Cowles Foundation Paper 722

Econometric Theory, 5, 1989, 95–131. Printed in the United States of America.

STATISTICAL INFERENCE IN REGRESSIONS WITH INTEGRATED PROCESSES: PART 2

JOON Y. PARK AND PETER C.B. PHILLIPS
Cowles Foundation for Research in Economics, Yale University

This paper continues the theoretical investigation of Park and Phillips [7]. We develop an asymptotic theory of regression for multivariate linear models that accommodates integrated processes of different orders, nonzero means, drifts, time trends, and cointegrated regressors. The framework of analysis is general but has a common architecture that helps to simplify and codify what would otherwise be a myriad of isolated results. A good deal of earlier research by the authors and by others comes within the new framework. Special models of some importance are considered in detail, such as VAR systems with multiple lags and cointegrated variates.

1. INTRODUCTION

In Part 1 of this work (Park and Phillips [7]), we embarked on a simple and unifying analysis of multivariate regressions with integrated processes. We showed how all of the major asymptotic distributions in such models can be represented in a common form which provides a simple groundwork for subsequent analysis. This common form helps to simplify the presentation of rather complicated results and it illuminates earlier research findings by clarifying their specialized structure within a much broader context. Our attention in Part 1 was devoted to processes that are integrated of order one. As in Part 1, we call a time series $\{X_t\}$ an integrated process of order k (an $I(k)$ process) if the time series of k-th differences $\{\Delta^k X_t\}$ is stationary (an $I(0)$ process).

The object of this sequel is to show how the regression theory in Part 1 lays the groundwork of an asymptotic theory for regressions with processes that are integrated of different orders. The simplest extension allows for the presence of stationary regressors as well as $I(1)$ processes. The stationary regressors may be jointly dependent variables or exogenous variables; the $I(1)$ processes may be lagged dependent variables or other time series with unit roots; there may be nonzero means, drifts, and possibly time trends in the formulation; and, since the setting is multivariate, we may wish to allow for cointegration among the $I(1)$ regressors. At the next level of generality

Our thanks go to Glena Ames for her skill and effort in typing the manuscript of this paper and to the NSF for research support under Grant No. SES 8519595. This paper was completed while Joon Y Park was an Alfred P Sloan Doctoral Dissertation Fellow

 0266-4666/89 $5.00 + .00

we wish to include $I(0)$, $I(1)$, and $I(2)$ processes as regressors. This allows for vector autoregressions (VAR's) with unit roots and additional $I(1)$ regressors. Once this model is extended to allow for nonzero means, drifts, possible time trends, and cointegrated regressors, the framework is sufficiently broad to provide a fairly complete picture of the asymptotic theory of regression for integrated processes of different orders. Note that the inclusion of $I(k)$ regressors of different orders itself requires that the theory must accommodate cointegration since, for example, $I(0)$ and $I(1)$ processes are trivially cointegrated. Moreover, once the theory is completely developed for $I(0)$, $I(1)$, and $I(2)$ regressors, generalizations to higher orders are straightforward. Note also that the presence of $I(0)$ regressors in the framework allows us to treat higher order VAR systems with many lags as a simple special case of the general theory. The paper thus includes as a special case recent results in Sims, Stock, and Watson [12], which deals with the asymptotic normality of coefficients in a VAR with unit roots. To sum up, our aim in this sequel is to bring together a set of results into one general framework which will give a comprehensive asymptotic theory of regression for models of this type.

The plan of the paper follows the same lines as Part 1. The models that are central to our study are discussed together with some preliminary theory in Section 2. Section 3 develops the asymptotic theory for least squares in the new context and relates the results to our earlier theory in Part 1. Hypothesis testing is the subject of Section 4, and we develop and study extensions of the G- and H-statistics of Part 1. Specializations of our theory are examined in Section 5. These include regressions with strictly exogenous regressors, general linear models with cointegrated regressors, VAR systems with exogenous regressors and general VAR's with many lags, unit roots, and cointegrated variates. Some concluding remarks are made in Section 6. Proofs are given in the Appendix.

2. THE MODELS AND PRELIMINARY THEORY

Let $\{y_t\}_1^\infty$ be an n-dimensional multiple time series generated either by

$$y_t = A_1 x_{1t} + A_2 x_{2t} + u_t \tag{1}$$

or by

$$y_1 = A_1 x_{1t} + A_2 x_{2t} + A_3 x_{3t} + u_t \tag{2}$$

where A_1, A_2, and A_3 are, respectively, $n \times m_1$, $n \times m_2$, and $n \times m_3$ coefficient matrices and where

$$x_{1t} = v_{1t}, \quad \Delta x_{2t} = v_{2t}, \quad \Delta^2 x_{3t} = v_{3t}. \tag{3}$$

In (3), Δ is the standard first difference operator. Initializations of the processes $\{x_{2t}\}_0^\infty$ and $\{x_{3t}\}_{-1}^\infty$ at $t = 0$ and $t = -1,0$, respectively, do not affect our results, and any random initial values are permissible. As in Part 1, we shall also consider as direct extensions of (1) and (2) time series $\{y_t\}_1^\infty$ that are generated by

$$y_t = \mu + A_1 x_{1t} + A_2 x_{2t} + u_t \tag{1$'$}$$

$$y_t = \mu + \theta t + A_1 x_{1t} + A_2 x_{2t} + u_t \tag{1$''$}$$

and

$$y_t = \mu + A_1 x_{1t} + A_2 x_{2t} + A_3 x_{3t} + u_t \tag{2$'$}$$

$$y_t = \mu + \theta t + A_1 x_{1t} + A_2 x_{2t} + A_3 x_{3t} + u_t. \tag{2$''$}$$

The conditions for the innovation sequence $w_t' = (u_t', v_{1t}', v_{2t}', v_{3t}')$ are entirely analogous to those in Part 1. We require that the partial sum process $S_t = \Sigma_1^t w_j$ satisfy a multivariate invariance principle. Thus, if for $r \in [0,1]$,

$$X_T(r) = T^{-1/2} S_{[Tr]}$$

then as $T \uparrow \infty$

$$X_T(r) \Rightarrow B(r) \tag{4}$$

where, as usual, the symbol "$\Rightarrow$" signifies weak convergence of the associated probability measures. In (4),

$$\begin{array}{cccccc} & & n & m_1 & m_2 & m_3 \\ B(r) & = & (B_0(r), & B_1(r), & B_2(r), & B_3(r)) \end{array}$$

is k-vector Brownian motion ($k = n + m_1 + m_2 + m_3$) with covariance matrix

$$\Omega = \begin{array}{c} \left[\begin{array}{cccc} \Omega_0 & \Omega_{01} & \Omega_{02} & \Omega_{03} \\ \Omega_{10} & \Omega_1 & \Omega_{12} & \Omega_{13} \\ \Omega_{20} & \Omega_{21} & \Omega_2 & \Omega_{23} \\ \Omega_{30} & \Omega_{31} & \Omega_{32} & \Omega_3 \end{array}\right] \end{array} \begin{array}{l} n \\ m_1 \\ m_2 \\ m_3 \end{array} \tag{5}$$

$$= \lim_{T\to\infty} \frac{1}{T} E(S_T S_T')$$

$$= \Sigma + \Lambda + \Lambda'$$

where

$$\Sigma = \lim_{T\to\infty} \frac{1}{T} \Sigma_1^T E(w_t w_t') \tag{6}$$

and

$$\Lambda = \lim_{T\to\infty} \frac{1}{T}\sum_{t=2}^{T}\sum_{j=1}^{t-1} E(w_j w_t') \tag{7}$$

Σ_{ij} and Λ_{ij} for $i,j = 0,\ldots,3$ are defined to be submatrices of Σ and Λ corresponding to Ω_{ij} in (5). We assume the diagonal submatrices Ω_0, Ω_2, and Ω_3 of Ω, and Σ_0 and Σ_1 of Σ to be positive definite but we do not require Ω to be positive definite. Our theory also requires weak convergence to matrix stochastic integrals of the form $\int_0^1 BdB'$ which, as well as (4), holds for a wide class of sequences $\{w_t\}$ that are weakly dependent and possibly heterogeneously distributed under quite general conditions. For these conditions, see Part 1 and the references given therein. Finally, we require weak laws to apply to the second sample moments of $\{w_t\}$; more specifically we need $T^{-1}\Sigma_1^T w_t w_t' \to \Sigma$ in probability which holds under very general weak dependence and moment conditions (see, for example, McLeish [5]).

When $\{w_t\}$ is weakly stationary, $\{x_{3t}\}$ is an $I(2)$ process, $\{x_{2t}\}$ is $I(1)$, and $\{x_{1t}\}$ is $I(0)$. In this case, (6) and (7) reduce to

$$\Sigma = E(w_1 w_1')$$

and

$$\Lambda = \sum_{j=2}^{\infty} E(w_1 w_j').$$

Moreover, if the series defining Λ is absolutely convergent, then the process $\{w_t\}$ has a continuous spectral density matrix $f_{ww}(\lambda)$ and

$$\Omega = 2\pi f_{ww}(0). \tag*{(5)$'$}$$

Throughout this paper it will be convenient to refer to $\{x_{1t}\}$ as a stationary process and to $\{x_{2t}\}$ and $\{x_{3t}\}$ as integrated processes, although, strictly speaking, our theory allows for somewhat greater generality.

The model defined by (1) or (2) (together with (3)) may be regarded as a simultaneous equations system in which we have both stationary and integrated regressors. In this case, none of the common exogeneity conditions for the regressors is presumed. Many other models that have been previously studied in isolation come within the framework of (1)-(1)$''$ or (2)-(2)$''$. Indeed, our subsequent theory is applicable to a wide range of important linear models in statistics and econometrics. Some of these will be individually discussed in later sections in detail. We list a few of the most relevant specializations below:

Model (1) + (3) can be specialized to: (1a) first order VAR systems with unit roots and additional stationary regressors; (1b) higher order VAR systems with single unit roots; and (1c) multiple time series regressions with regressors that are cointegrated of order $CI(1,1)$ in the terminology of Engle and Granger [2].

Model (2) + (3) similarly specializes to: (2a) first order VAR systems with unit roots and additional $I(1)$ regressors; (2b) higher order VAR systems with double unit roots; and (2c) multiple time series regressions with regressors that are cointegrated of order $CI(2,1)$ or $CI(2,2)$.

Nankervis and Savin [6] study (1a) and (2a) by Monte Carlo methods applied to the subcase of a simple scalar stochastic difference equation with an exogenous variable. (1c) was considered by Sims [11,12]. Fuller, Hasza, and Goebel [2] examined (1a) and (1b) in a simple scalar case and a univariate version of (2b) was considered by Hasza and Fuller [4].

The notation introduced in (4) to (7) will be used throughout the paper. Particularly, $B_0(r)$, $B_1(r)$, $B_2(r)$, and $B_3(r)$ denote, unless otherwise stated, four vector Brownian motions which are, respectively, n, m_1, m_2, and m_3 dimensional with covariance matrices given by the corresponding diagonal submatrices of Ω in (5). We also define

$$\bar{B}_3(r) = \int_0^r B_3(s)ds$$

and

$$\Delta_{20} = \Sigma_{20} + \Lambda_{20}, \qquad \Delta_{21} = \Sigma_{21} + \Lambda_{21}.$$

Moreover, we shall frequently write these and other stochastic processes without the argument for notational brevity when there is no risk of misunderstanding.

The following lemma will be used extensively in the derivation of our subsequent results:

LEMMA 2.1.

(a) (i) $T^{-5/2}\Sigma_1^T x_{3t} \Rightarrow \int_0^1 \bar{B}_3$, (ii) $T^{-3/2}\Sigma_1^T x_{2t} \Rightarrow \int_0^1 B_2$,

(iii) $T^{-1/2}\Sigma_1^T x_{1t} \Rightarrow B_1(1)$, (iv) $T^{-1/2}\Sigma_1^T u_1 \Rightarrow B_0(1)$;

(b) (i) $T^{-7/2}\Sigma_1^T t x_{3t} \Rightarrow \int_0^1 r\bar{B}_3$, (ii) $T^{-5/2}\Sigma_1^T t x_{2t} \Rightarrow \int_0^1 rB_2$,

(iii) $T^{-3/2}\Sigma_1^T t x_{1t} \Rightarrow \int_0^1 rdB_1 = B_1(1) - \int_0^1 B_1$,

(iv) $T^{-3/2}\Sigma_1^T t u_t \Rightarrow \int_0^1 rdB_0 = B_0(1) - \int_0^1 B_0$;

(c) (i) $T^{-4}\Sigma_1^T x_{3t}x_{3t}' \Rightarrow \int_0^1 \bar{B}_3\bar{B}_3'$, (ii) $T^{-3}\Sigma_1^T x_{3t}x_{2t}' \Rightarrow \int_0^1 \bar{B}_3 B_2'$,

(iii) $T^{-2}\Sigma_1^T x_{2t}x_{2t}' \Rightarrow \int_0^1 B_2B_2'$,

(iv) $T^{-2}\Sigma_1^T x_{3t}x_{1t}' \Rightarrow \int_0^1 \bar{B}_3 dB_1' = \bar{B}_3(1)B_1(1)' - \int_0^1 B_3B_1'$,

(v) $T^{-2}\Sigma_1^T x_{3t}u_t' \Rightarrow \int_0^1 \bar{B}_3 dB_0' = \bar{B}_3(1)B_0(1)' - \int_0^1 B_3B_0'$,

(vi) $T^{-1}\Sigma_1^T x_{2t}x_{1t}' \Rightarrow \int_0^1 B_2 dB_1' + \Delta_{21}$,

(vii) $T^{-1}\Sigma_1^T x_{2t}u_t' \Rightarrow \int_0^1 B_2 dB_0' + \Delta_{20}$.

Joint weak convergence of all the above also applies.

3. LEAST SQUARES ESTIMATION

As in Part 1, we consider three multiple least squares regressions corresponding, respectively, to (1), (1)′, and (1)″:

$$y_t = \hat{A}_1x_{1t} + \hat{A}_2x_{2t} + \hat{u}_t \tag{8}$$

$$y_t = \bar{\mu} + \bar{A}_1x_{1t} + \bar{A}_2x_{2t} + \bar{u}_t \tag{9}$$

$$y_t = \tilde{\mu} + \tilde{\theta}t + \tilde{A}_1x_{1t} + \tilde{A}_2x_{2t} + \tilde{u}_t. \tag{10}$$

Similarly, for (2), (2)′ and (2)″ we define

$$y_t = \hat{A}_1x_{1t} + \hat{A}_2x_{2t} + \hat{A}_3x_{3t} + \hat{u}_t \tag{11}$$

$$y_t = \bar{\mu} + \bar{A}_1x_{1t} + \bar{A}_2x_{2t} + \bar{A}_3x_{3t} + \bar{u}_t \tag{12}$$

$$y_t = \tilde{\mu} + \tilde{\theta}t + \tilde{A}_1x_{1t} + \tilde{A}_2x_{2t} + \tilde{A}_3x_{3t} + \tilde{u}_t. \tag{13}$$

We let $A = (A_1,A_2)$ in (1) or $A = (A_1,A_2,A_3)$ in (2) with analogous definitions of $\hat{A}$, $\bar{A}$, and $\tilde{A}$ in (8) to (13). Similarly, we let $x_t' = (x_{1t}',x_{2t}')$ in (1) or $x_t' = (x_{1t}',x_{2t}',x_{3t}')$ in (2). Define $x_t^{1'} = (1,x_t')$, $x_t^{2'} = (1,t,x_t')$, and given a sample of size T, define

$$X' = (x_1,\ldots,x_T), \quad X^{1'} = (x_1^1,\ldots,x_T^1), \quad X^{2'} = (x_1^2,\ldots,x_T^2). \tag{14}$$

With this notation we have:

$$\hat{A} = Y'X(X'X)^{-1}, \quad \bar{A}^1 = Y'X^1(X^{1'}X^1)^{-1}, \quad \tilde{A}^2 = Y'X^2(X^{2'}X^2)^{-1}$$

where $\bar{A}^1 = (\bar{\mu},\bar{A})$ and $\tilde{A}^2 = (\tilde{\mu},\tilde{\theta},\tilde{A})$. The least squares estimator of the covariance matrix is given for each regression equation by

$$\hat{\Sigma}_0 = \frac{1}{T}Y'(I - P_X)Y, \quad \bar{\Sigma}_0 = \frac{1}{T}Y'(I - P_{X^1})Y, \quad \tilde{\Sigma}_0 = \frac{1}{T}Y'(I - P_{X^2})Y$$

where $P_C = C(C'C)^{-1}C'$ for any matrix C of full column rank (with probability one, if it is random).

As is well known, the inclusion of the constant term in (9) or (12) and of the time trend in (10) or (13) has the same effect on estimation of the coefficient matrix A as demeaning and detrending the series $\{x_t\}$ prior to regression (8) or (11). The estimates $\tilde{A}$ in (10) or (13) are thus easily seen to be invariant with respect to the introduction of a nonzero constant mean for $\{v_{2t}\}$, which turns $\{x_{2t}\}$ in (3) into a random walk with drift. We can also have a similar invariance to an unknown mean of $\{v_{3t}\}$ simply by taking a regression with a quadratic rather than a linear time trend (i.e. add the regressor t^2 to (1)″ and (2)″).

We can expect under quite general conditions that

$$T^{-1/2}\Sigma_1^T(w_t w_t' - \Sigma) \Rightarrow N(0, V^0). \tag{15}$$

For example, if we let $\xi_t = w_t \otimes w_t - E(w_t \otimes w_t)$ and assume $\{\xi_t\}_1^\infty$ is a weakly stationary process for which the invariance principle (4) holds, then it is not difficult to show (see, for example, Theorem 3.4 of Part 1) that (15) also holds, with covariance matrix given by

$$V^0 = P_D\left[\Phi_0 + \sum_{k=1}^{\infty} (\Phi_j + \Phi_j')\right]P_D \tag{16}$$

where

$$\Phi_j = E(w_t w_{t+j}' \otimes w_t w_{t+j}') - (\text{vec }\Sigma)(\text{vec }\Sigma)'$$

and D is the $k^2 \times k(k+1)/2$ duplication matrix. We now define submatrices

$$V_{ij}^0 = (S_i \otimes S_j)V^0(S_i' \otimes S_j') \quad (i,j = 0,1,2,3)$$

where the S_i $(i = 0,1,2,3)$ are selector matrices which select subvectors of w_t corresponding to the component vectors u_t, v_{1t}, v_{2t}, and v_{3t}, respectively. Thus, for example, V_{10}^0 is the limiting covariance matrix of $T^{-1/2}\Sigma_1^T x_{1t}u_t'$. When $\{x_{1t}\}$ and $\{u_t\}$ are contemporaneously uncorrelated and $\Sigma_{10} = 0$, as is often assumed in the standard regression theory, the submatrix V_{10}^0 of V^0 in (15) may simplify to

$$V_{10}^0 = \Sigma_1 \otimes \Sigma_0. \tag{17}$$

This holds, for example, if we further assume that $\{(v_{1t}, u_t)\}_1^\infty$ is a martingale difference sequence. The asymptotic normality of $T^{-1/2}\Sigma_1^T x_{1t}u_t'$ which is given in (15), of course, holds under more general conditions, especially when $\Sigma_{10} = 0$. For an introductory and unified exposition of this subject, see Chapter 5 of White [13].

The next theorem characterizes the asymptotic behavior of the least squares estimators for A_1, the coefficient matrix of the stationary regressors, in regressions (8) to (13).

THEOREM 3.1. *We have in regressions (8) to (10) or in (11) to (13)*

(a) $\hat{A}_1, \bar{A}_1, \tilde{A}_1 \xrightarrow{P} A_1^*$

where $A_1^* = A_1 + \Sigma_{10}'\Sigma_1^{-1}$. *Moreover, if (15) holds, then*

(b) $\sqrt{T}(\hat{A}_1 - A_1^*), \sqrt{T}(\bar{A}_1 - A_1^*), \sqrt{T}(\tilde{A}_1 - A_1^*) \Rightarrow N(0, V)$.

The covariance matrix V is given by

$$V = (J_1 \otimes J_2)V^0(J_1' \otimes J_2')$$

where

$$J_1 = \begin{array}{ccc} n & m_1 & m_2 + m_3 \\ (I, & -\Sigma_{10}'\Sigma_1^{-1}, & 0), \end{array} \qquad J_2 = \begin{array}{ccc} n & m_1 & m_2 + m_3 \\ (0, & \Sigma_1^{-1}, & 0). \end{array}$$

The above theorem shows, in short, that the least squares estimators of A_1 in (8) to (10) or in (11) to (13) are asymptotically equivalent to the regression coefficients from the regression of y_{2t} on x_{1t}, where $y_{2t} = y_t - A_2x_{2t}$ or $y_{2t} = y_t - A_2x_{2t} - A_3x_{3t}$ corresponding to the underlying data generating mechanism (1) and (2). It also can be shown that a similar result holds when there are regressors which are integrated of a higher order. Therefore, the standard regression theory for the stationary variables applies to the least squares estimators of A_1: these estimators are consistent if $\Sigma_{10} = 0$ and otherwise, are inconsistent. Moreover, adding integrated variables as well as a time trend to these regressions only induces $0_p(T^{-1})$ changes in the least squares estimators of A_1 and does not affect the limiting distributions given in Theorem 3.1(b). We, of course, only need joint asymptotic normality of $T^{-1/2}(\Sigma_1^T x_{1t}u_t' - \Sigma_{10})$ and $T^{-1/2}(\Sigma_1^T x_{1t}x_{1t}' - \Sigma_1)$ for Theorem 3.1(b) to be valid. Notice also that when $\Sigma_{10} = 0$, the covariance matrix V is reduced to $(I \otimes \Sigma_1^{-1})V_{10}^0(I \otimes \Sigma_1^{-1})$.

The asymptotic results for the remainder of the regression coefficients in (8) to (10) are given below in terms of the functional

$$f(B,M,E) = \left[\int_0^1 dBM' + E'\right]\left[\int_0^1 MM'\right]^{-1}$$

which we introduced in Part 1.

THEOREM 3.2. *The limiting distributions of* $T(\hat{A}_2 - A_2)$, $T(\bar{A}_2 - A_2)$, $\sqrt{T}(\bar{\mu} - \mu)$, $T(\tilde{A}_2 - A_2)$, $\sqrt{T}(\tilde{\mu} - \mu)$ *and* $T^{3/2}(\tilde{\theta} - \theta)$ *in the least squares regressions (8) to (10) can be represented in the form* $f(P, M(B_2), E(\Pi, B_2))$, *where*

$$P = B_0 - \Sigma_{10}'\Sigma_1^{-1}B_1, \quad \Pi = \Delta_{20} - \Delta_{21}\Sigma_1^{-1}\Sigma_{10}$$

and $M(\cdot)$ *and* $E(\cdot,\cdot)$ *are given for each estimator precisely as in Theorems 3.1 to 3.3 of Part 1.*

All the least squares estimators considered in Theorem 3.2 are consistent even if $\Sigma_{10} \neq 0$, and the stationary regressors are contemporaneously correlated with the regression errors. We obviously allow the integrated regressors to be correlated with the errors as well. Note particularly that the rates of convergence for these estimators are exactly the same as those for the least squares estimators in regressions with only integrated regressors, as considered in Part 1. Thus, as far as consistency is concerned, including stationary regressors in a regression with integrated variables is innocuous irrespective of possible correlation with the regression errors.

When $\Sigma_{10} = 0$, we have $P = B_0$ and $\Pi = \Delta_{20}$ in Theorem 3.2, and each of the estimators there has the same limiting distribution as the corresponding least squares estimator in the same regressions with only integrated regressors. This can be easily seen by comparing Theorem 3.2 with Theorems 3.1–3.3 of Part 1. More specifically, the least squares estimators of A_2, μ, and θ in (8) to (10) behave asymptotically, if $\Sigma_{10} = 0$, just as the regression coefficients in

$$y_{1t} = \hat{A}_2 x_{2t} + \hat{u}_t \tag{8$'$}$$

$$y_{1t} = \bar{\mu} + \bar{A}_2 x_{2t} + \bar{u}_t \tag{9$'$}$$

$$y_{1t} = \tilde{\mu} + \tilde{\theta} t + \tilde{A}_2 x_{2t} + \tilde{u}_t \tag{10$'$}$$

where $y_{1t} = y_t - A_1 x_{1t}$; the asymptotic results given in Part 1 apply directly to the estimators in (8)′ to (10)′.

The limiting distributions of the least squares estimators in (11) to (13), except for those of A_1 which are given by Theorem 3.1, can also be represented simply in terms of the functional f. In the formulae below, these limiting distributions are explicitly given in terms of $f(P,N,E)$ where $P(r)$ is the n-vector process given in Theorem 3.2 and $N = N(B_2,B_3)$ is a function of the two Brownian motions B_2 and B_3.

THEOREM 3.3. *We have in regressions (11) to (13)*

(a) $T(\hat{A}_2 - A_2) \Rightarrow f(P,Q_1,\Pi), \quad T^2(\hat{A}_3 - A_3) \Rightarrow f(P,Q_2,\Gamma)$

(b) $T(\bar{A}_2 - A_2) \Rightarrow f(P,Q_1^*,\Pi), \quad T^2(\bar{A}_3 - A_3) \Rightarrow f(P,Q_2^*,\Gamma^*)$

$\sqrt{T}(\bar{\mu} - \mu) \Rightarrow f(P,\bar{P}_1,\gamma_1)$

(c) $T(\tilde{A}_2 - A_2) \Rightarrow f(P,Q_1^{**},\Pi), \quad T^2(\tilde{A}_3 - A_3) \Rightarrow f(P,Q_2^{**},\Gamma^{**})$

$\sqrt{T}(\tilde{\mu} - \mu) \Rightarrow f(P,\bar{P}_2,\gamma_2), \quad T^{3/2}(\tilde{\theta} - \theta) \Rightarrow f(P,\bar{P}_3,\gamma_3).$

Here

$$Q_1(r) = B_2(r) - \left[\int_0^1 B_2\bar{B}_3'\left(\int_0^1 \bar{B}_3\bar{B}_3'\right)^{-1}\right]\bar{B}_3(r)$$

$$Q_2(r) = \bar{B}_3(r) - \left[\int_0^1 \bar{B}_3 B_2'\left(\int_0^1 B_2 B_2'\right)^{-1}\right]B_2(r)$$

and if we define $R(r)$ by $R(r)' = (Q_1(r)', \bar{B}_3(r)')$, then

$$\bar{P}_1(r) = 1 - \left[\int_0^1 R'\left(\int_0^1 RR'\right)^{-1}\right]R(r)$$

$$\bar{P}_2(r) = 1 - \frac{3}{2}r - \left[\int_0^1 R^{+\prime}\left(\int_0^1 R^+R^{+\prime}\right)^{-1}\right]R^+(r)$$

$$\bar{P}_3(r) = r - \frac{1}{2} - \left[\int_0^1 sR^{*\prime}\left(\int_0^1 R^*R^{*\prime}\right)^{-1}\right]R^*(r)$$

where $R^*(r) = R(r) - \int_0^1 R$ and $R^+(r) = R(r) - 3r\int_0^1 sR$. Moreover, $Q_j^*(r)$ is defined from $B_2^*(r) = B_2(r) - \int_0^1 B_2$ and $\bar{B}_3^*(r) = \bar{B}_3(r) - \int_0^1 \bar{B}_3$, and $Q_j^{**}(r)$ is defined from $B_2^{**}(r) = B_2^*(r) - 12r\int_0^1 sB_2^*$ and $\bar{B}_3^{**}(r) = \bar{B}_3^*(r) - 12r\int_0^1 s\bar{B}_3^{**}$ in the same way as $Q_j(r)$ for $j = 1,2$. Also, if we let $B_2^+(r) = B_2(r) - 3r\int_0^1 sB_2$ and $\bar{B}_3^+(r) = \bar{B}_3(r) - 3r\int_0^1 s\bar{B}_3$, and define $Q_1^+(r)$ accordingly, then for Π defined in Theorem 3.2,

$$\Gamma = -\left(\int_0^1 \bar{B}_3B_2'\right)\left(\int_0^1 B_2B_2'\right)^{-1}\Pi, \quad \gamma_1 = -\int_0^1 Q_1'\left(\int_0 Q_1Q_1'\right)^{-1}\Pi,$$

$$\gamma_2 = -\int_0^1 Q_1^{+\prime}\left(\int_0^1 Q_1^+Q_1^{+\prime}\right)^{-1}\Pi, \quad \gamma_3 = -\int_0^1 sQ_1^{*\prime}\left(\int_0^1 Q_1^*Q_1^{*\prime}\right)^{-1}\Pi.$$

Finally, Γ^* and Γ^{**} are defined, respectively, from $(B_2^*(r), \bar{B}_3^*(r))$ and $(B_2^{**}(r), \bar{B}_3^{**}(r))$ in a manner analogous to that of Γ from $(B_2, \bar{B}_3)$.

The above theorem together with Theorem 3.1 completely specifies the asymptotic behavior of the regression coefficients in (11) to (13). Theorem 3.3 gives asymptotic results for the various least squares estimators in a very general and simple functional form with differences only in the respective arguments. Similar representations arose in our results in Part 1, and this again shows how useful this functional is for the study of asymptotic theory in regressions with integrated processes. Many of the interesting results that follow later in the paper can be deduced quite easily from the representations given in Theorem 3.3.

It is interesting to note that the stochastic processes defined in Theorem 3.3 can be interpreted as the projection residuals in an appropriate Hilbert space. Thus, treating $C[0,1]$ as a subspace of the Hilbert space $L^2[0,1]$ with inner product $\int_0^1 g_1 g_2$ for square integrable functions g_1 and g_2 defined on $[0,1]$, we find that each element of $Q_1(r)$ is just the residual from the projection of the corresponding element of $B_2(r)$ onto the subspace of the Hilbert space spanned by $\{\bar{B}_{3j}\}_{j=1}^{m_3}$ for a given realization of these stochastic processes. We may equivalently define $Q_1(r)$ to be simply the residual from the continuous time regression

$$B_2(r) = \hat{A}\bar{B}_3(r) + Q_1(r)$$

where $\hat{A}$ minimizes the continuous time least squares criterion

$$\int_0^1 (B_2(r) - \hat{A}\bar{B}_3(r))'(B_2(r) - \hat{A}\bar{B}_3(r))dr.$$

Thus the projection operation is preserved in a well defined sense under the asymptotics. All the other stochastic processes that occur in the statement of Theorem 3.3 can be obtained in a similar fashion.

When the stationary regressors are excluded from regressions (11) to (13) and if $A_1 = 0$ in (2), (2)′, and (2)″, the results in Theorem 3.3 remain valid if we just replace $P(r)$ by $B_0(r)$, and Π by Δ_{20}. This can be seen easily from the proof of the theorem. The limiting distributions of the least squares estimators in this case are therefore only special cases of those given in Theorem 3.3 when $\Sigma_{10} = 0$. Including stationary variables in the regressions thus does not affect the asymptotics of the least squares estimators as long as they are not contemporaneously correlated with the regression errors. This parallels a similar result in Theorem 3.2.

Once again, all the least squares estimators in the regression equations (11) to (13), except for those of A_1, are consistent regardless of the correlation between regressors and regression errors. The estimators of A_2 are $0_p(T^{-1})$-consistent as in the case of regressions (8) to (10). We can therefore expect in all of these regressions that consistency of these estimators holds even if we relax some of our conditions imposed on the underlying models, for example, the zero mean condition for $\{u_t\}$. The rate of convergence for the estimators of A_3 in (11) to (13) is even faster and is of order $0_p(T^{-2})$. It thus seems natural that more aberrant regression errors are permissible for the consistent estimation of A_3. It is easy to show that all three least squares estimators $\hat{A}_3$, $\bar{A}_3$, and $\tilde{A}_3$ of A_3 are, in fact, consistent even if the mean of $\{u_t\}$ has a time trend. This, of course, implies that $\hat{A}_3$ is still consistent if $\{y_t\}$ is generated by (2)″.

The coefficient matrix A_1 of stationary variables can be consistently estimated even when $\Sigma_{10} \neq 0$ and the least squares estimators of A_1 in (8) to (13) are inconsistent. One obvious way to obtain a consistent estimator is to use instrumental variables in a simple two-step procedure. More precisely, we first estimate the coefficients of the integrated variables as well as the constant and the time trend from a least squares regression without the stationary variables (in (8) to (13)). In the next stage we estimate A_1 using appropriate instruments for x_{1t}, as if the estimates for the regression coefficients obtained in the first step were true values. When the instruments satisfy the same condition as x_{1t}, we can easily show from Theorem 3.2 and Theorem 3.3 that this substitution of the estimates for the true parameters in the second stage only affects the asymptotics through terms of $0_p(T^{-1})$ for all our models.

Suppose now that $\{v_t\}$ has nonzero mean and we assume, by redefining $\{v_t\}$, that (3) is replaced by

$$x_{1t} = \pi_1 + v_{1t}, \quad \Delta x_{2t} = \pi_2 + v_{2t}, \quad \Delta^2 x_{3t} = \pi_3 + v_{3t}. \tag{3)'$$

Then we have:

$$x_{2t} = \pi_2 t + \sum_{j=1}^{t} v_{2j} = \pi_2 t + x_{2t}^0$$

$$x_{3t} = \pi_3 t^2 + \sum_{k=1}^{t}\sum_{j=1}^{k} v_{3j} = \pi_3 t^2 + x_{3t}^0$$

ignoring initial conditions (which would determine the constant term in $\{x_{2t}\}$ and the linear time trend term in $\{x_{3t}\}$, noting that these may be random, of course). Now the processes $\{x_{2t}\}$ and $\{x_{3t}\}$ are driven by deterministic trends as well as stochastic trends, which we have denoted by $\{x_{2t}^0\}$ and $\{x_{3t}^0\}$. The deterministic component, however, apparently dominates asymptotic behavior in both processes since the stochastic trends are of lower order in both instances.

Asymptotic results for the least squares estimators in regressions (8) to (13), when $\{x_t\}$ is driven by (3)′ instead of (3), can be obtained without difficulty by first considering appropriately transformed models as we do in Part 1. We will, however, not report the detailed asymptotics here since the results are too long and also somewhat obvious given the methodology and the results in Part 1. Instead, we briefly outline below the main effects of introducing a nonzero mean in $\{v_t\}$ on the asymptotic behavior of the least squares estimators of A in our regressions.

If $m_2 = 1$ in regressions (8) to (10), then $\{x_{2t}\}$ behaves asymptotically as if it were $\pi_2 t$ and more conventional regression theory applies. It is in fact easy to show, for example, that

$$T^{3/2}(\bar{A}_2 - A_2) \Rightarrow N[0, (12/\pi_2^2)V]$$

where $V = \Omega_0 - \Sigma_{10}'\Sigma_1^{-1}\Omega_{10} - \Omega_{10}'\Sigma_1^{-1}\Sigma_{10} + \Sigma_{10}'\Sigma_1^{-1}\Omega_1\Sigma_1^{-1}\Sigma_{10}$, which reduces to Ω_0 if $\Sigma_{10} = 0$. When $\pi_1 = 0$, then $\hat{A}_2$ satisfies a similar result with covariance matrix $(3/\pi_2^2)V$. When $m_2 > 1$ quite a different picture emerges. Most interestingly, if $\pi_1 \neq 0$, $\sqrt{T}(\hat{A}_1 - A_1)$ is no longer asymptotically normal even under the ideal condition $\Sigma_{10} = 0$. This can be clearly seen from the fact that the limiting distribution $\sqrt{T}(\hat{A}_1 - A_1)\pi_1$ is essentially equivalent to that of $\sqrt{T}(\tilde{\mu} - \mu)$ in Theorem 3.6(a) of Part 1 when $\Sigma_{10} = 0$, and it is nonnormal. The limiting distribution of $T(\hat{A}_2 - A_2)$ is also given similarly as Theorem 3.5 or Theorem 3.6 of Part 1, depending on whether $\pi_1 = 0$ or $\pi_1 \neq 0$. If $\Sigma_{10} \neq 0$ and $m_1 > 1$, the final expressions involve appropriately redefined $P(r)$ and Π in Theorem 3.2. A similar argument goes through for $T(\bar{A}_2 - A_2)$, the asymptotics being given by Theorem 3.6 of Part 1. The estimators $\bar{A}$ and $\tilde{A}$ are invariant, with respect to π_1 and both π_1 and π_2, respectively.

Standard regression asymptotics apply to (11) to (13) if $\pi_2 \neq 0$, $\pi_3 \neq 0$, and $m_2 = m_3 = 1$ in which case $\{x_{2t}\}$ and $\{x_{3t}\}$ may well be regarded, respectively, as $\pi_2 t$ and $\pi_3 t^2$ for the purpose of asymptotic theory. Thus $T^{3/2}(\tilde{A}_2 - A_2)$ and $T^{5/2}(\tilde{A}_3 - A_3)$ both are asymptotically normal with respective covariance matrices $(192/\pi_2^2)V$ and $(180/\pi_3^2)V$, where V is given above. The results for $\hat{A}_2$ and $\hat{A}_3$ can be obtained analogously as above. Once again, if $\pi_1 \neq 0$, $\hat{A}_1$ is not asymptotically normal unless $m_2 = m_3 = 1$. Finally, we note that all these estimators are consistent if (3) is replaced by (3)′ in our models.

THEOREM 3.4. *We have in regressions (8) to (10) or in (11) to (13)*

(a) $\hat{\Sigma}_0, \bar{\Sigma}_0, \tilde{\Sigma}_0 \xrightarrow{p} \Sigma_0 - \Sigma_{10}'\Sigma_1^{-1}\Sigma_{10}$.

Moreover, if (15) holds with $\Sigma_{10} = 0$ *then*

(b) $\sqrt{T}(\hat{\Sigma}_0 - \Sigma_0), \sqrt{T}(\bar{\Sigma}_0 - \Sigma_0), \sqrt{T}(\tilde{\Sigma}_0 - \Sigma_0) \Rightarrow N(0, V_0^0)$

where V_0^0 *is a submatrix of* V^0 *in (15) as defined previously.*

The least squares estimators of Σ_0 are inconsistent if $\Sigma_{10} \neq 0$, and otherwise, they are consistent in regressions (8) to (10) or in (11) to (13). Again correlation between the integrated variables and the regression errors does not affect consistency. Moreover, if $\Sigma_{10} = 0$ and (15) holds, the effect on the estimation of the error covariance matrix of including stationary variables in the regressions is at most of order $0_p(T^{-1})$ in all six regression equations. The asymptotic normality of these estimators therefore follows exactly as in Part 1 (Theorem 3.4) where all the regressors are integrated.

4. HYPQTHESIS TESTING

In this section we shall derive asymptotic results for tests of linear hypotheses that involve the regression coefficients in (8) to (13). The main case considered here is where x_{1t} is contemporaneously uncorrelated with the regression errors. If $\Sigma_{10} \neq 0$, then estimators of the covariance matrices Σ_0 and Ω_0, which are based on least squares regression residuals, are inconsistent (due to the inconsistency of estimates of A_1). However, consistent estimators of these covariance matrices can be obtained even in this case from regression residuals if we consistently estimate the coefficient matrix A_1 of the stationary variables, for example, by the two-step procedure outlined earlier in Section 3. Note that when $\Sigma_{10} \neq 0$, standard testing procedures for A_1 based on the least squares estimators in (8) to (13) make little sense since these estimators are not consistent. Appropriate tests may be constructed using the instrumental variable procedure. The usual chi-square test statistic can also be used in this case since substituting estimates for the unknown coefficients of the integrated regressors (as well as the time trend and the constant term) in the second step only affects the estimation of A_1 at order

$0_p(T^{-1})$ as mentioned earlier. Moreover, as long as we are interested only in regression coefficients other than A_1, tests can always be based on the least squares regressions without the stationary regressors, which are, in fact, the first stage of the two-step procedure for the estimation of A_1. The results for these regressions are given, if we simply redefine the regression errors $\{u_t\}$, in Part 1 for models (1), (1)$'$, and (1)$''$, and in Theorem 4.2 below for (2), (2)$'$, and (2)$''$.

We first look at the regression equations (8) to (10) and consider null hypotheses of the form

$$R_1 \operatorname{vec} A_1 = r_1 \tag{18}$$

$$R_2 \operatorname{vec} A_2 = r_2 \tag{19}$$

where R_j and r_j are known $q_j \times nm_j$, $q_j \times 1$ matrices, respectively, and R_j is of full rank q_j for $j = 1,2$. Other hypotheses of interest are

$$\mu = \mu_0 \text{ and } \theta = \theta_0. \tag{20}$$

We commonly employ the Wald statistic for testing hypotheses (18) to (20). Denoting by $\operatorname{diag}(K)$ the block diagonal entries of a square matrix K, define

$$\operatorname{diag}(X'X)^{-1} = \begin{array}{cc} m_1 & m_2 \\ (M_{11}, & M_{12}), \end{array} \qquad \operatorname{diag}(X^{1\prime}X^1)^{-1} = \begin{array}{ccc} 1 & m_1 & m_2 \\ (m_{21}, & M_{21}, & M_{22}), \end{array}$$

$$\operatorname{diag}(X^{2\prime}X^2)^{-1} = \begin{array}{cccc} 1 & 1 & m_1 & m_2 \\ (m_{31}, & m_{32}, & M_{31}, & M_{32}) \end{array} \tag{21}$$

where X^1 and X^2 are given by (14). We will not need to be specific about off block diagonal entries in (21). In (8) the Wald statistics for testing (18) and (19) are given by

$$F_1(\hat{A}_j) = (R_j \operatorname{vec} \hat{A}_j - r_j)'[R_j(\hat{\Sigma}_0 \otimes M_{1j})R_j']^{-1}(R_j \operatorname{vec} \hat{A}_j - r_j)$$

for $j = 1,2$. The statistics for the same hypotheses in regressions (9) and (10), which will be denoted by $F_1(\bar{A}_j)$ and $F_1(\tilde{A}_j)$, respectively, can be constructed similarly by substituting $\bar{A}_j$ or $\tilde{A}_j$ for $\hat{A}_j$, $\bar{\Sigma}_0$ or $\tilde{\Sigma}_0$ for $\hat{\Sigma}_0$, and finally M_{2j} or M_{3j} for M_{1j} in (22). Also, the tests of (20) in (10) can be based on

$$F_1(\tilde{\mu}) = m_{31}^{-1}(\tilde{\mu} - \mu_0)'\tilde{\Sigma}_0^{-1}(\tilde{\mu} - \mu_0), \quad F_1(\tilde{\theta}) = m_{32}^{-1}(\tilde{\theta} - \theta_0)'\tilde{\Sigma}_0^{-1}(\tilde{\theta} - \theta_0)$$

and $F_1(\bar{\mu})$ is defined similarly with m_{21} and $\bar{\Sigma}_0$. Moreover, we may want to perform joint tests, and we denote, for example, by $F_1(\hat{A}_1, \hat{A}_2)$ the Wald statistic for the joint test of (18) and (19) in regression equation (8).

As in Part 1 and following the same general nomenclature therein, we also define G_1-statistics for the tests of (19) and (20) (not for (18)) from the corresponding F_1-statistics simply by replacing $\hat{\Sigma}_0$, $\bar{\Sigma}_0$, and $\tilde{\Sigma}_0$ with $\tilde{\Omega}_0$, a consistent estimator of Ω_0. Our subsequent theory for tests of hypotheses (19) and (20) is focused on the G_1-statistics rather than the standard Wald statistics. For more discussion on the rationale behind the G_1-statistics and for the consistent estimation of Ω_0, see Part 1. If $\{u_t\}$ is a white noise, or a martingale difference sequence, then the two statistics are asymptotically equivalent.

The following theorem summarizes the asymptotic results for these statistics:

THEOREM 4.1. *Suppose* $\Sigma_{10} = 0$ *in regressions (8) to (10).*

(a) *If (15) holds with the covariance matrix given by (17), then*

$F_1(\hat{A}_1),\ F_1(\bar{A}_1),\ F_1(\tilde{A}_1) \Rightarrow \chi^2_{q_1}$.

(b) *The limiting distributions of the* G_1*-statistics for testing (19) or (20) are identical to those of the corresponding G-statistics defined in Part 1.*
(c) *Given part (a) above, the conventional Wald tests on* A_1 *are asymptotically independent of the* G_1 *(and* F_1*)-statistics for testing (19) or (20).*

Parts (a) and (b) of the above theorem are what we would expect from Theorems 3.1 and 3.2. Theorem 3.1 shows that the least squares estimators of A_1 in (8) to (10) behave asymptotically exactly as if there were no integrated regressors in either models (1) to (1)″ or the fitted regressions (8) to (10). Theorem 4.1(a) just confirms that this is also true for the Wald statistics, which have limiting chi-square distributions under ideal conditions. This observation also makes it plain that other testing procedures are possible for the hypothesis (18) and remain valid for the more general case considered here, since the problem is equivalent asymptotically to hypothesis testing in the standard regression model. For a detailed treatment of this subject, see White [13].

Given (15) and $\Sigma_{10} = 0$ (as is often assumed in standard regression theory) a test of hypothesis (18), for example from regression (8), can be based on the (asymptotic chi-squared) statistic

$$(R_1 \operatorname{vec} \hat{A}_1 - r_1)'[R(I \otimes \hat{\Sigma}_1^{-1})\hat{V}_{10}^0(I \otimes \hat{\Sigma}_1^{-1})R']^{-1}(R_1 \operatorname{vec} \hat{A}_1 - r_1)$$

where $\hat{\Sigma}_1 = T^{-1}\Sigma_1^T x_{1t}x_{1t}'$ and $\hat{V}_{10}^0$ is a consistent estimate of V_{10}^0. If V^0 is given by (16), then the consistent estimation of this covariance matrix is essentially equivalent to that of Ω_{10} and may be conducted in the usual way (see Part 1).

Theorem 4.1(b) similarly implies that including stationary variables in the regressions, as long as they are contemporaneously uncorrelated with the regression errors, does not affect the limiting distributions of the Wald statistics as well as those of the least squares estimators. Thus, the tests of (19) and (20) in regressions (8) to (10) are asymptotically equivalent to the corresponding tests in (8)′ to (10)′, the asymptotic results for which are given in Theorem 4.1 of Part 1. Now it is obvious that the transformations introduced in Part 1 to eliminate or reduce the nuisance parameter dependencies remain valid (with only trivial change in notation), and that the limiting distributions of the (so transformed) H-statistics are as given by Theorem 4.2 of Part 1.

Finally, in view of Lemma A.1 in Part 1, which shows that the Wald statistics for the joint tests can be written simply as the sum of the two Wald statistics for each individual test, Theorem 4.1(c) completely specifies the limiting distributions of the test statistics for the joint tests, for example $F_1(\hat{A}_1, \hat{A}_2)$.

We now consider in the regression equations (11) to (13) the tests for the null hypotheses (18) to (20) and

$$R_3 \operatorname{vec} A_3 = r_3 \tag{22}$$

where R_3 and r_3 are defined as their counterparts in (18) and (19). As in (21) we also define

$$
\begin{aligned}
\operatorname{diag}(X'X)^{-1} &= (\overset{m_1}{M_{11}},\ \overset{m_2}{M_{12}},\ \overset{m_3}{M_{13}}),\\
\operatorname{diag}(X^{1\prime}X)^{-1} &= (\overset{1}{m_{21}},\ \overset{m_1}{M_{21}},\ \overset{m_2}{M_{22}},\ \overset{m_3}{M_{23}}),\\
\operatorname{diag}(X^{2\prime}X^2)^{-1} &= (\overset{1}{m_{31}},\ \overset{1}{m_{32}},\ \overset{m_1}{M_{31}},\ \overset{m_2}{M_{32}},\ \overset{m_3}{M_{33}})
\end{aligned} \tag{23}
$$

where X^1 and X^2 are given in (14). Again, it will not be necessary to specify off block diagonal entries in (23). The notation given in (23) replaces that used earlier and should not be a source of confusion. The Wald statistics for testing (18) to (20) and (22) are then constructed in an analogous fashion as for (8) to (10). We denote these new Wald statistics, based on the least squares regressions (11) to (13), by $F_2(\cdot)$. We also define corresponding G_2-statistics for tests of (19), (20), and (22) again by replacing $\hat{\Sigma}_0$, $\bar{\Sigma}_0$, and $\tilde{\Sigma}_0$ with $\tilde{\Omega}_0$ as before.

In the following theorem, the limiting distributions of the G_2-statistics for testing hypotheses (19), (20), and (22) are presented in terms of the functional

$$
\begin{aligned}
g_R(B,N,E) = {} & \left(\int_0^1 dB \otimes N + e\right)' \left[I \otimes \left(\int_0^1 NN'\right)^{-1}\right] R' \\
& \cdot \left\{R\left[\Omega \otimes \left(\int_0^1 NN'\right)^{-1}\right]R'\right\} R\left[I \otimes \left(\int_0^1 NN'\right)^{-1}\right]\left(\int_0^1 dB \otimes N + e\right)
\end{aligned}
$$

where $e = \operatorname{vec} E'$ and Ω is the covariance matrix of the vector Brownian motion B. The functional g_R was introduced in Part 1, and the only differ-

ence here is that N is now a function of two Brownian motions, i.e. $N = N(B_2,B_3)$, as in Theorem 3.3. We also define $g(B,N,E) = g_I(B,N,E)$.
We have

THEOREM 4.2. *Let $\Sigma_{10} = 0$ and consider regressions (11)–(13). First,*

(a) *If (15) holds with covariance matrix given by (17), then*

$F_2(\hat{A}_1),F_2(\bar{A}_1),F_2(\tilde{A}_1) \Rightarrow \chi^2_{q_1}$.

Also, for the other tests,

(b) $G_2(\hat{A}_2) \Rightarrow g_{R_2}(B_0,Q_1,\Delta_{20}),\quad G_2(\hat{A}_3) \Rightarrow g_{R_3}(B_0,Q_2,\Gamma_0)$

(c) $G_2(\bar{A}_2) \Rightarrow g_{R_2}(B_0,Q_1^*,\Delta_{20}),\quad G_2(\bar{A}_3) \Rightarrow g_{R_3}(B_0,Q_2^*,\Gamma_0^*)$

$\quad G(\bar{\mu}) \Rightarrow g(B_0,\bar{P}_1,\gamma_1)$

(d) $G_2(\tilde{A}_2) \Rightarrow g_{R_2}(B_0,Q_1^{**},\Delta_{20}),\quad G_2(\tilde{A}_3) \Rightarrow g_{R_3}(B_0,Q_2^{**},\Gamma_0^{**})$

$\quad G_2(\tilde{\mu}) \Rightarrow g(B_0,\bar{P}_2,\gamma_2),\quad G_2(\tilde{\theta}) \Rightarrow g(B_0,\bar{P}_3,\gamma_3)$

where notation is defined in Theorem 3.3, except for Γ_0, Γ_0^*, and Γ_0^{**} which are defined analogously to Γ, Γ^*, and Γ^{**} but with Δ_{20} replacing Π in these definitions. Additionally,

(e) *Given part (a) the conventional Wald tests on A_1 are asymptotically independent of the G_2 (and F_2)-statistics in (b)–(d).*

Part (a) of the above theorem is entirely analogous to that of Theorem 4.1(a). It implies that hypothesis testing on the coefficients of the stationary regressors in (11)–(13) reduces, at least in terms of the relevant asymptotics, to testing in a regression with only stationary variables. Now one can expect this to be true in general for regressions that involve both stationary variables and integrated processes. In fact, it is not difficult to show that the same result holds in regressions with higher order integrated regressors. Again, other procedures for testing (18) are possible using appropriately constructed (asymptotically) chi-square statistics, as suggested previously.

The limiting distributions of the G_2-statistics for testing (19), (20), and (22) are explicitly given in (b)–(d) for regressions (11)–(13). The results for the case where there are no stationary variables in the regressions are easily obtained from Theorem 4.2(b)–(d) just by taking $\Sigma_{10} = 0$. Moreover, by virtue of Lemma A.1 of Part 1, the limiting distributions of the G_2-statistics for the joint tests of (19), (20), and (22) can also be easily deduced from Theorem 4.2(b)–(d). For example, assuming $R_2 = I$ in (20) to avoid unnecessary complications, we have

$$G_2(\hat{A}_2,\hat{A}_3) \Rightarrow g(B_0,Q_1,\Delta_{20}) + g_{R_3}(B_0,\bar{B}_3,0)$$

$$G_2(\bar{A}_2,\bar{A}_3) \Rightarrow g(B_0,Q_1^*,\Delta_{20}) + g_{R_3}(B_0,\bar{B}_3^*,0)$$

$$G_2(\tilde{A}_2,\tilde{A}_3) \Rightarrow g(B_0,Q_1^{**},\Delta_{20}) + g_{R_3}(B_0,\bar{B}_3^{**},0)$$

where notation is defined in Theorem 3.3. The limiting distributions of other joint test statistics can be represented similarly and will not be reported here. Finally, Theorem 4.2(e) again completely specifies the asymptotic results for joint tests which also include the hypothesis (18).

We are now in a position to define the transformations which give rise to the H_2-statistics. These statistics are analogous to the H-statistics developed in Part 1. In particular, the motivations are identical and the limiting distributions of the H_2-statistics for the simple null hypotheses of the form $R_2 = I$ and $R_3 = I$ do not depend on Δ_{20}. The nuisance parameter dependency of the asymptotic distributions is therefore concentrated in the covariance matrix of the Brownian motions, and this dependency disappears or reduces to a parsimoniously tractable form in some special cases that we will consider in Section 5. For further discussion on this point, see Part 1.

For tests involving regression coefficients from (11) we define:

$$\begin{aligned}
H_2(\hat{A}_2) &= G_2(\hat{A}_2) - 2T \operatorname{tr} \tilde{\Omega}_0^{-1}\tilde{\Delta}_{20}'(\hat{A}_2 - A_2) \\
&\quad + T^2 \operatorname{tr} \tilde{\Omega}_0^{-1}\tilde{\Delta}_{20}'[X_2'(I - P_{X_3})X_2]^{-1}\tilde{\Delta}_{20} \\
&= G_2(\hat{A}_2) + a, \text{ say} \\
H_2(\hat{A}_3) &= G_2(\hat{A}_3) - 2T^2 \operatorname{tr} \tilde{\Omega}_0^{-1}\tilde{\Gamma}_1'(\hat{A}_3 - A_3) \\
&\quad + T^4 \operatorname{tr} \tilde{\Omega}_0^{-1}\tilde{\Gamma}_1'[X_3'(I - P_{X_2})X_3]^{-1}\tilde{\Gamma}_1 \\
H_2(\hat{A}_2,\hat{A}_3) &= G_2(\hat{A}_2,\hat{A}_3) + a.
\end{aligned}$$

For (12) with the constant term, we set:

$$\begin{aligned}
H_2(\bar{A}_2) &= G_2(\bar{A}_2) - 2T \operatorname{tr} \tilde{\Omega}_0^{-1}\tilde{\Delta}_{20}'(\bar{A}_2 - A_2) \\
&\quad + T^2 \operatorname{tr} \tilde{\Omega}_0^{-1}\tilde{\Delta}_{20}'[X_2^{*\prime}(I - P_{X_3^*})X_2^*]^{-1}\tilde{\Delta}_{20} \\
&= G_2(\hat{A}_2) + b, \text{ say} \\
H_2(\bar{A}_3) &= G_2(\bar{A}_3) - 2T^2 \operatorname{tr} \tilde{\Omega}_0^{-1}\tilde{\Gamma}_2'(\bar{A}_3 - A_3) \\
&\quad + T^4 \operatorname{tr} \tilde{\Omega}_0^{-1}\tilde{\Gamma}_2'[X_3^{*\prime}(I - P_{X_2^*})X_3^*]^{-1}\tilde{\Gamma}_2 \\
H_2(\bar{\mu}) &= G_2(\bar{\mu}) - 2\sqrt{T}\tilde{\gamma}_1\tilde{\Omega}_0^{-1}(\bar{\mu} - \mu) + T\tilde{\gamma}_1\tilde{\Omega}_0^{-1}\tilde{\gamma}_1' RSS_1^{-1} \\
H_2(\bar{A}_2,\bar{A}_3) &= G_2(\bar{A}_2,\bar{A}_3) + b, \\
H_2(\bar{\mu},\bar{A}_2,\bar{A}_3) &= G_2(\bar{\mu},\bar{A}_2,\bar{A}_3) + b.
\end{aligned}$$

Similarly, for regression (13) with the time trend we define:

$$\begin{aligned}
H_2(\tilde{A}_2) &= G_2(\tilde{A}_2) - 2T \operatorname{tr} \tilde{\Omega}_0^{-1}\tilde{\Delta}_{20}'(\tilde{A}_2 - A_2) \\
&\quad + T^2 \operatorname{tr} \tilde{\Omega}_0^{-1}\tilde{\Delta}_{20}'[X_3^{**\prime}(I - P_{X_3^{**}})X_2^{**}]^{-1}\tilde{\Delta}_{20} \\
&= G_2(\tilde{A}_2) + c, \text{ say}
\end{aligned}$$

$$H_2(\tilde{A}_3) = G_2(\tilde{A}_3) - 2T^2 \operatorname{tr} \tilde{\Omega}_0^{-1}\tilde{\Gamma}_3'(\tilde{A}_3 - A_3) + T^4 \operatorname{tr} \tilde{\Omega}_0^{-1}\tilde{\Gamma}_3'[X_3^{**\prime}(I - P_{X_2^{**}})X_3^{**}]^{-1}\tilde{\Gamma}_3$$

$$H_2(\tilde{\mu}) = G_2(\tilde{\mu}) - 2\sqrt{T}\tilde{\gamma}_2\tilde{\Omega}_0^{-1}(\tilde{\mu} - \mu) + T\tilde{\gamma}_2\tilde{\Omega}_0^{-1}\tilde{\gamma}_2' RSS_2^{-1}$$

$$H_2(\tilde{\theta}) = G_2(\tilde{\theta}) - 2T^{3/2}\tilde{\gamma}_3\tilde{\Omega}_0^{-1}(\tilde{\theta} - \theta) + T^3\tilde{\gamma}_3\tilde{\Omega}_0^{-1}\tilde{\gamma}_3' RSS_3^{-1}$$

$$H_2(\tilde{A}_2,\tilde{A}_3) = G_2(\tilde{A}_2,\tilde{A}_3) + c,$$

$$H_2(\tilde{\mu},\tilde{\theta},\tilde{A}_2,\tilde{A}_3) = G_2(\tilde{\mu},\tilde{\theta},\tilde{A}_2,\tilde{A}_3) + c.$$

Here $\tilde{\Delta}_{20}$ is a consistent estimate of Δ_{20} which can be obtained as in Part 1, X^* and X^{**} are matrices, the t-th rows of which are, respectively, deviations of x_t from the sample mean and the fitted time trend. Also, we define for computational convenience the least squares regression equations

(i) $x_{3t} = \hat{\Pi}_1 x_{2t} + e_{1t}$

(ii) $x_{3t} = \hat{\mu} + \hat{\Pi}_2 x_{2t} + e_{2t}$

(iii) $x_{3t} = \hat{\mu} + \hat{\theta}t + \hat{\Pi}_3 x_{2t} + e_{3t}$

for a sample of size T, and let $\tilde{\Gamma}_j = T^{-1}\hat{\Pi}_j\tilde{\Delta}_{20}$ $(j = 1,2,3)$. Similarly, in the regressions

(i) $1 = x_{2t}'\hat{\beta}_1 + x_{3t}'\hat{\delta}_1 + e_{1t}$

(ii) $1 = \hat{\theta}t + x_{2t}'\hat{\beta}_2 + x_{3t}'\hat{\delta}_2 + e_{2t}$

(iii) $t = \hat{\mu} + x_{2t}'\hat{\beta}_3 + x_{3t}'\hat{\delta}_3 + e_{3t}$

we let RSS_j be the residual sum of squares from the j-th regression ($j = 1$, 2, 3) and let $\tilde{\gamma}_j = -T^{1/2}\hat{\beta}_j'\tilde{\Delta}_{20}$ ($j = 1$, 2) and $\tilde{\gamma}_3 = -T^{-1/2}\hat{\beta}_3'\tilde{\Delta}_{20}$.

If $\Delta_{20} = 0$, then $\Gamma_0 = 0$ and $\gamma_j = 0$ for all $j = 1$, 2, 3. Therefore, if the x_{2t}'s are strictly exogenous or lagged variables whose innovations are only contemporaneously correlated with the regression errors, then the above transformations are unnecessary and the H_2- and G_2-statistics are asymptotically equivalent. Note that we do not need a correction for possible serial correlation between the regression errors and the innovations of $\{x_{3t}\}$. Finally, we conclude this section by presenting the limiting distributions of the H_2-statistics in terms of a functional h, which is given by

$$h(B,N) = g(B,N,0) = g_I(B,N,0)$$

as in Part 1.

THEOREM 4.3. *Let $\Sigma_{10} = 0$. Then for the regression equations (11)–(13) we have*

(a) $H_2(\hat{A}_2) \Rightarrow h(B_0,Q_1)$, $\quad H_2(\hat{A}_3) \Rightarrow h(B_0,Q_2)$,

$\quad H_2(\hat{A}_2,\hat{A}_3) \Rightarrow h(B_0,Q_1) + h(B_0,\bar{B}_3')$

(b) $H_2(\bar{A}_2) \Rightarrow h(B_0,Q_1^*),\quad H_2(\bar{A}_3) \Rightarrow h(B_0,Q_2^*)$

$H_2(\bar{\mu}) \Rightarrow h(B_0,\bar{P}_1)$

$H_2(\bar{A}_2,\bar{A}_3) \Rightarrow h(B_0,Q_1^*) + h(B_0,\bar{B}_3^*)$

$H_2(\bar{\mu},\bar{A}_2,\bar{A}_3) \Rightarrow h(B_0,1) + h(B_0,Q_1^*) + h(B_0,\bar{B}_3^*)$

(c) $H_2(\tilde{A}_2) \Rightarrow h(B_0,Q_1^{**}),\quad H_2(\tilde{A}_3) \Rightarrow h(B_0,Q_2^{**})$

$H_2(\tilde{\mu}) \Rightarrow h(B_0,\bar{P}_2),\quad H_2(\tilde{\theta}) \Rightarrow h(B_0,\bar{P}_3)$

$H_2(\tilde{A}_2,\tilde{A}_3) \Rightarrow h(B_0,Q_1^{**}) + h(B_0,\bar{B}_3^{**})$

$H_2(\tilde{\mu},\tilde{\theta},\tilde{A}_2,\tilde{A}_3) \Rightarrow h(B_0,1) + h(B_0,r^*) + h(B_0,Q_1^{**}) + h(B_0,\bar{B}_3^{**})$

where $r^* = r - \frac{1}{2}$ *and other notation is defined in Theorem 3.3.*

5. SPECIALIZATIONS

5.1. Regressions with Strictly Exogenous Integrated Regressors

In view of Theorem 3.2 and Theorem 4.1, it is easy to see that when $\Sigma_{10} = 0$ Theorem 5.2 and Theorem 5.4 of Part 1 apply to the regression coefficients of the integrated variables as well as the constant and the time trends in (8)–(10), since they are asymptotically equivalent, both for estimation and hypothesis testing, to regressions (8)′–(10)′. Hence all the theory in Section 5.1 of Part 1 remains valid.

We can, of course, expect similar results for regression equations (11)–(13). More explicitly, we assume that the integrated regressors in these regressions are strictly exogenous or, in other words, that the processes $\{v_{2t}\}$ and $\{v_{3t}\}$, which drive $\{x_{2t}\}$ and $\{x_{3t}\}$, are generated independently from the regression error process $\{u_t\}$. It follows then that both B_2 and B_3 in the results of Theorem 3.3 become independent of B_0 since $\Omega_{20} = \Omega_{30} = 0$.

The following theorem, which is parallel to Theorem 5.2 of Part 1, characterizes the asymptotic behavior of the least squares estimators in (11)–(13) when the integrated regressors are strictly exogenous. It can be easily deduced from Lemma 5.1 of Part 1.

THEOREM 5.1. *Suppose* $\Sigma_{10} = 0$, $\Omega_{20} = \Delta_{20} = 0$, *and* $\Omega_{30} = 0$. *Then we have in regressions (11)–(13)*

(a) $(\hat{A}_2 - A_2)M_{12}^{-1/2},(\bar{A}_2 - A_2)M_{22}^{-1/2},(\tilde{A}_2 - A_2)M_{32}^{-1/2} \Rightarrow N(0,\Omega_0 \otimes I_{m_2})$

(b) $(\hat{A}_3 - A_3)M_{13}^{-1/2},(\bar{A}_3 - A_3)M_{23}^{-1/2},(\tilde{A}_3 - A_3)M_{33}^{-1/2} \Rightarrow N(0,\Omega_0 \otimes I_{m_3})$

(c) $(\bar{\mu} - \mu)m_{21}^{-1/2},(\tilde{\mu} - \mu)m_{31}^{-1/2},(\tilde{\theta} - \theta)m_{32}^{-1/2} \Rightarrow N(0,\Omega_0)$

where notation is defined in (23).

Theorem 5.1 once again confirms that upon appropriate standardization the conventional asymptotic theory applies to regressions with strictly exogenously integrated regressors. It seems almost trivial to extend this result to the case of higher order integrated processes. We can write the results given in Theorem 5.1 heuristically in a form which is more compatible with classical regression theory, as in Part 1. For example, we have

$$\hat{A}_2 - A_2 \sim N(0, \Omega_0 \otimes (X_2'(I - P_{X_3})X_2)^{-1})$$

and

$$\hat{A}_3 - A_3 \sim N(0, \Omega_0 \otimes (X_3'(I - P_{X_2})X_3)^{-1})$$

conditionally on a realization of $\{x_{2t}\}$ and $\{x_{3t}\}$ for $t = 1, \ldots, T$. Similar expressions are, of course, possible for the other estimators.

The next result also follows from our previous theory and Corollary 5.3 of Part 1.

THEOREM 5.2. *Suppose* $\Sigma_{10} = 0$, $\Omega_{20} = \Delta_{20} = 0$, *and* $\Omega_{30} = 0$. Then in the regression *equations (11)–(13):*

(a) *The limiting distributions of the G_2-statistics for the hypotheses, possibly joint, of (19), (20), and (22) are chi-square, with degrees of freedom given by the number of restrictions for each test; and*

(b) *If $\Omega = \Sigma$ and (15) holds with covariance matrix given by (17), then the F_2- (and G_2)-statistics for the joint hypothesis of (18) with any of (19), (20), and (22) are also asymptotically chi-square.*

Theorem 5.2(a) is just a natural extension of Theorem 5.4 in Part 1, and (b) is an immediate consequence of Theorem 4.2(c). Thus, (b) also applies to the F_1- (and G_1)-statistics in regressions (8)–(10) in view of Theorem 4.1(c). Notice that we do not impose any restriction on Ω_{23} in either the above theorem or Theorem 5.1. We therefore allow $\{x_{2t}\}$ and $\{x_{3t}\}$ to be driven by the processes that are both serially and contemporaneously intercorrelated as long as the invariance principle (4) holds. Also, if the innovation $\{w_t\}$ is a square integrable martingale difference sequence and $\Omega = \Sigma$ as in Theorem 5.2(b), then $\Sigma_{10} = 0$, $\Sigma_{20} = 0$, $\Sigma_{30} = 0$ is sufficient to ensure that the above two theorems hold.

It may be worth noting that Theorems 5.1 and 5.2 are valid when in addition to our underlying assumptions

$$E(u_t | F_{t-1}) = 0 \quad \text{and} \quad E(u_t u_t' | F_{t-1}) = \Sigma_0$$

where $F_{t-1} = \sigma(x_1, \ldots, x_t, y_{t-1})$, a σ-field representing the information accumulated up to time $t - 1$ (i.e., including x_t which is assumed to be predetermined here). The standard statistical procedure for the linear regression is, of course, not legitimate if the above conditions are violated and the model is misspecified. Our results developed in previous sections are intended to apply in quite general situations which allow for misspecification in this sense. By their very nature they therefore include correctly specified models in which orthogonality and homoskedasticity conditions obtain. Under these conditions, the classical regression theory is fully applicable as is implied by Theorems 5.1 and 5.2.

5.2. Regressions with Cointegrated Regressors

We shall consider the n-variate linear model

$$y_t = Az_t + u_t \tag{24}$$

where the m-vector process $\{z_t\}$ is cointegrated in the sense of Engle and Granger [2]; that is, the regressors are integrated individually, but there are certain linear combinations (or cointegrating vectors) which lower their order of integration. The regression theory for the above model when $\{z_t\}$ is $I(1)$ is considered in Part 1. These results, however, do not apply to the above model as was mentioned in Part 1, since the covariance matrix of the Brownian motion which asymptotically represents $\{z_t\}$ becomes singular (see Phillips [8] and Phillips and Ouliaris [10]).

In the following, we show how the general regression theory for (24) can be readily deduced from our previous results. Now we assume in (24) that the vector process $\{z_t\}$ is cointegrated, while each variable is either $I(1)$ or $I(2)$, though it is not difficult to extend the subsequent theory to the case of higher order integrated processes. First transform (24) as:

$$y_t = A_1 x_{1t} + A_2 x_{2t} + A_3 x_{3t} + u_t \tag{25}$$

where $x_{jt} = H_j' z_t$, $A_j = AH_j$ $(j = 1,2,3)$, and $H = (H_1, H_2, H_3)$ is an orthogonal matrix with the convention that $x_{3t} \equiv 0$ if $\{z_t\}$ is $I(1)$. It is assumed that the above transformed model satisfies the conditions underlying (1) or (2) introduced in Section 2.

The asymptotic theory developed in Section 3 and Section 4 is directly applicable to the least squares regression corresponding to (25). The asymptotic results for certain linear combinations of the regression coefficients $\hat{A}$ of the original model (24), viz. $\hat{A}H_j$ $(j = 1,2,3)$, are given by the theorems in previous sections. We can, of course, consider extensions of (24) to regressions with a constant or a time trend in a similar way.

In particular, each column h of the matrix H_1 is a cointegrating vector of $\{z_t\}$ such that $\{h'z_t\}$ is stationary. We have assumed in (25) that the columns of H_1 are a set of orthonormal vectors. Being specific about the matrix constructed from such cointegrating vectors, however, may require unnecessary effort in getting the explicit limiting distribution of $\hat{A}$ for a specific model such as a VAR with unit roots. Thus, let C be any matrix (of full column rank) such that $R(C) = R(H_1)$ where $R(K)$ denotes the linear space spanned by the column vectors of the matrix K; and set $x_{1t} = C'z_t$ in

(25) with a conformable definition for A_1. We denote by $R(K)^{\perp}$ the orthogonal complement of $R(K)$.

We assumed that (15) holds, and hence $\sqrt{T}(\hat{A}_1 - A_1^*)$ has a limiting normal distribution with covariance matrix $V = (J_1 \otimes J_2)V^0(J_1' \otimes J_2')$, as given in Theorem 3.1. Our next result now follows:

THEOREM 5.3.
(a) $\sqrt{T}(\hat{A} - A^*) \Rightarrow N(0, (I \otimes C)V(I \otimes C'))$

where $A^* = A + \Sigma_{10}'\Sigma_1^{-1}C'$. *Moreover, if* V *is positive definite,* $R(R') \cap R(I \otimes C)^{\perp} = \phi$, *and* $R \operatorname{vec} A^* = r$ *for a given* $q \times nm$ *matrix* R *of rank* q, *then*

(b) $(R \operatorname{vec} \hat{A} - r)'[R(I \otimes C)V(I \otimes C')R']^{-1}(R \operatorname{vec} \hat{A} - r) \Rightarrow \chi_q^2$.

The above theorem is a direct consequence of Theorem 3.1. It is easy to show that the results are also valid for models which extend (24) by including a constant or a time trend term. The asymptotic normality of $\hat{A}$ with a singular covariance matrix was earlier found by Sims [11] in some special cases. Theorem 5.3(a) implies that for any vector $\delta \notin R(C)^{\perp}$,

$$\sqrt{T}(\hat{A} - A^*)\delta \Rightarrow N(0, (I \otimes \delta' C)V(I \otimes C'\delta)).$$

If $\delta \in R(C)^{\perp}$, then the limiting distribution of $(\hat{A} - A^*)\delta = (\hat{A} - A)\delta$ is, upon restandardization, easily obtained from Theorem 3.2 or Theorem 3.4 and is nonnormal.

In the simple case where $\Sigma_{10} = 0$ and V_{10}^0 is given by (17), we have $V = \Sigma_0 \otimes \Sigma_1^{-1}$ and the limiting covariance matrix of $\sqrt{T}(\hat{A} - A)$ in Theorem 5.3(a) reduces to $\Sigma_0 \otimes C\Sigma_1^{-1}C'$. Further, by Theorem 3.4, the least squares estimator $\hat{\Sigma}_0$ of Σ_0 from (24) is consistent. Hence, the covariance matrix can be consistently estimated by

$$\hat{\Sigma}_0 \otimes C\left\{C'\left[\frac{Z'Z}{T}\right]C\right\}^{-1}C' \tag{26}$$

if all the cointegrating vectors are known. However, this is not true in many interesting cases and the following result should then be useful.

Proposition 5.4:

$$\left[\frac{Z'Z}{T}\right]^{-1} = C\left\{C'\left[\frac{Z'Z}{T}\right]C\right\}^{-1}C' + O_p\left(\frac{1}{T}\right).$$

Proposition 5.4, of course, implies that if $V = \Sigma_0 \otimes \Sigma_1^{-1}$, then the limiting covariance matrix of $\sqrt{T}(\hat{A} - A)$ can be consistently estimated by $\hat{\Sigma}_0 \otimes (T^{-1}Z'Z)^{-1}$, instead of (26) which is parallel to the classical regression theory. Finally, observe that the following standard Wald statistic for testing $R \operatorname{vec} A = r$, where R is given as in Theorem 5.3(b),

$$(R \operatorname{vec} \hat{A} - r)' \{ R[\hat{\Sigma}_0 \otimes (Z'Z)^{-1}]R' \}^{-1} (R \operatorname{vec} \hat{A} - r)$$

has a limiting chi-squared distribution that is exactly the same as in the classical regression theory. All of the above results obviously hold for regressions with a constant or a time trend as well.

5.3. VAR Systems with Exogenous Variables

As mentioned earlier, our theory also applies to first order VAR systems with exogenous variables when the roots are unity. We explicitly write the models as:

$$y_t = A_1 x_{1t} + A_2 y_{t-1} + u_t, \quad A_2 = I \tag{27}$$

and

$$y_t = A_1 x_{1t} + A_2 x_{2t} + A_3 y_{t-1} + u_t, \quad A_3 = I \tag{28}$$

where $\{x_{1t}\}$ and $\{x_{2t}\}$ are generated by (3). Notice that models (27) and (28) correspond, respectively, to (1) and (2) introduced in Section 2. We call x_t "exogenous" simply because it is generated by a mechanism distinct from $\{u_t\}$. We do not presume that any conventional exogeneity condition applies to $\{x_t\}$. A special case of the above models (viz, that for which $n = m_1 = m_2 = 1$ and $\{x_{1t}\}$ and $\{x_{2t}\}$ are strictly exogenous) was studied in Nankervis and Savin [6] using Monte Carlo methods.

In the VAR system (27) with stationary exogenous variables, it is easy to see that Δ_{20} reduces to $A_1 \Lambda_{10} + \Lambda_0$, Δ_{21} to $A_1 \Lambda_1 + \Lambda_{01}$ and $B_2 = A_1 B_1 + B_0$. If the matrix

$$\begin{bmatrix} \Omega_0 & \Omega_{10}' \\ \Omega_{10} & \Omega_1 \end{bmatrix} \tag{29}$$

is positive definite, then so is the covariance matrix of B_2, and therefore the asymptotic results for the least squares regressions (8)–(10) for $x_{2t} = y_{t-1}$ are given by Theorems 3.1 and 3.2 with the substitutions indicated above. Of course, we set $\mu = \theta = 0$ here. It may be worth noting that Theorem 3.1 implies that the regression coefficient of x_{1t} aysmptotically behaves exactly as if the existence of the unit roots were known.

Moreover, when $\Sigma_0 = 0$, the limiting distributions of the G-statistics are easily deduced. Contrary to the VAR system considered in Part 1, the asymptotic distributions of the transformed H-statistics, however, are dependent upon A_1 and the parameters in (29). It is interesting to recall that this parameter dependency becomes one dimensional when $n = 1$ as shown in

Section 5.3 of Part 1. More specifically, the distributions of the various H-statistics asymptotically depend on the composite parameter

$$\rho = \frac{\omega_0^2 + a_1'\omega_{10}}{\omega_0(\omega_0^2 + 2\alpha_1'\omega_{10} + \alpha_1'\Omega_1\alpha_1)^{1/2}}$$

where (and henceforth) we use lower case letters to denote scalars and vectors. Thus, if $\Sigma_{10} = 0$, the existence of a unit root can be tested by first estimating ρ and using the table (for $m = 1$) provided at the end of Part 1, although the test based on a regression of y_t on y_{t-1} yields a much simpler procedure. Finally, if x_{1t} is a set of differenced y_t's and is not "exogenous," then the matrix (29) is singular. This case will be considered in Section 5.4.

In (28), the exogenous variables are integrated of order one and $\{y_t\}$ is effectively driven by $\{x_{2t}\}$ which dominates lower order terms. The asymptotic theory for regressions (11)–(13) with $x_{3t} = y_{t-1}$ (and $\theta = 0$) is thus given by Theorem 3.1 and Theorem 3.3 with the replacement of B_3 by A_2B_2, if A_2 is of full row rank and $A_2\Omega_2A_2'$ is positive definite. Deficiency in the row rank of A_2 implies the existence of cointegration in $\{y_t\}$, and this case can be analyzed as in the previous section. It also follows that the strict exogeneity of $\{x_{2t}\}$ together with $\Sigma_{10} = 0$ is sufficient for Theorem 5.1 and Theorem 5.2 to hold for the model (28).

This confirms some of the conjectures made by Nankervis and Savin [6] based on their extensive Monte Carlo study. They found, for example, that the standard t-statistic for testing a unit root in (28) with $a_1 = 0$ is asymptotically normal if $a_2 \neq 0$. The same result obviously does not hold in (27). One might expect, however, the limiting distribution of the t-statistic for the null hypothesis $a_2 = 1$ in (27) can be well approximated by normal if $\{x_{1t}\}$ is "nearly" integrated and strictly exogenous, and if $a_1 \neq 0$. This explains how Nankervis and Savin [6] mistakenly concluded that the statistic has a limiting normal distribution in such a case.

More precisely, the asymptotic distribution of the t-statistic is given, in this simple setup, by

$$\frac{\int_0^1 V_1\,dW_0}{\left(\int_0^1 V_1^2\right)^{1/2}} \tag{30}$$

where $V_1 = a_1\omega_1W_1 + \omega_0W_0$ and, W_0 and W_1 are two independent Brownian motions with unit variance. It is now clear that the normal approximation of (30) becomes more satisfactory as a_1 or ω_1 gets large relative to ω_0 (see Lemma 5.1 of Part 1), which is likely to occur if $\{x_{1t}\}$ is an autoregressive process with a coefficient near to unity. In fact, it is easy to see that Nankervis and Savin [6] set $\omega_1^2 = 25\omega_0^2$ in their experiment. When $a_1 = 0$, then (30) reduces to the limiting distribution of Dickey and Fuller's τ statistic, as we would expect.

When $n = m_2 = 1$ and $\Sigma_{10} = 0$ in (28) we can easily deduce from Theorem 4.3 that the limiting distributions of the various H_2-statistics depend only on $\rho = \omega_{20}/(\omega_0\omega_2)$, which is parallel to Lemma 5.6 of Part 1. The asymptotic distributions of some of the H_2-statistics in the case of $\rho = 1$ are tabulated in Hasza and Fuller [4], where they considered a (scalar) stochastic difference equation with a double unit root when the innovations are independently generated. A generalization of their model to the multivariate case with possibly serially dependent innovations will be explored in the next section.

5.4. General VAR System with Unit Roots

Based upon our previous results, a very general theory for VAR systems with unit roots and with lags of an arbitrary order can now be established. We first consider a VAR with simple unit roots which is appropriately transformed and written as:

$$y_t = A_{11}(y_{t-1} - y_{t-2}) + \ldots + A_{1p}(y_{t-p} - y_{t-p-1}) + A_2 y_{t-1} + u_t. \tag{31}$$

Here $A_2 = I$ and all the roots of the determinantal equation $\det A_1(z) = 0$, where $A_1(z) = I - \Sigma_{j=1}^{p} A_{1j} z^j$, are assumed to lie outside the unit circle. We write $A_1 = (A_{11}, \ldots, A_{1p})$, $x_{1t}' = (y_{t-1}' - y_{t-2}', \ldots, y_{t-p}' - y_{t-p-1}')$ and $x_{2t} = y_{t-1}$, and assume (31) in this notation satisfies the conditions introduced for (1) in Section 2.

With this formulation (31) would seem to fall in the framework of (27), except that the included stationary variables are not "exogenous" and the covariance matrix given by (29) is in general singular. The results for (27) therefore do not apply. In fact, if we assume $\{u_t\}$ is a weakly stationary process which has a spectral density matrix, and (5)′ holds, then the following relationship is satisfied by the submatrices of (29):

$$\Omega_1 = \iota_p\iota_p' \otimes \underline{\Omega}_1, \quad \underline{\Omega}_1 = J\Omega_0 J'$$

$$\Omega_{10} = \iota_p \otimes J\Omega_0$$

where ι_p is the p-vector of ones and $J = A_1(1)^{-1}$. It is now not difficult to show that

$$B_1 = \iota_p \otimes JB_0 \text{ a.s.}$$

and

$$B_2 = A_1 B_1 + B_0 = JB_0 \text{ a.s.}$$

since the Brownian motions are of almost surely continuous sample paths.

The asymptotic theory for the regressions (8)–(10) when the data are generated by (31) now can be easily deduced from Theorems 3.1 and 3.2 with these substitutions. The true parameter values for μ and θ are, of course, assumed to be zero. From the viewpoint of statistical inference on A_1, the regressions (8)–(10) for the model (31) are thus asymptotically equivalent to the VAR using differenced data. This is common in practice when the roots are believed to be unity. The asymptotic theory for stationary VAR's, which is well developed in the literature (e.g. Hannan [3]), may therefore be applied for inference on A_1.

When the lagged differences in (31) are uncorrelated with the regression errors, as is conventionally assumed in empirical VAR models, our covariance matrix estimates are consistent. The functional $g_R(B_0,M(B_2),E)$ in Theorem 4.1, which represents the limiting distributions of the G_1-statistics for regressions (8)–(10), reduces in this case to $g_R(B_0,M(JB_0),E)$. Also, the asymptotic distributions of the transformed H-statistics can be represented as $h(B_0,M(JB_0))$ with various functionals $M(\cdot)$ introduced in Part 1. It can now be easily shown for these $M(\cdot)$ that

$$h(B_0,M(JB_0)) \equiv h(W,M(W)) \tag{32}$$

where W denotes n-vector standard Brownian motion with covariance matrix I_n. (32) is parallel to the result of Lemma 5.5 in Part 1, and implies that the tests for unit roots and other tests, which may include the constant term and the coefficient of the time trend, in (31) are asymptotically equivalent to the corresponding tests in the first order VAR system with unit roots considered in Part 1. The condition $\Sigma_{10} = 0$ imposed here, is often violated. It may also be worth noting that this condition seems likely to imply $\Delta_{21} = \Delta_{20} = 0$ in many cases of practical interest (although not necessarily). To the extent that these conditions are satisfied, however, the motivation for the transformations to construct the H-statistics decreases accordingly.

Secondly, we consider VAR's with double unit roots which we transform as:

$$y_t = A_{11}\Delta^2 y_{t-1} + \ldots + A_{1p}\Delta^2 y_{t-p} + A_2\Delta y_{t-1} + A_3 y_{t-2} + u_t \tag{33}$$

where $A_2 = A_3 = I$ and the same condition as in (31) is satisfied for A_{1j}, $j = 1,\ldots,p$. By appropriately defining x_{1t}, x_{2t}, and x_{3t}, we suppose the conditions assumed for (2) hold for the above model. The univariate version of (33) with independent errors was studied by Hasza and Fuller [4].

It follows from Theorem 3.1 that statistical inference on A_1 in (33) based upon regressions (11)–(13) (again with $\mu = \theta = 0$) is asymptotically equivalent to that based on the stationary VAR using second differenced data. If the second differenced stationary component in (33) is either uncorrelated with the innovation sequence $\{u_t\}$, or is excluded from the regression, the hypothesis of unit roots can be tested using our H_2-statistics. The limiting distributions of these statistics can be easily obtained from Theorem 4.3 by substituting $B_3 = B_2 = JB_0$, where $J = A_1(1)^{-1}$ as defined earlier for (31). We also have

LEMMA 5.5. *For the various functionals $N(\cdot,\cdot)$ introduced in Theorem 3.3,*

$$h(B_0, N(JB_0, JB_0)) \equiv h(W, N(W, W))$$

where W denotes n-vector standard Brownian motion (i.e., Brownian motion with covariance matrix $= I_n$).

Lemma 5.5 implies that the null distributions of the H_2-statistics for the unit root tests and other tests including the constant term and the time trend coefficients are asymptotically invariant within a wide class of $\{u_t\}$, which is allowed to be weakly dependent and possibly heterogeneous. No transformation is needed if the innovations are martingale differences. When $n = 1$, the limiting distributions of some of the H_2-statistics, viz. $H_2(\hat{A}_2, \hat{A}_3)$, $H_2(\bar{A}_2, \bar{A}_3)$, $H_2(\bar{\mu}, \bar{A}_2, \bar{A}_3)$, $H_2(\tilde{A}_2, \tilde{A}_3)$, and $H_2(\tilde{\mu}, \tilde{\theta}, \tilde{A}_2, \tilde{A}_3)$, are tabulated in Hasza and Fuller [4]. Similarly, as in the case of single unit roots, spurious demeaning, and detrending seem to yield statistics whose limiting distributions have thicker tails. The direction of the bias in size resulting from the tests based nominally on chi-square tables, however, is not certain. Interestingly, their table shows that all the statistics considered there, except $H_2(\tilde{A}_2, \tilde{A}_3)$ would lead us to under-rejection of the null hypotheses if the decision were based on the chi-square table. This is in sharp contrast to the simple test of a single unit root.

For a VAR with unit roots but not transformed as in (31) or (33), the results of Section 5.2 apply and correspond to those given above for the appropriately transformed models. From the practical point of view, the asymptotic normality given in Theorem 5.3 becomes more important as we have more lags and less unit roots, because then the limiting covariance matrix, roughly speaking, becomes less singular and the overall asymptotic behavior of $\hat{A}$ is mainly determined by the contribution from the stationary part. In the context of the VAR system considered here, the matrix C defined in Section 5.2 takes the form $I_n \otimes C_2$. for the p-th order VAR with simple unit roots, for example, C_2 is given by

$$C_2 = \begin{bmatrix} 1 & & \\ -1 & 1 & 0 \\ & \ddots & \ddots \end{bmatrix}$$

and therefore Theorem 5.3(b) holds for any matrix R except for R of the form $R_1 \otimes \iota_p$ where ι_p is as before the p-vector of ones. Similarly, where there are double unit roots, $R(C_1)^{\perp}$ is two dimensional and is spanned by ι_p and τ_p, where $\tau_p = (1,2,\ldots,p)'$.

We have assumed so far that Ω_0 is positive definite and $\{y_t\}$ is itself not cointegrated. If this is not true and there exists a matrix C_1 such that $\{C_1'y_t\}$ is stationary, then each column of $C_1 \otimes I_p$ is a cointegrating vector for the p-th order VAR. Combining this with the above results, one can easily deduce the theory for the general VAR system. Finally notice that if the innovation sequence $\{u_t\}$ is i.i.d. or a martingale difference sequence as is in the standard VAR system, Proposition 5.4 implies that the usual Wald statistic has a limiting chi-square distribution as long as the restriction matrix R satisfies the condition given in Theorem 5.3. Our results here therefore give a rather complete answer to Sims's [11] problem concerning asymptotic normality of the regression coefficients in a VAR with unit roots, which has recently been explored in Sims, Stock, and Watson [12].

6. CONCLUSIONS

The central purpose of this paper and its companion, Part 1, has been to achieve a simple and unifying asymptotic analysis of multivariate regressions with integrated processes. The framework of analysis we have developed is a general one but it has a common architecture that helps to simplify and codify what would otherwise be a myriad of isolated results. The models we have looked at in detail show well the scope of the underlying theory. They have been deliberately selected with two objectives in mind: (i) to help provide an overall picture of how additional complications, such as multiple lags, cointegrated regressors, and regressors with drift, may be comfortably accommodated within the theory; and (ii) to provide explicit results for models of obvious empirical relevance such as VAR's with some unit roots and some cointegrated variates. It is our hope that the framework we have developed will provide a useful conduit for further research in this field as well.

REFERENCES

1. Engle, R.F. & C.W.J. Granger. Cointegration and error correction: Representation, estimation and testing. *Econometica* 55 (1987): 251–276.
2. Fuller, W.A., D.P. Hasza, & J.J. Goebel. Estimation of parameters of stochastic difference equations. *Annals of Statistics* 9 (1981): 531–543.
3. Hannan, E.J. *Multiple Time Series.* New York: John Wiley, 1970.
4. Hasza, D.P. & W.A. Fuller. Estimation of autoregressive processes with unit roots. *Annals of Statistics* 7 (1979): 1106–1120.
5. McLeish, D.L. A maximal inequality and dependent strong laws. *Annals of Probability* 3 (1975): 829–839.
6. Nankervis, J.C. & N.E. Savin. Finite sample distributions of t and F statistics in an AR(1) model with an exogenous variable. Mimeographed, University of Iowa, 1986.

7. Park, J.Y. & P.C.B. Phillips. Statistical inference in regressions with integrated processes: Part 1. Cowles Foundation Discussion Paper No. 811, Yale University, 1986.
8. Phillips, P.C.B. Understanding spurious regressions in econometrics. *Journal of Econometrics* 33 (1986): 311-340.
9. Phillips, P.C.B. & S.N. Durlauf. Multiple time series regression with integrated processes. *Review of Economic Studies* 53 (1986): 473-496.
10. Phillips, P.C.B. & S. Ouliaris. Testing for cointegration using principal components methods. Cowles Foundation Discussion Paper No. 809R, Yale University, 1986/1987.
11. Sims, C.A. Least squares estimation of autoregressions with some unit roots. CERDE Discussion Paper No. 78-95, University of Minnesota, 1978.
12. Sims, C.A., J.H. Stock, & M.W. Watson. Inference in linear time series models with some unit roots. Mimeographed, Minnesota University, 1986.
13. White, H. *Asymptotic Theory for Econometricians.* New York: Academic Press, 1984.

MATHEMATICAL APPENDIX

1. Proof of Lemma 2.1. The initial values of $\{x_t\}$ do not affect our asymptotic results and are set to be zero with no loss of generality. To prove (a)(i), (b)(i), and (c)(i)–(ii), we first notice for $r \in [(t-1)/T, t/T]$ that

$$T^{-3/2}x_{3t} = \int_0^r X_{3T}(s)ds + T^{-3/2}S_{3t} + T^{-1/2}\left(\frac{t}{T} - r\right)S_{3(t-1)} \tag{A1}$$

using an obvious subscript notation for components of X_T and S_t. It follows from (A1) that

$$T^{-5/2}x_{3t} = \int_{(t-1)/T}^{t/T}\int_0^r X_{3T}(s)dsdr + T^{-5/2}S_{3t} + \frac{1}{2}T^{-5/2}S_{3(t-1)}$$

for all $t = 1, \ldots, T$ and, since $\Sigma_1^T S_t = O_p(T^{3/2})$

$$\begin{aligned} T^{-5/2}\Sigma_1^T x_{3t} &= \int_0^1\int_0^r X_{3T}(s)dsdr + o_p(1) \\ &\Rightarrow \int_0^1\int_0^r B_3(s)dsdr = \int_0^1 \bar{B}_3 \end{aligned}$$

by the continuous mapping theorem. This proves (a)(i). By the same token (A1) yields:

$$\begin{aligned} T^{-5/2}tx_{3t} &= r\int_0^r X_{3T}(s)ds + T^{-3/2}\left(\frac{t}{T} - r\right)\Sigma_1^{t-1}S_{3j} - T^{-1/2}\left(\frac{t}{T} - r\right)^2 S_{3(t-1)} \\ &\quad + T^{-5/2}tS_{3t} + T^{-3/2}\left(\frac{t}{T} - r\right)tS_{3(t-1)} \end{aligned}$$

and

$$T^{-7/2}tx_{3t} = \int_{(t-1)/T}^{t/T} r\int_0^r X_{3T}(s)dsdr + \frac{1}{2}T^{-7/2}\Sigma_1^{t-1}S_{3j}$$
$$- \frac{1}{3}T^{-7/2}S_{3(t-1)} + T^{-7/2}tS_{3t} + \frac{1}{2}T^{-7/2}tS_{3(t-1)}.$$

Notice (a)(i) and that $\Sigma_1^T tS_t = O_p(T^{5/2})$ to get

$$T^{-7/2}\Sigma_1^T tx_{3t} = \int_0^1 r\int_0^r X_{3T}(s)dsdr + o_p(1)$$
$$\Rightarrow \int_0^1 r\bar{B}_3$$

as stated in (b)(i). The proof of (c)(i) is also immediate from (A1), i.e.,

$$T^{-4}\Sigma_1^T x_{3t}x_{3t}' = \int_0^1\left[\int_0^r X_{3T}(s)ds\right]\left[\int_0^r X_{3T}(s)ds\right]'dr$$
$$+ T^{-4}\Sigma_1^T[(\Sigma_1^{t-1}S_{3j})S_{3t}' + S_{3t}(\Sigma_1^{t-1}S_{3j})']$$
$$+ \frac{1}{2}T^{-4}\Sigma_1^T[(\Sigma_1^{t-1}S_{3j})S_{3(t-1)}' + S_{3(t-1)}(\Sigma_1^{t-1}S_{3j})']$$
$$- \frac{1}{3}T^{-4}\Sigma_1^T S_{3(t-1)}S_{3(t-1)}'$$
$$= \int_0^1\left[\int_0^r X_{3T}(s)ds\right]\left[\int_0^r X_{3T}(s)ds\right]'dr + o_p(1)$$
$$\Rightarrow \int_0^1 \bar{B}_3\bar{B}_3'.$$

Here we use the fact that $\Sigma_1^T(\Sigma_1^{t-1}S_j)S_t' = O_p(T^3)$. This can be easily seen from (c)(ii), which will be proved next. We have

$$T^{-1/2}x_{2t} = X_{2T}(r) + T^{-1/2}v_{2t}$$

for $r \in [(t-1)/T, t/T)$ and therefore,

$$T^{-3}\Sigma_1^T x_{3t}x_{2t}' = \int_0^1\left[\left(\int_0^r X_{3T}(s)ds\right)X_{2T}(r)'\right]dr$$
$$+ T^{-3}\Sigma_1^T(\Sigma_1^{t-1}S_{3j})v_{2t}' + \frac{1}{2}T^{-3}\Sigma_1^T S_{3(t-1)}S_{2(t-1)}' + T^{-3}\Sigma_1^T S_{3t}S_{2t}'$$
$$= \int_0^1\left[\left(\int_0^r X_{3T}(s)ds\right)X_{2T}(r)'\right]dr + o_p(1)$$

$$\Rightarrow \int_0^1 \bar{B}_3 B_2'$$

since $\Sigma_1^T(\Sigma_1^{t-1}S_j)v_t' = O_p(T^2)$ as is easily seen again from what follows. For (c)(iv), we have

$$\begin{aligned}
T^{-2}\Sigma_1^T x_{3t}x_{1t}' &= T^{-2}\Sigma_1^T(\Sigma_1^{t-1}S_{3j})v_{1t}' + T^{-2}\Sigma_1^T S_{3t}v_{1t}' \\
&= T^{-2}\Sigma_1^T S_{3t}(S_{1T} - S_{1t})' + o_p(1) \\
&= (T^{-3/2}\Sigma_1^T S_{3t})(T^{-1/2}S_{1T})' - T^{-2}\Sigma_1^T S_{3t}S_{1t}' + o_p(1) \\
&\Rightarrow \left[\int_0^1 B_3(r)dr\right]B_1(1)' - \int_0^1 B_3(r)B_2(r)'dr \\
&\equiv \int_0^1 \bar{B}_3 dB_1'
\end{aligned}$$

by integration by parts. The proof of (c)(v) is entirely analogous and is omitted. For the remaining results, see Part 1.

2. Proof of Theorem 3.1. In the regression equation (8), we have

$$\begin{aligned}
\hat{A}_1 &= A_1 + U'(I - P_{X_2})X_1[X_1'(I - P_{X_2})X_1]^{-1} \\
&= A_1 + \left(\frac{U'X_1}{T}\right)\left(\frac{X_1'X_1}{T}\right)^{-1} + O_p(T^{-1}) \\
&\xrightarrow{P} A_1 + \Sigma_{10}'\Sigma_1^{-1}
\end{aligned}$$

and given (15),

$$\begin{aligned}
\hat{A}_1 - A_1^* &= \left(\frac{U'X_1}{T}\right)\left(\frac{X_1'X_1}{T}\right)^{-1} - \Sigma_{10}'\Sigma_1^{-1} + O_p(T^{-1}) \\
&= \left(\frac{U'X_1}{T} - \Sigma_{10}'\right)\left(\frac{X_1'X_1}{T}\right)^{-1} - \Sigma_{10}'\left(\frac{X_1'X_1}{T}\right)^{-1}\left(\frac{X_1'X_1}{T} - \Sigma_1\right)\Sigma_1^{-1} \\
&\quad + O_p(T^{-1}) \\
&= \left(\frac{U'X_1}{T} - \Sigma_{10}'\right)\Sigma_1^{-1} - \Sigma_{10}'\Sigma_1^{-1}\left(\frac{X_1'X_1}{T} - \Sigma_1\right)\Sigma_1^{-1} + o_p(T^{-1/2}).
\end{aligned}$$

It follows from (15) that

$$\begin{aligned}
T^{-1/2}(\hat{A}_1 - A_1^*) &= J_1(T^{-1/2}\Sigma_1^T(w_t w_t' - \Sigma))J_2' + o_p(1) \\
&\Rightarrow (J_1 \otimes J_2)N(0, V^0) \equiv N(0, V)
\end{aligned}$$

as required. The results for other regressions can be deduced similarly.

3. Proof of Theorem 3.2. Write the least squares regression equations (8)–(10) as

$$\tilde{Y}' = \hat{A}_2\tilde{X}_2' + \hat{U}'$$

$$\tilde{Y}' = \bar{\mu}\tilde{\iota}' + \bar{A}_2\tilde{X}_2' + \bar{U}'$$

$$\tilde{Y}' = \tilde{\mu}\tilde{\iota}' + \tilde{\theta}\tilde{\tau}' + \tilde{A}_2\tilde{X}_2' + \tilde{U}'$$

where $\iota = (1,1,\dots,1)'$, $\tau = (1,2,\dots,T)'$, and $\tilde{\iota} = (I - P_{X_1})\iota$, $\tilde{\tau} = (I - P_{X_1})\tau$, $\tilde{X}_2 = (I - P_{X_1})X_2$, $\tilde{Y} = (I - P_{X_1})Y$. We have from Lemma 2.1

$$\frac{U'\tilde{\iota}}{T} \Rightarrow \int_0^1 dP, \quad \frac{\tilde{U}'\tau}{T^{3/2}} \Rightarrow \int_0^1 r dP, \quad \frac{U'\tilde{X}_2}{T} \Rightarrow \int_0^1 dPB_2' + \Pi'. \tag{A2}$$

Moreover, it is easy to see that

$$\frac{\tilde{\iota}'\tilde{\iota}}{T} = 1 + O_p\left(\frac{1}{T}\right), \quad \frac{\tilde{\tau}'\tilde{\tau}}{T^3} = \frac{\tau'\tau}{T^3} + O_p\left(\frac{1}{T}\right), \quad \frac{\tilde{X}_2'\tilde{X}_2}{T^3} = \frac{X_2'X_2}{T^3} + O_p\left(\frac{1}{T}\right),$$

$$\frac{\tilde{\iota}'\tilde{\tau}}{T^2} = \frac{\iota'\tau}{T^2} + O_p\left(\frac{1}{T}\right), \quad \frac{\tilde{\iota}'\tilde{X}_2}{T^{3/2}} = \frac{\iota'X_2}{T^{3/2}} + O_p\left(\frac{1}{T}\right), \quad \frac{\tilde{\tau}'\tilde{X}_2}{T^{5/2}} = \frac{\tau'X_2}{T^{5/2}} + O_p\left(\frac{1}{T}\right). \tag{A3}$$

Now the stated result follows easily from the proofs of Theorem 3.1–Theorem 3.3 in Part 1.

4. Proof of Theorem 3.3. In notation similar to that above, we rewrite regressions (11)–(13) as

$$\tilde{Y}' = \hat{A}_2\tilde{X}_2' + \hat{A}_3\tilde{X}_3' + \hat{U}'$$

$$\tilde{Y}' = \bar{\mu}\tilde{\iota}' + \bar{A}_2\tilde{X}_2' + \bar{A}_3\tilde{X}_3' + \bar{U}'$$

$$\tilde{Y}' = \tilde{\mu}\tilde{\iota}' + \tilde{\theta}\tilde{\tau}' + \tilde{A}_2\tilde{X}_2' + \tilde{A}_3\tilde{X}_3' + \tilde{U}'.$$

Notice that

$$\frac{U'\tilde{X}_3}{T^2} \Rightarrow \int_0^1 dPB_3' \tag{A4}$$

and

$$\frac{\tilde{\iota}'\tilde{X}_3}{T^{3/2}} = \frac{\iota'X_3}{T^{5/2}} + O_p\left(\frac{1}{T}\right), \quad \frac{\tilde{\tau}'\tilde{X}_3}{T^{7/2}} = \frac{\tau'X_3}{T^{7/2}} + O_p\left(\frac{1}{T}\right), \quad \frac{\tilde{X}_2'\tilde{X}_3}{T^3} = \frac{X_2'X_3}{T^3} + O_p\left(\frac{1}{T}\right),$$

$$\frac{\tilde{X}_3'\tilde{X}_3}{T^4} = \frac{X_3'X_3}{T^4} + O_p\left(\frac{1}{T}\right). \tag{A5}$$

The stated results are immediate from (A2)–(A5) and Lemma 2.1. Notice also that

$$\int_0^1 R'\left(\int_0^1 RR'\right)^{-1} R = \int_0^1 \bar{B}_3'\left(\int_0^1 \bar{B}_3\bar{B}_3'\right)^{-1}\bar{B}_3 + \int_0^1 Q_1'\left(\int_0^1 Q_1Q_1'\right)^{-1}Q_1$$

because $\int_0^1 \bar{B}_3 Q_1' = 0$ a.s.

5. Proof of Theorem 3.4. It can be easily shown that in regressions (8)–(10) or in (11)–(12)

$$\frac{\hat{U}'\hat{U}}{T},\frac{\bar{U}'\bar{U}}{T},\frac{\tilde{U}'\tilde{U}}{T} = \frac{U'U}{T} - \left(\frac{U'X_1}{T}\right)\left(\frac{X_1'X_1}{T}\right)^{-1}\left(\frac{X_1'U}{T}\right) + O_p\left(\frac{1}{T}\right)$$

$$\xrightarrow{P} \Sigma_0 - \Sigma_{10}'\Sigma_1^{-1}\Sigma_{10}.$$

Moreover, if $\Sigma_{10} = 0$ and (15) holds, then $T^{-1/2}U'X_1 = O_p(1)$ and

$$\frac{\hat{U}'\hat{U}}{T},\frac{\bar{U}'\bar{U}}{T},\frac{\tilde{U}'\tilde{U}}{T} = \frac{U'U}{T} + O_p\left(\frac{1}{T}\right)$$

from which the stated results follow immediately.

6. Proof of Theorem 4.1. If $\Sigma_{10} = 0$, it follows from Theorem 3.4(a) that $\hat{\Sigma}_0,\bar{\Sigma}_0$, $\tilde{\Sigma}_0 \xrightarrow{P} \Sigma_0$. Therefore, by Theorem 3.1 it suffices to that

$$M_{i1} \xrightarrow{P} \Sigma_1^{-1}, \quad i = 1,2,3$$

for part (a). Also, if $\Sigma_{10} = 0$, $P(r)$, and Π in (A2)–(A4) reduce, respectively, to $B_0(r)$ and Δ_{20}. Part (b) is now immediate in light of (A3). Finally, the limiting variate Z (say) in (15) depends on a quadratic function of the elements of w_t, whereas $B(r)$ depends on partial sums which are linear in w_t's. Hence Z and $B(r)$ are uncorrelated and, being Gaussian, are therefore independent. The independence, of course, carries over to any two statistics whose limiting distributions are, respectively, represented by functionals of Z and $B(r)$, as in our case. This proves part (c).

7. Proof of Theorem 4.2. The proofs of part (a) and (e) are entirely analogous, respectively, to those of parts (a) and (c) of Theorem 4.1. For (b)–(d), notice that

$$T^2M_{12} \Rightarrow \left(\int_0^1 Q_1Q_1'\right)^{-1}, \quad T^3M_{13} \Rightarrow \left(\int_0^1 Q_2Q_2'\right)^{-1}$$

$$T^2M_{22} \Rightarrow \left(\int_0^1 Q_1^*Q_1^{*\prime}\right)^{-1}, \quad T^3M_{23} \Rightarrow \left(\int_0^1 Q_2^*Q_2^{*\prime}\right)^{-1}$$

$$T^2M_{32} \Rightarrow \left(\int_0^1 Q_1^{**}Q_1^{**\prime}\right)^{-1}, \quad T^3M_{33} \Rightarrow \left(\int_0^1 Q_2^{**}Q_2^{**\prime}\right)^{-1}$$

$$Tm_{21} \Rightarrow \left(\int_0^1 \bar{P}_1^2\right)^{-1}, \quad Tm_{31} \Rightarrow \left(\int_0^1 \bar{P}_2^2\right)^{-1}, \quad T^3m_{32} \Rightarrow \left(\int_0^1 \bar{P}_3^2\right)^{-1}. \tag{A6}$$

The stated results are immediate from the above results, and those of Theorem 3.3. Recall that if $\Sigma_{10} = 0$, then Π reduces to Δ_{20}.

8. Proof of Theorem 4.3. We have

$$\tilde{\Gamma}_1 \Rightarrow \left(\int_0^1 \bar{B}_3B_2'\right)\left(\int_0^1 B_2B_2'\right)^{-1}\Delta_{20}$$

$$\tilde{\Gamma}_2 \Rightarrow \left(\int_0^1 \bar{B}_3^* B_2^{*\prime}\right)\left(\int_0^1 B_2^* B_2^{*\prime}\right)^{-1} \Delta_{20}$$

$$\tilde{\Gamma}_3 \Rightarrow \left(\int_0^1 \bar{B}_3^{**} B_2^{**\prime}\right)\left(\int_0^1 B_2^{**} B_2^{**\prime}\right)^{-1} \Delta_{20}.$$

Also,

$$\tilde{\gamma}_1 \Rightarrow -\int_0^1 Q_1'\left(\int_0^1 Q_1 Q_1'\right)^{-1} \Delta_{20}$$

$$\tilde{\gamma}_2 \Rightarrow -\int_0^1 Q_1^{+\prime}\left(\int_0^1 Q_1^+ Q_1^{+\prime}\right)^{-1} \Delta_{20}$$

$$\tilde{\gamma}_3 \Rightarrow -\int_0^1 sQ_1^{*\prime}\left(\int_0^1 Q_1^* Q_1^{*\prime}\right)^{-1} \Delta_{20}$$

and $RSS_1^{-1} = m_{21}$, $RSS_2^{-1} = m_{31}$, and $RSS_3^{-1} = m_{32}$. Now all the stated results for the statistics for testing single hypotheses follow immediately from the above results, (A6), Theorem 3.3, and Theorem 4.2(b)–(d). For the joint tests, apply Lemma A.1 in Part 1. Thus, for example, in the case of $R_2 = I$ and $R_3 = I$,

$$G_2(\hat{A}_2,\hat{A}_3) \Rightarrow g(B_0,Q_1,\Delta_{20}) + h(B_0,\bar{B}_3)$$

and the stated result for $H_2(\hat{A}_2,\hat{A}_3)$ follows easily from the fact that

$$a \Rightarrow h(B_0,Q_1) - g(B_0,Q_1,\Delta_{20}),$$

which may be deduced from the earlier result for $H_2(\hat{A}_2)$. The proofs for the other joint test statistics are analogous and are omitted.

9. Proof of Theorem 5.1. To facilitate the proofs, we define a functional k by

$$k(B,N) = \int_0^1 dBN'\left(\int_0^1 NN'\right)^{-1/2}$$

as in Part 1. It is easy to see from Theorem 3.3 and (A9)–(A12) that the limiting distributions of the given statistics can be represented in the generic form $k(B_0,N(B_2,B_3))$ where the function $N(\cdot,\cdot)$ of B_2 and B_3 is given for each estimator in Theorem 3.3. Since $\Omega_{20} = 0$ and $\Omega_{30} = 0$, and $B(r)$ is Gaussian, B_0 is independent of both B_2 and B_3, and, hence, any function of B_2 and B_3, including $N = N(B_2,B_3)$. All the stated results are now simple consequences of Lemma 5.1 in Part 1.

10. Proof of Theorem 5.2. When $\Sigma_{10} = 0$ and $\Delta_{20} = 0$, the limiting distributions of the G_2-statistics in Theorem 4.2 become of the form $h_R(B_0,N)$, where $h_R(\cdot,\cdot) = g_R(\cdot,\cdot,0)$ as in part 1, and N is a function of B_2 and B_3 appropriately defined for each test. If $\Omega_{20} = 0$ and $\Omega_{30} = 0$, then $N = N(B_2,B_3)$ is independent of B_0 and the results for the statistics for the single hypothesis follow easily from Corollary 5.3(a) in Part 1. For the tests of joint hypotheses among (19), (20), and (22), we assume momentarily that (19) and (22) are of simple form and that $R_2 = I$ and $R_3 = I$. Now, by virtue of Lemma A.1 of Part 1, all the G_2-statistics for the

joint tests can simply be written as sums of the statistics for the individual null hypotheses. Therefore, if $\Delta_{20} = 0$, we have for example

$$G_2(\hat{A}_2,\hat{A}_3) \Rightarrow h(B_0,Q_1) + h(B_0,\bar{B}_3).$$

Now notice that

$$\int_0^1 Q_1 \bar{B}_3' = 0 \text{ a.s.}$$

and apply Corollary 5.3(b) of Part 1 to get the desired result. The proofs for the other joint tests, and for the tests of subsets of parameters are entirely analogous. The latter case obviously extends our proof to the general case where R_2 and R_3 are arbitrary matrices of full row rank, which can be easily seen by appropriate transformation of the models where necessary. This proves part (a). For (b), we assume $R_1 = I$. It is easy to see that this causes no loss of generality. Then again by Lemma A.1 in Part 1, we can write, for example,

$$F_2(\hat{A}_1,\hat{A}_2,\hat{A}_3) = F_2(\hat{A}_1) + \tilde{F}_2(\hat{A}_2,\hat{A}_3)$$

where $\tilde{F}_2(\hat{A}_2,\hat{A}_3)$ is the Wald statistic for the joint test of (19) and (22) based on regressions with no stationary variables. We have, however, seen that $\tilde{F}_2(\hat{A}_2,\hat{A}_3)$ is asymptotically equivalent to $F_2(\hat{A}_2,\hat{A}_3)$, the Wald statistic constructed from regression (11). Now the stated result for $F_2(\hat{A}_1,\hat{A}_2,\hat{A}_3)$ follows from Theorem 4.2(a), (e), and part (a) above. The proofs for the other Wald statistics for the joint tests of (18)–(20) and (22) are analogous and are therefore omitted. Finally, notice that, under the given conditions, $\Omega = \Sigma$ and the F_2- and G_2-statistics have the same asymptotic distributions.

11. Proof of Theorem 5.3. To prove part (a), we first let $C = H_1$. Then we have

$$\begin{aligned}\sqrt{T}(\hat{A} - A^*) &= \sqrt{T}(\hat{A}_1 - A_1^*)H_1' + \frac{1}{\sqrt{T}}[T(\hat{A}_2 - A_2)H_2'] \\ &\quad + \frac{1}{T^{3/2}}[T^2(\hat{A}_3 - A_3)H_3'] \\ &= \sqrt{T}(\hat{A}_1 - A_1^*)H_1' + o_p(1).\end{aligned}$$

Notice that $A^*H = (A_1^*,0,0)$. Now the result follows directly from Theorem 3.1(b). If C is a matrix such that $R(C) = R(H_1)$, then there exists an invertible matrix L such that $H_1 = CL$. The stated result in (a) therefore follows easily by redefining V^0, Σ_1, and Σ_{10} appropriately. For part (b), simply note that, under the given conditions, $R(I \otimes C)$ is of full row rank, and so $R(I \otimes C)V(I \otimes C')R'$ is invertible.

12. Proof of Proposition 5.4. We assume w.l.o.g. that $C = H_1$. It is easy to show

$$\left(\frac{X'X}{T}\right)^{-1} = K_1\left(\frac{X_1'X_1}{T}\right)^{-1}K_1' + O_p\left(\frac{1}{T}\right)$$

where $K_1' = (\overset{m_1}{I}, \overset{m_2 + m_3}{0})$, which is immediate from Lemma 2.1 by inverting a partitioned matrix. Now, notice that

$$\begin{aligned}\left(\frac{Z'Z}{T}\right)^{-1} &= H\left(\frac{X'X}{T}\right)^{-1} H' \\ &= H_1\left\{H_1'\left(\frac{Z'Z}{T}\right)H_1\right\}^{-1} H_1' + O_p\left(\frac{1}{T}\right).\end{aligned}$$

to get the stated result.

13. Proof of Lemma 5.5. Write

$$h(B_0, N) = \operatorname{tr}\left\{\Omega_0^{-1/2} \int_0^1 dB_0 N' \left(\int_0^1 NN'\right)^{-1} \int_0^1 N dB_0' \Omega_0^{-1/2}\right\}$$

and notice that

$$\begin{aligned}&N(JB_0, JB_0)'\left(\int_0^1 N(JB_0, JB_0)N(JB_0, JB_0)'\right)^{-1} N(JB_0, JB_0) \\ &= N(B_0, B_0)'\left(\int_0^1 N(B_0, B_0)N(B_0, B_0)'\right)^{-1} N(B_0, B_0) \\ &= N(W, W)'\left(\int_0^1 N(W, W)N(W, W)'\right)^{-1} N(W, W)\end{aligned}$$

upon the transformation $B_0 \to \Omega_0^{-1/2} B_0 = W$ for the various functions $N(\cdot, \cdot)$ introduced in Theorem 3.3.