Cowles Foundation Paper #728

Econometric Theory, 5, 1989, 181–240. Printed in the United States of America

PARTIALLY IDENTIFIED ECONOMETRIC MODELS

P.C.B. PHILLIPS
Cowles Foundation for Research in Economics
Yale University

This paper studies a class of models where full identification is not necessarily assumed. We term such models partially identified. It is argued that partially identified systems are of practical importance since empirical investigators frequently proceed under conditions that are best described as apparent identification. One objective of the paper is to explore the properties of conventional statistical procedures in the context of identification failure. Our analysis concentrates on two major types of partially identified model: the classic simultaneous equations model under rank condition failures; and time series spurious regressions. Both types serve to illustrate the extensions that are needed to conventional asymptotic theory if the theory is to accommodate partially identified systems. In many of the cases studied, the limit distributions fall within the class of compound normal distributions. They are simply represented as covariance matrix or scalar mixtures of normals. This includes time series spurious regressions, where representations in terms of functionals of vector Brownian motion are more conventional in recent research following earlier work by the author [23]. These asymptotic results are covered by a limit theory that we describe as a limiting mixed Gaussian (LMG) family. Extensions of the LMG family are also explored. These are designed to embrace all of our asymptotic results. A new theory is put forward that is based on a limiting Gaussian functional (LGF) condition. This leads to the required extensions. It is distinguished from the LMG theory in several ways: first, in form, since it involves functionals of random elements on function spaces rather than functions of finite-dimensional random vectors; second, in generality, since it accommodates unit root limit theory as well as LMG; and third, in its implications, since it allows for a certain type of variable random information in the limit distribution when applied to maximum likelihood estimators. Some applications are discussed including the Gaussian AR(1) for stable, explosive, and unit root coefficients. The latter example illustrates well the need for a theory such as LGF.

1. INTRODUCTION

The subject of my talk today is partially identified models. This is the term that I shall use to describe models that are identified in some parts while

This paper is based on an invited address given to the Australasian Meetings of the Econometric Society, Christchurch, New Zealand, August 26–29, 1987. It was drafted in July 1987 and revised in August 1988 I am grateful to two referees and P Jeganathan for very helpful comments on the first version of the paper. My thanks also go to Glena Ames for her skill and effort in keyboarding the manuscript and to the National Science Foundation for support under Grant SES 8519595

 0266-4666/89 $5.00 + 00

being unidentified in others. This will include totally identified and totally unidentified systems, so that the term is rather comprehensive.

Identification, as we presently study it, is a mathematical property of a model and it is a subject that we tend to treat very much in isolation. Prior to estimation, it is conventional to assume that the model is fully identified or, at least, that we are working with estimable functions. Such assumptions may be explicit or implicit but they underpin all of the commonly used theories of inference. Asymptotic statistical theory, in particular, is almost invariably developed in this way. The approach is well illustrated by the modern theory of inference in nonlinear regressions. Here, the identifiability conditions are often rather strong. They are designed, with attendant regularity conditions, to ensure that the objective criterion converges almost surely and uniformly to a nonrandom function with a unique optimum at the true value or ultimate point of consistency. In some cases, this good behavior of the objective function is even directly assumed.

This approach has always seemed heroic and rather unjustified to me. It does go a long way towards simplifying asymptotic theory and inference, but in doing so it rules out many interesting possibilities. For example, in nonergodic models and totally unidentified models, the objective criterion does not behave in this way. Instead, the criterion, upon suitable standardization, converges weakly to a random function whose optimum may also be random. In such examples, a different and more general approach to the development of an asymptotic theory is needed.

These issues have a major bearing on empirical work. Here, investigators typically proceed under conditions that are best described as apparent identification. That is, estimation and testing goes ahead in practice as if the model were fully identified but with no certainty that this is true. As we might expect, the properties of the statistical procedures that we employ hinge crucially on whether the system is identified or not. When there is identification failure in a model, the properties of our estimators and tests undergo important changes. This is especially true of properties that rely on an asymptotic theory of inference. It seems important that we should understand the implications of identification failure for statistical inference. Yet, this is a subject that seems to be virtually untouched in the literature. The primary aim of the present paper is to investigate this rather neglected class of problem.

There are many situations where we might expect identification failures to arise in econometrics. We shall concentrate our discussion on two major types of partially identified model. These will serve to illustrate most of the problems that can occur and to provide guidelines for the development of a general theory. The first of these is the classic simultaneous equations model. In single equation estimation, identification failures can result in some coefficients being identified and others unidentified. Those that are identified may be regarded as asymptotically estimable functions. In systems estimation, some equations in the system may be identified while others are not; and those equations that are unidentified may contain some identifiable coefficients.

The second type of partially identified model with which we will be concerned is a time series spurious regression. This could be a single equation regression in which some components are spurious and others are not. Or it could be a system where some regression equations are spurious (possibly only partially spurious) and others are not. As our analysis will show, such time series spurious regressions have a close formal similarity with partially identified structural equations. The results we obtain for both types of model indicate the nature of the extensions that are needed to conventional asymptotics to accommodate partially identified systems.

A second aim of the paper is to develop such extensions and thereby tie together the asymptotic theory for unidentified, partially identified, and fully identified systems. As we shall show, some of the cases that we consider come within what we shall call a limiting mixed Gaussian (LMG) family. However, there are many exceptions. Prominent among these are certain time series regressions involving integrated processes and unit root autoregressions. A second level of generalization that is designed to accommodate this rather important class of exceptions to the LMG family will also be developed.

The plan of the paper is as follows. Section 2 deals with structural estimation in simultaneous systems. We first develop a finite-sample theory for partially identified structural equations and then show how a general asymptotic theory follows through the operation of an invariance principle. Properties of conventional statistical tests and estimators are considered in detail. Section 3 deals with time series regressions. This includes spurious regressions, partially spurious regressions, and cointegrating regressions. Representations of the limit distributions that go beyond functionals of Brownian motion are pursued and, in most cases, these turn out to be simple scale mixtures of normals. A close formal similarity between spurious regressions and totally unidentified structural equations is discovered. Section 4 explores the LMG theory and the extensions that are necessary to embrace all of our earlier asymptotic results. Some conclusions are drawn and some topics for further research are discussed in Section 5. Proofs are given in the Appendix.

A word on notation. We use the symbol "$\Rightarrow$" to signify weak convergence, the symbol "$\equiv$" to signify equality in distribution, and the inequality "> 0" to signify positive definite when applied to matrices. Stochastic processes such as the Brownian motion $B(r)$ on $[0,1]$ are frequently written as B to achieve notational economy. Similarly, we write integrals with respect to Lebesgue measure such as $\int_0^1 B(s)ds$ more simply as $\int_0^1 B$. Vector Brownian motion with covariance matrix Ω is written "$BM(\Omega)$." We use $O(n)$ to denote the orthogonal group of $n \times n$ matrices, $V_{k,n}$ to denote the Stiefel manifold $\{H_1(n \times k) : H_1'H_1 = I_k\}$, $U(V_{k,n})$ to signify the uniform distribution on $V_{k,n}$, and the abbreviation "a.s." for almost surely. We use $r(\Pi)$ to signify the rank of the matrix Π, P_Π to signify the orthogonal projection onto the range space of Π (with $Q_\Pi = I - P_\Pi$), and $\|\Pi\|$ to signify the Euclidean norm $\{\mathrm{tr}(\Pi'\Pi)\}^{1/2}$ of the matrix Π. Finally, $I(1)$ is used to denote an integrated process of order one and $I(0)$ denotes stationarity.

2. STRUCTURAL ESTIMATION

2.1 Partially Identified Structural Equations

We will work with the structural equation

$$y_1 = Y_2\beta + Z_1\gamma + u = W\delta + u, \tag{1}$$

where $y_1(T \times 1)$ and $Y_2(T \times n)$ contain observations of $n + 1$ endogenous variables, $Z_1(T \times k_1)$ is an observation matrix of k_1 included exogenous variables, and u is a random disturbance vector. The reduced form of (1) is written in partitioned format as

$$[y_1, Y_2] = [Z_1, Z_2]\begin{bmatrix} \pi_1 & \Pi_1 \\ \pi_2 & \Pi_2 \end{bmatrix} + [v_1, V_2], \tag{2}$$

or

$$Y = Z\Pi + V,$$

where Z_2 is a $T \times k_2$ matrix of exogenous variables excluded from (1). It is assumed that $k_2 \geq n$ so that (1) is "apparently" identified by order conditions and that Z is of full column rank $k = k_1 + k_2$. We also assume that (2) is in canonical form (see [18] for details of the necessary transformations) so that the rows of V are independent and identically distributed (i.i.d.) $(0, I_m)$, $m = n + 1$. Use of the canonical form helps to simplify subsequent arguments, involves no loss of generality [18], and is a convenient route to results for the unstandardized model. For the development of a finite-sample theory, we will also require the more specific:

(C1) $\quad V \equiv N_{T,m}(0, I)$,

which will be removed later when we study asymptotic behavior. However, as we shall see, many elements of the finite-sample distribution theory persist in infinite samples and, moreover, retain their validity under general conditions when the normality assumption is removed. This is an instance of the operation of the invariance principle (see, for example, [3, p. 72]) and is an interesting and important feature of partially identified systems.

To complete the specification of (2) and to facilitate the development of an asymptotic theory, we assume the conventional condition:

(C2) $\quad T^{-1}Z'Z \to M > 0$

as $T \to \infty$. We partition M conformably with Z as

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}.$$

The identifying relations connecting the parameters of (1) and (2) are

$$\pi_1 - \Pi_1 \beta = \gamma, \tag{3}$$

$$\pi_2 - \Pi_2 \beta = 0. \tag{4}$$

We know that (1) is identified if and only if $r(\Pi_2) = n \le k_2$. We call this the fully identified case. The polar opposite occurs when

$$\Pi_2 = 0 \tag{5}$$

and $r(\Pi_2) = 0$. This is often called the leading case in econometric distribution theory [17,19]. Note that the parameter vector β is totally unidentified. Interestingly, however, the structural equation (1) is not totally unidentified even in this case, provided that $k_1 > 0$. As is clear from (3) when $\Pi_1 = 0$, for example, the entire coefficient vector $\gamma = \pi_1$ is identified and equals a subset of the reduced form coefficients.

Suppose $\Pi_1 \ne 0$ and $r(\Pi_1) = k_{12}$. Define an orthogonal matrix

$$R = [\overset{k_{11}}{R_1}, \quad \overset{k_{12}}{R_2}] \in O(k_1), \tag{6}$$

where R_1 spans the null space of Π_1', $k_1 = k_{11} + k_{12}$, and $O(\ )$ denotes the orthogonal group. We now use R to rotate the coordinate system in the space of the included exogenous variables in (1). Under this rotation we obtain:

$$y_1 = Y_2\beta + Z_1 RR'\gamma + u,$$

or

$$y_1 = Y_2\beta + Z_{11}\gamma_1 + Z_{12}\gamma_2 + u, \tag{7}$$

where $Z_{11} = Z_1 R_1$, $Z_{12} = Z_1 R_2$, and the new coefficients are given by

$$\gamma_1 = R_1'\gamma = R_1'\pi_1, \tag{8}$$

$$\gamma_2 = R_2'\gamma = R_2'\pi_1 - R_2'\Pi_1\beta. \tag{9}$$

In the new coordinate system γ_1 is identified and γ_2 is totally unidentified (since $R_2'\Pi_1$ has full row rank). γ_1 may be regarded as a vector of asymptotically estimable functions of the coefficients in (1). In terms of the former coefficients, we have

$$\gamma = R_1\gamma_1 + R_2\gamma_2. \tag{10}$$

In the general case where Π_2 and Π_1 are of arbitrary rank, we may rotate coordinates in both the space of endogenous variables Y_2 and the space of exogenous variables Z_1 to isolate estimable functions. Suppose $r(\Pi_2) = n_1$ and define

$$S = \overset{\displaystyle n_1}{[S_1,} \quad \overset{\displaystyle n_2}{S_2]} \in O(n), \tag{11}$$

where S_2 spans the null space of Π_2, $n = n_1 + n_2$, and $\Pi_{21} = \Pi_2 S_1$ has full rank n_1. Let

$$\beta_1 = S_1'\beta, \qquad \beta_2 = S_2'\beta,$$

and then (4) becomes

$$\pi_2 - \Pi_2 SS'\beta = \pi_2 - \Pi_{21}\beta_1 = 0$$

in the new coordinates, so that β_1 is identifiable. Similarly, under this rotation, we have

$$\Pi_1\beta = \Pi_1 S_1\beta_1 + \Pi_1 S_2\beta_2 = \Pi_{11}\beta_1 + \Pi_{12}\beta_2 \text{ (say).}$$

Now define

$$R = \overset{\displaystyle k_{11}}{[R_1,} \quad \overset{\displaystyle k_{12}}{R_2]} \in O(k_1), \tag{12}$$

where R_1 spans the null space of Π_{12}' and let

$$\gamma_1 = R_1'\gamma = R_1'\pi_1 - R_1'\Pi_{11}\beta_1,$$

$$\gamma_2 = R_2'\gamma = R_2'\pi_1 - R_2'\Pi_{11}\beta_1 - R_2'\Pi_{12}\beta_2.$$

Under the simultaneous action of (11) and (12), the structural equation (1) becomes

$$y_1 = Y_2SS'\beta + Z_1RR'\gamma + u,$$

or

$$y_1 = Y_{21}\beta_1 + Y_{22}\beta_2 + Z_{11}\gamma_1 + Z_{12}\gamma_2 + u. \tag{13}$$

In (13), the coefficients (β_1,γ_1) are identified and (β_2,γ_2) are totally unidentified. The original coefficients are recovered from the equations:

$$\beta = S_1\beta_1 + S_2\beta_2,$$

$$\gamma = R_1\gamma_1 + R_2\gamma_2.$$

Equation (13) represents the estimable function format of a general partially identified structural equation. Systems of equations involve no new difficulties and comprise a set of structural equations, each of which falls into the general format of (13) upon appropriate transformation of coordinates. However, it is important to observe that in general there will be no single rotation of the coordinate system which will transform the model into a system in which each equation is in estimable function format as in (13).

2.2 Distribution Theory

To fix ideas, it will be convenient to work with the structural equation (1) in the partially identified case (5). As we have seen, this equation may be written in the estimable function format (7), where $\gamma_1 = R_1'\gamma$ is identified and $(\beta, \gamma_2 = R_2'\gamma)$ is totally unidentified. The main ideas are then well illustrated by considering the instrumental variables estimator:

$$\hat{\delta} = \operatorname{argmin}_\delta (y - W\delta)' P_H (y - W\delta),$$

where $H = [Z_1, Z_3]$ is a $T \times (k_1 + k_3)$ matrix of instruments with Z_3 a submatrix of Z_2 formed by column selection. We require $k_3 \geq n$, so that the order condition of sufficient instruments is satisfied. Define $k_* = k_1 + k_3$.

Subvector coefficient estimates are given by:

$$\hat{\beta} = [Y_2'(P_H - P_{Z_1})Y_2]^{-1}[Y_2'(P_H - P_{Z_1})y_1], \tag{14}$$

$$\hat{\gamma} = (Z_1'Z_1)^{-1}Z_1'y_1 - (Z_1'Z_1)^{-1}Z_1'Y_2\hat{\beta}, \tag{15}$$

with

$$\hat{\gamma}_1 = R_1'\hat{\gamma}, \qquad \hat{\gamma}_2 = R_2'\hat{\gamma}.$$

We now have the following theorem.

THEOREM 2.1. *Under (C1)*

(a)
$$\hat{\beta} \equiv \int_{S>0} N(0, S^{-1}) pdf(S) dS, \qquad S \equiv W_n(k_3, I), \tag{16}$$

$$\equiv \int_{z>0} N(0, zI) pdf(z) dz, \qquad 1/z \equiv \chi_q^2, \qquad q = k_3 - n + 1, \tag{17}$$

$$\equiv q^{-1/2} t_q, \tag{18}$$

$$= r \text{ (say)},$$

where t_q denotes an n-vector multivariate t distribution with q degrees of freedom. The density of $\hat{\beta}$ is

$$pdf(r) = \frac{c}{(1 + r'r)^{(q+n)/2}}, \tag{19}$$

where

$$c = \Gamma((q+n)/2)/\Gamma(q/2)\pi^{n/2}. \tag{20}$$

(b) $$\hat{\gamma}_1 \equiv \int_{R^n} N(\gamma_1, (1+r'r)G_1)pdf(r)dr, \tag{21}$$

$$\equiv \int_{m>0} N(\gamma_1, (1+m)G_1)pdf(m)dm, \tag{22}$$

$$= s_1 \text{ (say)},$$

where

$$G_1 = R_1'(Z_1'Z_1)^{-1}R_1,$$

$$m \equiv B'\left(\frac{n}{2}, \frac{k_3-n+1}{2}\right),$$

and B' denotes a beta-prime distribution with the stated degrees of freedom, i.e.,

$$pdf(m) = \left[B\left(\frac{n}{2}, \frac{k_3-n+1}{2}\right)\right]^{-1} m^{n/2-1}(1+m)^{-(k_3+1)/2}. \tag{23}$$

(c) $$\hat{\gamma}_2 \equiv \int_{R^n} N(R_2'\pi_1 - R_2'\Pi_1 r, (1+r'r)G_2)pdf(r)dr, \tag{24}$$

$$= s_2 \text{ (say)},$$

where

$$G_2 = R_2'(Z_1'Z_1)^{-1}R_2.$$

(d) $$\hat{\gamma} \equiv R_1 s_1 + R_2 s_2,$$

$$= s \text{ (say)}.$$

COROLLARY 2.2. *Under (C1) and (C2)*

(a) $$\hat{\beta} \Rightarrow r \equiv \int_{z>0} N(0, zI)pdf(z)dz, \tag{25}$$

(b) $$\sqrt{T}(\hat{\gamma}_1 - \gamma_1) \Rightarrow \int_{m>0} N(0, (m+1)\bar{G}_1)pdf(m)dm, \tag{26}$$

where

$\bar{G}_1 = R_1' M_{11}^{-1} R_1.$

$$\text{(c)} \qquad \hat{\gamma}_2 \Rightarrow R_2'\pi_1 - R_2'\Pi_1 r \equiv \int_{z>0} N(R_2'\pi_1, zR_2'\Pi_1\Pi_1'R_2)\, pdf(z)dz, \tag{27}$$

$$\text{(d)} \qquad \hat{\gamma} \Rightarrow \int_{z>0} N(\bar{\gamma}, zP_2\Pi_1\Pi_1'P_2)\, pdf(z)dz, \tag{28}$$

$$= \bar{s} \text{ (say)},$$

where

$\bar{\gamma} = R_1\gamma_1 + P_2\pi_1,$

$P_2 = R_2R_2',$

and the scale variates z and m are as in Theorem 2.1.

Remarks. (i) Note that the density of $\hat{\beta}$ given by (19) is independent of β. The fact that this distribution carries no information about β is consonant with and, indeed, is a consequence of the fact that β is totally unidentified.

(ii) Note also that the density (19) is independent of T and is the limiting distribution of $\hat{\beta}$, as indicated in (25). Thus, the distribution of $\hat{\beta}$ is the same in finite and infinite samples. This distributional invariance to T and the nondegeneracy of the limiting distribution are manifestations of the uncertainty about β that is implicit in its lack of identification. The phenomenon is further discussed in earlier work [17,19,21] where (19) and (25) were first derived.

(iii) $\hat{\gamma}_1$ is consistent for γ_1, the identified exogenous variable coefficients. In finite samples, $\hat{\gamma}_1$ is distributed about γ_1 as a variance mixture of normals, represented by (22). This mixture distribution persists in the limit. Indeed, the limiting distribution of $\sqrt{T}(\hat{\gamma}_1 - \gamma_1)$ is a variance mixture of normals with the same mixture distribution (23) as in finite samples. Interestingly, this limit distribution differs from the conventional asymptotic theory for consistent estimators of identified coefficients. Rather than the usual normal theory, we have here a mixture of normals where the mixture distribution carries the effect of the lack of identifiability of β into the asymptotic distribution of $\hat{\gamma}_1$.

(iv) γ_2 is totally unidentified and $\hat{\gamma}_2$, like $\hat{\beta}$, has a nondegenerate limit distribution. Unlike $\hat{\beta}$, however, the finite sample and asymptotic distributions of $\hat{\gamma}_2$ given by (24) and (28), respectively, are not the same. This is explained by the fact that some sources of variation in $\hat{\gamma}_2$ that are present in finite samples are eliminated as $T \to \infty$: in particular, the variation that results from the estimation of the reduced form coefficients (π_1, Π_1). The source of variation that remains as $T \to \infty$ is the uncertainty about β arising from its lack of identification and this is embodied in the limit variate r that appears in (27).

(v) As we shall see in Section 4, all of the limiting distributions given by (25)–(28) fall within the LMG family. In particular, they are all represented as scale mixtures of normals. Interestingly, this conclusion holds for both the identified and the unidentified coefficients, with the different rates of convergence that apply in the two cases. Note also that in several instances, such as (16), the distributions may also be written as covariance matrix mixtures of normals, which are still within the LMG family. However, in all cases these are easily reduced to scale mixtures of normals as shown in the final results (25)–(28).

(vi) The above results refer explicitly to the model (1) and (2) as formulated in canonical form. Here, the rows of V are i.i.d.$(0, I)$. In the general case, the rows of V are i.i.d.$(0,\Omega)$ with $\Omega > 0$. The transformations that reduce the general case to canonical form are given in Phillips [18, Theorem 3.3.1]. We use an asterisk to signify coefficients (and associated estimators) in the general case and partition Ω as

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{21}' \\ \omega_{21} & \Omega_{22} \end{bmatrix}.$$

Then

$$\beta = \omega_{11\cdot 2}^{-1/2}\Omega_{22}^{1/2}(\beta^* - \Omega_{22}^{-1}\omega_{21}),$$

$$\gamma = \omega_{11\cdot 2}^{-1/2}\gamma^*,$$

where $\omega_{11\cdot 2} = \omega_{11} - \omega_{21}'\Omega_{22}^{-1}\omega_{21}$. The corresponding IV estimators of β^* and γ^* satisfy the equations

$$r^* = \omega_{11\cdot 2}^{1/2}\Omega_{22}^{-1/2}r + \Omega_{22}^{-1}\omega_{21}, \tag{29}$$

$$s^* = \omega_{11\cdot 2}^{1/2}s. \tag{30}$$

We deduce from the above correspondence and (25)–(28) that

$$\hat{\beta}^* \Rightarrow r^* \equiv \int_{z>0} N(\Omega_{22}^{-1}\omega_{21},\ z\omega_{11\cdot 2}\Omega_{22}^{-1})pdf(z)dz, \tag{31}$$

$$\hat{\gamma}^* \Rightarrow \bar{s}^* \equiv \int_{z>0} N(\bar{\gamma}^*,\ z\omega_{11\cdot 2}P_2\Pi_1\Pi_1'P_2)pdf(z)dz, \tag{32}$$

with

$$\bar{\gamma}^* = \omega_{11\cdot 2}^{1/2}\bar{\gamma}.$$

Analogous results hold for the component estimators $\sqrt{T}(\hat{\gamma}_1^* - \gamma_1^*)$ and $\hat{\gamma}_2^*$.

Note particularly with regard to (31) that the limit distribution r^* is now a scale mixture of normals centered at $\Omega_{22}^{-1}\omega_{21}$, which is the regression coefficient of y_1 on Y_2 for a population with covariance matrix Ω.

2.3 General Asymptotic Theory

The results given in Corollary 2.2 are obtained under (C1). It is easy to see that they apply for all i.i.d. $(0,I)$ error distributions as well as those which are i.i.d. $N(0,I)$. In fact, a somewhat stronger result is possible. Suppose the rows of V form a martingale difference sequence with the natural filtration and assume that the differences are stationary, ergodic, and have conditional covariance matrix I_m. Define

$$D = [Q_{Z_1}Z_3(Z_3'Q_{Z_1}Z_3)^{-1/2},\ Z_1(Z_1'Z_1)^{-1/2}] = [D_1,D_2]. \tag{33}$$

Then we have under (C2):

LEMMA 2.3.

$$D'V \Rightarrow N_{k_*,m}(0,I). \tag{34}$$

THEOREM 2.4. *If (C2) holds and if the rows of V form a sequence of stationary, ergodic martingale differences with conditional covariance matrix I_m, then (a)–(d) of Corollary 2.2 continue to hold.*

Remarks. (i) Theorem 2.4 is an instance of the operation of an invariance principle. Theorem 2.1 and Corollary 2.2 were obtained under the $N(0,I)$ error condition (C1). According to the invariance principle, the asymptotic results should hold for a much wider class of errors. Broadly speaking, each statistic of interest can be written in the form $f_T(D'V)$, where f_T is a sequence of continuous functions that converge to a continuous function f. We know from Lemma 2.3 that $D'V \Rightarrow N(0,I)$ for a general class of errors V. By an extension of the continuous mapping theorem ([3, Theorem 5.5]), we deduce that

$$f_T(D'V) \Rightarrow f(N(0,I)), \tag{35}$$

and the results obtained under the $N(0,I)$ error condition (C1) now apply for the wider class of errors.

(ii) Note that the operation of the invariance principle described above increases the value of the finite-sample distribution theory performed under $N(0,I)$ errors. In particular, for unidentified coefficients such as β in (7), the exact finite-sample distribution of $\hat{\beta}$ under $N(0,I)$ errors is also the asymptotic distribution in the wider class. Thus, the finite-sample theory under normal errors explicitly addresses the distribution of the functional $f(N(0,I))$ that represents the asymptotic theory in the more general case (35). A special case of this phenomenon was given earlier in [22].

2.4 Statistical Tests

The properties of conventional statistical tests in partially identified structural equations are also of interest. We shall concentrate our discussion on Wald tests of hypotheses relating to the coefficients of the endogenous and exogenous regressors in (1). Each of these make use of the equation error variance estimator

$$\hat{\sigma}^2 = T^{-1}(y_1 - W_1\hat{\delta})'(y_1 - W_1\hat{\delta}) = T^{-1}(y_1 - Y_2\hat{\beta})'Q_{Z_1}(y_1 - Y_2\hat{\beta}).$$

To test

$$H_\beta : A\beta = a,$$

where A is $p_a \times n$ matrix of rank p_a $(\leq n)$, we would use the statistic

$$W_\beta = (A\hat{\beta} - a)'\{A[Y_2'(P_H - P_{Z_1})Y_2]^{-1}A'\}^{-1}(A\hat{\beta} - a)/\hat{\sigma}^2.$$

Similarly, to test

$$H_\gamma : B\gamma = b,$$

where B is $p_b \times k_1$ matrix of rank p_b $(\leq k_1)$, we have the statistic

$$W_\gamma = (B\hat{\gamma} - b)'[B(Z_1'QZ_1)^{-1}B']^{-1}(B\hat{\gamma} - b)/\hat{\sigma}^2,$$

where

$$Q = P_H - P_H Y_2(Y_2'P_H Y_2)^{-1}Y_2'P_H. \tag{36}$$

When the structural equation (1) is fully identified and (C2) holds, these statistics are conventional asymptotic χ^2 criteria and

$$W_\beta \Rightarrow \chi^2_{p_a}, \qquad W_\gamma \Rightarrow \chi^2_{p_b}, \tag{37}$$

under the null hypotheses. When the equation is partially identified, the limit theory (37) breaks down and we get quite different results.

As before, we will work with the leading case (5) to illustrate the effects of departures from the standard theory. The results are especially interesting in the important subcase where both β and γ are totally unidentified. Here, we have $r(\Pi_1) = k_1 \leq n$, and we introduce a rotation

$$L \in O(n), \qquad L = \begin{array}{c} \quad n - k_1 \quad k_1 \\ [\ L_1\ ,\ L_2\ ] \end{array}$$

with the properties that

$$\Pi_{11} = \Pi_1 L_1 = 0,$$

$$\Pi_{12} = \Pi_1 L_2, \qquad r(\Pi_{12}) = k_1.$$

The following preliminary results are useful. They hold under the conditions stated in Theorem 2.4.

LEMMA 2.5.

$\hat{\sigma}^2 \Rightarrow 1 + r'r,$

where r is the random vector given in (25).

LEMMA 2.6.

$$\text{(a)} \qquad (Y_2'P_HY_2)^{-1} \Rightarrow L_1(L_1'\xi'Q_F\xi L_1)^{-1}L_1', \qquad n > k_1,$$

$$\text{(b)} \qquad (T^{-1}Y_2'P_HY_2)^{-1} \underset{p}{\rightarrow} (\Pi_1'M_{11}\Pi)^{-1}, \qquad n = k_1,$$

where

$\xi \equiv N_{k_*,n}(0,I),$

$$F = \begin{array}{c} k_1 \\ \left[\begin{array}{c} 0 \\ M_{11}^{1/2}\Pi_{12} \end{array}\right] \end{array} \begin{array}{c} n - k_1 \\ k_1 \end{array}.$$

LEMMA 2.7.

$$Z_1'QZ_1 \Rightarrow \Pi_{12}'^{-1}\xi_2'\{Q_F - Q_F\xi_1(\xi_1'Q_F\xi_1)^{-1}\xi_1'Q_F\}\xi_2\Pi_{12}^{-1},$$

$$\equiv \Pi_{12}'^{-1}W_{k_1}(k_* - n, I)\Pi_{12}^{-1}, \tag{38}$$

where

$[\xi_1, \xi_2] = \xi[L_1, L_2] = \xi L \equiv N_{k_*,n}(0,I).$

Remarks. (i) Lemma 2.5 shows that, in contrast to identified structural equations, the standard error of regression converges weakly to a random variable, whose distribution depends on the limiting distribution of the structural coefficient estimator.

(ii) Lemma 2.6 shows that when $n > k_1$ there is a singularity in the limit of the inverse of the sample moment matrix $(Y_2'P_HY_2)^{-1}$. Moreover, the limit matrix

$$L_1(L_1'\xi'Q_F\xi L_1)^{-1}L_1 = L_1(\xi_1'Q_F\xi_1)^{-1}L_1$$

is random and its distribution depends on the inverse of

$$\xi_1'Q_F\xi_1 \equiv W_{n-k_1}(k_* - k_1, I). \tag{39}$$

The rank of the limit matrix (39) is $n - k_1$. When $n = k_1$, the matrix L_1 has no columns and (39) may be interpreted as the zero matrix. In fact, rescaling is required to avoid degeneracy and part (b) of the lemma gives the appropriate result for $T(Y_2'P_HY_2)^{-1}$ in this case. Note that part (a) is very different from conventional theory for simultaneous systems where we would expect $(Y_2'P_HY_2)^{-1} = O_p(T^{-1})$. The differences arise not only because of the lack of identifiability of β and γ but also because of the column rank deficiency of Π_1. The latter ensures that there are certain linear combinations of Y_2 which depend only on the errors V_2 and not the exogenous variables Z_1. The limit behavior of idempotent quadratic forms in such linear combinations is quite different from those involving the exogenous variables. This follows directly from Lemma 2.3.

(iii) When $n = k_1$, there is no rank deficiency in Π_1 and the sample moment matrix $T^{-1}Y_2'P_HY_2$ has a nonsingular probability limit $\Pi_1'M_{11}\Pi_1$. The limiting behavior in cases (a) and (b) of Lemma 2.6 is therefore quite distinct. Note also that the standardization is different in the two cases.

(iv) Lemma 2.7 describes the limiting behavior of the matrix $Z_1'QZ_1$. Note that this is proportional to the inverse of the usual estimate of the asymptotic covariance matrix of $\hat{\gamma}$. We see that $Z_1'QZ_1$ converges weakly to a random matrix of full rank k_1 a.s. Thus, the limit of $(Z_1'QZ_1)^{-1}$ is given by

$$\{\Pi_{12}'^{-1}W_{k_1}(k_* - n, I)\Pi_{12}^{-1}\}^{-1} \equiv \{W_{k_1}(k_* - n, (\Pi_{12}\Pi_{12}')^{-1})\}^{-1}, \tag{40}$$

and this random matrix properly represents the uncertainty about $\hat{\gamma}$ that is implicit in its (total) lack of identification in this case.

(v) Note also that (38) holds for all $n \geq k_1$ in spite of the different limiting behavior of $Y_2'P_HY_2$ in the two cases $n > k_1$ and $n = k_1$ given in Lemma 2.6.

(vi) The proof of Lemma 2.7 is of some independent interest. Observe that under (C2), $Z_1'Z_1$ is $O(T)$. The limiting behavior of $Z_1'QZ_1$ therefore involves a degeneracy in which the leading term is zero. The proof in the Appendix shows how to develop an expansion that yields the next dominant term. In the present case, the next term is $O_p(1)$ and $Z_1'QZ_1$ converges weakly to the nondegenerate random matrix (38) in the limit.

THEOREM 2.8. *Under the conditions of Theorem 2.4,*

(a) $$W_\beta \Rightarrow (Ar - a)'W(A)(Ar - a)/(1 + r'r), \tag{41}$$

(b) $$W_\gamma \Rightarrow (\bar{B}r - \bar{b})'W(\bar{B})(\bar{B}r - \bar{b})/(1 + r'r), \tag{42}$$

where

$$W(A) \equiv W_{p_a}(k_3 - n + p_a, (AA')^{-1}),$$
$$W(\bar{B}) \equiv W_{p_b}(k_3 - n + p_b, (\bar{B}\bar{B}')^{-1}),$$
$$\bar{B} = -B\Pi_1, \qquad \bar{b} = b - B\pi_1.$$

In these representations r, $W(A)$ and $W(B)$ are dependent variables.

Remarks. (i) The limiting distributions represented by (41) and (42) are not chi-squared, so that conventional theory under the null is obviously inappropriate. We also observe that (41) and (42) continue to hold under the alternative hypotheses

$$H_\alpha : A\beta \neq a, \qquad H_\gamma : \beta\gamma \neq b.$$

The tests are therefore inconsistent. This squares with the fact that β and γ are unidentified. Even an infinite sample of data delivers no information about these parameters, so that data-based tests cannot discriminate between the null and the alternative hypothesis.

(ii) Note also that the distributions given in (41) and (42) are invariant to the true values of β and γ. Thus, the distributions themselves are invariant under the null and alternative hypotheses.

The polar case in which the coefficient vector γ is fully identified is also of interest. Here, $\Pi_1 = 0$, and hence $Y_2 = V_2$. The limiting behavior of the sample moment matrix $T^{-1}Z_1'QZ_1$ is now quite different from (38). We have instead:

LEMMA 2.9. *If* $\Pi_1 = 0$, *then as* $T \to \infty$,

$$T^{-1}Z_1'QZ_1 \Rightarrow M_{11}^{1/2}\Theta_{21}\Theta_{21}'M_{11}^{1/2}, \tag{43}$$

where Θ_{21} *is the* $k_1 \times (k_* - n)$ *submatrix of*

$$\Theta = \begin{array}{cc} \begin{array}{cc} k_* - n & n \end{array} & \\ \begin{bmatrix} \Theta_{11} & \Theta_{12} \\ \Theta_{21} & \Theta_{22} \end{bmatrix} & \begin{array}{c} k_3 \\ k_1 \end{array} \end{array} \in O(k_*)$$

in the displayed partition and

$$\Theta \equiv U(O(k_*)),$$

i.e., Θ *is uniform on the orthogonal group* $O(k_*)$.

THEOREM 2.10. *If* $\Pi_1 = 0$, *then as* $T \to \infty$,

$$W_\gamma \Rightarrow \bar{s}'B'\{B(M_{11}^{1/2}\Theta_{21}\Theta_{21}'M_{11}^{1/2})^{-1}B'\}^{-1}B\bar{s}/(1 + r'r) \tag{44}$$

under the null hypothesis H_γ*; and* W_γ *diverges when* H_γ *is false. Here*

$$\bar{s} \equiv \int_{m>0} N(0,(1 + m)M_{11}^{-1})pdf(m)dm,$$

as in (26) (with $R_1 = I$*),* r *is given by (25) and* Θ_{21} *is given in (43).*

Remarks. (i) Theorem 2.10 shows that when γ is identified, a Wald test of H_γ is consistent. However, use of conventional chi-squared critical values leads to a size distortion in the test which persists asymptotically. This distortion is caused by the nonidentifiability of β, which induces: (a) a random limit for the error variance estimator $\hat{\sigma}^2$; (b) a nonnormal limit distribution for the scaled error in the coefficient $\sqrt{T}(\hat{\gamma} - \gamma)$; and (c) a random limit for the covariance matrix estimator $T^{-1}Z_1'QZ_1$. Each of these effects figure in the nonstandard limit distribution as it is expressed in (44).

(ii) Notwithstanding the above remark, the numerator quadratic form in

(44) is a standard χ^2 variate. To see this, it is simplest to work directly from W_γ. Under the null,

$$B\hat{\gamma} - b = B(\hat{\gamma} - \gamma) = B(Z_1'QZ_1)^{-1}Z_1'Qu,$$

where

$$u = v_1 - V_2\beta$$

is the structural equation error in (1). When $\Pi_1 = 0$, we have $Y_2 = V_2$ and so $QV_2 = 0$. Thus, $Z_1'Qu = Z_1'Qv_1$. In view of Lemma 2.3,

$$D'V = D'[v_1, V_2] \Rightarrow [X_1, X_2] \equiv N_{k_*,m}(0, I),$$

and from (A11) in the Appendix, we have

$$T^{-1/2}Z_1'D \to [0, M_{11}^{1/2}] = E \text{ (say)}$$

as $T \to \infty$. We write, under the null,

$$\begin{aligned}(B\hat{\gamma} - b)'&[B(Z_1'QZ_1)^{-1}B']^{-1}(B\hat{\gamma} - b)\\ &= v_1'QZ_1(Z_1'QZ_1)^{-1}B'[B(Z_1'QZ_1)^{-1}B']^{-1}B(Z_1'QZ_1)^{-1}Z_1'Qv_1,\\ &= v_1'Q(B)v_1 \text{ (say)}.\end{aligned}$$

Here, $Q(B)$ is idempotent of rank p_b and depends on $D'Y_2 = D'V_2$. Under (C1), it is obvious that conditional on $D'V_2$

$$v_1'Q(B)v_1|_{D'V_2} \equiv \chi^2_{p_b},$$

and, being independent of $D'V_2$, this is also the unconditional distribution. The same argument applies in the limit as $T \to \infty$ in the general case. Since $Q = P_HQ = QP_H$, we simply write

$$v_1'Q(B)v_1 = v_1'D(D'Q(B)D)D'v_1 = v_1'D\bar{Q}D'v_1.$$

Here $\bar{Q} = \bar{Q}(D'V_2, T^{-1/2}Z_1'D)$ is idempotent of rank p_b and depends only on the random matrix $D'V_2$ and the nonrandom matrix $T^{-1/2}Z_1'D$. But $D'V_2 \Rightarrow X_2$, $D'v_1 \Rightarrow X_1$, and X_1 and X_2 are independent. Thus,

$$v_1'D\{\bar{Q}(D'V_2, T^{-1/2}Z_1'D)\}D'v_1 \Rightarrow X_1'\{\bar{Q}(X_2, E)\}X_1 \equiv \chi^2_{p_b}$$

by the argument above. It follows that (44) may be reduced to the simpler form

$$W_\gamma \Rightarrow \chi^2_{p_b}/(1 + r'r). \tag{45}$$

However, it is easy to see that the numerator and the denominator of (45) are statistically dependent. In fact, partitioning $D = [D_1, D_2]$ as in (33), we may write conformably

$$D'V_2 = \begin{bmatrix} D_1'V_2 \\ D_2'V_2 \end{bmatrix} \Rightarrow \begin{bmatrix} X_{21} \\ X_{22} \end{bmatrix} = X_2,$$

and

$$D'v_1 = \begin{bmatrix} D_1'v_1 \\ D_2'v_1 \end{bmatrix} \Rightarrow \begin{bmatrix} X_{11} \\ X_{12} \end{bmatrix} = X_1.$$

Then we have

$$\hat{\beta} \Rightarrow r = (X_{21}X_{21}')^{-1}X_{21}X_{11},$$

which makes explicit the dependence in the ratio (45).

(iii) The form of (45) is simple and rather interesting. It even suggests the possibility that inferences about γ might be performed conditional on estimates of the unidentified coefficients β. To examine this possibility further, consider the form of $\hat{\gamma}$ given in (15), namely,

$$\hat{\gamma} = (Z_1'Z_1)^{-1}Z_1'Y\begin{pmatrix} 1 \\ -\hat{\beta} \end{pmatrix}.$$

Under $\Pi_1 = 0$, we have $\gamma = \pi_1$ and

$$\hat{\gamma} - \gamma = (Z_1'Z_1)^{-1}Z_1'(v_1 - V_2\hat{\beta}).$$

Conditional on $\hat{\beta}$ (or, equivalently, $D_1'V$) and under (C1), we have

$$\hat{\gamma} - \gamma|_{\hat{\beta}} \equiv N(0, (1 + \hat{\beta}'\hat{\beta})(Z_1'Z_1)^{-1}).$$

To test $H_\gamma : B\gamma = b$, consider the statistic

$$\hat{W}_\gamma = (B\hat{\gamma} - b)'\{(1 + \hat{\beta}'\hat{\beta})B(Z_1'Z_1)^{-1}B'\}^{-1}(B\hat{\gamma} - b).$$

Now

$$\hat{W}_\gamma|_{\hat{\beta}} \equiv \chi^2_{p_b}.$$

When (C1) is relaxed, we obtain, using Lemma 2.3, the same result in the limit, namely,

$$\hat{W}_\gamma|_{\hat{\beta}} \Rightarrow \chi^2_{p_b}. \tag{46}$$

The relation (46) may be used to make valid conditional inferences about γ. In effect, we estimate (1) and then conduct statistical tests conditional on the estimated values of the unidentified coefficients. Note that in the present context, we may regard $\hat{\beta}$ as an ancillary statistic. Its distribution, as we have seen, does not depend on any parameter other than the degrees of freedom $q = k_3 - n + 1$. In particular, the distribution of $\hat{\beta}$ does not depend on γ. In making inferences about γ, it is therefore appropriate to proceed conditionally on the observed estimate of $\hat{\beta}$. This approach leads directly to an asymptotic χ^2 test based on (46). Some of the conceptual issues involved in performing conditional inferences of this type have recently been discussed by Lehmann [13].

(iv) Since $\gamma = \pi_1$ when $\Pi_1 = 0$, we may conduct tests of H_γ that are based directly on reduced-form estimates such as

$$\hat{\pi}_1 = (Z_1'Q_2Z_1)^{-1}Z_1'Q_2y_1, \qquad Q_2 = I - Z_2(Z_2'Z_2)^{-1}Z_2'.$$

Thus, we have the Wald test

$$W_\pi = (B\hat{\pi}_1 - b)'[B(Z_1'Q_2Z_1)^{-1}B']^{-1}(B\hat{\pi}_1 - b)/\hat{\sigma}_v^2,$$

where

$$\hat{\sigma}_v^2 = T^{-1}y_1'\{Q_2 - Q_2Z_1(Z_1'Q_2Z_1)^{-1}Z_1'Q_2\}y_1.$$

This leads to conventional asymptotic tests based on

$$W_\pi \Rightarrow \chi^2_{p_b}.$$

(v) We must observe that the procedures outlined in Remarks (iii) and (iv) above both depend on the knowledge that $\Pi_1 = 0$ and, in the case of (iii), that β is unidentified ($\Pi_2 = 0$). This information is not available in practice, although, if it were suspected, pretests of $\Pi_2 = 0$ and $\Pi_1 = 0$ could be carried out using reduced-form estimates of these coefficient matrices. In effect, such tests would assess the empirical support for the total lack of identification of β and the identifiability of γ. In the absence of such information, we can expect that tests based on W_γ will be conducted and then the results of Theorem 2.10 apply.

(vi) The analysis of this section may be extended to the case where $0 < r(\Pi_1) < k_1$. The algebra is somewhat more complicated and will not be reported here. The polar cases of $r(\Pi_1) = 0$ (γ identified) and $r(\Pi_1) = k_1$ (γ unidentified) serve well in illustrating the main conclusions.

3. TIME SERIES REGRESSIONS

3.1 Spurious Regressions

Let $\{z_t\}_0^\infty$ be an m-vector integrated process with generating mechanism

$$z_t = z_{t-1} + \xi_t, \qquad t = 1,2,\ldots. \tag{47}$$

The initial value z_0 in (47) may be any random variable, including a constant. The sequence $\{\xi_t\}$ is strictly stationary and ergodic with zero mean, finite variance, and continuous spectral density matrix $f_{\xi\xi}(\lambda)$. We further assume that the partial sum process constructed from $\{\xi_t\}$ satisfies a multivariate invariance principle. In effect, for $r \in [0,1]$ and as $T \to \infty$, we require:

(C3) $X_T(r) = T^{-1/2}\Sigma_1^{[Tr]}\xi_t \Rightarrow B(r)$,

where $B(r) \equiv BM(\Omega)$, i.e., Brownian motion with covariance matrix

$$\Omega = 2\pi f_{\xi\xi}(0) = \Omega_0 + \Omega_1 + \Omega_1' \tag{48}$$

with

$$\Omega_0 = E(\xi_0\xi_0'), \qquad \Omega_1 = \sum_{k=1}^{\infty} E(\xi_0\xi_k').$$

More explicit conditions under which (C3) holds are discussed in detail in earlier work by the author [23,25,29].

We now partition $z_t = (y_t, x_t')'$ into the scalar variate y_t and the n-vector x_t $(m = n + 1)$ with the following conformable partitions of Ω and $B(r)$:

$$\Omega = \begin{array}{cc} \begin{array}{cc} 1 & n \end{array} & \\ \begin{bmatrix} \omega_{11} & \omega_{21}' \\ \omega_{21} & \Omega_{22} \end{bmatrix} & \begin{array}{c} 1 \\ n \end{array} \end{array}, \qquad B(r) = \begin{bmatrix} B_1(r) \\ B_2(r) \end{bmatrix} \begin{array}{c} 1 \\ n \end{array}.$$

In this section and in Section 3.2, we shall assume that $\Omega > 0$. We shall further use the following conformable block triangular decomposition of Ω:

$$\Omega = L'L, \qquad L = \begin{bmatrix} l_{11} & 0 \\ l_{21} & L_{22} \end{bmatrix}$$

with

$$l_{11} = (\omega_{11} - \omega_{21}'\Omega_{22}^{-1}\omega_{21})^{1/2}, \qquad l_{21} = \Omega_{22}^{-1/2}\omega_{21}, \qquad L_{22} = \Omega_{22}^{1/2}.$$

Our object of study is the linear least-squares regression

$$y_t = \hat{\beta}'x_t + \hat{u}_t. \tag{49}$$

When $\{\xi_t\}$ is i.i.d. $N(0,I)$, (49) reduces to the Granger and Newbold prototype of a spurious regression. In this context, y_t and x_t are independent, integrated processes. Yet the regression (49) typically leads to apparently significant correlations in conventional regression significance tests, thereby justifying the nomenclature spurious regression. The properties of such time series regressions in the general stochastic environment determined by (47) have recently been analyzed in detail by the author [23].

The limiting behavior of $\hat{\beta}$ in (49) is a simple consequence of (C3). In particular, as shown in Phillips [23], when $T \to \infty$

$$\hat{\beta} \Rightarrow \left(\int_0^1 B_2B_2'\right)^{-1}\left(\int_0^1 B_2B_1\right), \tag{50}$$

a matrix quotient of quadratic functionals of the Brownian motion B. As it stands, the representation (50) is simple and elegant, but not very helpful in terms of setting the limiting distribution of $\hat{\beta}$ in the wider context of general asymptotic theory. The following results help to do just that.

Let the m-dimensional Brownian motion B be defined on the probability space (Ω, F, P) and let F_2 be the sub-σ-field of F that is generated by $\{B_2(r) : 0 \le r \le 1\}$. We use the symbol "$\cdot\,|F_2$" to signify the conditional distribution relative to F_2.

LEMMA 3.1.

$$B_1|_{F_2} \equiv \omega_{21}'\Omega_{22}^{-1}B_2 + l_{11}W_1, \tag{51}$$

where W_1 is an independent standard Brownian motion, i.e., $W_1 \equiv BM(1)$ and is independent of B_2.

THEOREM 3.2.

$$\hat{\beta} \Rightarrow \int_{V>0} N(\Omega_{22}^{-1}\omega_{21}, \omega_{11\cdot 2}V(B_2))dP(V), \tag{52}$$

$$\equiv \int_{v>0} N(\Omega_{22}^{-1}\omega_{21}, v\omega_{11\cdot 2}\Omega_{22}^{-1})dP(v), \tag{53}$$

where

$$V = \int_0^1\int_0^1 C(r)(r \wedge s)C(s)'drds,$$

$$v = \int_0^1\int_0^1 U(r)(r \wedge s)U(s)drds,$$

$$C(r) = \left(\int_0^1 B_2B_2'\right)^{-1} B_2(r),$$

$$U(r) = e_1'\left(\int_0^1 W_2W_2'\right)^{-1} W_2(r),$$

$$e_1' = (1,0,\ldots,0), \qquad (n \times 1),$$

and $W_2 \equiv BM(I_n)$ or n-vector standard Brownian motion. $P(V)$ denotes the probability measure on the covariance matrix $V = V(B_2) > 0$ that is induced by the vector Brownian motion B_2; and $P(v)$, similarly, is the probability measure on the variance $v = v(W_2) > 0$ that is induced by the standard Brownian motion W_2.

Remarks. (i) Lemma 3.1 shows how to represent a conditional Brownian motion as unconditional Brownian motion about a given Brownian path. The latter is the conditional mean of the new process and the conditional variance is simply proportional to $l_{11}^2 = \omega_{11\cdot 2} = \omega_{11} - \omega_{21}'\Omega_{22}^{-1}\omega_{21}$, the conventional conditional variance from a multivariate normal distribution. In fact, (51) may be regarded as the Gaussian process analogue of familiar theory from normal multivariate analysis. It has many useful applications.

(ii) Theorem 3.2 gives the limiting distributions of $\hat{\beta}$ as a covariance matrix mixture of normals in (52) and as a simpler variance mixture of normals in (53). Note that the covariance matrix of the process $z_t' = (y_t, x_t')$ is approximately $t\Omega$. Theorem 3.2 shows that the limit distribution of the sample regression coefficient $\hat{\beta}$ is a scale mixture of normals centered at $\Omega_{22}^{-1}\omega_{21}$, which may be interpreted as the population regression coefficient of y on x for the time series $\{z_t\}$ with asymptotic covariance matrix $t\Omega$.

(iii) As discussed in Phillips [23], the limit distribution of $\hat{\beta}$ is nondegenerate. This is a manifestation of the spurious nature of the regression. In effect, the noise in the regression (49) is as strong as the signal and is, moreover, contaminated with it. This leads to a persistent indeterminacy in the regression which is reflected in the dispersion of the limit distribution of $\hat{\beta}$.

(iv) There is a striking relationship between the results in Theorem 3.2 for the time series regression (49) and those obtained earlier in Section 2.2 for the structural equation estimator when the coefficient vector β is totally unidentified. For the latter case, we obtained (see (31) above):

$$\hat{\beta}^* \Rightarrow \int_{z>0} N(\Omega_{22}^{-1}\omega_{21}, z\omega_{11\cdot 2}\Omega_{22}^{-1})\, pdf(z)\,dz, \tag{54}$$

where $\hat{\beta}^*$ is the IV estimator of the structural coefficient β in (1) and where Ω is the covariance matrix of the endogenous variables that appear in the equation. The similarity between (53) and (54) is indeed striking; and it goes deeper than the apparent similarity in the formulae. This is because both regressions share a fundamental indeterminacy: the structural-equation case in view of the total lack of identification of the coefficients; and the time series regression since the signal is persistently swamped by the strength of the noise. In neither case is the signal delivered by the regressors sufficiently clear and uncontaminated by noise to provide determinacy. Our results indicate that the spurious-regression case may be regarded as a time series analogue of the structural equation regression under lack of identification.

(v) Suppose, for example, that for some constant n-vector β, we defined $u_t = (1,-\beta')z_t$ with z_t generated by (47). Then, we have

$$y_t = \beta' x_t + u_t, \tag{55}$$

and the regression (49) might purport to estimate (55). However, there is no information about β in the generating mechanism (47) so that β is clearly unidentified. In this sense, the time series model (55) with reduced form (47) may be reinterpreted as an unidentified structural equation in a simultaneous system. Note that there are no identifiability relations corresponding to (3) and (4) because in the present case $\Pi = 0$. Indeed, we may write the reduced form (47) simply as

$$z_t = v_t, \qquad v_t = v_{t-1} + \xi_t. \tag{56}$$

Here, the noise v_t in the reduced form is itself an integrated process and there is no systematic component (i.e., $\Pi = 0$). We then obtain $u_t = (1,-\beta')v_t$ in (55), and the noise in (55) is as strong as the signal x_t and is, in general, contaminated with it. Since $\Pi = 0$ in (56), the data carry no information about β and the vector is unidentified.

(vi) The above results may be readily extended to regressions such as (49) with a fitted intercept or time trend. The formulae derived still apply but with B_2 (and, hence W_2) replaced by demeaned or detrended Brownian motion, namely,

$$\bar{B}_2 = B_2(r) - \int_0^1 B_2, \tag{57}$$

$$\tilde{B}_2 = B_2(r) - \tilde{\alpha}_0 - \tilde{\alpha}_1 r, \tag{58}$$

where

$$\begin{bmatrix} \tilde{\alpha}_0 \\ \tilde{\alpha}_1 \end{bmatrix} = \begin{bmatrix} 0 & \int_0^1 s \\ \int_0^1 s & \int_0^1 s^2 \end{bmatrix}^{-1} \begin{bmatrix} \int_0^1 B_2 \\ \int_0^1 sB_2 \end{bmatrix} = \begin{bmatrix} 4\int_0^1 B_2 - 6\int_0^1 sB_2 \\ 12\int_0^1 sB_2 - 6\int_0^1 B_2 \end{bmatrix}$$

(and $\overline{W}_2, \tilde{W}_2$, respectively, in the case of W_2).

3.2 Partially Spurious Regressions

When the generating mechanism (47) involves a systematic component such as a drift, then regressions such as (49) are only partially spurious. Let us suppose that, in place of (47), we have the corresponding model with a drift vector μ, namely,

$$z_t = \mu + z_{t-1} + \xi_t, \tag{59}$$

or in partitioned form

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \xi_{1t} \\ \xi_{2t} \end{bmatrix}. \tag{60}$$

We assume that $\mu_2 \neq 0$ and to simplify notation we further assume that the system (60) has been scaled so that $\mu_2'\mu_2 = 1$. Note that (59) and (60) may also be written as

$$z_t = \mu t + z_t^0, \qquad y_t = \mu_1 t + y_t^0, \qquad x_t = \mu_2 t + x_t^0,$$

where the superscript zero signifies a driftless $I(1)$ process, i.e.,

$$z_t^0 = z_{t-1}^0 + \xi_t, \qquad y_t^0 = y_{t-1}^0 + \xi_{1t}, \qquad x_t^0 = x_{t-1}^0 + \xi_{2t}.$$

Since $\mu_2 \neq 0$, we have

$$y_t - \mu_1\mu_2' x_t = y_t^0 - \mu_1\mu_2' x_t^0 = u_t \text{ (say)}, \tag{61}$$

where u_t is a driftless $I(1)$ process also.

We now write (61) as

$$y_t = \beta' x_t + u_t, \qquad \beta = \mu_1 \mu_2, \tag{62}$$

and the least-squares regression equation (49) may be interpreted as an estimate of (62). The following theorem gives the asymptotic properties of these regression coefficients for data generated by (59). It is convenient to introduce the rotation

$$H = [\mu_2, H_2] \in O(n)$$

and transform coordinates of the regressors in (62) according to

$$\begin{aligned} y_t &= \beta' H H' x_t + u_t, \\ &= \beta_1 x_{1t} + \beta_2' x_{2t} + u_t, \end{aligned} \tag{63}$$

where

$$\beta_1 = \beta' \mu_2 = \mu_1, \qquad \beta_2 = H_2' \beta = \mu_1 H_2' \mu_2 = 0,$$

and

$$x_{1t} = t + \mu_2' x_t^0,$$

$$x_{2t} = H_2' x_t^0 = x_{2t-1} + H_2' \xi_{2t}.$$

Let $\hat{\beta}_1$ and $\hat{\beta}_2$ be the least-squares regression coefficients of β_1 and β_2 in (63). Then

$$\hat{\beta} = H\hat{\underline{\beta}} = \mu_2 \hat{\beta}_1 + H_2 \hat{\beta}_2, \tag{64}$$

and

$$\beta = H\underline{\beta} = \mu_2 \beta_1 + H_2 \beta_2 = \mu_2 \mu_1. \tag{65}$$

Assuming that $\{\xi_t\}$ satisfies (C3), we construct from $B(r)$ the n-vector Brownian motion

$$\underline{B}(r) = \begin{bmatrix} \underline{B}_1(r) \\ \underline{B}_2(r) \end{bmatrix} = \begin{bmatrix} 1 & -\beta' \\ 0 & H_2' \end{bmatrix} B(r) \equiv BM(\underline{\Omega}),$$

with

$$\underline{\Omega} = \begin{bmatrix} \underline{\omega}_{11} & \underline{\omega}_{21}' \\ \underline{\omega}_{21} & \underline{\Omega}_{22} \end{bmatrix},$$

where

$$\underline{\omega}_{11} = \omega_{11} - 2\beta' \omega_{21} + \beta' \Omega_{22} \beta,$$

$$\underline{\omega}_{21} = H_2(\omega_{21} - \Omega_{22}\beta),$$

$$\underline{\Omega}_{22} = H_2 \Omega_{22} H_2'.$$

We now have the following theorem.

THEOREM 3.3.

(a)
$$\sqrt{T}(\hat{\beta}_1 - \beta_1) \Rightarrow \left(\int_0^1 \zeta^2\right)^{-1}\left(\int_0^1 \zeta \underline{\mathbf{B}}_1\right), \tag{66}$$

$$= \int_{v_1>0} N(0, \underline{\omega}_{11\cdot 2} v_1) dP(v_1), \tag{67}$$

(b)
$$\hat{\beta}_2 \Rightarrow \left(\int_0^1 \eta\eta'\right)^{-1}\left(\int_0^1 \eta \underline{\mathbf{B}}_1\right), \tag{68}$$

$$\equiv \int_{V_2>0} N(\underline{\Omega}_{22}^{-1}\underline{\omega}_{21}, \underline{\omega}_{11\,2} V_2) dP(V_2), \tag{69}$$

$$\equiv \int_{v_2>0} N(\underline{\Omega}_{22}^{-1}\underline{\omega}_{21}, v_2 \underline{\omega}_{11\cdot 2}\underline{\Omega}_{22}^{-1}) dP(v_2), \tag{70}$$

where

$$\zeta(r) = r - \left(\int_0^1 s\underline{\mathbf{B}}_2'\right)\left(\int_0^1 \underline{\mathbf{B}}_2\underline{\mathbf{B}}_2'\right)^{-1}\underline{\mathbf{B}}_2(r),$$

$$\eta(r) = \underline{\mathbf{B}}_2(r) - \left(\int_0^1 \underline{\mathbf{B}}_2 s\right)\left(\int_0^1 s^2\right)^{-1} r,$$

$$v_1 = \int_0^1\int_0^1 U_1(r)(r \wedge s)U_1(s)drds,$$

$$U_1(r) = \left(\int_0^1 \zeta^2\right)^{-1}\zeta(r),$$

$$V_2 = \int_0^1\int_0^1 C_2(r)(r \wedge s)C_2'(s)drds,$$

$$C_2(r) = \left(\int_0^1 \eta\eta'\right)^{-1}\eta(r),$$

$$v_2 = \int_0^1\int_0^1 U_2(r)(r \wedge s)U_2(s)drds,$$

$$U_2(r) = e_1'\left(\int_0^1 \underline{\eta}\underline{\eta}'\right)^{-1}\underline{\eta}(r),$$

$$\underline{\eta}(r) = \underline{\mathbf{W}}_2(r) - \left(\int_0^1 \underline{\mathbf{W}}_2 r\right)\left(\int_0^1 r^2\right)^{-1} r,$$

$$\underline{\mathbf{W}}_2(r) \equiv BM(I_{n-1}).$$

COROLLARY 3.4.

$$\hat{\beta} - \beta \Rightarrow H_2\left(\int_0^1 \eta\eta'\right)^{-1}\left(\int_0^1 \eta \underline{B}_1\right),$$

$$\equiv H_2 \int_{v_2>0} N(\underline{\Omega}_{22}^{-1}\underline{\omega}_{21}, v_2 \underline{\omega}_{11\cdot 2}\underline{\Omega}_{22}^{-1})dP(v_2). \tag{71}$$

Remarks. (i) We see from Theorem 3.3(a) that $\hat{\beta}_1 = \beta_1 + O_p(T^{-1/2})$, so that $\hat{\beta}_1$ is a consistent estimator of β_1. The component $\beta_1 = \mu_2'\beta = \mu_1$ may be interpreted as an (asymptotically) estimable function of the vector β in (62).

(ii) Again, we have an analogy with structural equation estimation. In this case, (62) may be viewed as a structural equation in which (y_t, x_t) are endogenous with reduced form given by

$$z_t = \mu t + z_t^0, \qquad z_t^0 = z_{t-1}^0 + \xi_t. \tag{72}$$

The systematic component in the reduced form (72) is μt. The analogue of the identifiability relation (4) is

$$(1, -\beta')\mu = 0,$$

or

$$\mu_1 - \beta'\mu_2 = 0.$$

This relation not only determines the conditions under which a particular component of β is identifiable (namely, $\mu_2 \neq 0$) but also indicates what that component is, namely, $\beta_1 = \mu_2'\beta = \mu_1$.

(iii) Note that, when $\mu_2 \neq 0$, x_t itself involves a drift and that we may write $x_t = \mu_2 t + x_t^0$, where x_t^0 is a driftless $I(1)$ process. Thus, $x_t = O_p(t)$ and this trending component of x_t dominates the stochastic trend in $x_t^0 = O_p(\sqrt{t})$. Moreover, the presence of this trending component in x_t ensures that the regression (49) results in a consistent estimate of μ_1, the trend in $y_t = \mu_1 t + y_t^0$. The remaining components in the regression are spurious. This leads us to the nomenclature, partially spurious regression. In effect, there is a component in the signal x_t which dominates the noise and it is this component that is consistently estimated in the regression.

(iv) The limiting distribution of $\sqrt{T}(\hat{\beta}_1 - \beta_1)$ is mixed normal as we see from (67). Note especially that when $n > 1$, this distribution is nonnormal. When $n = 1$, we have $\zeta(r) = r$ (the component of $\zeta(r)$ involving $\underline{B}_2(r)$ is annihilated) and

$$v_1 = \int_0^1 \int_0^1 rs(r \wedge s)drds \bigg/ \left(\int_0^1 r^2\right)^2 = 3$$

is a constant. In this special case, we have

$$\sqrt{T}(\hat{\beta}_1 - \beta_1) \Rightarrow N(0, 3\omega_{11\cdot 2}).$$

(cf. Park and Phillips [15, p. 17]).

(v) Those components of β which are unidentified when $n > 1$, namely, $\beta_2 = H_2'\beta = 0$, are estimated by $\hat{\beta}_2$. For these components, we have results that are entirely analogous to those that apply in a spurious regression. Indeed, the representations of the limiting distribution of $\hat{\beta}_2$ given by (69) and (70) closely mirror the earlier representations (52) and (53). We see that this asymptotic theory again falls in the compound normal family and may be regarded either as a covariance matrix mixture or a scalar mixture of underlying normals.

(vi) The representations (66) and (68) depend on the functionals $\zeta(r)$ and $\eta(r)$ of the Brownian motion $\underline{B}_2(r)$. These functionals have simple interpretations. In the space $L_2[0,1]$, $\zeta(r)$ is the projection of r on the orthogonal complement of the space spanned by the components of $\underline{B}_2$. Similarly, in $L_2[0,1]^{n-1}$, $\eta(r)$ is the projection of $\underline{B}_2$ on the orthogonal complement of the space spanned by the trend r. These functionals preserve, in the asymptotic representations (66) and (68), characteristics of the finite-sample construction of the statistics $\hat{\beta}_1$ and $\hat{\beta}_2$ that are evident from regression formulae. See, in particular, formulae (A20) and (A21) in the Appendix.

3.3 Cointegrating Regressions

In Sections 3.1 and 3.2, we have assumed that $\Omega > 0$. When Ω is singular, a different theory applies. In this case, the variables in z_t are said to be cointegrated [5] and the generating mechanism (47) has a deficient set of unit roots. An asymptotic theory for regression has been investigated in other work [23,29,30,32]. When the submatrix $\Omega_{22} > 0$, we know that $\gamma' = (1, -\omega_{21}'\Omega_{22}^{-1})$ is a cointegrating vector, that $\Omega\gamma = 0$, and that $\hat{\beta} \to_p \Omega_{22}^{-1}\omega_{21}$ (see [23]). We write the new cointegrated system as

$$y_t = \beta' x_t + \xi_{1t}, \tag{73}$$

$$x_t = x_{t-1} + \xi_{2t}, \tag{74}$$

where $\beta = \Omega_{22}^{-1}\omega_{21}$ and where $\xi_t' = (\xi_{1t}, \xi_{2t}')$ satisfies (C3) as before. We have the following limiting distribution theory from Phillips [30]:

$$T(\hat{\beta} - \beta) \Rightarrow \left(\int_0^1 B_2 B_2'\right)^{-1} \left(\int_0^1 B_2 dB_1 + \lambda\right), \tag{75}$$

where

$$\lambda = \sum_{k=0}^{\infty} E(\xi_{20}\xi_{1k}). \tag{76}$$

Here, (75) depends on the theory of weak convergence to the stochastic integral $\int_0^1 B_2 dB_1$ (see [4] and [26]) and allows also for a bias term λ. Note that λ is nonzero even for i.i.d. sequences $\{\xi_t\}$ when $\omega_{21} \neq 0$.

The term λ arises because of the correlation between x_t and ξ_{1t} and it carries what is, in effect, a second-order simultaneous equations bias. Because of the strength of the signal x_t (an $I(1)$ process) relative to the error ξ_{1t} (an $I(0)$ process), $\hat{\beta}$ is consistent for β; and the bias term λ has only a second-order effect on the asymptotic distribution of $\hat{\beta}$. This is in contrast to the first-order effect of conventional simultaneous equations bias for models with $I(0)$ regressors, where the bias induces an inconsistency in the least-squares estimate $\hat{\beta}$.

The presence of the bias term in (75) does not of itself seem of major significance. Nevertheless, it turns out to be of importance (i) in matters of inference because of the nuisance parameters carried in λ and (ii) when it comes to determining the general asymptotic family to which (75) belongs. When $\lambda = \omega_{21} = 0$, (75) falls within the compound normal distribution family; when $\lambda \neq 0$, it does not. To see this we note that, when $\omega_{21} = 0$, B_1 and B_2 are independent, and when $\lambda = 0$, we have

$$\left(\int_0^1 B_2 B_2'\right)^{-1} \int_0^1 B_2 dB_1 \bigg|_{F_2} \equiv N\left(0, \omega_{11\cdot 2}\left(\int_0^1 B_2 B_2'\right)^{-1}\right), \qquad \omega_{11\cdot 2} = \omega_{11}. \tag{77}$$

Upon integration with respect to the probability measure $P(G)$ on $G = \int_0^1 B_2 B_2' > 0$ that is induced by B_2, it is clear that (77) becomes

$$\int_{G>0} N(0, \omega_{11} G^{-1}) dP(G) \equiv \int_{g>0} N(0, g\omega_{11}\Omega_{22}^{-1}) dP(g), \tag{78}$$

a compound normal distribution. On the right side of (78) (which is proved in the same way as (53) of Theorem 3.2), we have

$$g = e_1'\left(\int_0^1 W_2 W_2'\right)^{-1} e_1,$$

where $e_1' = (1,0,\ldots,0)$ and $W_2 \equiv BM(I_n)$. This proves the stated result for $\lambda = 0$.

When $\lambda \neq 0$ and $\omega_{21} = 0$, the limit distribution is a convolution of (78) and $\left(\int_0^1 B_2B_2'\right)^{-1}\lambda$. In fact, conditional on F_2, we have

$$\left(\int_0^1 B_2B_2'\right)^{-1}\left(\int_0^1 B_2 dB_1 + \lambda\right)\bigg|_{F_2}$$

$$\equiv N\left(\left(\int_0^1 B_2B_2'\right)^{-1}\lambda, \omega_{11}\left(\int_0^1 B_2B_2'\right)^{-1}\right),$$

and integrating over $G > 0$, we find the unconditional distribution

$$\left(\int_0^1 B_2B_2'\right)^{-1}\left(\int_0^1 B_2 dB_1 + \lambda\right) \equiv \int_{G>0} N(G^{-1}\lambda, \omega_{11}G^{-1})dP(G). \tag{79}$$

Relation (79) does not belong to the compound normal family. It is a mean *and* covariance matrix mixture of normals, and as such it belongs to a more general family that we describe as limiting mixed Gaussian in the next section.

In the general case where $\omega_{21} \neq 0$ and $\lambda \neq 0$, we may write (following (51) above)

$$B_1 \equiv \omega_{21}'\Omega_{22}^{-1}B_2 + l_{11}W_{11},$$

where $W_1 \equiv BM(1)$, independent of B_2. Then (75) is distributionally equivalent to

$$\left(\int_0^1 B_2B_2'\right)^{-1}\int_0^1 B_2 dB_2'\Omega_{22}^{-1}\omega_{21} + \left(\int_0^1 B_2B_2'\right)^{-1}\lambda$$

$$+ l_{11}\left(\int_0^1 B_2B_2'\right)^{-1}\int_0^1 B_2 dW_1.$$

Conditional on F_2, this is

$$N(\Psi\Omega_{22}^{-1}\omega_{21} + G^{-1}\lambda, \omega_{11\cdot 2}G^{-1}),$$

where

$$\Psi = \left(\int_0^1 B_2B_2'\right)^{-1}\left(\int_0^1 B_2 dB_2'\right),$$

$$G = \int_0^1 B_2B_2'.$$

Upon integration with respect to the joint probability measure $P(\Psi, G)$, we get

$$\left(\int_0^1 B_2 B_2'\right)^{-1}\left(\int_0^1 B_2 dB_1 + \lambda\right)$$
$$\equiv \int N(\Psi\Omega_{22}^{-1}\omega_{21} + G^{-1}\lambda, \omega_{11\cdot 2}G^{-1})dP(\Psi, G), \tag{80}$$

where the integral is over $G > 0$ and $\Psi \in R^{n^2}$.

We observe that Ψ is a matrix version of the classic unit root distribution ([24, Eq. (10)]), so that (80) is to be distinguished from (79) in that it involves both mixed Gaussian and unit root elements. When $\omega_{21} \neq 0$, the latter are eliminated only by explicitly incorporating into the estimation the information on the presence of unit roots in (74). This can be achieved in various ways; for example, by the use of maximum likelihood methods on (73) and (74) jointly. This is an approach that is explored in detail in subsequent work [27]. We shall have occasion to refer to it again in Section 4.3(iv) below.

4. LIMITING MIXED GAUSSIAN (LMG) AND LIMITING GAUSSIAN FUNCTIONAL (LGF) FAMILIES

4.1 The LMG Family

The limit distributions obtained in earlier sections of this paper have a simple general form involving matrix ratios of random elements. In Section 2, the limit distributions involved functions of finite-dimensional Gaussian random elements, while in Section 3, they involved functionals of Gaussian random processes. The form of the results suggests that the criterion function underlying estimation may in each case admit a related linear-quadratic asymptotic approximation that involves the same random elements.

To fix ideas, let $\Lambda_T(h)$ denote a sample objective criterion used in the estimation of a parameter vector $\theta \in R^n$ and suitably centered and scaled so that its argument h measures scaled deviations from some fixed parameter value θ_0, say. The examples given below will make this formulation more transparent. Optimization of Λ_T leads to an optimization estimator $\hat\theta$ and the associated deviation is $\hat h_T = \delta_T^{-1}(\hat\theta - \theta_0)$ for some sequence of (diagonal matrix) scale factors δ_T. When $\hat\theta$ is a consistent estimator of θ_0, we have $\|\delta_T\| \to 0$, but for estimators whose elements converge with probability zero we can set $\delta_T = I_n$ for all T.

Following a suggestion of a referee, we shall call $\{\Lambda_T(h) : h \in R^n\}$ a limiting mixed Gaussian (LMG) family if (C4) and (C5) hold as $T \to \infty$:

(C4) $\quad \Lambda_T(h) - \left[h'(S_T^{1/2}Z_T + \lambda) - \frac{1}{2}h'V_T h\right] \underset{p}{\to} 0,$

where λ is a constant vector and

(C5) $\quad (Z_T, S_T, V_T) \Rightarrow (Z, S, V),$

with $Z \equiv N(0, I_n)$, Z independent of (S, V) and $S > 0$, $V > 0$ a.s.
In view of (C5), we have

$$\Lambda_T(h) \Rightarrow \Lambda(h) = h'(S^{1/2}Z + \lambda) - \frac{1}{2}h'Vh. \tag{81}$$

Since

$$\hat{h}_T = \underset{h}{\operatorname{argmax}}\, \Lambda_T(h),$$

we obtain

$$\begin{aligned} \hat{h}_T \Rightarrow \underset{h}{\operatorname{argmax}}\, \Lambda(h) &= V^{-1}(S^{1/2}Z + \lambda), \\ &\equiv \int_{S>0, V>0} (\lambda V^{-1}, V^{-1}SV^{-1})\, dP(S, V), \end{aligned} \tag{82}$$

where $P(S, V)$ is the joint probability measure of (S, V). With one exception, which we shall discuss later, the LMG family and the limit distribution (82) include all of the asymptotic results obtained in Sections 2 and 3. Note that (82) is, in general, both a mean and a covariance matrix mixture of normals. But when $\lambda = 0$, it reduces to a simple covariance matrix mixture.

Quadratic approximations, such as that implied by (C4), are in no way new. They appear in a general form in LeCam [11; 12, p. 210] in the context of log-likelihood ratio criteria and in the work of Jeganathan [8], Davies [5], and Basawa and Scott [1] on locally asymptotically mixed normal (LAMN) families, again in the context of likelihood objective functions. The LAMN theory, in particular, involves a linear-quadratic approximation condition that is quite closely related to (C4). It will be helpful to our discussion if we give the conditions of the LAMN theory here. We shall use the treatment in Jeganathan [8] as the basis of our outline.

Let $\{E_T\}_{T=1}^{\infty} = \{\Omega_T, A_T, P_{\theta,T}; \theta \in \Theta\}_{T=1}^{\infty}$ be a sequence of probability spaces (or experiments) whose probability measures are indexed by $\theta \in R^n$. We denote the log-likelihood ratio by

$$\Lambda_T(\psi, \theta) = \ln(dP_{\psi,T}/dP_{\theta,T}),$$

where the symbol "dP/dQ" signifies the Radon–Nikodym derivative of the Q-continuous part of P with respect to Q. Then from Jeganathan [8], $\{E_T\}$ satisfies LAMN condition at $\theta = \theta^0$ if, as $T \to \infty$,

(C6) $\quad \Lambda_T(\theta^0 + \delta_T h, \theta^0) - [h'S_T(\theta^0)^{1/2}Z_T(\theta^0) - (1/2)h'S_T(\theta^0)h] \underset{p}{\to} 0,$

under $P_{\theta^0, T}$,

and

(C7) $\quad (Z_T(\theta^0), S_T(\theta^0)) \Rightarrow (Z, S(\theta^0)), \qquad$ under $P_{\theta^0,T}$,

where $Z_T(\theta^0)$ $(n \times 1)$ and $S_T(\theta^0)$ $(n \times n)$ are A_T-measurable matrices with $S_T(\theta^0) > 0$ a.s. $(P_{\theta^0,T})$, $h \in R^n$ is any constant vector, and δ_T is a sequence of matrices for which $\|\delta_T\| \to 0$ as $T \to \infty$. The limit random matrix $S(\theta^0) > 0$ a.s., the limit random vector $Z \equiv N(0, I_n)$ and Z is independent of S.

Now let $\hat{\theta}$ be the maximum likelihood estimate of θ^0 and set

$$\hat{h}_T = \underset{h}{\operatorname{argmax}}\ \Lambda_T(\theta^0 + \delta_T h, \theta^0). \tag{83}$$

Then, $\hat{h}_T = \delta_T^{-1}(\hat{\theta} - \theta_0)$. From (C6) and (C7), we have

$$\begin{aligned} \Lambda_T(\theta^0 + \delta_T h, \theta^0) &\Rightarrow h'S(\theta^0)^{1/2}Z - \frac{1}{2}\, h'S(\theta^0)h, \\ &= \Lambda(h) \text{ (say).} \end{aligned} \tag{84}$$

Let

$$\hat{h} = \underset{h}{\operatorname{argmax}}\ \Lambda(h). \tag{85}$$

We deduce from (83)–(85) that $\hat{h}_T \Rightarrow \hat{h}$ or equivalently

$$\delta_T^{-1}(\hat{\theta} - \theta^0) \Rightarrow S(\theta^0)^{-1/2}Z \equiv \int_{S>0} N(0, S(\theta^0)^{-1})dP(S). \tag{86}$$

In this case, therefore, the limit distribution is a covariance matrix mixture of normals and is a good deal simpler than (82) when $\lambda \neq 0$.

Note that the quadratic approximation in (C4) is the same as that in (C6) when $\lambda = 0$ and $V_T = S_T$. These are important additional elements in the LAMN theory. First, maximum likelihood takes into account all information in the system and for correctly specified likelihoods this ensures that $\lambda = 0$ (no bias effects). Second, when $V_T = S_T$, we have in the limit

$$E\left[\exp\left\{h'S^{1/2}Z - \frac{1}{2}\, h'Sh\right\}\right] = 1,$$

and this ensures that the sequences of measures $\{P_{\theta_0,T}\}$ and $\{P_{\theta_0+\delta_T h,T}\}$ are contiguous for all h (see, for example [31, pp. 98–99]). Finally, Jeganathan [8, Theorem 3] shows that the contiguity of these sequences and the weak convergence

$$S_T(\theta^0) \Rightarrow S(\theta^0), \qquad \text{under } P_{\theta^0+\delta_T h,T}, \tag{87}$$

are necessary and sufficient conditions for (86). We shall return to this point later in Section 4.4.

The class of estimators that we wish to consider is larger than maximum likelihood. We also wish to allow for situations of misspecification which will of their very nature induce bias effects. For these reasons, we shall focus our attention on the LMG family and take as examples some of the estimators considered earlier in the paper.

Example 1 (Unidentified Structural Estimation). To begin let

$$J_T(\beta) = (y_1 - Y_2\beta)'(P_H - P_{Z_1})(y_1 - Y_2\beta).$$

Then the case of the structural equation IV estimator $\hat{\beta}$ given in (14) satisfies

$$\hat{\beta} = \operatorname{argmin} J_T(\beta).$$

Define

$$\Lambda_T(h) = -(1/2)(J_T(h) - J_T(0)),$$

and note that

$$\hat{h}_T = \underset{h}{\operatorname{argmax}}\, \Lambda_T(h) = \hat{\beta}.$$

Now

$$\Lambda_T(h) = h'Y_2'(P_H - P_{Z_1})y_1 - (1/2)h'Y_2'(P_H - P_{Z_1})Y_2h, \tag{88}$$

and we write

$$\begin{aligned} Y_2'(P_H - P_{Z_1})y_1 &= Y_2'D_1D_1'y_1 \\ &= [(Y_2'D_1D_1'Y_2)^{1/2}][(Y_2'D_1D_1'Y_2)^{-1/2}Y_2'D_1D_1'y_1] \\ &= S_T^{1/2}Z_T, \\ Y_2'(P_H - P_Z)Y_2 &= Y_2'D_1D_1'Y_2 = S_T. \end{aligned}$$

When (5) holds, we know from Lemma 2.3 that

$$Z_T = (Y_2'D_1D_1'Y_2)^{-1/2}Y_2'D_1D_1'y_1 \Rightarrow Z \equiv N(0, I_n),$$

$$S_T = Y_2'D_1D_1'Y_2 \Rightarrow S \equiv W_n(k_3, I),$$

and Z and S are independent. It follows that

$$\Lambda_T(h) = h'S_T^{1/2}Z_T - \frac{1}{2}h'S_Th, \tag{89}$$

and this satisfies (C4) and (C5) with $V_T = S_T$ and $\lambda = 0$. Thus, $\Lambda_T(h)$ belongs to the LMG family with $\delta_T = I$.

Example 2 (Partially Identified Structural Estimation). Consider the case in Section 2 where $\Pi_2 = 0$ and $\Pi_1 = 0$. Here, β is unidentified and $\gamma = \pi_1$ is identified in (1). Let

$$J_T(\beta,\gamma) = (y_1 - Y_2\beta - Z_1\gamma)'P_H(y_1 - Y_2\beta - Z_1\gamma),$$

and define

$$\begin{aligned}\Lambda_T(h,\pi_1 + l/\sqrt{T}) &= -(1/2)\{J_T(h,\pi_1 + l/\sqrt{T}) - J_T(0,\pi_1)\},\\ &= [h',l']\begin{bmatrix} Y_2'P_H \\ T^{-1/2}Z_1'P_H\end{bmatrix}(y_1 - Z_1\pi_1)\\ &\quad - (1/2)[h',l']\begin{bmatrix} Y_2'P_HY_2 & T^{-1/2}Y_2'P_HZ_1 \\ T^{-1/2}Z_1'P_HY_2 & T^{-1}Z_1'Z_1\end{bmatrix}\begin{bmatrix} h \\ l\end{bmatrix}.\end{aligned}$$

Write $P_H = DD'$ as before and define

$$W_{1T} = D'(y_1 - Z_1\pi_1) = D'v_1 \Rightarrow X_1,$$

$$W_{2T} = \begin{bmatrix} Y_2'D \\ T^{-1/2}Z_1'D\end{bmatrix} \Rightarrow \begin{bmatrix} X_2 \\ \underline{\mathrm{M}}_{11}\end{bmatrix},$$

where

$$[X_1, X_2'] \equiv N_{k_*,m}(0,I),$$

$$\underline{\mathrm{M}}_{11} = [0, M_{11}^{1/2}].$$

Also define

$$Z_T = (W_{2T}W_{2T}')^{-1/2}W_{2T}W_{1T},$$

$$S_T = W_{2T}W_{2T}',$$

and note that

$$Z_T \Rightarrow Z \equiv N(0,I),$$

$$S_T \Rightarrow S = \begin{bmatrix} X_2X_2' & X_2\underline{\mathrm{M}}_{11}' \\ \underline{\mathrm{M}}_{11}X_2' & M_{11}\end{bmatrix}.$$

Then

$$\Lambda_T(h,\pi_1 + l/\sqrt{T}) = (h',l')S_T^{1/2}Z_T - (1/2)(h',l')S_T\begin{pmatrix} h \\ l\end{pmatrix}.$$

We see that the criterion Λ_T belongs to the LMG family with $V_T = S_T$, $\lambda = 0$, and $\delta_T = I$ as in Example 1. We have

$$\begin{aligned}\Lambda_T &\Rightarrow [h',l']S^{1/2}Z - (1/2)[h',l']S\begin{bmatrix} h \\ l\end{bmatrix},\\ &= \Lambda(h,l) \text{ (say)}.\end{aligned}$$

Noting that $\gamma = \pi_1$, we write

$$\begin{bmatrix} \hat{h}_T \\ \hat{l}_T \end{bmatrix} = \begin{bmatrix} \hat{\beta} \\ \sqrt{T}(\hat{\gamma} - \gamma) \end{bmatrix} = \operatorname{argmax} \Lambda_T,$$

and

$$\begin{bmatrix} \hat{h}_T \\ \hat{l}_T \end{bmatrix} \Rightarrow \begin{bmatrix} \hat{h} \\ \hat{l} \end{bmatrix} = \operatorname{argmax} \Lambda.$$

The vector $(\hat{h}', \hat{l}')$ satisfies the system

$$\begin{bmatrix} X_2X_2' & X_2\underline{\mathbf{M}}_{11}' \\ \underline{\mathbf{M}}_{11}X_2' & \underline{\mathbf{M}}_{11}\underline{\mathbf{M}}_{11}' \end{bmatrix} \begin{bmatrix} \hat{h} \\ \hat{l} \end{bmatrix} = \begin{bmatrix} X_2X_1 \\ \underline{\mathbf{M}}_{11}X_1 \end{bmatrix},$$

and thus

$$\hat{h} = (X_2 Q_{\underline{\mathbf{M}}_{11}'} X_2)^{-1} X_2 Q_{\underline{\mathbf{M}}_{11}'} X_1,$$

$$\equiv \int_{S>0} N(0, S^{-1}) dP(S),$$

with

$$S = X_2 Q_{\underline{\mathbf{M}}_{11}'} X_2' \equiv W_n(k_3, I),$$

corresponding to earlier results in Section 2. Similarly,

$$\hat{l} = (\underline{\mathbf{M}}_{11} Q_{X_2'} \underline{\mathbf{M}}_{11}')^{-1} (\underline{\mathbf{M}}_{11} Q_{X_2'} X_1),$$

$$\equiv \int_{V>0} N(0, V^{-1}) dP(V),$$

where

$$V = \underline{\mathbf{M}}_{11} Q_{X_2'} \underline{\mathbf{M}}_{11}' = M_{11} K K' M_{11}',$$

where K is uniform on the Stiefel manifold $V_{k_*-n,k_*} = \{K(k_* \times k_* - n) : K'K = I_{k_*-n}\}$. Note that we may also write

$$\hat{l} = (\underline{\mathbf{M}}_{11}\underline{\mathbf{M}}_{11}')^{-1}\underline{\mathbf{M}}_{11}X_1 - (\underline{\mathbf{M}}_{11}\underline{\mathbf{M}}_{11}')^{-1}\underline{\mathbf{M}}_{11}X_2'\hat{h},$$

$$= (\underline{\mathbf{M}}_{11}\underline{\mathbf{M}}_{11}')^{-1}\underline{\mathbf{M}}_{11}[X_1, X_2'] \begin{bmatrix} 1 \\ \hat{h} \end{bmatrix}.$$

Since $\hat{h}$ depends on $Q_{\underline{\mathbf{M}}_{11}'}[X_1, X_2']$, which is independent of $\underline{\mathbf{M}}_{11}[X_1, X_2']$, we deduce that

$$\hat{l} \equiv \int N(0, (1 + \hat{h}'\hat{h})(\underline{\mathbf{M}}_{11}\underline{\mathbf{M}}_{11}')^{-1}) dP(\hat{h}),$$

$$= \int N(0, (1 + \hat{h}'\hat{h}) M_{11}^{-1}) dP(\hat{h}),$$

again corresponding to earlier results in Section 2 (specifically, (26) with $R_1 = I_{k_1}$).

Example 3 (Spurious Regressions). We take the case of (49) above. Using the notation of Section 3.1, we have

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}}\, J_T(\beta),$$

with

$$J_T(\beta) = T^{-2}\Sigma_1^T(y_t - \beta' x_t)^2 = T^{-2}(y - X\beta)'(y - X\beta)$$

in conventional regression notation. Define $\bar{\beta} = \Omega_{22}^{-1}\omega_{21}$ and set

$$\begin{aligned}\Lambda_T(h) &= -(1/2)\{J_T(\bar{\beta} + h) - J_T(\bar{\beta})\},\\ &= T^{-2}\{h'X'(y - X\bar{\beta}) - (1/2)h'X'Xh\}.\end{aligned} \tag{90}$$

In view of (C3), we have

$$T^{-2}X'(y - X\bar{\beta}) \Rightarrow \int_0^1 B_2(B_1 - B_2\bar{\beta}),$$

$$T^{-2}X'X \Rightarrow \int_0^1 B_2B_2'$$

and from Lemma 3.1,

$$\begin{aligned}\int_0^1 B_2(B_1 - B_2\bar{\beta})|_{F_2} &\equiv l_{11}\int_0^1 B_2W_1 \equiv N\left(0, \omega_{11\cdot 2}\int_0^1\int_0^1 B_2(r)(r\wedge s)B_2'(s)\right),\\ &\equiv \left(\int_0^1\int_0^1 B_2(r)(r\wedge s)B_2'(s)\right)^{1/2} N(0, \omega_{11\cdot 2}I).\end{aligned}$$

Moreover, simple calculations show that

$$T^{-4}X'A_TX \Rightarrow \int_0^1\int_0^1 B_2(r)(r\wedge s)B_2(s)'drds,$$

where

$$A_T = [(\min(i,j))_{ij}]_{T\times T},$$

and

$$(T^{-4}X'A_TX)^{-1/2}(T^{-2}X'(y - X\bar{\beta})) \Rightarrow N(0, \omega_{11\cdot 2}I),$$

where the limit distribution is independent of B_2. We may therefore write (90) in the form

$$\begin{aligned}\Lambda_T(h) &= h'\{(T^{-4}X'A_TX)^{1/2}[(T^{-4}X'A_TX)^{-1/2}(T^{-2}X'(y - X\bar{\beta}))]\}\\ &\quad - (1/2)h'(T^{-2}X'X)h,\\ &= h'S_T^{1/2}Z_T - (1/2)h'V_Th \text{ (say).}\end{aligned} \tag{91}$$

Here

$$Z_T = (T^{-4}X'A_TX)^{-1/2}(T^{-2}X'(y - X\bar{\beta})) \Rightarrow Z \equiv N(0, \omega_{11 \cdot 2}I), \tag{92}$$

$$S_T \Rightarrow S \equiv \int_0^1 \int_0^1 B_2(r)(r \wedge s)B_2'(s), \tag{93}$$

$$V_T \Rightarrow V \equiv \int_0^1 B_2B_2', \tag{94}$$

and Z is independent of (S, V). Clearly, $\Lambda_T(h)$ belongs to the LMG family and satisfies conditions (C4) and (C5) with $\delta_T = I$.

As shown earlier in Theorem 3.2, the limit distribution of $\hat{\beta}$ is mixed normal. Indeed, from (91)–(94), we have

$$\Lambda_T(h) \Rightarrow \Lambda(h) = h'S^{1/2}Z - (1/2)h'Vh, \tag{95}$$

and setting $\hat{h} = \text{argmax } \Lambda(h)$, we obtain

$$\hat{\beta} - \bar{\beta} = \hat{h}_T \Rightarrow \hat{h} = V^{-1}S^{1/2}Z \equiv \int_{V(B_2)>0} N(0, \omega_{11 \cdot 2}V(B_2))dP(V(B_2)), \tag{96}$$

where

$$V(B_2) = \left(\int_0^1 B_2B_2'\right)^{-1} \left(\int_0^1 \int_0^1 B_2(r)(r \wedge s)B_2(s)'\right) \left(\int_0^1 B_2B_2'\right)^{-1}.$$

In place of (95), we may write the weak convergence directly in terms of functionals of Brownian motion, as in Section 3. Thus,

$$\begin{aligned} \Lambda_T(h) &= h'(T^{-2}X'(y - X\bar{\beta})) - (1/2)h'(T^{-2}X'X)h, \\ &\Rightarrow h'\int_0^1 B_2(B_2 - B_2\bar{\beta}) - (1/2)h'\left(\int_0^1 B_2B_2'\right)h = \Lambda(h). \end{aligned} \tag{97}$$

This representation of $\Lambda(h)$ is suggestive. It indicates the possibility of extending the LMG family of limit distributions in terms of Gaussian functionals. Indeed, the form of (97) may plausibly be interpreted as a continuous stochastic process extension of (81) or (84) where the limits are functions of finite-dimensional random elements. The need for such extensions will become more apparent below.

4.2 The LGF Family

Following up this idea of extending the LMG family, we shall say that the criterion function $\Lambda_T(h)$ satisfies the *limiting Gaussian functional* (LGF) condition if

(C8) $\quad \Lambda_T(h) - \{h'W_T - (1/2)h'S_Th\} \xrightarrow[p]{} 0$

for some n-vector W_T and $n \times n$ matrix S_T; and

$$\text{(C9)} \qquad (W_T, S_T) \Rightarrow \left(\int_0^1 MdN + \lambda, \int_0^1 MM'\right).$$

In (C9), the elements of M are square integrable and lie in $D[0,1]$, the space of all real-valued functions on [0,1] that are right continuous and have finite left limits; $N(r)$ is a Gaussian random function whose sample paths lie in $C[0,1]$; and λ is a constant vector.

The following special cases will help to clarify the relationship between the LGF and LMG families. We suppose that $\Lambda_T(h)$ is LGF with limit function

$$\Lambda(h) = h'\left(\int_0^1 MdN + \lambda\right) - \frac{1}{2} h'\left(\int_0^1 MM'\right)h. \tag{98}$$

(i) If $N(p) = \int_0^p G(r)dr$, where $G(r)$ is a Gaussian process with covariance kernel matrix $K(r,s)$ and is independent of M, then LGF reduces to LMG with

$$S = \int_0^1 \int_0^1 M(r)K(r,s)M(s)'drds,$$

and

$$V = \int_0^1 MM'.$$

In this way LMG may be regarded as a special case of LGF.

(ii) If M and N are independent with $N \equiv BM(I)$, then LGF reduces to LMG with

$$S = V = \int_0^1 MM'.$$

When $\lambda = 0$, this corresponds also with (84).

(iii) LGF need not always reduce to LMG. For example, if $N = B_1$, $M = B_2$, and $B = (B_1', B_2')' \equiv BM(\Omega)$ with $\Omega > 0$, then

$$\begin{aligned} \Lambda(h) &= h'\left(\int_0^1 B_2 dB_1 + \lambda\right)\frac{1}{2} h'\left(\int_0^1 B_2 B_2'\right)h, \\ &= \left[h'\left(\int_0^1 B_2 dWl_{11} + \lambda\right) - \frac{1}{2} h'\left(\int_0^1 B_2 B_2'\right)\right] + h'\int_0^1 B_2 dB_2' \Omega_{22}^{-1} \omega_{21}. \end{aligned} \tag{99}$$

The term in square parentheses belongs to the LMG family as in (ii) above. Thus, when $\omega_{21} = 0$, $\Lambda_T(h)$ is LMG. But when $\omega_{21} \neq 0$, the linear term in (99) cannot be included in the LMG family. This is precisely what happens in the general case of a linear least-squares cointegrating regression as in (75) above. To see this, note that when $B_2 \equiv BM(1)$, we have

$$\int_0^1 B_2 dB_2 = \frac{1}{2}\,(B_2(1)^2 - 1) \equiv \frac{1}{2}\,(\chi_1^2 - 1),$$

whose distribution is skewed, whereas the distribution of $S^{1/2}Z$ in the linear term of (81) is always symmetric. We shall consider other examples where this arises below.

4.3 Applications of the LGF Family

We shall now look at some specific applications of the LGF family. The first of these also fall within the LMG family but are worth considering because their treatment is instructive and helps to demonstrate the flexibility of LGF.

(i) *Partially spurious regressions:* In the notation of Section 3.2, define

$\underline{\beta}_2 = \underline{\Omega}_{22}^{-1}\underline{\omega}_{21}$.

Let

$J_T(\beta_1,\beta_2) = T^{-2}(y - x_1\beta_1 - X_2\beta_2)'(y - x_1\beta_1 - X_2\beta_2),$

and

$\Lambda_T(h,l) = -(1/2)\{J_T(\beta_1 + T^{-1/2}h, \underline{\beta}_2 + l) - J_T(\beta_1, \underline{\beta}_2)\}.$

Then

$$\Lambda_T(h,l) \Rightarrow [h,l']\begin{bmatrix} \int_0^1 r(\underline{B}_1 - \underline{B}_2'\underline{\beta}_2) \\ \int_0^1 \underline{B}_2(\underline{B}_1 - \underline{B}_2'\underline{\beta}_2) \end{bmatrix}$$

$$- (1/2)\,[h,l']\begin{bmatrix} \int_0^1 r^2 & \int_0^1 r\underline{B}_2' \\ \int_0^1 \underline{B}_2 r & \int_0^1 \underline{B}_2\underline{B}_2' \end{bmatrix}\begin{bmatrix} h \\ l \end{bmatrix},$$

$$= \Lambda(h,l).$$

We now set $\lambda = 0$,

$$M(r) = \begin{bmatrix} r \\ \underline{B}_2(r) \end{bmatrix},$$

and

$$N(r) = \int_0^1 (\underline{B}_1(s) - \underline{B}_2'(s)\underline{\beta}_2)ds.$$

This example falls within the LGF framework of (C8) and (C9). Note that by optimizing $\Lambda(h)$, we obtain directly

$$\begin{bmatrix} \sqrt{T}(\hat{\beta}_1 - \beta_1) \\ \hat{\beta}_2 - \beta_2 \end{bmatrix} \Rightarrow \begin{bmatrix} \hat{h} \\ \hat{l} \end{bmatrix} = \operatorname{argmax} \Lambda(h,l)$$

$$= \begin{bmatrix} \left(\int_0^1 \zeta^2\right)^{-1} \int_0^1 \zeta \underline{B}_1 \\ \left(\int_0^1 \eta\eta'\right)^{-1} \int_0^1 \eta(\underline{B}_1 - \underline{B}_2'\underline{\beta}_2) \end{bmatrix}$$

consistent with the earlier results (66) and (68).

(ii) *Structural estimation:* We shall take the case of the IV estimator $\hat{\beta}$ given in (14). As seen in (88), we have

$$\Lambda_T(h) = h'Y_2'(P_H - P_{Z_1})y_1 - (1/2)h'Y_2'(P_H - P_{Z_1})h,$$

and in place of (89), we may write this as

$$\Lambda_T(h) = h'S_TZ_T - \frac{1}{2}h'S_TS_T'h, \tag{100}$$

$$\Rightarrow h'SZ - \frac{1}{2}h'SS'h = \Lambda(h),$$

where

$$S_T = Y_2'D_1 \Rightarrow S \equiv N_{n,k_3}(0,I),$$

$$Z_T = D_1'y_1 \Rightarrow Z \equiv N(0,I_{k_3}).$$

Now partition

$$S = [S_1,S_2,\ldots,S_{k_3}], \qquad Z' = [Z_1,Z_2,\ldots,Z_{k_3}],$$

and then

$$SZ = \sum_{j=1}^{k_3} S_jZ_j \equiv \sum_1^{k_3} \int_{(j-1)/k_3}^{j/k_3} MdN = \int_0^1 MdN, \tag{101}$$

where we define

$$
\begin{aligned}
N(r) &= W(r) \equiv BM(1), \\
M(r) &= k_3^{1/2} S_1, \qquad 0 \le r < 1/k_3, \\
&= k_3^{1/2} S_2, \qquad 1/k_3 \le r < 2/k_3, \\
&\ \ \vdots \\
&= k_3^{1/2} S_k, \qquad (k_3 - 1)/k_3 \le r < 1.
\end{aligned}
$$

Note that

$$
\begin{aligned}
\int_{(j-1)/k_3}^{j/k_3} MdN &= k_3^{1/2} S_j \{ W(j/k_3) - W((j-1)/k_3) \}, \\
&\equiv S_j Z_j,
\end{aligned}
$$

where $Z_j \equiv N(0,1)$ and is independent of $S_j \equiv N(0, I_n)$. This justifies (101). We also have

$$SS' = \sum_1^{k_3} S_j S_j' = \int_0^1 MM'.$$

It follows that

$$\Lambda(h) = h' \left(\int_0^1 MdN \right) - (1/2) h' \left(\int_0^1 MM' \right) h,$$

which, together with (100), gives us an alternative way of looking at $(\Lambda_T(h), \Lambda(h))$ in terms of the LGF family.

(iii) *The Gaussian AR(1):* Let $\{X_t\}$ be generated by the AR(1)

$$X_t = \theta X_{t-1} + u_t, \tag{102}$$

where u_t is i.i.d. $N(0,1)$ and $X_0 = 0$. The asymptotic behavior of the coefficient estimator

$$\hat{\theta} = \sum_1^T X_t X_{t-1} \Big/ \sum_1^T X_{t-1}^2 \tag{103}$$

is well-known to depend on whether the model (102) is stable ($|\theta| < 1$), explosive ($|\theta| > 1$), or has a unit root ($\theta = 1$). Let θ_0 be the true value of θ in (102) and define the log-likelihood ratio

$$
\begin{aligned}
\Lambda_T(h) &= \ln\{ pdf(X;\theta)/pdf(X;\theta_0) \}, \qquad \theta = \theta_0 + \delta_T h, \\
&= -(1/2) \sum_1^T (X_t - \theta X_{t-1})^2 + (1/2) \sum_1^T (X_t - \theta_0 X_{t-1})^2, \\
&= h \left(\delta_T \sum_1^T X_{t-1} u_t \right) - (1/2) h^2 \left(\delta_T^2 \sum_1^T X_{t-1}^2 \right).
\end{aligned} \tag{104}
$$

The limiting behavior of $\Lambda_T(h)$ is also well-known and may be characterized as follows for the three distinct cases. We remark that case 1 is classical. Case 2 is covered by Basawa and Brockwell [2] and has recently been extended to explosive and partially explosive Gaussian AR(p)'s by Jeganathan [9]. Case 3 has been recently studied in detail in Phillips [24], Chan and Wei [4], and Jeganathan [10]. We have

$$\Lambda_T(h) \Rightarrow \Lambda(h).$$

Case 1 ($|\theta_0| < 1$, $\delta_T = T^{-1/2}$).

$$\Lambda(h) = hY(\theta_0)Z - (1/2)h^2Y(\theta_0)^2, \tag{105}$$

with $Z \equiv N(0,1)$, $Y(\theta_0) = (1 - \theta_0^2)^{-1/2}$.

Case 2 ($|\theta_0| > 1$, $\delta_T = (\theta_0^2 - 1)/\theta_0^T$).

$$\Lambda(h) = hYZ - (1/2)h^2Y^2, \tag{106}$$

with $Z \equiv N(0,1)$ independent of $Y \equiv N(0,1)$.

Case 3 ($\theta_0 = 1$, $\delta_T = T^{-1}$).

$$\Lambda(h) = h\left(\int_0^1 BdB\right) - (1/2)h^2\left(\int_0^1 B^2\right), \tag{107}$$

with $B(r) \equiv BM(1)$ on $C[0,1]$.

Each of these cases comes within the general LGF family defined in (C8) and (C9). To see this, let

$$1(r) = 1, \qquad 0 \le r \le 1,$$

be a constant function on $C[0,1]$. Then we have

$$M(r) = 1(r)Y(\theta_0), \qquad \text{in Case 1,}$$

$$M(r) = 1(r)Y, \qquad \text{in Case 2,}$$

$$M(r) = B(r) \equiv BM(1), \text{ in Case 3,}$$

with

$$N(r) = B(r)$$

in all three cases and where B is independent of Y in Case 2. We may then write

$$\Lambda(h) = h\left(\int_0^1 MdN\right) - (1/2)h^2\int_0^1 M^2,$$

embracing all three cases within the LGF family.

It is apparent that Cases 1 and 2 also fall within the LMG family. However, Case 3 is not covered by LMG. This is because the stochastic integral $\int_0^1 BdB$ cannot be written in the form of a simple scale mixture of normals, as required for LMG. Thus, in this unit root case there is a real need for a family that is more general than LMG.

It may be remarked at this point that, since the objective criterion $\Lambda_T(h)$ is a log-likelihood ratio and $\hat{\theta}$ is the maximum likelihood estimate (MLE), $\Lambda_T(h)$ also falls within the LAMN family in Cases 1 and 2. These cases have been studied earlier [2,9]. However, to the best of our knowledge, no theory has until now been put forward which accommodates the unit root Case 3 as well as Cases 1 and 2. We shall examine why the unit root case is not covered by the LAMN theory more fully in Section 4.4 below.

(iv) *Cointegrating regressions:* In the notation of Section 3.3, λ is given directly by (74), $M(r) = B_2(r)$, and $N(r) = B_1(r)$. We write

$$J_T(\beta) = (y - X\beta)'(y - X\beta).$$

Then

$$\Lambda_T(h) = -(1/2)\{J_T(\beta + T^{-1}h) - J_T(\beta)\},$$

$$\Rightarrow h'\left(\int_0^1 B_2 dB_1 + \lambda\right) - (1/2)h'\left(\int_0^1 B_2 B_2'\right)h, \tag{108}$$

so that $\Lambda_T(h)$ falls directly within the framework of (C8) and (C9).

We emphasize that this result applies to the least-squares estimator $\hat{\beta}$ derived by minimizing the objective function $J_T(\beta)$. There are many other ways of estimating β in the cointegrated system (73) and (74). We remark that the full maximum likelihood estimate (MLE) of β requires complete estimation of the generating mechanism of the innovations ξ_t. Such estimation is difficult if, as is typically the case, ξ_t is modeled by a vector ARMA process for which the orders of the polynomial lags must also be estimated. However, the MLE ($\tilde{\beta}$, let us say) has powerful advantages over $\hat{\beta}$ for inferential purposes. Complete estimation removes the bias term λ in (108) and purges B_1 of its dependence on B_2 (arising directly from the endogeneity of the regressor x_t in (73)). These effects bring the log-likelihood ratio criterion for $\tilde{\beta}$, $\tilde{\Lambda}_T(h)$ (let us say), within the LAMN family. The reader is referred to a subsequent paper by the author [27] for a detailed study of this case. We should also remark that these apparently rather favorable results for the MLE $\tilde{\beta}$ arise only when (73) and (74) are estimated as specified with the unit roots of (74) explicitly incorporated. When any of these unit roots are estimated, as they can be in unrestricted vector AR or ARMA specifications for z_t, the limit theory is analogous to the AR(1) Case 3 and the LMG (and here LAMN) families no longer apply. We shall now look into the reason for this breakdown in the presence of unit roots.

4.4 Why the LGF Family Is Needed When There Are Estimated Unit Roots

Our analysis is facilitated by the very detailed study of the LAMN condition in the recent work of Jeganathan [8–10]. We shall focus our attention on the Gaussian AR(1) example considered above, since this includes cases where the LAMN condition holds (namely, $|\theta_0| < 1$ and $|\theta_0| > 1$) and where it does not ($\theta = 1$).

In [8, Theorem 3], Jeganathan gives necessary and sufficient conditions for the pair (W_T, S_T) that appear in (C8) to satisfy

$$(W_T, S_T) \Rightarrow (S^{1/2}Z, S), \tag{109}$$

where $Z \equiv N(0, I_n)$ and is independent of $S > 0$ (a.s.). In the context of the LAMN conditions (C6) and (C7), we have the parametric dependencies

$$(W_T, S_T, S) = (W_T(\theta_0), S_T(\theta_0), S(\theta_0))$$

and an underlying sequence of experiments $\{E_T\}_{T=1}^{\infty} = \{(\Omega_T, A_T, P_{\theta,T}) : \theta \in \Theta\}_{T=1}^{\infty}$. We write $\theta = \theta_0 + \delta_T h$, where δ_T is a sequence of matrices for which $\|\delta_T\| \to 0$, and we construct a sequence of associated probability measures $\{P_{\theta_0+\delta_T h, T}\}_{T=1}^{\infty}$ adjacent to the sequence $\{P_{\theta_0}\}_{T=1}^{\infty}$. Now Jeganathan [8, Theorem 3] proves that (109) applies if and only if the two following conditions hold:

(C10) $\{P_{\theta_0,T}\}, \{P_{\theta_0+\delta_T h, T}\}$ are contiguous for all $h \in R^n$,

(C11) $S_T(\theta_0) \Rightarrow S(\theta_0)$ under $P_{\theta_0+\delta_T h, T}$ for all $h \in R^n$.

Using this result, we find the following.

Proposition 4.1. *If $\{X_t\}$ is generated by the* AR(1) *given in* (102) *with $\theta_0 = 1$ and if $\theta = \theta_0 + T^{-1}h$ defines an adjacent parameter sequence with associated probability measures $\{P_{\theta_0+T^{-1}h, T}\}$, then* (C11) *fails. In particular,*

$$\begin{aligned} S_T = T^{-2}\sum_1^T X_{t-1}^2 \Rightarrow \int_0^1 J_h^2 &= S(\theta_0, h) \text{ (say)}, \\ &\neq S(\theta_0), \qquad \text{for } h \neq 0, \end{aligned} \tag{110}$$

where

$$J_h(r) = \int_0^r e^{(r-s)h} dB(s)$$

is a diffusion process on $C[0,1]$ and $B \equiv BM(1)$.

Remarks. (i) Proposition 4.1 clarifies why the LAMN condition breaks down for the unit root case $\theta_0 = 1$. In effect, there is more variability in the "random information" component S_T of $\Lambda_T(h)$ than the LAMN framework permits. This is shown directly in (110), which specifies the way in which the limiting *random information* depends on h or the extent of the deviation from $\theta_0 = 1$.

(ii) The phenomenon noted in the previous remark may be described as *variable random information*. Changes in the sequence of probability measures $\{P_{\theta_0+T^{-1}h,T}\}$ brought about by changes in h induce changes in the limiting random information measure $S(\theta_0, h)$. In effect, the quadratic approximation to $\Lambda(h)$ itself varies for different contiguous sequences $\{P_{\theta_0+T^{-1}h,T}\}$.

(iii) Recently, Jeganathan [10] has studied the AR(p) model with roots on or near the unit circle under quite a general condition on the density of the errors. His results (especially [10, Theorem 14]) are more general than (110) and are used to establish the contiguity condition (C10) and to construct an asymptotic approximation to the likelihood ratio. These results also fall within the LGF family rather than LAMN and for the same reason, namely, the failure of (C11).

5. CONCLUSION

This paper has covered a good deal of ground. Our primary aim has been to open up, for theoretical study and asymptotic analysis, models that are partially identified. The most obvious candidate for investigation in this area is structural estimation under rank condition failure, the subject of our study in Section 2. Spurious regressions present another major application, as we found in Section 3. Other examples include errors in variables systems under identification (or instrument) failure and ARMA model estimation in the presence of degenerate common factors. Similar problems can also arise in microeconometric models with endogenous regressors, such as models with self-selectivity. In such models, where two-step procedures and instrumental variables are routinely used, partial identification occurs because of instrument failures. That is, the instruments fail to satisfy what might be called the *relevance condition*. This condition requires that the asymptotic correlation matrix between the instruments and the regressors be of full rank. If the instruments fail, then the model is only partly identified and conventional asymptotics break down. Much of the ongoing literature on econometric estimation places great stress on the orthogonality condition for instrument validity. The relevance condition is equally important but is seldom discussed. In microeconometric settings, instrument failures through the breakdown of the relevance condition deserve particular attention because of the low explanatory power of so many regressions with micro data sets. Very low R^2's in the companion regressions which form the instruments in such cases point to the possibility of instrument failure and the associated breakdown of conventional asymptotics.

Our second aim has been to develop the extensions to conventional asymptotic theory which are needed to embrace partially identified systems. In most of our applications, the limit distributions come within the class of compound normal distributions and are simply represented as covariance matrix or scalar mixtures of normals. We have put forward two limit theories for optimization estimators: one based on the LMG conditions (C4) and (C5) and the other based on the LGF conditions (C8) and (C9). The LMG and LGF conditions are very similar but they differ in that, in the limit, the latter involves functionals of Banach-valued random elements whereas the former involves functions of finite-dimensional random vectors. The LGF theory seems to have a particularly wide range of interesting applications including models in which there are estimated unit roots.

REFERENCES

1. Basawa, I.V. & D J. Scott. *Asymptotic optimal inference for nonergodic models.* New York: Springer Verlag, 1983
2. Basawa, I V & P.J Brockwell. Asymptotic conditional inference for regular nonergodic models with an application to autoregressive processes. *Annals of Statistics* 12 (1984): 161–171
3. Billingsley, P. *Convergence of probability measures.* New York: Wiley, 1968
4. Chan, N.H. & C.Z. Wei. Limiting distributions of least-squares estimates of unstable autoregressive processes. *Annals of Statistics* 16 (1988): 367–401
5. Davies, R.B. Asymptotic inference when the amount of information is random. In L.M. LeCam & R.A. Olshen (eds.), *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer,* Vol. II, Wadsworth Inc., 1986.
6. Engle, R.F. & C.W J. Granger. Cointegration and error correction: representation, estimation, and testing. *Econometrica* 55 (1987): 251–276.
7. Hall, P. & C.C. Heyde. *Martingale limit theory and its application.* New York: Academic Press, 1980.
8. Jeganathan, P. An extension of a result of L. LeCam concerning asymptotic normality. *Sankhya Series A* 42 (1980) 146–160.
9. Jeganathan, P. On the strong approximation of distributions of estimators in linear stochastic models, I and II: Stationary and explosive models. University of Michigan, mimeographed, 1986
10. Jeganathan, P. On the asymptotic behavior of least-squares estimators in AR time series with roots near the unit circle: Parts I and II. University of Michigan Technical Report #150, 1987.
11. LeCam, L. Locally asymptotically normal families of distributions. *University of California Publications in Statistics* 3 (1960): 37–98.
12. LeCam, L *Asymptotic methods in statistical decision theory.* New York: Springer, 1986.
13. Lehmann, E *Testing statistical hypotheses.* New York: Wiley (2nd Ed.), 1986.
14. Muirhead, R. *Aspects of multivariate statistical theory.* New York: Wiley, 1982.
15. Park, J.Y. & P.C.B. Phillips. Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory* 4(1988): 468–497.
16. Park, J.Y. & P.C.B. Phillips. Statistical inference in regressions with integrated processes: Part 2. *Econometric Theory* 5 (1989): 95–131.
17. Phillips, P.C B. Small sample distribution theory in econometric models of simultaneous equations. Cowles Foundation Discussion Paper No. 617, 1982

18. Phillips, P.C.B. Exact small-sample theory in the simultaneous equations model. In M.D. Intriligator & Z. Griliches (eds.), *Handbook of Econometrics*, Chapter 8 and pp. 449–516. Amsterdam: North-Holland, 1983.
19. Phillips, P.C.B. The exact distribution of LIML: I. *International Economic Review* 25 (1984): 249–261.
20. Phillips, P.C.B. The exact distribution of exogenous variable coefficient estimators. *Journal of Econometrics* 26 (1984): 387–398.
21. Phillips, P.C.B. The exact distribution of LIML: II. *International Economic Review* 26 (1985): 21–36.
22 Phillips, P.C.B. A nonnormal limiting distribution. *Econometric Theory* 1 (1985): 145
23. Phillips, P.C.B. Understanding spurious regressions in econometrics. *Journal of Econometrics* 33 (1986): 311–340.
24. Phillips, P.C.B. Time series regression with a unit root. *Econometrica* 55 (1987): 277–301.
25. Phillips, P.C.B. Asymptotic expansions in nonstationary vector autoregressions. *Econometric Theory* 3 (1987): 45–68.
26. Phillips, P.C.B. Weak convergence to the matrix stochastic integral $\int_0^1 BdB'$. *Journal of Multivariate Analysis* 24 (1988): 252–264.
27. Phillips, P.C.B. Optimal inference in cointegrated systems. Cowles Foundation Discussion Paper No. 866, March 1988.
28. Phillips, P.C.B. Towards a unified asymptotic theory of autoregression. *Biometrika* 74 (1987): 535–547.
29. Phillips, P.C.B. Regression theory for near integrated time series. *Econometrica* 56 (1988): 1021–1044.
30. Phillips, P.C.B. & S.N. Durlauf. Multiple time series regression with integrated processes. *Review of Economic Studies* 53 (1986): 473–496.
31. Prakasa Rao, B.L.S. *Asymptotic theory of statistical inference*. New York: Wiley, 1987.
32 Stock, J.H. Asymptotic properties of least-squares estimators of cointegrating vectors. *Econometrica* 55 (1987): 1035–1056.

APPENDIX

Proof of Theorem 2.1. We first write $P_H - P_{Z_1} = D_1 D_1'$, where D_1 is a $T \times k_3$ matrix of orthonormal vectors spanning $R(H) \cap R(Z_1)^{\perp}$; for example,

$$D_1 = Q_{Z_1} Z_3 (Z_3' Q_{Z_1} Z_3)^{-1/2}.$$

Then it is simple to deduce that

$$(Y_2' D_1 D_1' Y_2)^{-1/2} Y_2' D_1 D_1' y_1 \,|_{Y_2} \equiv N(0, I_n),$$

and this is also the unconditional distribution since it is independent of Y_2. Furthermore,

$$S = Y_2' D_1 D_1' Y_2 \equiv W_n(k_3, I),$$

so that

$$\hat{\beta} \equiv [W_n(k_3, I)]^{-1/2} N(0, I),$$

$$\equiv \int_{S>0} N(0, S^{-1}) pdf(S) dS,$$

as required for (16). We now note that the conditional characteristic function of $\hat{\beta}$ given S is

$$cf(t) = \exp\left\{-\frac{1}{2} t' S^{-1} t\right\},$$

$$= \exp\left\{-\frac{1}{2} t' t\, h' S^{-1} h\right\},$$

where $h = t/(t't)^{1/2}$. Set $z = h'S^{-1}h$ and note that

$$1/z = (h'S^{-1}h)^{-1} \equiv W_1(k_3 - n + 1, 1) \equiv \chi^2_{k_3 - n + 1},$$

so that

$$\hat{\beta} = \int_{z>0} N(0, zI) pdf(z) dz,$$

as required for (17). Equations (18) and (19) now follow directly from standard multivariate theory (e.g., [14, p. 33]).

To prove (b), we note from (15) that

$$\hat{\gamma}_1 = R_1'\hat{\gamma} = R_1'\pi_1 + R_1'(Z_1'Z_1)^{-1} Z_1' V \begin{pmatrix} 1 \\ -\hat{\beta} \end{pmatrix}.$$

Since $Z_1'V$ and $\hat{\beta} \equiv r$ are independent, we have

$$\hat{\gamma}_1|_r \equiv N(R_1'\pi_1, (1 + r'r) R_1'(Z_1'Z_1)^{-1} R_1),$$

so that

$$\hat{\gamma}_1 \equiv \int_{R^n} N(\gamma_1, (1 + r'r) G_1) pdf(r),$$

as required for (21). We now transform $r \to (m, h)$ using the decomposition $r = hm^{1/2}$ with $m = r'r$ and $h = r/(r'r)^{1/2}$. The measure transforms as

$$dr = (1/2) m^{n/2-1} dm (dh),$$

where (dh) is the invariant measure on the sphere $S_n = \{h : h'h = 1\}$. We obtain

$$\hat{\gamma}_1 \equiv \int_{m>0} N(\gamma_1, (1 + m) G_1) \frac{(c/2) m^{n/2-1}}{(1 + m)^{(k_3+1)/2}} dm \int_{S_n} (dh),$$

where the constant c is given in (20) and

$$\int_{S_n} (dh) = 2\pi^{n/2} / \Gamma\left(\frac{n}{2}\right).$$

This leads immediately to (22).

To prove (c), we note that

$$\hat{\gamma}_2 = R_2'\hat{\gamma} = R_2'\pi_1 - R_2'\Pi_1 r + R_2'(Z_1'Z_1)^{-1}Z_1'V\begin{pmatrix}1\\ -r\end{pmatrix},$$

so that

$$\hat{\gamma}_2 \equiv \int_{R^n} N(R_2'\pi_1 - R_2'\Pi_1 r,\ (1 + r'r)G_2)\,pdf(r)dr,$$

giving (24) as stated. Part (d) follows directly from (b) and (c). ∎

Proof of Corollary 2.2. Part (a) follows from (17) since the distribution of r is independent of T. Part (b) follows from (22) by noting that

$$T^{-1}G_1 \to \bar{G}_1$$

under (C2). Parts (c) and (d) then follow from (a) and (b). ∎

Proof of Lemma 2.3. Take the scalar case with

$$d'v = \sum_{j=1}^{T} d_{Tj}v_j, \qquad d'd = 1.$$

Let $X_{Tj} = d_{Tj}v_j$, and define the system of σ-fields $F_{Ti} = \sigma(X_{T1},\ldots,X_{Ti})$, $i = 1,\ldots,T$. Then (S_{Ti},F_{Ti}) with $S_{Ti} = \sum_{j=1}^{i} X_{Tj}$ is a martingale array. Its conditional variance is

$$V_T = \sum_{j=1}^{T} E(X_{Tj}^2|F_{Tj-1}) = \sum_{1}^{T} d_{Tj}^2 = 1, \tag{A1}$$

and for $\epsilon > 0$, we have

$$\begin{aligned}
&\sum_{j=1}^{T} E\{X_{Tj}^2 I(|X_{Tj}| > \epsilon)|F_{Tj-1}]\\
&\quad = \sum_{1}^{T} d_{Tj}^2 E\left\{v_j^2 I\left(|v_j| > \frac{\epsilon}{|d_{Tj}|}\right)|F_{Tj-1}\right\},\\
&\quad \le \left(\sum_{1}^{T} d_{Tj}^2\right) E\left\{v_1^2 I\left(|v_1| > \frac{\epsilon}{\max_j |d_{Tj}|}\right)\right\},\\
&\quad = E\left\{v_1^2 I\left(|v_1| > \frac{\epsilon}{\max_j |d_{Tj}|}\right)\right\},\\
&\quad \to 0
\end{aligned} \tag{A2}$$

as $T \to \infty$ since

$$\max_j |d_{Tj}| \to 0, \qquad \text{as } T \to \infty.$$

(Note that under (C2), the elements of D are $O(T^{-1/2})$.) In view of (A1), $V_T \to 1$ a.s., and in view of (A2), the conditional Lindeberg condition is satisfied. Thus, $d'v \Rightarrow N(0,1)$ by the martingale central limit theorem (e.g., [7, p. 58]). The matrix case is handled in a similar way by treating an arbitrary linear combination of the elements of $D'V$. The stated result then follows by the Cramér–Wold device. ∎

Proof of Theorem 2.4. From the proof of Theorem 2.1, it is clear that we can write $\hat{\beta} = f(D'V)$, where f is a continuous function of the elements of $D'V$. It follows from Lemma 2.3 and the continuous mapping theorem that

$$\hat{\beta} \Rightarrow f(N(0,I)) \equiv r,$$

as given by (25). In the case of (b), we have

$$\begin{aligned}\sqrt{T}(\hat{\gamma}_1 - \gamma_1) &= R_1'(T^{-1}Z_1'Z_1)^{-1/2}f(D'V),\\ &\Rightarrow R_1'M_{11}^{-1/2}f(N(0,I))\end{aligned}$$

by the continuous mapping theorem, again for a suitably defined continuous function $f(\ )$. This yields (26) directly and the other results follow in an analogous fashion. ∎

Proof of Lemma 2.5.

$$\begin{aligned}\hat{\sigma}^2 &= T^{-1}(y_1 - Y_2\hat{\beta})'Q_{Z_1}(y_1 - Y_2\hat{\beta}),\\ &= (1,-\hat{\beta}')(T^{-1}Y'Q_{Z_1}Y)\begin{pmatrix}1\\ -\hat{\beta}\end{pmatrix},\\ &\Rightarrow (1 + r'r),\end{aligned}$$

as required. ∎

Proof of Lemma 2.6. Transform

$$Y_2 = Z_1\Pi_1 + V_2$$

on the right by L giving

$$[Y_{21}, Y_{22}] = Z[\Pi_{11},\Pi_{12}] + [V_{21}, V_{22}], \qquad \Pi_{11} = 0, \tag{A3}$$

where

$$\begin{aligned}Y_2L &= Y_2[L_1,L_2] = [Y_{21}, Y_{22}],\\ V_2L &= V_2[L_1,L_2] = [V_{21}, V_{22}].\end{aligned}$$

Then

$$\begin{aligned}L'Y_2'P_HY_2L &= \begin{bmatrix}V_{21}'\\ Y_{22}'\end{bmatrix}P_H[V_{21}, Y_{22}],\\ &= \begin{bmatrix}\bar{V}_{21}'\bar{V}_{21} & \bar{V}_{22}'\bar{Y}_{22}\\ \bar{Y}_{22}'\bar{V}_{21} & \bar{Y}_{22}'\bar{Y}_{22}\end{bmatrix},\end{aligned} \tag{A4}$$

where $\bar{V}_{21} = \bar{Y}_{21} = D'V_{21}$, $\bar{Y}_{22} = D'Y_{22}$, and D is given in (33). Write the partitioned inverse of (A4) as

$$L'(Y_2'P_H Y_2)^{-1}L = \begin{bmatrix} G_{11} & G_{21}' \\ G_{21} & G_{22} \end{bmatrix}, \tag{A5}$$

with

$$G_{11} = (\bar{V}_{21}' Q_{\bar{Y}_{22}} \bar{V}_{21})^{-1},$$

$$G_{21} = -(\bar{Y}_{22}'\bar{Y}_{22})^{-1}\bar{Y}_{22}'\bar{V}_{21}G_{11},$$

$$G_{22} = (\bar{Y}_{22}'\bar{Y}_{22})^{-1} + (\bar{Y}_{22}'\bar{Y}_{22})^{-1}\bar{Y}_{22}'\bar{V}_{21}G_{11}\bar{V}_{21}'\bar{Y}_{22}(\bar{Y}_{22}'\bar{Y}_{22})^{-1}.$$

Now

$$T^{-1/2}\bar{Y}_{22} = T^{-1/2}D'Z_1\Pi_{12} + O_p(T^{-1/2}) \underset{p}{\to} F,$$

$$T^{-1}\bar{Y}_{22}'\bar{Y}_{22} \underset{p}{\to} \Pi_{12}'M_{11}\Pi_{12} > 0,$$

$$Q_{\bar{Y}_{22}} \underset{p}{\to} I - F(F'F)^{-1}F' = Q_F,$$

$$\bar{V}_{21} = D'V_{21} = D'V_2L_1 \Rightarrow \xi L_1 \equiv N_{k_*,n-k_1}(0,I),$$

where

$$F' = [0, F_2'] = [0, \Pi_{12}'M_{11}^{1/2}],$$

$$\xi \equiv N_{k_*,n}(0,I),$$

and

$$(F'F)^{-1} = (\Pi_{12}'M_{11}\Pi_{12})^{-1}.$$

We deduce from these results and (A5) that

$$L'(Y_2'P_H Y_2)^{-1}L \Rightarrow \begin{bmatrix} (L_1'\xi'Q_F\xi L_1)^{-1} & 0 \\ 0 & 0 \end{bmatrix},$$

and thus

$$(Y_2'P_H Y_2)^{-1} \Rightarrow L_1(L_1'\xi'Q_F\xi L_1)^{-1}L_1',$$

as required for part (a).

To prove part (b), we observe that when $n = k_1$

$$Y_2 = Y_{22} = Z_1\Pi_1 + V_2,$$

and $\Pi_1 = \Pi_{12}$ is $k_1 \times k_1$ and nonsingular. We deduce directly that

$$T^{-1}Y_2'P_H Y_2 \underset{p}{\to} \Pi_1'M_{11}\Pi_1$$

as stated. ∎

Proof of Lemma 2.7. Recall that

$$Q = P_H - P_H Y_2(Y_2'P_H Y_2)^{-1}Y_2'P_H,$$

and $P_H = DD'$, so that

$$Z_1'QZ_1 = Z_1'D[I - D'Y_2(Y_2'P_HY_2)^{-1}Y_2'D]D'Z_1.$$

Now, when $n > k$, we have

$$\begin{aligned}
D'Y_2(Y_2'P_HY_2)^{-1}Y_2'D &= D'Y_2LL'(Y_2'P_HY_2)^{-1}LL'Y_2'D, \\
&= [\bar{Y}_{21}, \bar{Y}_{22}]L'(Y_2'P_HY_2)^{-1}L\begin{bmatrix} \bar{Y}_{21}' \\ \bar{Y}_{22}' \end{bmatrix}, \\
&= \bar{Y}_{21}G_{11}\bar{Y}_{21}' - P_{\bar{Y}_{22}}\bar{Y}_{21}G_{11}\bar{Y}_{21}' - \bar{Y}_{21}G_{11}\bar{Y}_{21}'P_{\bar{Y}_{22}} \\
&\quad + P_{\bar{Y}_{22}} + P_{\bar{Y}_{22}}\bar{Y}_{21}G_{11}\bar{Y}_{21}'P_{\bar{Y}_{22}}, \\
&= Q_{\bar{Y}_{22}}\bar{Y}_{21}G_{11}\bar{Y}_{21}'Q_{\bar{Y}_{22}} + P_{\bar{Y}_{22}},
\end{aligned}$$

and thus

$$Z_1'QZ_1 = Z_1'D[Q_{\bar{Y}_{22}} - Q_{\bar{Y}_{22}}\bar{Y}_{21}(\bar{Y}_{21}'Q_{\bar{Y}_{22}}\bar{Y}_{21})^{-1}\bar{Y}_{21}'Q_{\bar{Y}_{21}}]D'Z_1. \tag{A6}$$

But

$$\begin{aligned}
T^{-1/2}\bar{Y}_{22} &= T^{-1/2}D'Z_1\Pi_{12} + T^{-1/2}D'V_{22}, \\
&= F_T + T^{-1/2}\bar{V}_{22} \text{ (say)}.
\end{aligned}$$

Simple manipulations now show that

$$\begin{aligned}
P_{\bar{Y}_{22}} &= P_{F_T} - T^{-1/2}F_T(F_T'F_T)^{-1}\bar{V}_{22}'P_{F_T} \\
&\quad - T^{-1/2}P_{F_T}\bar{V}_{22}(F_T'F_T)^{-1}F_T' + T^{-1/2}\bar{V}_{22}(F_T'F_T)^{-1}F_T' \\
&\quad + T^{-1/2}F_T(F_T'F_T)^{-1}\bar{V}_{22}' + O_p(T^{-1}),
\end{aligned}$$

and so

$$\begin{aligned}
F_T'Q_{\bar{Y}_{22}} &= T^{-1/2}\bar{V}_{22}'P_{F_T} - T^{-1/2}\bar{V}_{22}' + O_p(T^{-1}), \\
&= -T^{-1/2}\bar{V}_{22}'Q_{F_T} + O_p(T^{-1}).
\end{aligned} \tag{A7}$$

It follows that

$$F_T'Q_{\bar{Y}_{22}}F_T = T^{-1}\bar{V}_{22}'Q_{F_T}\bar{V}_{22} + O_p(T^{-3/2}). \tag{A8}$$

Now $T^{-1/2}D'Z_1 = F_T\Pi_{12}^{-1}$ and so, from (A6)–(A8), we deduce that

$$\begin{aligned}
Z_1'QZ_1 &= T\{\Pi_{12}'^{-1}F_T'[Q_{\bar{Y}_{22}} - Q_{\bar{Y}_{22}}\bar{Y}_{21}(\bar{Y}_{21}'Q_{\bar{Y}_{22}}\bar{Y}_{21})^{-1}\bar{Y}_{21}'Q_{\bar{Y}_{22}}]F_T\Pi_{22}^{-1}\}, \\
&= T\Pi_{12}'^{-1}\{T^{-1}\bar{V}_{22}'Q_{F_T}\bar{V}_{22} - T^{-1}\bar{V}_{22}'Q_{F_T}\bar{Y}_{21}(\bar{Y}_{21}'Q_{\bar{Y}_{22}}\bar{Y}_{21})^{-1}\bar{Y}_{21}'Q_{F_T}\bar{V}_{22} \\
&\quad + O_p(T^{-3/2})\}\Pi_{12}^{-1}, \\
&= \Pi_{12}'^{-1}\bar{V}_{22}'[Q_{F_T} - Q_{F_T}\bar{Y}_{21}(\bar{Y}_{21}'Q_{\bar{Y}_{22}}\bar{Y}_{21})^{-1}\bar{Y}_{21}'Q_{F_T}]\bar{V}_{22}\Pi_{12}^{-1} + O_p(T^{-1/2}), \\
&= \Pi_{12}'^{-1}\bar{V}_{22}'[Q_F - Q_F\bar{V}_{21}(\bar{V}_{21}'Q_F\bar{V}_{21})^{-1}\bar{V}_{21}'Q_F]\bar{V}_{22} + o_p(1).
\end{aligned}$$

But

$$[\bar{V}_{21}, \bar{V}_{22}] = D'V_2[L_1, L_2] \Rightarrow \xi[L_1, L_2] \equiv N_{k_*,n}(0, I),$$

and, writing

$$[\xi_1,\xi_2] = \xi[L_1,L_2],$$

we deduce that

$$Z_1'QZ_1 \Rightarrow \Pi_{12}'^{-1}\xi_2'[Q_F - Q_F\xi_1(\xi_1'Q_F\xi_1)^{-1}\xi_1'Q_F]\xi_2\Pi_{12}^{-1},$$
$$\equiv \Pi_{12}'^{-1}W_{k_1}(k_* - n, I)\Pi_{12}^{-1}.$$

The last line is obtained by noting that conditional on ξ_1 the matrix quadratic form in ξ_2 is Wishart with degrees of freedom equal to

$$\mathrm{tr}\{Q_F - Q_F\xi_1(\xi_1'Q_F\xi_1)^{-1}\xi_1'Q_F\} = k_* - k_1 - (n - k_1) = k_* - n.$$

When $n = k_1$, we have $D\ Y_2 = D'Y_{22} = \bar{Y}_{22}$ and

$$D'Y_2(Y_2'P_HY_2)^{-1}Y_2'D = P_{\bar{Y}_{22}}. \tag{A9}$$

In this case, therefore, we have

$$Z_1'QZ_1 = T\{\Pi_{12}'^{-1}F_T'Q_{\bar{Y}_{22}}F_T\Pi_{12}^{-1}\},$$

and from (A8) and (A9), we obtain

$$Z_1'QZ_1 = \Pi_{12}'^{-1}\bar{V}_{22}'Q_{F_T}\bar{V}_{22}\Pi_{12}^{-1} + o_p(1),$$
$$\Rightarrow \Pi_{12}'^{-1}\xi'Q_F\xi\Pi_{12}^{-1},$$
$$\equiv \Pi_{12}'^{-1}W_{k_1}(k_* - k_1, I)\Pi_{12}^{-1},$$

so that the stated result holds for all $n \geq k_1$. ∎

Proof of Theorem 2.8. Note that by Lemmas 2.3, 2.5, and Theorem 2.4,

$$A\hat{\beta} - a \Rightarrow Ar - a,$$
$$\hat{\sigma}^2 \Rightarrow 1 + r'r,$$
$$Y_2'(P_H - P_{Z_1})Y_2 = Y_2'D_1D_1'Y_2 \Rightarrow W_n(k_3, I),$$

and joint weak convergence also applies. We also have

$$A[Y_2'(P_H - P_{Z_1})Y_2]^{-1}A' \Rightarrow A[W_n(k_3, I)]^{-1}A',$$

so that

$$\{A[Y_2'(P_H - P_{Z_1})Y_2]^{-1}A'\}^{-1} \Rightarrow W(A) \equiv W_{p_a}(k_3 - n + p_a, (AA')^{-1})$$

by standard theory of the Wishart distribution (for example, [14, Theorem 3.2.11]). Part (a) follows directly.

To prove (b), we note that by Corollary 2.2 (using $R_2 = I$, $\hat{\gamma} = \hat{\gamma}_2$) and Lemma 2.7

$$\hat{\gamma} \Rightarrow \pi_1 - \Pi_1 r,$$
$$Z_1'QZ_1 \Rightarrow W_{k_1}(k_* - n, (\Pi_{12}\Pi_{12}')^{-1}).$$

Moreover,

$$[B(Z_1'QZ_1)^{-1}B']^{-1} \Rightarrow [B\{W_{k_1}(k_* - n, (\Pi_{12}\Pi_{12}')^{-1})\}^{-1}B']^{-1},$$
$$\equiv W_{p_b}(k_3 - n + p_b, (B\Pi_{12}\Pi_{12}'B')^{-1}),$$
$$= W(\bar{B}),$$

since $\Pi_{12}\Pi_{12}' = \Pi_1\Pi_1'$. We deduce that

$$W_\gamma = (B\hat{\gamma} - b)'[B(Z_1'QZ_1)^{-1}B']^{-1}(B\hat{\gamma} - b)/\hat{\sigma}^2,$$
$$\Rightarrow (\bar{B}r - \bar{b})'W(B)(\bar{B}r - \bar{b})/(1 + r'r),$$

where

$$\bar{B} = -B\Pi_1,$$
$$\bar{b} = b - B\pi_1,$$

as required for part (b). ■

Proof of Lemma 2.9. Observe that

$$T^{-1}Z_1'QZ_1 = (T^{-1/2}Z_1'D)\{I - D'Y_2(Y_2'DD'Y_2)^{-1}Y_2'D\}(T^{-1/2}D'Z_1), \tag{A10}$$
$$T^{-1/2}Z_1'D = [0, (T^{-1}Z_1'Z_1)^{1/2}] \rightarrow [0, M_{11}^{1/2}], \tag{A11}$$

and since $\Pi_1 = 0$, we have

$$D'Y_2 = D'V_2 = \bar{V}_2 \text{ (say)},$$

and

$$\bar{V}_2 \equiv N_{k_*,n}(0, I).$$

Now

$$\Theta_2 = \bar{V}_2(\bar{V}_2'\bar{V}_2)^{-1/2} \equiv U(V_{n,k_*}),$$

i.e., Θ_2 is uniformly distributed on the Stiefel manifold $V_{n,k_*} = \{\Theta_2(k_* \times n) : \Theta_2'\Theta_2 = I_n\}$. Construct the orthogonal matrix

$$\Theta = [\Theta_1, \Theta_2] \in O(k_*),$$

and partition as

$$\Theta = \begin{array}{cc} \begin{array}{cc} k_* - n & n \end{array} & \\ \begin{bmatrix} \Theta_{11} & \Theta_{21} \\ \Theta_{21} & \Theta_{22} \end{bmatrix} & \begin{array}{c} k_3 \\ k_1 \end{array} \end{array},$$

so that

$$\Theta_1\Theta_1' = I - \Theta_2\Theta_2'. \tag{A12}$$

From (A10)–(A12), we deduce that

$$T^{-1}Z_1'QZ_1 = (T^{-1/2}Z_1'D)\Theta_1\Theta_1(T^{-1/2}D'Z_1),$$
$$\Rightarrow M_{11}^{1/2}\Theta_{21}\Theta_{21}'M_{11}^{1/2},$$

as required. ■

Proof of Theorem 2.10. Since $\Pi_1 = 0$, we have $R_1 = I$ and from Corollary 2.2

$$\sqrt{T}(\hat{\gamma} - \gamma) \Rightarrow \bar{s} \equiv \int_{m>0} N(0,(1+m)M_{11}^{-1})pdf(m)dm,$$

where $pdf(m)$ is given in (23). Under the null $H_\gamma : B\gamma = b$, we deduce that

$$\sqrt{T}(B\hat{\gamma} - b) = \sqrt{T}B(\hat{\gamma} - \gamma) \Rightarrow B\bar{s}.$$

The stated result now follows from Lemmas 2.5, 2.9, and the continuous mapping theorem, noting that joint weak convergence of the component variates ($\hat{\beta}$, $\sqrt{T}(\hat{\gamma} - \gamma), \Theta_{21}$) also applies. When H_γ is false, $\sqrt{T}(B\hat{\gamma} - b)$ diverges and so too does the statistic W_γ. ■

Proof of Lemma 3.1. Consider first the finite-dimensional distributions. For fixed r, we have

$$B(r) = \begin{bmatrix} B_1(r) \\ B_2(r) \end{bmatrix} \equiv N(0, r\Omega).$$

By the conventional theory for conditional distributions of the multivariate normal, we obtain

$$B_1(r)|_{F_2} \equiv N(\omega_{21}'\Omega_{22}^{-1}B_2(r), r\omega_{11\cdot 2}),$$
$$\equiv \omega_{21}'\Omega_{22}^{-1}B_2(r) + l_{11}W_1(r),$$

where

$$\omega_{11\cdot 2} = \omega_{11} - \omega_{21}'\Omega_{22}^{-1}\omega_{21} = l_{11}^2$$

is the conditional variance of $B_1(r)$ given $B_2(r)$ and

$$W_1(r) \equiv BM(1),$$

standard Brownian motion independent of $B_2(r)$. Similarly, for $0 < r < s \le 1$ we have

$$B(r) \equiv N(0, r\Omega), \qquad B(s) - B(r) \equiv N(0, (r-s)\Omega)$$

and the distributions are independent. Thus,

$$(B_1(r), B_1(s) - B_1(r))|_{F_2} \equiv N\left(\omega_{21}'\Omega_{22}^{-1}(B_2(r), B_2(s) - B_2(r)), \omega_{11\cdot 2}\begin{bmatrix} r & 0 \\ 0 & s-r \end{bmatrix}\right),$$
$$\equiv \omega_{21}'\Omega_{22}^{-1}(B_2(r), B_2(s) - B_2(r))$$
$$+ l_{11}(W_1(r), W_1(s) - W_1(r)).$$

Higher-dimensional distributions follow in the same way. It follows that the finite-dimensional distributions of $B_1|F_2$ are equivalent to those of

$$\omega_{21}'\Omega_{22}^{-1}B_2 + l_{11}W_1$$

given any realization of B_2 in F_2. Since the finite-dimensional distributions are a determining class on $C[0,1]$, the space of continuous functions on the (0,1] interval (see [3, p. 35]), we deduce that

$$B_1|_{F_2} \equiv \omega_{21}'\Omega_{22}^{-1}B_2 + l_{11}W_1,$$

as required. ■

Proof of Theorem 3.2. From (50), we have

$$\hat{\beta} \Rightarrow \left(\int_0^1 B_2B_2'\right)^{-1}\left(\int_0^1 B_2B_1\right),$$

and by Lemma 3.1,

$$\int_0^1 B_2B_1|_{F_2} \equiv \left(\int_0^1 B_2B_2'\right)\Omega_{22}^{-1}\omega_{21} + l_{11}\int_0^1 B_2W_1.$$

It follows that

$$\left(\int_0^1 B_2B_2'\right)^{-1}\int_0^1 B_2B_1|_{F_2} \equiv \Omega_{22}^{-1}\omega_{21} + l_{11}\left(\int_0^1 B_2B_2'\right)^{-1}\left(\int_0^1 BW_1\right). \tag{A13}$$

However, B_2 is independent of W_1 so that

$$\left(\int_0^1 B_2B_2'\right)^{-1}\int_0^1 BW_1|_{F_2} \equiv N(0, V(B_2)), \tag{A14}$$

where

$$V(B_2) = \left(\int_0^1 B_2B_2'\right)^{-1}\left(\int_0^1\int_0^1 B_2(r)(r \wedge s)B_2(s)\right)\left(\int_0^1 B_2B_2'\right)^{-1},$$

and

$$r \wedge s = \min(r,s) = E(W_1(r)W_1(s))$$

is the covariance kernel of the Brownian motion W_1. Relation (52) follows directly from (A13) and (A14) by integrating the conditional distribution with respect to the probability measure of $V(B_2)$ induced by B_2.

To prove (53), we note first that the conditional characteristic function corresponding to (A14) is

$$cf(s|F_2) = \exp\left\{-\frac{1}{2}s'V(B_2)s\right\}. \tag{A15}$$

Since $B_2 \equiv BM(\Omega_{22})$, we may write

$$B_2 \equiv \Omega_{22}^{1/2}W_2,$$

where $W_2 \equiv BM(I_n)$, independent of W_1. Now

$$s'V(B_2)s = (s'\Omega_{22}^{-1}s)(\bar{s}'V(B_2)\bar{s}), \tag{A16}$$

where

$$\bar{s} = s/(s'\Omega_{22}^{-1}s)^{1/2},$$

and

$$\begin{aligned}\bar{s}'V(B_2)\bar{s} &\equiv \bar{s}'\Omega_{22}^{-1/2}V(W_2)\Omega_{22}^{-1/2}\bar{s}\\ &= h'V(W_2)h,\end{aligned} \tag{A17}$$

with

$$h = \Omega_{22}^{-1/2}\bar{s} = \Omega_{22}^{-1/2}s/(s'\Omega_{22}^{-1}s)^{1/2}.$$

The vector h lies on the unit sphere in R^n. We construct an orthogonal matrix

$$H = [h, H_2] \in O(n),$$

and noting that

$$\overline{W}_2 = H'W_2 \equiv W_2 \equiv BM(I_n),$$

we deduce that

$$\begin{aligned}h'V(W_2)h &\equiv h'HV(H'W_2)H'h,\\ &= e_1'V(\overline{W}_2)e_1.\end{aligned} \tag{A18}$$

It follows from (A16)–(A18) that the conditional characteristic function may equivalently be written as

$$cf(s|F_2') = \exp\left\{-\frac{1}{2}s'\Omega_{22}^{-1}sv\right\}, \tag{A19}$$

where $F_2' = \sigma\{W_2(r): 0 \le r \le 1\}$, and

$$\begin{aligned}v &= e_1'V(W_2)e_1,\\ &= e_1'\left\{\left(\int_0^1 W_2W_2'\right)^{-1}\left(\int_0^1\int_0^1 W_2(r)(r \wedge s)W_2(s)\right)\left(\int_0^1 W_2W_2'\right)\right\}e_1.\end{aligned}$$

The stated result (53) is obtained directly by integrating the conditional density that corresponds to (A19) with respect to the probability measure of v induced by W_2. ∎

Proof of Theorem 3.3. Least-squares regression on (63) yields

$$\hat{\beta}_1 = (x_1'Q_2x_1)^{-1}(x_1'Q_2y), \tag{A20}$$

$$\hat{\beta}_2 = (X_2'Q_1X_2)^{-1}(X_2'Q_1y), \tag{A21}$$

in conventional notation for partitioned regressions. Now

$$T^{-3}x_1'Q_2x_1 = T^{-3}x_1'x_1 - (T^{-5/2}x_1'X_2)(T^{-2}X_2'X_2)^{-1}(T^{-5/2}X_2'x_1), \tag{A22}$$

$$\Rightarrow \int_0^1 \zeta^2, \tag{A23}$$

where

$$\zeta(r) = r - \int_0^1 r\underline{B}_2'\left(\int_0^1 \underline{B}_2\underline{B}_2\right)^{-1}\underline{B}_2(r),$$

and $\underline{B} \equiv BM(H_2'\Omega_{22}H_2) = BM(\underline{\Omega}_{22})$. We may readily verify (A23) by observing the joint convergence

$$\begin{bmatrix} T^{-3}x_1'x_1 \\ T^{-5/2}x_1'X_2 \\ T^{-2}X_2'X_2 \end{bmatrix} \Rightarrow \begin{bmatrix} \int_0^1 r^2 \\ \int_0^1 r\underline{B}_2' \\ \int_0^1 \underline{B}_2\underline{B}_2' \end{bmatrix},$$

and by applying the continuous mapping theorem. Relation (A23) then follows directly from the construction of the process $\zeta(r)$. In a similar way, we obtain

$$T^{-5/2}x_1'Q_2u = T^{-5/2}x_1'u - (T^{-5/2}x_1'X_2)(T^{-2}X_2'X_2)^{-1}(T^{-2}X_2'u),$$

$$= \int_0^1 \zeta\underline{B}_1,$$

where $\underline{B}_1$ is the first component of the n-dimensional Brownian motion:

$$\underline{B}(r) = \begin{bmatrix} \underline{B}_1(r) \\ \underline{B}_2(r) \end{bmatrix} = \begin{bmatrix} 1 & -\beta' \\ 0 & H_2' \end{bmatrix}\begin{bmatrix} B_1(r) \\ B_2(r) \end{bmatrix} \equiv BM(\underline{\Omega}),$$

where

$$\underline{\Omega} = \begin{bmatrix} \underline{\omega}_{11} & \underline{\omega}_{21}' \\ \underline{\omega}_{21} & \underline{\Omega}_{22} \end{bmatrix}^{(n\times n)},$$

and

$$\underline{\omega}_{11} = \omega_{11} - 2\beta'\omega_{21} + \beta'\Omega_{22}\beta,$$

$$\underline{\omega}_{21} = H_2(\omega_{21} - \Omega_{22}\beta),$$

$$\underline{\Omega}_{22} = H_2\Omega_{22}H_2'.$$

Relation (66) now follows directly. To prove (67), we note from Lemma 3.1 that

$$\underline{B}_1|_{\underline{F}_2} = \underline{\omega}_{21}'\underline{\Omega}_{22}^{-1}\underline{B}_2 + \underline{l}_{11}W_1, \tag{A24}$$

where $W_1 \equiv BM(1)$, $\underline{F}_2$ is the σ-field generated by $\{\underline{B}_2(r): 0 \le r \le 1\}$ and $\underline{l}_{11} = \underline{\omega}_{11\cdot 2}^{1/2} = (\underline{\omega}_{11} - \underline{\omega}_{22}'\underline{\Omega}_{22}^{-1}\underline{\omega}_{21})^{1/2}$. Note also that ζ is orthogonal to the components of $\underline{B}_2$ in $L_2[0,1]$. It follows that

$$\int_0^1 \zeta \underline{B}_1 |_{\underline{F}_2} \equiv \underline{l}_{11} \int_0^1 \zeta W_1,$$

and

$$\left(\int_0^1 \zeta^2\right)^{-1} \left(\int_0^1 \zeta B_1\right)\Big|_{\underline{F}_2} \equiv \underline{l}_{11} \left(\int_0^1 \zeta^2\right)^{-1} \int_0^1 \zeta W_1,$$

$$\equiv N(0, \underline{\omega}_{11\cdot 2} v_1), \tag{A25}$$

where

$$v_1 = \left(\int_0^1 \zeta^2\right)^{-1} \left(\int_0^1 \int_0^1 \zeta(r)(r \wedge s)\zeta(s)\right) \left(\int_0^1 \zeta^2\right)^{-1}.$$

Integrating (A25) with respect to the probability measure $P(v_1)$ induced by $\underline{B}_2$, we obtain (67) as required.

To prove (68), we work from (A21) in a similar fashion, finding

$$T^{-2}X_2'QX_2 = T^{-2}X_2'X_2 - (T^{-5/2}X_2'x_1)(T^{-3}x_1'x_1)^{-1}(T^{-5/2}x_1'X_2),$$

$$\Rightarrow \int_0^1 \underline{B}_2\underline{B}_2' - \left(\int_0^1 \underline{B}_2 r\right)\left(\int_0^1 r^2\right)^{-1}\left(\int_0^1 r\underline{B}_2'\right),$$

$$= \int_0^1 \eta\eta'.$$

and

$$T^{-2}X_2'Q_1u = (T^{-2}X_2'u) - (T^{-5/2}X_2'x_1)(T^{-3}x_1'x_1)^{-1}(T^{-5/2}x_1'u),$$

$$\Rightarrow \int_0^1 \eta \underline{B}_1,$$

where

$$\eta(r) = \underline{B}_2(r) - \left(\int_0^1 \underline{B}_2 r\right)\left(\int^1 r^2\right)^{-1} r.$$

We deduce that

$$\hat{\beta}_2 = (X_2'Q_1X_2)^{-1}(X_2'Q_1y),$$

$$= (X_2'Q_1X_2)^{-1}(X_2Q_1u),$$

$$\Rightarrow \left(\int_0^1 \eta\eta'\right)^{-1}\left(\int_0^1 \eta\underline{B}_1\right),$$

giving (68). Noting that

$$\int_0^1 \eta \underline{B}_2' = \int_0^1 \underline{B}_2\underline{B}_2' - \left(\int_0^1 \underline{B}_2 r\right)\left(\int_0^1 r^2\right)^{-1}\left(\int_0^1 r\underline{B}_2\right) = \int_0^1 \eta\eta',$$

and again using the conditional Brownian motion argument based on (A24), we find

$$\left(\int_0^1 \eta\eta'\right)^{-1}\left(\int_0^1 \eta\underline{B}_1\right)\bigg|_{\underline{F}_2} \equiv \underline{\Omega}_{22}^{-1}\underline{\omega}_{21} + \underline{l}_{11}\left(\int_0^1 \eta\eta'\right)^{-1}\int_0^1 \eta W_1,$$

$$\equiv N(\underline{\Omega}_{22}^{-1}\underline{\omega}_{21}, \underline{\omega}_{11\cdot 2}V_2), \quad \textbf{(A26)}$$

where

$$V_2 = \left(\int_0^1 \eta\eta'\right)^{-1}\left(\int_0^1\int_0^1 \eta(r)(r \wedge s)\eta(s)'\right)\left(\int_0^1 \eta\eta'\right)^{-1}.$$

Integrating (A26) with respect to the probability measure $P(V_2)$ induced on $V_2 > 0$ by $\underline{B}_2$, we deduce the stated result (69).

To prove (70), we write

$$\underline{B}_2(r) \equiv \underline{\Omega}_{22}^{1/2}\underline{W}_2,$$

where $\underline{W}_2 \equiv BM(I_{n-1})$. The conditional characteristic function corresponding to (A26) is

$$cf(s|\underline{F}_2) = \exp\left\{i\underline{\omega}_{21}'\underline{\Omega}_{22}^{-1}s - \frac{1}{2}\underline{\omega}_{11\cdot 2}s'V_2 s\right\}.$$

Now

$$s'V_2 s = (s'\underline{\Omega}_{22}^{-1}s)(\bar{s}'V_2\bar{s}), \qquad \bar{s} = s/(s'\underline{\Omega}_{22}^{-1}s)^{1/2}$$

and

$$\bar{s}'V_2\bar{s} \equiv \bar{s}\underline{\Omega}_{22}^{-1/2}V_2(\underline{W}_2)\underline{\Omega}_{22}^{-1/2}\bar{s} = h'V_2(\underline{W}_2)h,$$

where

$$h = \underline{\Omega}_{22}^{-1/2}\bar{s} = \underline{\Omega}_{22}^{-1/2}s/(s'\underline{\Omega}_{22}^{-1}s)^{1/2}$$

lies on the unit sphere in R^{n-1}. Using the same argument as that leading to (A18) and (A19), we find that the conditional characteristic function has the equivalent representation

$$cf(s|\underline{F}_2) = \exp\left\{i\underline{\omega}_{21}'\underline{\Omega}_{22}^{-1}s - \frac{1}{2}\underline{\omega}_{11\cdot 2}v_2 s'\underline{\Omega}_{22}^{-1}s\right\},$$

where

$$v_2 = e_1'V_2(\underline{W}_2)e_1.$$

Note that $\underline{W}_2(r)$ is the $n-1$ dimensional standard Brownian motion (as distinct from n-vector Brownian motion in the proof of Theorem 3.2). Integration of the conditional density with respect to the probability measure $P(v_2)$ yields the stated result (70). ■

Proof of Corollary 3.4. From (64) and (65),

$$\begin{aligned}\hat{\beta} - \beta &= H(\hat{\underline{\beta}} - \underline{\beta}) = \mu_2(\hat{\beta}_1 - \beta_1) + H_2(\hat{\beta}_2 - \beta_2),\\ &= H_2\hat{\beta}_2 + O_p(T^{-1/2}),\\ &\Rightarrow H_2\int_{v_2>0} N(\underline{\Omega}_{22}^{-1}\underline{\omega}_{21}, v_2\omega_{11\cdot 2}\underline{\Omega}_{22}^{-1})dP(v_2),\end{aligned}$$

as required for (71). ∎

Proof of Proposition 4.1. Note that under $P_{\theta_0+T^{-1}h,T}$, the model (102) is given by

$$X_t = (1 + h/T)X_{t-1} + u_t, \tag{A27}$$

where $\theta_0 = 1$. In the terminology of Phillips [28,29], where models such as (A27) are studied in detail, $\{X_t\}$ is a near-integrated process. From Lemma 1 of Phillips [28], we have

$$T^{-2}\sum_1^T X_{t-1}^2 \Rightarrow \int_0^1 J_h(r)^2 dr = S(\theta_0, h),$$

where

$$J_h(r) = \int_0^r e^{(r-s)h}dB(s),$$

as required. Clearly,

$$S(\theta_0, h) \neq S(\theta_0) = \int_0^1 B(r)^2 dr, \qquad h \neq 0,$$

and (C9) is violated. ∎