Econometric Theory, **18**, 2002, 1309–1335. Printed in the United States of America.
DOI: 10.1017/S0266466602186026

REGRESSION THEORY FOR NEARLY COINTEGRATED TIME SERIES

Michael Jansson
University of California, Berkeley

Niels Haldrup
University of Aarhus

This paper proposes a notion of near cointegration and generalizes several existing results from the cointegration literature to the case of near cointegration. In particular, the properties of conventional cointegration methods under near cointegration are characterized, thereby investigating the robustness of cointegration methods. In addition, we obtain local asymptotic power functions of five cointegration tests that take cointegration as the null hypothesis.

1. INTRODUCTION

In a highly influential Monte Carlo study, Granger and Newbold (1974) considered regressions of independent random walks on each other and found that the usual significance test based on the regression F-statistic tends to over-reject the null. To describe this phenomenon, the term *spurious regression* was coined.[1] The numerical findings of Granger and Newbold are given an analytical explanation by Phillips (1986), whereas Park, Ouliaris, and Choi (1988) and Park (1990) provide further clarification. These authors consider regressions involving quite general integrated processes and show that the asymptotic properties of the appropriate F-statistic depend crucially on ρ^2, the squared multiple correlation coefficient computed from the long-run covariance matrix of the underlying innovation sequence. If $\rho^2 < 1$, the F-statistic diverges at rate T (where T is the sample size) whereas $T^{-1} \times F$ has a nondegenerate limiting distribution, which only depends on the dimension of the system. In other words, the regression is spurious whenever the coefficient of correlation is less than unity. In contrast, when $\rho^2 = 1$ the series are cointegrated and $F = O_p(1)$ with a complicated limiting distribution. Conventional asymptotic results therefore depend discontinuously on ρ^2.

This paper is an abridged version of Chapter 1 of the first author's Ph.D. dissertation at University of Aarhus. The paper has benefited from the comments of H. Peter Boswijk, Bruce Hansen (the co-editor), two anonymous referees, and seminar participants at Penn State University, Tinbergen Institute, University of California (Riverside, San Diego), the 1998 European Meeting of the Econometric Society, and the 1999 NOS-S conference on macroeconomic transmission mechanisms. Address correspondence to: Michael Jansson, Department of Economics, University of California, Berkeley, 549 Evans Hall #3880, Berkeley, CA 94720-3880, USA; e-mail: mjansson@econ.berkeley.edu.

On the other hand, it is quite obvious that the finite sample distribution of the F-statistic depends continuously on ρ^2. As a consequence, there is reason to believe that spurious regression asymptotics provide a poor approximation to the finite sample behavior of the F-statistic when the processes are "nearly" cointegrated in the sense that ρ^2 is "close" to unity. More generally, finite sample approximations based on spurious regression theory are likely to be of limited usefulness whenever the limiting behavior of the object of interest (e.g., an estimator or a test statistic) exhibits a discontinuity at $\rho^2 = 1$ and values of ρ^2 close to unity are of particular interest. In contrast, a model of near cointegration in which ρ^2 is a sequence of parameters lying in a shrinking neighborhood of unity as T tends to infinity is much more appealing in such situations.

Motivated by these considerations, the present paper introduces a model in which ρ^2 is a primitive parameter and uses this model to propose a notion of near cointegration.[2] By construction, the limiting behavior of the F-statistic depends continuously on ρ^2 in our setup, and the model of near cointegration can therefore be used to bridge the gap between spurious regression and cointegration with respect to the limiting behavior of the F-statistic.[3] We use the model of near cointegration to generalize several existing results from the cointegration literature to the case of near cointegration. In particular, the robustness of cointegration methods is investigated. We characterize the limiting behavior under near cointegration of the usual Wald statistic devised to test hypotheses on a cointegrating vector and show that the limiting distribution obtained under near cointegration stochastically dominates the χ^2 distribution applicable under cointegration. This result complements Elliott's recent study (1998), where the implications of near integration in exactly cointegrated models are examined. In addition to studying the robustness of cointegration methods, we characterize the behavior of five regression based cointegration tests under local alternatives and compute the corresponding local asymptotic power functions. Among the tests under study, three are found to have virtually identical local asymptotic power properties, whereas the remaining two are significantly inferior in terms of local asymptotic power.

The paper proceeds as follows. In Section 2, we introduce our model. Section 3 discusses the behavior of regression estimators under near cointegration, and Section 4 contains the corresponding results for inference procedures based on these estimators. In Section 5, we report the behavior of several cointegration tests under near cointegration. Finally, Section 6 offers a few concluding remarks. Proofs of all results of the paper are outlined in the Appendix.

Before we begin, a word on notation. The inequality > 0 signifies positive definiteness when applied to square matrices, and $\|\cdot\|$ is the Euclidean norm. For any symmetric $A > 0$, $A^{-1/2} = (A^{1/2})^{-1}$ and $A^{1/2}$ is the upper triangular matrix with positive diagonal elements such that $A^{1/2}A^{1/2\prime} = A$. The symbols $\overset{\mathcal{L}}{=}$, $\to_d$, and $\to_p$ signify equality in law, convergence in distribution, and convergence in probability, respectively. Finally, to simplify the notation integrals such as $\int_0^1 W(r)\,dr$ and stochastic integrals such as $\int_0^1 W(r)\,dW(r)'$ are typically written as $\int W$ and $\int W dW'$, respectively.

2. THE MODEL AND ASSUMPTIONS

Suppose $\{z_{Tt}: 0 \leq t \leq T, T \geq 1\}$ is an m-vector triangular array generated by

$$\Delta z_{Tt} = C_T(L)e_t, \tag{1}$$

where each $C_T(L) = \sum_{i=0}^{\infty} C_{Ti} L^i$ is an $m \times m$ matrix lag polynomial, whereas $\{e_t: t \in \mathbb{Z}\}$ is independent and identically distributed (i.i.d.) with $E(e_t) = 0$ and $E(e_t e_t') > 0$. Triangular arrays are considered to be able to define a notion of near cointegration similar to those of Tanaka (1993, 1996). Our objective is to do so by means of a parameterization of $\{C_T(L): T \geq 1\}$ and $E(e_t e_t')$ that involves minimal loss of generality for any fixed T and depends on T in the simplest possible fashion.

Applying the BN (Beveridge and Nelson, 1981) decomposition[4] to $C_T(L)$, we can write z_{Tt} as

$$z_{Tt} = C_T(1)\xi_t + \tilde{C}_T(L)e_t + z_{T0} - \tilde{C}_T(L)e_0,$$

where $\xi_t = \sum_{s=1}^{t} e_s$. We assume that $C_T(1)$ is upper triangular (with nonnegative diagonal elements) and $E(e_t e_t') = I_m$. For any fixed T, these assumptions entail no loss of generality.[5] Partition z_{Tt} into $m_y = 1$ and $m_x = m - 1$ components as $z_{Tt} = (y_{Tt}, x_{Tt}')'$. Conformably, partition $C_T(1)$ and $\tilde{C}_T(L)$ after the first row as $C_T(1) = (C_T^y(1)', C_T^x(1)')'$ and $\tilde{C}_T(L) = (\tilde{C}_T^y(L)', \tilde{C}_T^x(L)')'$, respectively. Assuming $C_T^y(1)$ and $C_T^x(1)$ have full row rank, $C_T(1)$ can be written as

$$C_T(1) = \begin{pmatrix} \omega_{yy,T}^{1/2}(1-\rho_T^2)^{1/2} & (\rho_T \Omega_{xx,T}^{-1/2}\bar{\omega}_{xy,T})' \\ 0 & \Omega_{xx,T}^{1/2} \end{pmatrix}, \tag{2}$$

where $\omega_{yy,T} > 0$ is a scalar, $\Omega_{xx,T} > 0$ is an $m_x \times m_x$ matrix, $\bar{\omega}_{xy}$ is an m_x-vector such that $\bar{\omega}_{xy,T}' \Omega_{xx,T}^{-1} \bar{\omega}_{xy,T} = \omega_{yy,T}$ and $0 \leq \rho_T \leq 1$. Under this parameterization, the rank of $C_T(1)$ depends solely on the scalar ρ_T. Indeed, $C_T(1)$ has full rank whenever $0 \leq \rho_T < 1$, whereas $C_T(1)$ has (deficient) rank $m-1$ if $\rho_T = 1$. This feature is very convenient for our purposes, as it enables us to define a notion of near cointegration by modeling $\{\rho_T\}$ as a sequence lying in a shrinking neighborhood of unity as T increases without bound.

In recognition of the fact that our main emphasis is on the cointegration properties of $\{z_{Tt}\}$, we henceforth make the simplifying assumption that ρ_T is the only parameter of the model that varies with T. To make this assumption explicit, the redundant subscript T will be omitted in expressions involving the parameters $C_T^x(L)$, $\tilde{C}_T(L)$, $\omega_{yy,T}$, $\Omega_{xx,T}$, and $\bar{\omega}_{xy,T}$ that do not vary with T. Likewise, x_{Tt} will be written as x_t.

Let $u_t = C^u(L)e_t$, where $C^u(L) = (1, -\beta')\tilde{C}(L)$ and $\beta = \Omega_{xx}^{-1}\bar{\omega}_{xy}$. Define $w_t = (u_t, \Delta x_t')$ and let Ω_{ww} be the long-run covariance matrix of w_t, namely,

$$\Omega_{ww} = \begin{pmatrix} \omega_{uu} & \omega_{xu}' \\ \omega_{xu} & \Omega_{xx} \end{pmatrix} = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} E(w_t w_s'),$$

where the partitioning is in conformity with w_t. The development of formal results will proceed under the following assumptions.

A1. $C_T(L) = C_T(1) + \tilde{C}(L)(1-L)$, where $\tilde{C}(L) = \sum_{i=0}^{\infty} \tilde{C}_i L^i$ is a lag polynomial with $\sum_{i=0}^{\infty} i\|\tilde{C}_i\| < \infty$, and $C_T(1)$ is defined as in (2) with $\omega_{yy}, \bar{\omega}_{xy}$, and Ω_{xx} fixed and $1-\rho_T^2 = T^{-2}\lambda^2\omega_{uu.x}/\omega_{yy}$ for some $\lambda \geq 0$, where $\omega_{uu.x} = \omega_{uu} - \omega_{xu}'\Omega_{xx}^{-1}\omega_{xu} > 0$ and $C^u(1)(1,0')' > 0$.

A2. $\{e_t\}$ is i.i.d. with $E(e_t) = 0$ and $E(e_t e_t') = I_m$.

A3. $z_{T0} = \tilde{C}(L)e_0$.

The long-run covariance matrix of Δz_{Tt} is $\Omega_{zz} = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} E(\Delta z_{Tt}\Delta z_{Ts}')$. In view of the relation

$$\Omega_{zz} = \lim_{T\to\infty} C_T(1)C_T(1)' = \lim_{T\to\infty}\begin{pmatrix} \omega_{yy} & \rho_T\bar{\omega}_{xy}' \\ \rho_T\bar{\omega}_{xy} & \Omega_{xx}\end{pmatrix} = \begin{pmatrix} \omega_{yy} & \bar{\omega}_{xy}' \\ \bar{\omega}_{xy} & \Omega_{xx}\end{pmatrix},$$

the parameters ρ_T, ω_{yy}, Ω_{xx}, and $\bar{\omega}_{xy}$ in A1 all have natural interpretations. Indeed, it is apparent that $C_T(1)$ is parameterized directly in terms of the elements of Ω_{zz} and the scalar ρ_T. In turn, ρ_T^2 is the squared coefficient of multiple correlation computed from $C_T(1)C_T(1)'$.

Under A1, $\{z_{Tt}\}$ is nearly cointegrated in the sense that ρ_T^2 lies in a shrinking neighborhood of unity as T increases. Of course, near cointegration reduces to cointegration when $\lambda = 0$ in A1. In that case, the assumption $\omega_{uu.x} > 0$ states that the cointegration is regular in the sense of Park (1992). On the other hand, when $\lambda \neq 0$, the condition $C^u(1)(1,0')' > 0$ can be interpreted as an identification assumption.[6] The assumption $\Omega_{xx} > 0$ implies that $\{x_t\}$ is a noncointegrated integrated process. Although somewhat restrictive, the assumption of noncointegrated regressors is fairly standard in the related literature,[7] so to facilitate comparisons with existing results we shall maintain this assumption throughout. Assumption A3 is introduced to avoid any complications that might arise as a result of a nonzero mean in z_{Tt} and/or a "remote past" initialization of the $\{z_{Tt}\}$ process (e.g., Canjels and Watson, 1997).

The parameter λ introduced in A1 will play a prominent role in the sequel. Because

$$\lambda = \frac{T\cdot\omega_{yy}^{1/2}(1-\rho_T^2)^{1/2}}{\omega_{uu.x}^{1/2}}, \tag{3}$$

λ can be interpreted as a signal-to-noise ratio. Specifically, the numerator in (3) is proportional to $\omega_{yy}^{1/2}(1-\rho_T^2)^{1/2}$, which is the long-run standard deviation of Δy_{Tt} conditional on Δx_t in the case where ρ_T does not vary with T. The denominator, $\omega_{uu.x}^{1/2}$, is the long-run standard deviation of u_t conditional on Δx_t. Under cointegration, the former is zero and $\lambda = 0$. Under spurious regression (when $\rho_T < 1$ is fixed), on the other hand, the right-hand side of (3) diverges.

Near cointegration corresponds to the intermediate case where the numerator and denominator of (3) are of the same order of magnitude.

Our model can be written in triangular form as

$$y_{Tt} = \beta' x_t + \delta_T' \xi_t + C^u(L) e_t,$$

$$\Delta x_t = C^x(L) e_t, \tag{4}$$

where $\delta_T = (\omega_{yy}^{1/2}(1-\rho_T^2)^{1/2}, (\rho_T - 1)\bar{\omega}_{xy}' \Omega_{xx}^{-1/2\prime})'$. As it turns out, the presence of the additional "error term" $\delta_T'\xi_t$ on the right-hand side of (4) leads to an increase in the asymptotic variance of estimators of β. On the other hand, because $\delta_T'\xi_t$ is asymptotically uncorrelated with "the regressor" x_t in (4) in the sense that the long-run covariance between Δx_t and $T\delta_T' e_t$, namely, $\lim_{T\to\infty} T \cdot C^x(1)\delta_T$, equals zero,[8] the presence of $\delta_T'\xi_t$ does not affect the asymptotic bias of estimators of β.

Notions of near cointegration that are closely related to ours have been proposed by Tanaka (1993, 1996).[9,10] The near cointegration model of Tanaka (1993, equation (20)) can be written in triangular form as

$$y_{Tt} = \beta' x_t + \delta_T' \xi_t + C_T^u(L) e_t,$$

$$\Delta x_t = C^x(L) e_t,$$

where $\xi_t = \sum_{s=1}^t e_s$, $\{e_t\}$ satisfies A2, $C_T^u(L) = C^u(L) + T^{-1}\dot{C}^u(L)$, and $\lim_{T\to\infty} T \cdot C^x(1)\delta_T = 0$. The asymptotic distributions of interest are unaffected by the presence of the term $T^{-1}\dot{C}^u(L)e_t$, and Tanaka's model (1993) yields results that coincide with the results of the present paper.

Tanaka's near cointegration model (1996, equations (11.68) and (11.70)) can be written as

$$y_{Tt} = \beta_T' x_t + \delta_T' \xi_t + C_T^u(L) e_t,$$

$$\Delta x_t = C^x(L) e_t,$$

where $\xi_t = \sum_{s=1}^t e_s$, $\{e_t\}$ satisfies A2, $\beta_T = \beta + T^{-1}\dot{\beta}$, $C_T^u(L) = C^u(L) + T^{-1}\dot{C}^u(L)$, and $\lim_{T\to\infty} T \cdot C^x(1)\delta_T = 0$. The term $\dot{\beta} = T(\beta_T - \beta)$ is nonzero in general and gives rise to a drift term in the limiting distribution of estimators of β (such as $\hat{\beta}_T$ and $\hat{\beta}_T^\dagger$ defined in Section 3) but does not affect the limiting distribution of cointegration tests, the objects of interest in Tanaka (1996).

Unlike the parameterizations employed by Tanaka (1993, 1996), the parameterization of near cointegration proposed here is explicitly one-dimensional (involving only the scalar parameter λ) and therefore leads to simpler representations of the limiting distributions of interest, which facilitates the interpretation of the results.

3. BEHAVIOR OF REGRESSION ESTIMATORS

Let $\hat{\alpha}_T$ and $\hat{\beta}_T$ be the ordinary least squares (OLS) estimators in the multiple regression

$$y_{Tt} = \hat{\alpha}_T' d_t + \hat{\beta}_T' x_t + \hat{u}_{Tt} \qquad (t = 1, \ldots, T), \tag{5}$$

where $d_t = (1, \ldots, t^{m_d - 1})'$ for some $m_d \geq 1$. Let $\Psi_T = \operatorname{diag}(T^{1/2}, \ldots, T^{m_d - 1/2}, T \cdot \iota_{m_x}')$, where ι_{m_x} is an m_x-vector of ones. As is well known (e.g., Phillips and Durlauf, 1986), the limiting distribution of $\Psi_T(\hat{\alpha}_T', (\hat{\beta}_T - \beta)')'$ under cointegration is rather complicated and depends on the parameters Ω_{ww} and Γ_{ww}, where Ω_{ww} was defined in Section 2, whereas

$$\Gamma_{ww} = (\Gamma_{\cdot u} \quad \Gamma_{\cdot x}) = \begin{pmatrix} \gamma_{uu} & \gamma_{ux} \\ \gamma_{xu} & \Gamma_{xx} \end{pmatrix} = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{t} E(w_t w_s').$$

A similar situation occurs under near cointegration.

LEMMA 1. *Suppose $\{z_{Tt}\}$ is generated by (1) and suppose A1–A3 hold. Then*

$$\Psi_T \begin{pmatrix} \hat{\alpha}_T \\ \hat{\beta}_T - \beta \end{pmatrix} \to_d \left(\int Q_x Q_x' \right)^{-1} \left(\omega_{uu.x}^{1/2} \int Q_x \, dU_\lambda + \int Q_x \, dX' \Omega_{xx}^{-1} \omega_{xu} + \begin{pmatrix} 0 \\ \gamma_{ux}' \end{pmatrix} \right),$$

where $Q_x(r)' = (D(r)', X(r)')$, $D(r) = (1, r, \ldots, r^{m_d - 1})'$, $X(r) = \Omega_{xx}^{1/2} V(r)$, and $U_\lambda(r) = \lambda \int_0^r U(s)\, ds + U(r)$, whereas V and U are independent Wiener processes of dimension m_x and 1, respectively.

In addition to $(\hat{\alpha}_T', \hat{\beta}_T')'$, we want to study an estimator that has a compound normal limiting distribution under cointegration. For concreteness, we consider the canonical cointegration regression (CCR) estimator proposed by Park (1992).[11] To construct this estimator, we need consistent estimators of Ω_{ww}, Γ_{ww}, and $\Sigma_{ww} = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E(w_t w_t')$. Let $\hat{w}_{Tt}' = (\hat{u}_{Tt}, \Delta x_t')$, where $\{\hat{u}_{Tt}\}$ are the residuals from (5). We can estimate Σ_{ww} by $\hat{\Sigma}_{ww} = T^{-1} \sum_{t=1}^{T} \hat{w}_{Tt} \hat{w}_{Tt}'$, whereas Ω_{ww} and Γ_{ww} can be estimated by kernel estimators of the form

$$\hat{\Omega}_{ww} = T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} k\left(\frac{|t - s|}{\hat{b}_T} \right) \hat{w}_{Tt} \hat{w}_{Ts}'$$

and

$$\hat{\Gamma}_{ww} = T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{t} k\left(\frac{|t - s|}{\hat{b}_T} \right) \hat{w}_{Tt} \hat{w}_{Ts}',$$

where $k(\cdot)$ is a kernel function and $\{\hat{b}_T\}$ is a sequence of (possibly sample-dependent) bandwidth parameters.[12] As shown in Lemma 5 in the Appendix, the estimators $\hat{\Omega}_{ww}$, $\hat{\Gamma}_{ww}$, and $\hat{\Sigma}_{ww}$ are consistent under A1–A4, where

A4.

(i) $k(0) = 1, k(\cdot)$ is continuous at zero, $\sup_{s\geq 0}|k(s)| < \infty$, and $\int_0^\infty \bar{k}(r)\,dr < \infty$, where $\bar{k}(r) = \sup_{s\geq r}|k(s)|$ (for all $r \geq 0$)

(ii) $\hat{b}_T = \hat{a}_T b_T$, where $\hat{a}_T$ and b_T are positive with $\hat{a}_T + \hat{a}_T^{-1} = O_p(1)$ and $b_T^{-1} + T^{-1/2} b_T = o(1)$.

The CCR estimator $(\hat{\alpha}_T^{\dagger\prime}, \hat{\beta}_T^{\dagger\prime})'$ is the OLS estimator obtained from the multiple regression

$$y_{Tt}^{\dagger} = \hat{\alpha}_T^{\dagger\prime} d_t + \hat{\beta}_T^{\dagger\prime} x_{Tt}^{\dagger} + \hat{u}_{Tt}^{\dagger} \qquad (t = 1, \ldots, T), \tag{6}$$

where $y_{Tt}^{\dagger} = y_{Tt} - \hat{\beta}_T' \hat{\Gamma}_{\cdot x}' \hat{\Sigma}_{ww}^{-1} \hat{w}_{Tt} - \hat{\omega}_{xu}' \hat{\Omega}_{xx}^{-1} \Delta x_t$, $x_{Tt}^{\dagger} = x_t - \hat{\Gamma}_{\cdot x}' \hat{\Sigma}_{ww}^{-1} \hat{w}_{Tt}$, and $\hat{\beta}_T$ is the OLS estimator from (5).

LEMMA 2. *Suppose $\{z_{Tt}\}$ is generated by (1) and suppose A1–A4 hold. Then*

$$\Psi_T \begin{pmatrix} \hat{\alpha}_T^{\dagger} \\ \hat{\beta}_T^{\dagger} - \beta \end{pmatrix} \to_d \left(\int Q_x Q_x' \right)^{-1} \left(\omega_{uu.x}^{1/2} \int Q_x \, dU_\lambda \right),$$

where Ψ_T, Q_x, and U_λ are defined as in Lemma 1. The limiting distribution is compound normal:

$$\left(\int Q_x Q_x' \right)^{-1} \left(\omega_{uu.x}^{1/2} \int Q_x \, dU_\lambda \right) \Bigg|_{\mathcal{F}_V}$$
$$\stackrel{\mathcal{L}}{=} \mathcal{N}\left(0, \omega_{uu.x} \left(\int Q_x Q_x' \right)^{-1} \left(\int Q_{x,\lambda} Q_{x,\lambda}' \right) \left(\int Q_x Q_x' \right)^{-1} \right),$$

where $Q_{x,\lambda}(r) = \lambda \int_r^1 Q_x(s)\,ds + Q_x(r)$ and $\cdot|_{\mathcal{F}_V}$ signifies the conditional distribution relative to $\mathcal{F}_V = \sigma(V(r) : 0 \leq r \leq 1)$.

Tanaka (1993) obtains a result equivalent to the first half of Lemma 2.[13] In important respects, the near cointegration case closely resembles the cointegration case. For instance, $\hat{\beta}_T$ and $\hat{\beta}_T^{\dagger}$ are superconsistent estimators of β. Moreover, the limiting distribution of $T(\hat{\beta}_T^{\dagger} - \beta)$ is compound normal. The mean of this compound normal distribution is zero even under near cointegration, and the presence of $\delta_T' \xi_t$ on the right-hand side of (4) therefore does not introduce a bias term in the limiting distribution of $\hat{\beta}_T^{\dagger}$. Under cointegration (i.e., when $\lambda = 0$), $\int Q_{x,\lambda} Q_{x,\lambda}' = \int Q_x Q_x'$ and the covariance matrix in the mixture representation in Lemma 2 reduces to $\omega_{uu.x}(\int Q_x Q_x')^{-1}$. Otherwise, if $\lambda \neq 0$, the covariance matrix is of "sandwich" form. As pointed out by the co-editor, this suggests that OLS-type inference (based on the CCR estimates) will be misleading under near cointegration. Indeed, defining $\underline{Q_x}(r) = \int_r^1 Q_x(s)\, ds$, we have

$$\int Q_{x,\lambda} Q'_{x,\lambda} - \int Q_x Q'_x = \lambda \underline{Q_x}(0)\underline{Q_x}(0)' + \lambda^2 \int \underline{Q_x}\,\underline{Q_x}' > 0.$$

The presence of $\delta'_T \xi_t$ on the right-hand side of (4) therefore leads to an increase in the asymptotic variance of $\hat{\beta}^\dagger_T$, suggesting that overrejection of true null hypotheses (on β) is likely to occur under near cointegration. A more precise statement corroborating this conjecture will be provided in the next section.

4. INFERENCE ON REGRESSION COEFFICIENTS

This section is concerned with inference on regression coefficients. Particular attention is given to the robustness of conventional cointegration procedures under near cointegration. Consider a general linear hypothesis of the form $H_0 : \Phi_\beta \beta = \phi_\beta$, where Φ_β is a $p \times m_x$ matrix of rank p and ϕ_β is a p-vector.[14] Define

$$G_T = \hat{\omega}^{-1}_{uu.x} (\Phi_\beta \hat{\beta}^\dagger_T - \phi_\beta)' \left[\Phi_\beta \left(\sum_{t=1}^T x^\dagger_{Tt,d} x^{\dagger\prime}_{Tt,d} \right)^{-1} \Phi'_\beta \right]^{-1} (\Phi_\beta \hat{\beta}^\dagger_T - \phi_\beta), \tag{7}$$

where $\hat{\omega}_{uu.x} = \hat{\omega}_{uu} - \hat{\omega}'_{xu} \hat{\Omega}^{-1}_{xx} \hat{\omega}_{xu}$ and $x^\dagger_{Tt,d} = x^\dagger_{Tt} - (\sum_{s=1}^T x^\dagger_{Ts} d'_s)(\sum_{s=1}^T d_s d'_s)^{-1} d_t$.

THEOREM 3. *Suppose $\{z_{Tt}\}$ is generated by (1) and suppose A1–A4 hold.*

(a) When $H_0 : \Phi_\beta \beta = \phi_\beta$ is true,

$$G_T \to_d \left\| \int \widetilde{V^p_D}\, dU_\lambda \right\|^2,$$

where $\widetilde{V^p_D}(r) = (\int V^p_D V^{p\prime}_D)^{-1/2} V^p_D(r)$, $V^p_D(r) = V_{1D}(r) - (\int V_{1D} V'_{2D}) \times (\int V_{2D} V'_{2D})^{-1} V_{2D}(r)$, and $V_D(r) = (V_{1D}(r)', V_{2D}(r)')' = V(r) - (\int V D') \times (\int D D')^{-1} D(r)$, where the partitioning is after the pth row, whereas V, D, and U_λ are defined as in Lemma 1.

(b) The limiting distribution in part (a) satisfies

$$\left\| \int \widetilde{V^p_D}\, dU_\lambda \right\|^2 \Bigg|_{\mathcal{F}_V} \overset{\mathcal{L}}{=} \sum_{i=1}^p (1 + \lambda^2 \cdot \mu_i) \chi^2_i(1),$$

where $\{\chi^2_i(1)\}_{i=1}^p$ are i.i.d. $\chi^2(1)$ variates and $0 \le \mu_1 \le \cdots \le \mu_p$ are the eigenvalues of the matrix $\int_0^1 \underline{\widetilde{V^p_D}}(r) \underline{\widetilde{V^p_D}}(r)' dr$, where $\underline{\widetilde{V^p_D}}(r) = \int_r^1 \widetilde{V^p_D}(s)\, ds$. In particular, the family $\{\mathcal{L}(\| \int \widetilde{V^p_D}\, dU_\lambda \|^2) : \lambda \ge 0\}$ is stochastically increasing in λ, where $\mathcal{L}(\cdot)$ denotes the probability law of the argument.

Under cointegration, the limiting distribution of G_T is $\chi^2(p)$ when H_0 is true. In a recent paper, Elliott (1998) investigates the robustness of this remarkable result by considering a model in which the regressors are nearly integrated whereas some linear combination of the regressand and the regressor is exactly stationary. It turns out that the χ^2 result can break down when the regressors are not exactly integrated. Theorem 3 enables us to conduct a complimentary experiment: we can investigate the robustness of cointegration methods in a model where the regressors are exactly integrated whereas some linear combination of the regressand and the regressors is nearly stationary. It follows from Theorem 3(b) that tests based on the distribution applicable under cointegration (the $\chi^2(p)$ distribution) are oversized (asymptotically) under near cointegration. That is,

$$\lim_{T\to\infty} \Pr(G_T > t) = \Pr\left(\left\| \int \widetilde{V_D^p}\, dU_\lambda \right\|^2 > t\right) > \Pr\left(\left\| \int \widetilde{V_D^p}\, dU_0 \right\|^2 > t\right)$$
$$= \Pr(\chi^2(p) > t)$$

for all $t > 0$ whenever $\lambda > 0$.

For concreteness, consider the case where $m_d = 1, \Phi_\beta = I_{m_x}$, and $\phi_\beta = \beta_0$. In other words, consider the null hypothesis $H_0 : \beta = \beta_0$ in a regression of $y_{Tt}^\dagger$ on $x_{Tt}^\dagger$ and a constant. To illustrate the magnitude of the size distortions encountered under near cointegration, we have simulated the limiting distribution of G_T for $m_x = 1, \ldots, 4$ and for various values of λ. Specifically, we have made 20,000 draws from the distribution of the discrete approximations (using 2,000 steps) to the limiting random variables. Figure 1 plots the rejection frequencies corresponding to a test with a nominal size of 5%.

The evidence presented in Figure 1 suggests that severe size distortions can occur if conventional cointegration methods are being used when the series are nearly cointegrated rather than exactly cointegrated. In fact, the size increases dramatically as (the absolute value of) λ increases from 0, and substantial size distortions are encountered even for values of λ in the range 5–10. Whether or not this is a problem obviously depends on whether or not researchers can be expected to be able to detect such departures from exact cointegration. It is therefore of interest to know whether or not tests for cointegration can be expected to reject the null hypothesis of cointegration when λ is equal to 10, say. A partial answer to this question is provided in the next section, where we illustrate how to obtain the local asymptotic power functions of tests for cointegration.

5. LOCAL ASYMPTOTIC POWER OF COINTEGRATION TESTS

During the last decade, numerous cointegration tests taking cointegration as the null hypothesis have been proposed. These test procedures utilize different properties of cointegrated systems, and it therefore seems desirable to investigate what, if anything, can be said about the power properties of the different tests. In this section, we characterize the behavior of five regression based cointegration tests under local alternatives.[15] Moreover, we obtain the correspond-

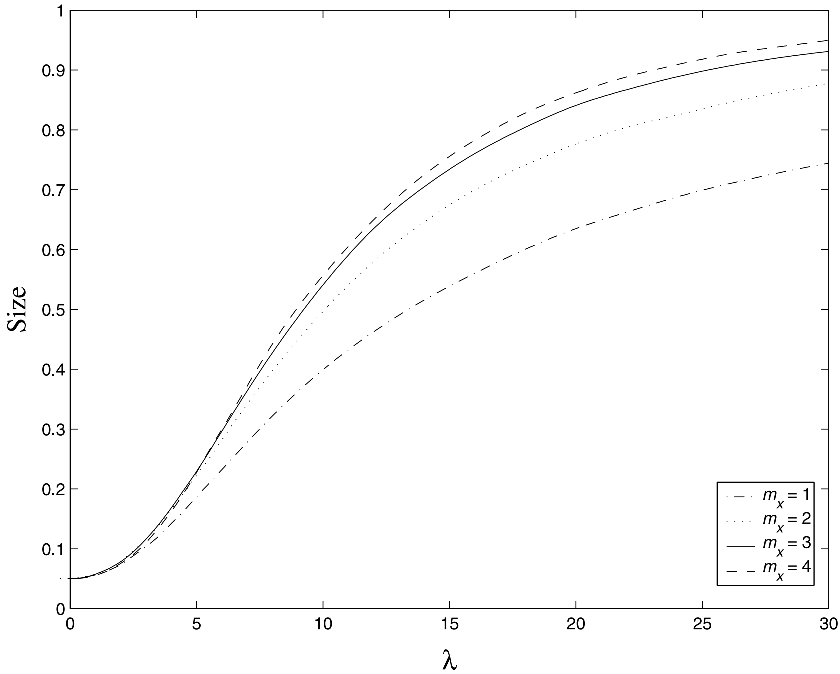

Figure 1. Rejection rates for G_T. Nominal size is 5%.

ing local asymptotic power functions and use these to address the following questions.

(i) Does any one of these tests dominate the others in terms of local asymptotic power?
(ii) Can cointegration tests be expected to detect those departures from cointegration that seriously distort the size of conventional cointegration procedures (cf. Section 4)?

The variable addition test proposed by Park (1990) is computed as follows. Let k_1 and k_2 be arbitrary nonnegative integers such that $k = k_1 + k_2 \geq 1$ and for $t = 1, \ldots, T$, let $r_{1t} = (t^{m_d}, \ldots, t^{m_d + k_1 - 1})'$ (if $k_1 \geq 1$) and (if $k_2 \geq 1$) let $\{r_{2t}\}$ be a k_2-dimensional computer generated random walk such that $\{\Delta r_{2t}\} \sim i.i.d.$ $\mathcal{N}(0, I_{k_2})$.[16] Finally, let $r_t' = (r_{1t}', r_{2t}')$. Based on the multiple regressions (6) and

$$y_{Tt}^{\dagger} = \ddot{\alpha}_T^{\dagger\prime} d_t + \ddot{\beta}_T^{\dagger\prime} x_{Tt}^{\dagger} + \ddot{\gamma}_T^{\dagger\prime} r_t + \ddot{u}_{Tt}^{\dagger} \qquad (t = 1, \ldots, T), \tag{8}$$

construct the statistic $J_T(k_1, k_2) = \hat{\omega}_{uu.x}^{-1}[\sum_{t=1}^{T}(\hat{u}_{Tt}^{\dagger})^2 - \sum_{t=1}^{T}(\ddot{u}_{Tt}^{\dagger})^2]$. This is simply the Wald test used to test the significance of the regressor r_t in (8). As a consequence, $J_T(k_1,k_2) \to_d \chi^2(k)$ under the null hypothesis of cointegration (Park, 1990).

Several cointegration tests based on partial score sums have been proposed. We consider the test due to Shin (1994).[17] That test is based on $CI_T = \hat{\omega}_{uu.x}^{-1} T^{-2} \sum_{t=1}^{T} \hat{S}_{Tt}^2$, where $\hat{S}_{Tt} = \sum_{s=1}^{t} \hat{u}_{Ts}^{\dagger}$.[18] Evidently, CI_T is constructed by applying the stationarity test proposed by Kwiatkowski, Phillips, Schmidt, and Shin (1992) to the residuals $\{\hat{u}_{Tt}^{\dagger}\}$ from (6).

Cointegration tests also can be based on the residuals $\{\check{S}_{Tt}\}$ from the multiple regression

$$S_{Tt}^{y} = \check{\alpha}_T' S_t^d + \check{\beta}_T' S_{Tt}^x + \check{S}_{Tt},$$

where $S_{Tt}^y = \sum_{s=1}^{t} y_{Ts}^{\dagger}, S_t^d = \sum_{s=1}^{t} d_s$, and $S_{Tt}^x = \sum_{s=1}^{t} x_{Ts}^{\dagger}$ for $1 \le t \le T$. Choi and Ahn (1995) propose the statistics $LM_T^I = [\hat{\omega}_{uu.x}^{-1}(T^{-1}\sum_{t=2}^{T} \check{S}_{T,t-1} \Delta \check{S}_{Tt} - \hat{\gamma}_{uu.x}^{+})]^2$, $LM_T^{II} = (\hat{\omega}_{uu.x}^{-1} \cdot T^{-2} \sum_{t=2}^{T} \check{S}_{T,t-1}^2)^{-1} \cdot LM_T^I$, and $SBDH_T^I = \hat{\omega}_{uu.x}^{-1} T^{-2} \sum_{t=1}^{T} \check{S}_{Tt}^2$, where $\hat{\gamma}_{uu.x}^{+} = (1, -\hat{\omega}_{xu}' \hat{\Omega}_{xx}^{-1})(\hat{\Gamma}_{ww} - \hat{\Sigma}_{ww})(1, -\hat{\omega}_{xu}' \hat{\Omega}_{xx}^{-1})'$. These tests are intimately related to the stationarity tests of Choi and Ahn (1998).

THEOREM 4. *Suppose $\{z_{Tt}\}$ is generated by (1) and suppose A1–A4 hold. Then*

$$J_T(k_1,k_2) \to_d \left\| \int \widetilde{R_Q}\, dU_\lambda \right\|^2,$$

$$CI_T \to_d \int \hat{U}_\lambda^2,$$

$$SBDH_T^I \to_d \int U_{\lambda,\bar{Q}}^2,$$

$$LM_T^I \to_d \left(\int U_{\lambda,\bar{Q}}\, dU_{\lambda,\bar{Q}} \right)^2,$$

$$LM_T^{II} \to_d \frac{\left(\int U_{\lambda,\bar{Q}}\, dU_{\lambda,\bar{Q}} \right)^2}{\int U_{\lambda,\bar{Q}}^2},$$

where $\widetilde{R_Q}(r) = (\int R_Q R_Q')^{-1/2} R_Q(r)$, $R_Q(r) = R(r) - (\int RQ')(\int QQ')^{-1} Q(r)$, $Q' = (D', V')$, $R' = (R_1', R_2')$, $R_1(r) = (r^{m_d}, \dots, r^{m_d+k_1-1})'$, R_2 is a k_2-dimensional Wiener process independent of Q and U_λ, $\hat{U}_\lambda(r) = U_\lambda(r) - (\int Q dU_\lambda)'(\int QQ')^{-1} \bar{Q}(r)$, and $U_{\lambda,\bar{Q}}(r) = U_\lambda(r) - (\int \bar{Q} U_\lambda)'(\int \overline{QQ'})^{-1} \bar{Q}(r)$, $\bar{Q}(r) = \int_0^r Q(s)\, ds$, whereas D, V, and U_λ are defined as in Lemma 1.

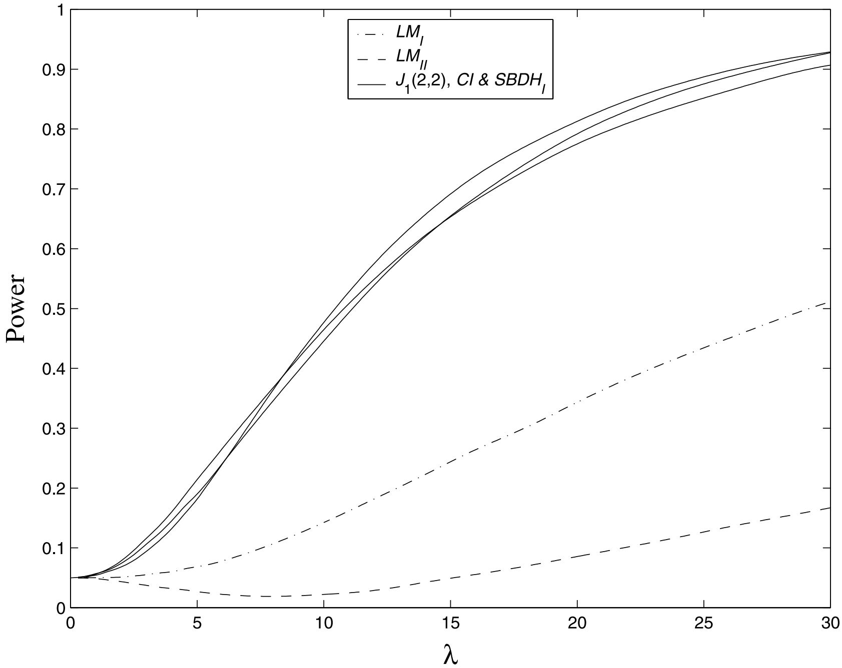

FIGURE 2. Local power of tests for cointegration; $m_d = 1$, $m_x = 1$.

An expression equivalent to the representation of the limiting distribution of CI_T is obtained by Tanaka (1996, Theorem 11.11).[19] To obtain local asymptotic power functions, we have simulated (the discrete time counterparts of) the limiting distributions of the $J_T(2,2)$,[20] CI_T, LM_T^I, LM_T^{II}, and $SBDH_T^I$ test statistics in the case where $m_d = 1$ and $m_x = 1$.[21] As in Section 4, we have used 2,000 steps and have repeated the procedure 20,000 times. Figure 2 shows the local asymptotic power functions of tests with size 5%.

The local asymptotic power properties of $J_T(2,2)$, CI_T, and $SBDH_T^I$ are very similar, whereas LM_T^{II} and (in particular) LM_T^I are remarkably inferior in terms of local asymptotic power. Because the local asymptotic power properties of $J_T(2,2)$, CI_T, and $SBDH_T^I$ are almost indistinguishable, our tentative conclusion is that the choice among these tests should be guided by finite sample considerations concerning size distortions.

Remark 1. Notice that $LM_T^{II} = (SBDH_T^I)^{-1} \cdot LM_T^I + o_p(1)$. Under fixed alternatives (i.e., under spurious regression), LM_T^I diverges at a faster rate than $SBDH_T^I$, and a test based on LM_T^{II} is therefore consistent (Choi and Ahn, 1995). In contrast, because both LM_T^I and $SBDH_T^I$ are $O_p(1)$ under near cointegration, LM_T^{II} might be expected to have rather disastrous local asymptotic power properties if the local asymptotic power of $SBDH_T^I$ is higher than the local asymptotic power of LM_T^I. Figure 2 confirms this conjecture. More generally, our findings illustrate the obvious, but important, point that the (local asymptotic) power properties of a test cannot be deduced from the rate of divergence under fixed alternatives. In the present example, e.g., LM_T^I and $SBDH_T^I$ diverge at the same rate under fixed alternatives and LM_T^I diverges faster than both of these (Choi and Ahn, 1995). Evidently, Figure 2 tells an entirely different story.

Remark 2. The local asymptotic power properties of the tests depend solely on λ. In particular, our distributional results do not depend on the particular estimator used to estimate nuisance parameters such as $\omega_{uu.x}$. In fact, the asymptotic results are the same as if these nuisance parameters were known. As pointed out by a referee, this is somewhat unfortunate, because there is ample (simulation) evidence documenting that the finite sample size properties of tests can be very sensitive to the choice of nuisance parameter estimation method (see McCabe, Leybourne, and Shin, 1997, and references therein). We share this view and encourage the reader to interpret the local asymptotic power curves presented here as approximations to the finite sample size-adjusted power curves (as opposed to the true power curves) of the corresponding tests.

In the previous section, we argued that Wald tests based on conventional cointegration methods can encounter severe size distortions when the series are nearly cointegrated and λ exceeds 5. On the other hand, the evidence presented in Figure 2 indicates that even when $\lambda = 10$ the power of the tests for cointegration can be well below 50%. This suggests that even if the departure from (exact) cointegration is substantial (in the sense that it severely affects the size of the conventional tests), tests for cointegration cannot be expected to detect such departures very frequently. Therefore, whenever a researcher rejects a structural hypothesis (on the coefficient β) using cointegration methods, the result should be interpreted carefully. Indeed, it might be the case that the structural hypothesis is correct, whereas the (possibly auxiliary) assumption of cointegration is not. This of course leaves open the question of how to interpret the coefficient vector in a noncointegrated system, a question that we shall not attempt to answer here.[22]

6. CONCLUDING REMARKS

Based on a new representation, a notion of near cointegration was proposed. The notion of near cointegration was used to generalize several existing results from the cointegration literature to the case of near cointegration. Throughout, we have deliberately studied the properties of known inference procedures under near cointegration rather than proposed new methods. As a result, several extensions are possible. For instance, a companion paper by one of us (Jansson, 2001c) takes the analysis of Section 5 one step further and uses a model of near cointegration to propose a new cointegration test with (essentially) optimal local asymptotic power properties.

NOTES

1. Earlier, Yule (1926) used the term *nonsense correlation* to describe a similar phenomenon.

2. In the aforementioned papers, ρ^2 is computed from a long-run covariance matrix that is itself defined by taking limits as $T \to \infty$. Therefore, it is not immediately obvious how to model ρ^2 as a sequence of parameters that lie in a shrinking neighborhood of unity as T increases without bound. By working with a representation where ρ^2 is a primitive parameter, we circumvent this potential problem.

3. A previous version of this article (Jansson and Haldrup, 2000) contains a detailed study of the F-statistic.

4. $C_T(L) = C_T(1) + \tilde{C}_T(L)(1-L)$, where $\tilde{C}_T(L) = \sum_{i=0}^{\infty} \tilde{C}_{Ti} L^i$ is a lag polynomial with coefficients $\tilde{C}_{Ti} = -\sum_{j=i+1}^{\infty} C_{Tj}$. These coefficients satisfy $\sum_{i=0}^{\infty} i \|\tilde{C}_{Ti}\| < \infty$ (as required under Assumption A1, which appears later in this section) if $\sum_{i=0}^{\infty} i^2 \|C_{Ti}\| < \infty$.

5. Suppose $\bar{C}(L) = \sum_{i=0}^{\infty} \bar{C}_i L^i$ is a matrix lag polynomial and suppose $\{\bar{e}_t : t \in \mathbb{Z}\}$ is i.i.d. with $E(\bar{e}_t) = 0$ and $E(\bar{e}_t \bar{e}_t') = \bar{\Sigma} > 0$. We can construct an orthogonal matrix $\mathcal{O}$ such that $\bar{C}(1)\bar{\Sigma}^{1/2}\mathcal{O}$ is upper triangular (with nonnegative diagonal elements). Given any such $\mathcal{O}$, define $C(L) = \bar{C}(L)\bar{\Sigma}^{1/2}\mathcal{O}$ and $e_t = \mathcal{O}'\Sigma^{-1/2}\bar{e}_t$. By construction, $C(1)$ is upper triangular and $\{e_t\}$ is i.i.d. with $E(e_t) = 0$ and $E(e_t e_t') = \mathcal{O}'\bar{\Sigma}^{-1/2}\bar{\Sigma}\bar{\Sigma}^{-1/2\prime}\mathcal{O} = I_m$.

6. For $T \geq 1$, let

$$D_T(L) = \begin{pmatrix} 1 & -\rho_T \beta' \\ 0 & I_{m_x} \end{pmatrix} C_T(L).$$

It is not hard to show that $C^u(1)(1,0')' > 0$ holds under the identification/invertibility condition $\inf\{|z| : |D_T(z)| = 0\} > 1 \quad \forall T, \lambda > 0$.

7. Notable exceptions include Park and Phillips (1989, Sec. 5.2), Choi (1994), and McCabe et al. (1997). See also Phillips (1995) and Chang and Phillips (1995).

8. Indeed, $T \cdot C^x(1)\delta_T = T(\rho_T - 1)\bar{\omega}_{xy} = O(T^{-1})$ under A1.

9. Alternative conditions of near cointegration have appeared in Quintos and Phillips (1993, Sec. 5) and Phillips (1998a, p. 1025). The (multivariate extension of the) notion of near cointegration introduced by Quintos and Phillips (1993) is more general than the notion suggested in the present paper. On the other hand, the notion of near cointegration discussed in Phillips (1998a) is fundamentally different from ours, because the series $\{h' y_t\}$ generated by equation (5) of that paper is nearly integrated.

10. Details concerning the derivation of the triangular form of Tanaka's models are available from the authors upon request.

11. Alternative estimators with identical asymptotic properties include the estimators proposed by Johansen (1988, 1991), Phillips (1991), Phillips and Hansen (1990), Saikkonen (1991, 1992), and Stock and Watson (1993).

12. For convenience, we do not make the dependence of $\hat{\Sigma}_{ww}$, $\hat{\Omega}_{ww}$, and $\hat{\Gamma}_{ww}$ on T and $\hat{b}_T$ explicit.

13. Using Tanaka's (1993) notation, the representation of the limiting distribution of $T(\tilde{\beta}_j - \beta)$ in Theorem 6 of that paper should read

$$(A_1' W_j A_1)^{-1}[A_1' W_j \delta + A_1' V_j (g - A_1(A_1' A_1)^{-1} A_1' g)].$$

Upon subtraction of $(A_1' A_1)^{-1} A_1' \delta$, we arrive at the representation

$$(A_1' W_j A_1)^{-1}[A_1' W_j (\delta - A_1(A_1' A_1)^{-1} A_1' \delta) + A_1' V_j (g - A_1(A_1' A_1)^{-1} A_1' g)],$$

which is equivalent to the result in Lemma 2. The difference in the location parameter is due to the fact that Tanaka (1993, p. 49) defines the population value of the regression coefficient as $\beta = (A_1'A_1)^{-1}(A_1'A_2 - A_1'\delta/T)$, whereas our β equals $(A_1'A_1)^{-1}A_1'A_2$ in Tanaka's (1993) notation.

14. It is a simple matter to generalize Theorem 3 to the case of nonlinear hypotheses. To conserve space, we shall not do so here.

15. Harris (1997) and Snell (1998) propose tests for cointegration that utilize principal component methods. These tests are not considered here.

16. This particular choice of superfluous regressors is advocated by Park (1990). On the other hand, little guidance on the optimal choice of k_1 and k_2 is provided although Remark c of the paper suggests that $k_1 + k_2 \geq 2$ is preferable.

17. Closely related tests have been proposed by Hansen (1992), Harris and Inder (1994), Kuo (1998), Leybourne and McCabe (1993), McCabe et al. (1997), Quintos and Phillips (1993), and Tanaka (1996). In Jansson and Haldrup (2000), we also study Hansen's L_c test (1992). The local asymptotic power properties of that test are very similar to those of Shin's test (1994), as are the local asymptotic power properties of the test due to Xiao (1999) (Jansson, 2001a).

18. In its original formulation, Shin's test (1994) uses Saikkonen's estimator (1991). The formulation based on the CCR estimator is due to Choi and Ahn (1995).

19. To see the equivalence, notice that rows 1 through $q-1$ in the expression

$$\int_0^t \widetilde{w}(s)\,ds - \int_0^1 \widetilde{w}(s)\widetilde{w}_1'(s)\,ds\left(\int_0^1 \widetilde{w}_1(s)\widetilde{w}_1'(s)\,ds\right)^{-1}\int_0^t \widetilde{w}_1(s)\,ds$$

in Theorem 11.11 of Tanaka (1996) are identically zero. As a consequence, the limiting distribution of Tanaka's $\tilde{S}_{T2}$ (1996) depends on the vector $c(J_2', J_3)/(\gamma(1)J_3)$ only through the scalar $c/\gamma(1)$. Indeed, the variate $cZ_2(t)$ appearing in the statement of Tanaka (1996) has the following simple representation:

$$\frac{c}{\gamma(1)}\left\{\int_0^t \widetilde{w}_2(s)\,ds - \int_0^1 \widetilde{w}_2(s)\widetilde{w}_1'(s)\,ds\left(\int_0^1 \widetilde{w}_1(s)\widetilde{w}_1'(s)\,ds\right)^{-1}\int_0^t \widetilde{w}_1(s)\,ds\right\}.$$

20. That is, $r_{1t} = (t, t^2)'$ and r_{2t} is a two-dimensional random walk in (8). Changing the values of k_1 and k_2 does not seem to affect the local power of the J_T test much.

21. Results for $2 \leq m_x \leq 4$ are qualitatively similar and can be found in Jansson and Haldrup (2000).

22. For recent contributions to this discussion, see Phillips (1998) and Phillips and Moon (1999).

REFERENCES

Beveridge, S. & C.R. Nelson (1981) A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the "business cycle." *Journal of Monetary Economics* 7, 151–174.

Canjels, E. & M.W. Watson (1997) Estimating deterministic trend in the presence of serially correlated errors. *Review of Economics and Statistics* 79, 184–200.

Chang, Y. & P.C.B. Phillips (1995) Time series regression with mixtures of integrated processes. *Econometric Theory* 11, 1033–1094.

Choi, I. (1994) Spurious regressions and residual-based tests for cointegration when regressors are cointegrated. *Journal of Econometrics* 60, 313–320.

Choi, I. & B.C. Ahn (1995) Testing for cointegration in a system of equations. *Econometric Theory* 11, 952–983.

Choi, I. & B.C. Ahn (1998) Testing the null of stationarity for multiple time series. *Journal of Econometrics* 88, 41–77.

Elliott, G. (1998) On the robustness of cointegration methods when regressors almost have unit roots. *Econometrica* 66, 149–158.

Granger, C.W.J. & P. Newbold (1974) Spurious regressions in econometrics. *Journal of Econometrics* 2, 111–120.

Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and Its Application*. New York: Academic Press.

Hansen, B.E. (1992) Tests for parameter instability in regressions with I(1) processes. *Journal of Business and Economic Statistics* 10, 321–335.

Harris, D. (1997) Principal components analysis of cointegrated time series. *Econometric Theory* 13, 529–557.

Harris, D. & B. Inder (1994) A test of the null hypothesis of cointegration. In C. Hargreaves (ed.), *Nonstationary Time Series Analysis and Cointegration*, pp. 133–152. Oxford: Oxford University Press.

Jansson, M. (2001a) Another Test of the Null Hypothesis of Cointegration. Manuscript, University of California, Berkeley.

Jansson, M. (2001b) Consistent covariance matrix estimation for linear processes. *Econometric Theory*, forthcoming.

Jansson, M. (2001c) Point Optimal Tests of the Null Hypothesis of Cointegration. Manuscript, University of California, Berkeley.

Jansson, M. & N. Haldrup (2000) Spurious Regression, Cointegration, and Near Cointegration: A Unifying Approach. Department of Economics Working paper 2000-14, University of California, San Diego.

Johansen, S. (1988) Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12, 231–254.

Johansen, S. (1991) Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551–1580.

Kuo, B.-S. (1998) Test for partial parameter instability in regressions with I(1) processes. *Journal of Econometrics* 86, 337–368.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, & Y. Shin (1992) Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics* 54, 159–178.

Leybourne, S.J. & B.P.M. McCabe (1993) A simple test for cointegration. *Oxford Bulletin of Economics and Statistics* 55, 97–103.

McCabe, B.P.M., S.J. Leybourne, & Y. Shin (1997) A parametric approach to testing the null of cointegration. *Journal of Time Series Analysis* 18, 395–413.

Park, J.Y. (1990) Testing for unit roots and cointegration by variable addition. In T.B. Fomby & G.F. Rhodes, Jr. (eds.), *Advances in Econometrics: Co-integration, Spurious Regression, and Unit Roots,* pp. 107–133. Greenwich, Connecticut: JAI Press.

Park, J.Y. (1992) Canonical cointegrating regressions. *Econometrica* 60, 119–143.

Park, J.Y., S. Ouliaris, & B. Choi (1988) Spurious Regressions and Tests for Cointegration. CAE Working paper 88-07, Cornell University.

Park, J.Y. & P.C.B. Phillips (1989) Statistical inference in regressions with integrated processes, part 2. *Econometric Theory* 5, 95–131.

Phillips, P.C.B. (1986) Understanding spurious regressions in econometrics. *Journal of Econometrics* 33, 311–340.

Phillips, P.C.B. (1988a) Regression theory for near-integrated time series. *Econometrica* 56, 1021–1043.

Phillips, P.C.B. (1988b) Weak convergence of sample covariance matrices to stochastic integrals via martingale approximations. *Econometric Theory* 4, 528–533.

Phillips, P.C.B. (1991) Spectral regression for cointegrated time series. In W.A. Barnett, J. Powell, & G.E. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Economics and Statistics,* pp. 413–435. Cambridge: Cambridge University Press.

Phillips, P.C.B. (1995) Fully modified least squares and vector autoregression. *Econometrica* 63, 1023–1078.

Phillips, P.C.B. (1998) New tools for understanding spurious regressions. *Econometrica* 66, 1299–1325.

Phillips, P.C.B. & S.N. Durlauf (1986) Multiple time series regression with integrated processes. *Review of Economic Studies* 53, 473–495.

Phillips, P.C.B. & 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. & H.R. Moon (1999) Linear regression limit theory for nonstationary panel data. *Econometrica* 67, 1057–1111.

Phillips, P.C.B. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.

Quintos, C.E. & P.C.B. Phillips (1993) Parameter constancy in cointegrating regressions. *Empirical Economics* 18, 675–706.

Saikkonen, P. (1991) Asymptotically efficient estimation of cointegration regressions. *Econometric Theory* 7, 1–21.

Saikkonen, P. (1992) Estimation and testing of cointegrated systems by an autoregressive approximation. *Econometric Theory* 8, 1–27.

Shin, Y. (1994) A residual-based test of the null of cointegration against the alternative of no cointegration. *Econometric Theory* 10, 91–115.

Snell, A. (1998) Testing for r versus $r - 1$ cointegrating vectors. *Journal of Econometrics* 88, 151–191.

Stock, J.H. & M.W. Watson (1993) A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61, 783–820.

Tanaka, K. (1993) An alternative approach to the asymptotic theory of spurious regression, cointegration, and near cointegration. *Econometric Theory* 9, 36–61.

Tanaka, K. (1996) *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*. New York: Wiley.

Xiao, Z. (1999) A residual based test for the null hypothesis of cointegration. *Economics Letters* 64, 133–141.

Yule, G.U. (1926) Why do we sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time-series. *Journal of the Royal Statistical Society* 89, 1–63.

APPENDIX

This Appendix outlines the proofs of the main results of the paper. To facilitate the proofs, we start with two preliminary lemmas.

LEMMA 5. *Suppose* $\{z_{Tt}\}$ *is generated by (1) and suppose A1–A4 hold. Then* $\hat{\Omega}_{ww} \to_p \Omega_{ww}, \hat{\Gamma}_{ww} \to_p \Gamma_{ww}$, *and* $\hat{\Sigma}_{ww} \to_p \Sigma_{ww}$.

For $1 \le t \le T$, *let* $q_t' = (d_t', x_t')$, $v_{Tt} = y_{Tt} - \beta' x_t$, $q_{Tt}^{\dagger\prime} = (d_t', x_{Tt}^{\dagger\prime})$, *and* $v_{Tt}^{\dagger} = y_{Tt}^{\dagger} - \beta' x_{Tt}^{\dagger}$. *Moreover, let* $\Upsilon_T = \mathrm{diag}(T^{m_d+1/2}, \ldots, t^{m_d+k_1-1/2}, T \cdot \iota_{k_2}')$, *where* ι_{k_2} *is a* k_2*-vector of ones.*

LEMMA 6. *Suppose* $\{z_{Tt}\}$ *is generated by (1) and suppose A1–A3 hold. Then*

(a) $T^{1/2} \Psi_T^{-1} q_{\lfloor T \cdot \rfloor} \to_d Q_x(\cdot)$,
(b) $\Psi_T^{-1} \sum_{t=1}^{T} q_t v_{Tt} \to_d \omega_{uu.x}^{1/2} \int Q_x \, dU_\lambda + \int Q_x \, dX' \Omega_{xx}^{-1} \omega_{xu} + (0, \gamma_{ux})'$,

where $\lfloor\cdot\rfloor$ *denotes the integer part of the argument. Moreover, if A4 holds, then*

(c) $T^{1/2}\Psi_T^{-1}q^{\dagger}_{T,\lfloor T\cdot\rfloor} \to_d Q_x(\cdot)$,
(d) $T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor} v^{\dagger}_{Tt} \to_d \omega_{uu.x}^{1/2}U_\lambda(\cdot)$,
(e) $\Psi_T^{-1}\sum_{t=1}^{T} q^{\dagger}_{Tt}v^{\dagger}_{Tt} \to_d \omega_{uu.x}^{1/2}\int Q_x\,dU_\lambda$,
(f) $T^{-1}\sum_{t=2}^{T}(\sum_{s=1}^{t-1} v^{\dagger}_{Ts})v^{\dagger}_{Tt} \to_d \omega_{uu.x}\int U_\lambda\,dU_\lambda + \gamma^{+}_{uu.x}$,
(g) $T^{1/2}\Upsilon_T^{-1}r_{\lfloor T\cdot\rfloor} \to_d R(\cdot)$,
(h) $\Upsilon_T^{-1}\sum_{t=1}^{T} r_t v^{\dagger}_{Tt} \to_d \omega_{uu.x}^{1/2}\int R\,dU_\lambda$,

where $\gamma^{+}_{uu.x} = (1,-\omega_{xu}'\Omega_{xx}^{-1})(\Gamma_{ww}-\Sigma_{ww})(1,-\omega_{xu}'\Omega_{xx}^{-1})'$, *whereas* Ψ_T, Q_x, U_λ, *X, and R are defined as in the text.*

Proof of Lemma 5. For $1 \le t \le T$, let $\hat{u}^{*}_{Tt} = C^u(L)e_t - \hat{\alpha}_T' d_t + (\hat{\beta}_T - \beta)'x_t$ and $\hat{u}^{**}_{Tt} = \delta_T'\xi_t$. Because $\hat{u}_{Tt} = \hat{u}^{*}_{Tt} + \hat{u}^{**}_{Tt}$, we can write $\hat{w}_{Tt}$ as the sum of $\hat{w}^{*}_{Tt} = (\hat{u}^{*}_{Tt},\Delta x_t')'$ and $\hat{w}^{**}_{Tt} = (\hat{u}^{**}_{Tt},0')'$. Using notation typified by

$$\hat{\Gamma}^{*,**}_{ww} = T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{t} k\left(\frac{|t-s|}{\hat{b}_T}\right)\hat{w}^{*}_{Tt}\hat{w}^{**\prime}_{Ts},$$

the corresponding decomposition of $\hat{\Gamma}_{ww}$ is $\hat{\Gamma}_{ww} = \hat{\Gamma}^{*,*}_{ww} + \hat{\Gamma}^{**,**}_{ww} + \hat{\Gamma}^{*,**}_{ww} + \hat{\Gamma}^{**,*}_{ww}$. Now, $\hat{\Gamma}^{*,*}_{ww} \to_p \Gamma_{ww}$ by Corollary 4 of Jansson (2001b), which is applicable because A5(i) of that paper holds by Lemma 1 and Lemma 6(a) of the present paper. To show $\hat{\Gamma}_{ww} \to_p \Gamma_{ww}$ it therefore suffices to show that $\hat{\Gamma}^{**,**}_{ww}$, $\hat{\Gamma}^{*,**}_{ww}$, and $\hat{\Gamma}^{**,*}_{ww}$ are $o_p(1)$.

The idea of the proof is the following. Each of the matrices $\hat{\Gamma}^{**,**}_{ww}$, $\hat{\Gamma}^{*,**}_{ww}$, and $\hat{\Gamma}^{**,*}_{ww}$ can be written as $T^{-1}\sum_{i=0}^{T-1} k(\hat{b}_T^{-1}i)M_{Ti}$, where $\{M_{Ti}: 0 \le i \le T-1; T \ge 1\}$ is a triangular array of random matrices. The event $\{\|T^{-1}\sum_{i=0}^{T-1} k(\hat{b}_T^{-1}i)M_{Ti}\| > \varepsilon\}$ is a subset of

$$\{\hat{a}_T \notin [\underline{a},\bar{a}]\} \cup \left\{\sup_{\underline{a}\le a\le\bar{a}}\left\|T^{-1}\sum_{i=0}^{T-1} k\left(\frac{i}{a\cdot b_T}\right)M_{Ti}\right\| > \varepsilon\right\}.$$

Under A4, $\inf_{T\ge T_0}\Pr(\hat{a}_T \notin [\underline{a},\bar{a}])$ can be made arbitrarily close to zero for sufficiently large T_0 and appropriately selected $0 < \underline{a} \le \bar{a} < \infty$. We therefore have $T^{-1}\sum_{i=0}^{T-1} k(\hat{b}_T^{-1}i)M_{Ti} = o_p(1)$ if

$$\sup_{\underline{a}\le a\le\bar{a}}\left\|T^{-1}\sum_{i=0}^{T-1} k\left(\frac{i}{a\cdot b_T}\right)M_{Ti}\right\| = o_{p^*}(1) \tag{A.1}$$

for every $0 < \underline{a} \le \bar{a} < \infty$, where $o_{p^*}(1)$ denotes convergence to zero in outer probability. (To avoid measurability complications, we consider convergence in outer probability rather than convergence in probability.) The proof proceeds by applying additive decompositions to $\hat{\Gamma}^{**,**}_{ww}$, $\hat{\Gamma}^{*,**}_{ww}$, and $\hat{\Gamma}^{**,*}_{ww}$ and establishing (A.1) for each element of these decompositions. In each instance, we make use of the fact that

$$\sup_{\underline{a}\le a\le\bar{a}} T^{-1}\sum_{i=0}^{T-1}\left|k\left(\frac{i}{a\cdot b_T}\right)\right| m_{Ti} = o_{p^*}(1)$$

if $\{m_{Ti}: 0 \le i \le T-1; T \ge 1\}$ is a triangular array of nonnegative random variables with $\max_{0\le i\le T-1} E(m_{Ti}) = O(T^{1/2})$.

Indeed, by Markov's inequality and the properties of $\Pr^*$, E^*, and $\bar{k}(\cdot)$, we have

$$\begin{aligned}
&\Pr^*\left(\sup_{\underline{a}\le a\le\bar{a}} T^{-1}\sum_{i=0}^{T-1}\left|k\left(\frac{i}{a\cdot b_T}\right)\right| m_{Ti}>\varepsilon\right)\\
&\quad\le \varepsilon^{-1}E^*\left(\sup_{\underline{a}\le a\le\bar{a}} T^{-1/2}\sum_{i=0}^{T-1}\left|k\left(\frac{i}{a\cdot b_T}\right)\right| T^{-1/2}m_{Ti}\right)\\
&\quad\le \varepsilon^{-1}E\left(T^{-1/2}\sum_{i=0}^{T-1}\bar{k}\left(\frac{i}{\bar{a}b_T}\right)T^{-1/2}m_{Ti}\right)\\
&\quad\le \varepsilon^{-1}T^{-1/2}\sum_{i=0}^{T-1}\bar{k}\left(\frac{i}{\bar{a}b_T}\right)\left(T^{-1/2}\max_{0\le j\le T-1}E(m_{Tj})\right)\\
&\quad\le \varepsilon^{-1}\left(T^{-1/2}\max_{0\le i\le T-1}E(m_{Ti})\right)\left(T^{-1/2}\sum_{i=0}^{T-1}\bar{k}\left(\frac{i}{\bar{a}b_T}\right)\right)=o(1)
\end{aligned}$$

for any $\varepsilon>0$, where $\Pr^*(\cdot)$ and $E^*(\cdot)$ denote outer probability and outer expectation, respectively, and the last equality uses the assumption on $\max_{0\le i\le T-1}E(m_{Ti})$ along with the fact that $T^{-1/2}\sum_{i=0}^{T-1}\bar{k}((\bar{a}b_T)^{-1}i)=o(1)$ under A5 (Jansson, 2001b).

Defining $i=t-s$ and applying the decomposition $\hat{w}_{Tt}^{**}=\hat{w}_{Ts}^{**}+(\hat{w}_{Tt}^{**}-\hat{w}_{Ts}^{**})$, $\hat{\Gamma}_{ww}^{**,**}$ can be written as $\hat{\Gamma}_{ww,1}^{**,**}(\hat{a}_T)+\hat{\Gamma}_{ww,2}^{**,**}(\hat{a}_T)$, where

$$\hat{\Gamma}_{ww,1}^{**,**}(a)=T^{-1}\sum_{i=0}^{T}k\left(\frac{i}{a\cdot b_T}\right)\sum_{s=1}^{T-i}\hat{w}_{Ts}^{**}\hat{w}_{Ts}^{**\prime},$$

$$\hat{\Gamma}_{ww,2}^{**,**}(a)=T^{-1}\sum_{i=0}^{T}k\left(\frac{i}{a\cdot b_T}\right)\sum_{s=1}^{T-i}(\hat{w}_{T,s+i}^{**}-\hat{w}_{Ts}^{**})\hat{w}_{Ts}^{**\prime}.$$

By subadditivity of $\|\cdot\|$, $\|\hat{\Gamma}_{ww,1}^{**,**}(a)\|\le T^{-1}\sum_{i=0}^{T}|k((a\cdot b_T)^{-1}i)|\cdot\sum_{s=1}^{T-i}\|\hat{w}_{Ts}^{**}\hat{w}_{Ts}^{**\prime}\|$, so $\hat{\Gamma}_{ww,1}^{**,**}(\hat{a}_T)=o_p(1)$ if $\max_{0\le i\le T-1}E(\sum_{s=1}^{T-i}\|\hat{w}_{Ts}^{**}\hat{w}_{Ts}^{**\prime}\|)=O(T^{1/2})$. Using A2 and the relation $\|\delta_T\|=O(T^{-1})$, we have

$$\begin{aligned}
\max_{0\le i\le T-1}E\left(\sum_{s=1}^{T-i}\|\hat{w}_{Ts}^{**}\hat{w}_{Ts}^{**\prime}\|\right)&=E\left(\sum_{s=1}^{T}\|\hat{w}_{Ts}^{**}\hat{w}_{Ts}^{**\prime}\|\right)=\|\delta_T\|^2\sum_{s=1}^{T}E(\xi_s'\xi_s)\\
&=\|\delta_T\|^2\sum_{s=1}^{T}m\cdot s=O(1).
\end{aligned}$$

Next, $\|\hat{\Gamma}_{ww,2}^{**,**}(a)\|\le T^{-1}\sum_{i=0}^{T}|k((a\cdot b_T)^{-1}i)|\cdot\|\sum_{s=1}^{T-i}(\hat{w}_{T,s+i}^{**}-\hat{w}_{Ts}^{**})\hat{w}_{Ts}^{**\prime}\|$. Therefore, $\hat{\Gamma}_{ww,2}^{**,**}(\hat{a}_T)=o_p(1)$ if $\max_{0\le i\le T-1}E\|\sum_{s=1}^{T-i}(\hat{w}_{T,s+i}^{**}-\hat{w}_{Ts}^{**})\hat{w}_{Ts}^{**\prime}\|=O(T^{1/2})$. By the law of iterated expectations and A2,

$$\begin{aligned}
E[(\xi_{s+i}-\xi_s)'(\xi_{t+i}-\xi_t)\xi_s'\xi_t]&=E(E[(\xi_{s+i}-\xi_s)'(\xi_{t+i}-\xi_t)|\mathcal{G}_{\max(s,t)}]\xi_s'\xi_t)\\
&=E[m\max(i-|t-s|,0)\xi_s'\xi_t]\\
&=m^2\max(i-|t-s|,0)\min(t,s)\\
&\le m^2 i1\{|t-s|<i\}s
\end{aligned}$$

for any $1 \leq s, t \leq T$ and $i \geq 0$, where $\mathcal{G}_t = \sigma(e_s : s \leq t)$ for any $t \geq 1$ and $1\{\cdot\}$ is the indicator function. Using this relation, the Cauchy–Schwarz inequality, and $\|\delta_T\| = O(T^{-1})$,

$$
\begin{aligned}
\left(E\left\|\sum_{s=1}^{T-i}(\hat{w}^{**}_{T,s+i}-\hat{w}^{**}_{Ts})\hat{w}^{**\prime}_{Ts}\right\|\right)^2 &\leq E\left(\left\|\sum_{s=1}^{T-i}(\hat{w}^{**}_{T,s+i}-\hat{w}^{**}_{Ts})\hat{w}^{**\prime}_{Ts}\right\|^2\right)\\
&= \sum_{s=1}^{T-i}\sum_{t=1}^{T-i} E\left[(\hat{w}^{**}_{T,s+i}-\hat{w}^{**}_{Ts})'(\hat{w}^{**}_{T,t+i}-\hat{w}^{**}_{Tt})\hat{w}^{**\prime}_{Ts}\hat{w}^{**}_{Tt}\right]\\
&\leq \sum_{s=1}^{T-i}\sum_{t=1}^{T-i} m^2\|\delta_T\|^4 i 1\{|t-s|<i\}s\\
&\leq \sum_{s=1}^{T-i} m^2\|\delta_T\|^4 2i^2 s\\
&\leq m^2\|\delta_T\|^4 i^2(T-i)(T-i+1)\\
&\leq m^2\|\delta_T\|^4 T^4 = O(1)
\end{aligned}
$$

for any $0 \leq i \leq T-1$. Therefore, $\hat{\Gamma}^{**,**}_{ww,2}(\hat{a}_T) = o_p(1)$ and $\hat{\Gamma}^{**,**}_{ww} = o_p(1)$, as was to be shown.

Next, consider $\hat{\Gamma}^{*,**}_{ww}$, which can be written as $\hat{\Gamma}^{*,**}_{ww,1}(\hat{a}_T) + \hat{\Gamma}^{*,**}_{ww,2}(\hat{a}_T)$, where

$$
\hat{\Gamma}^{*,**}_{ww,1}(a) = T^{-1}\sum_{i=0}^{T} k\left(\frac{i}{a\cdot b_T}\right)\sum_{s=1}^{T-i}(\hat{w}^{*}_{T,s+i}-w_{s+i})\hat{w}^{**\prime}_{Ts},
$$

$$
\hat{\Gamma}^{*,**}_{ww,2}(a) = T^{-1}\sum_{i=0}^{T} k\left(\frac{i}{a\cdot b_T}\right)\sum_{s=1}^{T-i} w_{s+i}\hat{w}^{**\prime}_{Ts}.
$$

Because

$$
\hat{w}^{*}_{T,s+i} - w_{s+i} = \left(\begin{pmatrix}\hat{\alpha}_T & 0\\ \hat{\beta}_T-\beta & 0\end{pmatrix}'\Psi_T\right)\left(T^{1/2}\Psi_T^{-1}\begin{pmatrix}d_{s+i}\\ x_{s+i}\end{pmatrix}\right)T^{-1/2}
$$

and $\|AB\| \leq \|A\|\cdot\|B\|$ for conformable A and B, an upper bound on $\|\hat{\Gamma}^{*,**}_{ww,1}(a)\|$ is given by

$$
\begin{aligned}
&\left\|\Psi_T\begin{pmatrix}\hat{\alpha}_T\\ \hat{\beta}_T-\beta\end{pmatrix}\right\|(T\|\delta_T\|)\\
&\quad\times T^{-1}\sum_{i=0}^{T}\left|k\left(\frac{i}{a\cdot b_T}\right)\right| T^{-1}\sum_{s=1}^{T-i}\left\|T^{1/2}\Psi_T^{-1}\begin{pmatrix}d_{s+i}\\ x_{s+i}\end{pmatrix}\right\|\|T^{-1/2}\xi_s\|.
\end{aligned}
$$

By Lemma 1 and $T\|\delta_T\| = O(1)$, $\|\Psi_T(\hat{\alpha}_T', (\hat{\beta}_T-\beta)')'\|(T\|\delta_T\|) = O_p(1)$, so the following condition is sufficient for $\hat{\Gamma}^{*,**}_{ww,1}(\hat{a}_T) = o_p(1)$:

$$
\max_{0\leq i\leq T-1} T^{-1}\sum_{s=1}^{T-i} E(\|T^{1/2}\Psi_T^{-1}(d_{s+i}', x_{s+i}')'\|\|T^{-1/2}\xi_s\|) = O(T^{1/2}).
$$

By A2, $\max_{1\le t\le T} E(\|T^{-1/2}\xi_t\|^2) = E(\|T^{-1/2}\xi_T\|^2) = m$. Moreover, it can be shown that $\max_{1\le t\le T} E(\|T^{1/2}\Psi_T^{-1}(d_t' \quad x_t')'\|^2) = O(1)$. Using these relations and the Cauchy–Schwarz inequality, the proof of $\hat{\Gamma}^{*,**}_{ww,1}(\hat{a}_T) = o_p(1)$ is completed as follows:

$$\begin{aligned}
&\max_{0\le i\le T-1} E\left(T^{-1}\sum_{s=1}^{T-i}\left\| T^{1/2}\Psi_T^{-1}\begin{pmatrix} d_{s+i}\\ x_{s+i}\end{pmatrix}\right\| \|T^{-1/2}\xi_s\|\right)\\
&\quad\le \max_{0\le i\le T-1} T^{-1}\sum_{s=1}^{T-i}\sqrt{E(\|T^{-1/2}\xi_s\|^2)}\sqrt{E(\|T^{1/2}\Psi_T^{-1}(d_{s+i}', x_{s+i}')'\|^2)}\\
&\quad\le \sqrt{\max_{1\le t\le T} E(\|T^{-1/2}\xi_t\|^2)}\sqrt{\max_{1\le t\le T} E(\|T^{1/2}\Psi_T^{-1}(d_t', x_t')'\|^2)} = O(1).
\end{aligned}$$

Let $\sum_{k=0}^{\infty} C_k^w L^k = (C^u(L), C^x(L)')'$. Because $\hat{w}^{**}_{Tt} = \delta_T'\sum_{l=0}^{\infty} 1\{l \le t-1\} e_{t-l}$ and $w_t = \sum_{k=0}^{\infty} C_k^w e_{t-k}$, $\hat{\Gamma}^{*,**}_{ww,2}(\hat{a}_T)$ can be written as $\sum_{j=1}^{3}\hat{\Gamma}^{*,**}_{ww,2,j}(\hat{a}_T)$, where

$$\hat{\Gamma}^{*,**}_{ww,2,j}(a) = T^{-1}\sum_{i=0}^{T} k((a\cdot b_T)^{-1}i)M^{*,**}_{T,i,j}, \qquad 1\le j\le 3,$$

with

$$M^{*,**}_{T,i,1} = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} C_k^w \sum_{s=1}^{T-i} e_{s+i-k}e_{s-l}' 1\{l\le s-1\}1\{k\ne l+i\}(\delta_T, 0),$$

$$M^{*,**}_{T,i,2} = \sum_{l=0}^{\infty} C_{l+i}^w \sum_{s=1}^{T-i} (e_{s-l}e_{s-l}' - I_m)1\{l\le s-1\}(\delta_T, 0),$$

$$M^{*,**}_{T,i,3} = \sum_{l=0}^{\infty} C_{l+i}^w \sum_{s=1}^{T-i} I_m 1\{l\le s-1\}(\delta_T, 0).$$

By the law of iterated expectations and A2,

$$\begin{aligned}
&E\left(\left\|T^{-1}\sum_{s=1}^{T-i} e_{s+i-k}e_{s-l}' 1\{l\le s-1\}1\{k\ne l+i\}\right\|^2\right)\\
&\quad = T^{-2}\sum_{s=l+1}^{T-i}\sum_{t=l+1}^{T-i} E(e_{s+i-k}'e_{t+i-k}e_{s-l}'e_{t-l}1\{k\ne l+i\})\\
&\quad = T^{-2}\sum_{s=l+1}^{T-i}\sum_{t=l+1}^{T-i} E[E(e_{s+i-k}'e_{t+i-k}e_{s-l}'e_{t-l}1\{k\ne l+i\}|\mathcal{G}_{\max(s,t)+\max(i-k,-l)-1})]\\
&\quad = 1\{k\ne l+i\}T^{-2}\sum_{s=l+1}^{T-i} E(e_{s+i-k}'e_{s+i-k})E(e_{s-l}'e_{s-l})\\
&\quad \le T^{-2}\sum_{s=l+1}^{T-i} m^2\\
&\quad \le m^2 T^{-1}1\{l\le T-1\}
\end{aligned}$$

for any $i, l \geq 0$. The conclusion $\hat{\Gamma}^{*,**}_{ww,2,1}(\hat{a}_T) = o_p(1)$ now follows because

$$\begin{aligned}
&\max_{0\leq i\leq T-1} E(\|M^{*,**}_{T,i,1}\|) \\
&\quad \leq T\|\delta_T\| \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \|C^w_k\| \max_{0\leq i\leq T-1} E\left(\left\| T^{-1}\sum_{s=1}^{T-i} e_{s+i-k}e'_{s-l}1\{l\leq s-1\}1\{k\neq l+i\}\right\|\right) \\
&\quad \leq T\|\delta_T\| \sum_{k=0}^{\infty} \|C^w_k\| \sum_{l=0}^{\infty} mT^{-1/2}1\{l\leq T-1\} \\
&\quad = (T\|\delta_T\|)mT^{1/2}\sum_{k=0}^{\infty}\|C^w_k\| = O(T^{1/2}),
\end{aligned}$$

where the second inequality uses the Cauchy–Schwarz inequality and the previous display, whereas the last equality uses $T\|\delta_T\| = O(1)$ and the fact that $\sum_{k=0}^{\infty}\|C^w_k\| < \infty$ under A1.

Next, $\|M^{*,**}_{T,i,2}\| \leq T\|\delta_T\|\sum_{l=0}^{\infty}\|C^w_{l+i}\|\|T^{-1}\sum_{s=1}^{T-i}(e_{s-l}e'_{s-l} - I_m)1\{l\leq s-1\}\|$ for any $i \geq 0$ and

$$E\left(\left\|T^{-1}\sum_{s=1}^{T-i}(e_{s-l}e'_{s-l} - I_m)1\{l\leq s-1\}\right\|\right) = E\left(\left\|T^{-1}\sum_{s=1}^{T-i-l}(e_s e'_s - I_m)\right\|\right)$$

for any $i, l \geq 0$. By A2, $\{\mathrm{vec}(e_t e'_t - I_m): t\geq 1\}$ is a uniformly integrable martingale difference sequence, so

$$\begin{aligned}
&\sup_{l\geq 0}\max_{0\leq i\leq T-1} E\left(\left\|T^{-1}\sum_{s=1}^{T-i}(e_{s-l}e'_{s-l} - I_m)1\{l\leq s-1\}\right\|\right) \\
&\quad = \sup_{l\geq 0} E\left(\left\|T^{-1}\sum_{s=1}^{T-l}(e_s e'_s - I_m)\right\|\right) = o(1)
\end{aligned}$$

by an argument analogous to the proof of Theorem 2.22 of Hall and Heyde (1980). In particular,

$$\begin{aligned}
&\max_{0\leq i\leq T-1} E(\|M^{*,**}_{T,i,2}\|) \\
&\quad \leq T\|\delta_T\|\sum_{l=0}^{\infty}\|C^w_{l+i}\| \max_{0\leq i\leq T-1} E\left(\left\|T^{-1}\sum_{s=1}^{T-i}(e_{s-l}e'_{s-l} - I_m)1\{l\leq s-1\}\right\|\right) \\
&\quad \leq T\|\delta_T\|\left[\sup_{l\geq 0}\max_{0\leq i\leq T-1} E\left(\left\|T^{-1}\sum_{s=1}^{T-i}(e_{s-l}e'_{s-l} - I_m)1\{l\leq s-1\}\right\|\right)\right]\sum_{l=0}^{\infty}\|C^w_l\| \\
&\quad = o(1),
\end{aligned}$$

so $\hat{\Gamma}^{*,**}_{ww,2,2}(\hat{a}_T) = o_p(1)$. Finally, $\hat{\Gamma}^{*,**}_{ww,2,3}(\hat{a}_T) = o(1)$ because

$$\begin{aligned}\max_{0\le i\le T-1} \|M^{*,**}_{T,i,3}\| &\le T\|\delta_T\| \max_{0\le i\le T-1}\left[\sum_{l=0}^{\infty}\|C^w_{l+i}\|\left(T^{-1}\sum_{s=1}^{T-i}1\{l\le s-1\}\right)\right]\\ &\le T\|\delta_T\| \max_{0\le i\le T-1}\left(\sum_{l=0}^{\infty}\|C^w_{l+i}\|\right)\\ &= T\|\delta_T\|\sum_{l=0}^{\infty}\|C^w_l\| = O(1).\end{aligned}$$

The proof of $\hat{\Gamma}^{**,*}_{ww} = o_p(1)$ is analogous to that of $\hat{\Gamma}^{*,**}_{ww} = o_p(1)$ and is omitted to conserve space. The proof of $\hat{\Sigma}_{ww} \to_p \Sigma_{ww}$ is a special case (with $\Sigma_{i=0}^{T-1}$ replaced by $\Sigma_{i=0}^{0}$ throughout) of the proof of $\hat{\Gamma}_{ww} \to_p \Gamma_{ww}$. Finally, $\hat{\Omega}_{ww} = \hat{\Gamma}_{ww} + \hat{\Gamma}'_{ww} - \hat{\Sigma}_{ww}$ is consistent for $\Omega_{ww} = \Gamma_{ww} + \Gamma'_{ww} - \Sigma_{ww}$, because $\hat{\Gamma}_{ww} \to_p \Gamma_{ww}$ and $\hat{\Sigma}_{ww} \to_p \Sigma_{ww}$. ∎

Proof of Lemma 6. Let $\{w_t\}$, $\{\xi_t\}$, and $\{r_{2t}\}$ be defined as in the text. By a multivariate analogue of Phillips and Solo (1992, Theorem 3.4), we have

$$\begin{pmatrix} T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor} w_t \\ T^{-1/2}\xi_{\lfloor T\cdot\rfloor} \\ T^{-1/2}r_{2,\lfloor T\cdot\rfloor}\end{pmatrix} \to_d \begin{pmatrix}\Omega^{1/2}_{ww} & 0\\ I_m & 0\\ 0 & I_{k_2}\end{pmatrix}\begin{pmatrix}U(\cdot)\\ V(\cdot)\\ R_2(\cdot)\end{pmatrix}, \tag{A.2}$$

where U, V, and R_2 are independent Wiener processes of dimension 1, m_x, and k_2, respectively, and

$$\Omega^{1/2}_{ww} = \begin{pmatrix}\omega^{1/2}_{uu.x} & \omega'_{xu}\Omega^{-1/2\prime}_{xx}\\ 0 & \Omega^{1/2}_{xx}\end{pmatrix}.$$

Moreover,

$$T^{-1}\sum_{t=1}^{T}\left(\sum_{s=1}^{t} w_s\right)w'_t \to_d \Omega^{1/2}_{ww}\int_0^1\begin{pmatrix}U\\V\end{pmatrix}d\begin{pmatrix}U\\V\end{pmatrix}'\Omega^{1/2\prime}_{ww} + \Gamma'_{ww}, \tag{A.3}$$

$$T^{-1}\sum_{t=1}^{T} r_{2t}w'_t \to_d \int_0^1 R_2\, d\begin{pmatrix}U\\V\end{pmatrix}'\Omega^{1/2\prime}_{ww}, \tag{A.4}$$

by Phillips (1988b), whereas

$$T^{-1}\sum_{t=1}^{T} w_t w'_t \to_p \Sigma_{ww} \tag{A.5}$$

in view of the law of large numbers.

Part (a) is standard. For $1 \le t \le T$, $v_{Tt} = u_t + \delta'_T\xi_t$, where u_t, δ_T, and ξ_t are defined as in Section 2. By (10), (11), and part (a),

$$\Psi_T^{-1}\sum_{t=1}^{T} q_t u_t \to_d \omega^{1/2}_{uu.x}\int Q_x\, dU + \int Q_x\, dX'\Omega^{-1}_{xx}\omega_{xu} + (0,\gamma_{ux})'.$$

Moreover,

$$\Psi_T^{-1}\sum_{t=1}^{T} q_t \xi_t' \delta_T \to_d \lambda\omega_{uu.x}^{1/2}\int Q_x U,$$

by (A.2), part (a), the relation $\lim_{T\to\infty} T\delta_T' = (\lambda\omega_{uu.x}^{1/2}, 0')$, and the continuous mapping theorem (CMT). Because $\int Q_x\, dU_\lambda = \int Q_x\, dU + \lambda\int Q_x U$, part (b) is obtained by combining the preceding displays.

Part (c) is standard. Define $q_t^{\dagger\dagger} = (d_t', x_t^{\dagger\dagger\prime})'$ and $v_{Tt}^{\dagger\dagger} = u_t^{\dagger\dagger} + \delta_T'\xi_t$, where $x_t^{\dagger\dagger} = x_t - \Gamma_{\cdot x}'\Sigma_{ww}^{-1}w_t$ and $u_t^{\dagger\dagger} = u_t - \omega_{xu}'\Omega_{xx}^{-1}\Delta x_t = (1, -\omega_{xu}'\Omega_{xx}^{-1})w_t$. Let $(c^{\dagger\dagger}) - (h^{\dagger\dagger})$ denote the counterparts of parts (c)–(h) in which $q_{Tt}^{\dagger}$ and $v_{Tt}^{\dagger}$ are replaced with $q_t^{\dagger\dagger}$ and $v_{Tt}^{\dagger\dagger}$, respectively. Part (c††) follows from part (a). By (10), $\lim_{T\to\infty} T\delta_T' = (\lambda\omega_{uu.x}^{1/2}, 0')$, and CMT,

$$T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor} u_t^{\dagger\dagger} = T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}(1, -\omega_{xu}'\Omega_{xx}^{-1})w_t \to_d \omega_{uu.x}^{1/2}U(\cdot),$$

$$T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}\delta_T'\xi_t = (T\delta_T)'T^{-3/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}\xi_t \to_d \omega_{uu.x}^{1/2}\lambda\int_0^{\cdot} U(\tau)\,d\tau.$$

Combining these results, part (d††) is obtained. Next,

$$(\mathrm{diag}(T^{1/2},\ldots,T^{m_d-1/2}))^{-1}\sum_{t=1}^{T} d_t v_t^{\dagger\dagger} \to_d \omega_{uu.x}^{1/2}\int D\,dU_\lambda$$

by (a), (d††), and CMT, whereas

$$T^{-1}\sum_{t=1}^{T} x_t^{\dagger\dagger}u_t^{\dagger\dagger} = \left(T^{-1}\sum_{t=1}^{T} x_t w_t' - \Gamma_{\cdot x}'\Sigma_{ww}^{-1}T^{-1}\sum_{t=1}^{T} w_t w_t'\right) \times \begin{pmatrix} 1 \\ -\Omega_{xx}^{-1}\omega_{xu}\end{pmatrix} \to_d \omega_{uu.x}^{1/2}\int X\,dU$$

and

$$T^{-1}\sum_{t=1}^{T} x_t^{\dagger\dagger}\xi_t'\delta_T = \left(T^{-2}\sum_{t=1}^{T} x_t^{\dagger\dagger}\xi_t'\right)(T\delta_T) \to_d \omega_{uu.x}^{1/2}\lambda\int XU$$

in view of (A.3), (A.5), (c††), (A.2), and CMT. Part (e††) is established by using these results and the relation

$$\Psi_T^{-1}\sum_{t=1}^{T} q_t^{\dagger\dagger}v_{Tt}^{\dagger\dagger} = \begin{pmatrix} (\mathrm{diag}(T^{1/2},\ldots,T^{m_d-1/2}))^{-1}\sum_{t=1}^{T} d_t v_t^{\dagger\dagger} \\ T^{-1}\sum_{t=1}^{T} x_t^{\dagger\dagger}u_t^{\dagger\dagger} + T^{-1}\sum_{t=1}^{T} x_t^{\dagger\dagger}\xi_t'\delta_T \end{pmatrix}.$$

Because

$$T^{-1}\sum_{t=2}^{T}\left(\sum_{s=1}^{t-1}v_{Ts}^{\dagger\dagger}\right)v_{Tt}^{\dagger\dagger}$$
$$= T^{-1}\sum_{t=2}^{T}\left(\sum_{s=1}^{t-1}u_s^{\dagger\dagger}\right)u_t^{\dagger\dagger} + T^{-1}\delta_T'\sum_{t=2}^{T}\left(\sum_{s=1}^{t-1}\xi_s\right)u_t^{\dagger\dagger} + T^{-1}\sum_{t=2}^{T}\left(\sum_{s=1}^{t-1}u_s^{\dagger\dagger}\right)\xi_t'\delta_T,$$

part ($\mathrm{f}^{\dagger\dagger}$) can be obtained by combining the following results, each of which is obtained in standard fashion using (A.2), (A.3), (A.5), and CMT:

$$T^{-1}\sum_{t=2}^{T}\left(\sum_{s=1}^{t-1}u_s^{\dagger\dagger}\right)u_t^{\dagger\dagger} = T^{-1}\sum_{t=2}^{T}\left(\sum_{s=1}^{t}u_s^{\dagger\dagger}\right)u_t^{\dagger\dagger} - T^{-1}\sum_{t=2}^{T}u_t^{\dagger\dagger}u_t^{\dagger\dagger}$$
$$\to_d \omega_{uu.x}\int U\,dU + \gamma_{uu.x}^{+},$$

$$T^{-1}\delta_T'\sum_{t=2}^{T}\left(\sum_{s=1}^{t-1}\xi_s\right)u_t^{\dagger\dagger} \to_d \omega_{uu.x}\lambda\int \bar{U}\,dU,$$

$$T^{-1}\sum_{t=2}^{T}\left(\sum_{s=1}^{t-1}u_s^{\dagger\dagger}\right)\xi_t'\delta_T \to_d \omega_{uu.x}\lambda\int U_\lambda U,$$

where $\bar{U}(r) = \int_0^r U(\tau)\,d\tau$. Parts (g) and ($\mathrm{h}^{\dagger\dagger}$) are proved in the same way as (a) and ($\mathrm{e}^{\dagger\dagger}$), respectively.

Now,

$$x_t^{\dagger\dagger} - x_{Tt}^{\dagger} = (\hat{\Gamma}_{\cdot x}'\hat{\Sigma}_{ww}^{-1} - \Gamma_{\cdot x}'\Sigma_{ww}^{-1})w_t + \hat{\Gamma}_{\cdot x}'\hat{\Sigma}_{ww}^{-1}(\hat{w}_{Tt} - w_t)$$

and

$$v_{Tt}^{\dagger\dagger} - v_{Tt}^{\dagger} = (\hat{\beta}_T - \beta)\hat{\Gamma}_{\cdot x}'\hat{\Sigma}_{ww}^{-1}\hat{w}_{Tt} + (\hat{\omega}_{xu}'\hat{\Omega}_{xx}^{-1} - \omega_{xu}'\Omega_{xx}^{-1})\Delta x_t.$$

By Lemma 5, $\hat{\Gamma}_{\cdot x}'\hat{\Sigma}_{ww}^{-1} \to_p \Gamma_{\cdot x}'\Sigma_{ww}^{-1}$ and $\hat{\omega}_{xu}'\hat{\Omega}_{xx}^{-1} \to_p \omega_{xu}'\Omega_{xx}^{-1}$. Furthermore, $\hat{\beta}_T - \beta = O_p(T^{-1})$ and $\max_{1\le t\le T}\|\hat{w}_{Tt} - w_t\| = O_p(T^{-1/2})$ by Lemma 1, the proof of which only uses (a)–(b) of the present lemma. Using these facts, the proof of (d) is completed as follows:

$$T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}(v_{Tt}^{\dagger\dagger} - v_{Tt}^{\dagger})$$
$$= (\hat{\beta}_T - \beta)\hat{\Gamma}_{\cdot x}'\hat{\Sigma}_{ww}^{-1}T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}\hat{w}_{Tt} + (\hat{\omega}_{xu}'\hat{\Omega}_{xx}^{-1} - \omega_{xu}'\Omega_{xx}^{-1})T^{-1/2}x_{\lfloor T\cdot\rfloor} \to_d 0.$$

Analogously, the proof of (e)–(h) can be completed by using elementary manipulations to show that $\Psi_T^{-1}\sum_{t=1}^{T}(q_{Tt}^{\dagger}v_{Tt}^{\dagger} - q_t^{\dagger\dagger}v_{Tt}^{\dagger\dagger}) \to_d 0$, $T^{-1}\sum_{t=2}^{T}(\sum_{s=1}^{t-1}v_{Ts}^{\dagger})v_{Tt}^{\dagger} - T^{-1}\sum_{t=2}^{T}(\sum_{s=1}^{t-1}v_{Ts}^{\dagger\dagger})v_{Tt}^{\dagger\dagger} \to_d 0$, etc. ∎

Proof of Lemma 1. In view of the relation

$$\Psi_T\begin{pmatrix}\hat{\alpha}_T\\ \hat{\beta}_T-\beta\end{pmatrix}=\left(\Psi_T^{-1}\sum_{t=1}^{T}q_tq_t'\Psi_T^{-1}\right)^{-1}\left(\Psi_T^{-1}\sum_{t=1}^{T}q_tv_{Tt}\right),$$

the stated result follows from Lemma 6(a)–(b) and CMT. ■

Proof of Lemma 2. We have

$$\Psi_T\begin{pmatrix}\hat{\alpha}_T^{\dagger}\\ \hat{\beta}_T^{\dagger}-\beta\end{pmatrix}=\left(\Psi_T^{-1}\sum_{t=1}^{T}q_{Tt}^{\dagger}q_{Tt}^{\dagger\prime}\Psi_T^{-1}\right)^{-1}\left(\Psi_T^{-1}\sum_{t=1}^{T}q_{Tt}^{\dagger}v_{Tt}^{\dagger}\right)$$

$$\to_d\left(\int Q_xQ_x'\right)^{-1}\left(\omega_{uu.x}^{1/2}\int Q_x\,dU_\lambda\right),$$

where the limiting distribution is obtained using Lemma 6(c) and (e) and CMT. It follows from integration by parts that $\int Q_x\,dU_\lambda\overset{\mathcal{L}}{=}\int Q_{x,\lambda}\,dU$. The mixture representation is obtained by noting that $\int Q_{x,\lambda}\,dU|_{\mathcal{F}_V}\overset{\mathcal{L}}{=}\mathcal{N}(0,\int Q_{x,\lambda}Q_{x,\lambda}')$ by the properties of the Itô integral. ■

Proof of Theorem 3. The statistic G_T can be written as

$$G_T=\left\|\left[\Phi_\beta\left(T^{-2}\sum_{t=1}^{T}x_{Tt,d}^{\dagger}x_{Tt,d}^{\dagger\prime}\right)^{-1}\Phi_\beta'\right]^{-1/2}\hat{\omega}_{uu.x}^{-1/2}\Phi_\beta T(\hat{\beta}_T^{\dagger}-\beta)\right\|^2.$$

By Lemma 5, Lemma 6(c), Lemma 2, and CMT,

$$\Phi_\beta\left(T^{-2}\sum_{t=1}^{T}x_{Tt,d}^{\dagger}x_{Tt,d}^{\dagger\prime}\right)^{-1}\Phi_\beta'\to_d\Phi_\beta\left(\int X_DX_D'\right)^{-1}\Phi_\beta'=\int X_D^{\Phi_\beta}X_D^{\Phi_\beta\prime}$$

and

$$\hat{\omega}_{uu.x}^{-1/2}\Phi_\beta T(\hat{\beta}_T^{\dagger}-\beta)\to_d\Phi_\beta\left(\int X_DX_D'\right)^{-1}\int X_D\,dU_\lambda=\int X_D^{\Phi_\beta}\,dU_\lambda,$$

where $X_D=\Omega_{xx}^{1/2}V_D$ and

$$X_D^{\Phi_\beta}(r)=\Phi_\beta\left(\int X_DX_D'\right)^{-1}X_D(r)=(\Phi_\beta\Omega_{xx}^{-1/2\prime})\left(\int V_DV_D'\right)^{-1}V_D(r).$$

As a consequence, $G_T\to_d\|\int\widetilde{X_D^{\Phi_\beta}}\,dU_\lambda\|^2$, where $\widetilde{X_D^{\Phi_\beta}}(r)=(\int X_D^{\Phi_\beta}X_D^{\Phi_\beta\prime})^{-1/2}X_D^{\Phi_\beta}(r)$.

The distribution of $\widetilde{X_D^{\Phi_\beta}}$ depends on Φ_β and $\Omega_{xx}^{-1/2\prime}$ through $\Phi_\beta\Omega_{xx}^{-1/2\prime}$. By the partitioned inverse formula, $\widetilde{X_D^{\Phi_\beta}}=\widetilde{V_D^p}$ when $\Omega_{xx}=I_{m_x}$ and $\Phi_\beta=(I_p,0)$ and the proof of part (a) can therefore be completed by showing that no generality is lost by assuming $\Phi_\beta\Omega_{xx}^{-1/2\prime}=(I_p,0)$. $\mathcal{L}(\widetilde{X_D^{\Phi_\beta}})$ is invariant under transformations of the form $\Phi_\beta\Omega_{xx}^{-1/2\prime}\to K\Phi_\beta\Omega_{xx}^{-1/2\prime}\mathcal{O}$, where K is a nonsingular $p\times p$ matrix and $\mathcal{O}$ is an orthogonal $m_x\times m_x$ matrix. Take $\mathcal{O}$ such that $\Phi_\beta\Omega_{xx}^{-1/2\prime}\mathcal{O}=(L,0)$, where L is lower triangular. Setting $K=L^{-1}$, we arrive at the desired conclusion.

To establish part (b), it suffices to show that

$$\int \widetilde{V_D^p}\, dU_\lambda \bigg|_{\mathcal{F}_V} \overset{\mathcal{L}}{=} \mathcal{N}\left(0, I_p + \lambda^2 \int_0^1 \underline{\widetilde{V_D^p}}(r) \underline{\widetilde{V_D^p}}(r)'\, dr\right).$$

Using integration by parts, $\int \widetilde{V_D^p}\, dU_\lambda \overset{\mathcal{L}}{=} \int \widetilde{V_{D,\lambda}^p}\, dU$, where $\widetilde{V_{D,\lambda}^p}(r) = \lambda \underline{\widetilde{V_D^p}}(r) + \widetilde{V_D^p}(r)$ and $\underline{\widetilde{V_D^p}}(r) = \int_r^1 \widetilde{V_D^p}(s)\, ds$. By the properties of the Itô integral,

$$\int \widetilde{V_{D,\lambda}^p}\, dU \bigg|_{\mathcal{F}_V} \overset{\mathcal{L}}{=} \mathcal{N}\left(0, \int_0^1 \widetilde{V_{D,\lambda}^p}(r) \widetilde{V_{D,\lambda}^p}(r)'\, dr\right).$$

Now, $\int_0^1 \widetilde{V_D^p}(r) \widetilde{V_D^p}(r)'\, dr = I_p$, and the result follows because

$$\int_0^1 \underline{\widetilde{V_D^p}}(r) \widetilde{V_D^p}(r)'\, dr + \int_0^1 \widetilde{V_D^p}(r) \underline{\widetilde{V_D^p}}(r)'\, dr = \left(\int_0^1 \widetilde{V_D^p}(r)\, dr\right)\left(\int_0^1 \widetilde{V_D^p}(r)\, dr\right)' = 0,$$

where the equality uses integration by parts and the relation $\int_0^1 V_D(r)\, dr = 0$, respectively. ■

Proof of Theorem 4. Let $r_{Tt,q}^{\dagger} = r_t - (\sum_{s=1}^T r_s q_{Ts}^{\dagger\prime})(\sum_{s=1}^T q_{Ts}^{\dagger} q_{Ts}^{\dagger\prime})^{-1} q_{Tt}^{\dagger}$. By Lemma 5, Lemma 6(c), (e), (g), and (h), and CMT,

$$\begin{aligned} J_T(k_1, k_2) &= \left\| \left(\Upsilon_T^{-1} \sum_{t=1}^T r_{Tt,q}^{\dagger} r_{Tt,q}^{\dagger\prime} \Upsilon_T^{-1}\right)^{-1/2} \hat{\omega}_{uu.x}^{-1/2} \Upsilon_T^{-1} \sum_{t=1}^T r_{Tt,q}^{\dagger} v_{Tt}^{\dagger} \right\|^2 \\ &\to_d \left\| \left(\int R_Q R_Q'\right)^{-1/2} \left(\int R_Q\, dU_\lambda\right) \right\|^2 = \left\| \left(\int \tilde{R}_Q\, dU_\lambda\right) \right\|^2. \end{aligned}$$

Next, by Lemma 5, Lemma 6(c)–(e), and CMT, $\hat{\omega}_{uu.x}^{-1/2} T^{-1/2} \hat{S}_{T,\lfloor T\cdot \rfloor} \to_d \hat{U}_\lambda(\cdot)$, so $CI_T = \hat{\omega}_{uu.x}^{-1} T^{-2} \sum_{t=1}^T \hat{S}_{Tt}^2 \to_d \int \hat{U}_\lambda^2$, as claimed.

Finally, using Lemma 5, Lemma 6(c)–(f), and CMT, it is not hard to show that $\hat{\omega}_{uu.x}^{-1/2} T^{-1/2} \check{S}_{T,\lfloor T\cdot \rfloor} \to_d U_{\lambda,\bar{Q}}(\cdot)$ and $\hat{\omega}_{uu.x}^{-1}(T^{-1} \sum_{t=2}^T \check{S}_{t-1} \Delta \check{S}_t - \hat{\gamma}_{uu.x}^{+}) \to_d \int U_{\lambda,\bar{Q}} dU_{\lambda,\bar{Q}}$, where the notation $\int U_{\lambda,\bar{Q}} dU_{\lambda,\bar{Q}}$ is shorthand for $\int U_{\lambda,\bar{Q}} dU_\lambda - (\int U_\lambda \bar{Q}')(\int \bar{Q}\bar{Q}')^{-1} \int U_{\lambda,\bar{Q}} Q$. As a consequence, $SBDH_T^I = \hat{\omega}_{uu.x}^{-1} T^{-2} \sum_{t=1}^T \check{S}_{Tt}^2 \to_d \int U_{\lambda,\bar{Q}}^2$, $LM_T^I \to_d (\int U_{\lambda,\bar{Q}} dU_{\lambda,\bar{Q}})^2$, and $LM_T^{II} = (SBDH_T^I)^{-1} \cdot LM_T^I + o_p(1) \to_d (\int U_{\lambda,\bar{Q}}^2)^{-1} (\int U_{\lambda,\bar{Q}} dU_{\lambda,\bar{Q}})^2$. ■