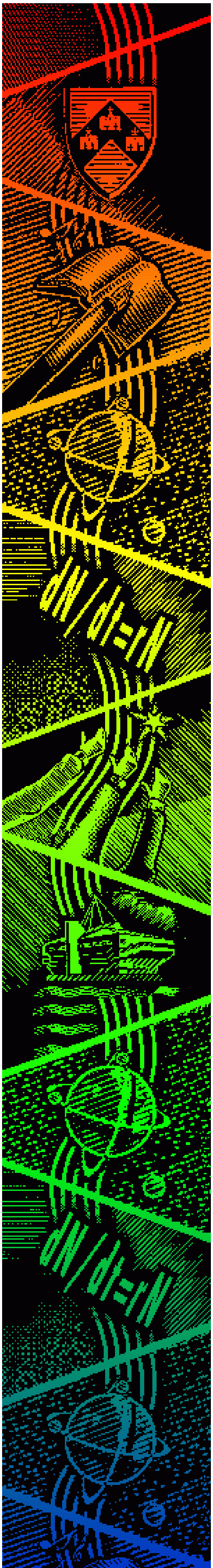

THE UNIVERSITY of York

Discussion Papers in Economics

No. 2006/19

Constructing Optimal Tests on a Lagged Dependent Variable

by

Patrick Marsh

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

Constructing Optimal Tests on a Lagged Dependent Variable

Patrick Marsh

Department of Economics, University of York[1]

3rd October 2006

[1] Address for Correspondence: Department of Economics, University of York, Heslington, York, YO10 5DD, U.K., e-mail: pwnm1@york.ac.uk, tel. +44 1904 433084, fax: +44 1904 433759.

Abstract:

Via the leading unit root case, the problem of testing on a lagged dependent variable is characterized by a nuisance parameter which is present only under the alternative (see Andrews and Ploberger (1994)). This has proven a barrier to the construction of optimal tests. Moreover, in their absence it is impossible to objectively assess the absolute power properties of existing tests. Indeed, feasible tests based upon the optimality criteria used here are found to have numerically superior power properties to both the original Dickey and Fuller (1981) statistics and the efficient detrended versions suggested by Elliott, Rothenberg and Stock (1996) and analysed in Burridge and Taylor (2000).

Keywords: nuisance parameter, invariant test, unit root.

1 Introduction

This paper proposes methods by which optimal tests on a lagged dependent variable in a linear regression model may be constructed. Both the need for and difficulties associated with inference on a lagged dependent variable are highlighted via the leading unit root case, as considered by Dickey and Fuller (1979, 1981). Although we will ultimately consider all cases, initially suppose that

$$y_t = \beta_1 + \beta_2 t + \rho y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim iid(0, \sigma^2), \quad t = 1, 2, ., T. \tag{1}$$

Naive testing for a unit root in (1) has rightly been criticized, as in Schmidt and Phillips (1992), as the degree of deterministic trending is different under the unit root. Instead consider testing

$$H_0 : \rho = 1 \cap \beta_2 = 0 \quad \text{vs.} \quad H_1 : |\rho| < 1 \cap \beta_2 \neq 0, \tag{2}$$

so that the degree of trending is linear under either hypothesis. In DeJong, Nankervis, Savin and Whiteman (1992), imposing the restrictions on the parameters in (1)

$$\beta_1 = (\alpha_1(1-\rho) + \beta_2\rho)\,, \quad \beta_2 = \alpha_2(1-\rho), \tag{3}$$

implies a model of the form

$$y_t = \alpha_1 + \alpha_2 t + u_t \quad ; \quad u_t = \rho u_{t-1} + \varepsilon_t, \tag{4}$$

which has formed the basis for the majority of recent unit root tests, such as those in Dufour and King (1991) and Elliott, Rothenberg and Stock (1996). Indeed the latter provides GLS-type tests which have become benchmark, see Burridge and Taylor (2000).

Here we do not impose these restrictions. None-the-less under H_0 in (2) the two formulations coincide. Under trend stationary alternatives they differ, in that in (1) the mean of y_t depends on the autoregressive parameter, while in (4) it does not. For any given data set, in the event of rejection, it will not be possible to say which trend stationary formulation generated it. Consequently, it is necessary that we have unit root tests which are powerful against trend stationary processes characterized by (1).

This analysis also highlights the key difficulty with testing on a lagged dependent variable. Since in (1) the mean of y_t depends on ρ, the alternative distribution of any reasonable test statistic will depend not only on ρ, but also the parameters β_1 and β_2. Thus, while we are easily able to construct tests having known size, under the alternative β_1 becomes a nuisance parameter. That is, as in Andrews and Ploberger (1994), there is a nuisance parameter present only under the alternative. This remains true for testing that ρ is any value, including zero, and regardless of any additional restrictions imposed by the null. This difficulty has, so far, prevented the construction of any optimality theory in this case. Thus objective assessment of the tests we do have is not possible.

As in Andrews and Ploberger (1994) the solution is to provide tests which are weighted optimal, with the influence of the nuisance parameter on power integrated out. Specifically, within the semi-parametric elliptically symmetric family, we achieve the following. It is shown that integrating out in power is equivalent to applying optimality criteria to the integrated density of Berger, Liseo and Wolpert (1999). There nuisance parameters are directly integrated out of the sample density, before constructing estimators and tests for interest parameters. We then provide a method for finding such integrated densities, in the elliptic family, which avoids all of the technical difficulties usually associated with integrating out such parameters. Weighted optimal tests such as point or locally optimal ones, follow by applying the appropriate criteria to the integrated density.

The numerical evidence presented in the paper focuses on the unit root case. It is shown that, appropriate to their context, the original Dickey-Fuller (1979) t-tests have powers close to a weighted power envelope, and thus we can suggest no improvement. On the other hand, for joint hypotheses such as in (2) their (1981) F-tests are short of optimal, with a feasible point optimal test having significantly superior power properties. Currently, the favoured test seems to be the GLS Dickey-Fuller (DF_{GLS}) test as described in Elliott, Rothenberg and Stock (1996). For testing a unit root in (1) via (2) our feasible test slightly outperforms this competitor. In formulation (4), this is reversed. However, the DF_{GLS} is known to have power which evaporates as the initial condition grows. Here, our test is shown to significantly out-perform this test for both moderate and large initial conditions, in both formulations of the problem.

The plan for the paper is as follows. The next section details the model and assumptions. Section 3 presents and discusses the main results in the context of the simple unit root test on ρ, while Section 4 discusses the results and gives the numerical power comparisons. Technical proofs and some tables are placed in an appendix.

2 Model and Motivation

To formalize our treatment we shall consider models which generalize those in (1) and (4), with

$$M_1 \; : \; y_t = \rho y_{t-1} + x_t'\beta + \varepsilon_t \quad \text{and}$$
$$M_2 \; : \; y_t = x_t'\alpha + u_t \quad ; \quad u_t = \rho u_{t-1} + \varepsilon_t,$$

where now x_t is any k vector of regressors and β and α are k vectors of unknowns. To continue let $y = (y_1, ..., y_T)'$, $\varepsilon = (\varepsilon_1, ..., \varepsilon_T)'$, $X = (x_1', .., x_T')'$ so these may be written as:

$$M_1 \; : \; \Delta_\rho y = X\beta + \varepsilon \quad \text{and} \tag{5}$$

$$M_2 \quad : \quad \Delta_\rho y = \Delta_\rho X\alpha + \varepsilon, \tag{6}$$

where $\Delta_\rho = I_N - \rho L$ and L is the $T \times T$ matrix lag-operator, having one's on the first lower off-diagonal and zero's elsewhere. We shall proceed under the following assumption on the joint density of the innovations ε;

Assumption 1 *Let $\mathcal{F}(\mu, \Sigma)$ denote the elliptically symmetric family with mean μ and variance Σ, then assume that the density of ε, $f(\varepsilon) \in \mathcal{F}(0, \sigma^2\Omega)$, $|\Omega| = 1$, with*

$$\mathcal{F}(0, \Omega) = \left\{ f : f(\varepsilon) = q\left[\varepsilon'\Omega^{-1}\varepsilon\right)\right] \right\},$$

where $q(.)$ is a nonincreasing convex, measurable function on $[0, \infty)$. ■

Given Assumption 1 and since for either model y is a linear transformation of the innovations then both the data y and also the differenced data $\Delta_1 y = \{y_t - y_{t-1}\}_{t=1}^T$ is also elliptically symmetric, with

$$\begin{aligned} M_1 \quad &: \quad \Delta_1 y \sim \mathcal{F}\left(\Delta_1\Delta_\rho^{-1}X\beta, \sigma^2\Sigma_\rho\right) \quad ; \quad M_2 : \Delta_1 y \sim \mathcal{F}\left(\Delta_1 X\alpha, \sigma^2\Sigma_\rho\right), \\ \Sigma_\rho \quad &= \quad \Delta_1\Delta_\rho^{-1}\Omega\left(\Delta_\rho^{-1}\right)'\Delta_1', \end{aligned}$$

so that the difference between the models is characterized by the dependence of the mean upon the parameter ρ.

To illustrate this difference, suppose that $\Omega = I_T$ and first consider the simple unit root test

$$H_0 : \rho = 1 \quad \text{vs.} \quad H_0 : |\rho| < 1 \tag{7}$$

in both M_1 and M_2. Under H_0 we have

$$M_1 : E[\Delta_1 y] = X\beta \quad ; \quad M_2 : E[\Delta_1 y] = \Delta_1 X\alpha,$$

while under H_1

$$M_1 : E[\Delta_1 y] = \Delta_1\Delta_\rho^{-1}X\beta \quad ; \quad M_2 : E[\Delta_1 y] = \Delta_1 X\alpha,$$

that is for M_2 the mean of the data does not depend on ρ. Thus, for M_2, construction of either invariant (as in King (1980) or Dufour and King (1991)) or similar (assuming Gaussian innovations as in Hillier (1987)) tests is straightforward.

First define a matrix C_2 by

$$C_2 C_2' = M_Z = I - (\Delta_1 X)(X'\Delta_1'\Delta_1 X)^{-1} X'\Delta_1' \quad ; \quad C_2' C_2 = I_{N-k},$$

and then

$$v_2 = \frac{w_2}{|w_2|} \quad ; \quad w_2 = C_2'\Delta_1 y \quad \text{and } A_2 = C_2'\Delta_1\Delta_\rho^{-1}\left(\Delta_\rho^{-1}\right)'\Delta_1' C_2,$$

so v_2 is the maximal invariant, of which all invariant tests are functions. It has density, with respect to normalized Haar measure on the surface of the $N-k$ unit sphere $\mathbb{S}_{N-k}$, given by King (1980), as

$$pdf(v_2;\rho) = \frac{\int_{a>0} pdf(aw_2|\, H_1)da}{\int_{a>0} pdf(aw_2|\, H_0)da} = |A_2|^{-1/2}\left(v' A_2^{-1} v\right)^{-(N-k)/2}.$$

Optimal tests then follow by applying optimality criteria to the density of the maximal invariant. Choices include point optimal tests (of which the Elliott, Rothenberg and Stock (1996) test is an example), locally best tests which maximize the slope of power at the null hypothesis (as in Dufour and King (1991)) or weighted average power tests which maximize power averaged over all ρ under the alternative (Forchini (2005)). Yet more criteria are examined in Forchini and Marsh (2000).

All of this is possible only because in M_2 the mean of the data does not depend on ρ. For M_1 since the mean of the data does depend on ρ the problem itself is not invariant. However, it is easy to construct tests which have known size. To do so define the matrix C_1 by

$$C_1 C_1' = M_X = I - X(X'X)^{-1}X' \quad ; \quad C_1' C_1 = I_{N-k},$$

and then

$$v_1 = \frac{w_1}{|w_1|} \quad ; \quad w_1 = C_1'\Delta_1 y,$$

so that the density of v_1 is constant on $\mathbb{S}_{N-k}$ under H_0. Any test which is a function of v_1 will therefore have known size.

An optimal test is designed to maximize (some function of) power. To fully illustrate the associated difficulties with attempting this for M_1, suppose briefly that the ε are Gaussian, and so

$$w_1 = C_1'\Delta_1 y \sim N(\tilde{X}_\rho\beta, \sigma^2 A_1),$$

where

$$\tilde{X}_\rho = C_1' \Delta_1 \Delta_\rho^{-1} X \quad \text{and} \quad A_1 = C_1' \Delta_1 \Delta_\rho^{-1} \left(\Delta_\rho^{-1}\right)' \Delta_1' C_1. \tag{8}$$

Consequently, we have for v_1

$$pdf(v_1; \rho) = \frac{\int_{a>0} pdf(aw_1 | H_1) da}{\int_{a>0} pdf(aw_1 | H_0) da} = |A_1|^{-1/2} \left(v_1' A_1^{-1} v_1\right)^{-(T-k)/2} \times h(v_1, \tilde{X}_\rho \beta, \rho), \tag{9}$$

where

$$\begin{aligned} h(v_1, \tilde{X}_\rho \beta, \rho) &= e^{-\xi/2} \int_{a>0} e^{-\frac{a}{2}} a^{\frac{T-k}{2}-1} \exp\left\{ \sqrt{a} \frac{v_1' A_1^{-1} \tilde{X}_\rho \beta}{\sqrt{v_1' A^{-1} v_1}} \right\} da, \\ \xi &= \beta' \tilde{X}_\rho' A_1^{-1} \tilde{X}_\rho \beta. \end{aligned}$$

Thus the density of v_1 equals a quantity which is equivalent to the density in the straightforward case M_2, multiplied by a term depending on $\tilde{X}_\rho \beta$ and the value of ρ. Since $pdf(v_1, 1) = 1$, then the second term only applies under H_1. It is in this sense that β is a nuisance parameter which is present only under the alternative, as in Andrews and Ploberger (1994). Moreover, as there, we will construct optimal tests by integrating out its influence on power. However, before proceeding to do this, we should note the following differences in the set-up here and in the latter paper.

First, even in the simplest Gaussian case, finding an explicit expression for power is not feasible, since we don't even have a resolved expression for the density. Consequently, a direct approach for the elliptic family as a whole will not work. Second, at least for the unit root case, hypotheses such as (7) make less sense than (2). In this case β becomes a mixture of parameters of interest and nuisance, thus representing a subtly different problem.

3 Weighted optimal inference on a lagged dependent variable.

In this section we provide weighted optimal tests which are applicable to either type of null hypothesis and fully workable in the semi-parametric elliptically symmetric family. The first obstacle to providing such tests is that even when we assume Gaussian innovations, we have no closed form for power.

We first show that weighted optimal tests follow from applying standard criteria to the integrated density of Berger, Liseo and Wolpert (1999). Here we will focus on providing point optimal ω^{PO} and locally best ω^{LB} tests, defined by

$$\omega^{PO} = \arg \max_{\omega \in \mathbb{S}_{N-k}} P_\omega = \int_\omega pdf(v_1;\rho)\,(dv) \quad \text{and} \tag{10}$$

$$\omega^{LB} = \arg \max_{\omega \in \mathbb{S}_{N-k}} \left.\frac{\partial P_\omega}{\partial \rho}\right|_{\rho=1}, \tag{11}$$

where $\mu = \mu(\tilde{X}_\rho, \beta)$ represents the nuisance parameter under the alternative. As in Andrews and Ploberger (1994) we will provide tests which are weighted optimal. That is, for some weight-function $\pi(\mu) : \mathbb{R}^{N-k} \to \mathbb{R}$, on the nuisance parameter μ, satisfying

$$\int_{\mathbb{R}^{N-k}} \pi(\mu) d\mu = 1, \qquad \text{for all } \beta \text{ and } \rho, \tag{12}$$

weighted versions of the tests in (10) and (11) are

$$\omega_\pi^{PO} = \arg \max_{\omega \in \mathbb{S}_{N-k}} \int_\mu P_\omega\, \pi(\mu) d\mu \quad \text{and}$$

$$\omega_\pi^{LB} = \arg \max_{\omega \in \mathbb{S}_{N-k}} \left.\frac{\partial \int_\mu P_\omega\, \pi(\mu) d\mu}{\partial \rho}\right|_{\rho=1}.$$

To proceed, follow Berger, Liseo and Wolpert (1999) and define the integrated density, first for v_1 by

$$\overline{pdf}(v_1;\rho) = \int_{\mu_1} pdf(v_1;\rho)\pi(\mu)d\mu, \tag{13}$$

and also for $w_1 = C_1' \Delta_1 y$, by

$$\overline{pdf}(w_1;\rho) = \int_{\mu_1} pdf(w_1;\rho)\pi(\mu)d\mu.$$

Moreover we can define a derived density for v_1, as

$$\widetilde{pdf}(v_1;\rho) = \frac{\int_{a>0} \overline{pdf}(aw_1|\, H_1)da}{\int_{a>0} \overline{pdf}(aw_1|\, H_0)da}, \tag{14}$$

that is $\widetilde{pdf}(v_1;\rho)$ is an integrated density derived from the integrated density of w_1.

We are now in a position to state and prove a theorem which clearly demonstrates the possibility of constructing weighted optimal tests in the semi-parametric case, even though no resolved expression for power exists.

Theorem 1 *Suppose that the data y is generated such that Assumption 1 is satisfied, then weighted optimal unit root tests on the lagged dependent variable are found via*

$$\omega_\pi^{PO} = \arg\max_{\omega \in \mathbb{S}_{N-k}} \overline{P}_\omega = \int_\omega \widetilde{pdf}(v_1;\rho)\,(dv) \quad \textit{and}$$
$$\omega_\pi^{LB} = \arg\max_{\omega \in \mathbb{S}_{N-k}} \left.\frac{\partial \overline{P}_\omega}{\partial \rho}\right|_{\rho=1}. \quad \blacksquare$$

The importance of Theorem 1 is twofold. First it establishes that weighted optimal tests can be found by applying standard optimality criteria to an integrated density. Thus every strictly optimal test has an immediate weighted analogue. Second, we need actually only consider finding an appropriate integrated density for $w_1 = C'\Delta_1 y$, rather than for v_1. This is crucial in providing a mechanism for constructing weighted optimal tests which circumvents the rather obvious obstacle of feasibility.

Since we have no resolved expressions for the density of v_1 it is far from clear how to choose a weighting function so as to integrate out the nuisance parameter. However, as Theorem 1 demonstrates, we need only consider weight functions which integrate out the nuisance parameter from the density of w_1. Moreover, since w_1 is a linear transformation of y it has a distribution in the elliptically symmetric family. Therefore, we can exploit the relationships between joint, conditional and marginal densities within that family to provide both the weight function and the resulting integrated density.

Before proceeding we need to be explicit about the two types of hypotheses to be considered; those which restrict only ρ under the null and those which also restrict part or all of β. In order to maintain consistency of notation throughout we will consider the extended regression model;

$$y_t = \rho y_{t-1} + x_t'\beta + z_t'\gamma + \varepsilon_t, \quad t = 1, 2, .., T, \tag{15}$$

where now z_t and γ are l-vectors of covariates and parameters, respectively.

We will consider testing both single and joint unit root hypotheses in the context of (15), specifically tests of

$$\begin{aligned} H_0 &: \rho = 1 \cap \gamma = 0 \quad \text{vs.} \quad H_1^S : |\rho| < 1 \cap \gamma = 0 \quad \text{and} \\ H_0 &: \rho = 1 \cap \gamma = 0 \quad \text{vs.} \; H_1^J : |\rho| < 1 \cap \gamma \neq 0, \end{aligned} \tag{16}$$

so that these tests differ only under the alternative. For example, testing (2) in (1) as in the introduction, is characterized by testing H_0 versus H_1^J with $x_t = 1$, $z_t = t$.

Given model (15) tests having known size can be characterized by the vector

$$v_1 = \frac{w_1}{|w_1|} \quad ; \quad w_1 = C_1' \Delta_1 y,$$

where the matrix C_1 is as defined above. Under the null hypothesis the distribution of v_1 is constant on $\mathbb{S}_{N-k}$. Under the alternative hypotheses, however, its distribution is not known, and will in general be different for each case. Thus, for $\tilde{X}_\rho$ and A_1 defined in (8), under H_1^S,

$$w_1 |\, H_1^S \sim \mathcal{F}(\tilde{X}_\rho \beta, \sigma^2 A_1).$$

Here, because the set of regressors is unchanged from null to alternative we shall treat the whole of $\tilde{X}_\rho \beta$ as the nuisance parameter, that is

$$\mu_S = E[w_1] = \tilde{X}_\rho \beta,$$

is the nuisance parameter present under the alternative H_1^S.

On the other hand, putting $W = (X, Z)$ and $\lambda = (\beta', \gamma')'$, then under the joint alternative,

$$w_1 |\, H_1^J \sim \mathcal{F}(\tilde{W}_\rho \lambda, \sigma^2 A_1),$$

where now $\tilde{W}_\rho = C_1' \Delta_1 \Delta_\rho^{-1} W$. In this case the regressor set changes from null to alternative, and thus not all of $\tilde{W}_\rho \delta$ is nuisance. Moreover, any 'optimal' test should depend upon this change. Although there are a number of ways to achieve this, here we will assume that here the nuisance parameter under H_1^J, to be integrated out is

$$\mu_J = \lambda,$$

that is just the parameter set. Notice that this approach is equivalent to providing a weighted average most powerful test over values of δ under the alternative, similar to Forchini (2005).

The following theorem, proved in the appendix, gives both the weight function and the weighted point optimal and locally best test for testing H_0 against each alternative.

Theorem 2 *Suppose that the data y is generated according to model (15) such that its distribution satisfies Assumption 1, and let* $q_S(.)$ *and* $q_J(.)$ *be convex non-negative functions defining particular elliptically symmetric families, then:*
(i) For testing against H_1^S*, the appropriate weight function, for* $|\rho| < 1$ *is*

$$\pi_S = \pi(\mu_S) = \left|\sigma^2 A_1\right|^{-1/2} q_S\left(\mu_S' \left(\sigma^2 A_1\right)^{-1} \mu_S\right),$$

and hence weighted point optimal and locally best tests are given by

$$\begin{aligned} \omega_{\pi_S}^{PO} &: \textit{reject } H_0 \textit{ if } \quad v_1' A_1^{-1} v_1 \leq k_1 \quad \textit{and} \\ \omega_{\pi_S}^{LB} &: \textit{reject } H_0 \textit{ if } \quad v_1' \left.\frac{\partial \left(A_1^{-1}\right)}{\partial \rho}\right|_{\rho=1} v_1 \leq k_2, \end{aligned} \tag{17}$$

where k_1 *and* k_2 *are constants chosen so that the size of each test is fixed at* α*.*
(ii) For testing against H_1^J*, the appropriate weight function, for* $|\rho| < 1$ *is*

$$\pi(\mu_J) = \left|\sigma^{-2}\left(W'W\right)\right|^{1/2} q_J\left(\sigma^{-2}\mu_J'\left(W'W\right)\mu_J\right),$$

and hence weighted point optimal and locally best tests are given by

$$\begin{aligned} \omega_{\pi_J}^{PO} &: \textit{reject } H_0 \textit{ if } \quad v_1' \left(A_1 + \tilde{P}_W\right)^{-1} v_1 \leq k_3 \quad \textit{and} \\ \omega_{\pi_J}^{LB} &: \textit{reject } H_0 \textit{ if } \quad v_1' \left.\frac{\partial \left(A_1 + \tilde{P}_W\right)^{-1}}{\partial \rho}\right|_{\rho=1} v_1 \leq k_4, \end{aligned} \tag{18}$$

where $\tilde{P}_W = \tilde{W}_\rho \left(W'W\right)^{-1} \tilde{W}_\rho'$ *and* k_3 *and* k_4 *are constants chosen so that the size of each test is fixed at* α*.* ■

Theorem 2 gives weighted optimal tests for each formulation of the unit root hypothesis as applied to a lagged dependent variable. Both the theoretical and numerical properties of the resulting tests are analyzed in the following section.

4 Analysis

4.1 Discussion

Choosing a particular prior or weight function is always open to the criticism of it being a mere contrivance. However, it is important to note the following. Together the theorems set out a clear procedure for deriving weighted optimal tests. Specifically, such tests follow immediately by applying standard techniques (such as those in King (1980) and Dufour and King (1991)) to the integrated density for w_1. Exploiting the marginalisation properties of the elliptically symmetric family provides both the appropriate weight function and hence the resultant tests. Practitioners are free to exploit the general results to derive their own weight functions and tests.

The given tests enjoy precisely the properties we would desire. Consider the density of v_1 given in (9), written as

$$pdf(v_1, \rho) = |A_1|^{-1/2} \left(v_1' A_1^{-1} v_1\right)^{-(T-k)/2} \times h(\mu_S),$$

and notice that since

$$\overline{pdf}(v_1; \rho) = |A_1|^{-1/2} \left(v' A_1^{-1} v\right)^{-(N-k)/2}, \tag{19}$$

then

$$\int_{\mu_S} h(\mu_S)\, \pi(\mu_S) d\mu_S = 1.$$

That is we have very precisely integrated out of the density that part which was unresolved. The implication is that every test satisfying some optimality criteria for M_2 has a precise (weighted) analogue for testing H_0^S in M_1. In fact the only difference between them will be that for M_2 we project the data orthogonal to the columns of the differenced regressors, while for M_1 we project orthogonal to the columns themselves.

For the tests for the joint hypothesis H_0^J, notice that

$$A_1 + \tilde{P}_W = C_1' \Delta_1 \Delta_\rho^{-1} \left(I_{N-k} + W \left(W'W\right)^{-1} W'\right) (\Delta_\rho^{-1})' \Delta_1' C_1,$$

so that any resulting test will be a function of two components. One represents the simpler case where only ρ is restricted, while the other represents a projection on the space spanned by the columns of $W = (X, Z)$. This also would seem to be a desirable property for such tests to enjoy.

In practice the assumption that the covariance of the innovations is scalar may not be warranted. However, and with some generality, it is possible to make robust the procedures described above and deliver operational testing procedures. To do so assume now that the $(\varepsilon_i)_1^T$ are a stationary ergodic process, having covariance structure,

$$E[\varepsilon\varepsilon'] = \sigma^2\Omega, \ \ |\Omega| = 1.$$

Consequently, let $\hat{\Omega}$ be any parametric or semi-parametric estimator with

$$||\Omega - \hat{\Omega}|| = o_p(1), \tag{20}$$

where $||.||$ is any matrix norm, (note that all norms are equivalent on the space of symmetric positive definite matrices). Hence, for

$$A_\Omega = C_1'\Delta_1\Delta_\rho^{-1}\Omega\left(\Delta_\rho^{-1}\right)'\Delta_1'C_1,$$

and on account of (20),

$$\left(v'A_\Omega^{-1}v - v'A_{\hat{\Omega}}^{-1}v\right) = o_p(1)$$

then asymptotically robust tests can be derived via optimality criteria applied to $v'A_{\hat{\Omega}}^{-1}v$. The need to estimate $\hat{\Omega}$ consistently restricts the class of models somewhat, generally defined weak mixing process are ruled out. However, neither the density nor the precise nature of the correlation structure in Ω need be specified. Since the (ε_i) are stationary, then via Wald's decomposition, we need only construct a consistent estimator for their transfer function, for example via the consistent augmented autoregression of Ng and Perron (2001).

The extension to the case of testing the hypotheses $H_0 : \rho = \rho_0$ is immediate. We can define

$$v_0 = \frac{w_0}{|w_0|} \quad ; \quad w_0 = C'\Delta_0 y,$$

$$C_1C_1' = M_X = I - X(X'X)^{-1}X' \quad ; \quad C_1'C_1 = I_{N-k},$$

which, following precisely the steps taken to arrive at (19), gives

$$\overline{pdf}(v_0;\rho) = |A_0|^{-1/2}\left(v_0'A_0^{-1}v_0\right)^{-(N-k)/2},$$

where $A_0 = C_1'\Delta_0\Delta_\rho^{-1}\left(\Delta_\rho^{-1}\right)'\Delta_0 C_1$.

Hillier (1987) characterizes the class of similar (and consequently under our assumptions invariant) tests for the significance of a lagged dependent variable, i.e. tests for $H_0 : \rho = 0$. However, no optimal procedures were developed. Here, we are able to characterize weighted optimal tests, as in the following corollary to Theorem 1.

Corollary 1 *Suppose that the data y is generated according to model M_1 such that its distribution satisfies Assumption 1 with $\Omega = I_T$, and suppose that we are testing the hypotheses $H_0 : \rho = 0$. Then weighted point optimal and locally best tests are given by*

$$\omega_\pi^{PO} : \textit{reject } H_0 \textit{ if } \quad \frac{y' C_1 A^{-1} C_1' y}{y' M_X y} \leq k_5 \quad \textit{and}$$

$$\omega_\pi^{LB} : \textit{reject } H_0 \textit{ if } \quad \frac{y' C_1 \left. \frac{\partial \left(A^{-1}\right)}{\partial \rho} \right|_{\rho=0} C_1' y}{y' M_X y} v_1' \leq k_6,$$

where k_5 and k_6 are chosen so that the size of each test is α, and

$$A = C_1' \Delta_\rho^{-1} \left(\Delta_\rho^{-1}\right)' C_1.$$

Importantly, both of these tests are identical to the optimal procedures derived for the same hypothesis in M_2, see for example King and Hillier (1985).

4.2 Numerical Results

All of the tests proposed in this paper take the form of quadratic forms on the surface of the unit sphere. Such forms can always be written as ratios of quadratic forms in y. As a result, the densities and distributions (under either hypothesis) are, in principle, available via a variety of numerical methods, see also DeJong, Nankervis, Savin and Whiteman (1992). Alternatively, convenient asymptotic approximations to these (as opposed to the limiting forms of the statistics themselves) are available in the form of saddlepoint approximations, as do Forchini and Marsh (2000). Given this, and also that the focus of the paper is upon power optimality, the numerical work will focus on comparing weighted optimal tests with those currently available in the literature.

We shall do so in the context of the simple model

$$y_t = \beta_1 + \beta_2 t + \rho y_{t-1} + \varepsilon_t \quad ; \quad \varepsilon_t \sim iid(0, \sigma^2).$$

In this context the literature has not bettered the original Dickey-Fuller (1979, 1981) statistics, although refinements to their procedures in more general settings are many. We will consider testing the following sets of hypotheses

$$\begin{aligned}
H_0^1 &: \rho = 1 \quad \text{vs.} \quad H_1^1 : |\rho| < 1, \\
H_0^2 &: \rho = 1 \cap \beta_2 = 0 \quad \text{vs.} \quad H_1^2 : |\rho| < 1 \cap \beta_2 = 0, \\
H_0^3 &: \rho = 1 \cap \beta_1 = \beta_2 = 0 \quad \text{vs.} \quad H_1^3 : |\rho| < 1 \cap \beta_2 = 0, \\
H_0^4 &: \rho = 1 \cap \beta_1 = \beta_2 = 0 \quad \text{vs.} \quad H_1^4 : |\rho| < 1 \quad \text{and} \\
H_0^5 &: \rho = 1 \cap \beta_2 = 0 \quad \text{vs.} \quad H_1^5 : |\rho| < 1.
\end{aligned}$$

Each of these hypotheses has associated with it a particular Dickey-Fuller test. For hypotheses H_0^1 and H_0^2 these are the pairs $\hat{\rho}_\tau - 1, \hat{\tau}_\tau$ and $\hat{\rho}_\mu - 1, \hat{\tau}_\mu$, i.e. the OLS estimator for $\rho - 1$ and its t-ratio, respectively. For hypotheses H_0^3, H_0^4 and H_0^5, these are Φ_1, Φ_2 and Φ_3, i.e. the likelihood ratio (or a monotone function of the F-ratio) test for each respective hypothesis. In addition we consider the t-tests based upon efficiently detrended data, as in Elliott, Rothenberg and Stock (1996) and Burridge and Taylor (2000); DF_{GLS}^{μ} for H_0^3 and DF_{GLS}^{τ} for H_0^4 and H_0^5, respectively.

First we compare the power of these statistics with the weighted power envelope, obtained as the power of the weighted point optimal test, at each appropriate value of ρ. For hypotheses H_0^1 and H_0^2 the point optimal tests are given in (17), while for H_0^3, H_0^4 and H_0^5 they are given in (18). All of the Monte Carlo experiments were performed according to the following specifications. Wherever the values of β_1 and β_2 are not specified, by either hypothesis, they were set equal to 0.1. For a sample size of $T = 100$ and for 20000 replications the appropriate critical value was simulated under each null hypothesis. For a variety of alternative values of ρ the rejection frequencies of each of the tests were simulated. These outcomes are reported in Tables 1 through 7 in the appendix.

Tables 1 and 2 contain a comparison between the $\hat{\rho}_\tau - 1, \hat{\tau}_\tau$ and $\hat{\rho}_\mu - 1, \hat{\tau}_\mu$ tests and the weighted envelopes, PE_τ and PE_μ, respectively. The powers of the Dickey-Fuller tests are close to their respective envelopes. Indeed, here we report no further comparisons for these first two hypotheses. The current tests, by criteria as objective as can be achieved in this context, have powers which cannot be significantly improved upon, if at all. Since these tests form the basis of the augmented Dickey-Fuller tests, or the procedures of Phillips and Perron (1988) and Ng and Perron (2001), this perhaps gives also some additional confidence in those procedures. Moreover, although not reported, it is the case that t-tests for each value of ρ between 0 and ± 1 inclusive tend to share this property.

The outcomes of experiments for the further three hypotheses stand in stark contrast. For these cases and both sample sizes the powers of the Φ_1, Φ_2 and Φ_3 tests are a small fraction of the relevant envelopes, denoted here by PE_1, PE_2 and PE_3, respectively for hypotheses H_0^3, H_0^4 and H_0^5 and reported in Tables 3,4 and 5. The Dickey-Fuller tests have particularly low powers when a trend is included in the alternative, as previous studies have reported. However, the literature has not yet provided feasible tests having significantly greater power in these circumstances. Moreover, these tests are completely outperformed by the DF_{GLS}^τ, despite this test not being designed for these alternatives.

We also simulate the power of the locally best test, $\omega_{\pi_J}^{LB}$ given in (18), and also a feasible, nearly efficient test, based on a principle similar to that employed by Elliott, Rothenberg and Stock (1996) for tests in M_2. The test is chosen so that it is asymptotically point optimal for the value at which the (weighted) asymptotic power envelope is 0.5. As in the latter paper, this point, say $c^* = T(\rho - 1)$ is approximated via Monte Carlo simulation, for a sample size of $T = 500$. On the basis of 5000 replications, the appropriate values of c^* were, to the nearest integer

$$H_0^3 : c_3^* = 7, \quad H_0^4 : c_4^* = 10, \quad H_0^5 : c_5^* = 13.$$

The powers of the resultant feasible tests, $\omega_{\pi_J}(c^*)$ are presented in Tables 3 to 5. The locally best tests perform adequately only very close the null hypothesis. On the other hand the $\omega_{\pi_J}(c^*)$ tests have powers very close to their respective envelopes over the range of alternatives and outperform both the F-tests and efficient t-tests.

A final set of experiments compare the performance of the efficient Dickey-Fuller test (DF^{τ}_{GLS}) with our feasible point optimal test ($\omega_{\pi_J}(c^*)$) in both the formulation considered here (M_1) and that for which it was designed (M_2). Moreover, we will examine the power performances as the initial condition deviates from its assumed value of 0.

First we will consider

$$y_t = \beta_1 + \beta_2 t + \rho y_{t-1} + \varepsilon_t; \quad y_0 = y^* + \beta_1, \quad y^* \neq 0. \tag{21}$$

and test the joint hypothesis given above as H_0^5. Performing the experiments as outlined above, the powers as we vary ρ and y^* are given in Tables 6a (for $\omega_{\pi_J}(c^*)$) and 6b (for (DF^{τ}_{GLS})). For small y^* the powers of DF^{τ}_{GLS} are close to those of $\omega_{\pi_J}(c^*)$, though naturally smaller. As is well known the power of the DF^{τ}_{GLS} test collapses for large y^*. This is not the case for $\omega_{\pi_J}(c^*)$. Although it is not unbiased, over a range of ρ values its power is either stable or increases slightly with y^*.

In order that these comparisons are fair these experiments were repeated in the model

$$y_t = \beta_1 + \beta_2 t + u_t \quad ; \quad u_t = \rho u_{t-1} + \varepsilon_t; \quad y_0 = y^* + \beta_1, \quad y^* \neq 0, \tag{22}$$

here testing the simple hypothesis given above as H_0^1. As should be expected for small deviations in the initial condition the DF^{τ}_{GLS} test is more powerful than $\omega_{\pi_J}(c^*)$. Once again though the power of DF^{τ}_{GLS} collapses as y^* increases while, generally, that of $\omega_{\pi_J}(c^*)$ remains stable.

5 Conclusions

This paper has demonstrated that the problem of testing on a lagged dependent variable, including the relevant unit root test, is generally characterized by the existence of a nuisance parameter, present only under the alternative. As in Andrews and Ploberger (1994) optimal tests can be defined as optimizing some weighted function of power. This is equivalent to applying standard optimality criteria to the integrated density of Berger, Liseo and Wolpert (1999). In the elliptically symmetric family this is shown to be straightforward to accomplish, with the obvious technical difficulties completely circumvented.

For the leading unit root case, where no further restrictions are imposed under the null, the resulting weighted criteria are directly analogous to those applied with much success in the alternative framework of Dufour and King (1991). Numerical evidence here shows that the t-tests of Dickey and Fuller (1979) have objectively good power properties in this context. Indeed, it turns out that generally it is difficult to improve on the t-test for any hypothesized parameter value.

On the other hand when additional restrictions are imposed, here we can provide weighted optimal tests which are much more powerful than the relevant F-type tests of Dickey and Fuller (1981). Some power superiority is also evident over the currently favoured efficient-detrended version of the Dickey-Fuller t-test. Moreover, comparisons in both the current and the Elliott, Rothenberg and Stock (1996) framework demonstrates that the power of a feasible weighted point optimal test is stable as the initial condition deviates from its hypothesized value while that of the efficient Dickey-Fuller test collapses. That is if there is uncertainty over the initial condition one might prefer the proposed test, even in M_2. For M_1 the caveat is not necessary.

Acknowledgments: Thanks are due to Francesco Bravo, Les Godfrey, Grant Hillier, Peter Phillips and participants at seminars at the Universities of Monash and York. The author has particularly benefited from detailed comments from both Giovanni Forchini and Rob Taylor.

References

Andrews, D.W.K. and W. Ploberger (1994) Optimal Tests when a Nuisance Parameter is Present only under the Alternative. *Econometrica*, **62**, 1383-1414.

Berger, J.O., B. Liseo and R.L. Wolpert (1999) Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science*, **14**, 1-28.

Burridge, P. and A.M.R. Taylor (2000) On the power of GLS-type unit root tests. *Oxford Bulletin of Economics and Statistics*, **62**, 633-645.

DeJong, D.N., J.C. Nankervis, N.E. Savin and C.H. Whiteman (1992) Integration versus trend stationarity in time series. *Econometrica,* **60**, 423-433.

Dickey, D.A. and W.A. Fuller (1979) Distribution of the Estimators for Autoregressive Series with a Unit Root. *Journal of the American Statistical Society*, **74**, 427-431.

Dickey, D.A. and W.A. Fuller (1981) Likelihood Ratio Statistics for Autoregressive Series with a Unit Root. *Econometrica*, **49**, 1057-1072.

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

Elliott, G., T.J. Rothenberg and J.H. Stock (1996) Efficient tests for an autoregressive unit root. *Econometrica*, **64**, 813-836.

Forchini, G. (2005) Optimal weighted average power similar tests for the covariance structure in the linear regression model. *Journal of Econometrics*, **124**, 253–267.

Forchini, G. and P.W. Marsh (2000) Exact inference for the unit root hypothesis. *Discussion Paper in Economics,* ***00/54,*** University of York.

Hillier, G.H. (1987) Classes of similar regions and their power properties for some econometric testing problems. *Econometric Theory*, **3**, 1-44.

Kariya, T. (1980) Locally robust tests for serial correlation in least squares regression. *Annals of Statistics*, **8**, 1065-1070.

King, M. L. (1980) Robust tests for spherical symmetry and their application to least squares regression. *Annals of Statistics*, **8**, 1265–1271.

King, M.L. and G.H. Hillier (1985) Locally best invariant tests of the error covariance matrix of the linear regression model. *Journal of the Royal Statistical Society (B),* **47**, 98-102.

Ng, S. and P. Perron (2001) Lag length selection and the construction of unit root tests with good size and power. *Econometrica,* **69**, 1519–1554.

Phillips, P.C.B. and P. Perron (1988) Testing for a unit root in time series regression. *Biometrika,* **75**, 335-346.

Schmidt, P. and P.C.B. Phillips (1992) LM Tests for a Unit Root in the Presence of Deterministic Trends. *Oxford Bulletin of Economics and Statistics,* **54**, 257-287.

Appendix

I) Proof of Theorem 1: Following Andrews and Ploberger (1994), for a given measurable, non-negative weight function $\pi(\mu)$, weighted power is given by

$$\begin{aligned} \overline{P}_\omega &= \int_\mu P_\omega \, \pi(\mu) \, d\mu \\ &= \int_\mu \left(\int_\omega pdf(v_1; \rho) \, (dv) \, \pi(\mu) \right) d\mu, \end{aligned}$$

since $pdf(v_1; \rho)$ is a density and thus non-negative, Tonelli's Theorem implies we may change the order of integration, to obtain

$$\begin{aligned} \overline{P}_\omega &= \int_\omega \left(\int_\mu pdf(v_1; \rho) \, \pi(\mu) \, d\mu \right) (dv) \\ &= \int_\omega \overline{pdf}(v_1; \rho)(dv), \end{aligned} \tag{23}$$

so that the optimality criteria applied to the integrated density immediately yield weighted optimal tests.

From (14) we define an integrated likelihood for v as

$$\begin{aligned} \widetilde{pdf}(v_1; \rho) &= \frac{\int_{a>0} \overline{pdf}(aw_1|H_1)da}{\int_{a>0} \overline{pdf}(aw_1|H_0)da} \\ &= \frac{\int_{a>0} a^{\frac{N-k}{2}-1} \int_{\mathbb{R}^{N-k}} pdf(w_1; \rho) \; \pi(\mu) \, d\mu \; da}{\int_{a>0} a^{\frac{N-k}{2}-1} \int_{\mathbb{R}^{N-k}} pdf(w_1; 1) \, \pi(\mu) \, d\mu \, da}, \end{aligned} \tag{24}$$

since, for $N - k > 2$, Tonelli's Theorem applies and so, again, we may interchange the order of integration. Consequently, noticing that

$$\int_{\mathbb{R}^{N-k}} pdf(aw_1; 1) \, \pi(\mu) d\mu = pdf(aw_1; 1) \, ,$$

which follows from

$$\int_{\mathbb{R}^{N-k}} \pi(\mu) d\mu = 1, \quad \text{for all } \beta \text{ and } \rho,$$

then

$$\begin{aligned} \widetilde{pdf}(v; \rho) &= \int_{\mathbb{R}^{N-k}} \frac{\int_0^\infty a^{\frac{N-k}{2}-1} pdf(aw_1; \rho) da}{\int_0^\infty a^{\frac{N-k}{2}-1} pdf(aw_1; 1) da} \, \pi(\mu_1) \, d\mu_1 \\ &= \int_{\mathbb{R}^{N-k}} pdf(v; \rho) \, \pi(\mu_1) \, d\mu_1 = \overline{pdf}(v; \rho), \end{aligned}$$

as required. That is for a given weight function $\pi(\mu)$ any integrated density for w immediately induces an equivalent integrated density for v. Substituting into (23) immediately gives the result. ■

II) Proof of Theorem 2:

Part (i): The integrated density for w_1 is given, for any weight function $\pi(\mu)$, by

$$\overline{pdf}(w;\rho) = \int_{\mu_1 \in \mathbb{R}^{N-k}} pdf(w;\rho)\,\pi(\mu) d\mu, \tag{25}$$

and in order that the integrated density for v_1 exists always (because $\lim_{\rho\to 1}\mu = 0$), we will suppose that at $\rho = 1$,

$$\pi(\mu) = \delta_\mu(1), \tag{26}$$

the delta function taking the value 1 if $\mu = 0$, 0 otherwise.

For $\rho \neq 1$, from Kariya (1980, Section 3) and the fact that marginal densities in the elliptically symmetric family are themselves elliptically symmetric, we have

$$w_1 \sim \mathcal{F}(\mu, \sigma^2 A_1),$$

which we will interpret as the conditional distribution of w_1 given the nuisance parameter $\mu = \mu_S = \tilde{X}_\rho \beta$.

The method of this paper is to choose the weight function so that under H_1^S, $(w_1', \mu)'$ are jointly elliptically symmetric in that

$$\begin{pmatrix} w_1 \\ \mu \end{pmatrix} \sim \mathcal{F}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} \Sigma_w & \Sigma_{w,\mu} \\ \Sigma_{w,\mu} & \Sigma_\mu \end{pmatrix}\right]. \tag{27}$$

Immediately then, there exists a weight function given by

$$\pi(\mu) = |\Sigma_\mu|^{-1/2}\, q_S\left(\mu' \Sigma_\mu^{-1} \mu\right),$$

so that marginally $\mu \sim F(0, \Sigma_\mu)$ and the condition (26) ensures that the requirement (12) is met.

To determine this weight function notice that conditionally, we have

$$w_1|\,\mu \sim \mathcal{F}\left(\mu,\, \sigma^2 A_1\right),$$

and so the matrices in (27) must therefore satisfy,

$$\begin{aligned} \Sigma_{w,\mu}\Sigma_\mu^{-1}\mu &= \mu \\ \Sigma_w - \Sigma_{w,\mu}\Sigma_\mu^{-1}\Sigma_{w,\mu} &= A_1. \end{aligned} \tag{28}$$

Since the choice of what is essentially a prior for the nuisance parameter is, and should be, somewhat user dependent, (28) has many solutions. Indeed it is not possible to identify both $\Sigma_{w,\mu}$ and Σ_{μ}^{-1} individually, however we must have that

$$\Sigma_{w,\mu}\Sigma_{\mu}^{-1}\tilde{X}_{\rho} = \tilde{X}_{\rho},$$

since neither matrix can depend on β. For H_1^S we will take the simplest solution (in particular so as to avoid use of a non-singular elliptical weight function) which has

$$\Sigma_{\mu} = \Sigma_{w,\mu} = A_1, \tag{29}$$

which determines the weight function precisely. Moreover, from (28) we have

$$\Sigma_w = 2A_1,$$

and so the integrated density for w is simply

$$\overline{pdf}(w_1;\rho) = \left|2\sigma^2 A_1\right|^{-1/2} \bar{q}\left(w_1'\left(2\sigma^2 A_1\right)^{-1} w_1\right),$$

from which the integrated density for v then immediately follows from King (1980), as

$$\widetilde{pdf}(v_1;\rho) = |A_1|^{-1/2}\left(v' A_1^{-1} v\right)^{-(N-k)/2}.$$

Applying the results of Theorem 1 and using the definitions of ω^{PO} and ω^{LP} the weighted optimal tests then follow. ■

Part (ii): Under H_1^J we interpret the distribution of w_1 as being conditional upon the nuisance parameter λ, which is given by

$$w_1|\,\lambda \sim \mathcal{F}\left(\tilde{W}_{\rho}\lambda, \sigma^2 A_1\right).$$

Notice that λ does not vanish under the null hypothesis, unlike μ above, since it is a mixture of interest and nuisance parameters, and so no restrictions on the weight function are needed.

Once again we suppose that the data w_1 and nuisance parameter λ are jointly distributed as

$$\begin{pmatrix} w_1 \\ \lambda \end{pmatrix} \sim \mathcal{F}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} \Sigma_w & \Sigma_{w,\lambda} \\ \Sigma_{w,\lambda} & \Sigma_{\lambda} \end{pmatrix}\right],$$

The weight function is, similar to before, determined by matrices which satisfy

$$\begin{aligned} \Sigma_{w,\lambda}\Sigma_{\lambda}^{-1}\tilde{W}_{\rho}\delta &= \tilde{W}_{\rho}\delta \\ \Sigma_{w} - \Sigma_{w,\lambda}\Sigma_{\lambda}^{-1}\Sigma_{w,\lambda} &= A_1. \end{aligned}$$

The straight-forward solution, again avoiding non-singular solutions, is to set

$$\begin{aligned} \Sigma_{w,\lambda} &= \tilde{W}_{\rho}(W'W)^{-1} = C_1'\Delta_1\Delta_{\rho}^{-1}W(W'W)^{-1} \\ \Sigma_{\lambda} &= (W'W)^{-1}, \end{aligned}$$

so the weight function is given as

$$\pi(\lambda) = \left|\sigma^{-2}(W'W)\right|^{1/2} q_J\left(\sigma^{-2}\lambda'(W'W)\lambda\right).$$

For the integrated density of w_1 we have that,

$$\Sigma_w = A_1 + \tilde{P}_W, \quad \tilde{P}_W = \tilde{W}_{\rho}(W'W)^{-1}\tilde{W}_{\rho}',$$

and so

$$\overline{pdf}(w_1;\rho) = \left|\sigma^2(A_1 + \tilde{P}_W)\right|^{-1/2} \bar{q}\left(w_1'\left(\sigma^2(A_1 + \tilde{P}_W)\right)^{-1} w_1\right),$$

and so from King (1980), we have

$$\overline{pdf}\,(v_1,\rho) = |A_1 + \tilde{P}_W|^{-1/2}\left(v'\left(A_1 + \tilde{P}_W\right)^{-1} v\right)^{-(N-k)/2},$$

and once again the results of Theorem 1 and the definitions for ω^{PO} and ω^{LP} yield the given weighted optimal tests. ■

III) Tables

The results presented here represent outcomes of 20000 Monte Carlo replications, both for the critical values and the rejection frequencies given here. All experiments were performed using Mathematica 4.1 on a 3.0Ghz Pentium IV PC.

Table 1: Powers of tests for H_0^1 vs. H_1^1, $T = 100$

ρ	.975	.950	.925	.900	.875	.850	.825	.800
PE_{τ}	.077	.124	.223	.316	.514	.599	.777	.851
$\hat{\tau}_{\tau}$	.073	.108	.188	.279	.432	.519	.695	.787
$\hat{\rho}_{\tau} - 1$	.081	.117	.211	.313	.488	.593	.771	.843

Table 2: Powers of tests for H_0^2 vs. H_1^2, $T = 100$

ρ	.975	.950	.925	.900	.875	.850	.825	.800
PE_μ	.077	.169	.268	.438	.583	.728	.834	.924
$\hat{\tau}_\mu$	.075	.138	.237	.396	.488	.653	.818	.883
$\hat{\rho}_\mu - 1$	.077	.155	.260	.435	.580	.719	.826	.919

Table 3: Powers of tests for H_0^3 vs. H_1^3, $T = 100$

ρ	.975	.950	.925	.900	.875	.850	.825	.800
PE_1	.098	.248	.458	.693	.845	.928	.973	.994
ϕ_1	.057	.083	.142	.233	.371	.521	.669	.795
DF_{GLS}^μ	.058	.117	.207	.352	.518	.672	.801	.889
$\omega_{\pi_J}(c^*)$	.094	.234	.451	.673	.829	.915	.971	.989
$\omega_{\pi_J}^{LB}$	.098	.144	.196	.255	.295	.369	.424	.487

Table 4: Powers of tests for H_0^4 vs. H_1^4, $T = 100$

ρ	.975	.950	.925	.900	.875	.850	.825	.800
PE_2	.055	.088	.248	.491	.706	.869	.941	.985
ϕ_2	.050	.053	.055	.096	.133	.207	.290	.416
DF_{GLS}^τ	.055	.079	.234	.375	.552	.710	.833	.919
$\omega_{\pi_J}(c^*)$	.053	.085	.237	.482	.645	.794	.901	.957
$\omega_{\pi_J}^{LB}$	.055	.069	.100	.136	.167	.204	.253	.309

Table 5: Powers of tests for H_0^5 vs. H_1^5, $T = 100$

ρ	.975	.950	.925	.900	.875	.850	.825	.800
PE_3	.058	.105	.197	.333	.438	.611	.754	.846
ϕ_3	.056	.070	.105	.147	.206	.321	.472	.571
DF_{GLS}^τ	.058	.100	.188	.288	.416	.592	.722	.837
$\omega_{\pi_J}(c^*)$	.056	.095	.191	.308	.425	.599	.737	.844
$\omega_{\pi_J}^{LB}$	.058	.076	.099	.126	.152	.182	.215	.267

Table 6a: Powers of the $\omega_{\pi_J}(c^*)$ test for H_0^6 vs. H_1^6, in (21) with different y^*, $T = 100$

	ρ							
y^*	.975	.950	.925	.900	.875	.850	.825	.800
1	.056	.104	.183	.314	.451	.610	.745	.875
2	.051	.097	.177	.292	.452	.609	.754	.856
3	.044	.089	.171	.298	.455	.617	.761	.868
4	.040	.084	.167	.297	.459	.629	.778	.883
5	.032	.072	.158	.291	.468	.642	.788	.900
6	.026	.066	.155	.292	.470	.651	.799	.905
7	.020	.060	.147	.288	.478	.664	.815	.911
8	.014	.053	.139	.293	.482	.682	.831	.924
9	.012	.045	.131	.281	.489	.692	.842	.929
10	.008	.037	.125	.283	.501	.708	.857	.941

Table 6b: Powers of the DF^{τ}_{GLS} test for H_0^6 vs. H_1^6, in (21) with different y^*, $T = 100$

	ρ							
y^*	.975	.950	.925	.900	.875	.850	.825	.800
1	.046	.089	.176	.304	.446	.615	.744	.856
2	.035	.075	.151	.271	.422	.602	.735	.849
3	.023	.056	.117	.234	.380	.552	.708	.835
4	.014	.039	.091	.182	.323	.496	.664	.800
5	.008	.022	.066	.136	.259	.423	.599	.751
6	.003	.012	.038	.093	.204	.352	.528	.694
7	.001	.006	.021	.062	.143	.274	.447	.624
8	.001	.002	.009	.036	.091	.206	.356	.545
9	.000	.001	.005	.016	.057	.138	.283	.455
10	.000	.000	.001	.008	.032	.090	.202	.361

Table 7a: Powers of the $\omega_{\pi_J}(c^*)$ test for H_0^1 vs. H_1^1, in (22) with different y^*, $T = 100$

	ρ							
y^*	.975	.950	.925	.900	.875	.850	.825	.800
1	.098	.159	.278	.444	.608	.754	.851	.908
2	.093	.157	.272	.429	.601	.762	.850	.915
3	.088	.152	.263	.416	.588	.745	.852	.913
4	.085	.156	.245	.408	.570	.717	.834	.916
5	.077	.140	.249	.408	.569	.728	.851	.920
6	.068	.128	.244	.400	.579	.735	.860	.928
7	.063	.125	.242	.412	.601	.758	.868	.946
8	.048	.116	.247	.407	.604	.768	.883	.952
9	.043	.104	.230	.419	.609	.785	.896	.958
10	.036	.099	.232	.414	.627	.800	.915	.961

Table 7b: Powers of the DF_{GLS}^{τ} test for H_0^1 vs. H_1^1, in (22) with different y^*, $T = 100$

	ρ							
y^*	.975	.950	.925	.900	.875	.850	.825	.800
1	.090	.168	.313	.505	.685	.823	.911	.960
2	.066	.148	.287	.466	.651	.799	.889	.943
3	.050	.121	.246	.417	.599	.764	.865	.932
4	.033	.085	.194	.353	.540	.716	.822	.901
5	.018	.055	.141	.265	.451	.634	.786	.883
6	.011	.035	.093	.204	.371	.553	.721	.844
7	.004	.017	.053	.141	.278	.454	.639	.787
8	.002	.009	.030	.087	.201	.365	.552	.725
9	.000	.002	.012	.049	.133	.253	.461	.636
10	.000	.000	.006	.023	.078	.175	.348	.537