_n' H' Q_n \varepsilon_n \mid \leq \sum_{i=1}^{P} \epsilon_i^2$$

and the first bound follows from Lemma 8. The second bound can be established with the same method.

Proof of (f): Writing $\sigma_n^{-1} z_n = B_n \varepsilon_n$, where B_n is a $T \times T$ matrix,

$$B_n = \begin{bmatrix} 0 & 1 & \beta_n & \cdots & \beta_n^{T-2} \\ 0 & 0 & 1 & \cdots & \beta_n^{T-3} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

the first bound can be established as in Lemma 9(d) once it is noted that $B_n' \bar{M}_n B_n$ has $T-D$ nonzero eigenvalues, the smallest nonzero eigenvalue being bounded from below by $\frac{1}{4}$ for any value of β_n or T.

Writing

$$T^{-2} \sigma_n^{-2} z_n' \bar{M}_n H H' \bar{M}_n z_n = \varepsilon_n' \left(T^{-2} B_n' \bar{M}_n H H' \bar{M}_n B_n\right) \varepsilon_n.$$

By the Poincare Separation Theorem

$$tr(T^{-2} B_n' \bar{M}_n H H' \bar{M}_n B_n) \leq tr(T^{-2} B_n' H H' B_n)$$

where

$$tr(T^{-2} B_n' H H' B_n) \leq 4 \psi_n(1)^2 \gamma_{n0}.$$

Lemma 7 then establishes the second bound.

QED.

Lemma 10 *If* $\left(\sum_{j=0}^{\infty} j \mid \psi_{nj} \mid\right)$ *is finite for each* n *and*

$$(i) \quad N^{-1} \sum_{n=1}^{N} \left(\sum_{j=0}^{\infty} j \mid \psi_{nj} \mid \right)^2 < \infty$$

$$(ii) \quad N^{-2} \sum_{n=1}^{N} \left(\sum_{j=0}^{\infty} j \mid \psi_{nj} \mid \right)^4 \to 0$$

for $0 \leq k \leq 1$, *then*

$$\begin{aligned} p\lim(N^{-1}T^{-2} \sum_{n=1}^{N} \hat{\sigma}_n^{-2} y'_{n-} \bar{M}_n y_{n-}) &= p\lim(N^{-1}T^{-2} \sum_{n=1}^{N} \sigma_n^{-2} y'_{n-} \bar{M}_n y_{n-}) \\ &= \frac{1}{2} \lim_{N \to \infty} N^{-1} \sum_{n=1}^{N} \psi_n(1)^2. \end{aligned}$$

Proof: Using (16)

$$\begin{aligned} &\hat{\sigma}_n^{-2} y'_{n-} \bar{M}_n y_{n-} = \\ &\frac{\hat{\sigma}_n^{-2}}{\sigma_n^{-2}} \left[\psi_n(1)^2 \varepsilon'_n H' \bar{M}_n H \varepsilon_n + 2\psi_n(1) \varepsilon'_n H' \bar{M}_n u_{n-} + u'_{n-} \bar{M}_n u_{n-} \right]. \end{aligned}$$

Applying Billingsley's (1986; Theorem 6.2) weak law of large numbers for triangular arrays term by term.

$$\begin{aligned} &Var \left(N^{-1}T^{-2} \sum_{n=1}^{N} \frac{\hat{\sigma}_n^{-2}}{\sigma_n^{-2}} \psi_n(1)^2 \varepsilon'_n H' \bar{M}_n H \varepsilon_n \right) \\ \leq \quad & N^{-2} \sum_{n=1}^{N} E \left(T^{-2} \frac{\hat{\sigma}_n^{-2}}{\sigma_n^{-2}} \psi_n(1)^2 \varepsilon'_n H' \bar{M}_n H \varepsilon_n \right)^2 \\ = \quad & O(1) N^{-2} \sum_{n=1}^{N} \psi_n(1)^4 \end{aligned}$$

using independence across groups, Lemma 9(a), 8(d) and Hölder's inequality, which implies that

$$E \left(T^{-2} \frac{\hat{\sigma}_n^{-2}}{\sigma_n^{-2}} \psi_n(1)^2 \varepsilon'_n H' \bar{M}_n H \varepsilon_n \right)^2 \leq \left[E \left(\frac{\hat{\sigma}_n^{-2}}{\sigma_n^{-2}} \right)^4 E \left(T^{-2} \psi_n(1)^2 \varepsilon'_n H' \bar{M}_n H \varepsilon_n \right)^4 \right]^{\frac{1}{2}}.$$

Using the fact that $\sum_{j=0}^{\infty}\alpha_{nj}^2 \leq \sum_{j=0}^{\infty} j \mid \psi_{nj} \mid$ (obviously $\psi_n(1) \leq \sum_{j=0}^{\infty} j \mid \psi_{nj} \mid$), the same steps imply

$$Var\left(N^{-1}T^{-2}\sum_{n=1}^{N}\frac{\hat{\sigma}_n^{-2}}{\sigma_n^{-2}}\psi_n(1)\varepsilon_n' H' \bar{M}_n u_{n-}\right) \leq O(T^{-1})N^{-2}\sum_{n=1}^{N}\left[\sum_{j=0}^{\infty} j \mid \psi_{nj} \mid\right]^3$$

$$Var\left(N^{-1}T^{-2}\sum_{n=1}^{N}\frac{\hat{\sigma}_n^{-2}}{\sigma_n^{-2}}\psi_n(1)u_{n-}' \bar{M}_n u_{n-}\right) \leq O(T^{-2})N^{-2}\sum_{n=1}^{N}\left[\sum_{j=0}^{\infty} j \mid \psi_{nj} \mid\right]^2$$

From these expressions it is immediate that for each term $Var(\cdot) \to 0$ as $N, T \to \infty$ under the conditions of the lemma.

By Lemma 9(a)-(d)

$$E\left(N^{-1}T^{-2}\sum_{n=1}^{N}\frac{\hat{\sigma}_n^{-2}}{\sigma_n^{-2}}\psi_n(1)\varepsilon_n' H' \bar{M}_n u_{n-}\right) \leq O(T^{-\frac{1}{2}})N^{-1}\sum_{n=1}^{N}\left[\sum_{j=0}^{\infty} j \mid \psi_{nj} \mid\right]^{\frac{3}{2}}$$

$$E\left(N^{-1}T^{-2}\sum_{n=1}^{N}\frac{\hat{\sigma}_n^{-2}}{\sigma_n^{-2}}\psi_n(1)u_{n-}' \bar{M}_n u_{n-}\right) \leq O(T^{-1})N^{-1}\sum_{n=1}^{N}\left[\sum_{j=0}^{\infty} j \mid \psi_{nj} \mid\right]$$

From these expressions it is immediate that for each term $E(\cdot) \to 0$ as $N, T \to \infty$ under the conditions of the lemma. By Billingsley's weak law of large numbers this implies that

$$p\lim(N^{-1}T^{-2}\sum_{n=1}^{N}\hat{\sigma}_n^{-2}y_{n-}'\bar{M}_n y_{n-}) =$$
$$\lim_{N,T\to\infty} E\left(N^{-1}T^{-2}\sum_{n=1}^{N}\frac{\hat{\sigma}_n^{-2}}{\sigma_n^{-2}}\psi_n(1)^2\varepsilon_n' H' \bar{M}_n H\varepsilon_n\right)$$

Since it has been established that $\lim_{T\to\infty} E\left(T^{-2}\frac{\hat{\sigma}_n^{-2}}{\sigma_n^{-2}}\psi_n(1)^2\varepsilon_n' H' \bar{M}_n H\varepsilon_n\right)^2$ is finite then by Theorem 4.52. of Chung (1974)

$$\lim_{T\to\infty} E\left(T^{-2}\frac{\hat{\sigma}_n^{-2}}{\sigma_n^{-2}}\psi_n(1)^2\varepsilon_n' H' \bar{M}_n H\varepsilon_n\right) = E\left(\lim_{T\to\infty} T^{-2}\frac{\hat{\sigma}_n^{-2}}{\sigma_n^{-2}}\psi_n(1)^2\varepsilon_n' H' \bar{M}_n H\varepsilon_n\right)$$

and this establishes that

$$\lim_{N,T\to\infty} E\left(N^{-1}T^{-2}\sum_{n=1}^{N}\sigma_n^{-2}y_{n-}'\bar{M}_n y_{n-}\right) = \frac{1}{2}\lim_{N\to\infty} N^{-1}\sum_{n=1}^{N}\psi_n(1)^2.$$

The same steps obviously imply

$$p\lim(N^{-1}T^{-2}\sum_{n=1}^{N}\sigma_n^{-2}y_{n-}'\bar{M}_n y_{n-}) = \frac{1}{2}\lim_{N\to\infty} N^{-1}\sum_{n=1}^{N}\psi_n(1)^2.$$

QED.

Lemma 11 *If in addition to the conditions of Lemma 10, $\frac{N}{T} \to 0$ then*

$$p\lim(N^{-\frac{1}{2}}T^{-1}\sum_{n=1}^{N}\left(\hat{\sigma}_n^{-2}-\sigma_n^{-2}\right)y'_{n-}\bar{M}_n e_n)=0$$

Proof: Using (16)

$$\left(\hat{\sigma}_n^{-2}-\sigma_n^{-2}\right)y'_{n-}\bar{M}_n e_n=\left(\frac{\hat{\sigma}_n^{-2}}{\sigma_n^{-2}}-1\right)\left(\psi_n(1)\varepsilon'_n H'\bar{M}_n\varepsilon_n+u'_{n-}\bar{M}_n\varepsilon_n\right)$$

where i_T is a $T\times 1$ column vector of ones. Expanding $E\left(\frac{\hat{\sigma}_n^{-2}}{\sigma_n^{-2}}-1\right)^q$ and using the proof of Lemma 9(d), it is straightforward to show

$$E\left(\frac{\hat{\sigma}_n^{-2}}{\sigma_n^{-2}}-1\right)^q\le\max_{0\le k\le P+2}E\left[\left(\frac{\hat{\sigma}_n^{-2}}{\sigma_n^{-2}}-1\right)^q\mid rank(z_n)=k\right]\equiv\gamma(T,q)<\infty$$

for $T-P-2>2q$ and that $\gamma(T,q)=O(T^{-1})$. It is also easily shown that

$$\begin{aligned}E\left(T^{-1}\psi_n(1)\varepsilon'_n H'\bar{M}_n\varepsilon_n\right)^{2q}&\le c_{21}(T,q)\left(\sum_{j=0}^{\infty}j\mid\psi_{nj}\mid\right)^{2q}\\E\left(T^{-1}u'_{n-}\bar{M}_n\varepsilon_n\right)^{2q}&\le c_{22}(T,q)\left(\sum_{j=0}^{\infty}j\mid\psi_{nj}\mid\right)^{q}\end{aligned}$$

where $c_{21}(T,q)$, and $c_{22}(T,q)$ are bounded.

By independence across groups and Hölder's inequality

$$\begin{aligned}&Var\left(N^{-\frac{1}{2}}T^{-1}\sum_{n=1}^{N}\left(\frac{\hat{\sigma}_n^{-2}}{\sigma_n^{-2}}-1\right)\psi_n(1)\varepsilon'_n H'\bar{M}_n\varepsilon_n\right)\le\\&N^{-1}\sum_{n=1}^{N}\left[E\left(\frac{\hat{\sigma}_n^{-2}}{\sigma_n^{-2}}-1\right)^4\right]^{\frac{1}{2}}\left[E\left(T^{-1}\psi_n(1)\varepsilon'_n H'\bar{M}_n\varepsilon_n\right)^4\right]^{\frac{1}{2}}\\=\;&O(T^{-\frac{1}{2}})N^{-1}\sum_{n=1}^{N}\left(\sum_{j=0}^{\infty}j\mid\psi_{nj}\mid\right)^2\end{aligned}$$

In the same way

$$Var\left(N^{-\frac{1}{2}}T^{-1}\sum_{n=1}^{N}\left(\frac{\hat{\sigma}_n^{-2}}{\sigma_n^{-2}}-1\right)u'_{n-}\bar{M}_n e_n\right)\le O(T^{-\frac{1}{2}})N^{-1}\sum_{n=1}^{N}\left(\sum_{j=0}^{\infty}j\mid\psi_{nj}\mid\right)$$

For both terms under the stated conditions $Var(\cdot) \to 0$ for $N, T \to \infty$.

In the same way

$$E \mid N^{-\frac{1}{2}}T^{-1}\sum_{n=1}^{N}\left(\frac{\hat{\sigma}_n^{-2}}{\sigma_n^{-2}}-1\right)\psi_n(1)\varepsilon_n' H'\bar{M}_n\varepsilon_n \mid \leq O(T^{-\frac{1}{2}})N^{-\frac{1}{2}}\sum_{n=1}^{N}\left(\sum_{j=0}^{\infty} j \mid \psi_{nj} \mid\right)$$

$$= \left(\frac{N}{T}\right)^{\frac{1}{2}}\left\{O(1)N^{-1}\sum_{n=1}^{N}\left(\sum_{j=0}^{\infty} j \mid \psi_{nj} \mid\right)\right\}$$

$$E \mid N^{-\frac{1}{2}}T^{-1}\sum_{n=1}^{N}\left(\frac{\hat{\sigma}_n^{-2}}{\sigma_n^{-2}}-1\right)u_{n-}'\bar{M}_n e_n \mid \leq \left(\frac{N}{T}\right)^{\frac{1}{2}}\left\{O(1)N^{-1}\sum_{n=1}^{N}\left(\sum_{j=0}^{\infty} j \mid \psi_{nj} \mid\right)^{\frac{1}{2}}\right\}$$

The expression in brackets are finite for all N, T establishing that

$$E\left(N^{-\frac{1}{2}}T^{-1}\sum_{n=1}^{N}\left(\hat{\sigma}_n^{-2}-\sigma_n^{-2}\right)y_{n-}'\bar{M}_n e_n\right) \to 0$$

if $\frac{N}{T} \to 0$. QED

Proof of Theorem 7:

Step 1: Known nuisance parameters.

Consistent with the treatment of y_t as fixed for $t < 1$, define the maximal invariant (for large T) with respect to location, $\bar{M}_n y_n$. If β_n and σ_n ($n = 1 \ldots N$) are known, the log-likelihood for $\bar{\theta}$ is (ignoring constants)

$$\begin{aligned} \ell(\bar{\theta}) &= -\frac{1}{2}\sum_{n=1}^{N}\left(\bar{M}_n e_n\right)'\left(\sigma_n^2 \bar{M}_n \bar{M}_n'\right)^{-}\left(\bar{M}_n e_n\right) \\ &= -\frac{1}{2}\sum_{n=1}^{N}\sigma_n^{-2}\left(\triangle y_n - \bar{\theta} y_{n-} - z_n\beta_n\right)'\bar{M}_n\left(\triangle y_n - \bar{\theta} y_{n-} - z_n\beta_n\right) \end{aligned}$$

and the maximum likelihood estimate is

$$\hat{\theta}_0 = \frac{\sum_{n=1}^{N}\sigma_n^{-2}\left(\triangle y_n - z_n\beta_n\right)'\bar{M}_n y_{n-}}{\sum_{n=1}^{N}\sigma_n^{-2}y_{n-}'\bar{M}_n y_{n-}}$$

therefore

$$\ell(\hat{\theta}_0) = \ell(0) + \frac{1}{2}\hat{\theta}_0^2\left(\sum_{n=1}^{N}\sigma_n^{-2}y_{n-}'\bar{M}_n y_{n-}\right)$$

The proof of Theorem 4 demonstrates that the relevant neighborhood of the null hypothesis $\bar{\theta} = 0$ is $O(N^{-\frac{1}{2}}T^{-1})$. Writing

$$\ell(N^{\frac{1}{2}}T\hat{\theta}_0) = \ell(0) + \frac{1}{2}\left(N^{\frac{1}{2}}T\hat{\theta}_0\right)^2\left(N^{-1}T^{-2}\sum_{n=1}^{N}\sigma_n^{-2}y_{n-}'\bar{M}_n y_{n-}\right)$$

The claim will be established as in Theorem 3 for β_n and σ_n known if

$$p\lim(N^{-1}T^{-2}\sum_{n=1}^{N}\sigma_n^{-2}y'_{n-}\bar{M}_n y_{n-}) = C$$

where C is some positive, finite constant, but this is established in Lemma 10.

Step 2: Unknown nuisance parameters.

If β_n and σ_n $(n = 1 \ldots N)$ are unknown then the theorem will be established if there is an invariant statistic $\hat{\theta}_e$ such that

$$N^{\frac{1}{2}}T\left(\hat{\theta}_e - \hat{\theta}_0\right) = o_p(1) \tag{17}$$

Define

$$\hat{\theta}_e = \left(\sum_{n=1}^{N}\hat{\sigma}_n^{-2}y'_{n-}\bar{M}_n y_{n-}\right)^{-1}\left(\sum_{n=1}^{N}\hat{\sigma}_n^{-2}y'_{n-}M_n \triangle y_n\right)$$

where

$$\hat{\sigma}_n^2 = T^{-1}\triangle y'_n M_n \triangle y_n.$$

The statistic $\hat{\theta}_e$ is invariant to $\bar{\beta}$ since it can be written in terms of the invariant statistic $\bar{M}_n y_n$. If $N^{-1}\sum_{n=1}^{N}\left(\sum_{j=0}^{\infty} j \mid \psi_{nj} \mid\right)^2 < \infty$, $N^{-2}\sum_{n=1}^{N}\left(\sum_{j=0}^{\infty} j \mid \psi_{nj} \mid\right)^4 \to 0$ then Lemma 10 establishes that

$$p\lim\left(N^{-1}T^{-2}\sum_{n=1}^{N}\hat{\sigma}_n^{-2}y'_{n-}\bar{M}_n y_{n-}\right) = p\lim\left(N^{-1}T^{-2}\sum_{n=1}^{N}\sigma_n^{-2}y'_{n-}\bar{M}_n y_{n-}\right).$$

Under the null

$$\sum_{n=1}^{N}\hat{\sigma}_n^{-2}y'_{n-}M_n \triangle y_n = \sum_{n=1}^{N}\hat{\sigma}_n^{-2}y'_{n-}\bar{M}_n e_n - \sum_{n=1}^{N}\hat{\sigma}_n^{-2}y'_{n-}Q_n e_n$$

If in addition to the conditions of Lemma 10, $\frac{N}{T} \to 0$ then Lemma 11 establishes that

$$p\lim\left(N^{-\frac{1}{2}}T^{-1}\sum_{n=1}^{N}\hat{\sigma}_n^{-2}y'_{n-}\bar{M}_n e_n\right) = p\lim\left(N^{-\frac{1}{2}}T^{-1}\sum_{n=1}^{N}\sigma_n^{-2}y'_{n-}\bar{M}_n e_n\right).$$

and a proof similar to the one found in Lemma 9 can easily establish

$$p\lim\left(N^{-\frac{1}{2}}T^{-1}\sum_{n=1}^{N}\left(\hat{\sigma}_n^{-2} - \sigma_n^{-2}\right)y'_{n-}Q_n e_n\right) = 0$$

The proof is therefore concluded if $N^{-\frac{1}{2}}T^{-1}\sum_{n=1}^{N}\sigma_n^{-2}y_{n-}'Q_n e_n$ converges weakly to zero. Using (16)

$$\sigma_n^{-2}y_{n-}'Q_n e_n = \psi_n(1)\varepsilon_n' H' Q_n \varepsilon_n + u_{n-}'Q_n\varepsilon_n,$$

Lemma 9(e) establishes that

$$\begin{aligned} Var(N^{-\frac{1}{2}}T^{-1}\sum_{n=1}^{N}u_{n-}'Q_n\varepsilon_n) &\leq O(T^{-1})N^{-1}\sum_{n=1}^{N}\left(\sum_{j=0}^{\infty}j \mid \psi_{nj}\mid\right) \\ E \mid N^{-\frac{1}{2}}T^{-1}\sum_{n=1}^{N}u_{n-}'Q_n\varepsilon_n) \mid &\leq \left(\frac{N}{T}\right)^{\frac{1}{2}}\left\{O(1)N^{-1}\sum_{n=1}^{N}\left(\sum_{j=0}^{\infty}j \mid \psi_{nj}\mid\right)^{\frac{1}{2}}\right\} \end{aligned}$$

therefore under the conditions of Lemma 11 $p\lim(u_{n-}'Q_n\varepsilon_n)=0$. Lemma 9(e) is not strong enough to establish $p\lim(T^{-1}\psi_n(1)\varepsilon_n'H'Q_n\varepsilon_n)=0$, however this can be established for $P=1$.

If $P=1$ then $T^{-\frac{1}{2}}\varepsilon_n'H'Q_n\varepsilon_n = \delta\epsilon^2$ where δ is the single nonzero eigenvalue of $\left(T^{-\frac{1}{2}}\frac{H'Q_n+Q_nH}{2}\right)$. Using the Poincare Separation Theorem $\mid\delta\mid\leq \mu(T^{-1}H'\bar{M}_n z_n(z_n'\bar{M}_n z_n)^{-}z_n'\bar{M}_n H)$, where $\mu(A)$ represents the maximum eigenvalue of matrix A. But if $rank(z_n)=1$ then

$$\mu(T^{-1}H'\bar{M}_n z_n(z_n'\bar{M}_n z_n)^{-}z_n'\bar{M}_n H) = (T^{-1}\sigma_n^{-2}z_n'\bar{M}_n z_n)^{-}(T^{-2}\sigma_n^{-2}z_n'\bar{M}_n HH'\bar{M}_n z_n).$$

This expression in conjunction with Lemma 9(f) and Hölder's inequality implies

$$E \mid T^{-\frac{1}{2}}\psi_n(1)\varepsilon_n'H'Q_n\varepsilon_n\mid^q \leq c(T,q)\psi_n(1)^{2q}\gamma_{n0}^{\frac{q}{2}}$$

where $c(T,q)=O(1)$ for $q<\infty$. Therefore

$$\begin{aligned} Var(N^{-\frac{1}{2}}T^{-1}\sum_{n=1}^{N}\psi_n(1)\varepsilon_n'H'Q_n\varepsilon_n) &\leq \frac{N}{T}\left\{O(1)N^{-2}\sum_{n=1}^{N}\psi_n(1)^4\gamma_{n0}\right\} \\ E \mid N^{-\frac{1}{2}}T^{-1}\sum_{n=1}^{N}\psi_n(1)\varepsilon_n'H'Q_n\varepsilon_n) \mid &\leq \left(\frac{N}{T}\right)^{\frac{1}{2}}\left\{O(1)N^{-1}\sum_{n=1}^{N}\psi_n(1)^2\gamma_{n0}^{\frac{1}{2}}\right\} \end{aligned}$$

and $p\lim(T^{-1}\psi_n(1)\varepsilon_n'H'Q_n\varepsilon_n)=0$ if

$$N^{-2}\sum_{n=1}^{N}\psi_n(1)^4\gamma_{n0}<\infty,\ N^{-1}\sum_{n=1}^{N}\psi_n(1)^2\gamma_{n0}^{\frac{1}{2}}<\infty$$

and $\frac{N}{T}\to 0$. QED.

8 Appendix-Details of Monte Carlo Experiments

All reported Monte Carlo exercises were based on twenty thousand independent replications for each point of the power functions. All data were generated from AR(1) models with standard normal disturbances, where initial values y_0 (and y_{-1} for Tables 1-3, where higher order AR models were estimated) were fixed at 0 and no deterministic components were present (although deterministic components were estimated in Tables 2 and 3) . For Figure 6 random normal disturbances were generated using the GAUSS "rndn" function and adjusted to have covariance matrix $\Omega = I_N + i_N i_N'$. In all other exercises i.i.d disturbances were generated using the MATLAB "randn" routine. All power calculations used size-adjusted critical values so that reported powers are for tests with exact 5% size.

The parameters of Bayes tests that are not fully described in the text are:

Figure 5, top left panel:$\theta = (0, -.015)$, $\pi(\theta) = (0, 1)$.

Figure 5, top right panel:$\theta = (0, -.015 - .24)$, $\pi(\theta) = (.2, .78, .02)$.

Figure 5, middle left panel: $\theta = (0, -.23)$, $\pi(\theta) = (1, 0)$.

Figure 5, middle right panel:

$\theta = (0, -.075 - .15 - .24)$, $\pi(\theta) = (.82, .07, .04, .07)$.

Figure 5, bottom panel: $\theta = (0, -.015 - .075 - .18)$, $\pi(\theta) = (.7, .28, .016, .004)$.

Table 1, Bayes1: Same as Figure 5, bottom panel

Table 1, Bayes2:

$\theta = (0, -.015, -.3, -.45, -.6, -.75)$, $\pi(\theta) = (.31, .11, .15, .08, .33, .02)$.

(This is similar in power to the test in Figure 5, middle right panel, however it does a bit less well against heterogeneous alternatives for small N, and better for homogeneous alternatives.)

Because the inclusion of a constant or a constant and time trend affects the overall shape of the power envelope, Bayes tests need to be adjusted to place weight on alternatives further away from the null to keep good power. The tests used in Tables 2 and 3 were:

Table 2, Bayes1:

$\theta = (0, -.13, -.25, -.5, -.55, -.65)$, $\pi(\theta) = (.19, .15, .16, 0.21, .14, .15)$.

Table 2, Bayes2:

$\theta = (0, -.2, -.45, -.85, -.95)$, $\pi(\theta) = (.21, .01, 0.09, 0.17, 0.52)$.

Table 3, Bayes1:

$\theta = (0, -.125, -.35, -.45, -.5, -.85)$, $\pi(\theta) = (.04, .06, .07, .21, .35, .27)$.

Table 3, Bayes2:

$\theta = (0, -.035, -.5, -.6, -.9, -1)$, $\pi(\theta) = (.125, .05, .165, .2, .29, .17)$.

References

Abuaf, Niso and Phillippe Jorion (1990): "Purchasing Power Parity in the Long Run," *Journal of Finance*, 45, 157-174.

Bernard, A. and C. Jones (1996): "Productivity Across Industries and Countries; Time Series Theory and Evidence," *The Review of Economic Statistics*, 135-146.

Billingsley, Patrick (1986): *Probability and Measure*, New York: John Wiley & Sons.

Canzoneri, Matthew B., Robert E. Cumby, and Behzad Diba (1999): "Relative Labor Productivity and the Real Exchange Rate in the Long Run: Evidence for a Panel of OECD Countries," *Journal of International Economics,* 47, 245-66.

Chung, Kai Lai (1974): *A Course in Probability Theory*, San Diego: Academic Press.

Davies, R. B. (1969): "Beta Optimal Tests and an Application to the Summary Evaluation of Experiments," *Journal of the Royal Statistical Society, Series B,* 31, 524-38.

Eaton, Morris L. (1970): "A Complete Class Theorem for Multidimensional One-Sided Alternatives," *Annals of Mathematical Statistics*, 41, no. 6, 1884-1888.

Effron, Bradley (1975): "Defining the Curvature of a Statistical Problem (with Applications to Second Order Efficiency," *The Annals of Statistics*, 3, no.6, 1189-1242.

Effron, Bradley (1978): "The Geometry of Exponential Families," *The Annals of Statistics*, 6, no. 2, 362-76.

Elliott, Graham (1993): "Efficient Tests for a Unit Root when the Initial Observation is Drawn from its Unconditional Distribution," unpublished manuscript.

Elliott, Graham, Thomas J. Rothenberg, and James H. Stock (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, no. 3 (May), 813-36.

Evans, P. and G. Karras (1996): "Convergence Revisited," forthcoming, *Journal of Monetary Economics*.

Fan, K. and A. J. Hoffman (1955): "Some Metric Inequalities in the Space of Matrices," *Proceedings of the American Mathematical Society*, 6, 111-16.

Fisher, R. A. (1932): *Statistical Methods for Research Workers*, Edinburgh: Oliver and Boyd.

Frankel, J. and A. Rose (1995): "A Panel Project on Purchasing Power Parity: Mean Reversion Within and Between Countries," *Journal of International Economics*, 40, 209-24.

Fuller, Wayne (1976): *Introduction to Statistical Time Series*, New York: John Wiley & Sons.

Hamilton, James D. (1994): *Time Series Analysis*. Princeton: Princeton University Press.

Im, Kyung So, M. Hashem Pesaran, and Yongcheol Shin (1997): "Testing for Unit Roots in Heterogeneous Panels," Manuscript, University of Cambridge.

Lehmann, E. L. (1959): *Testing Statistical Hypotheses*, New York: John Wiley & Sons.

Andrew Levin and Chien-Fu Lin (1993): "Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties," Manuscript, University of California, San Diego Discussion Paper 93-56.

Lon-Mu Liu and George C. Tiao (1980): "Random Coefficient First-Order Autoregressive Models," *Journal of Econometrics*, 13, 305-325.

Maddala, G.S. and Shaowen Wu, (1997) Fs"A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test," Manuscript, Ohio State University.

Matthes, T.K. and Truax, D.R. (1967): "Tests of Composite Hypotheses for the Multivariate Exponential Family," *Annals of Mathematical Statistics*, 38, 681-697.

Nankervis, J.C. and N.E. Savin (1988): "The Exact Moments of the Least-Squares Estimator for the Autoregressive Model," *Journal of Econometrics*, 37, 381-388.

O'Connell, Paul (1998): "The Overvaluation of Purchasing Power Parity," *Journal of International Economics*, 44, 1-19.

Oh, K (1996): "Purchasing Power Parity and Unit Root Tests Using Panel Data," *Journal of International Money and Finance,* 15, 405-18.

Papell, D (1997): "Searching for Stationarity: Purchasing Power Parity Under the Current Float," *Journal of International Economics,* 43, 313-32.

Pedroni, Peter (1997): "Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis," manuscript, Indiana University.

Daniel Quah (1994): "Exploiting Cross-Section Variations for Unit Root Inference in Dynamic Data," *Economics Letters*, 44, 9-19.

Rao, R. Ranga (1962): "Relations Between Weak and Uniform Convergence of Measures with Applications," *Annals of Mathematical Statistics*, 33, 659-680.

Sawa, Takamitsu (1978): "The Exact Moments of the Least Squares Estimator for the Autoregressive Model," *Journal of Econometrics*, 8, 159-72.

Stein, Charles (1956): "The Admissability of Hotelling's T^2-Test," Annals of Mathematical Statistics, 27. 616-623.

Abraham Wald (1950): *Statistical Decision Functions*, New York: Wiley.

Wei, Shang Jin and David Parsley (1995): "Purchasing Power Dis-Parity During the Floating Rate Period: Exchange Rate Volatility, Trade Barriers, and Other Culprits," manuscript, Harvard University.

Wu, Y (1996): "Are Real Exchange Rates Nonstationary? Evidence from a Panel-Data Test," *Journal of Money, Credit and Banking*, 28, no. 1, 54-63.

Table 1 ($\bar{z}_{nt} = 0$)
Size Adjusted Powers for N Groups, K of which have $\theta = \bar{\theta}$ (5% size)

		$\bar{\theta} = -.05$					$\bar{\theta} = -.10$				
		N=5	N=10	N=15	N=20	N=25	N=5	N=10	N=15	N=20	N=25
LL	K=1	0.11	0.08	0.07	0.07	0.06	0.12	0.09	0.08	0.07	0.07
	3	0.41	0.19	0.16	0.13	0.11	0.48	0.22	0.16	0.14	0.13
	5	1.00	0.41	0.28	0.22	0.18	1.00	0.46	0.32	0.24	0.21
	10	.	1.00	0.77	0.59	0.47	.	1.00	0.82	0.64	0.53
	15	.	.	1.00	0.93	0.82	.	.	1.00	0.95	0.86
	20	.	.	.	1.00	0.98	.	.	.	1.00	0.99
	25	.	.	.	.	1.00	.	.	.	.	1.00
IPS-ADF	K=1	0.13	0.10	0.10	0.08	0.08	0.22	0.15	0.13	0.11	0.10
	3	0.55	0.33	0.26	0.21	0.19	0.89	0.60	0.44	0.37	0.31
	5	0.98	0.69	0.53	0.42	0.36	1.00	0.96	0.83	0.72	0.62
	10	.	1.00	0.99	0.93	0.86	.	1.00	1.00	1.00	0.99
	15	.	.	1.00	1.00	1.00	.	.	1.00	1.00	1.00
	20	.	.	.	1.00	1.00	.	.	.	1.00	1.00
	25	.	.	.	.	1.00	.	.	.	.	1.00
Bayes1	K=1	0.13	0.10	0.09	0.08	0.08	0.23	0.17	0.14	0.12	0.12
	3	0.54	0.32	0.24	0.20	0.17	0.85	0.59	0.47	0.38	0.34
	5	0.99	0.67	0.50	0.40	0.33	1.00	0.95	0.83	0.72	0.64
	10	.	1.00	0.99	0.92	0.84	.	1.00	1.00	1.00	0.99
	15	.	.	1.00	1.00	1.00	.	.	1.00	1.00	1.00
	20	.	.	.	1.00	1.00	.	.	.	1.00	1.00
	25	.	.	.	.	1.00	.	.	.	.	1.00
Bonferoni	K=1	0.11	0.08	0.07	0.07	0.06	0.32	0.21	0.17	0.15	0.12
	3	0.23	0.14	0.11	0.10	0.09	0.65	0.45	0.36	0.30	0.26
	5	0.33	0.20	0.15	0.13	0.11	0.82	0.61	0.51	0.43	0.39
	10	.	0.33	0.24	0.20	0.17	.	0.84	0.74	0.66	0.60
	15	.	.	0.33	0.26	0.23	.	.	0.87	0.79	0.73
	20	.	.	.	0.32	0.28	.	.	.	0.88	0.83
	25	.	.	.	.	0.33	.	.	.	.	0.89
Bayes2	K=1	0.12	0.09	0.07	0.08	0.08	0.28	0.21	0.18	0.15	0.14
	3	0.39	0.18	0.13	0.13	0.12	0.76	0.52	0.44	0.37	0.33
	5	0.95	0.34	0.22	0.21	0.18	1.00	0.80	0.67	0.59	0.54
	10	.	0.99	0.58	0.47	0.40	.	1.00	0.98	0.93	0.89
	15	.	.	1.00	0.88	0.72	.	.	1.00	1.00	0.99
	20	.	.	.	1.00	0.98	.	.	.	1.00	1.00
	25	.	.	.	.	1.00	.	.	.	.	1.00

		$\bar{\theta} = -.20$					$\bar{\theta} = -.40$				
LL	K=1	0.13	0.09	0.09	0.07	0.06	0.13	0.09	0.08	0.07	0.07
	3	0.51	0.25	0.18	0.14	0.13	0.53	0.25	0.18	0.14	0.12
	5	1.00	0.50	0.34	0.25	0.21	1.00	0.52	0.33	0.26	0.22
	10	.	1.00	0.85	0.67	0.55	.	1.00	0.85	0.68	0.57
	15	.	.	1.00	0.96	0.88	.	.	1.00	0.97	0.89
	20	.	.	.	1.00	0.99	.	.	.	1.00	0.99
	25	.	.	.	.	1.00	.	.	.	.	1.00
IPS-ADF	K=1	0.37	0.22	0.18	0.15	0.14	0.60	0.37	0.28	0.23	0.20
	3	1.00	0.87	0.72	0.61	0.53	1.00	0.99	0.95	0.88	0.81
	5	1.00	1.00	0.99	0.95	0.89	1.00	1.00	1.00	1.00	0.99
	10	.	1.00	1.00	1.00	1.00	.	1.00	1.00	1.00	1.00
	15	.	.	1.00	1.00	1.00	.	.	1.00	1.00	1.00
	20	.	.	.	1.00	1.00	.	.	.	1.00	1.00
	25	.	.	.	.	1.00	.	.	.	.	1.00
Bayes1	K=1	0.57	0.43	0.36	0.31	0.28	0.98	0.94	0.89	0.83	0.78
	3	1.00	0.96	0.91	0.86	0.81	1.00	1.00	1.00	1.00	1.00
	5	1.00	1.00	1.00	0.99	0.98	1.00	1.00	1.00	1.00	1.00
	10	.	1.00	1.00	1.00	1.00	.	1.00	1.00	1.00	1.00
	15	.	.	1.00	1.00	1.00	.	.	1.00	1.00	1.00
	20	.	.	.	1.00	1.00	.	.	.	1.00	1.00
	25	.	.	.	.	1.00	.	.	.	.	1.00
Bonferoni	K=1	0.86	0.72	0.64	0.57	0.54	1.00	1.00	1.00	0.99	0.99
	3	1.00	0.97	0.95	0.91	0.89	1.00	1.00	1.00	1.00	1.00
	5	1.00	1.00	0.99	0.98	0.97	1.00	1.00	1.00	1.00	1.00
	10	.	1.00	1.00	1.00	1.00	.	1.00	1.00	1.00	1.00
	15	.	.	1.00	1.00	1.00	.	.	1.00	1.00	1.00
	20	.	.	.	1.00	1.00	.	.	.	1.00	1.00
	25	.	.	.	.	1.00	.	.	.	.	1.00
Bayes2	K=1	0.80	0.71	0.64	0.57	0.53	1.00	1.00	1.00	0.99	0.99
	3	1.00	0.99	0.97	0.96	0.95	1.00	1.00	1.00	1.00	1.00
	5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	10	.	1.00	1.00	1.00	1.00	.	1.00	1.00	1.00	1.00
	15	.	.	1.00	1.00	1.00	.	.	1.00	1.00	1.00
	20	.	.	.	1.00	1.00	.	.	.	1.00	1.00
	25	.	.	.	.	1.00	.	.	.	.	1.00

Table 2 ($\bar{z}_{nt} = 1$)
Size Adjusted Powers for N Groups, K of which have $\theta = \bar{\theta}$ (5% size)

		$\bar{\theta} = -.05$					$\bar{\theta} = -.10$				
		N=5	N=10	N=15	N=20	N=25	N=5	N=10	N=15	N=20	N=25
LL	K=1	0.07	0.06	0.06	0.06	0.06	0.09	0.07	0.06	0.06	0.06
	3	0.17	0.11	0.09	0.08	0.08	0.33	0.16	0.12	0.10	0.09
	5	0.46	0.19	0.14	0.12	0.11	0.97	0.33	0.21	0.17	0.13
	10	.	0.82	0.42	0.28	0.23	.	1.00	0.66	0.46	0.35
	15	.	.	0.95	0.62	0.46	.	.	1.00	0.86	0.69
	20	.	.	.	0.99	0.79	.	.	.	1.00	0.95
	25	.	.	.	.	1.00	.	.	.	.	1.00
IPS-ADF	K=1	0.08	0.07	0.06	0.07	0.07	0.13	0.10	0.09	0.08	0.08
	3	0.19	0.14	0.12	0.11	0.10	0.50	0.30	0.24	0.20	0.17
	5	0.41	0.25	0.20	0.18	0.16	0.94	0.63	0.48	0.38	0.32
	10	.	0.72	0.52	0.43	0.37	.	1.00	0.97	0.89	0.81
	15	.	.	0.88	0.75	0.66	.	.	1.00	1.00	0.99
	20	.	.	.	0.96	0.89	.	.	.	1.00	1.00
	25	.	.	.	.	0.98	.	.	.	.	1.00
Bayes1	K=1	0.09	0.08	0.07	0.07	0.07	0.15	0.11	0.11	0.10	0.09
	3	0.23	0.17	0.13	0.13	0.11	0.63	0.42	0.33	0.28	0.24
	5	0.46	0.31	0.23	0.21	0.18	0.95	0.79	0.66	0.55	0.47
	10	.	0.73	0.58	0.51	0.42	.	1.00	0.99	0.98	0.95
	15	.	.	0.86	0.79	0.71	.	.	1.00	1.00	1.00
	20	.	.	.	0.94	0.90	.	.	.	1.00	1.00
	25	.	.	.	.	0.98	.	.	.	.	1.00
Bonferoni	K=1	0.07	0.05	0.05	0.05	0.05	0.11	0.08	0.07	0.06	0.06
	3	0.09	0.07	0.06	0.06	0.05	0.22	0.15	0.11	0.09	0.09
	5	0.12	0.08	0.07	0.07	0.06	0.32	0.21	0.16	0.13	0.11
	10	.	0.11	0.10	0.08	0.07	.	0.34	0.24	0.20	0.17
	15	.	.	0.12	0.10	0.09	.	.	0.33	0.27	0.22
	20	.	.	.	0.12	0.10	.	.	.	0.33	0.28
	25	.	.	.	.	0.11	.	.	.	.	0.33
Bayes2	K=1	0.08	0.07	0.06	0.06	0.06	0.14	0.12	0.10	0.09	0.09
	3	0.15	0.11	0.09	0.09	0.09	0.42	0.31	0.25	0.22	0.21
	5	0.23	0.16	0.13	0.13	0.12	0.69	0.53	0.44	0.39	0.37
	10	.	0.32	0.25	0.24	0.22	.	0.91	0.84	0.79	0.74
	15	.	.	0.41	0.38	0.35	.	.	0.98	0.96	0.94
	20	.	.	.	0.53	0.48	.	.	.	0.99	0.99
	25	.	.	.	.	0.61	.	.	.	.	1.00
		$\bar{\theta} = -.20$					$\bar{\theta} = -.40$				
LL	K=1	0.11	0.08	0.07	0.07	0.06	0.12	0.08	0.07	0.07	0.07
	3	0.48	0.20	0.15	0.12	0.11	0.54	0.23	0.16	0.13	0.12
	5	1.00	0.44	0.27	0.21	0.16	1.00	0.51	0.32	0.24	0.20
	10	.	1.00	0.80	0.60	0.46	.	1.00	0.85	0.67	0.54
	15	.	.	1.00	0.95	0.83	.	.	1.00	0.97	0.89
	20	.	.	.	1.00	0.99	.	.	.	1.00	1.00
	25	.	.	.	.	1.00	.	.	.	.	1.00
IPS-ADF	K=1	0.24	0.17	0.13	0.12	0.11	0.49	0.31	0.24	0.20	0.18
	3	0.93	0.67	0.50	0.42	0.36	1.00	0.97	0.88	0.79	0.69
	5	1.00	0.98	0.89	0.80	0.71	1.00	1.00	1.00	0.99	0.98
	10	.	1.00	1.00	1.00	1.00	.	1.00	1.00	1.00	1.00
	15	.	.	1.00	1.00	1.00	.	.	1.00	1.00	1.00
	20	.	.	.	1.00	1.00	.	.	.	1.00	1.00
	25	.	.	.	.	1.00	.	.	.	.	1.00
Bayes1	K=1	0.35	0.24	0.19	0.17	0.15	0.83	0.61	0.49	0.42	0.37
	3	1.00	0.92	0.80	0.70	0.62	1.00	1.00	1.00	1.00	0.99
	5	1.00	1.00	1.00	0.99	0.95	1.00	1.00	1.00	1.00	1.00
	10	.	1.00	1.00	1.00	1.00	.	1.00	1.00	1.00	1.00
	15	.	.	1.00	1.00	1.00	.	.	1.00	1.00	1.00
	20	.	.	.	1.00	1.00	.	.	.	1.00	1.00
	25	.	.	.	.	1.00	.	.	.	.	1.00
Bonferoni	K=1	0.41	0.28	0.23	0.19	0.17	0.97	0.92	0.87	0.83	0.80
	3	0.77	0.60	0.48	0.43	0.37	1.00	1.00	1.00	0.99	0.99
	5	0.91	0.78	0.66	0.59	0.53	1.00	1.00	1.00	1.00	1.00
	10	.	0.95	0.88	0.82	0.76	.	1.00	1.00	1.00	1.00
	15	.	.	0.96	0.92	0.88	.	.	1.00	1.00	1.00
	20	.	.	.	0.97	0.94	.	.	.	1.00	1.00
	25	.	.	.	.	0.97	.	.	.	.	1.00
Bayes2	K=1	0.48	0.35	0.32	0.28	0.24	0.97	0.94	0.90	0.85	0.81
	3	0.97	0.91	0.86	0.80	0.75	1.00	1.00	1.00	1.00	1.00
	5	1.00	1.00	0.99	0.98	0.97	1.00	1.00	1.00	1.00	1.00
	10	.	1.00	1.00	1.00	1.00	.	1.00	1.00	1.00	1.00
	15	.	.	1.00	1.00	1.00	.	.	1.00	1.00	1.00
	20	.	.	.	1.00	1.00	.	.	.	1.00	1.00
	25	.	.	.	.	1.00	.	.	.	.	1.00

Table 3 ($\bar{z}_{nt} = \{1, t\}$)

Size Adjusted Powers for N Groups, K of which have $\theta = \bar{\theta}$ (5% size)

		$\bar{\theta} = -.05$					$\bar{\theta} = -.10$				
		N=5	N=10	N=15	N=20	N=25	N=5	N=10	N=15	N=20	N=25
LL	K=1	0.07	0.06	0.06	0.06	0.06	0.09	0.07	0.07	0.06	0.07
	3	0.12	0.09	0.08	0.08	0.08	0.27	0.15	0.12	0.11	0.10
	5	0.20	0.13	0.12	0.10	0.10	0.69	0.30	0.21	0.18	0.16
	10	.	0.34	0.25	0.20	0.17	.	0.95	0.63	0.47	0.39
	15	.	.	0.49	0.36	0.29	.	.	0.99	0.85	0.71
	20	.	.	.	0.59	0.48	.	.	.	1.00	0.95
	25	.	.	.	.	0.69	.	.	.	.	1.00
IPS-ADF	K=1	0.06	0.06	0.06	0.06	0.06	0.10	0.09	0.08	0.07	0.07
	3	0.11	0.10	0.08	0.08	0.07	0.30	0.20	0.15	0.15	0.12
	5	0.18	0.14	0.11	0.10	0.10	0.64	0.39	0.29	0.25	0.21
	10	.	0.31	0.23	0.20	0.18	.	0.92	0.75	0.63	0.54
	15	.	.	0.41	0.33	0.29	.	.	0.98	0.93	0.86
	20	.	.	.	0.51	0.44	.	.	.	1.00	0.98
	25	.	.	.	.	0.60	.	.	.	.	1.00
Bayes1	K=1	0.06	0.06	0.06	0.06	0.06	0.11	0.09	0.08	0.08	0.07
	3	0.11	0.09	0.09	0.08	0.08	0.31	0.21	0.17	0.15	0.14
	5	0.19	0.13	0.12	0.11	0.10	0.61	0.40	0.31	0.27	0.23
	10	.	0.30	0.24	0.21	0.19	.	0.90	0.77	0.66	0.57
	15	.	.	0.42	0.35	0.30	.	.	0.98	0.93	0.88
	20	.	.	.	0.53	0.45	.	.	.	1.00	0.99
	25	.	.	.	.	0.61	.	.	.	.	1.00
Bonferoni	K=1	0.06	0.05	0.05	0.05	0.05	0.09	0.07	0.06	0.06	0.06
	3	0.07	0.06	0.06	0.05	0.05	0.14	0.10	0.08	0.08	0.08
	5	0.09	0.07	0.06	0.06	0.06	0.21	0.14	0.11	0.10	0.09
	10	.	0.09	0.08	0.07	0.07	.	0.23	0.17	0.14	0.13
	15	.	.	0.09	0.08	0.07	.	.	0.22	0.19	0.17
	20	.	.	.	0.09	0.08	.	.	.	0.23	0.21
	25	.	.	.	.	0.09	.	.	.	.	0.24
Bayes2	K=1	0.06	0.06	0.05	0.06	0.06	0.11	0.08	0.08	0.07	0.06
	3	0.09	0.07	0.06	0.07	0.07	0.22	0.14	0.13	0.11	0.10
	5	0.11	0.09	0.08	0.08	0.07	0.32	0.23	0.19	0.17	0.15
	10	.	0.14	0.12	0.12	0.10	.	0.42	0.37	0.33	0.29
	15	.	.	0.15	0.15	0.13	.	.	0.55	0.50	0.45
	20	.	.	.	0.18	0.17	.	.	.	0.64	0.59
	25	.	.	.	.	0.20	.	.	.	.	0.71

		$\bar{\theta} = -.20$					$\bar{\theta} = -.40$				
LL	K=1	0.12	0.09	0.08	0.08	0.07	0.15	0.10	0.09	0.08	0.08
	3	0.52	0.26	0.17	0.16	0.13	0.69	0.34	0.24	0.20	0.18
	5	1.00	0.55	0.36	0.29	0.23	1.00	0.71	0.49	0.38	0.33
	10	.	1.00	0.91	0.76	0.64	.	1.00	0.97	0.89	0.81
	15	.	.	1.00	0.99	0.94	.	.	1.00	1.00	0.99
	20	.	.	.	1.00	1.00	.	.	.	1.00	1.00
	25	.	.	.	.	1.00	.	.	.	.	1.00
IPS-ADF	K=1	0.18	0.15	0.12	0.10	0.10	0.42	0.27	0.21	0.18	0.17
	3	0.79	0.54	0.39	0.31	0.29	1.00	0.94	0.83	0.72	0.64
	5	1.00	0.91	0.76	0.64	0.57	1.00	1.00	1.00	0.98	0.96
	10	.	1.00	1.00	0.99	0.98	.	1.00	1.00	1.00	1.00
	15	.	.	1.00	1.00	1.00	.	.	1.00	1.00	1.00
	20	.	.	.	1.00	1.00	.	.	.	1.00	1.00
	25	.	.	.	.	1.00	.	.	.	.	1.00
Bayes1	K=1	0.25	0.17	0.14	0.13	0.11	0.68	0.46	0.34	0.29	0.24
	3	0.84	0.62	0.50	0.42	0.37	1.00	1.00	0.98	0.95	0.90
	5	1.00	0.94	0.84	0.76	0.68	1.00	1.00	1.00	1.00	1.00
	10	.	1.00	1.00	1.00	0.99	.	1.00	1.00	1.00	1.00
	15	.	.	1.00	1.00	1.00	.	.	1.00	1.00	1.00
	20	.	.	.	1.00	1.00	.	.	.	1.00	1.00
	25	.	.	.	.	1.00	.	.	.	.	1.00
Bonferoni	K=1	0.25	0.18	0.14	0.12	0.11	0.85	0.76	0.69	0.65	0.61
	3	0.53	0.38	0.30	0.26	0.23	1.00	0.98	0.96	0.95	0.93
	5	0.71	0.53	0.43	0.37	0.34	1.00	1.00	1.00	0.99	0.99
	10	.	0.78	0.65	0.59	0.54	.	1.00	1.00	1.00	1.00
	15	.	.	0.79	0.72	0.68	.	.	1.00	1.00	1.00
	20	.	.	.	0.82	0.77	.	.	.	1.00	1.00
	25	.	.	.	.	0.84	.	.	.	.	1.00
Bayes2	K=1	0.33	0.23	0.19	0.16	0.15	0.91	0.81	0.72	0.66	0.57
	3	0.74	0.61	0.54	0.49	0.45	1.00	1.00	1.00	1.00	1.00
	5	0.91	0.84	0.79	0.74	0.72	1.00	1.00	1.00	1.00	1.00
	10	.	0.99	0.98	0.97	0.97	.	1.00	1.00	1.00	1.00
	15	.	.	1.00	1.00	1.00	.	.	1.00	1.00	1.00
	20	.	.	.	1.00	1.00	.	.	.	1.00	1.00
	25	.	.	.	.	1.00	.	.	.	.	1.00

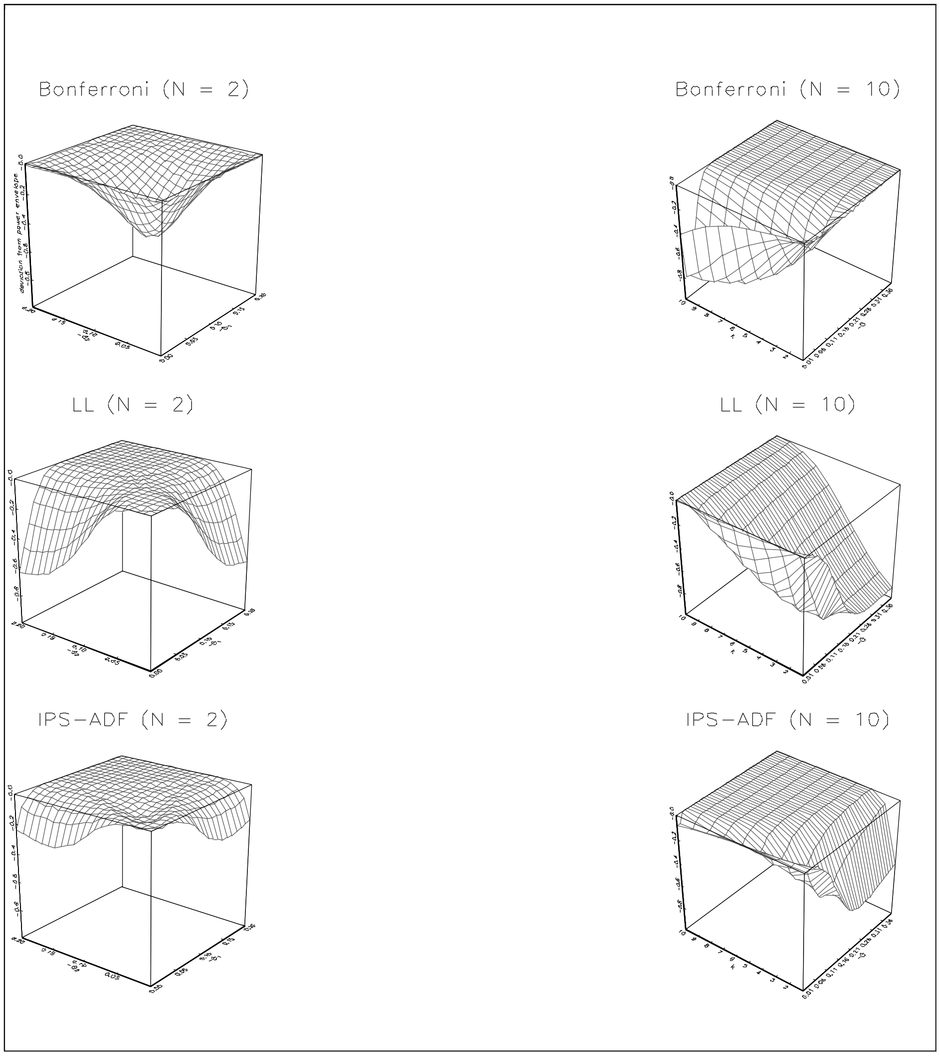

Figure 3: Deviations of power functions of the Bonferroni, LL, and IPS tests from symmetric power envelope for N = 2 (left panels) and N = 10 (right panels).

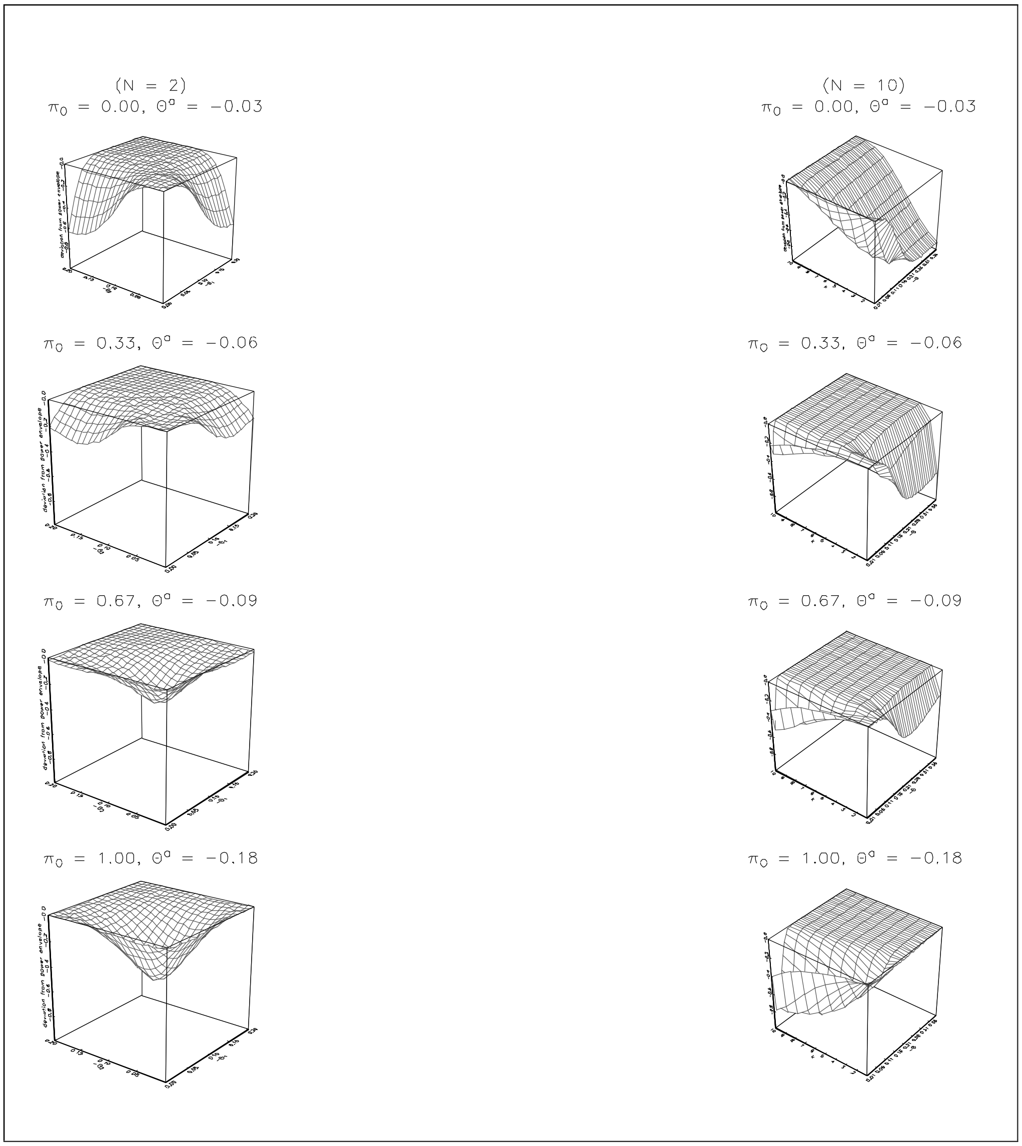

Figure 4: Bayes tests as parameters π and θ^a are varied.

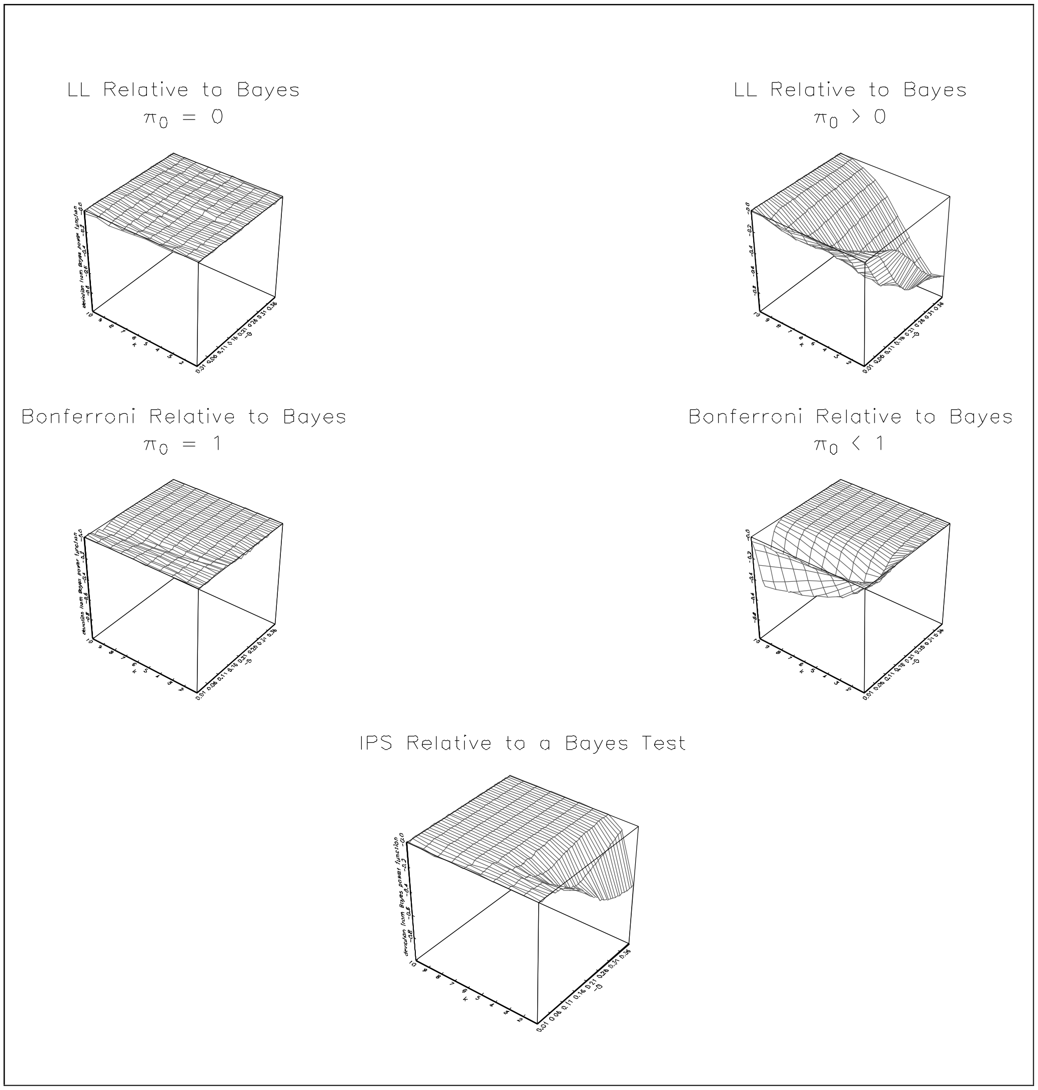

Figure 5: Comparison of Powers of Various Bayes Tests to IPS, LL, and Bonferroni Tests (T=100, 5% Size. N Groups, K of which have $\rho < 1$.)

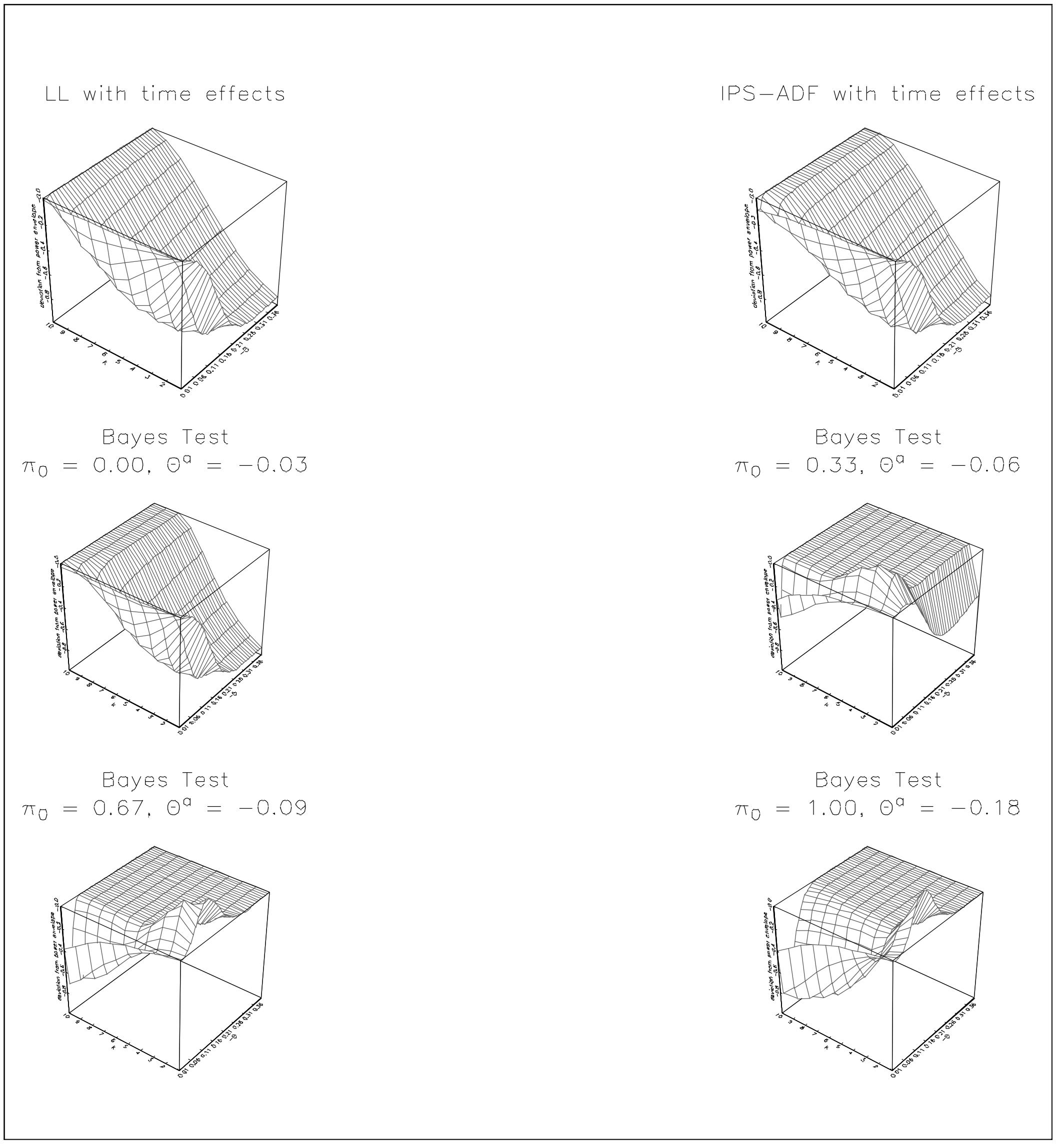

Figure 6: Comparison of Powers of LL, IPS and Bayes Tests when covarinace matrix is nondiagonal (T=100, 5% Size. N Groups, K of which have $\rho < 1$.)