Optimal Inference in Regression Models with Nearly Integrated Regressors

Michael Jansson and Marcelo J. Moreira*
UC Berkeley and Harvard University

April 21, 2004

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

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 studying 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.[1] 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 maximizing 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 satisfying 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.

---

*e-mail: mjansson@econ.berkeley.edu (Jansson) and moreira@kuznets.fas.harvard.edu (Moreira).
We are grateful to Guido Imbens, Jim Powell, Tom Rothenberg, Paul Ruud, Jim Stock, George Tauchen, Ed Vytlacil, Mark Watson, and seminar participants at Berkeley, Iowa State, Montréal, Princeton, Stanford, and the 2003 NBER Summer Institute for comments.

[1] 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 (2003), Cavanagh, Elliott, and Stock (1995), Lanne (2002), Stock and Watson (1996), and Wright (1999, 2000).

In an attempt to provide an affirmative answer to that question, the present paper develops attainable finite sample and asymptotic efficiency bounds (power envelopes) under the assumption that the latent errors of the regression model are Gaussian white noise. In addition, the normality assumption is shown to be least favorable in the sense 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.

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 (Efron (1975, 1978)). To overcome the complications caused by this nonstandard feature, we adopt an approach similar to the conventional approach in cross section econometrics, where inference in regression models with random, exogenous regressors is conducted conditional on the regressors. 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 whose distribution does not depend on the parameter of interest. 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), Le Cam and Yang (2000)) to show that the limiting experiment associated with our regression model inherits the statistical properties of the finite sample model.

The bulk of this paper studies 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.[2] 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, hereby enabling us to introduce the main ideas of the paper without the use of asymptotics. 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. We illustrate this point by showing that results extend in a straightforward way to a cointegration-type model accommodating correlation between the (potentially) serially correlated

[2]Recent papers studying predictive regressions include Ang and Bekaert (2003), Campbell and Yogo (2003), Ferson, Sarkissian, and Simin (2003), Lanne (2002), Lewellen (2002), Polk, Thompson, and Vuolteenaho (2004), and Torus, Valkanov, and Yan (2001). See also Stambaugh (1999) and the references therein.

error term of the equation of interest and current (and lagged) values of the nearly integrated regressor.

Section 2 introduces the model used in the bulk of the paper. 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 feasible testing procedures, 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 studies a cointegration-type model and shows that the limiting experiment of that model coincides with the limiting experiment of the model of Sections 3 and 4, implying that the asymptotic optimality results of Section 4 can be extended to cointegration-type models. Mathematical derivations have been relegated to two Appendices.

2. Predictive Regression Model

Following Cavanagh, Elliott, and Stock (1995), Sections 3-5 consider a bivariate model in which the observed data $\left\{(y_t, x_t)' : 1 \leq t \leq T\right\}$ is generated by the recursive system

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

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

where[3]

A1. $v_0^x = 0$.

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

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

By design, the model $(1) - (2)$ captures the salient features of the predictive regression model, a popular model in empirical finance.[4] Our goal is to construct one- and two-sided tests of the null hypothesis $\beta = \beta_0$, treating α and γ as unknown nuisance parameters. Regarding the nuisance parameter γ, particular attention will 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" to (but not necessarily equal to) unity.

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

[4]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 $\beta = 0$.

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 feasible testing procedures that enjoy asymptotic optimality properties under the assumptions of Section 4 and are asymptotically valid under A1-A3 and local-to-unity asymptotics.

3. Optimal Inference with Gaussian Errors: Finite Sample Theory

Consider the Gaussian model

$$\begin{aligned} y_t &= \alpha + \beta x_{t-1} + \varepsilon_t^y, &(3)\\ x_t &= \gamma x_{t-1} + \varepsilon_t^x, &(4)\end{aligned}$$

where

A1*. $x_0 = 0$.

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

Under these assumptions, the log likelihood function $L(\cdot)$ satisfies

$$\begin{aligned} -2L(\alpha,\beta,\gamma) &= \sigma_{yy.x}^{-1} \sum_{t=1}^{T} \left[y_t - \alpha - \beta x_{t-1} - \sigma_{xx}^{-1}\sigma_{xy}(x_t - \gamma x_{t-1}) \right]^2 \\ &\quad + \sigma_{xx}^{-1} \sum_{t=1}^{T} (x_t - \gamma x_{t-1})^2, \end{aligned}$$

where $\sigma_{yy.x} = \sigma_{yy} - \sigma_{xx}^{-1}\sigma_{xy}^2$ and Σ has been partitioned conformably with ε_t.

Testing problems involving β are invariant under transformations of the form

$$(y_t, x_t) \to (y_t + a, x_t), \qquad a \in \mathbb{R}.$$

The statistic $[(y_1, \ldots, y_T)\,\iota_\perp, (x_1, \ldots, x_T)]'$ is a maximal invariant under this group of transformations, where $\iota_\perp$ is a $T \times (T-1)$ matrix whose columns form an orthonormal basis for the set of T-vectors that sum to zero (i.e. are orthogonal to $\iota = (1, \ldots, 1)'$).

Apart from an additive constant, the log density of this maximal invariant is given by $\mathcal{L}(\beta,\gamma) = \max_\alpha L(\alpha,\beta,\gamma)$.

Now,

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

where

$$\begin{aligned}
S_\beta &= \sigma_{yy.x}^{-1}\sum_{t=1}^{T} x_{t-1}^{\mu}\left(y_t - \sigma_{xx}^{-1}\sigma_{xy}x_t\right), \qquad x_{t-1}^{\mu} = x_{t-1} - T^{-1}\sum_{s=1}^{T} x_{s-1}, \\
S_\gamma &= \sigma_{xx}^{-1}\sum_{t=1}^{T} x_{t-1}x_t - \sigma_{xx}^{-1}\sigma_{xy}S_\beta, \\
S_{\beta\beta} &= \sigma_{yy.x}^{-1}\sum_{t=1}^{T} x_{t-1}^{\mu 2}, \\
S_{\gamma\gamma} &= \sigma_{xx}^{-1}\sum_{t=1}^{T} x_{t-1}^{2}.
\end{aligned}$$

It follows from (6) that $S = (S_\beta, S_\gamma, S_{\beta\beta}, S_{\gamma\gamma})'$ is a sufficient statistic for the distribution of the maximal invariant. When 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 sense of Efron (1975, 1978)), the minimal sufficient statistic being of dimension four whereas the parameter vector (β,γ) is two-dimensional. (A precise statement is provided in Lemma 1 (a) below.) As a consequence, conventional optimality theory for exponential families (e.g., Lehmann (1994)) does not apply. 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 are sufficient to remove the statistical curvature from the problem.

Since 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 sense of Basu (1977)). A conditionality argument therefore suggests that inference on β should be based on the conditional distribution of (S_β, S_γ) given $(S_{\beta\beta}, S_{\gamma\gamma})$. 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) – (4) and suppose A1*-A2* hold.*

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

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

where $f_S^0(\cdot)$ is a density of S when $\beta = \gamma = 0$ and $K_S(\cdot)$ 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

$$\begin{aligned} f_{S_\beta,S_\gamma|S_{\beta\beta},S_{\gamma\gamma}}(s_\beta, s_\gamma|s_{\beta\beta}, s_{\gamma\gamma};\beta,\gamma) &= g_S(\beta,\gamma)\, h_S(s_\beta, s_\gamma|s_{\beta\beta}, s_{\gamma\gamma}) \\ &\quad \times \exp(\beta s_\beta + \gamma s_\gamma) \end{aligned}$$

for some functions $g_S(\cdot)$ and $h_S(\cdot)$.

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 (1994) to find UMP conditionally unbiased (UMPCU) tests for one- and two-sided testing problems concerning β.

First, consider the one-sided testing problem[5]

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

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

$$\begin{aligned} E_{\beta_0,\gamma}\left[\phi(S)\,|S_{\beta\beta}, S_{\gamma\gamma}\right] &\le \eta \qquad \forall\gamma \in \mathbb{R}, \\ E_{\beta,\gamma}\left[\phi(S)\,|S_{\beta\beta}, S_{\gamma\gamma}\right] &\ge \eta \qquad \forall\beta > \beta_0, \gamma \in \mathbb{R}. \end{aligned}$$

[5]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.

Any conditionally unbiased test is conditionally similar in the sense that

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

On the other hand, the properties of exponential families (e.g., Lehmann (1994, Theorem 2.9)) can be used to show that a test is UMP among conditionally similar tests only if it is unbiased. As a consequence, a test is UMPCU if and only if it is UMP among tests satisfying (7).

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

$$\phi_\eta^*\left(s\right)=1\left[s_\beta>C_\eta\left(s_\gamma,s_{\beta\beta},s_{\gamma\gamma}\right)\right], \tag{8}$$

where $1[\cdot]$ is the indicator function, $C_\eta(\cdot)$ satisfies

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

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) below that the test associated with $\phi_\eta^*(\cdot)$ is the UMPCU level η test.

Next, consider the two-sided testing problem

$$H_0:\beta=\beta_0 \qquad \text{vs.} \qquad H_1:\beta\neq\beta_0.$$

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

$$\begin{aligned} E_{\beta_0,\gamma}\left[\phi\left(S\right)|S_{\beta\beta},S_{\gamma\gamma}\right] &\leq \eta \qquad \forall\gamma\in\mathbb{R}, \\ E_{\beta,\gamma}\left[\phi\left(S\right)|S_{\beta\beta},S_{\gamma\gamma}\right] &\geq \eta \qquad \forall\beta\neq\beta_0,\gamma\in\mathbb{R}. \end{aligned}$$

It follows from Lemma 1 (b) and the properties of exponential families (e.g., Lehmann (1994, Theorem 2.9)) that a level η test is conditionally unbiased only if its test function $\phi(\cdot)$ satisfies

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

In turn, this condition holds if and only if

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

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

Theorem 2 (b) shows that a level η test is UMPCU if its test function is given by

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

where $\underline{C}_\eta\left(\cdot\right)$ and $\overline{C}_\eta\left(\cdot\right)$ satisfy

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

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

Theorem 2. *Let* $\left\{(y_t,x_t)'\right\}$ *be generated by* $(3)-(4)$ *and suppose A1*-A2* hold.*

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

$$E_{\beta,\gamma}\phi\left(S\right)\leq E_{\beta,\gamma}\phi_\eta^{*}\left(S\right)$$

for every $\beta\geq\beta_0$ *and every* $\gamma\in\mathbb{R}$.

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

$$E_{\beta,\gamma}\phi\left(S\right)\leq E_{\beta,\gamma}\phi_\eta^{**}\left(S\right)$$

for every $\beta\in\mathbb{R}$ *and every* $\gamma\in\mathbb{R}$.

Remarks. (i) Qualitatively similar results can be obtained if the intercept term α is known. For concreteness, suppose $\alpha = 0$. In that case, the log likelihood function $L^0(\cdot)$ satisfies

$$L^0(\beta,\gamma) - L^0(0,0) = \beta S_\beta^0 + \gamma S_\gamma^0 - \frac{1}{2}\left[\gamma^2 + \left(\beta - \sigma_{xx}^{-1}\sigma_{xy}\gamma\right)^2 \sigma_{xx}\sigma_{yy.x}^{-1}\right] S_{\gamma\gamma},$$

where $S_\beta^0 = \sigma_{yy.x}^{-1}\sum_{t=1}^T x_{t-1}\left(y_t - \sigma_{xx}^{-1}\sigma_{xy}x_t\right)$ and $S_\gamma^0 = \sigma_{xx}^{-1}\sum_{t=1}^T x_{t-1}x_t - \sigma_{xx}^{-1}\sigma_{xy}S_\beta^0$. The distribution of the minimal sufficient statistic $\left(S_\beta^0, S_\gamma^0, S_{\gamma\gamma}\right)'$ is a curved exponential family, but the curvature disappears once we condition the specific ancillary $S_{\gamma\gamma}$. Optimality results analogous to Theorem 2 are therefore readily available, but we omit a precise statement because the case where α is known appears to be of limited empirical relevance.

(ii) 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. By the properties of exponential families (e.g., Lemma 1 (b) and Lehmann (1994, Theorem 4.1)), the condition

$$E_{\beta_0,\gamma}\left[\phi(S)\,|S_{\beta\beta}, S_{\gamma\gamma}\right] = \eta \qquad \forall\gamma\in\Gamma$$

is equivalent to (7) whenever Γ contains an open interval. 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.

(iii) Any conditionally similar test is similar in the sense that

$$E_{\beta_0,\gamma}\phi(S) = \eta \qquad \forall\gamma\in\mathbb{R}.$$

Using the results of Wijsman (1958), it can be shown that the converse does not hold. As a consequence, the class of similar tests is strictly greater than the class of conditionally similar tests. It is an open question whether the test based on $\phi_\eta^*(\cdot)$ is UMP within the class of unbiased tests.

(iv) Studying a more general (but closely related) model, Stock and Watson (1996) investigated tests maximizing 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 $\phi(\cdot)$ satisfying

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

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 satisfying (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 the weighting function $G(\cdot)$ (and Γ), implying that no UMP test exists among tests satisfying (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.

4. Optimal Inference with Gaussian Errors: Asymptotic Theory

This section develops an asymptotic counterpart of Theorem 2. Whereas the finite sample results of the previous section require no specific assumptions about the range of possible values of the persistence parameter γ (cf. remark (ii) following Theorem 2), the asymptotic properties of the model $(1)-(2)$ 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 likelihood ratios (e.g., Choi, Hall, and Shick (1996)) is applicable. 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 constant c.

Because the statistical curvature does not vanish when $\gamma = \gamma_T(c)$, testing problems concerning β exhibit nonstandard large-sample properties under 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)).[6] Procedures that are asymptotically valid when γ is local-to-unity have been proposed by Campbell and Dufour (1997), Campbell and Yogo (2003), Cavanagh, Elliott, and Stock (1995), and Lanne (2002), but all of these existing testing procedures

[6] 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 non-robustness result can also be used to establish the invalidity of two-step procedures based on unit root pretests (e.g., Stock and Watson (1996)).

are asymptotically biased.[7] By developing an asymptotic counterpart of Theorem 2, this section demonstrates by example that (non-trivial) 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 $\beta = \beta_T(b) = \beta_0 + T^{-1}\sigma_{xx}^{-1/2}\sigma_{yy.x}^{1/2}b$, b being a fixed constant. 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:

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

where

$$\begin{aligned} R_\beta &= \sigma_{xx}^{-1/2}\sigma_{yy.x}^{-1/2}T^{-1}\sum_{t=1}^{T} x_{t-1}^{\mu}\left(y_t - \beta_0 x_{t-1} - \sigma_{xx}^{-1}\sigma_{xy}\Delta x_t\right), \\ R_\gamma &= \sigma_{xx}^{-1}T^{-1}\sum_{t=1}^{T} x_{t-1}\Delta x_t - \frac{\rho}{\sqrt{1-\rho^2}}R_\beta, \qquad \rho = \sigma_{xy}\sigma_{xx}^{-1/2}\sigma_{yy}^{-1/2}, \\ R_{\beta\beta} &= \sigma_{xx}^{-1}T^{-2}\sum_{t=1}^{T} x_{t-1}^{\mu 2}, \\ R_{\gamma\gamma} &= \sigma_{xx}^{-1}T^{-2}\sum_{t=1}^{T} x_{t-1}^{2}. \end{aligned}$$

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 of Section 3, it turns out to be convenient to work with R. The following lemmas give some useful properties of its limiting distribution.

---

[7] The tests proposed by Campbell and Dufour (1997), Campbell and Yogo (2003) 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)).

Lemma 3. *Let* $\{(y_t, x_t)'\}$ *be generated by* (3) − (4) *and suppose A1*-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*

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

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

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

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* $W_{x,c}$ *an Ornstein-Uhlenbeck process satisfying the stochastic differential equation* $dW_{x,c}(r) = cW_{x,c}(r)\, dr + dW_x(r)$ *with initial condition* $W_{x,c}(0) = 0$.

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*

$$\begin{aligned}
f_{\mathcal{R}}(r; b, c, \rho) &= K_{\mathcal{R}}(b,c,\rho)\, f_{\mathcal{R}}^0(r;\rho)\\
&\quad \times \exp\left[br_{\beta} + cr_{\gamma} - \frac{1}{2}\left(b - \frac{\rho}{\sqrt{1-\rho^2}}c\right)^2 r_{\beta\beta} - \frac{1}{2}c^2 r_{\gamma\gamma}\right],
\end{aligned}$$

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

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

$$f_{\mathcal{R}_{\beta}^{\rho},\mathcal{R}_{\gamma}^{\rho}|\mathcal{R}_{\beta\beta},\mathcal{R}_{\gamma\gamma}}\left(r_{\beta},r_{\gamma}|r_{\beta\beta},r_{\gamma\gamma};b,c,\rho\right) = g_{\mathcal{R}}\left(b,c,\rho\right)h_{\mathcal{R}}\left(r_{\beta},r_{\gamma}|r_{\beta\beta},r_{\gamma\gamma};\rho\right) \times \exp\left(br_{\beta}+cr_{\gamma}\right)$$

for some functions $g_{\mathcal{R}}\left(\cdot\right)$ *and* $h_{\mathcal{R}}\left(\cdot\right)$.

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 $f_{\mathcal{R}}^{0}\left(\cdot;\rho\right)$ 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 to obtain an integral representation of $f_{\mathcal{R}}^{0}\left(\cdot;\rho\right)$ which is useful for computational purposes.

In view of Lemma 4, the functional $\mathcal{R}^{\rho}\left(b,c\right)$ 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 $\mathcal{R}_{\beta\beta}\left(c\right)$ and $\mathcal{R}_{\gamma\gamma}\left(c\right)$ are specific ancillaries in the limiting experiment. It therefore seems plausible that appropriately constructed asymptotic counterparts of $\phi_{\eta}^{*}\left(\cdot\right)$ and $\phi_{\eta}^{**}\left(\cdot\right)$ should enjoy asymptotic optimality properties analogous to the finite sample optimality properties enjoyed by $\phi_{\eta}^{*}\left(\cdot\right)$ and $\phi_{\eta}^{**}\left(\cdot\right)$. 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 $\pi\left(\cdot\right)$ such that the probability of rejecting H_0 equals $\pi\left(r\right)$ whenever $R = r$. This test function satisfies $\phi = \pi \circ \zeta$, where $\phi\left(\cdot\right)$ is the test function associated with S and $\zeta\left(\cdot\right)$ is any mapping such that $\zeta\left(S\right) = R$ (with probability one).

Asymptotic optimality results for the one-sided testing problem

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

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

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

where $C_b\left(\mathbb{R}^2\right)$ denotes the set of bounded, continuous, real-valued functions on $\mathbb{R}^2$. The advantage of this characterization of conditional 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 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 similar (at level η) if

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

In perfect analogy with Theorem 2 (a), Theorem 5 (a) below 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

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

where $\mathcal{C}_\eta(\cdot)$ is the continuous function satisfying

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

and $\mathcal{R}^\rho=\left(\mathcal{R}_\beta,\mathcal{R}_\gamma^\rho,\mathcal{R}_{\beta\beta},\mathcal{R}_{\gamma\gamma}\right)'=\mathcal{R}^\rho\left(0,0\right)$.[8]

An attainable efficiency bound for the two-sided testing problem

$$H_0: b=0 \qquad \text{vs.} \qquad H_1: b\neq 0,$$

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

[8] 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 11 of Appendix A. The domain of $\mathcal{C}_\eta(\cdot)$ is a set $\mathbb{S}\subseteq\mathbb{R}^4$ satisfying $\Pr\left[\left(\mathcal{R}_\gamma^\rho,\mathcal{R}_{\beta\beta},\mathcal{R}_{\gamma\gamma};\rho\right)\in\mathbb{S}\right]=1$.

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

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

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

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

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

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

Theorem 5. *Let* $\left\{ (y_t, x_t)' \right\}$ *be generated by* $(3) - (4)$ *and suppose A1*-A2* hold.*

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

$$\begin{aligned} \overline{\lim}_{T\to\infty} E_{\beta_T(b),\gamma_T(c)} \pi_T(R) &\leq \lim_{T\to\infty} E_{\beta_T(b),\gamma_T(c)} \pi_\eta^{*}(R;\rho) \\ &= E\left[\pi_\eta^{*}\left(\mathcal{R}^\rho(b,c);\rho\right)\right] \end{aligned}$$

for every $b \geq 0$ *and every* $c \in \mathbb{R}$.

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

$$\begin{aligned} \overline{\lim}_{T\to\infty} E_{\beta_T(b),\gamma_T(c)} \pi_T(R) &\leq \lim_{T\to\infty} E_{\beta_T(b),\gamma_T(c)} \pi_\eta^{**}(R;\rho) \\ &= E\left[\pi_\eta^{**}\left(\mathcal{R}^\rho(b,c);\rho\right)\right] \end{aligned}$$

for every $b \in \mathbb{R}$ *and every* $c \in \mathbb{R}$.

In view of Theorem 5, the maximal attainable (by tests satisfying the restrictions we impose) local asymptotic power against the local alternative $\beta = \beta_T(b)$ depends on the values of two nuisance parameters, the persistence parameter c and ρ, the coefficient of correlation computed from Σ.

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

$$\begin{aligned}\varphi_\eta^*(b,c;\rho) &= E\left[\pi_\eta^*\left(\mathcal{R}^\rho(b,c);\rho\right)\right] \\ &= \Pr\left[\mathcal{R}_\beta^\rho(b,c) > \mathcal{C}_\eta\left(\mathcal{R}_\gamma^\rho(b,c), \mathcal{R}_{\beta\beta}(c), \mathcal{R}_{\gamma\gamma}(c);\rho\right)\right],\end{aligned}$$

$$\begin{aligned}\varphi_\eta^{**}(b,c;\rho) &= E\left[\pi_\eta^{**}\left(\mathcal{R}^\rho(b,c);\rho\right)\right] \\ &= \Pr\left[\mathcal{R}_\beta^\rho(b,c) < \underline{\mathcal{C}}_\eta\left(\mathcal{R}_\gamma^\rho(b,c), \mathcal{R}_{\beta\beta}(c), \mathcal{R}_{\gamma\gamma}(c);\rho\right)\right] \\ &\quad + \Pr\left[\mathcal{R}_\beta^\rho(b,c) > \overline{\mathcal{C}}_\eta\left(\mathcal{R}_\gamma^\rho(b,c), \mathcal{R}_{\beta\beta}(c), \mathcal{R}_{\gamma\gamma}(c);\rho\right)\right].\end{aligned}$$

The next section proposes feasible 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) Theorem 5 (a) remains true if the requirement (19) is replaced with the following condition:

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

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 (1994, Theorem 4.1)) 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 $\pi(\cdot)$ satisfying

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

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

(ii) In view of remark (i), the function $\varphi_\eta^*(\cdot)$ constitutes a suitable power envelope also if c is treated as an unknown, non-positive 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 was 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. Feasible Inference

This section considers the general case where $\{(y_t, x_t)'\}$ is generated by (1) − (2), assumptions A1-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 are well behaved more generally. This will be accomplished by constructing a statistic $\hat{R}$, which is asymptotically equivalent to R under the assumptions of the previous section and is well behaved more generally.

Let $x_0 = x_1$, $\hat{v}_0^x = 0$, and define $\hat{v}_t^x = x_t - x_1$ and $\hat{x}_{t-1}^\mu = x_{t-1} - T^{-1}\sum_{s=1}^T x_{s-1}$ (for $t = 1, \ldots, T$). Let

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

denote the long-run variance of $(\varepsilon_t^y, \psi(L)\,\varepsilon_t^x)'$ and let $\hat{\Omega}$ be a consistent estimator thereof. Finally, let

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

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 $\hat{R}=\left(\hat{R}_{\beta},\hat{R}_{\gamma},\hat{R}_{\beta\beta},\hat{R}_{\gamma\gamma}\right)'$ is invariant under transformations of the form $(y_t,x_t)\to(y_t+a,x_t)$, where $a\in\mathbb{R}$.[9] Under the assumptions of Section 4, $\hat{R}$ is asymptotically equivalent to R. More generally, we have:

Theorem 6. *Let $\left\{(y_t,x_t)'\right\}$ be generated by $(1)-(2)$, suppose A1-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,*

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

and

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

In view of Theorem 6, the Gaussian 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).

[9] 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 $m_x\in\mathbb{R}$.

Construction of consistent long-run variance estimators is a problem that has received considerable attention and there is no shortage of estimators satisfying the high-level assumption $\hat{\Omega} \to_p \Omega$ of Theorem 6.[10] To describe a consistent kernel estimator of Ω, let $\hat{u}_t = \left(y_t^\mu - \hat{\beta}\hat{x}_{t-1}^\mu, \hat{v}_t^x - \hat{\gamma}\hat{v}_{t-1}^x\right)'$, where $y_t^\mu = y_t - T^{-1}\sum_{s=1}^T y_s$, $\hat{\beta} = \left(\sum_{t=1}^T \hat{x}_{t-1}^{\mu 2}\right)^{-1}\sum_{t=1}^T \hat{x}_{t-1}^\mu y_t$, and $\hat{\gamma} = \left(\sum_{t=1}^T \hat{v}_{t-1}^{x2}\right)^{-1}\sum_{t=1}^T \hat{v}_{t-1}^x \hat{v}_t^x$. 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

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

To implement the tests based on $\pi_\eta^*\left(\hat{R};\hat{\rho}\right)$ and $\pi_\eta^{**}\left(\hat{R};\hat{\rho}\right)$, 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 $\left(\mathcal{R}_\gamma^\rho, \mathcal{R}_{\beta\beta}, \mathcal{R}_{\gamma\gamma}\right)$. That distribution is non-standard 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 $\mathcal{R}^\rho$.

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

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

where

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

[10] 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).

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

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

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

6. Cointegration

A closely related problem to that considered in Sections 2-5 is the problem of conducting inference on long run economic relationships between highly persistent variables. The concept of cointegration (Engle and Granger (1987)) provides a useful framework for doing inference when the individual time series are known to have unit roots. A celebrated result from cointegration analysis shows that standard Wald-type test statistics based on asymptotically efficient estimators of cointegrating coefficients have asymptotic χ^2 null distributions when the individual time series have unit roots.[11] Elliott (1998) has shown that the validity of the χ^2 result depends crucially on the validity of the unit root assumption. Indeed, the χ^2 result breaks down when the largest autoregressive root is modeled as local to unity.[12]

Recognizing the importance of Elliott's (1998) result, Stock and Watson (1996) and Wright (1999, 2000) have developed cointegration-type inference procedures that

[11] The class of asymptotically efficient (in the sense of Phillips (1991a) and Saikkonen (1991)) estimators includes the estimators proposed by Johansen (1988, 1991), Park (1992), Phillips and Hansen (1990), Phillips (1991a, 1991b), Saikkonen (1991, 1992), and Stock and Watson (1993). For a review, see Watson (1994).

[12] Kauppi (2004) establishes a similar non-robustness result about Phillips's (1995) fully modified VAR method.

are asymptotically valid in the absence of (exact) unit root assumptions on the individual time series. Although the test proposed by Stock and Watson (1996) enjoys certain optimality properties, it is asymptotically biased, as are the tests due to Wright (1999, 2000). In contrast, this section shows that results analogous to Theorems 5 and 6 are available for cointegration-type models accommodating correlation between the (potentially) serially correlated error term of the equation of interest and current (and lagged) values of the nearly integrated regressor.

6.1. Optimal Inference with Gaussian Errors.

Consider the model

$$\begin{aligned} y_t &= \alpha + \beta x_t + \varepsilon_t^y, && (26)\\ x_t &= \gamma x_{t-1} + \varepsilon_t^x, && (27) \end{aligned}$$

where

B1*. $x_0 = 0$.

B2*. $\varepsilon_t = (\varepsilon_t^y, \varepsilon_t^x)' \sim i.i.d.\ \mathcal{N}(0, \Sigma)$, where Σ is a known, positive definite matrix.

Utilizing the methods developed in Sections 3-4, it is straightforward to develop asymptotically unbiased cointegration-type testing procedures with demonstrable optimality properties.

When $\left\{(y_t, x_t)'\right\}$ is generated by $(26) - (27)$, testing problems involving β are invariant under transformations of the form $(y_t, x_t) \to (y_t + a, x_t)$, where $a \in \mathbb{R}$. The log density $\mathcal{L}^{CI}(\cdot)$ of the maximal invariant $[(y_1, \ldots, y_T)\,\iota_\perp, (x_1, \ldots, x_T)]'$ satisfies

$$\begin{aligned} &\mathcal{L}^{CI}(\beta, \gamma) - \mathcal{L}^{CI}(0,0) \\ &= \beta S_\beta^{CI} + \gamma S_\gamma^{CI} - \frac{1}{2}\begin{pmatrix} \beta \\ -\sigma_{xx}^{-1}\sigma_{xy}\gamma \end{pmatrix}' S_{\beta\beta}^{CI} \begin{pmatrix} \beta \\ -\sigma_{xx}^{-1}\sigma_{xy}\gamma \end{pmatrix} - \frac{1}{2}\gamma^2 S_{\gamma\gamma}^{CI}, \end{aligned} \tag{28}$$

where

$$
\begin{aligned}
S_{\beta}^{CI} &= \sigma_{yy.x}^{-1}\sum_{t=1}^{T} x_t^m\left(y_t-\sigma_{xx}^{-1}\sigma_{xy}x_t\right), \qquad x_t^m = x_t - T^{-1}\sum_{s=1}^{T}x_s, \\
S_{\gamma}^{CI} &= \sigma_{xx}^{-1}\sum_{t=1}^{T} x_{t-1}x_t - \sigma_{xx}^{-1}\sigma_{xy}\sigma_{yy.x}^{-1}\sum_{t=1}^{T} x_{t-1}^{\mu}\left(y_t-\sigma_{xx}^{-1}\sigma_{xy}x_t\right), \\
S_{\beta\beta}^{CI} &= \begin{pmatrix} \sigma_{yy.x}^{-1}\sum_{t=1}^{T}x_t^{m2} & \sigma_{yy.x}^{-1}\sum_{t=1}^{T}x_t^m x_{t-1}^{\mu} \\ \sigma_{yy.x}^{-1}\sum_{t=1}^{T}x_t^m x_{t-1}^{\mu} & \sigma_{yy.x}^{-1}\sum_{t=1}^{T}x_{t-1}^{\mu 2}\end{pmatrix}, \\
S_{\gamma\gamma}^{CI} &= \sigma_{xx}^{-1}\sum_{t=1}^{T}x_{t-1}^2,
\end{aligned}
$$

$x_t^m = x_t - T^{-1}\sum_{s=1}^{T}x_s$, $x_{t-1}^{\mu} = x_{t-1} - T^{-1}\sum_{s=1}^{T}x_{s-1}$, and Σ has been partitioned in conformity with ε_t.

In view of (28) , the model admits a minimal sufficient statistic whose distribution is a curved exponential family. Moreover, the curvature disappears when we condition on specific ancillaries, but in this case the dimension of the specific ancillary is greater than two, suggesting that inference is more complicated in this case than in the case of the predictive regression model of Sections 3-4. Because the individual elements of $S_{\beta\beta}^{CI}$ are asymptotically equivalent (in the appropriate sense), the source of this potential complication vanishes asymptotically. In fact, it turns out that the limiting experiment associated with the cointegration model of this section is exactly the same as the limiting experiment associated with the predictive regression model.

Lemma 8. *Let* $\left\{(y_t, x_t)'\right\}$ *be generated by* $(26)-(27)$ *and suppose B1*-B2* hold.*

(a) For every bounded sequence $\left\{(b_T, c_T)'\right\}$,

$$
\mathcal{L}^{CI}\left[\beta_T(b_T), \gamma_T(c_T)\right] - \mathcal{L}^{CI}\left[\beta_T(0), \gamma_T(0)\right]
$$

$$
= b_T R_{\beta}^{CI} + c_T R_{\gamma}^{CI} - \frac{1}{2}\left(b_T - \frac{\rho}{\sqrt{1-\rho^2}}c_T\right)^2 R_{\beta\beta}^{CI} - \frac{1}{2}c_T^2 R_{\gamma\gamma}^{CI} + o_{p_0}(1),
$$

where

$$
\begin{aligned}
R_{\beta}^{CI} &= \sigma_{xx}^{-1/2}\sigma_{yy.x}^{-1/2}T^{-1}\sum_{t=1}^{T}x_{t-1}^{\mu}\left(y_t-\beta_0 x_t-\sigma_{xx}^{-1}\sigma_{xy}\Delta x_t\right), \\
R_{\gamma}^{CI} &= \sigma_{xx}^{-1}T^{-1}\sum_{t=1}^{T}x_{t-1}\Delta x_t-\frac{\rho}{\sqrt{1-\rho^2}}R_{\beta}^{CI}, \qquad \rho=\sigma_{xy}\sigma_{xx}^{-1/2}\sigma_{yy}^{-1/2}, \\
R_{\beta\beta}^{CI} &= \sigma_{xx}^{-1}T^{-2}\sum_{t=1}^{T}x_{t-1}^{\mu 2}, \\
R_{\gamma\gamma}^{CI} &= \sigma_{xx}^{-1}T^{-2}\sum_{t=1}^{T}x_{t-1}^{2},
\end{aligned}
$$

and "$o_{p_0}(1)$" is shorthand for "$o_p(1)$ when $(\beta,\gamma)=(\beta_0,1)$".

(b) 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 $R^{CI}=\left(R_{\beta}^{CI},R_{\gamma}^{CI},R_{\beta\beta}^{CI},R_{\gamma\gamma}^{CI}\right)'\to_d\mathcal{R}^{\rho}(b,c)$ as $T\to\infty$, where $\mathcal{R}^{\rho}(b,c)$ is defined as in Section 4.

It follows from Lemma 8 that R^{CI} is an asymptotically sufficient statistic, implying that there is no loss of attainable asymptotic power from confining attention to tests based on R^{CI}. Because the limiting representation of R^{CI} coincides with the limiting representation of the minimal sufficient statistic R of the predictive regression model, the limiting experiment of the cointegration model coincides with the limiting experiment of the predictive regression model. As a consequence, the asymptotic optimality theory of Section 4 carries over in an obvious way.

Theorem 9. *Let $\{(y_t,x_t)'\}$ be generated by $(26)-(27)$ and suppose B1*-B2* hold.*

(a) If $\{\pi_T(\cdot)\}$ satisfies

$$
\lim_{T\to\infty}E_{\beta_T(0),\gamma_T(c)}\left[\left(\pi_T\left(R^{CI}\right)-\eta\right)g\left(R_{\beta\beta}^{CI},R_{\gamma\gamma}^{CI}\right)\right]=0 \tag{29}
$$

for every $c\in\mathbb{R}$ and every $g\in C_b(\mathbb{R}^2)$, then

$$
\begin{aligned}
\overline{\lim}_{T\to\infty}E_{\beta_T(b),\gamma_T(c)}\pi_T\left(R^{CI}\right) &\leq \lim_{T\to\infty}E_{\beta_T(b),\gamma_T(c)}\pi_{\eta}^{*}\left(R^{CI};\rho\right) \\
&= E\left[\pi_{\eta}^{*}\left(\mathcal{R}^{\rho}(b,c);\rho\right)\right]
\end{aligned}
$$

for every $b \geq 0$ and every $c \in \mathbb{R}$.

(b) If $\{\pi_T(\cdot)\}$ satisfies (29) and

$$\lim_{T\to\infty} E_{\beta_T(0),\gamma_T(c)}\left[\left(\pi_T\left(R^{CI}\right)-\eta\right)R_\beta^{CI}\cdot g\left(R_{\beta\beta}^{CI},R_{\gamma\gamma}^{CI}\right)\right]=0 \tag{30}$$

for every $c \in \mathbb{R}$ and every $g \in C_b(\mathbb{R}^2)$, then

$$\begin{aligned}\overline{\lim}_{T\to\infty}E_{\beta_T(b),\gamma_T(c)}\pi_T\left(R^{CI}\right) &\leq \lim_{T\to\infty}E_{\beta_T(b),\gamma_T(c)}\pi_\eta^{**}\left(R^{CI};\rho\right)\\ &= E\left[\pi_\eta^{**}\left(\mathcal{R}^\rho(b,c);\rho\right)\right]\end{aligned}$$

for every $b \in \mathbb{R}$ and every $c \in \mathbb{R}$.

Remark. The proof of Theorem 9 is based on Lemma 8 and the methods developed in Section 4. In an attempt to shed additional light on the asymptotic equivalence between the model $(26)-(27)$ and the predictive regression model $(3)-(4)$, we briefly sketch an alternative method of proof. Replacing y_t with $y_t-\beta_0 x_t$ if necessary, there is no loss of generality from assuming that the null hypothesis of interest is $H_0:\beta=0$. As in Jeganathan (1997), the model can be written as

$$y_t = \alpha+\tilde{\beta}x_{t-1}+\tilde{\varepsilon}_t^y, \tag{31}$$

$$x_t = \gamma x_{t-1}+\varepsilon_t^x, \tag{32}$$

where $\tilde{\beta}=\beta\gamma$, $\tilde{\varepsilon}_t^y=\varepsilon_t^y+\beta\varepsilon_t^x$, $x_0=0$, and

$$\begin{pmatrix}\tilde{\varepsilon}_t^y\\ \varepsilon_t^x\end{pmatrix}\sim i.i.d.\ \mathcal{N}\left(0,\tilde{\Sigma}\right),\qquad \tilde{\Sigma}=\begin{pmatrix}1&\beta\\0&1\end{pmatrix}\Sigma\begin{pmatrix}1&\beta\\0&1\end{pmatrix}'.$$

If $\tilde{\Sigma}$ is viewed as a fixed parameter (rather than a function of β (and Σ)), this reparameterized model is in the form considered in Sections 3-4 and the null hypothesis can be expressed as a restriction on $\tilde{\beta}$ (because $\beta=0$ if and only if $\tilde{\beta}=0$). Because $\tilde{\Sigma}=\Sigma+O(T^{-1})$ under contiguous alternatives (i.e. when $\beta=O(T^{-1})$), it can be shown that the dependence of $\tilde{\Sigma}$ on β is asymptotically negligible (in the appropriate sense), implying that an alternative proof of Theorem 9 can proceed by applying Theorem 5 to the reparameterized model $(31)-(32)$.

6.2. Feasible Inference. Suppose $\left\{(y_t, x_t)'\right\}$ is generated by the following generalization of the model considered in the previous subsection:

$$y_t = \alpha + \beta x_t + u_t^y, \tag{33}$$
$$x_t = \mu_x + v_t^x, \qquad v_t^x = \gamma v_{t-1}^x + u_t^x, \tag{34}$$

where

$$u_t = (u_t^y, u_t^x)' = \Psi(L)\,\varepsilon_t \tag{35}$$

and

B1. $v_0^x = 0$.

B2. $E(\varepsilon_t|\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots) = 0$, $E(\varepsilon_t\varepsilon_t'|\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots) = \Sigma$ for some positive definite matrix Σ, and $\sup_t E\left[\|\varepsilon_t\|^{2+\varrho}\right] < \infty$ for some $\varrho > 0$.

B3. $\Psi(L) = I_2 + \sum_{i=1}^{\infty}\Psi_i L^i$, where $|\Psi(1)| \neq 0$ and $\sum_{i=1}^{\infty} i\,\|\Psi_i\| < \infty$.

By accommodating correlation between the (potentially) serially correlated error term of the equation of interest and current (and lagged) values of the nearly integrated regressor, the model (33) $-$ (35) generalizes the predictive regression model (1)$-$(2) in a couple of ways. Nevertheless, it turns out to be relatively straightforward to construct testing procedures with desirable large-sample properties. Once again, the idea is to base inference on a statistic whose limiting distribution (under B1-B3 and local reparameterizations of β and γ) is that of $\mathcal{R}^{\rho}(b, c)$. For completeness, we give a precise description.

Let $x_0 = x_1$, $\hat{v}_0^x = 0$, and define $\hat{v}_t^x = x_t - x_1$ and $\hat{x}_{t-1}^{\mu} = x_{t-1} - T^{-1}\sum_{s=1}^{T} x_{s-1}$ (for $t = 1, \ldots, T$). Define

$$\Omega = \begin{pmatrix} \omega_{yy} & \omega_{yx} \\ \omega_{xy} & \omega_{xx} \end{pmatrix} = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} E(u_t u_s')$$

and

$$\delta_{xy} = \lim_{T\to\infty} T^{-1}\sum_{t=2}^{T}\sum_{s=1}^{t-1} E(u_t^x u_s^{y\prime}),$$

and let $\hat{\Omega}$ and $\hat{\delta}_{xy}$ be consistent estimators thereof. Finally, let

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

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.

Theorem 10. *Let $\left\{(y_t,x_t)'\right\}$ be generated by* (33) – (35) *, suppose B1-B3 hold, and suppose $b=T\left(\beta-\beta_0\right)\omega_{xx}^{1/2}\omega_{yy.x}^{-1/2}$ and $c=T\left(\gamma-1\right)$ are fixed as T increases without bound, where $\omega_{yy.x}=\omega_{yy}-\omega_{xx}^{-1}\omega_{xy}^{2}$. If $\left(\hat{\Omega},\hat{\delta}_{xy}\right)\to_p\left(\Omega,\delta_{xy}\right)$, then $\hat{R}^{CI}\to_d\mathcal{R}^{\rho}\left(b,c\right)$ as $T\to\infty$, where $\rho=\omega_{xy}\omega_{xx}^{-1/2}\omega_{yy}^{-1/2}$. Moreover,*

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

and

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

Remark. Let $\hat{u}_t=\left(\hat{u}_t^{y},\hat{u}_t^{x}\right)'=\left(y_t^{\mu}-\hat{\beta}\hat{x}_{t-1}^{\mu},\hat{v}_t^{x}-\hat{\gamma}\hat{v}_{t-1}^{x}\right)'$, where $y_t^{\mu}=y_t-T^{-1}\sum_{s=1}^{T}y_s$, $\hat{\beta}=\left(\sum_{t=1}^{T}\hat{x}_{t-1}^{\mu 2}\right)^{-1}\sum_{t=1}^{T}\hat{x}_{t-1}^{\mu}y_t$, and $\hat{\gamma}=\left(\sum_{t=1}^{T}\hat{v}_{t-1}^{x2}\right)^{-1}\sum_{t=1}^{T}\hat{v}_{t-1}^{x}\hat{v}_t^{x}$. Under the assumptions of Theorem 10 and fairly general conditions on $k\left(\cdot\right)$ and B_T, it follows from Jansson (2002) that

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

and

$$\hat{\delta}_{xy} = T^{-1} \sum_{t=2}^{T} \sum_{s=1}^{t-1} k\left(\frac{|t-s|}{B_T}\right) \hat{u}_t^x \hat{u}_s^{y\prime} \to_p \delta_{xy}.$$

7. Appendix A: Auxiliary Lemmas

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

Lemma 11. *Let* $\eta \in (0,1)$ *be given and define*

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

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

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

Lemma 11 is a special case of the following result, which gives general conditions under which critical value functions for one- and two-sided tests are continuous in their arguments.

Lemma 12. *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*

$$\begin{aligned} \int_{\underline{C}_\eta(\theta)}^{\overline{C}_\eta(\theta)} f(r; \theta)\, dr &= 1 - \eta, \\ \int_{\underline{C}_\eta(\theta)}^{\overline{C}_\eta(\theta)} r f(r; \theta)\, dr &= (1 - \eta) \int_{-\infty}^{\infty} r f(r; \theta)\, dr. \end{aligned}$$

If

$$\lim\nolimits_{\delta \downarrow 0} \sup\nolimits_{\{\theta \in \Theta : d_\theta(\theta_0, \theta) < \delta\}} \int_{-\infty}^{\infty} \max(1, |r|)\, |f(r; \theta_0) - f(r; \theta)|\, dr = 0, \qquad \text{(A.1)}$$

then $\underline{C}_\eta : \Theta \to \mathbb{R}$ *and* $\overline{C}_\eta : \Theta \to \mathbb{R}$ *are continuous at* θ_0.

(c) Condition $(A.1)$ *holds if, for any* $\varepsilon > 0$, *there exists a constant* $\delta > 0$, *a set* $\Delta \subseteq \mathbb{R}$, *and a function* $\bar{f} : \mathbb{R} \to \mathbb{R}_+$ *such that*

$$\sup_{\{\theta\in\Theta:d_\theta(\theta_0,\theta)<\delta\}} \int_\Delta \max(1,|r|)\, f(r;\theta)\, dr < \varepsilon, \tag{A.2}$$

$$\sup_{\{\theta\in\Theta:d_\theta(\theta_0,\theta)<\delta\}} f(r;\theta) \le \bar{f}(r) \qquad \forall r \in \mathbb{R}\setminus\Delta, \tag{A.3}$$

$$\int_{\mathbb{R}\setminus\Delta} \max(1,|r|)\, \bar{f}(r)\, dr < \infty. \tag{A.4}$$

Proof of Lemma 12. The proof is by contradiction.

Proof of (a). If $C_\eta(\cdot)$ is not continuous at θ_0, then there exists a sequence $\{\theta_n : n \ge 1\} \subseteq \Theta$ and a constant $\varepsilon > 0$ such that $\lim_{n\to\infty}\theta_n = \theta_0$ and either $\sup_n C_\eta(\theta_n) < C_\eta(\theta_0) - \varepsilon$ or $\inf_n C_\eta(\theta_n) > C_\eta(\theta_0) + \varepsilon$.

If $\sup_n C_\eta(\theta_n) < C_\eta(\theta_0) - \varepsilon$, then

$$\begin{aligned}
\eta &= \underline{\lim}_{n\to\infty} \int_{C_\eta(\theta_n)}^\infty f(r;\theta_n)\, dr \ge \underline{\lim}_{n\to\infty} \int_{C_\eta(\theta_0)-\varepsilon}^\infty f(r;\theta_n)\, dr \\
&\ge \int_{C_\eta(\theta_0)-\varepsilon}^\infty \left[\underline{\lim}_{n\to\infty} f(r;\theta_n)\right] dr = \int_{C_\eta(\theta_0)-\varepsilon}^\infty f(r;\theta_0)\, dr \\
&> \int_{C_\eta(\theta_0)}^\infty f(r;\theta_0)\, dr = \eta,
\end{aligned}$$

a contradiction. In the display, the first equality uses the defining property of $C_\eta(\cdot)$, the first inequality uses $\sup_n C_\eta(\theta_n) < C_\eta(\theta_0) - \varepsilon$, the second inequality uses Fatou's lemma, the second equality uses continuity of $f(r;\cdot)$, the third inequality uses uniqueness of $C_\eta(\theta_0)$, and the last equality uses the definition of $C_\eta(\cdot)$.

Analogous reasoning shows that $\inf_n C_\eta(\theta_n) > C_\eta(\theta_0) + \varepsilon$ is impossible, completing the proof of part (a).

Proof of (b). If $\int_{\underline{C}_\eta(\theta)}^{\overline{C}_\eta(\theta)} f(r;\theta)\, dr = 1 - \eta$ for every $\theta \in \Theta$, but $\left(\underline{C}_\eta(\cdot), \overline{C}_\eta(\cdot)\right)'$ is not continuous at θ_0, then there exists a sequence $\{\theta_n : n \ge 1\} \subseteq \Theta$ and a constant $\varepsilon > 0$ such that $\lim_{n\to\infty}\theta_n = \theta_0$ and either

$$\sup_n \underline{C}_\eta(\theta_n) < \underline{C}_\eta(\theta_0) - \varepsilon \text{ and } \sup_n \overline{C}_\eta(\theta_n) < \overline{C}_\eta(\theta_0) - \varepsilon \tag{A.5}$$

or

$$\inf_n \underline{C}_\eta(\theta_n) > \underline{C}_\eta(\theta_0) + \varepsilon \text{ and } \inf_n \overline{C}_\eta(\theta_n) > \overline{C}_\eta(\theta_0) + \varepsilon. \tag{A.6}$$

The equations defining $\underline{C}_\eta(\cdot)$ and $\overline{C}_\eta(\cdot)$ can be written as follows:

$$I(\theta, k) = J(\theta, k) \qquad \forall \theta \in \Theta, k \in \mathbb{R},$$

where $I(\theta, k) = \int_{\underline{C}_\eta(\theta)}^{\overline{C}_\eta(\theta)} (r-k) f(r;\theta)\, dr$ and $J(\theta, k) = (1-\eta)\int_{-\infty}^{\infty} (r-k) f(r;\theta)\, dr$. Now, by (A.1),

$$\begin{aligned} |J(\theta_0, k) - J(\theta_n, k)| &\leq (1-\eta)\int_{-\infty}^{\infty} |r-k|\, |f(r;\theta_0) - f(r;\theta_n)|\, dr \\ &\leq (1-\eta)(1+|k|)\int_{-\infty}^{\infty} \max(1,|r|)\, |f(r;\theta_0) - f(r;\theta_n)|\, dr \\ &\to 0 \end{aligned}$$

as $n \to \infty$. To complete a proof by contradiction it therefore suffices to find a value of k such that $I(\theta_0, k) - I(\theta_n, k) \nrightarrow 0$ as $n \to \infty$.

Suppose (A.5) holds. If $\overline{C}_\eta(\theta_n) < \underline{C}_\eta(\theta_0)$, then

$$\begin{aligned} & I\left(\theta_0, \underline{C}_\eta(\theta_0)\right) - I\left(\theta_n, \underline{C}_\eta(\theta_0)\right) \\ = & \int_{\underline{C}_\eta(\theta_0)}^{\overline{C}_\eta(\theta_0)} \left[r - \underline{C}_\eta(\theta_0)\right] f(r;\theta_0)\, dr + \int_{\underline{C}_\eta(\theta_n)}^{\overline{C}_\eta(\theta_n)} \left[\underline{C}_\eta(\theta_0) - r\right] f(r;\theta_n)\, dr \\ \geq & \int_{\underline{C}_\eta(\theta_0)}^{\overline{C}_\eta(\theta_0)} \left[r - \underline{C}_\eta(\theta_0)\right] f(r;\theta_0)\, dr > 0. \end{aligned}$$

Otherwise, if $\overline{C}_\eta(\theta_n) < \underline{C}_\eta(\theta_0)$, then

$$\begin{aligned} & I\left(\theta_0, \underline{C}_\eta(\theta_0)\right) - I\left(\theta_n, \underline{C}_\eta(\theta_0)\right) \\ = & \int_{\overline{C}_\eta(\theta_n)}^{\overline{C}_\eta(\theta_0)} \left[r - \underline{C}_\eta(\theta_0)\right] f(r;\theta_0)\, dr + \int_{\underline{C}_\eta(\theta_n)}^{\underline{C}_\eta(\theta_0)} \left[\underline{C}_\eta(\theta_0) - r\right] f(r;\theta_n)\, dr \\ & + \int_{\underline{C}_\eta(\theta_0)}^{\overline{C}_\eta(\theta_0)} \left[r - \underline{C}_\eta(\theta_0)\right] \left[f(r;\theta_0) - f(r;\theta_n)\right] dr, \end{aligned}$$

where

$$\int_{\overline{C}_\eta(\theta_n)}^{\overline{C}_\eta(\theta_0)} \left[r - \underline{C}_\eta(\theta_0)\right] f(r;\theta_0)\, dr \geq \int_{\overline{C}_\eta(\theta_0)-\varepsilon}^{\overline{C}_\eta(\theta_0)} \left[r - \underline{C}_\eta(\theta_0)\right] f(r;\theta_0)\, dr > 0,$$

$$\int_{\underline{C}_\eta(\theta_n)}^{\underline{C}_\eta(\theta_0)} \left[\underline{C}_\eta(\theta_0) - r\right] f(r;\theta_n)\, dr \geq 0,$$

and

$$\begin{aligned} &\left| \int_{\underline{C}_\eta(\theta_0)}^{\overline{C}_\eta(\theta_0)} \left[r - \underline{C}_\eta(\theta_0)\right] \left[f(r;\theta_0) - f(r;\theta_n)\right] dr \right| \\ \leq\ & \int_{-\infty}^{\infty} \left|r - \underline{C}_\eta(\theta_0)\right| \left|f(r;\theta_0) - f(r;\theta_n)\right| dr \to 0, \end{aligned}$$

where the last line uses (A.1). As a consequence,

$$\begin{aligned} &\underline{\lim}_{n\to\infty} \left[I\left(\theta_0, \underline{C}_\eta(\theta_0)\right) - I\left(\theta_n, \underline{C}_\eta(\theta_0)\right)\right] \\ \geq\ & \int_{\max\left(\underline{C}_\eta(\theta_0), \overline{C}_\eta(\theta_0)-\varepsilon\right)}^{\overline{C}_\eta(\theta_0)} \left[r - \underline{C}_\eta(\theta_0)\right] f(r|\theta_0)\, dr > 0 \end{aligned}$$

if (A.5) holds.

Analogous reasoning shows that

$$\overline{\lim}_{n\to\infty} \left[I\left(\theta_0, \overline{C}_\eta(\theta_0)\right) - I\left(\theta_n, \overline{C}_\eta(\theta_0)\right)\right] < 0$$

if (A.6) holds. The proof of (b) is therefore complete.

Proof of (c). To show that (A.1) is implied by (A.2) – (A.4) under the assumptions of the lemma, suppose (A.1) does not hold. Then there exists a sequence $\{\theta_n : n \geq 1\} \subseteq \Theta$ and a constant $\varepsilon > 0$ such that $\lim_{n\to\infty} \theta_n = \theta_0$ and

$$\int_{-\infty}^{\infty} \max(1, |r|) \left|f(r;\theta_0) - f(r;\theta_n)\right| dr > 2\varepsilon \qquad \forall n.$$

Assuming we can pick a constant $\delta > 0$, a set $\Delta \subseteq \mathbb{R}$, and a function $\bar{f} : \mathbb{R} \to \mathbb{R}_+$ such that (A.2) – (A.4) hold, we have:

$$\overline{\lim}_{n\to\infty} \sup_{\{\theta\in\Theta: d_\theta(\theta_0,\theta)<d_\theta(\theta_0,\theta_n)\}} \int_\Delta \max(1,|r|)\, f(r;\theta)\, dr \leq \varepsilon$$

and

$$\overline{\lim}_{n\to\infty} \int_{\mathbb{R}\backslash\Delta} \max(1,|r|)\, |f(r;\theta_0) - f(r;\theta_n)|\, dr = 0,$$

where the first display uses $\lim_{n\to\infty} d_\theta(\theta_0,\theta_n) = 0$ and (A.2), while the second display uses $\lim_{n\to\infty} d_\theta(\theta_0,\theta_n) = 0$, continuity of $f(r;\cdot)$ at θ_0, and the dominated convergence theorem (the applicability of which follows from (A.3) – (A.4)). As a consequence,

$$\begin{aligned}
& \overline{\lim}_{n\to\infty} \int_{-\infty}^{\infty} \max(1,|r|)\, |f(r;\theta_0) - f(r;\theta_n)|\, dr \\
\leq\; & 2\overline{\lim}_{n\to\infty} \sup_{\{\theta\in\Theta: d_\theta(\theta_0,\theta)<d_\theta(\theta_0,\theta_n)\}} \int_\Delta \max(1,|r|)\, f(r;\theta)\, dr \\
& + \overline{\lim}_{n\to\infty} \int_{\mathbb{R}\backslash\Delta} \max(1,|r|)\, |f(r;\theta_0) - f(r;\theta_n)|\, dr \\
\leq\; & 2\varepsilon,
\end{aligned}$$

a contradiction. ■

Proof of Lemma 11. The proof of Lemma 11 constructs a conditional probability density function of $\mathcal{R}_\beta$ given $\left(\mathcal{R}_\gamma^\rho, \mathcal{R}_{\beta\beta}, \mathcal{R}_{\gamma\gamma}\right)$ that satisfies the conditions of Lemma 12. Let $\mathcal{Z}_\beta = \mathcal{R}_{\beta\beta}^{-1/2}\mathcal{R}_\beta$, $\mathcal{Z}_\gamma = W_x(1)$, $\mathcal{Z}_{\beta\beta} = \int_0^1 W_x(r)\, dr$, $\mathcal{Q}_\gamma = \mathcal{Z}_\gamma^2$, $\mathcal{Q}_{\beta\beta} = \mathcal{Z}_{\beta\beta}^2$, and $\mathcal{Q}_{\gamma\gamma} = \mathcal{R}_{\gamma\gamma}$, where $\mathcal{R}^\rho = \left(\mathcal{R}_\beta, \mathcal{R}_\gamma^\rho, \mathcal{R}_{\beta\beta}, \mathcal{R}_{\gamma\gamma}\right)' = \mathcal{R}^\rho(0,0)$. For any $(r_\gamma, r_{\beta\beta}, r_{\gamma\gamma}, \rho) \in \mathbb{S}$, let

$$\mathbb{S}_\beta(r_\gamma, r_{\beta\beta}, r_{\gamma\gamma}; \rho) = \left\{ r_\beta : r_\gamma + \frac{\rho}{\sqrt{1-\rho^2}} r_\beta > -\frac{1}{2} \right\}.$$

By construction, $\mathbb{S}$ and $\mathbb{S}_\beta$ satisfy

$$\Pr\left[\mathcal{R}_\beta \in \mathbb{S}_\beta\left(\mathcal{R}_\gamma^\rho, \mathcal{R}_{\beta\beta}, \mathcal{R}_{\gamma\gamma}; \rho\right), \left(\mathcal{R}_\gamma^\rho, \mathcal{R}_{\beta\beta}, \mathcal{R}_{\gamma\gamma}, \rho\right) \in \mathbb{S}\right] = 1.$$

We now obtain a characterization of $f_{\mathcal{R}}^{0}(\cdot;\rho)$, the density of $\mathcal{R}^{\rho}$. Because

$$(\mathcal{Z}_\beta, \mathcal{Q}_\gamma, \mathcal{Q}_{\beta\beta}, \mathcal{Q}_{\gamma\gamma}) = \left(\frac{\mathcal{R}_\beta}{\sqrt{\mathcal{R}_{\beta\beta}}}, 2\mathcal{R}_\gamma^\rho + 2\frac{\rho}{\sqrt{1-\rho^2}}\mathcal{R}_\beta + 1, \mathcal{R}_{\gamma\gamma} - \mathcal{R}_{\beta\beta}, \mathcal{R}_{\gamma\gamma} \right),$$

it follows from the changes of variables formula that

$$f_{\mathcal{R}}^{0}(r;\rho) = \frac{2}{\sqrt{r_{\beta\beta}}} f_{\mathcal{Z}_\beta,\mathcal{Q}} \left(\frac{r_\beta}{\sqrt{r_{\beta\beta}}}, 2r_\gamma + 2\frac{\rho}{\sqrt{1-\rho^2}} r_\beta + 1, r_{\gamma\gamma} - r_{\beta\beta}, r_{\gamma\gamma} \right),$$

where $f_{\mathcal{Z}_\beta,\mathcal{Q}}(\cdot)$ is the density of $(\mathcal{Z}_\beta, \mathcal{Q}')' = (\mathcal{Z}_\beta, \mathcal{Q}_\gamma, \mathcal{Q}_{\beta\beta}, \mathcal{Q}_{\gamma\gamma})'$.

Now, $\mathcal{Z}_\beta \sim \mathcal{N}(0,1)$ and $\mathcal{Z}_\beta$ is independent of $\mathcal{Q}$. As a consequence,

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

where $f_{\mathcal{Q}}(\cdot)$ is the density of $\mathcal{Q}$.

Since

$$\begin{pmatrix} \mathcal{Z}_\gamma \\ \mathcal{Z}_{\beta\beta} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{pmatrix} \right),$$

the density of $(\mathcal{Q}_\gamma, \mathcal{Q}_{\beta\beta})'$ is given by

$$f_{\mathcal{Q}_\gamma,\mathcal{Q}_{\beta\beta}}(q_\gamma, q_{\beta\beta}) = \frac{1}{2\sqrt{q_\gamma}\sqrt{q_{\beta\beta}}} \left[f_{\mathcal{Z}_\gamma,\mathcal{Z}_{\beta\beta}}\left(\sqrt{q_\gamma}, \sqrt{q_{\beta\beta}}\right) + f_{\mathcal{Z}_\gamma,\mathcal{Z}_{\beta\beta}}\left(-\sqrt{q_\gamma}, \sqrt{q_{\beta\beta}}\right) \right],$$

where

$$f_{\mathcal{Z}_\gamma,\mathcal{Z}_{\beta\beta}}(z_\gamma, z_{\beta\beta}) = \frac{\sqrt{3}}{\pi} \exp\left[-\begin{pmatrix} z_\gamma \\ z_{\beta\beta} \end{pmatrix}' \begin{pmatrix} 2 & -3 \\ -3 & 6 \end{pmatrix} \begin{pmatrix} z_\gamma \\ z_{\beta\beta} \end{pmatrix} \right] \qquad \text{(A.7)}$$

is the density of $(\mathcal{Z}_\gamma, \mathcal{Z}_{\beta\beta})'$.

Moreover,

$$
\begin{aligned}
& \Pr\left(sign\left(\mathcal{Z}_{\gamma}\right)=sign\left(\mathcal{Z}_{\beta\beta}\right)|\mathcal{Z}_{\gamma}^{2}=q_{\gamma},\mathcal{Z}_{\beta\beta}^{2}=q_{\beta\beta}\right) \\
= & \ 1-\Pr\left(sign\left(-\mathcal{Z}_{\gamma}\right)=sign\left(\mathcal{Z}_{\beta\beta}\right)|\mathcal{Z}_{\gamma}^{2}=q_{\gamma},\mathcal{Z}_{\beta\beta}^{2}=q_{\beta\beta}\right) \\
= & \ \frac{f_{\mathcal{Z}_{\gamma},\mathcal{Z}_{\beta\beta}}\left(\sqrt{q_{\gamma}},\sqrt{q_{\beta\beta}}\right)}{f_{\mathcal{Z}_{\gamma},\mathcal{Z}_{\beta\beta}}\left(\sqrt{q_{\gamma}},\sqrt{q_{\beta\beta}}\right)+f_{\mathcal{Z}_{\gamma},\mathcal{Z}_{\beta\beta}}\left(-\sqrt{q_{\gamma}},\sqrt{q_{\beta\beta}}\right)}
\end{aligned}
$$

whenever $q_{\gamma}>0$ and $q_{\beta\beta}>0$.

Finally, we characterize the density of $\mathcal{Q}_{\gamma\gamma}$ conditional on $\left(\mathcal{Z}_{\gamma},\mathcal{Z}_{\beta\beta}\right)'$. Let the zero mean Gaussian process $\tilde{W}_{x}\left(\cdot\right)$ be given by

$$
\tilde{W}_{x}\left(r\right)=W_{x}\left(r\right)-p\left(r\right)'\left(\begin{array}{c}\mathcal{Z}_{\gamma}\\ \mathcal{Z}_{\gamma\gamma}\end{array}\right),\qquad p\left(r\right)=\left(\begin{array}{c}-2r+3r^{2}\\ 6r-6r^{2}\end{array}\right).
$$

By construction, $\left\{\tilde{W}_{x}\left(r\right):0\leq r\leq1\right\}$ is independent of $\left(\mathcal{Z}_{\gamma},\mathcal{Z}_{\beta\beta}\right)'$. The covariance kernel $\tilde{K}\left(\cdot\right)$ of $\tilde{W}_{x}\left(\cdot\right)$ is given by

$$
\begin{aligned}
\tilde{K}\left(r,s\right) & = Cov\left(\tilde{W}_{x}\left(r\right),\tilde{W}_{x}\left(s\right)\right) \\
& = \min\left(r,s\right)\left(1-\max\left(r,s\right)\right)-3\left(r-r^{2}\right)\left(s-s^{2}\right).
\end{aligned}
$$

It is easy to show that $g_{0}\left(r\right)=1$ is the only continuous function $g\left(\cdot\right)$ satisfying $\int_{0}^{1}g\left(r\right)^{2}dr=1$ and

$$
\int_{0}^{1}\tilde{K}\left(r,s\right)g\left(s\right)ds=0\qquad\forall r\in\left[0,1\right].
$$

It therefore follows from Mercer's Theorem (e.g. Shorack and Wellner (1986)) that

$$
\tilde{K}\left(r,s\right)=\sum_{j=1}^{\infty}\lambda_{j}g_{j}\left(r\right)g_{j}\left(s\right),
$$

where $\lambda_{j}>0$ for every j, $\lambda_{1}\geq\lambda_{2}\geq\ldots$ and $\left\{g_{j}\left(\cdot\right):j\geq0\right\}$ is an orthonormal basis of the set of square integrable functions defined on the unit interval. As a consequence, $\tilde{W}_{x}\left(r\right)$ can be represented as

$$\tilde{W}_x(r) = \sum_{j=1}^{\infty} \lambda_j^{1/2} \mathcal{Z}_j g_j(r),$$

where $\mathcal{Z}_j = \lambda_j^{-1/2} \int_0^1 \tilde{W}_x^{\mu}(r) g_j(r) dr \sim i.i.d.\ \mathcal{N}(0,1)$. Because the function $p(\cdot)$ can be written as

$$p(r) = \sum_{j=0}^{\infty} h_j g_j(r), \qquad h_j = \int_0^1 p(r) g_j(r) dr,$$

it follows from the preceding displays that

$$W_x(r) = (\mathcal{Z}_\gamma, \mathcal{Z}_{\beta\beta}) h_0 g_0(r) + \sum_{j=1}^{\infty} \lambda_j^{1/2} \left[\mathcal{Z}_j + \lambda_j^{-1/2} (\mathcal{Z}_\gamma, \mathcal{Z}_{\beta\beta}) h_j \right] g_j(r),$$

implying

$$\mathcal{Q}_{\beta\beta} = \int_0^1 W_x(r)^2 dr = [(\mathcal{Z}_\gamma, \mathcal{Z}_{\beta\beta}) h_0]^2 + \sum_{j=1}^{\infty} \lambda_j \left[\mathcal{Z}_j + \lambda_j^{-1/2} (\mathcal{Z}_\gamma, \mathcal{Z}_{\beta\beta}) h_j \right]^2,$$

where $\mathcal{Z}_j \sim i.i.d.\ \mathcal{N}(0,1)$ and $\{\mathcal{Z}_j : j \geq 1\}$ is independent of $(\mathcal{Z}_\gamma, \mathcal{Z}_{\beta\beta})'$. Conditional on $(\mathcal{Z}_\gamma, \mathcal{Z}_{\beta\beta}) = (z_\gamma, z_{\beta\beta})$, the characteristic function of $\mathcal{Q}_{\gamma\gamma}$ is therefore given by

$$\varkappa(t|z_\gamma, z_{\beta\beta}) = \exp\left(i [(z_\gamma, z_{\beta\beta}) h_0]^2 t\right) \prod_{j=1}^{\infty} \varkappa_j(t|z_\gamma, z_{\beta\beta}),$$

where

$$\varkappa_j(t|z_\gamma, z_{\beta\beta}) = (1 - i2\lambda_j t)^{-1/2} \exp\left(\frac{i [(z_\gamma, z_{\beta\beta}) h_j]^2 t}{1 - i2\lambda_j t} \right)$$

denotes the conditional characteristic function of $\lambda_j \left[\mathcal{Z}_j + \lambda_j^{-1/2} (\mathcal{Z}_\gamma, \mathcal{Z}_{\beta\beta}) h_j \right]^2$ given $(\mathcal{Z}_\gamma, \mathcal{Z}_{\beta\beta}) = (z_\gamma, z_{\beta\beta})$. It now follows from the inversion theorem for characteristic functions that the density of $\mathcal{Q}_{\gamma\gamma}$ conditional on $(\mathcal{Z}_\gamma, \mathcal{Z}_{\beta\beta}) = (z_\gamma, z_{\beta\beta})$ is given by

$$f_{\mathcal{Q}_{\gamma\gamma}|\mathcal{Z}_\gamma, \mathcal{Z}_{\beta\beta}}(q_{\gamma\gamma}|z_\gamma, z_{\beta\beta}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-i q_{\gamma\gamma} t) \varkappa(t|z_\gamma, z_{\beta\beta}) dt. \tag{A.8}$$

(The inversion theorem is applicable because

$$\begin{aligned}\int_{-\infty}^{\infty}\left|\varkappa\left(t|z_\gamma,z_{\beta\beta}\right)\right|dt &= \int_{-\infty}^{\infty}\left(\prod_{j=1}^{\infty}\left|\varkappa_j\left(t|z_\gamma,z_{\beta\beta}\right)\right|\right)dt\leq\int_{-\infty}^{\infty}\left(\prod_{j=1}^{\infty}\left|\varkappa_j\left(t|0,0\right)\right|\right)dt\\ &\leq \int_{-\infty}^{\infty}\left(\prod_{j=1}^{3}\left|\varkappa_j\left(t|0,0\right)\right|\right)dt\leq\int_{-\infty}^{\infty}\left|\varkappa_3\left(t|0,0\right)\right|^3dt<\infty.\end{aligned}$$

The first inequality in the display uses $\sup_{z_\gamma,z_{\beta\beta}}\left|\varkappa_j\left(t|z_\gamma,z_{\beta\beta}\right)\right|=\left|\varkappa_j\left(t|0,0\right)\right|$, which holds because

$$\operatorname{Re}\left(\frac{i\left[\left(z_\gamma,z_{\beta\beta}\right)h_j\right]^2t}{1-i2\lambda_jt}\right)=-\frac{2\lambda_j\left[\left(z_\gamma,z_{\beta\beta}\right)h_j\right]^2t^2}{1+4\lambda_j^2t^2}\leq 0.$$

The second inequality in the display uses $\sup_{j\geq 1,t\in\mathbb{R}}\left|\varkappa_j\left(t|0,0\right)\right|\leq 1$, the third inequality uses $\lambda_1\geq\lambda_2\geq\lambda_3$, and the last inequality uses $\lambda_3>0$.)

Combining pieces, we have:

$$f_{\mathcal{R}}^0\left(r;\rho\right)=\frac{2}{\sqrt{r_{\beta\beta}}}f_{\mathcal{Z}_\beta,\mathcal{Q}}\left(\frac{r_\beta}{\sqrt{r_{\beta\beta}}},2r_\gamma+2\frac{\rho}{\sqrt{1-\rho^2}}r_\beta+1,r_{\gamma\gamma}-r_{\beta\beta},r_{\gamma\gamma}\right),$$

where

$$\begin{aligned}&2f_{\mathcal{Z}_\beta,\mathcal{Q}}\left(z_\beta,q_\gamma,q_{\beta\beta},q_{\gamma\gamma}\right)\\ =\;&\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z_\beta^2}{2}\right)\frac{1}{\sqrt{q_\gamma}\sqrt{q_{\beta\beta}}}f_{\mathcal{Z}_\gamma,\mathcal{Z}_{\beta\beta}}\left(\sqrt{q_\gamma},\sqrt{q_{\beta\beta}}\right)f_{\mathcal{Q}_{\gamma\gamma}|\mathcal{Z}_\gamma,\mathcal{Z}_{\beta\beta}}\left(q_{\gamma\gamma}|\sqrt{q_\gamma},\sqrt{q_{\beta\beta}}\right)\\ &+\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z_\beta^2}{2}\right)\frac{1}{\sqrt{q_\gamma}\sqrt{q_{\beta\beta}}}f_{\mathcal{Z}_\gamma,\mathcal{Z}_{\beta\beta}}\left(-\sqrt{q_\gamma},\sqrt{q_{\beta\beta}}\right)f_{\mathcal{Q}_{\gamma\gamma}|\mathcal{Z}_\gamma,\mathcal{Z}_{\beta\beta}}\left(q_{\gamma\gamma}|-\sqrt{q_\gamma},\sqrt{q_{\beta\beta}}\right),\end{aligned}$$

and $f_{\mathcal{Z}_\gamma,\mathcal{Z}_{\beta\beta}}\left(\cdot\right)$ and $f_{\mathcal{Q}_{\gamma\gamma}|\mathcal{Z}_\gamma,\mathcal{Z}_{\beta\beta}}\left(\cdot\right)$ are defined in (A.7) and (A.8), respectively.

For any $\left(r_\gamma,r_{\beta\beta},r_{\gamma\gamma},\rho\right)\in\mathbb{S}$, let

$$\begin{aligned} & f^0_{\mathcal{R}_\beta|\mathcal{R}^\rho_\gamma,\mathcal{R}_{\beta\beta},\mathcal{R}_{\gamma\gamma}}\left(r_\beta|r_\gamma,r_{\beta\beta},r_{\gamma\gamma};\rho\right) \\ = & \ 1\left[r_\beta\in\mathbb{S}_\beta\left(r_\gamma,r_{\beta\beta},r_{\gamma\gamma};\rho\right)\right]\frac{f^0_{\mathcal{R}}\left(r_\beta,r_\gamma,r_{\beta\beta},r_{\gamma\gamma};\rho\right)}{\int_{\mathbb{S}_\beta\left(r_\gamma,r_{\beta\beta},r_{\gamma\gamma};\rho\right)}f_{\mathcal{R}}\left(r_\beta,r_\gamma,r_{\beta\beta},r_{\gamma\gamma};\rho\right)dr_\beta} \end{aligned}$$

be the conditional density of $\mathcal{R}_\beta$ given $\left(\mathcal{R}^\rho_\gamma,\mathcal{R}_{\beta\beta},\mathcal{R}_{\gamma\gamma}\right)=\left(r_\gamma,r_{\beta\beta},r_{\gamma\gamma}\right)$.

Let $\pi^*_\eta\left(r;\rho\right)=1\left[r_\beta>\mathcal{C}_\eta\left(r_\gamma,r_{\beta\beta},r_{\gamma\gamma};\rho\right)\right]$, where $\mathcal{C}_\eta:\mathbb{S}\to\mathbb{R}$ is implicitly given by

$$\int_{\mathcal{C}_\eta\left(r_\gamma,r_{\beta\beta},r_{\gamma\gamma};\rho\right)}^{\infty}f^0_{\mathcal{R}_\beta|\mathcal{R}^\rho_\gamma,\mathcal{R}_{\beta\beta},\mathcal{R}_{\gamma\gamma}}\left(r_\beta|r_\gamma,r_{\beta\beta},r_{\gamma\gamma};\rho\right)dr_\beta=\eta.$$

(Existence and uniqueness of $\mathcal{C}_\eta\left(\cdot\right)$ is easily shown.) By construction, $\pi^*_\eta\left(\cdot;\rho\right)$ satisfies (21). Moreover, Lemma 12 (a) can be used to used to show that $\mathcal{C}_\eta\left(\cdot\right)$ is continuous on $\mathbb{S}$.

Let $\pi^{**}_\eta\left(r;\rho\right)=1\left[r_\beta<\underline{\mathcal{C}}_\eta\left(r_\gamma,r_{\beta\beta},r_{\gamma\gamma};\rho\right)\right]+1\left[r_\beta>\overline{\mathcal{C}}_\eta\left(r_\gamma,r_{\beta\beta},r_{\gamma\gamma};\rho\right)\right]$, where the functions $\underline{\mathcal{C}}_\eta:\mathbb{S}\to\mathbb{R}$ and $\overline{\mathcal{C}}_\eta:\mathbb{S}\to\mathbb{R}$ satisfy

$$\int_{\underline{\mathcal{C}}_\eta\left(r_\gamma,r_{\beta\beta},r_{\gamma\gamma};\rho\right)}^{\overline{\mathcal{C}}_\eta\left(r_\gamma,r_{\beta\beta},r_{\gamma\gamma};\rho\right)}f^0_{\mathcal{R}_\beta|\mathcal{R}^\rho_\gamma,\mathcal{R}_{\beta\beta},\mathcal{R}_{\gamma\gamma}}\left(r_\beta|r_\gamma,r_{\beta\beta},r_{\gamma\gamma};\rho\right)dr_\beta=1-\eta$$

and

$$\begin{aligned} & \int_{\underline{\mathcal{C}}_\eta\left(r_\gamma,r_{\beta\beta},r_{\gamma\gamma};\rho\right)}^{\overline{\mathcal{C}}_\eta\left(r_\gamma,r_{\beta\beta},r_{\gamma\gamma};\rho\right)}r_\beta f^0_{\mathcal{R}_\beta|\mathcal{R}^\rho_\gamma,\mathcal{R}_{\beta\beta},\mathcal{R}_{\gamma\gamma}}\left(r_\beta|r_\gamma,r_{\beta\beta},r_{\gamma\gamma};\rho\right)dr_\beta \\ = & \ (1-\eta)\int_{-\infty}^{\infty}r_\beta f^0_{\mathcal{R}_\beta|\mathcal{R}^\rho_\gamma,\mathcal{R}_{\beta\beta},\mathcal{R}_{\gamma\gamma}}\left(r_\beta|r_\gamma,r_{\beta\beta},r_{\gamma\gamma};\rho\right)dr_\beta. \end{aligned}$$

(Existence and uniqueness of $\underline{\mathcal{C}}_\eta\left(\cdot\right)$ and $\overline{\mathcal{C}}_\eta\left(\cdot\right)$ follows from the development in Lehmann (1994, Section 4.2).) By construction, $\pi^{**}_\eta\left(\cdot;\rho\right)$ satisfies $(24)-(25)$. Moreover, Lemma 12 (b)-(c) can be used to used to show that $\underline{\mathcal{C}}_\eta\left(\cdot\right)$ and $\overline{\mathcal{C}}_\eta\left(\cdot\right)$ are continuous on $\mathbb{S}$. ■

8. Appendix B: Proofs

Proof of Lemma 1. Lemma 1 follows from equation (6) and the properties of exponential families (e.g., Lehmann (1994, Lemma 2.8)). ■

Proof of Theorem 2. By Lemma 1 (b) and Lehmann (1994, Theorem 4.3),

$$E_{\beta,\gamma}\left[\phi\left(S_\beta,S_\gamma,S_{\beta\beta},S_{\gamma\gamma}\right)|S_{\beta\beta},S_{\gamma\gamma}\right]\leq E_{\beta,\gamma}\left[\phi_\eta^*\left(S_\beta,S_\gamma,S_{\beta\beta},S_{\gamma\gamma}\right)|S_{\beta\beta},S_{\gamma\gamma}\right]$$

for any $\phi(\cdot)$ satisfying (7), any $\beta>\beta_0$, and any $\gamma\in\mathbb{R}$. Part (a) now follows from the law of iterated expectations.

Analogous reasoning establishes part (b) (including existence of the functions $\underline{C}_\eta(\cdot)$ and $\overline{C}_\eta(\cdot)$ satisfying $(12)-(13)$). ■

Proof of Lemma 3. Lemma 3 follows from standard weak convergence arguments (e.g., Phillips (1987, 1988a, 1988b) and Phillips and Solo (1992)) and straightforward algebra. ■

Proof of Lemma 4. Lemma 4 follows from (17), Lemma 3, Lehmann (1994, Lemma 2.8), 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,

$$\begin{aligned}&\mathcal{L}\left[\beta_T(b),\gamma_T(c)\right]-\mathcal{L}\left[\beta_T(0),\gamma_T(0)\right]\\ =\;& bR_\beta+cR_\gamma-\frac{1}{2}\left(b-\frac{\rho}{\sqrt{1-\rho^2}}c\right)^2R_{\beta\beta}-\frac{1}{2}c^2R_{\gamma\gamma}\rightarrow_{d_0}\Lambda^\rho(b,c),\end{aligned}$$

where

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

" $\rightarrow_{d_0}$ " is shorthand for " $\rightarrow_d$ when $(\beta,\gamma)=(\beta_0,1)$" and the convergence result follows from Lemma 3. ■

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

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

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

Proof of (a). Let $\Pi(\eta, \rho)$ denote the class of all functions $\pi(\cdot)$ satisfying

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

By construction, $\pi_\eta^*(\cdot;\rho) \in \Pi(\eta, \rho)$. Applying Lehmann (1994, Theorem 4.3) and the law of iterated expectations, it can be shown that $\pi_\eta^*(\cdot;\rho)$ satisfies

$$E\left[\pi(\mathcal{R}^\rho) e^{\Lambda^\rho(b,c)}\right] \le E\left[\pi_\eta^*(\mathcal{R}^\rho;\rho) e^{\Lambda^\rho(b,c)}\right] \qquad \forall b \ge 0, c \in \mathbb{R}, \pi \in \Pi(\eta, \rho).$$

Because $\mathcal{C}_\eta(\cdot)$ is continuous (Lemma 11), 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

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

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

$$\overline{\lim}_{T\to\infty} E_{\beta_T(b),\gamma_T(c)} \pi_T(R;\rho) = E\left[\pi(\mathcal{R}^\rho;\rho) e^{\Lambda^\rho(b,c)}\right]. \tag{B.1}$$

Let $\{\pi_T(\cdot)\}$, $b \ge 0$ and $c \in \mathbb{R}$ be given, and suppose $\{\pi_T(\cdot)\}$ satisfies (19). Let $\{\pi_{T(m)}(\cdot) : m \ge 1\}$ be any subsequence of $\{\pi_T(\cdot)\}$ satisfying

$$\lim_{m\to\infty} E_{\beta_T(b),\gamma_T(c)} \pi_{T(m)}(R;\rho) = \overline{\lim}_{T\to\infty} E_{\beta_T(b),\gamma_T(c)} \pi_T(R;\rho).$$

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

$$\left(\pi_{T(m(n))}, R\right) \to_{d_0} (\pi_\infty, \mathcal{R}^\rho) \tag{37}$$

as $n \to \infty$, where π_∞ is some random variable (defined on the same probability space as $\mathcal{R}^\rho$) and the dependence of R on n has been suppressed. Now,

$$\begin{aligned}\overline{\lim}_{T\to\infty} E_{\beta_T(b),\gamma_T(c)} \pi_T(R;\rho) &= \lim_{m\to\infty} E_{\beta_{T(m)}(b),\gamma_{T(m)}(c)} \pi_{T(m)}(R;\rho) \\ &= \lim_{n\to\infty} E_{\beta_{T(m(n))}(b),\gamma_{T(m(n))}(c)} \pi_{T(m(n))}(R;\rho) \\ &= E\left[\pi_\infty \exp\left(\Lambda^\rho(b,c)\right)\right] \\ &= E\left[\pi\left(\mathcal{R}^\rho;\rho\right) e^{\Lambda^\rho(b,c)}\right],\end{aligned}$$

where $\pi(\mathcal{R}^\rho) = E(\pi_\infty | \mathcal{R}^\rho)$, the third equality uses (37), 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

$$\begin{aligned}& E\left[\left(\pi\left(\mathcal{R}^\rho\right) - \eta\right) g\left(\mathcal{R}_{\beta\beta},\mathcal{R}_{\gamma\gamma}\right) e^{\Lambda^\rho(0,c)}\right] \\ =\ & E\left[\left(\pi_\infty - \eta\right) g\left(\mathcal{R}_{\beta\beta},\mathcal{R}_{\gamma\gamma}\right) e^{\Lambda^\rho(0,c)}\right] \\ =\ & \lim_{n\to\infty} E_{\beta_{T(m(n))}(0),\gamma_{T(m(n))}(c)}\left[\left(\pi_{T(m(n))}(R;\rho) - \eta\right) g\left(R_{\beta\beta},R_{\gamma\gamma}\right)\right] \\ =\ & 0\end{aligned}$$

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 (37), 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 $\pi(\cdot)$ satisfying $\pi \in \Pi(\eta,\rho)$ and

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

By construction, $\pi_\eta^{**}(\cdot;\rho) \in \Pi_0(\eta,\rho)$. Applying Lehmann (1994, Theorem 4.3) and the law of iterated expectations, it can be shown that $\pi_\eta^{**}(\cdot;\rho)$ satisfies

$$E\left[\pi\left(\mathcal{R}^{\rho}\right)e^{\Lambda^{\rho}(b,c)}\right] \leq E\left[\pi_{\eta}^{**}\left(\mathcal{R}^{\rho};\rho\right)e^{\Lambda^{\rho}(b,c)}\right] \qquad \forall b \in \mathbb{R}, c \in \mathbb{R}, \pi \in \Pi_0\left(\eta,\rho\right).$$

Because $\underline{\mathcal{C}}_{\eta}\left(\cdot\right)$ and $\underline{\mathcal{C}}_{\eta}\left(\cdot\right)$ are continuous (Lemma 11 (b)), it follows from Lemma 3 and CMT that $\pi_{\eta}^{**}\left(R;\rho\right) \rightarrow_{d_0} \pi_{\eta}^{**}\left(\mathcal{R}^{\rho};\rho\right)$. This convergence result, Le Cam's third lemma, and Billingsley (1999, Theorem 3.5) can be used to show that $\left\{\pi_{\eta}^{**}\right\}$ satisfies $\left(19\right)$, $\left(22\right)$, and

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

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

$$\overline{\lim}_{T\to\infty} E_{\beta_T(b),\gamma_T(c)}\pi_T\left(R;\rho\right) = E\left[\pi\left(\mathcal{R}^{\rho};\rho\right)e^{\Lambda^{\rho}(b,c)}\right]. \qquad \blacksquare$$

Proof of Theorem 6. The result $\hat{R} \rightarrow_d \mathcal{R}^{\rho}\left(b,c\right)$ follows from standard weak convergence arguments (e.g., Phillips (1987, 1988a, 1988b) and Phillips and Solo (1992)) and straightforward algebra. For instance,

$$\hat{R}_{\gamma\gamma} = \hat{\omega}_{xx}^{-1}T^{-2}\sum_{t=1}^{T}\hat{v}_{t-1}^{x2} = \omega_{xx}^{-1}T^{-2}\sum_{t=1}^{T}v_{t-1}^{x2} + o_p\left(1\right)$$

$$= \int_0^1 \left(\omega_{xx}^{-1/2}T^{-1/2}v_{\lfloor Tr\rfloor}^{x}\right)^2 dr + o_p\left(1\right) \rightarrow_d \int_0^1 W_{x,c}\left(r\right)^2 dr,$$

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

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

$$
\begin{aligned}
E_{\beta_T(b),\gamma_T(c)}\pi_\eta^*\left(\hat{R};\hat{\rho}\right) &= \Pr_{\beta_T(b),\gamma_T(c)}\left[\hat{R}_\beta > \mathcal{C}_\eta\left(\hat{R}_\gamma,\hat{R}_{\beta\beta},\hat{R}_{\gamma\gamma};\hat{\rho}\right)\right] \\
&\to \Pr\left[\mathcal{R}_\beta^\rho(b,c) > \mathcal{C}_\eta\left(\mathcal{R}_\gamma^\rho(b,c),\mathcal{R}_{\beta\beta}(c),\mathcal{R}_{\gamma\gamma}(c);\rho\right)\right] \\
&= \varphi_\eta^*(b,c;\rho),
\end{aligned}
$$

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

$$
\lim_{T\to\infty} E_{\beta_T(b),\gamma_T(c)}\pi_\eta^{**}\left(\hat{R};\hat{\rho}\right) \to \varphi_\eta^{**}(b,c;\rho) \qquad \forall(b,c)\in\mathbb{R}^2. \qquad \blacksquare
$$

Proof of Theorem 7. Let $\mathcal{Z}_\gamma$, $\mathcal{Z}_{\beta\beta}$, $\mathcal{Q}_{\gamma\gamma}$, and $\mathcal{Q}$ be defined as in the proof of Lemma 11 and recall that

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

where $f_{\mathcal{Q}}(\cdot)$ is the density of $\mathcal{Q}$ and

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

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

By the inversion theorem for characteristic functions,

$$
\begin{aligned}
& f_{\mathcal{Z}_\gamma,\mathcal{Z}_{\beta\beta},\mathcal{Q}_{\gamma\gamma}}\left(z_\gamma, z_{\beta\beta}, q_{\gamma\gamma}\right) \\
= \;& \frac{1}{(2\pi)^3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varkappa\left(t_\gamma, t_{\beta\beta}, t_{\gamma\gamma}\right)\exp\left[-i\left(t_\gamma z_\gamma + t_{\beta\beta}z_{\beta\beta} + t_{\gamma\gamma}q_{\gamma\gamma}\right)\right]dt_\gamma dt_{\beta\beta}dt_{\gamma\gamma} \\
= \;& \frac{1}{2\pi}\int_{-\infty}^{\infty}\varkappa_{\gamma\gamma}\left(t_{\gamma\gamma}; z_\gamma, z_{\beta\beta}\right)\exp\left[-it_{\gamma\gamma}q_{\gamma\gamma}\right]dt_{\gamma\gamma},
\end{aligned}
$$

where

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

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

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

where

$$
\begin{aligned}
l_1\left(t_\gamma, t_{\gamma\gamma}\right) &= -2t_\gamma^2\frac{\tanh\sqrt{-2it_{\gamma\gamma}}}{\sqrt{-2it_{\gamma\gamma}}} \\
&= -t_\gamma^2\frac{1}{\sqrt{|t_{\gamma\gamma}|}}\frac{\sinh 2\sqrt{|t_{\gamma\gamma}|} + \sin 2\sqrt{|t_{\gamma\gamma}|}}{\cosh 2\sqrt{|t_{\gamma\gamma}|} + \cos 2\sqrt{|t_{\gamma\gamma}|}} \\
&\quad -it_\gamma^2\frac{sign\left(t_{\gamma\gamma}\right)}{\sqrt{|t_{\gamma\gamma}|}}\frac{\sinh 2\sqrt{|t_{\gamma\gamma}|} - \sin 2\sqrt{|t_{\gamma\gamma}|}}{\cosh 2\sqrt{|t_{\gamma\gamma}|} + \cos 2\sqrt{|t_{\gamma\gamma}|}},
\end{aligned}
$$

$$
\begin{aligned}
l_2\left(t_\gamma, t_{\beta\beta}, t_{\gamma\gamma}\right) &= 2i\frac{t_\gamma t_{\beta\beta}}{t_{\gamma\gamma}}\left(\frac{1}{\cosh\sqrt{-2it_{\gamma\gamma}}}-1\right) \\
&= -2t_\gamma t_{\beta\beta}\frac{1}{|t_{\gamma\gamma}|}\frac{2\sinh\sqrt{|t_{\gamma\gamma}|}\sin\sqrt{|t_{\gamma\gamma}|}}{\cosh 2\sqrt{|t_{\gamma\gamma}|}+\cos 2\sqrt{|t_{\gamma\gamma}|}} \\
&\quad -2it_\gamma t_{\beta\beta}\frac{sign\left(t_{\gamma\gamma}\right)}{|t_{\gamma\gamma}|}\left(1-\frac{2\cosh\sqrt{|t_{\gamma\gamma}|}\cos\sqrt{|t_{\gamma\gamma}|}}{\cosh 2\sqrt{|t_{\gamma\gamma}|}+\cos 2\sqrt{|t_{\gamma\gamma}|}}\right),
\end{aligned}
$$

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

Now,

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

and

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

where

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

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

Using the properties of noncentral quadratic forms in normal random variables, it can be shown that

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

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

$$\begin{aligned} \varkappa_{\gamma\gamma}(t_{\gamma\gamma}; z_\gamma, z_{\beta\beta}) &= \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi(t_\gamma, t_{\beta\beta}, t_{\gamma\gamma}) \exp\left[-i\left(t_\gamma z_\gamma + t_{\beta\beta}z_{\beta\beta}\right)\right] dt_\gamma dt_{\beta\beta} \\ &= \frac{|A+iB|^{-1/2}}{\pi\sqrt{\cosh\sqrt{-2it_{\gamma\gamma}}}} \exp\left[-\begin{pmatrix} z_\gamma \\ z_{\beta\beta}\end{pmatrix}'\left(AB^{-1}A+B\right)^{-1}AB^{-1}\begin{pmatrix} z_\gamma \\ z_{\beta\beta}\end{pmatrix}\right] \\ &\quad \times \exp\left[+i\begin{pmatrix} z_\gamma \\ z_{\beta\beta}\end{pmatrix}'\left(B^{-1} - B^{-1}A\left(AB^{-1}A+B\right)^{-1}AB^{-1}\right)\begin{pmatrix} z_\gamma \\ z_{\beta\beta}\end{pmatrix}\right], \end{aligned}$$

where $A = A(t_{\gamma\gamma})$ and $B = B(t_{\gamma\gamma})$.

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

$$\begin{aligned} f_{\mathcal{Z}_\gamma,\mathcal{Z}_{\beta\beta},\mathcal{Q}_{\gamma\gamma}}(z_\gamma, z_{\beta\beta}, q_{\gamma\gamma}) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \varkappa_{\gamma\gamma}(t_{\gamma\gamma}; z_\gamma, z_{\beta\beta}) \exp\left[-it_{\gamma\gamma}q_{\gamma\gamma}\right] dt_{\gamma\gamma} \\ &= \frac{1}{\pi}\int_0^{\infty} \operatorname{Re}\left\{\varkappa_{\gamma\gamma}(t_{\gamma\gamma}; z_\gamma, z_{\beta\beta}) \exp\left[-it_{\gamma\gamma}q_{\gamma\gamma}\right]\right\} dt_{\gamma\gamma}. \qquad \blacksquare \end{aligned}$$

Proof of Lemma 8. Lemma 8 (a) follows simple algebra and the relations

$$\sigma_{xx}^{-1/2}\sigma_{yy.x}^{-1/2}T^{-1}\sum_{t=1}^{T}x_t^m\left(y_t-\beta_0x_t-\sigma_{xx}^{-1}\sigma_{xy}\Delta x_t\right)=R_\beta^{CI}+o_{p_0}(1)$$

and

$$\sigma_{xx}^{-1}T^{-2}\sum_{t=1}^{T}x_t^{m2}=\sigma_{xx}^{-1}T^{-2}\sum_{t=1}^{T}x_t^m x_{t-1}^\mu+o_{p_0}(1)=R_{\beta\beta}^{CI}+o_{p_0}(1).$$

Lemma 8 (b) follows from standard weak convergence arguments (e.g., Phillips (1987, 1988a, 1988b) and Phillips and Solo (1992)). ■

Proof of Theorem 9. Theorem 9 follows as Theorem 5 using Lemma 8 instead of Lemma 3. ■

Proof of Theorem 10. Theorem 10 follows from standard weak convergence arguments (e.g., Phillips (1987, 1988a, 1988b) and Phillips and Solo (1992)) and the proof of Theorem 6. ■

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 (2003): "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 Edition. 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 (2003): "Efficient Tests of Stock Return Predictability," NBER Working Paper No. 10026.

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. Shick (1996): "Asymptotically Uniformly Most Powerful Tests in Parametric and Semiparametric Models," *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.

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

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

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

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

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.

Engle, R. F., and C. W. J. Granger (1987): "Cointegration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55, 251–276.

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.

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.

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.

Johansen, S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.

——— (1991): "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59, 1551–1580.

Kauppi, H. (2004): "On The Robustness of Hypothesis Testing Based on Fully Modified Vector Autoregression When some Roots are Almost One," *Econometric Theory*, 20, 341–359.

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 Edition. New York: Springer-Verlag.

Lehmann, E. L. (1994): *Testing Statistical Hypotheses*. Second Edition. New York: Chapman and Hall.

Lewellen, J. (2002): "Predicting Returns with Financial Ratios," Journal of Financial Economics, forthcoming.

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.

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.

Park, J. Y. (1992): "Canonical Cointegrating Regressions," *Econometrica*, 60, 119–143.

Phillips, P. C. B. (1987): "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.

——— (1991a): "Optimal Inference in Cointegrated Systems," *Econometrica*, 59, 283–306.

——— (1991b): "Spectral Regression for Cointegrated Time Series," in *Nonparametric and Semiparametric Methods in Economics and Statistics*, ed. by W. A. Barnett, J. Powell, and G. E. Tauchen. Cambridge, U.K.: Cambridge University Press, pp. 413-435.

——— (1995): "Fully Modified Least Squares and Vector Autoregression," *Econometrica*, 63, 1023–1078.

Phillips, P. C. B., and B. E. Hansen (1990): "Statistical Inference in Instrumental Variables Regression with I(1) Variables," *Review of Economic Studies*, 57, 99–125.

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

Polk, C., S. Thompson, and T. Vuolteenaho (2004): "New Forecasts of the Equity Premium," Working paper, Harvard University.

Saikkonen, P. (1991): "Asymptotically Efficient Estimation of Cointegration Regressions," *Econometric Theory*, 7, 1–21.

——— (1992): "Estimation and Testing of Cointegrated Systems by an Autoregressive Approximation," *Econometric Theory*, 8, 1–27.

Shorack, G. R., and J. A. Wellner (1986): *Empirical Processes with Applications to Statistics*. New York: Wiley.

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, Volume IV*, ed. by R. F. Engle, and D. L. McFadden. New York: North Holland, pp. 2739-2841.

——— (1997): "Cointegration, Long-Run Comovements, and Long-Horizon Forecasting," in *Advances in Economics and Econometrics: Theory and Applications, Seventh World Congress, Volume 3*, ed. by D. Kreps, and K. F. Wallis. Cambridge: Cambridge University Press, pp. 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 (2001): "On Predicting Stock Returns with Nearly Integrated Explanatory Variables," Journal of Business, forthcoming.

Valkanov, R. (1999): "The Term Structure with Highly Persistent Interest Rates," Working paper, Anserson School of Management.

van der Vaart, A. W. (2002): "The Statistical Work of Lucien Le Cam," *Annals of Statistics*, 30, 631–682.

Watson, M. W. (1994): "Vector Autoregressions and Cointegration," in *Handbook of Econometrics, Volume IV*, ed. by R. F. Engle, and D. L. McFadden. New York: North Holland, pp. 2843-2915.

Wijsman, R. A. (1958): "Incomplete Sufficient Statistics and Similar Tests," *Annals of Mathematical Statistics*, 29, 1028–1045.

Wright, J. H. (1999): "A Simple Approach to Robust Inference in a Cointegrating System," International Finance Discussion Paper 654, Board of Goveners.

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