Econometrica, Vol. 74, No. 3 (May, 2006), 681–714

OPTIMAL INFERENCE IN REGRESSION MODELS WITH NEARLY INTEGRATED REGRESSORS

BY MICHAEL JANSSON AND MARCELO J. MOREIRA[1]

This paper considers the problem of conducting inference on the regression coefficient in a bivariate regression model with a highly persistent regressor. Gaussian asymptotic power envelopes are obtained for a class of testing procedures that satisfy a conditionality restriction. In addition, the paper proposes testing procedures that attain these power envelopes whether or not the innovations of the regression model are normally distributed.

KEYWORDS: Nearly integrated regressors, optimal inference, specific ancillarity.

1. INTRODUCTION

THIS PAPER CONSIDERS the problem of conducting inference on the regression coefficient in a bivariate regression model with a highly persistent regressor. Several papers that studied the problem of testing regression hypotheses in the presence of nearly integrated regressors have pointed out its nonstandard nature and/or proposed asymptotically valid testing procedures.[2] On the other hand, we know of only one paper, Stock and Watson (1996), that has obtained testing procedures with demonstrable optimality properties in a regression model with nearly integrated regressors.

Stock and Watson (1996) investigated tests that maximize a weighted average (local asymptotic) power criterion among tests of a certain level. The functional form of tests obtained by maximizing a weighted average power criterion depends on the underlying weighting function, implying that no uniformly most powerful (UMP) test exists among the class of all tests that satisfy only a level restriction. It therefore seems natural to ask whether it is possible to find "reasonable" restrictions subject to which a UMP test (of a hypothesis on the regression coefficient in a bivariate regression model with a nearly integrated regressor) can be derived.

In an attempt to provide an affirmative answer to that question, the present paper develops attainable finite sample and asymptotic efficiency bounds

[1]For comments and suggestions, we are grateful to a co-editor, three referees, Laura Chioda, Guido Imbens, Benoît Perron, Jim Powell, Tom Rothenberg, Paul Ruud, Jim Stock, George Tauchen, Ed Vytlacil, Mark Watson, and seminar participants at Berkeley, Harvard/MIT, Iowa State, Montréal, Princeton, Stanford, the Aarhus Econometrics Conference at Svinkløv, the 2003 NBER Summer Institute, and the 2005 SBFSIF conference. We thank Sam Thompson for providing the MATLAB code used to implement the Campbell and Yogo (2005) testing procedure.

[2]The problems caused by the presence of nearly integrated regressors have been pointed out by Cavanagh, Elliott, and Stock (1995), Elliott (1998), Elliott and Stock (1994), Jeganathan (1997), and Stock (1997). Inference procedures that are valid in the presence of nearly integrated regressors have been proposed by Campbell and Dufour (1997), Campbell and Yogo (2005), Cavanagh, Elliott, and Stock (1995), Lanne (2002), Stock and Watson (1996), and Wright (1999, 2000).

(power envelopes) under the assumption that the latent errors of the regression model are Gaussian white noise. In addition, it is shown that even if this distributional assumption is dropped, it is possible to construct testing procedures whose local asymptotic power functions coincide with the Gaussian power envelopes.[3]

Under the assumption of normality, the model exhibits the nonstandard feature of having a minimal sufficient statistic whose distribution belongs to a curved exponential family (in the terminology of Efron (1975, 1978)). Quite remarkably, it turns out that we can remove the statistical curvature from the inference problem by conducting the analysis conditional on the values of statistics that are specific ancillary in the sense that their distribution does not depend on the parameter of interest (but only on the nuisance parameter). It is this insight that enables us to develop finite sample optimality theory and motivates our asymptotic optimality theory, the development of which uses the theory of locally asymptotically quadratic (LAQ) likelihood ratios (Jeganathan (1995)) to show that the limiting experiment associated with our regression model inherits the statistical properties of the finite sample model.

We study a model in which the error term of the equation of interest is a martingale difference sequence with respect to its lags and to current and lagged values of the nearly integrated regressor. Although somewhat restrictive, this model is of empirical relevance insofar as it captures the salient features of the predictive regression model, a popular model in empirical finance.[4] The Gaussian version of the model enjoys the additional (expositional) advantage that its finite sample statistical properties are in one-to-one correspondence with the statistical properties of the associated limiting experiment, thereby enabling us to introduce the main ideas of the paper without the use of asymptotics.

The paper proceeds as follows. Section 2 introduces the model. Sections 3 and 4 develop finite sample and asymptotic optimality theory under the assumption that the latent errors of that model are Gaussian white noise. Section 5 constructs testing procedures, which are asymptotically optimal under the assumptions of Section 4, whose asymptotic validity requires less restrictive assumptions than the efficient testing procedures derived under the assumption of normality. Section 6 reports some numerical results, whereas Section 7 offers concluding remarks. Finally, all mathematical derivations have been relegated to the Appendix.

[3]In particular, the normality assumption is shown to be least favorable in the sense that no other distribution (of the latent independent and identically distributed errors) with mean zero and the same covariance matrix gives rise to a smaller power envelope than does the Gaussian distribution.

[4]Recent papers that studied have predictive regressions include Ang and Bekaert (2005), Campbell and Yogo (2005), Ferson, Sarkissian, and Simin (2003), Lanne (2002), Lewellen (2004), Polk, Thompson, and Vuolteenaho (2005), and Torus, Valkanov, and Yan (2005). See also Stambaugh (1999) and the references therein.

2. PREDICTIVE REGRESSION MODEL

Following Cavanagh, Elliott, and Stock (1995), we consider a bivariate model in which the observed data $\{(y_t, x_t)' : 1 \le t \le T\}$ are generated by the recursive system

$$y_t = \alpha + \beta x_{t-1} + \varepsilon_t^y, \tag{1}$$

$$x_t = \mu_x + v_t^x, \quad v_t^x = \gamma v_{t-1}^x + \psi(L)\varepsilon_t^x, \tag{2}$$

where the following assumptions are made.[5]

ASSUMPTION A1: *We have* $v_0^x = 0$.

ASSUMPTION A2: *We have* $E(\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots) = 0$, $E(\varepsilon_t \varepsilon_t' | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots) = \Sigma$ *for some positive definite matrix* Σ, *and* $\sup_t E[\|\varepsilon_t\|^{2+\varrho}] < \infty$ *for some* $\varrho > 0$, *where* $\varepsilon_t = (\varepsilon_t^y, \varepsilon_t^x)'$.

ASSUMPTION A3: *We have* $\psi(L) = 1 + \sum_{i=1}^{\infty} \psi_i L^i$, *where* $\psi(1) \neq 0$ *and* $\sum_{i=1}^{\infty} i|\psi_i| < \infty$.

By design, this model captures the salient features of the predictive regression model, a popular model in empirical finance.[6] Our goal is to construct one- and two-sided tests of the null hypothesis $\beta = \beta_0$, treating α, γ, and the ψ's as unknown nuisance parameters. Regarding the nuisance parameter γ, particular attention will be given to the (empirically relevant) case where the predetermined regressor x_{t-1} in (2) is highly persistent in the sense that γ is "close" (but not necessarily equal) to unity.

The development of inference procedures proceeds in three steps. First, Section 3 develops finite sample optimality theory under the assumption that $\mu_x = 0$, $\psi(L) = 1$, and ε_t is Gaussian white noise. Then, employing the same assumptions, Section 4 develops asymptotic optimality theory under the assumption that the persistence parameter γ is modeled as local-to-unity in the sense that $\gamma = 1 + T^{-1}c$ for some fixed constant c. Finally, Section 5 proposes testing procedures that enjoy asymptotic optimality properties under the assumptions of Section 4 and are asymptotically valid under Assumptions A1–A3 and local-to-unity asymptotics.

[5]In Assumption A2 and elsewhere in the paper, $\|\cdot\|$ denotes the Euclidean norm and (in)equalities that involve conditional expectations are assumed to hold almost surely.

[6]In a predictive regression, y_t denotes a stock return in period t, x_{t-1} is a predictor observed at time $t-1$, and the hypothesis of interest is $\beta = 0$.

3. OPTIMAL INFERENCE WITH GAUSSIAN ERRORS: FINITE SAMPLE THEORY

Consider the Gaussian model

$$y_t = \alpha + \beta x_{t-1} + \varepsilon_t^y, \tag{3}$$

$$x_t = \gamma x_{t-1} + \varepsilon_t^x, \tag{4}$$

where we make the following assumptions:

ASSUMPTION A1*: *We have* $x_0 = 0$.

ASSUMPTION A2*: *We have* $\varepsilon_t = (\varepsilon_t^y, \varepsilon_t^x)' \sim \text{i.i.d.}\,\mathcal{N}(0, \Sigma)$, *where* Σ *is a known, positive definite matrix.*

If β is unrelated to α (as is assumed here), then testing problems that involve β are invariant under location transformations of the form $(y_t, x_t) \to (y_t + a, x_t)$, where $a \in \mathbb{R}$. It therefore seems reasonable to consider only tests that are invariant under location transformations of the y's. The statistic

$$M_T = (y_2 - y_1, y_3 - y_1, \ldots, y_T - y_1, x_1, x_2, \ldots, x_T)' \tag{5}$$

is a maximal invariant under this group of transformations. The log likelihood $\mathcal{L}(\cdot)$ associated with M_T admits the quadratic expansion

$$\mathcal{L}(\beta, \gamma) - \mathcal{L}(0, 0) = \beta S_\beta + \gamma S_\gamma - \frac{1}{2}\left(\beta - \frac{\sigma_{xy}}{\sigma_{xx}}\gamma\right)^2 S_{\beta\beta} - \frac{1}{2}\gamma^2 S_{\gamma\gamma}, \tag{6}$$

where $\mathcal{L}(0, 0)$ is a constant (when interpreted as a function of β and γ) and

$$S_\beta = \sigma_{yy.x}^{-1} \sum_{t=1}^{T} x_{t-1}^{\mu}\left(y_t - \frac{\sigma_{xy}}{\sigma_{xx}} x_t\right), \quad S_\gamma = \sigma_{xx}^{-1} \sum_{t=1}^{T} x_{t-1} x_t - \frac{\sigma_{xy}}{\sigma_{xx}} S_\beta,$$

$$S_{\beta\beta} = \sigma_{yy.x}^{-1} \sum_{t=1}^{T} x_{t-1}^{\mu 2}, \quad S_{\gamma\gamma} = \sigma_{xx}^{-1} \sum_{t=1}^{T} x_{t-1}^2,$$

where $x_{t-1}^{\mu} = x_{t-1} - T^{-1} \sum_{s=1}^{T} x_{s-1}$, $\sigma_{yy.x} = \sigma_{yy} - \sigma_{xx}^{-1}\sigma_{xy}^2$, and Σ has been partitioned conformably with ε_t.[7]

It follows from (6) and the factorization criterion that $S = (S_\beta, S_\gamma, S_{\beta\beta}, S_{\gamma\gamma})'$ is a sufficient statistic for the distribution of the maximal invariant. When

[7] The log likelihood function $\mathcal{L}(\cdot)$ is simply a profile log likelihood function obtained by maximizing the log likelihood function associated with the entire data vector with respect to the location parameter α. The form of $\mathcal{L}(0, 0)$ is of no importance for the statistical analysis of the model, because $\mathcal{L}(0, 0)$ drops out of all expressions that involve (log) likelihood ratios.

studying invariant tests of $H_0 : \beta = \beta_0$, we can therefore restrict attention to tests based on S. Any such test can be represented by means of a $[0,1]$-valued function $\phi(\cdot)$ such that H_0 is rejected with probability $\phi(s)$ if $S = s = (s_\beta, s_\gamma, s_{\beta\beta}, s_{\gamma\gamma})'$. The associated probability of rejecting H_0 is $E_{\beta,\gamma}\phi(S)$, where the subscript on E indicates the distribution with respect to which the expectation is taken. Our aim is to explore the extent to which it is possible to maximize $E_{\beta,\gamma}\phi(S)$ uniformly in (β, γ) subject to "reasonable" restrictions on $\phi(\cdot)$.

The distribution of S is a curved exponential family (in the terminology of Efron (1975, 1978)), the minimal sufficient statistic being of dimension four, whereas the parameter vector (β, γ) is of dimension two. (A precise statement is provided in Lemma 1(a).) As a consequence, conventional optimality theory for exponential families (e.g., Lehmann and Romano (2005)) does not apply.[8] Nevertheless, it is possible to construct tests with interesting optimality properties because it turns out that a set of restrictions motivated by the conditionality principle is sufficient to remove the statistical curvature from the problem.

Because the distribution of $(S_{\beta\beta}, S_{\gamma\gamma})$ does not depend on β, the pair $(S_{\beta\beta}, S_{\gamma\gamma})$ is a specific ancillary for β (in the terminology of Basu (1977)). In other words, $(S_{\beta\beta}, S_{\gamma\gamma})$ is a statistic that would be ancillary if the value of the nuisance parameter γ were known. A conditionality argument therefore suggests that inference on β should be based on the conditional distribution of (S_β, S_γ) given $(S_{\beta\beta}, S_{\gamma\gamma})$.[9] A remarkable property of that conditional distribution is given in part (b) of the following lemma.

LEMMA 1: *Let* $\{(y_t, x_t)'\}$ *be generated by* (3) *and* (4) *and suppose Assumptions* A1* *and* A2* *hold*.

(a) *The joint distribution of S is a curved exponential family with density*

$$f_S(s; \beta, \gamma) = K(\beta, \gamma) f_S^0(s) \times \exp\left[\beta s_\beta + \gamma s_\gamma - \frac{1}{2}\left(\beta - \frac{\sigma_{xy}}{\sigma_{xx}}\gamma\right)^2 s_{\beta\beta} - \frac{1}{2}\gamma^2 s_{\gamma\gamma}\right],$$

where $f_S^0(\cdot)$ is a density of S when $\beta = \gamma = 0$ and $K(\cdot)$ is defined by the requirement $\int_{\mathbb{R}^4} f_S(s; \beta, \gamma)\, ds = 1$.

(b) *The conditional distribution of (S_β, S_γ) given $(S_{\beta\beta}, S_{\gamma\gamma})$ is a linear exponential family with density*

$$f_{S_\beta, S_\gamma | S_{\beta\beta}, S_{\gamma\gamma}}(s_\beta, s_\gamma | s_{\beta\beta}, s_{\gamma\gamma}; \beta, \gamma) = g(\beta, \gamma | s_{\beta\beta}, s_{\gamma\gamma}) h(s_\beta, s_\gamma | s_{\beta\beta}, s_{\gamma\gamma}) \exp(\beta s_\beta + \gamma s_\gamma)$$

[8]Proofs of optimality results in linear exponential families rely on the monotone likelihood ratio property and (in testing problems with nuisance parameters) on completeness of minimal sufficient statistics. Neither property holds in the curved exponential family studied here.

[9]Coincidentally, the specific ancillary $(S_{\beta\beta}, S_{\gamma\gamma})$ turns out to equal the observed Fisher information matrix, an object whose role in connection with conditional inference has been investigated by, e.g., Efron and Hinkley (1978) and Lindsay and Li (1997) in a different context.

for some functions $g(\cdot)$ *and* $h(\cdot)$.

In view of Lemma 1(b), we can remove the curvature from the testing problem by conditioning on the specific ancillary $(S_{\beta\beta}, S_{\gamma\gamma})$. It is this property that enables us to use the classical results of Lehmann and Romano (2005) to find UMP conditionally unbiased tests for one- and two-sided testing problems concerning β.

First, consider the one-sided testing problem[10]

$$H_0: \beta = \beta_0 \quad \text{vs.} \quad H_1: \beta > \beta_0.$$

A test with test function $\phi(\cdot)$ is conditionally η-unbiased if

$$E_{\beta_0,\gamma}[\phi(S)|S_{\beta\beta}, S_{\gamma\gamma}] \le \eta \quad \forall \gamma \in \mathbb{R},$$
$$E_{\beta,\gamma}[\phi(S)|S_{\beta\beta}, S_{\gamma\gamma}] \ge \eta \quad \forall \beta > \beta_0, \gamma \in \mathbb{R}.$$

Any conditionally η-unbiased test is conditionally η-similar in the sense that

$$E_{\beta_0,\gamma}[\phi(S)|S_{\beta\beta}, S_{\gamma\gamma}] = \eta \quad \forall \gamma \in \mathbb{R}. \tag{7}$$

On the other hand, the properties of exponential families (e.g., Lehmann and Romano (2005, Theorem 2.7.1)) can be used to show that a test is UMP among conditionally η-similar tests only if it is conditionally η-unbiased. As a consequence, a test is UMP conditionally η-unbiased if and only if it is UMP among tests that satisfy (7).

Consider the test function $\phi^*_\eta(\cdot)$ given by

$$\phi^*_\eta(s) = \mathbb{1}[s_\beta > C_\eta(s_\gamma, s_{\beta\beta}, s_{\gamma\gamma})], \tag{8}$$

where $\mathbb{1}[\cdot]$ is the indicator function, the conditional critical value function $C_\eta(\cdot)$ is implicitly (and essentially uniquely) defined by the requirement[11]

$$E_{\beta_0}[\phi^*_\eta(S)|S_\gamma, S_{\beta\beta}, S_{\gamma\gamma}] = \eta, \tag{9}$$

and the subscript γ on E has been omitted in recognition of the fact that the distribution of S_β conditional on $(S_\gamma, S_{\beta\beta}, S_{\gamma\gamma})$ does not depend on γ (because $(S_\gamma, S_{\beta\beta}, S_{\gamma\gamma})$ is sufficient for γ for any fixed value of β). By construction, the test based on $\phi^*_\eta(\cdot)$ satisfies (7). In fact, it follows from Theorem 2(a) that the test associated with $\phi^*_\eta(\cdot)$ is the UMP conditionally η-unbiased test.

[10] Results for the one-sided testing problem $H_0: \beta = \beta_0$ vs. $H_1: \beta < \beta_0$ are completely analogous and are omitted to conserve space.

[11] $C_\eta(\cdot)$ is "essentially unique" in the measure-theoretic sense. Specifically, any two conditional critical value functions that satisfy (9) agree almost everywhere on the support of $(S_\gamma, S_{\beta\beta}, S_{\gamma\gamma})$.

Next, consider the two-sided testing problem

$$H_0 : \beta = \beta_0 \quad \text{vs.} \quad H_2 : \beta \neq \beta_0.$$

In this case, a test is conditionally η-unbiased if its test function $\phi(\cdot)$ satisfies

$$E_{\beta_0,\gamma}[\phi(S)|S_{\beta\beta}, S_{\gamma\gamma}] \leq \eta \quad \forall \gamma \in \mathbb{R},$$
$$E_{\beta,\gamma}[\phi(S)|S_{\beta\beta}, S_{\gamma\gamma}] \geq \eta \quad \forall \beta \neq \beta_0, \gamma \in \mathbb{R}.$$

It follows from Lemma 1(b) and the properties of exponential families (e.g., Lehmann and Romano (2005, Theorem 2.7.1)) that a level η test is conditionally η-unbiased only if its test function $\phi(\cdot)$ satisfies

$$\left.\frac{\partial}{\partial\beta} E_{\beta,\gamma}[\phi(S)|S_{\beta\beta}, S_{\gamma\gamma}]\right|_{\beta=\beta_0} = 0 \quad \forall \gamma \in \mathbb{R}.$$

In turn, this condition holds if and only if

$$E_{\beta_0,\gamma}[\phi(S)S_\beta|S_{\beta\beta}, S_{\gamma\gamma}] = \eta E_{\beta_0,\gamma}[S_\beta|S_{\beta\beta}, S_{\gamma\gamma}] \quad \forall \gamma \in \mathbb{R}. \tag{10}$$

As a consequence, the class of test functions that satisfy (7) and (10) contains all test functions associated with tests that are conditionally η-unbiased. On the other hand, it can be shown that a test is UMP among tests that satisfy (7) and (10) only if it is conditionally η-unbiased.

Theorem 2(b) shows that a test is UMP conditionally η-unbiased if its test function is given by

$$\phi_\eta^{**}(s) = \mathbb{1}[s_\beta < \underline{C}_\eta(s_\gamma, s_{\beta\beta}, s_{\gamma\gamma})] + \mathbb{1}[s_\beta > \overline{C}_\eta(s_\gamma, s_{\beta\beta}, s_{\gamma\gamma})], \tag{11}$$

where $\underline{C}_\eta(\cdot)$ and $\overline{C}_\eta(\cdot)$ are implicitly (and essentially uniquely) defined by the requirements

$$E_{\beta_0}[\phi_\eta^{**}(S)|S_\gamma, S_{\beta\beta}, S_{\gamma\gamma}] = \eta, \tag{12}$$

$$E_{\beta_0}[\phi_\eta^{**}(S)S_\beta|S_\gamma, S_{\beta\beta}, S_{\gamma\gamma}] = \eta \cdot E_{\beta_0}[S_\beta|S_\gamma, S_{\beta\beta}, S_{\gamma\gamma}]. \tag{13}$$

THEOREM 2: *Let* $\{(y_t, x_t)'\}$ *be generated by* (3) *and* (4), *and suppose Assumptions* A1* *and* A2* *hold*.

(a) *If* $\phi(\cdot)$ *satisfies* (7), *then*

$$E_{\beta,\gamma}\phi(S) \leq E_{\beta,\gamma}\phi_\eta^*(S) \quad \forall \beta \geq \beta_0, \gamma \in \mathbb{R}.$$

(b) *If* $\phi(\cdot)$ *satisfies* (7) *and* (10), *then*

$$E_{\beta,\gamma}\phi(S) \leq E_{\beta,\gamma}\phi_\eta^{**}(S) \quad \forall \beta \in \mathbb{R}, \gamma \in \mathbb{R}.$$

REMARKS: (i) In most applications, the autoregressive parameter γ can be assumed to lie in some subset Γ of $\mathbb{R}$. In such cases, the condition (7) might appear excessively strong, a more reasonable condition being

$$E_{\beta_0,\gamma}[\phi(S)|S_{\beta\beta}, S_{\gamma\gamma}] = \eta \quad \forall \gamma \in \Gamma. \tag{14}$$

Provided Γ contains an open interval, the properties of exponential families (e.g., Lemma 1(b) and Lehmann and Romano (2005, Theorem 4.3.1)) can be used to show that (14) implies that $E_{\beta_0}[\phi(S)|S_\gamma, S_{\beta\beta}, S_{\gamma\gamma}] = \eta$. It is the latter property of conditionally similar tests that is used in the proof of Theorem 2. A similar remark applies to (10). Therefore, although the optimality results of Theorem 2 obviously reflect the fact that γ is assumed to be unknown, the (implicit) assumption that γ can take on any real value is not crucial. On the other hand, although our proofs go through for any open nonempty interval Γ, our results will be more important empirically when there is substantial uncertainty about the parameter γ.

(ii) Any conditionally η-similar test is η-similar in the sense that $E_{\beta_0,\gamma}\phi(S) = \eta$ for every $\gamma \in \mathbb{R}$. It can be shown that the converse does not hold. As a consequence, the class of η-similar tests is strictly greater than the class of conditionally η-similar tests. It is an open question whether the test based on $\phi^*_\eta(\cdot)$ is UMP within the class of η-unbiased tests.

(iii) Studying a more general (but closely related) model, Stock and Watson (1996) investigated tests that maximize a weighted average (local asymptotic) power criterion. When adapted to the model under consideration here, the approach of Stock and Watson (1996) involves maximization of

$$\int E_{\beta,\gamma}\phi(S)\, dG(\beta, \gamma) \tag{15}$$

among test functions $\phi(\cdot)$ that satisfy

$$E_{\beta_0,\gamma}\phi(S) \leq \eta \quad \forall \gamma \in \Gamma, \tag{16}$$

where Γ is some subset of $\mathbb{R}$ and $G(\cdot)$ is a weighting function defined on $[\beta_0, \infty) \times \Gamma$ (in the one-sided case) or $\mathbb{R} \times \Gamma$ (in the two-sided case). The class of tests that satisfies (16) depends on Γ, but is strictly larger than the class of conditionally similar tests. On the other hand, the test that maximizes (15) subject to (16) generally depends on (Γ and) the weighting function $G(\cdot)$, implying that no UMP test exists among tests that satisfy (16). Our approach to optimality theory therefore complements the approach of Stock and Watson (1996) in the sense that we are able to arrive at a stronger conclusion (existence of a UMP test) by confining attention to a strict subset of the set of testing procedures considered in the Stock and Watson (1996) approach.

(iv) Starting from the maximal invariant M_T, we employed two dimension reduction techniques to arrive at Theorem 2. First, sufficiency reduced the problem to one involving the vector S. Then, conditioning on specific ancillaries led to a further reduction of the dimension of the data, effectively removing the variation in $(S_{\beta\beta}, S_{\gamma\gamma})$ from the problem. Reducing by sufficiency before conditioning on (specific) ancillaries is consistent with the recommendations of Lehmann and Romano (2005, Chapter 10). Nevertheless, it might be tempting to attempt to condition on specific ancillaries before reducing by sufficiency. However, it can be shown that β is not identified from the distribution of the maximal invariant M_T given the specific ancillary $(x_1, \dots, x_T)'$. As a consequence, our model provides an illustration of the point that "it is desirable to reduce the data as far as possible through sufficiency, before attempting further reduction by means of (specific) ancillary statistics" (Lehmann and Romano (2005, Example 10.2.1)).

(v) The methodology developed herein can also be employed to develop a point estimator (of β) with explicit optimality properties. For details, see Eliasz (2004), who uses Lemma 1(b) and a result of Pfanzagl (1979) to obtain an optimal conditionally median unbiased estimator of β.

(vi) The results of this section extend readily to models with multiple regressors. This is so because the property that the statistical curvature can be removed from testing problems concerning β by conditioning on specific ancillaries is shared by models with multiple regressors. To be specific, suppose

$$y_t = \alpha + \beta' x_{t-1} + \varepsilon_t^y,$$
$$x_t = \gamma x_{t-1} + \varepsilon_t^x,$$

where the x's are multivariate, $v_0^x = 0$, and $\varepsilon_t = (\varepsilon_t^y, \varepsilon_t^{x\prime})' \sim \text{i.i.d.}\,\mathcal{N}(0, \Sigma)$, where Σ is a known, positive definite matrix. As in the scalar case, testing problems that involve β are invariant under location transformations of y's and the log likelihood $\mathcal{L}(\cdot)$ associated with the maximal invariant $(y_2 - y_1, y_3 - y_1, \dots, y_T - y_1, x_1', x_2', \dots, x_T')'$ admits a quadratic expansion

$$\mathcal{L}(\beta, \gamma) - \mathcal{L}(0, 0)$$
$$= \beta' S_\beta + \text{vec}(\gamma' \Sigma_{xx}^{-1})' S_\gamma + \delta_{\beta\beta}(\beta, \gamma)' S_{\beta\beta} + \delta_{\gamma\gamma}(\beta, \gamma)' S_{\gamma\gamma},$$

where $\mathcal{L}(0, 0)$ is a constant, $\delta_{\beta\beta}(\cdot)$ and $\delta_{\gamma\gamma}(\cdot)$ are some functions, and

$$S_\beta = \sigma_{yy.x}^{-1} \sum_{t=1}^{T} x_{t-1}^{\mu} (y_t - \sigma_{xy}' \Sigma_{xx}^{-1} x_t),$$

$$S_\gamma = \text{vec}\left(\sum_{t=1}^{T} x_{t-1} x_t' \right) - \text{vec}(S_\beta \sigma_{xy}'),$$

$$S_{\beta\beta} = \text{vech}\left(\sum_{t=1}^{T} x_{t-1}^{\mu} x_{t-1}^{\mu\prime} \right), \quad S_{\gamma\gamma} = \text{vech}\left(\sum_{t=1}^{T} x_{t-1} x_{t-1}' \right).$$

The quadratic terms in the expansion depend on the data only through the specific ancillaries $S_{\beta\beta}$ and $S_{\gamma\gamma}$. The curvature therefore disappears, thereby making the model amenable to analysis along the lines of Lehmann and Romano (2005, Chapter 4), once we condition on the specific ancillaries of the model. This feature is extremely attractive in the multivariate case, because it makes it straightforward to conduct inference on subsets of β. For specificity, consider the problem of testing

$$H_0 : \beta_1 = \beta_{1,0} \quad \text{vs.} \quad H_0 : \beta_1 > \beta_{1,0},$$

where β_1 is the first element of β. Among tests that are η-unbiased conditional on $(S_{\beta\beta}, S_{\gamma\gamma})$, it follows exactly as in Theorem 2(a) that the UMP test has test function $\phi^*_\eta(\cdot)$ given by

$$\phi^*_\eta(s) = \mathbb{1}[s_{\beta,1} > C_\eta(s_{\beta,2}, s_\gamma, s_{\beta\beta}, s_{\gamma\gamma})],$$

where $C_\eta(\cdot)$ is implicitly (and essentially uniquely) defined by the requirement

$$E_{\beta_{1,0}}[\phi^*_\eta(S) | S_{\beta,2}, S_\gamma, S_{\beta\beta}, S_{\gamma\gamma}] = \eta,$$

the statistic $S_\beta = (S_{\beta,1}, S'_{\beta,2})'$ has been partitioned after the first row, and notation recognizes the fact that the distribution of $S_{\beta,1}$ conditional on $(S_{\beta,2}, S_\gamma, S_{\beta\beta}, S_{\gamma\gamma})$ depends only on β_1.

4. OPTIMAL INFERENCE WITH GAUSSIAN ERRORS: ASYMPTOTIC THEORY

This section develops an asymptotic counterpart to Theorem 2. Whereas the finite sample results of the previous section require only mild assumptions about the range of possible values of the persistence parameter γ (cf. remark (i) following Theorem 2), the asymptotic properties of our model depend crucially on the assumptions made with respect to γ. When γ is bounded away from unity in absolute value, the curvature of the model vanishes asymptotically and standard large-sample optimality theory based on the theory of locally asymptotically normal (LAN) likelihood ratios (e.g., Choi, Hall, and Schick (1996)) is applicable. In particular, one-sided testing problems admit asymptotically UMP tests and two-sided testing problems admit asymptotically UMP unbiased tests. In contrast, Jeganathan (1997) has shown that the statistical curvature persists asymptotically when γ is modeled as local-to-unity in the sense that $\gamma = \gamma_T(c) = 1 + T^{-1}c$ for some fixed, unknown constant c.[12]

Because the statistical curvature does not vanish when $\gamma = \gamma_T(c)$, testing problems concerning β exhibit nonstandard large-sample properties under

[12]When $\gamma = \gamma_T(c) = 1 + T^{-1}c$ for some known constant c (e.g., when the unit root hypothesis $\gamma = 1$ is known to hold), the curvature also persists, but the situation is much simpler because the likelihood ratios are locally asymptotically mixed normal (LAMN) and the conditional optimality results of Feigin (1986) are applicable.

local-to-unity asymptotics. For instance, the t-test testing $\beta = \beta_0$ in (1) is not asymptotically pivotal under local-to-unity asymptotics (e.g., Cavanagh, Elliott, and Stock (1995), Elliott and Stock (1994)). Moreover, testing procedures developed under the assumption that $\gamma = 1$ are not robust to local departures from that assumption (e.g., Stock (1997)).[13] Procedures that are asymptotically valid when γ is local-to-unity have been proposed by Campbell and Dufour (1997), Campbell and Yogo (2005), Cavanagh, Elliott, and Stock (1995), and Lanne (2002), but all of these existing testing procedures are asymptotically biased.[14] In particular, these procedures are known to have power less than size for certain values of β close to its null value. By developing an asymptotic counterpart to Theorem 2, this section demonstrates by example that (nontrivial) asymptotically unbiased testing procedures can be constructed even when γ is local-to-unity.

Under the local-to-unity parameterization of γ, an appropriate parameterization of β is $\beta = \beta_T(b) = \beta_0 + T^{-1}\sigma_{xx}^{-1/2}\sigma_{yy.x}^{1/2}b$, where b is a fixed constant. In other words, $\beta_T(b) - \beta_0 = T^{-1}\sigma_{xx}^{-1/2}\sigma_{yy.x}^{1/2}b$, where the rate T^{-1} ensures contiguity of the associated probability measures (e.g., Jeganathan (1997)) and the scaling by $\sigma_{xx}^{-1/2}\sigma_{yy.x}^{1/2}$ gives rise to expressions that depend on the parameter b in a simple way. Expressed in terms of b, the null hypothesis is $b = 0$, while the one- and two-sided alternatives are $b > 0$ and $b \neq 0$, respectively.

Expanding $\mathcal{L}(\cdot)$ around $(\beta, \gamma) = (\beta_0, 1) = [\beta_T(0), \gamma_T(0)]$, we have

$$\begin{aligned} (17) \qquad & \mathcal{L}_T(b, c) - \mathcal{L}_T(0, 0) \\ & = bR_\beta + cR_\gamma - \frac{1}{2}\left(b - \frac{\rho}{\sqrt{1-\rho^2}}c\right)^2 R_{\beta\beta} - \frac{1}{2}c^2 R_{\gamma\gamma}, \end{aligned}$$

where $\mathcal{L}_T(b, c) = \mathcal{L}[\beta_T(b), \gamma_T(c)]$ and

$$R_\beta = \sigma_{xx}^{-1/2}\sigma_{yy.x}^{-1/2}T^{-1}\sum_{t=1}^{T} x_{t-1}^{\mu}(y_t - \beta_0 x_{t-1} - \sigma_{xx}^{-1}\sigma_{xy}\Delta x_t),$$

$$R_\gamma = \sigma_{xx}^{-1}T^{-1}\sum_{t=1}^{T} x_{t-1}\Delta x_t - \frac{\rho}{\sqrt{1-\rho^2}}R_\beta, \quad \rho = \sigma_{xy}\sigma_{xx}^{-1/2}\sigma_{yy}^{-1/2},$$

[13]Because tests of the unit root hypothesis $\gamma = 1$ are inconsistent against local-to-unity alternatives (e.g., Elliott, Rothenberg, and Stock (1996), Stock (1994)), this nonrobustness result can also be used to establish the invalidity of two-step procedures based on unit root pretests (e.g., Stock and Watson (1996)).

[14]The tests proposed by Campbell and Dufour (1997), Campbell and Yogo (2005), and Cavanagh, Elliott, and Stock (1995), respectively, are asymptotically biased because they are not asymptotically similar. In spite of being asymptotically similar, Lanne's (2002) test is also asymptotically biased (Wright (2000)).

$$R_{\beta\beta} = \sigma_{xx}^{-1} T^{-2} \sum_{t=1}^{T} x_{t-1}^{\mu 2}, \quad R_{\gamma\gamma} = \sigma_{xx}^{-1} T^{-2} \sum_{t=1}^{T} x_{t-1}^{2}.$$

As is S, the statistic $R = (R_\beta, R_\gamma, R_{\beta\beta}, R_{\gamma\gamma})'$ is minimal sufficient. When developing asymptotic counterparts of the results in Section 3, it turns out to be convenient to work with R. The following lemmas give some useful properties of its limiting distribution.

LEMMA 3: *Let* $\{(y_t, x_t)'\}$ *be generated by* (3) *and* (4), *and suppose Assumptions* A1* *and* A2* *hold. If* $b = T(\beta - \beta_0)\sigma_{xx}^{1/2}\sigma_{yy.x}^{-1/2}$ *and* $c = T(\gamma - 1)$ *are fixed as* T *increases without bound, then*

$$R \to_d \mathcal{R}^\rho(b, c) = \big(\mathcal{R}_\beta^\rho(b, c), \mathcal{R}_\gamma^\rho(b, c), \mathcal{R}_{\beta\beta}(c), \mathcal{R}_{\gamma\gamma}(c)\big)'$$

as $T \to \infty$, *where*

$$\mathcal{R}_\beta^\rho(b, c) = \int_0^1 W_{x,c}^\mu(r)\, dW_y(r) + \left(b - \frac{\rho}{\sqrt{1-\rho^2}} c\right) \int_0^1 W_{x,c}^\mu(r)^2\, dr,$$

$$\mathcal{R}_\gamma^\rho(b, c) = \int_0^1 W_{x,c}(r)\, dW_{x,c}(r) - \frac{\rho}{\sqrt{1-\rho^2}} \mathcal{R}_\beta^\rho(b, c),$$

$$\mathcal{R}_{\beta\beta}(c) = \int_0^1 W_{x,c}^\mu(r)^2\, dr, \quad \mathcal{R}_{\gamma\gamma}(c) = \int_0^1 W_{x,c}(r)^2\, dr,$$

W_x *and* W_y *are independent Wiener processes*, $W_{x,c}^\mu(r) = W_{x,c}(r) - \int_0^1 W_{x,c}(s)\, ds$, *and* $W_{x,c}$ *is an Ornstein–Uhlenbeck process that satisfies the stochastic differential equation* $dW_{x,c}(r) = cW_{x,c}(r)\, dr + dW_x(r)$ *with initial condition* $W_{x,c}(0) = 0$.[15]

LEMMA 4: *Let* $\mathcal{R}^\rho(b, c)$ *be defined as in Lemma* 3.

(a) *The joint distribution of* $\mathcal{R}^\rho(b, c)$ *is a curved exponential family with density*

$$f_{\mathcal{R}}^\rho(r; b, c) = K^\rho(b, c) f_{\mathcal{R}}^{\rho,0}(r) \times \exp\left[b r_\beta + c r_\gamma - \frac{1}{2}\left(b - \frac{\rho}{\sqrt{1-\rho^2}} c\right)^2 r_{\beta\beta} - \frac{1}{2} c^2 r_{\gamma\gamma}\right],$$

where $r = (r_\beta, r_\gamma, r_{\beta\beta}, r_{\gamma\gamma})'$, $f_{\mathcal{R}}^{\rho,0}(\cdot)$ *is a density of* $\mathcal{R}^\rho(0, 0)$, *and* $K^\rho(\cdot)$ *is defined by the requirement* $\int_{\mathbb{R}^4} f_{\mathcal{R}}^\rho(r; b, c)\, dr = 1$.

[15]The Wiener processes $W_x(\cdot)$ and $W_y(\cdot)$ are the weak limits of $\sigma_{xx}^{-1/2} T^{-1/2} \sum_{t=1}^{\lfloor T\cdot \rfloor} \varepsilon_t^x$ and $\sigma_{yy.x}^{-1/2} T^{-1/2} \sum_{t=1}^{\lfloor T\cdot \rfloor} (\varepsilon_t^y - \sigma_{xy}\sigma_{xx}^{-1}\varepsilon_t^x)$, respectively.

(b) *The conditional distribution of* $(\mathcal{R}^{\rho}_{\beta}(b,c), \mathcal{R}^{\rho}_{\gamma}(b,c))$ *given* $(\mathcal{R}_{\beta\beta}(c), \mathcal{R}_{\gamma\gamma}(c))$ *is a linear exponential family with density*

$$\begin{aligned} &f^{\rho}_{\mathcal{R}^{\rho}_{\beta},\mathcal{R}^{\rho}_{\gamma}|\mathcal{R}_{\beta\beta},\mathcal{R}_{\gamma\gamma}}(r_\beta, r_\gamma | r_{\beta\beta}, r_{\gamma\gamma}; b, c) \\ &\quad = g^{\rho}(b, c | r_{\beta\beta}, r_{\gamma\gamma}) h^{\rho}(r_\beta, r_\gamma | r_{\beta\beta}, r_{\gamma\gamma}) \exp(br_\beta + cr_\gamma) \end{aligned}$$

for some functions $g^{\rho}(\cdot)$ *and* $h^{\rho}(\cdot)$.

The characterizations of the limiting distribution of R given in Lemmas 3 and 4 serve complementary purposes. Lemma 4, which is based on the theory of LAQ likelihood ratios (Jeganathan (1995), Le Cam and Yang (2000)), forms the basis of the development of asymptotic counterparts of the results of the previous section. In particular, Lemma 4 (an asymptotic counterpart of Lemma 1) enables us to characterize one- and two-sided tests with asymptotic optimality properties. These characterizations, given in Theorem 5, are abstract in the sense that they involve the density $f^{\rho,0}_{\mathcal{R}}(\cdot)$ for which no closed form expression appears to be known. To help make the asymptotically optimal tests operational, Theorem 7 of Section 5 uses Lemma 3 and a result from Abadir and Larsson (2001) to obtain an integral representation of $f^{\rho,0}_{\mathcal{R}}(\cdot)$ that is useful for computational purposes.

In view of Lemma 4, the functional $\mathcal{R}^{\rho}(b,c)$ inherits those distributional properties of S that were exploited in the development of the finite sample optimality results of Section 3. By implication, the limiting experiment associated with the sequence of models under study here has the same basic structure as the finite sample experiments studied in Section 3. Specifically, the log likelihood ratios associated with the limiting experiment are quadratic; that is, the log likelihood ratios are LAQ in the sense of Jeganathan (1995). Moreover, the quadratic terms $\mathcal{R}_{\beta\beta}(c)$ and $\mathcal{R}_{\gamma\gamma}(c)$ are specific ancillaries in the limiting experiment. It therefore seems plausible that appropriately constructed asymptotic counterparts of $\phi^{*}_{\eta}(\cdot)$ and $\phi^{**}_{\eta}(\cdot)$ should enjoy asymptotic optimality properties analogous to the finite sample optimality properties enjoyed by $\phi^{*}_{\eta}(\cdot)$ and $\phi^{**}_{\eta}(\cdot)$. Theorem 5, the main result of the paper, verifies this conjecture.

Corresponding to any invariant test of $H_0 : b = 0$ based on R, there is a $[0,1]$-valued function $\pi(\cdot)$ such that the probability of rejecting H_0 equals $\pi(r)$ whenever $R = r$. This test function satisfies $\phi = \pi \circ \zeta$, where $\phi(\cdot)$ is the test function associated with S and $\zeta(\cdot)$ is any mapping such that $\zeta(S) = R$ (with probability 1).

Asymptotic optimality results for the one-sided testing problem

$$H_0 : b = 0 \quad \text{vs.} \quad H_1 : b > 0$$

can be obtained by restricting attention to test functions that satisfy an asymptotic conditional similarity condition. Our formulation of an asymptotic counterpart of the conditional η-similarity condition (7) is motivated by the fact that $\pi \circ \zeta$ satisfies (7) if and only if

$$E_{\beta_T(0),\gamma_T(c)}\big[(\pi(R)-\eta)g(R_{\beta\beta},R_{\gamma\gamma})\big]=0 \quad \forall c\in\mathbb{R}, g\in C_b(\mathbb{R}^2), \tag{18}$$

where $C_b(\mathbb{R}^2)$ denotes the set of bounded, continuous, real-valued functions on $\mathbb{R}^2$. The advantage of this characterization of conditional η-similarity is that it does not involve conditional distributions, implying that difficulties associated with conditional weak convergence (e.g., Sweeting (1989)) can be avoided by basing the formulation of an asymptotic conditional η-similarity condition on an asymptotic version of (18). Following Feigin (1986), who attributes the approach to Le Cam, we say that a sequence of tests with associated test functions $\{\pi_T(\cdot)\}$ is locally asymptotically conditionally η-similar if

$$\lim_{T\to\infty} E_{\beta_T(0),\gamma_T(c)}\big[(\pi_T(R)-\eta)g(R_{\beta\beta},R_{\gamma\gamma})\big]=0 \quad \forall c\in\mathbb{R}, g\in C_b(\mathbb{R}^2). \tag{19}$$

In perfect analogy with Theorem 2(a), Theorem 5(a) shows that a one-sided test of $b=0$ has maximal local asymptotic power among locally asymptotically conditionally similar tests if its testing function is given by

$$\pi_\eta^*(r;\rho)=\mathbb{1}[r_\beta>\mathcal{C}_\eta(r_\gamma,r_{\beta\beta},r_{\gamma\gamma};\rho)], \tag{20}$$

where $\mathcal{C}_\eta(\cdot)$ is the (unique) continuous function that satisfies[16,17]

$$E[\pi_\eta^*(\mathcal{R}^\rho;\rho)|\mathcal{R}_\gamma^\rho,\mathcal{R}_{\beta\beta},\mathcal{R}_{\gamma\gamma}]=\eta \tag{21}$$

and $\mathcal{R}^\rho=(\mathcal{R}_\beta,\mathcal{R}_\gamma^\rho,\mathcal{R}_{\beta\beta},\mathcal{R}_{\gamma\gamma})'=\mathcal{R}^\rho(0,0)$.

An attainable efficiency bound for the two-sided testing problem

$$H_0: b=0 \quad \text{vs.} \quad H_2: b\neq 0,$$

is available for the class of testing functions $\{\pi_T(\cdot)\}$ that satisfies (19) and the following asymptotic counterpart of (10):

$$\lim_{T\to\infty} E_{\beta_T(0),\gamma_T(c)}\big[(\pi_T(R)-\eta)R_\beta\cdot g(R_{\beta\beta},R_{\gamma\gamma})\big] = 0 \quad \forall c\in\mathbb{R}, g\in C_b(\mathbb{R}^2). \tag{22}$$

Indeed, it is shown in Theorem 5(b) that

$$\pi_\eta^{**}(r;\rho)=\mathbb{1}[r_\beta<\underline{\mathcal{C}}_\eta(r_\gamma,r_{\beta\beta},r_{\gamma\gamma};\rho)]+\mathbb{1}[r_\beta>\overline{\mathcal{C}}_\eta(r_\gamma,r_{\beta\beta},r_{\gamma\gamma};\rho)] \tag{23}$$

[16]The existence of the continuous function $\mathcal{C}_\eta(\cdot)$ (and the continuous functions $\underline{\mathcal{C}}_\eta(\cdot)$ and $\overline{\mathcal{C}}_\eta(\cdot)$ appearing in the definition of $\pi_\eta^{**}(\cdot)$) is established in Lemma 8 of the Appendix. The domain of $\mathcal{C}_\eta(\cdot)$ is a set $\mathbb{S}\subseteq\mathbb{R}^4$ that satisfies $\Pr[(\mathcal{R}_\gamma^\rho,\mathcal{R}_{\beta\beta},\mathcal{R}_{\gamma\gamma};\rho)\in\mathbb{S}]=1$.

[17]We have omitted the superscript ρ from the first element of $\mathcal{R}^\rho$ in recognition of the fact that $\mathcal{R}_\beta^\rho(b,c)$ does not depend on ρ when $c=0$.

is optimal among test functions that satisfy (19) and (22), where $\underline{\mathcal{C}}_\eta(\cdot)$ and $\overline{\mathcal{C}}_\eta(\cdot)$ are the (unique) continuous functions that satisfy

$$E[\pi_\eta^{**}(\mathcal{R}^\rho;\rho)|\mathcal{R}_\gamma^\rho,\mathcal{R}_{\beta\beta},\mathcal{R}_{\gamma\gamma}]=\eta, \tag{24}$$

$$E[\pi_\eta^{**}(\mathcal{R}^\rho;\rho)\cdot\mathcal{R}_\beta|\mathcal{R}_\gamma^\rho,\mathcal{R}_{\beta\beta},\mathcal{R}_{\gamma\gamma}]=\eta\cdot E[\mathcal{R}_\beta|\mathcal{R}_\gamma^\rho,\mathcal{R}_{\beta\beta},\mathcal{R}_{\gamma\gamma}]. \tag{25}$$

THEOREM 5: *Let* $\{(y_t,x_t)'\}$ *be generated by* (3) *and* (4), *and suppose Assumptions* A1* *and* A2* *hold*.

(a) *If* $\{\pi_T(\cdot)\}$ *satisfies* (19), *then*

$$\overline{\lim_{T\to\infty}}\, E_{\beta_T(b),\gamma_T(c)}\pi_T(R)\le\lim_{T\to\infty}E_{\beta_T(b),\gamma_T(c)}\pi_\eta^*(R;\rho)$$
$$=E\big[\pi_\eta^*(\mathcal{R}^\rho(b,c);\rho)\big]\quad\forall b\ge 0,c\in\mathbb{R}.$$

(b) *If* $\{\pi_T(\cdot)\}$ *satisfies* (19) *and* (22), *then*

$$\overline{\lim_{T\to\infty}}\, E_{\beta_T(b),\gamma_T(c)}\pi_T(R)\le\lim_{T\to\infty}E_{\beta_T(b),\gamma_T(c)}\pi_\eta^{**}(R;\rho)$$
$$=E\big[\pi_\eta^{**}(\mathcal{R}^\rho(b,c);\rho)\big]\quad\forall b\in\mathbb{R},c\in\mathbb{R}.$$

In view of Theorem 5, the maximal attainable (by tests that satisfy the restrictions we impose) local asymptotic power against the local alternative $\beta=\beta_T(b)$ depends on c and ρ, the persistence and correlation parameters.

Let $\varphi_\eta^*(\cdot)$ and $\varphi_\eta^{**}(\cdot)$ denote the asymptotic Gaussian power envelopes for one- and two-sided size η tests characterized in Theorem 5; that is, let

$$\varphi_\eta^*(b,c;\rho)=\Pr\big[\mathcal{R}_\beta^\rho(b,c)>\mathcal{C}_\eta\big(\mathcal{R}_\gamma^\rho(b,c),\mathcal{R}_{\beta\beta}(c),\mathcal{R}_{\gamma\gamma}(c);\rho\big)\big],$$
$$\varphi_\eta^{**}(b,c;\rho)=\Pr\big[\mathcal{R}_\beta^\rho(b,c)<\underline{\mathcal{C}}_\eta\big(\mathcal{R}_\gamma^\rho(b,c),\mathcal{R}_{\beta\beta}(c),\mathcal{R}_{\gamma\gamma}(c);\rho\big)\big]$$
$$+\Pr\big[\mathcal{R}_\beta^\rho(b,c)>\overline{\mathcal{C}}_\eta\big(\mathcal{R}_\gamma^\rho(b,c),\mathcal{R}_{\beta\beta}(c),\mathcal{R}_{\gamma\gamma}(c);\rho\big)\big].$$

The next section proposes one- and two-sided test functions that attain $\varphi_\eta^*(\cdot)$ and $\varphi_\eta^{**}(\cdot)$, respectively, under more general assumptions than those of Theorem 5.

REMARKS: (i) It is easy to show that Theorem 5 remains valid if Assumption A1* is replaced by the weaker assumption that $T^{-1/2}x_1=o_{p_0}(1)$, where $o_{p_0}(1)$ is shorthand for "$o_p(1)$ when $(\beta,\gamma)=(\beta_0,1)$." On the other hand, it is an almost immediate consequence of the results of Elliott (1999) and Müller and Elliott (2003) that Theorem 5 can fail to hold when $T^{-1/2}x_1$ has a limiting representation (under local-to-unity asymptotics) that depends on c in a nontrivial way. A more interesting question is whether the methodology developed in this paper can be used to obtain results analogous to Theorem 5 even if

Assumption A1* is replaced by an assumption of the Elliott (1999) and Müller and Elliott (2003) variety. Derivations available from the authors upon request show that this is in fact the case. Indeed, under weak assumptions on the initial condition, the limiting experiment associated with the maximal (location) invariant statistic is a curved exponential model, which can be "linearized" by conditioning on specific ancillaries.

(ii) It would be of interest to develop asymptotic power envelopes under weaker assumptions on the errors than those of Theorem 5. Two complementary generalizations of Assumption A2* seem particularly interesting. First, it would be of interest to accommodate serial correlation by studying the case where the errors are generated by a stationary Gaussian process. Adapting the methods of Jeganathan (1997, Section 3) to the present setup, it should be possible to show that the power envelopes for models with "smoothly" parameterized stationary Gaussian error processes are of the form $\varphi_\eta^{**}(\cdot;\rho)$ and $\varphi_\eta^{**}(\cdot;\rho)$, respectively, where ρ is the long-run (i.e., zero frequency) correlation of the errors. A second interesting generalization would retain the independent and identically distributed assumption on $(\varepsilon_t^y, \varepsilon_t^x)'$, but treat the error distribution as an unknown (infinite dimensional) nuisance parameter. It seems plausible that (semiparametric) power envelopes for a model of this kind can be obtained by employing methods similar to those of Jansson (2005). Semiparametric power envelopes obtained in this fashion can be no lower than $\varphi_\eta^{**}(\cdot;\rho)$ and $\varphi_\eta^{**}(\cdot;\rho)$, because it follows from Theorem 6 of the next section that $\varphi_\eta^{**}(\cdot;\rho)$ and $\varphi_\eta^{**}(\cdot;\rho)$ are attainable (with ρ being the correlation of the errors) even if the errors are non-Gaussian.

(iii) Theorem 5(a) remains true if the requirement (19) is replaced with the condition

$$\lim_{T\to\infty} E_{\beta_T(0),\gamma_T(c)}\big[(\pi_T(R)-\eta)g(R_{\beta\beta},R_{\gamma\gamma})\big]=0 \quad \forall c\in C, g\in C_b(\mathbb{R}^2),$$

where $C\subseteq\mathbb{R}$ contains an open interval. (A similar remark applies to Theorem 5(b).) The proof of this assertion is identical to the proof of Theorem 5(a) because it follows from the properties of exponential families (e.g., Lemma 4(b) and Lehmann and Romano (2005, Theorem 4.3.1)) that if $C\subseteq\mathbb{R}$ contains an open interval, then the class $\Pi(\eta,\rho)$ defined in the proof of Theorem 5(a) coincides with the class of all functions $\pi(\cdot)$ that satisfy

$$E\big[(\pi(\mathcal{R}^\rho)-\eta)g(\mathcal{R}_{\beta\beta},\mathcal{R}_{\gamma\gamma})\Lambda^\rho(0,c)\big]=0 \quad \forall c\in C, g\in C_b(\mathbb{R}^2),$$

where $\Lambda^\rho(\cdot)$ is defined as in the proof of Theorem 5(a).

(iv) In view of remark (iii), the function $\varphi_\eta^{*}(\cdot)$ constitutes a suitable power envelope also if c is treated as an unknown, nonpositive nuisance parameter—a plausible assumption in most empirical applications. On the other hand, the local asymptotic conditional similarity condition (19) would be unnecessarily restrictive if a consistent estimator of c were available. No such estimator exists under the assumptions of our model, but consistent estimation of c is feasible if c is treated as a known (continuous) function of β (e.g., Valkanov (1999)). (Consistent estimators of c are also available in certain panel versions of our model (e.g., Moon and Phillips (2000, 2004)).)

5. INFERENCE IN THE GENERAL CASE

This section considers the general case where $\{(y_t, x_t)'\}$ is generated by (1) and (2), Assumptions A1 and A3 hold, and local-to-unity asymptotics are employed. Our aim is to construct test functions with desirable large-sample properties. Specifically, we wish to develop test functions that do not require knowledge of any nuisance parameters, are asymptotically equivalent to $\pi_\eta^*(R;\rho)$ and $\pi_\eta^{**}(R;\rho)$ under the assumptions of Theorem 5, and have local asymptotic power functions of the form $\varphi_\eta^*(\cdot;\rho)$ and $\varphi_\eta^{**}(\cdot;\rho)$ more generally (i.e., under Assumptions A1–A3 and local-to-unity asymptotics). This will be accomplished by constructing a statistic $\hat{R}$, which is asymptotically equivalent to R under the assumptions of the previous section and has a limiting representation of the form $\mathcal{R}^\rho(b,c)$ more generally.

Let $x_0 = x_1$ and $\hat{v}_0^x = 0$, and define $\hat{v}_t^x = x_t - x_1$ (for $t = 1, \ldots, T$). Let $\hat{\Omega}$ be a consistent estimator of

$$\Omega = \begin{pmatrix} \omega_{yy} & \omega_{yx} \\ \omega_{xy} & \omega_{xx} \end{pmatrix} = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T}\sum_{s=1}^{T} E\left[\begin{pmatrix} \varepsilon_t^y \\ \psi(L)\varepsilon_t^x \end{pmatrix}\begin{pmatrix} \varepsilon_s^y \\ \psi(L)\varepsilon_s^x \end{pmatrix}'\right],$$

the long-run variance of $(\varepsilon_t^y, \psi(L)\varepsilon_t^x)'$. Finally, let

$$\hat{R}_\beta = \hat{\omega}_{yy.x}^{-1/2}\hat{\omega}_{xx}^{-1/2} T^{-1}\sum_{t=1}^{T} x_{t-1}^{\mu}(y_t - \beta_0 x_{t-1}) - \frac{\hat{\rho}}{\sqrt{1-\hat{\rho}^2}}\left[\frac{1}{2}(\hat{\omega}_{xx}^{-1}T^{-1}\hat{v}_T^{x2} - 1) - \hat{\omega}_{xx}^{-1}T^{-2}\hat{v}_T^x \sum_{t=1}^{T}\hat{v}_{t-1}^x\right],$$

$$\hat{R}_\gamma = \frac{1}{2}(\hat{\omega}_{xx}^{-1}T^{-1}\hat{v}_T^{x2} - 1) - \frac{\hat{\rho}}{\sqrt{1-\hat{\rho}^2}}\hat{R}_\beta,$$

$$\hat{R}_{\beta\beta} = \hat{\omega}_{xx}^{-1}T^{-2}\sum_{t=1}^{T} x_{t-1}^{\mu 2}, \qquad \hat{R}_{\gamma\gamma} = \hat{\omega}_{xx}^{-1}T^{-2}\sum_{t=1}^{T}\hat{v}_{t-1}^{x2},$$

where $\hat{\omega}_{yy.x} = \hat{\omega}_{yy} - \hat{\omega}_{xx}^{-1}\hat{\omega}_{xy}^2$, $\hat{\rho} = \hat{\omega}_{xy}\hat{\omega}_{xx}^{-1/2}\hat{\omega}_{yy}^{-1/2}$, and $\hat{\Omega}$ has been partitioned in the obvious way.

As is R, the statistic $\hat{R} = (\hat{R}_\beta, \hat{R}_\gamma, \hat{R}_{\beta\beta}, \hat{R}_{\gamma\gamma})'$ is invariant under transformations of the form $(y_t, x_t) \to (y_t + a, x_t)$, where $a \in \mathbb{R}$.[18] Under the assumptions of Section 4, $\hat{R}$ is asymptotically equivalent to R. More generally, we have the following theorem.

THEOREM 6: *Let* $\{(y_t, x_t)'\}$ *be generated by* (1) *and* (2), *suppose Assumptions* A1 *and* A3 *hold, and suppose* $b = T(\beta - \beta_0)\omega_{xx}^{1/2}\omega_{yy.x}^{-1/2}$ *and* $c = T(\gamma - 1)$ *are fixed as* T *increases without bound, where* $\omega_{yy.x} = \omega_{yy} - \omega_{xx}^{-1}\omega_{xy}^2$. *If* $\hat{\Omega} \to_p \Omega$, *then* $\hat{R} \to_d \mathcal{R}^\rho(b, c)$ *as* $T \to \infty$, *where* $\rho = \omega_{xy}\omega_{xx}^{-1/2}\omega_{yy}^{-1/2}$ *is the coefficient of correlation computed from* Ω. *Moreover*,

$$\lim_{T\to\infty} E_{\beta_T(b),\gamma_T(c)}\pi_\eta^*(\hat{R}; \hat{\rho}) = \varphi_\eta^*(b, c; \rho) \quad \forall\, b \geq 0, c \in \mathbb{R}$$

and

$$\lim_{T\to\infty} E_{\beta_T(b),\gamma_T(c)}\pi_\eta^{**}(\hat{R}; \hat{\rho}) = \varphi_\eta^{**}(b, c; \rho) \quad \forall\, b \in \mathbb{R}, c \in \mathbb{R}.$$

In view of Theorem 6, the Gaussian asymptotic power envelopes $\varphi_\eta^*(\cdot)$ and $\varphi_\eta^{**}(\cdot)$ are attainable whether or not the innovations of the regression model are normally distributed (with a known covariance matrix). Moreover, the presence of serial correlation does not affect our ability to attain the power envelope as long as Assumption A3 holds.[19]

Construction of consistent long-run variance estimators is a problem that has received considerable attention and there is no shortage of estimators that satisfy the high-level assumption $\hat{\Omega} \to_p \Omega$ of Theorem 6.[20]

To implement the tests based on $\pi_\eta^*(\hat{R}; \hat{\rho})$ and $\pi_\eta^{**}(\hat{R}; \hat{\rho})$, knowledge of the critical value functions $\mathcal{C}_\eta(\cdot)$, $\underline{\mathcal{C}}_\eta(\cdot)$, and $\overline{\mathcal{C}}_\eta(\cdot)$ is required. These critical value functions are implicitly defined in terms of the conditional distribution of $\mathcal{R}_\beta$ given $(\mathcal{R}_\gamma^\rho, \mathcal{R}_{\beta\beta}, \mathcal{R}_{\gamma\gamma})$. That distribution is nonstandard and does not appear to be available in closed form, but can easily be obtained (numerically) with the help of the following integral representation of the joint distribution of $\mathcal{R}^\rho$.

[18] In fact, $\hat{R}$ is invariant under transformations of the form $(y_t, x_t) \to (y_t + a, x_t + m_x)$, where $a \in \mathbb{R}$ and $m_x \in \mathbb{R}$.

[19] As usual, these predictions of asymptotic theory are not expected to be borne out in finite samples if the errors v_t^x are "nearly" I(-1) (i.e., if $|\psi(1)|$ is "small" relative to $\sum_{i=0}^\infty \psi_i^2$) or "nearly" I(1) (i.e., if $|\psi(1)|$ is "large" relative to $\sum_{i=0}^\infty \psi_i^2$).

[20] Important contributions to the literature on long-run variance estimation include Andrews (1991), Andrews and Monahan (1992), Hansen (1992), de Jong and Davidson (2000), and Newey and West (1987, 1994). A consistent estimator is described in remark (ii) at the end of this section.

THEOREM 7: *The joint distribution of $\mathcal{R}^\rho$ admits a density of the form*

$$f_{\mathcal{R}}^{\rho,0}(r) = \mathbb{1}\left\{r_\gamma + \frac{\rho}{\sqrt{1-\rho^2}} r_\beta > -\frac{1}{2}, 0 < r_{\beta\beta} < r_{\gamma\gamma}\right\}$$
$$\times \frac{1}{\sqrt{2\pi r_{\beta\beta}}} \exp\left(-\frac{r_\beta^2}{2r_{\beta\beta}}\right)$$
$$\times h\left(2r_\gamma + 2\frac{\rho}{\sqrt{1-\rho^2}} r_\beta + 1, r_{\gamma\gamma} - r_{\beta\beta}, r_{\gamma\gamma}\right),$$

where

$$h(q_\gamma, q_{\beta\beta}, q_{\gamma\gamma}) = \frac{1}{\pi^2\sqrt{q_\gamma}\sqrt{q_{\beta\beta}}} \int_0^\infty \mathrm{Re}\left\{\varkappa(t; \sqrt{q_\gamma}, \sqrt{q_{\beta\beta}}) \exp[-itq_{\gamma\gamma}]\right\} dt,$$

$$\varkappa(t; z_\gamma, z_{\beta\beta}) = \frac{|A+iB|^{-1/2}}{\sqrt{\cosh\sqrt{-2it}}} \exp\left[-\begin{pmatrix} z_\gamma \\ z_{\beta\beta}\end{pmatrix}'(AB^{-1}A+B)^{-1}AB^{-1}\begin{pmatrix} z_\gamma \\ z_{\beta\beta}\end{pmatrix}\right]$$
$$\times \exp\left[+i\begin{pmatrix} z_\gamma \\ z_{\beta\beta}\end{pmatrix}'\left(B^{-1} - B^{-1}A(AB^{-1}A+B)^{-1}AB^{-1}\right)\begin{pmatrix} z_\gamma \\ z_{\beta\beta}\end{pmatrix}\right]$$
$$+ \frac{|A+iB|^{-1/2}}{\sqrt{\cosh\sqrt{-2it}}}$$
$$\times \exp\left[-\begin{pmatrix} z_\gamma \\ -z_{\beta\beta}\end{pmatrix}'(AB^{-1}A+B)^{-1}AB^{-1}\begin{pmatrix} z_\gamma \\ -z_{\beta\beta}\end{pmatrix}\right]$$
$$\times \exp\left[+i\begin{pmatrix} z_\gamma \\ -z_{\beta\beta}\end{pmatrix}' \times \left(B^{-1} - B^{-1}A(AB^{-1}A+B)^{-1}AB^{-1}\right)\begin{pmatrix} z_\gamma \\ -z_{\beta\beta}\end{pmatrix}\right],$$

$$A = A(t) = \begin{pmatrix} \frac{1}{\sqrt{t}}\frac{\sinh 2\sqrt{t}+\sin 2\sqrt{t}}{\cosh 2\sqrt{t}+\cos 2\sqrt{t}} & \frac{1}{t}\frac{2\sinh\sqrt{t}\sin\sqrt{t}}{\cosh 2\sqrt{t}+\cos 2\sqrt{t}} \\ \frac{1}{t}\frac{2\sinh\sqrt{t}\sin\sqrt{t}}{\cosh 2\sqrt{t}+\cos 2\sqrt{t}} & \frac{1}{2t^{3/2}}\frac{\sinh 2\sqrt{t}-\sin 2\sqrt{t}}{\cosh 2\sqrt{t}+\cos 2\sqrt{t}} \end{pmatrix},$$

$$B = B(t) = \begin{pmatrix} \frac{1}{\sqrt{t}}\frac{\sinh 2\sqrt{t}-\sin 2\sqrt{t}}{\cosh 2\sqrt{t}+\cos 2\sqrt{t}} & \frac{1}{t}\left(1-\frac{2\cosh\sqrt{t}\cos\sqrt{t}}{\cosh 2\sqrt{t}+\cos 2\sqrt{t}}\right) \\ \frac{1}{t}\left(1-\frac{2\cosh\sqrt{t}\cos\sqrt{t}}{\cosh 2\sqrt{t}+\cos 2\sqrt{t}}\right) & \frac{1}{t}\left(1-\frac{1}{2\sqrt{t}}\frac{\sinh 2\sqrt{t}+\sin 2\sqrt{t}}{\cosh 2\sqrt{t}+\cos 2\sqrt{t}}\right) \end{pmatrix}.$$

REMARKS: (i) In $\hat{R}_\beta$ and $\hat{R}_\gamma$, the object $\frac{1}{2}(\hat{\omega}_{xx}^{-1}T^{-1}\hat{v}_T^{x2}-1)$ satisfies

$$\frac{1}{2}(\hat{\omega}_{xx}^{-1}T^{-1}\hat{v}_T^{x2}-1)\to_{d_0}\frac{1}{2}[W_x(1)^2-1],$$

where $\to_{d_0}$ is shorthand for "$\to_d$ when $(\beta,\gamma)=(\beta_0,1)$." This convergence result generalizes in an obvious way to higher dimensions, but the important equality $\frac{1}{2}[W_x(1)^2-1]=\int_0^1 W_x(r)\,dW_x(r)$ does not generalize. As pointed out by a referee, it may therefore seem more natural to employ a formulation that admits an obvious multivariate generalization whose limiting representation is of the form $\int_0^1 W(r)\,dW(r)'$. Our reason for not doing so is that Ng and Perron (2001), in their work on the finite sample size behavior of unit root tests, found that $\frac{1}{2}(\hat{\omega}_{xx}^{-1}T^{-1}\hat{v}_T^{x2}-1)$ tends to be better approximated by its asymptotic representation than are those objects that generalize most easily to higher dimensions.

(ii) Under the assumptions of Theorem 6 and fairly general conditions on the kernel $k(\cdot)$ and the bandwidth parameter B_T, it follows from Jansson (2002) that

$$\hat{\Omega}=T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}k\left(\frac{|t-s|}{B_T}\right)\hat{u}_t\hat{u}_s'\to_p\Omega,$$

where $\hat{u}_t=(y_t-T^{-1}\sum_{s=1}^{T}y_s-\hat{\beta}x_{t-1}^{\mu},\hat{v}_t^x-\hat{\gamma}\hat{v}_{t-1}^x)'$ and

$$\hat{\beta}=\left(\sum_{t=1}^{T}x_{t-1}^{\mu 2}\right)^{-1}\sum_{t=1}^{T}x_{t-1}^{\mu}y_t,\quad \hat{\gamma}=\left(\sum_{t=1}^{T}\hat{v}_{t-1}^{x2}\right)^{-1}\sum_{t=1}^{T}\hat{v}_{t-1}^{x}\hat{v}_t^{x}.$$

(iii) Because the function $|\varkappa(t;z_\gamma,z_{\beta\beta})|$ can be shown to exhibit exponential decay as $t\to\infty$, it is straightforward to obtain accurate numerical approximations to the integral that appears in the definition of $h(\cdot)$.

(iv) Asymptotic p-values for the one-sided test are given by the formula

$$p(\hat{R};\hat{\rho})=\frac{\int_{\hat{R}_\beta}^{\infty}f_{\mathcal{R}}^{0}(r_\beta,\hat{R}_\gamma,\hat{R}_{\beta\beta},\hat{R}_{\gamma\gamma};\hat{\rho})\,dr_\beta}{\int_{-\infty}^{\infty}f_{\mathcal{R}}^{0}(r_\beta,\hat{R}_\gamma,\hat{R}_{\beta\beta},\hat{R}_{\gamma\gamma};\hat{\rho})\,dr_\beta}. \tag{26}$$

In view of Theorem 7, numerical evaluation of $p(\hat{R};\hat{\rho})$ involves calculating two double integrals. In our experience, this calculation usually takes no more than 2–3 seconds on a contemporary computer. (MATLAB code for computing $p(\hat{R};\hat{\rho})$ is available from the authors upon request.)

6. SIMULATIONS

This section presents some simulation evidence that sheds light on the size and power properties of the one-sided test π^* and some of its rivals. Following Wright (2000), we simulate the simple model introduced in (3) and (4), and assume that ε_t^y and ε_t^x have unit variance and correlation denoted by ρ.

Table I reports the small sample ($T = 100$) size behavior across 1,000 replications for five testing procedures with nominal size equal to 5%. The tests considered are the one-sided t-tests based on the ordinary least squares (OLS) and dynamic OLS (DOLS) estimators of β, the L_2 test of Wright (2000), the refined Bonferroni test of Campbell and Yogo (2005) (labeled CYRB), and the test based on $\pi^*_{0.05}$.[21,22] We consider two values of ρ, namely -0.5 and 0.5. Five values of c, the local-to-unity parameter that governs the persistence of the regressor, are considered: $c = 0$ (corresponding to the exactly integrated regressors), $c \in \{-10, -20\}$ (corresponding to nearly integrated regressors), and $c \in \{-50, -100\} = \{-T/2, -T\}$ (corresponding to stationary regressors).

The t-test based on the OLS estimator, which has correct (asymptotic) size when the regressors are stationary, has null rejection probabilities close to 5% when the regressors are stationary, but its behavior is erratic when the regressors are nearly or exactly integrated with severe overrejections being observed for $\rho = -0.5$. The t-test based on the DOLS estimator (e.g., Stock and Watson (1993)), which has correct (asymptotic) size when the regressors are exactly integrated, has null rejection probabilities close to 5% when the regressors are exactly integrated, but in agreement with the theoretical results of Elliott

TABLE I

SIZE PROPERTIES

ρ	c	OLS	DOLS	L_2	CYRB	$\pi^*_{0.05}$
−0.5	0	15.1%	3.4%	4.8%	2.9%	4.3%
	−10	8.7%	0.3%	4.7%	3.2%	4.9%
	−20	7.8%	0.1%	4.2%	2.5%	4.1%
	−50	6.1%	0.0%	5.1%	0.7%	4.9%
	−100	4.7%	0.0%	5.2%	0.0%	3.9%
0.5	0	0.3%	3.9%	4.9%	0.8%	5.6%
	−10	3.4%	31.7%	5.9%	0.8%	4.4%
	−20	4.1%	55.3%	5.4%	0.9%	4.5%
	−50	5.7%	87.2%	5.6%	3.9%	5.8%
	−100	4.0%	99.3%	4.7%	60.9%	6.7%

[21]The CYRB test is designed to have asymptotic size equal to 5% when the persistence parameter c is bounded between 5 and -50. For details, see Campbell and Yogo (2005, Section 3.4).

[22]The $\pi^*_{0.05}$ test is implemented using the OLS estimator of Σ and employing high order recursive adaptive quadrature to numerically evaluate the conditional p-values using the formula in (26).

(1998) its behavior is found to be unsatisfactory when the regressors are nearly integrated (or stationary). Finally, of the remaining three tests, all being designed for the case of nearly integrated regressors, L_2 and $\pi^*_{0.05}$ exhibit nice behavior across the scenarios considered, while the Campbell and Yogo (2005) refined Bonferroni test is undersized in most cases.

In conclusion, Table I demonstrates that although t-tests based on the OLS and DOLS estimators exhibit unsatisfactory size behavior when the (endogenous) regressors are nearly integrated, at least three conceptually different methodologies can be used to obtain tests with good size properties across a range of values of the persistence parameter γ. We next explore the relative merits of these three methodologies from the point of view of power in models with nearly integrated regressors.

Table II reports the large-sample ($T = 1{,}000$) rejection rates across 500 replications (for a variety of values of the parameters ρ, c, and b) for the L_2 test, the Campbell and Yogo (2005) test, and $\pi^*_{0.05}$. Also reported are rejection rates for two additional testing procedures, labeled ORA and CYB, respectively,

TABLE II

POWER PROPERTIES

		$\rho = -0.5$					$\rho = 0.5$				
c	b	ORA	L_2	CYRB	CYB	$\pi^*_{0.05}$	ORA	L_2	CYRB	CYB	$\pi^*_{0.05}$
0	0	5.2%	7.4%	3.2%	1.6%	6.0%	5.4%	6.2%	0.8%	0.4%	5.8%
	5	54.0%	26.6%	43.4%	13.0%	42.8%	55.8%	26.8%	30.4%	25.2%	42.8%
	10	89.2%	55.0%	86.8%	45.6%	81.2%	91.6%	52.6%	71.4%	62.4%	61.8%
	15	97.6%	73.2%	97.0%	88.0%	95.6%	98.6%	71.4%	87.4%	84.4%	68.6%
	20	99.6%	80.4%	99.8%	98.2%	98.4%	99.6%	84.8%	95.0%	94.6%	75.5%
−5	0	5.4%	4.6%	2.4%	0.4%	5.6%	5.2%	4.2%	0.8%	0.0%	4.2%
	5	34.0%	2.8%	15.8%	1.6%	10.6%	32.4%	14.2%	30.4%	6.8%	13.2%
	10	71.8%	11.4%	54.4%	5.4%	34.4%	73.6%	30.0%	71.4%	31.2%	24.2%
	15	92.8%	25.0%	88.2%	23.8%	71.2%	91.8%	44.0%	87.4%	59.6%	28.6%
	20	98.8%	40.2%	98.0%	66.6%	91.8%	99.0%	58.4%	95.0%	81.0%	37.6%
−10	0	5.2%	6.4%	2.4%	0.4%	6.2%	6.0%	6.2%	0.4%	0.2%	2.2%
	5	22.6%	3.0%	10.2%	1.8%	9.4%	31.6%	10.4%	7.0%	3.6%	7.6%
	10	58.8%	4.6%	36.0%	4.8%	18.0%	64.4%	20.0%	26.2%	17.2%	11.0%
	15	84.2%	7.2%	70.2%	12.6%	40.4%	87.2%	35.4%	57.4%	39.2%	17.2%
	20	95.2%	16.8%	90.8%	33.0%	67.0%	97.4%	42.6%	73.0%	70.2%	22.0%
−15	0	4.8%	4.2%	1.4%	0.2%	8.6%	4.0%	5.0%	1.2%	0.0%	1.6%
	5	18.6%	2.0%	6.8%	1.0%	7.4%	19.8%	9.0%	6.4%	2.2%	4.4%
	10	55.2%	1.6%	29.2%	3.0%	12.8%	48.4%	17.2%	21.4%	13.2%	6.8%
	15	80.6%	1.6%	59.4%	7.8%	24.8%	79.2%	21.8%	41.4%	32.4%	12.6%
	20	93.6%	5.8%	83.0%	18.6%	41.4%	93.2%	32.2%	67.2%	55.0%	16.6%
−20	0	4.8%	5.0%	2.0%	0.4%	4.6%	5.2%	5.4%	0.8%	1.2%	2.2%
	5	16.6%	2.8%	6.2%	1.0%	7.6%	19.8%	8.2%	6.2%	1.6%	3.6%
	10	44.6%	1.0%	21.6%	2.6%	9.6%	40.4%	11.2%	16.6%	9.4%	4.4%
	15	70.8%	1.4%	43.4%	5.2%	15.2%	68.4%	17.4%	37.8%	27.6%	8.8%
	20	88.4%	3.0%	70.4%	12.8%	29.2%	84.2%	22.8%	54.6%	43.0%	13.2%

where ORA is an "oracle" t-test based on the estimator of β obtained from a regression of y_t on x_{t-1} and $x_t - \gamma_T(c)x_{t-1}$ (and therefore assumes knowledge of the persistence parameter c), and CYB is the (unrefined) Bonferroni test of Campbell and Yogo (2005). The latter has been included to allow comparison of our methodology to a Bonferroni procedure that (unlike CYRB) does not require partial knowledge of c (in the form of an upper bound on c).

The infeasible oracle t-test is seen to be strictly superior to all of its (feasible) competitors, implying that lack of knowledge of the persistence parameter c is associated with a nonnegligible loss of power. Being a conservative test (even asymptotically), the Campbell and Yogo (2005) refined Bonferroni test is inferior to $\pi^*_{0.05}$ (and L_2) for all values of b that are sufficiently close to zero. On the other hand, the CYRB test seems to dominate against alternatives that are not close to the null. (For the configurations considered here, CYRB dominates against alternatives for which $b \geq 10$.) Of the tests that do not require partial knowledge of c, the Campbell and Yogo (2005) (unrefined) Bonferroni test dominates L_2 in most cases, whereas the ranking of CYB and $\pi^*_{0.05}$ depends on the sign of ρ, with $\pi^*_{0.05}$ dominating CYB when $\rho = -0.5$, while CYB tends to outperform $\pi^*_{0.05}$ for alternatives away from the null when $\rho = 0.5$.

The results reported in Table II suggest three conclusions. First, the impressive performance of CYRB suggests that for the model (and parameter configurations) under consideration here, partial knowledge of the nuisance parameter is very valuable. Second, in applications where the practitioner is unwilling to assume partial knowledge of c, the ranking of CYB and $\pi^*_{0.05}$ depends on a single nuisance parameter, namely (the sign of) ρ. Finally, the fact that ρ is consistently estimable implies that in practice a simple, data-dependent method of choosing between CYB and $\pi^*_{0.05}$ is available.

7. CONCLUSION

This paper has proposed novel conditionality restrictions subject to which optimality results can be obtained for one- and two-sided testing problems that involve the regression coefficient in a bivariate regression model with a highly persistent regressor. We have developed finite sample and asymptotic optimality theory under the assumption of Gaussian errors and have shown the normality assumption to be least favorable. The derivation of finite sample optimality results uses classical statistical theory and the theory of (curved) exponential families, whereas the large-sample optimality results were obtained by using the finite sample optimality results and the theory of limits of experiments.

Because our asymptotic results depend on the underlying model only through the associated limiting experiment, they can be extended to models more general than the model in which the error term of the equation of interest is a martingale difference sequence with respect to its lags and to current and lagged values of the nearly integrated regressor. Jansson and Moreira

(2004) illustrate this point by showing that the results of this paper extend in a straightforward way to a (cointegration-type) model that accommodates correlation between the (potentially) serially correlated error term of the equation of interest and current (and lagged) values of the nearly integrated regressor.

Our asymptotic optimality results complement those available in the existing literature on limits of experiments. The optimality results currently available in that literature pertain almost exclusively to models that exhibit LAN or LAMN likelihood ratios. In contrast, our results are obtained for a model whose likelihood ratios are LAQ (but not LAMN) and differ from existing results in a nontrivial way.[23] In models with LAN likelihood ratios (such as (3) and (4) in the stationary case when $|\rho| < 1$), the commonly used Wald statistics are asymptotically optimal among tests with the same asymptotic level. Wald statistics also enjoy optimality properties in models with LAMN likelihood ratios (such as (3) and (4) in the unit root case when $\rho = 1$), being optimal among tests with correct asymptotic conditional size given the value of the observed information matrix.

In the LAMN context, conditioning on the observed information matrix seems natural because its asymptotic counterpart acts as an ancillary statistic in the limiting experiment. The latter property characterizes LAMN models within the class of LAQ models (Jeganathan (1995, Proposition 6)), implying that conditioning on ancillaries does not suffice if we want to develop optimality theory for LAQ models outside the class of LAMN models. This paper provides an example of a testing problem with nuisance parameters where the stronger requirement of conditioning on specific ancillaries (i.e., statistics that would be ancillary if the values of nuisance parameters were known) makes it possible to develop optimality results in a model with LAQ likelihood ratios. (Coincidentally, the specific ancillary in our example turns out to be given by the observed information matrix.) It would be of interest to explore whether the conditionality restriction proposed here can be applied to develop optimality results for other testing problems that involve nuisance parameters in models without LAMN structure.

Dept. of Economics, UC Berkeley, 549 Evans Hall 3880, Berkeley, CA 94720-3880, U.S.A.; mjansson@econ.berkeley.edu

and

Dept. of Economics, Harvard University, Littauer Center M-6, 1875 Cambridge Street, Cambridge, MA 02138, U.S.A.; moreira@fas.harvard.edu.

Manuscript received October, 2004; final revision received January, 2006.

[23] Another testing problem to which the theory of LAQ likelihood ratios applies but theory of LAMN likelihood ratios does not is the unit root testing problem. That testing problem has been extensively studied, celebrated results include those of Dickey and Fuller (1979, 1981), Elliott, Rothenberg, and Stock (1996), Phillips (1987a), and Phillips and Perron (1988). (For reviews, see Haldrup and Jansson (2005) and Stock (1994).)

APPENDIX: PROOFS

PROOF OF LEMMA 1: Lemma 1 follows from (6) and the properties of exponential families (e.g., Lehmann and Romano (2005, Lemma 2.7.2)). *Q.E.D.*

PROOF OF THEOREM 2: It follows from Lemma 1(b) and Lehmann and Romano (2005, Theorem 4.4.1) th at if $\phi(\cdot)$ satisfies (7), then

$$E_{\beta,\gamma}[\phi(S)|S_{\beta\beta}, S_{\gamma\gamma}] \leq E_{\beta,\gamma}[\phi^*_\eta(S)|S_{\beta\beta}, S_{\gamma\gamma}] \quad \forall \beta \geq \beta_0, \gamma \in \mathbb{R}.$$

Part (a) now follows from the law of iterated expectations.

Analogous reasoning establishes part (b) (including existence and essential uniqueness of the functions $\underline{C}_\eta(\cdot)$ and $\overline{C}_\eta(\cdot)$ that satisfying (12) and (13)). *Q.E.D.*

PROOF OF LEMMA 3: Lemma 3 follows from standard weak convergence arguments (e.g., Phillips (1987a, 1988a, 1988b)) and straightforward algebra. *Q.E.D.*

PROOF OF LEMMA 4: Lemma 4 follows from (17), Lemma 3, Lehmann and Romano (2005, Lemma 2.7.2), and Le Cam's third lemma (e.g., Jeganathan (1995, Proposition 1) and van der Vaart (2002, Lemma 3.1)). Le Cam's third lemma is applicable because the family of distributions associated with the maximal invariant has LAQ likelihood ratios at $(\beta, \gamma) = (\beta_0, 1)$. In particular, $\mathcal{L}_T(b, c) - \mathcal{L}_T(0, 0) \to_{d_0} \Lambda^\rho(b, c)$, where

$$\Lambda^\rho(b, c) = b\mathcal{R}_\beta + c\mathcal{R}^\rho_\gamma - \frac{1}{2}\left(b - \frac{\rho}{\sqrt{1-\rho^2}}c\right)^2 \mathcal{R}_{\beta\beta} - \frac{1}{2}c^2\mathcal{R}_{\gamma\gamma},$$

and the convergence result follows from Lemma 3. *Q.E.D.*

The proof of Theorem 5 makes use of the following lemma.

LEMMA 8: *Let* $\eta \in (0, 1)$ *be given and define*

$$\mathbb{S} = \{(r_\gamma, r_{\beta\beta}, r_{\gamma\gamma}; \rho) : r_\gamma \in \mathbb{R}, 0 < r_{\beta\beta} < r_{\gamma\gamma}, -1 < \rho < 1\}.$$

(a) *There exists a* (*unique*) *continuous function* $C_\eta : S \to \mathbb{R}$ *such that* $\pi^*_\eta(\cdot; \rho)$ *satisfies* (21), *where* $\pi^*_\eta(\cdot)$ *is defined as in Section* 4.

(b) *There exist* (*unique*) *continuous functions* $\underline{\mathcal{C}}_\eta : S \to \mathbb{R}$ *and* $\overline{\mathcal{C}}_\eta : S \to \mathbb{R}$ *such that* $\pi^{**}_\eta(\cdot; \rho)$ *satisfies* (24) *and* (25), *where* $\pi^{**}_\eta(\cdot)$ *is defined as in Section* 4.

A proof of Lemma 8 can be found in Jansson and Moreira (2004). That proof constructs a conditional probability density function of $\mathcal{R}_\beta$ given

$(\mathcal{R}_{\gamma}^{\rho}, \mathcal{R}_{\beta\beta}, \mathcal{R}_{\gamma\gamma})$ that satisfies the conditions of the following lemma, which gives general conditions under which critical value functions for one- and two-sided tests are continuous in their arguments.[24]

LEMMA 9: *Let (Θ, d_{θ}) be a metric space and let $\{f(\cdot;\theta):\theta \in \Theta\}$ be a family of probability density functions on $\mathbb{R}$. Let $\eta \in (0,1)$ and $\theta_0 \in \Theta$ be given, and suppose $f(r;\cdot)$ is continuous at θ_0 (with respect to the metric d_{θ}) for almost every $r \in \mathbb{R}$.*

(a) *Suppose that for every $\theta \in \Theta$ there is a unique number $C_{\eta}(\theta)$ such that*

$$\int_{C_{\eta}(\theta)}^{\infty} f(r;\theta)\,dr = \eta.$$

Then $C_{\eta}:\Theta \to \mathbb{R}$ is continuous at θ_0.

(b) *Suppose that for every $\theta \in \Theta$ there are unique numbers $\underline{C}_{\eta}(\theta)$ and $\overline{C}_{\eta}(\theta)$ such that*

$$\int_{\underline{C}_{\eta}(\theta)}^{\overline{C}_{\eta}(\theta)} f(r;\theta)\,dr = 1-\eta,$$

$$\int_{\underline{C}_{\eta}(\theta)}^{\overline{C}_{\eta}(\theta)} rf(r;\theta)\,dr = (1-\eta)\int_{-\infty}^{\infty} rf(r;\theta)\,dr.$$

If $\int_{-\infty}^{\infty}|r|f(r;\theta_0)\,dr < \infty$ and $\int_{-\infty}^{\infty}|r|f(r;\cdot)\,dr$ is continuous at θ_0, then $\underline{C}_{\eta}:\Theta \to \mathbb{R}$ and $\overline{C}_{\eta}:\Theta \to \mathbb{R}$ are continuous at θ_0.

PROOF OF THEOREM 5: The proof of Theorem 5 is based on Lemma 4 and the theory of LAQ likelihood ratios. Repeated use will be made of the fact that

$$\int_{\mathbb{R}^4} g(r) f_{\mathcal{R}}(r; b, c, \rho)\,dr = E\big[g(\mathcal{R}^{\rho}) e^{\Lambda^{\rho}(b,c)}\big] \quad \forall\, b \in \mathbb{R}, c \in \mathbb{R},$$

where $\Lambda^{\rho}(\cdot)$ is defined as in the proof of Lemma 4 and $g:\mathbb{R}^4 \to \mathbb{R}$ is any function such that either side of the equality is well defined.

PROOF OF (a): Let $\Pi(\eta,\rho)$ denote the class of all functions $\pi(\cdot)$ that satisfy

$$E\big[(\pi(\mathcal{R}^{\rho}) - \eta) g(\mathcal{R}_{\beta\beta}, \mathcal{R}_{\gamma\gamma}) e^{\Lambda^{\rho}(0,c)}\big] = 0 \quad \forall\, c \in \mathbb{R}, g \in C_b(\mathbb{R}^2).$$

By construction, $\pi_{\eta}^{*}(\cdot;\rho) \in \Pi(\eta,\rho)$. Applying Lehmann and Romano (2005, Theorem 4.4.1) and the law of iterated expectations, it can be shown that

[24]We are grateful to a referee for suggesting the present formulation of Lemma 9(b) and for pointing out that Lemma 9 is (essentially) a well-known result.

$\pi_\eta^*(\cdot;\rho)$ satisfies

$$\begin{aligned}&E\big[\pi(\mathcal{R}^\rho)e^{\Lambda^\rho(b,c)}\big]\\&\quad\le E\big[\pi_\eta^*(\mathcal{R}^\rho;\rho)e^{\Lambda^\rho(b,c)}\big]\quad\forall\, b\ge 0, c\in\mathbb{R}, \pi\in\Pi(\eta,\rho).\end{aligned}$$

Because $\mathcal{C}_\eta(\cdot)$ is continuous (Lemma 8), it follows from Lemma 3 and the continuous mapping theorem (CMT) that $\pi_\eta^*(R;\rho)\to_{d_0}\pi_\eta^*(\mathcal{R}^\rho;\rho)$. This convergence result, Le Cam's third lemma, and Billingsley (1999, Theorem 3.5) can be used to show that $\{\pi_\eta^*\}$ satisfies (19) and that

$$\lim_{T\to\infty} E_{\beta_T(b),\gamma_T(c)}\pi_\eta^*(R;\rho)=E\big[\pi_\eta^*(\mathcal{R}^\rho;\rho)e^{\Lambda^\rho(b,c)}\big]\quad\forall\, b\ge 0, c\in\mathbb{R}.$$

The proof of (a) will be completed by showing that for any $\{\pi_T(\cdot)\}$ that satisfies (19), any $b\ge 0$, and any $c\in\mathbb{R}$, there exists a $\pi\in\Pi(\eta,\rho)$ such that

$$\overline{\lim_{T\to\infty}}\, E_{\beta_T(b),\gamma_T(c)}\pi_T(R;\rho)=E\big[\pi(\mathcal{R}^\rho;\rho)e^{\Lambda^\rho(b,c)}\big].\tag{27}$$

Let $\{\pi_T(\cdot)\}$, $b\ge 0$, and $c\in\mathbb{R}$ be given, and suppose $\{\pi_T(\cdot)\}$ satisfies (19). Let $\{\pi_{T'}(\cdot)\}$ be any subsequence of $\{\pi_T(\cdot)\}$ that satisfies

$$\lim_{T'\to\infty} E_{\beta_{T'}(b),\gamma_{T'}(c)}\pi_{T'}(R;\rho)=\overline{\lim_{T\to\infty}}\, E_{\beta_T(b),\gamma_T(c)}\pi_T(R;\rho).$$

Because $\pi_{T'}=O_p(1)$, it follows from Prohorov's theorem (e.g., Billingsley (1999)) that there exists a further subsequence $\{\pi_{T''}(\cdot)\}$ such that

$$(\pi_{T''},R)\to_{d_0}(\pi_\infty,\mathcal{R}^\rho)\tag{28}$$

as $T''\to\infty$, where π_∞ is some random variable (defined on the same probability space as $\mathcal{R}^\rho$) and the dependence of R on T'' has been suppressed. Now,

$$\begin{aligned}\overline{\lim_{T\to\infty}}\, E_{\beta_T(b),\gamma_T(c)}\pi_T(R;\rho)&=\lim_{T''\to\infty}E_{\beta_{T''}(b),\gamma_{T''}(c)}\pi_{T''}(R;\rho)\\&=E\big[\pi_\infty e^{\Lambda^\rho(b,c)}\big]\\&=E\big[\pi(\mathcal{R}^\rho)e^{\Lambda^\rho(b,c)}\big],\quad \pi(\mathcal{R}^\rho)=E(\pi_\infty|\mathcal{R}^\rho),\end{aligned}$$

where the second equality uses (28), Le Cam's third lemma, and Billingsley (1999, Theorem 3.5), and the last equality uses the law of iterated expectations. The result $\pi\in\Pi(\eta,\rho)$ now follows because

$$\begin{aligned}&E\big[(\pi(\mathcal{R}^\rho)-\eta)g(\mathcal{R}_{\beta\beta},\mathcal{R}_{\gamma\gamma})e^{\Lambda^\rho(0,c)}\big]\\&\quad=E\big[(\pi_\infty-\eta)g(\mathcal{R}_{\beta\beta},\mathcal{R}_{\gamma\gamma})e^{\Lambda^\rho(0,c)}\big]\\&\quad=\lim_{T''\to\infty}E_{\beta_{T''}(0),\gamma_{T''}(c)}\big[(\pi_{T''}(R;\rho)-\eta)g(R_{\beta\beta},R_{\gamma\gamma})\big]\\&\quad=0\end{aligned}$$

for any $c \in \mathbb{R}$ and any $g \in C_b(\mathbb{R}^2)$, where the first equality uses the law of iterated expectations, the second equality uses (28), Le Cam's third lemma, Lemma 3, Billingsley (1999, Theorem 3.5) and CMT, and the last equality uses the fact that $\{\pi_T(\cdot)\}$ satisfies (19). This completes the proof of part (a).

PROOF OF (b): Let $\Pi_0(\eta,\rho) \subseteq \Pi(\eta,\rho)$ denote the class of all functions $\pi(\cdot)$ that satisfy $\pi \in \Pi(\eta,\rho)$ and

$$E\big[(\pi(\mathcal{R}^\rho) - \eta)\mathcal{R}_\beta \cdot g(\mathcal{R}_{\beta\beta}, \mathcal{R}_{\gamma\gamma})e^{\Lambda^\rho(0,c)}\big] = 0 \quad \forall c \in \mathbb{R}, g \in C_b(\mathbb{R}^2).$$

By construction, $\pi^{**}_\eta(\cdot;\rho) \in \Pi_0(\eta,\rho)$. Applying Lehmann and Romano (2005, Theorem 4.4.1) and the law of iterated expectations, it can be shown that $\pi^{**}_\eta(\cdot;\rho)$ satisfies

$$\begin{aligned} &E\big[\pi(\mathcal{R}^\rho)e^{\Lambda^\rho(b,c)}\big] \\ &\quad \le E\big[\pi^{**}_\eta(\mathcal{R}^\rho;\rho)e^{\Lambda^\rho(b,c)}\big] \quad \forall b \in \mathbb{R}, c \in \mathbb{R}, \pi \in \Pi_0(\eta,\rho). \end{aligned}$$

Because $\underline{\mathcal{C}}_\eta(\cdot)$ and $\overline{\mathcal{C}}_\eta(\cdot)$ are continuous (Lemma 8), it follows from Lemma 3 and CMT that $\pi^{**}_\eta(R;\rho) \to_{d_0} \pi^{**}_\eta(\mathcal{R}^\rho;\rho)$. This convergence result, Le Cam's third lemma, and Billingsley (1999, Theorem 3.5) can be used to show that $\{\pi^{**}_\eta\}$ satisfies (19), (22), and

$$\lim_{T\to\infty} E_{\beta_T(b),\gamma_T(c)}\pi^{**}_\eta(R;\rho) = E\big[\pi^{**}_\eta(\mathcal{R}^\rho;\rho)e^{\Lambda^\rho(b,c)}\big] \quad \forall b \in \mathbb{R}, c \in \mathbb{R}.$$

Finally, by proceeding as in the proof of (a) it can be shown that for any $\{\pi_T(\cdot)\}$ that satisfies (19) and (22), any $b \in \mathbb{R}$, and any $c \in \mathbb{R}$, there exists a $\pi \in \Pi_0(\eta,\rho)$ such that

$$\overline{\lim_{T\to\infty}} E_{\beta_T(b),\gamma_T(c)}\pi_T(R;\rho) = E\big[\pi(\mathcal{R}^\rho;\rho)e^{\Lambda^\rho(b,c)}\big]. \qquad \textit{Q.E.D.}$$

PROOF OF THEOREM 6: The result $\hat{R} \to_d \mathcal{R}^\rho(b,c)$ follows from standard weak convergence arguments (e.g., Phillips (1987a, 1988a, 1988b) and Phillips and Solo (1992)) and straightforward algebra. For instance,

$$\begin{aligned} \hat{R}_{\gamma\gamma} &= \hat{\omega}_{xx}^{-1}T^{-2}\sum_{t=1}^T \hat{v}^{x2}_{t-1} = \omega_{xx}^{-1}T^{-2}\sum_{t=1}^T v^{x2}_{t-1} + o_p(1) \\ &= \int_0^1 \big(\omega_{xx}^{-1/2}T^{-1/2}v^x_{\lfloor Tr\rfloor}\big)^2\,dr + o_p(1) \to_d \int_0^1 W_{x,c}(r)^2\,dr, \end{aligned}$$

where the second equality uses $\hat{\omega}_{xx} \to_p \omega_{xx}$ and $T^{-1/2}(x_1 - \mu_x) \to_p 0$, and the convergence result uses $\omega_{xx}^{-1/2}T^{-1/2}v^x_{\lfloor T\cdot\rfloor} \to_d W_{x,c}(\cdot)$ and CMT.

For any $b \geq 0$ and any $c \in \mathbb{R}$,

$$\begin{aligned}
&E_{\beta_T(b),\gamma_T(c)}\pi^*_\eta(\hat{R};\hat{\rho}) \\
&\quad = \Pr_{\beta_T(b),\gamma_T(c)}[\hat{R}_\beta > \mathcal{C}_\eta(\hat{R}_\gamma, \hat{R}_{\beta\beta}, \hat{R}_{\gamma\gamma};\hat{\rho})] \\
&\quad \to \Pr\big[\mathcal{R}^\rho_\beta(b,c) > \mathcal{C}_\eta\big(\mathcal{R}^\rho_\gamma(b,c), \mathcal{R}_{\beta\beta}(c), \mathcal{R}_{\gamma\gamma}(c);\rho\big)\big] \\
&\quad = \varphi^*_\eta(b,c;\rho),
\end{aligned}$$

where the convergence result uses $(\hat{R}', \hat{\rho})' \to_d (\mathcal{R}^\rho(b,c)', \rho)'$, continuity of $\mathcal{C}_\eta(\cdot)$, and CMT. An analogous argument shows that

$$\lim_{T\to\infty} E_{\beta_T(b),\gamma_T(c)}\pi^{**}_\eta(\hat{R};\hat{\rho}) \to \varphi^{**}_\eta(b,c;\rho) \quad \forall (b,c) \in \mathbb{R}^2. \qquad \textit{Q.E.D.}$$

PROOF OF THEOREM 7: Let $\mathcal{Z}_\gamma = W_x(1)$, $\mathcal{Z}_{\beta\beta} = \int_0^1 W_x(r)\,dr$, and $\mathcal{Q}_{\gamma\gamma} = \mathcal{R}_{\gamma\gamma}(0)$. Using changes of variables, it can be shown that

$$\begin{aligned}
f^{\rho,0}_{\mathcal{R}}(r) &= \frac{1}{\sqrt{2\pi r_{\beta\beta}}}\exp\left(-\frac{r^2_\beta}{2r_{\beta\beta}}\right) \\
&\quad \times 2f_{\mathcal{Q}}\left(2r_\gamma + 2\frac{\rho}{\sqrt{1-\rho^2}}r_\beta + 1, r_{\gamma\gamma} - r_{\beta\beta}, r_{\gamma\gamma}\right),
\end{aligned}$$

where, with $f_{\mathcal{Z}_\gamma,\mathcal{Z}_{\beta\beta},\mathcal{Q}_{\gamma\gamma}}(\cdot)$ denoting "the" density of $(\mathcal{Z}_\gamma, \mathcal{Z}_{\beta\beta}, \mathcal{Q}_{\gamma\gamma})'$,

$$\begin{aligned}
2f_{\mathcal{Q}}(q_\gamma, q_{\beta\beta}, q_{\gamma\gamma}) &= \frac{1}{\sqrt{q_\gamma}\sqrt{q_{\beta\beta}}} f_{\mathcal{Z}_\gamma,\mathcal{Z}_{\beta\beta},\mathcal{Q}_{\gamma\gamma}}(\sqrt{q_\gamma}, \sqrt{q_{\beta\beta}}, q_{\gamma\gamma}) \\
&\quad + \frac{1}{\sqrt{q_\gamma}\sqrt{q_{\beta\beta}}} f_{\mathcal{Z}_\gamma,\mathcal{Z}_{\beta\beta},\mathcal{Q}_{\gamma\gamma}}(\sqrt{q_\gamma}, -\sqrt{q_{\beta\beta}}, q_{\gamma\gamma}).
\end{aligned}$$

By the inversion theorem for characteristic functions,

$$\begin{aligned}
&f_{\mathcal{Z}_\gamma,\mathcal{Z}_{\beta\beta},\mathcal{Q}_{\gamma\gamma}}(z_\gamma, z_{\beta\beta}, q_{\gamma\gamma}) \\
&\quad = \frac{1}{(2\pi)^3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varkappa^*(t_\gamma, t_{\beta\beta}, t_{\gamma\gamma}) \\
&\qquad\qquad \times \exp[-i(t_\gamma z_\gamma + t_{\beta\beta}z_{\beta\beta} + t_{\gamma\gamma}q_{\gamma\gamma})]\,dt_\gamma\,dt_{\beta\beta}\,dt_{\gamma\gamma} \\
&\quad = \frac{1}{2\pi}\int_{-\infty}^{\infty} \varkappa(t; z_\gamma, z_{\beta\beta})\exp[-itq_{\gamma\gamma}]\,dt,
\end{aligned}$$

where

$$\varkappa(t; z_\gamma, z_{\beta\beta}) = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varkappa^*(t_\gamma, t_{\beta\beta}, t)\exp[-i(t_\gamma z_\gamma + t_{\beta\beta}z_{\beta\beta})]\,dt_\gamma\,dt_{\beta\beta}$$

and $\varkappa^*(\cdot)$ is the joint characteristic function of $(\mathcal{Z}_\gamma, \mathcal{Z}_{\beta\beta}, \mathcal{Q}_{\gamma\gamma})'$. It follows from Abadir and Larsson (2001) that

$$\varkappa^*(t_\gamma, t_{\beta\beta}, t_{\gamma\gamma}) = E\exp[it_\gamma\mathcal{Z}_\gamma + it_{\beta\beta}\mathcal{Z}_{\beta\beta} + it_{\gamma\gamma}\mathcal{Q}_{\gamma\gamma}] = \frac{\exp\left[\frac{1}{4}(l_1(t_\gamma, t_{\gamma\gamma}) + l_2(t_\gamma, t_{\beta\beta}, t_{\gamma\gamma}) + l_3(t_{\beta\beta}, it_{\gamma\gamma}))\right]}{\sqrt{\cosh\sqrt{-2it_{\gamma\gamma}}}},$$

where

$$\begin{aligned}
l_1(t_\gamma, t_{\gamma\gamma}) &= -2t_\gamma^2\frac{\tanh\sqrt{-2it_{\gamma\gamma}}}{\sqrt{-2it_{\gamma\gamma}}} \\
&= -t_\gamma^2\frac{1}{\sqrt{|t_{\gamma\gamma}|}}\frac{\sinh 2\sqrt{|t_{\gamma\gamma}|} + \sin 2\sqrt{|t_{\gamma\gamma}|}}{\cosh 2\sqrt{|t_{\gamma\gamma}|} + \cos 2\sqrt{|t_{\gamma\gamma}|}} \\
&\quad - it_\gamma^2\frac{\operatorname{sign}(t_{\gamma\gamma})}{\sqrt{|t_{\gamma\gamma}|}}\frac{\sinh 2\sqrt{|t_{\gamma\gamma}|} - \sin 2\sqrt{|t_{\gamma\gamma}|}}{\cosh 2\sqrt{|t_{\gamma\gamma}|} + \cos 2\sqrt{|t_{\gamma\gamma}|}}, \\
l_2(t_\gamma, t_{\beta\beta}, t_{\gamma\gamma}) &= 2i\frac{t_\gamma t_{\beta\beta}}{t_{\gamma\gamma}}\left(\frac{1}{\cosh\sqrt{-2it_{\gamma\gamma}}} - 1\right) \\
&= -2t_\gamma t_{\beta\beta}\frac{1}{|t_{\gamma\gamma}|}\frac{2\sinh\sqrt{|t_{\gamma\gamma}|}\sin\sqrt{|t_{\gamma\gamma}|}}{\cosh 2\sqrt{|t_{\gamma\gamma}|} + \cos 2\sqrt{|t_{\gamma\gamma}|}} \\
&\quad - 2it_\gamma t_{\beta\beta}\frac{\operatorname{sign}(t_{\gamma\gamma})}{|t_{\gamma\gamma}|}\left(1 - \frac{2\cosh\sqrt{|t_{\gamma\gamma}|}\cos\sqrt{|t_{\gamma\gamma}|}}{\cosh 2\sqrt{|t_{\gamma\gamma}|} + \cos 2\sqrt{|t_{\gamma\gamma}|}}\right), \\
l_3(t_{\beta\beta}, t_{\gamma\gamma}) &= i\frac{t_{\beta\beta}^2}{t_{\gamma\gamma}}\left(\frac{\tanh\sqrt{-2it_{\gamma\gamma}}}{\sqrt{-2it_{\gamma\gamma}}} - 1\right) \\
&= -t_{\beta\beta}^2\frac{1}{2|t_{\gamma\gamma}|^{3/2}}\frac{\sinh 2\sqrt{|t_{\gamma\gamma}|} - \sin 2\sqrt{|t_{\gamma\gamma}|}}{\cosh 2\sqrt{|t_{\gamma\gamma}|} + \cos 2\sqrt{|t_{\gamma\gamma}|}} \\
&\quad - it_{\beta\beta}^2\frac{\operatorname{sign}(t_{\gamma\gamma})}{|t_{\gamma\gamma}|} \\
&\quad \times\left(1 - \frac{1}{2\sqrt{|t_{\gamma\gamma}|}}\frac{\sinh 2\sqrt{|t_{\gamma\gamma}|} + \sin 2\sqrt{|t_{\gamma\gamma}|}}{\cosh 2\sqrt{|t_{\gamma\gamma}|} + \cos 2\sqrt{|t_{\gamma\gamma}|}}\right).
\end{aligned}$$

Now,

$$\begin{aligned}&\frac{1}{4}\operatorname{Re}[l_1(t_\gamma, t_{\gamma\gamma}) + l_2(t_\gamma, t_{\beta\beta}, t_{\gamma\gamma}) + l_3(t_{\beta\beta}, t_{\gamma\gamma})]\\ &\quad = -\frac{1}{4}\begin{pmatrix} t_\gamma \\ t_{\beta\beta}\end{pmatrix}' A(|t_{\gamma\gamma}|)\begin{pmatrix} t_\gamma \\ t_{\beta\beta}\end{pmatrix}\end{aligned}$$

and

$$\begin{aligned}&\frac{1}{4}\operatorname{Im}[l_1(t_\gamma, t_{\gamma\gamma}) + l_2(t_\gamma, t_{\beta\beta}, t_{\gamma\gamma}) + l_3(t_{\beta\beta}, t_{\gamma\gamma})] - (t_\gamma z_\gamma + t_{\beta\beta} z_{\beta\beta})\\ &\quad = -\operatorname{sign}(t_{\gamma\gamma})\frac{1}{4}\begin{pmatrix} t_\gamma \\ t_{\beta\beta}\end{pmatrix}' B(|t_{\gamma\gamma}|)\begin{pmatrix} t_\gamma \\ t_{\beta\beta}\end{pmatrix} - \begin{pmatrix} t_\gamma \\ t_{\beta\beta}\end{pmatrix}'\begin{pmatrix} z_\gamma \\ z_{\beta\beta}\end{pmatrix},\end{aligned}$$

where $A(\cdot)$ and $B(\cdot)$ are defined in the statement of Theorem 7. Using the properties of noncentral quadratic forms in normal random variables, it can be shown that

$$\begin{aligned}&\frac{1}{(2\pi)^2}\int_{\mathbb{R}^2}\exp\left(ix'z - \frac{1}{4}ix'\bar{B}x\right)\exp\left(-\frac{1}{4}x'\bar{A}x\right)dx\\ &\quad = \frac{|\bar{A} + i\bar{B}|^{-1/2}}{\pi}\exp[-z'(\bar{A}\bar{B}^{-1}\bar{A} + \bar{B})^{-1}\bar{A}\bar{B}^{-1}z]\\ &\qquad \times \exp\big[iz'\big(\bar{B}^{-1} - \bar{B}^{-1}\bar{A}(\bar{A}\bar{B}^{-1}\bar{A} + \bar{B})^{-1}\bar{A}\bar{B}^{-1}\big)z\big]\end{aligned}$$

for any $z \in \mathbb{R}^2$, any symmetric, nonsingular 2×2 matrix $\bar{B}$, and any symmetric, positive definite 2×2 matrix $\bar{A}$. As a consequence,

$$\begin{aligned}&\varkappa(t; z_\gamma, z_{\beta\beta})\\ &\quad = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varkappa^*(t_\gamma, t_{\beta\beta}, t)\exp[-i(t_\gamma z_\gamma + t_{\beta\beta} z_{\beta\beta})]\,dt_\gamma\,dt_{\beta\beta}\\ &\quad = \frac{|\bar{A} + i\bar{B}\cdot\operatorname{sign}(t)|^{-1/2}}{\pi\sqrt{\cosh\sqrt{-2it}}}\\ &\qquad \times \exp\left[-\begin{pmatrix} z_\gamma \\ z_{\beta\beta}\end{pmatrix}'(\bar{A}\bar{B}^{-1}\bar{A} + \bar{B})^{-1}\bar{A}\bar{B}^{-1}\begin{pmatrix} z_\gamma \\ z_{\beta\beta}\end{pmatrix}\right]\\ &\qquad \times \exp\left[+i\begin{pmatrix} z_\gamma \\ z_{\beta\beta}\end{pmatrix}'\big(\bar{B}^{-1} - \bar{B}^{-1}\bar{A}(\bar{A}\bar{B}^{-1}\bar{A} + \bar{B})^{-1}\bar{A}\bar{B}^{-1}\big)\right.\\ &\qquad\qquad \left.\times\begin{pmatrix} z_\gamma \\ z_{\beta\beta}\end{pmatrix}\cdot\operatorname{sign}(t)\right],\end{aligned}$$

where $\bar{A} = A(|t|)$ and $\bar{B} = B(|t|)$.

The stated result now follows because $\varkappa(t; z_\gamma, z_{\beta\beta}) = \overline{\varkappa(-t; z_\gamma, z_{\beta\beta})}$, implying that

$$f_{\mathcal{Z}_\gamma, \mathcal{Z}_{\beta\beta}, \mathcal{Q}_{\gamma\gamma}}(z_\gamma, z_{\beta\beta}, q_{\gamma\gamma}) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \varkappa(t; z_\gamma, z_{\beta\beta}) \exp[-itq_{\gamma\gamma}]\, dt$$
$$= \frac{1}{\pi}\int_0^{\infty} \operatorname{Re}\{\varkappa(t_{\gamma\gamma}; z_\gamma, z_{\beta\beta}) \exp[-itq_{\gamma\gamma}]\}\, dt.$$

Q.E.D.

REFERENCES

ABADIR, K. L., AND R. LARSSON (2001): "The Joint Moment Generating Function of Quadratic Forms in Multivariate Autoregressive Time Series," *Econometric Theory*, 17, 222–246.

ANDREWS, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

ANDREWS, D. W. K., AND J. C. MONAHAN (1992): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60, 953–966.

ANG, A., AND G. BEKAERT (2005): "Stock Return Predictability: Is It There?" Working Paper, Columbia Business School.

BASU, D. (1977): "On the Elimination of Nuisance Parameters," *Journal of the American Statistical Association*, 72, 355–366.

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

CAMPBELL, B., AND J.-M. DUFOUR (1997): "Exact Nonparametric Tests of Orthogonality and Random Walk in the Presence of a Drift Parameter," *International Economic Review*, 38, 151–173.

CAMPBELL, J. Y., AND M. YOGO (2005): "Efficient Tests of Stock Return Predictability," *Journal of Financial Economics*, forthcoming.

CAVANAGH, C. L., G. ELLIOTT, AND J. H. STOCK (1995): "Inference in Models with Nearly Integrated Regressors," *Econometric Theory*, 11, 1131–1147.

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

DE JONG, R. M., AND J. DAVIDSON (2000): "Consistency of Kernel Estimators of Heteroscedastic and Autocorrelated Covariance Matrices," *Econometrica*, 68, 407–423.

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

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

EFRON, B. (1975): "Defining the Curvature of a Statistical Problem (with Applications to Second Order Efficiency)," *The Annals of Statistics*, 3, 1189–1242.

——— (1978): "The Geometry of Exponential Families," *The Annals of Statistics*, 6, 362–376.

EFRON, B., AND D. V. HINKLEY (1978): "Assessing the Accuracy of the Maximum Likelihood Estimator: Observed versus Expected Fisher Information," *Biometrika*, 65, 457–487.

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

ELLIOTT, G. (1998): "On the Robustness of Cointegration Methods when Regressors Almost Have Unit Roots," *Econometrica*, 66, 149–158.

——— (1999): "Efficient Tests for a Unit Root when the Initial Observation Is Drawn from Its Unconditional Distribution," *International Economic Review*, 40, 767–783.

ELLIOTT, G., AND J. H. STOCK (1994): "Inference in Time Series Regression when the Order of Integration of a Regressor Is Unknown," *Econometric Theory*, 10, 672–700.
ELLIOTT, G., T. J. ROTHENBERG, AND J. H. STOCK (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, 813–836.
FEIGIN, P. D. (1986): "Asymptotic Theory of Conditional Inference for Stochastic Processes," *Stochastic Processes and Their Applications*, 22, 89–102.
FERSON, W. E., S. SARKISSIAN, AND T. SIMIN (2003): "Spurious Regressions in Financial Economics?" *Journal of Finance*, 58, 1393–1414.
HALDRUP, N., AND M. JANSSON (2005): "Improving Size and Power in Unit Root Testing," in *Palgrave Handbook of Econometrics*, forthcoming.
HANSEN, B. E. (1992): "Consistent Covariance Matrix Estimation for Dependent Heterogeneous Processes," *Econometrica*, 60, 967–972.
JANSSON, M. (2002): "Consistent Covariance Matrix Estimation for Linear Processes," *Econometric Theory*, 18, 1449–1459.
——— (2005): "Semiparametric Power Envelopes for Tests of the Unit Root Hypothesis," Manuscript, UC Berkeley.
JANSSON, M., AND M. J. MOREIRA (2004): "Optimal Inference in Regression Models with Nearly Integrated Regressors," Technical Working Paper 303, NBER.
JEGANATHAN, P. (1995): "Some Aspects of Asymptotic Theory with Applications to Time Series Models," *Econometric Theory*, 11, 818–887.
——— (1997): "On Asymptotic Inference in Linear Cointegrated Time Series Systems," *Econometric Theory*, 13, 692–745.
LANNE, M. (2002): "Testing the Predictability of Stock Returns," *Review of Economics and Statistics*, 84, 407–415.
LE CAM, L., AND G. L. YANG (2000): *Asymptotics in Statistics: Some Basic Concepts* (Second Ed.). New York: Springer-Verlag.
LEHMANN, E. L., AND J. P. ROMANO (2005): *Testing Statistical Hypotheses* (Third Ed.). New York: Springer-Verlag.
LEWELLEN, J. (2004): "Predicting Returns with Financial Ratios," *Journal of Financial Economics*, 74, 209–235.
LINDSAY, B. G., AND B. LI (1997): "On Second-Order Optimality of the Observed Fisher Information," *The Annals of Statistics*, 25, 2172–2199.
MOON, H. R., AND P. C. B. PHILLIPS (2000): "Estimation of Autoregressive Roots Near Unity Using Panel Data," *Econometric Theory*, 16, 927–997.
——— (2004): "GMM Estimation of Autoregressive Roots Near Unity with Panel Data," *Econometrica*, 72, 467–522.
MÜLLER, U. K., AND G. ELLIOTT (2003): "Tests for Unit Root and the Initial Condition," *Econometrica*, 71, 1269–1286.
NEWEY, W. K., AND K. D. WEST (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.
——— (1994): "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631–653.
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.
PFANZAGL, J. (1979): "On Optimal Median Unbiased Estimators in the Presence of Nuisance Parameters," *The Annals of Statistics*, 7, 187–193.
PHILLIPS, P. C. B. (1987a): "Time Series Regression with a Unit Root," *Econometrica*, 55, 277–301.
——— (1987b): "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535–547.
——— (1988a): "Regression Theory for Near-Integrated Time Series," *Econometrica*, 56, 1021–1043.

——— (1988b): "Weak Convergence of Sample Covariance Matrices to Stochastic Integrals via Martingale Approximations," *Econometric Theory*, 4, 528–533.

PHILLIPS, P. C. B., AND P. PERRON (1988): "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75, 335–346.

PHILLIPS, P. C. B., AND V. SOLO (1992): "Asymptotics for Linear Processes," *The Annals of Statistics*, 20, 971–1001.

POLK, C., S. THOMPSON, AND T. VUOLTEENAHO (2005): "Cross Sectional Forecasts of the Equity Premium," *Journal of Financial Economics*, forthcoming.

STAMBAUGH, R. F. (1999): "Predictive Regressions," *Journal of Financial Economics*, 54, 375–421.

STOCK, J. H. (1994): "Unit Roots, Structural Breaks and Trends," in *Handbook of Econometrics*, Vol. IV, ed. by R. F. Engle and D. L. McFadden. New York: North-Holland, 2739–2841.

——— (1997): "Cointegration, Long-Run Comovements, and Long-Horizon Forecasting," in *Advances in Economics and Econometrics: Theory and Applications, Seventh World Congress*, Vol. 3, ed. by D. Kreps and K. F. Wallis. Cambridge, U.K.: Cambridge University Press, 34–60.

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

——— (1996): "Confidence Sets in Regressions with Highly Serially Correlated Regressors," Working Paper, Harvard University.

SWEETING, T. J. (1989): "On Conditional Weak Convergence," *Journal of Theoretical Probability*, 2, 461–474.

TORUS, W., R. VALKANOV, AND S. YAN (2005): "On Predicting Stock Returns with Nearly Integrated Explanatory Variables," *Journal of Business*, 77, 937–966.

VALKANOV, R. (1999): "The Term Structure with Highly Persistent Interest Rates," Working Paper, Anderson School of Management.

VAN DER VAART, A. W. (2002): "The Statistical Work of Lucien Le Cam," *The Annals of Statistics*, 30, 631–682.

WRIGHT, J. H. (1999): "A Simple Approach to Robust Inference in a Cointegrating System," International Finance Discussion Paper 654, Board of Governors of the Federal Reserve System.

——— (2000): "Confidence Sets for Cointegrating Coefficients Based on Stationarity Tests," *Journal of Business & Economic Statistics*, 18, 211–222.