COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125 Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1047

NOTE: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than acknowledgment that a writer had access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

Fully Modified Least Squares and Vector Autoregression

by

Peter C.B. Phillips

May 1993

FULLY MODIFIED LEAST SQUARES AND VECTOR AUTOREGRESSION

by

Peter C. B. Phillips*

Cowles Foundation for Research in Economics,
Yale University

First version: March, 1993

Second version: May, 1993

*Prepared as the Marschak Lecture for the Far Eastern Meetings of the Econometric Society in Taipei, June 1993. Some of the results in Section 3 of this paper were first obtained by the author in 1990 and were circulated at Yale under the title "Time Series Regression with Cointegrated Regressors." All computations and graphics reported in the paper were performed by the author in programs written in GAUSS 3.0 on a 486-50 PC. My thanks go to the NSF for support under grant no. SES 9122142, and to Glena Ames for her skill and effort in keyboarding the manuscript.

0. ABSTRACT

Fully modified least squares (FM-OLS) regression was originally designed in work by Phillips and Hansen (1990) to provide optimal estimates of cointegrating regressions. The method modifies least squares to account for serial correlation effects and for the endogeneity in the regressors that results from the existence of a cointegrating relationship. Recent work by the author (1992) has shown that FM-OLS regression produces estimates of a unit root in time series regression that are hyperconsistent in the sense that their rate of convergence exceeds that of the OLS estimator, i.e. is faster than $O(T)$ as the sample size as $T \to \infty$. That result is extended here to multivariate unit root models and models with deterministic trends, where it is shown that estimates of the trend coefficients are also hypercosistent. The paper provides a general framework which makes it possible to study the asymptotic behavior of FM-OLS in models with full rank I(1) regressors, models with I(1) and I(0) regressors, models with unit roots, models with only stationary regressors and models with I(1) and I(0) regressors as well as deterministic trends. This framework enables us to consider the use of FM regression in the context of vector autoregressions (VAR's) with some unit roots and some cointegrating relations. The resulting FM-VAR regressions are shown to have some interesting, desirable and rather unexpected properties. For example, when there is some cointegration in the system, FM-VAR estimation has a limit theory that is normal for all of the stationary coefficients and mixed normal for *all* of the nonstationary coefficients. Thus, there are no unit root limit distributions even in the case of the unit root coefficient submatrix (i.e. I_{n-r}, for an n-dimensional VAR with r cointegrating vectors). When the system is stationary, the FM-VAR estimates are asymptotically equivalent to those of OLS. When the system has a full set of unit roots the FM-VAR estimator of the complete unit root matrix (i.e. I_n for an n-dimensional VAR) is hyperconsistent, just as in the single equation AR(1) case; and the FM-VAR estimates of the stationary part of the VAR is asymptotically equivalent to OLS. These results indicate that FM-VAR regression has some attractive features compared with conventional OLS levels VAR estimation.

The paper also develops an asymptotic theory for inference based on FM-OLS and FM-VAR regression. The limit theory for Wald tests that rely on the FM estimator is shown to involve a linear combination of independent chi-squared variates. This limit distribution is bounded above by the conventional chi-squared distribution with degrees of freedom equal to the number of restrictions. Thus, conventional critical values can be used to construct valid (but conservative) asymptotic tests in quite general FM time series regressions. This theory applies to causality testing in VAR's and is therefore potentially important in empirical applications.

JEL Classification No. 211

Keywords: Causality testing, cointegration; fully modified regression; fully modified vector autoregression; hyperconsistency; long-run covariance matrix; one-sided long-run covariance matrix; some unit roots

1. INTRODUCTION

In recognition of the fact that most economic time series have some nonstationary characteristics much recent attention in time series econometrics has been devoted to issues of modelling with, estimation for and inference from such data. As a direct consequence of this attention, a huge literature has emerged that seeks to confront these issues. Although the field is still very young (it is still under a decade old) the volume of contributions is so large that it is reasonable to think of it as having come a long way in a short time. Two early developments in this field opened up the area for subsequent research and are still of central importance as it begins to mature. One of these was the careful formulation of models that allow stationary and nonstationary time series to coexist in the same equation and that relate nonstationary series in long-run cointegrating relationships. Although there were many precursors to this research in empirical error correction modelling (see Hendry, 1993, for a recent overview), the paper by Engle and Granger (1987) was certainly the primary stimulus. The other early contribution that has since opened up many different avenues of research in this area was the development of an asymptotic theory of regression for nonstationary time series. There were precursors to this work too, coming from research in the statistical literature on univariate autoregression. But the development of a regression theory for multiple nonstationary time series came from work in econometrics on the asymptotics of unit roots and spurious regression (Phillips, 1986, 1987) and on multivariate functional central limit theory and its application to time series regression (Phillips and Durlauf, 1986). The arithmetic of I(1) and I(1)/I(0) asymptotic analysis, as we might now call this theory (see Phillips, 1988, for a general review of these techniques), enables us to study the asymptotic behavior of statistical procedures in the context of models that explicitly admit nonstationary time series. This means that we also have the apparatus to explore the statistical implications of one methodology, such as the use of error correction models, against those of another, like the use of unrestricted vector autoregression.

The present lecture is in one sense an extended illustration of this exercise. But it has a more basic and (what the author hopes is) ultimately a more important purpose. This is to develop an approach to regression for time series that takes advantage of data nonstationarity and potential cointegrating links between series without having to be explicit about their form and without preliminary pretesting. Cointegrating links between nonstationary series lead to endogeneities in the regressors that cannot be avoided by using vector autoregressions (VAR's) as if they were simply reduced forms. This is a point that was explained in earlier work (1991a) by the author and is illustrated here in Section 2. Nevertheless, we often do wish to use VAR's in empirical research without prefiltering to "induce" stationarity, without pretesting to determine the number of unit roots (or the dimension of the cointegration space), and without prior knowledge of either the directions in which the data may be stationary or the transformations that may be necessary to achieve this. However, least squares (OLS) regressions on levels VAR's which are treated as reduced forms do not have generally good properties in models of this type, especially with respect to the coefficients of (non redundant) nonstationary variables in the system. For example, as we explain in Section 2, OLS estimates of any cointegrating relations are asymptotically second order biased in the sense that their limit distributions are mislocated or shifted away from the true parameters, even though the estimates are consistent (or first order unbiased). The reason for this is simple. OLS regressions are not designed to take into account long-run endogeneities in the regressors and the presence of such endogeneities produces the aforementioned bias.

Ideally, we need a statistical estimation procedure that offers many of the advantages of an unrestricted levels VAR while at the same time allowing for potential long-run endogeneities. The procedure suggested in this paper is designed to achieve this marriage of the two principles. The method proposed here we call *fully modified vector autoregression* (FM-VAR) and is based on, but not identical to, a time series regression estimator known as fully modified least squares (FM-OLS) that was put forward in earlier research by Phillips and Hansen (1990).

The FM estimator was originally designed to estimate cointegrating relations directly by modifying traditional OLS with corrections that take account of endogeneity and serial correlation. One reason the method has proved useful in practice is that one can use the FM corrections to determine how important these effects are in an empirical application. This has helped to make the method less of a "black box" for practitioners. In cases where there are major differences with OLS the source or sources of those differences can usually be easily located and this in turn helps to provide the investigator with additional information about important features of the data. Recent simulation experience and empirical research indicates that the FM estimator performs very well in relation to other methods of estimating cointegrating relations -- see Cappuccio and Lubian (1992), Hansen and Phillips (1990), Hargreaves (1993), Phillips and Loretan (1991) and Rau (1992).

The present paper explores the use of the FM-OLS procedure in a much more general time series context than earlier research. Our framework includes vector autoregressions with some unit roots and some cointegrating vectors, without having to be explicit about the configuration or the dimension of the stationary and nonstationary components in the system and without the need to pretest the data concerning these characteristics. The resulting FM-VAR regression, as we call it, has some attractive and rather surprising properties that emerge from our analysis:

(i) First, when there is cointegration in the system the limit theory of the FM-VAR estimator is normal (and asymptotically equivalent to OLS) for the stationary coefficients, and mixed normal for all of the nonstationary coefficients including the unit roots. Thus, we get mixed normal limit theory for the FM-VAR estimates of the identified components of the cointegrating matrix, just like the optimal (maximum likelihood) estimates in Phillips (1991a) and Johansen (1988). But, in addition, the FM-VAR estimates of the unit root coefficient submatrix (I_{n-r} in the case of an n-dimensional VAR with an r dimensional cointegrating space and $n-r$ unit roots) also have a mixed normal limit theory. So there are no unit root distributions and there is no asymptotic bias in the estimation of the cointegration space in the FM-VAR limit theory.

(ii) When the system has a full set of unit roots, the FM-VAR estimator of the complete unit root matrix (I_n for an n-dimensional VAR) is hyperconsistent in the sense that the rate of convergence of the estimator exceeds the $O(T)$ rate of the OLS and MLE estimators. This extends earlier work by the author (1992), which showed that the FM-OLS estimator is hyperconsistent for a unit root in a single equation autoregression. We further show that when an autoregressive model with a unit root has deterministic trending regressors, the FM-OLS estimator of the coefficients of the deterministic trends is also hyperconsistent. This result is quite surprising and serves to illustrate the importance of the statistical dependence that exists in the limit between estimates of the coefficients of stochastic and deterministic trending regressors.

(iii) The normal and mixed normal limit distributions of FM-VAR estimates facilitate statistical inference in cointegrated VAR's. Wald tests that are based on the FM-VAR estimator are shown to have a limit distribution that is a linear combination of chi-squared variates. The limit variate is bounded above by the usual χ^2 distribution with degrees of freedom equal to the number of restrictions that are being tested. Thus, conventional critical values can be used to construct asymptotically valid (but conservative) tests in quite general FM-VAR regressions. This theory includes causality tests and therefore offers an alternative to sequential test procedures such as those in Toda-Phillips (1992), and to intentional model overfitting procedures like those in Toda and Yamamoto (1993).

The present work is related to some other recent research contained in papers by Phillips (1992) and Kitamura and Phillips (1992). Phillips (1992) demonstrates the hyperconsistency of FM-OLS in an autoregression with a unit root and the present paper extends that result to vector autoregressions, while at the same time considering models with less than a full set of unit roots. Kitamura and Phillips (1992) develop generalized method of moment (GMM) and generalized instrumental variable (GIVE) extensions of the FM regression procedure. The resulting FM-GMM and FM-GIVE estimators are designed to estimate cointegrated regression models, wherein the stationary components may also be endogenous and are consistently and efficiently estimated because of the GMM and GIVE features that are built into the FM-GMM and FM-GIVE procedures. Work on this problem relies on the fact that the FM procedure can be applied to models with cointegrated regressors and even stationary regressors without losing the method's good asymptotic properties. This result was originally shown by the author in some unpublished notes (1991b). Section 3 of the present paper extends those notes and provides a rather full treatment of the subject, giving a detailed analysis of the conditions under which the result holds and providing specific limits for the relevant (long-run) moment matrices. The treatment of this section is useful to a wide range of models including those where FM-GMM and

FM-GIVE procedures may be appropriate. Our main use of the treatment in the present paper will be to vector autoregressions.

The paper proceeds as follows. Section 2 provides an illustration and some background discussion of the relevant ideas that help to motivate the need for a modified VAR estimation procedure. Section 3 develops a general theory of FM-OLS asymptotics that covers models with I(1) and I(0) regressors, models with cointegrated regressors where the directions of cointegration are unknown, models with unit roots, models with only stationary regressors and models with I(1), I(0) and deterministic trending regressors. Autoregressions are studied in Section 4 and some simulations are reported that shed light on the finite sample performance of the FM-OLS estimator in the stationary and nonstationary AR(1) model. Section 5 develops an asymptotic theory of regression for the FM-VAR estimator and Section 6 derives the limit theory for Wald tests of restrictions, based on FM-VAR regression. Section 7 concludes the paper and summarizes our main results. Derivations and proofs are given in a technical appendix in Section 8.

The notation and terminology that we use in the paper for nonstationary regression asymptotics is based on earlier work by the author and has now become fairly standard in the time series econometrics literature. Thus, we call the matrix $\Omega = \Sigma_{k=-\infty}^{\infty} E(u u_0')$ the long-run variance matrix of the (covariance stationary) time series u_t and write lrvar(u_t) $= \Omega$. Similarly, we call $\Delta = \Sigma_{k=0}^{\infty} E(u_k u_0')$ the one-sided long-run variance matrix of u_t and write lrvar$_+(u_t) = \Delta$. In a similar way we designate long-run covariance matrices as lrcov($\cdot$) and lrcov$_+(\cdot)$. We use $BM(\Omega)$ to denote a vector Brownian motion with covariance matrix Ω and we usually write integrals like $\int_0^1 B(s)ds$ as $\int_0^1 B$ or simply $\int B$ when there is no ambiguity over limits. The notation $y_t \equiv$ I(1) signifies that the time series y_t is integrated of order one, so that $\Delta y_t \equiv$ I(0) and this requires that lrvar(Δy_t) > 0. In addition, the inequality "> 0" denotes positive definite when applied to matrices and the symbols "$\longrightarrow_d$", "$\longrightarrow_p$", "a.s.", "$\equiv$" and "$:=$" signify convergence in distribution, convergence in probability, almost surely, equality in distribution, and notational definition, respectively; and we use $\|A\|$ to signify the matrix norm $\{\mathrm{tr}(A'A)\}^{1/2}$, $|A|$ to denote the determinant of A, vec( ) to stack the rows of a matrix into a column vector, $[x]$ to denote the smallest integer $\leq x$ and all limits in the paper are taken as the sample size $T \to \infty$, except where otherwise noted.

2. BACKGROUND IDEAS AND MOTIVATION FOR MODIFIED VAR ESTIMATION

To illustrate some of the ideas that come into play in the present paper we will consider in this section the following first order n-vector autoregression

(1) $\quad y_t = Ay_{t-1} + \varepsilon_t , \quad t = 1, ..., T$

where $\varepsilon_t \equiv \text{iid}(0, \Sigma_{\varepsilon\varepsilon})$ with $\Sigma_{\varepsilon\varepsilon} > 0$ and the initialization y_0 is any random n-vector. Suppose the coefficient matrix A in (1) has the simple form

$$A = \begin{bmatrix} 0 & B \\ 0 & I_{n-r} \end{bmatrix} = (A_{ij}) , \quad say$$

for some $r \times (n-r)$ matrix B. Partitioning $y_t = (y_{1t}', y_{2t}')'$ conformably with A we have the following explicit form of (1)

(1a) $\quad y_{1t} = By_{2t-1} + \varepsilon_{1t} ,$

(1b) $\quad y_{2t} = y_{2t-1} + \varepsilon_{2t} ,$

showing that y_{2t} is a full rank I(1) process and that y_{1t} is cointegrated with y_{2t}. Thus, (1) is a simple VAR with some $(n-r)$ unit roots and some (r) cointegrating vectors that have the form $\beta' = [I \;\; -B]$. (This model extends a simple exercise given in Phillips (1992b).)

Premultiplication of (1) by β' gives the stationary relation

(1a′) $\quad \beta' y_t = y_{1t} - By_{2t} = \beta'\varepsilon_t = v_t , \quad$ say

which shows the directions in which the n-vector y_t is stationary. Since these directions (and indeed the form of the coefficient matrix A in (1)) are not known, we may well consider estimating the matrix A directly from (1) as a levels VAR. In such a regression y_{t-1} is treated as predetermined and the model is usually regarded as a "reduced form." However, because of the nonstationarity in the data, the endogeneity in the variable y_{2t} that is clear from the form of

(1a′) is also present in the lagged variable y_{2t-1}. This can most easily be seen by noting that (1a) is really just another way of writing (1a′) -- we simply add and subtract $B\varepsilon_{2t}$ to the right side of equation (1a).

To be more explicit we note that $E(\varepsilon_{1t}y_{2t-1}') = 0$, so that y_{2t-1} appears to satisfy the usual orthogonality condition of a "good" regressor or predetermined variable. Nevertheless, since y_{2t-1} is nonstationary the sample covariance $T^{-1}\Sigma_1^T\varepsilon_{1t}y_{2t-1}'$ does not converge to zero. Instead, we have, using standard weak convergence results (see Phillips, 1988),

(2) $$T^{-1}\Sigma_1^T\varepsilon_{1t}y_{2t-1}' \rightarrow_d \int_0^1 dB_1B_2' ,$$

where B_1 $(r \times 1)$ and B_2 $(n-r \times 1)$ are subvectors of the Brownian motion $B = (B_1', B_2')' \equiv BM(\Sigma_{\varepsilon\varepsilon})$. Now, although $E(\varepsilon_{1t}y_{2t-1}') = 0$, the limit processes B_1 and B_2 will be correlated Brownian motions whenever the contemporaneous correlation between ε_{1t} and ε_{2t} is nonzero (i.e., when $\Sigma_{\varepsilon\varepsilon}$ is not block diagonal). This correlation between B_2 (the limit process of $T^{-1/2}y_{2t-1}$) and B_1 (the limit process of partial sums of ε_{1t}) is the manifestation in the limit of the "endogeneity" of the regressor y_{2t-1} in (1a).

The effects of the "endogeneity" of the regressor y_{2t-1} on a levels VAR regression are simple to determine. It is most convenient to write the first subsystem of (1) as

$$\begin{aligned} y_{1t} &= A_{11}y_{1t-1} + A_{12}y_{2t-1} + \varepsilon_{1t} \\ &= A_{11}v_{t-1} + A_{12}y_{2t-1} + \varepsilon_{1t} , \end{aligned} \qquad (1a'')$$

since the true value of $A_{11} = 0$. Estimates of A_{11} and A_{12} from a levels VAR on (1) are equivalent to those obtained by OLS on the last equation above, i.e. on (1a″). Since v_{t-1} is stationary, the OLS estimator of A_{12} $(= B)$ in (1a″) is asymptotically equivalent to the OLS estimator of B in the restricted model (1a). The limit distribution is given by the following expression

$$T(\hat{B}-B) = (T^{-1}\Sigma_1^T\varepsilon_{1t}y_{2t-1}')\left(T^{-1}\Sigma_1^Ty_{2t-1}y_{2t-1}'\right)^{-1} \rightarrow_d \left(\int_0^1 dB_1B_2'\right)\left(\int_0^1 B_2B_2'\right)^{-1} ,$$

whose far right side we can decompose into two terms (following Phillips, 1991a) as

$$\left(\int_0^1 dB_{1\cdot2}B_2'\right)\left(\int_0^1 B_2B_2'\right)^{-1} + \Sigma_{12}\Sigma_{22}^{-1}\left(\int_0^1 dB_2B_2'\right)\left(\int_0^1 B_2B_2'\right)^{-1} ,$$

where $B_{1\cdot 2} = B_1 - \Sigma_{12}\Sigma_{22}^{-1}B_2 \equiv BM(\Sigma_{11\cdot 2})$ with $\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. The second term in the above expression is the "simultaneous equations bias" that results from the "endogeneity" of the nonstationary regressor y_{2t-1} in equation (1a). This term leads to a miscentering and skewness of the limit distribution of $\hat{B}$ and its dependence on nonscale nuisance parameters that are impossible to eliminate *in toto* at least in general VAR regressions. The first term in the above expression is the limit distribution of the optimal estimator under Gaussian errors ε_t in (1), as shown in Phillips (1991a).

To deal with the fact that levels VAR's are not "reduced forms" when some of the variables are nonstationary we need to find ways of dealing with potential endogeneities of the predetermined variables. Since these endogeneities arise from cointegrating linkages of the type (1a′), one way of proceeding is to pretest the data for the presence of cointegration and the rank of the cointegration space, which in the simple example above is just the rank of the coefficient matrix $I-A$. One can then perform a reduced rank regression to obtain an optimal estimate of the submatrix B (after suitable transformations), as in Johansen (1988). Other methods, such as those in Phillips (1991a, 1991c), are also possible.

This paper considers an alternate approach that is more in keeping with the principle of unrestricted levels VAR regression. Our proposal is to deal with potential endogeneities by making a correction to the OLS-VAR regression formula that adjusts for whatever endogeneities there may be in the predetermined variables that is due to their nonstationarity. We seek to make these adjustments without knowing in advance the directions in which the variables may be stationary and what the rank of the cointegration space may be. We also seek to avoid pretest or sequential inferential procedures so that our approach maintains the essential methodology of the unrestricted vector autoregression. In the absence of prior or pre-test information about the cointegration space, we need to allow for our correction to be sufficiently general to accommodate all potential endogeneities and our procedure must be capable of handling variables that are stationary in some directions and nonstationary in others without knowing these directions in advance and while preserving the usual VAR limit theory for the stationary components. Our method of achieving this is to use in the VAR context a version of the fully modified least squares (FM-OLS) procedure in Phillips and Hansen (1990). The precise details of our approach are laid out in Section 5. The next section shows how the asymptotic theory of FM-OLS regression can be extended to accommodate the type of situations that arise in general time series regressions where the dimension of the cointegration space is unknown. This theory is an essential element in dealing with the case of a general VAR with some unit roots.

3. FM-OLS REGRESSION WITH COINTEGRATED AND STATIONARY REGRESSORS

The basic model we will work with in this section has the form

(3) $$y_t = Ax_t + u_{0t} ,$$

where A is an $n \times m$ coefficient matrix and x_t is an $m = (m_1+m_2)$-dimensional vector of cointegrated or possibly stationary regressors that are specified according to the following equations

$$H_1'x_t = x_{1t} = u_{1t} , \quad (m_1 \times 1)$$
$$H_2'\Delta x_t = \Delta x_{2t} = u_{2t} . \quad (m_2 \times 1)$$

Here $H = [H_1, H_2]$ is $m \times m$ orthogonal and rotates the regressor space in (3) so that the model has the alternative form

(3′) $$y_t = A_1x_{1t} + A_2x_{2t} + u_{0t}$$

where $A_1 = AH_1$ and $A_2 = AH_2$.

Let $u_t = (u_{0t}', u_{1t}', u_{2t}')'$ and $\varphi_t = u_{0t} \otimes u_{1t}$. It is convenient for our development to assume that u_t is a linear process that satisfies

Assumption EC (*Error Condition*)

(a) $u_t = C(L)\varepsilon_t = \Sigma_{j=0}^{\infty}C_j\varepsilon_{t-j}$, $\Sigma_0^{\infty}j^a\|C_j\| < \infty$, $|C(1)| \neq 0$ *for some* $a > 1$.

(b) ε_t *is* iid *with zero mean, variance matrix* $\Sigma_\varepsilon > 0$ *and finite fourth order cumulants.*

(c) $E(\varphi_{t,j}) = E(u_{0t+j} \otimes u_{1t}) = 0$ *for all* $j \geq 0$. □

By a multivariate extension of Theorems 3.4 and 3.8 of Phillips-Solo (1992), Assumption EC ensures the validity of functional central limit theorems for u_t and $u_t u_t'$. In particular, we have

$$n^{-1/2}\Sigma_1^{[n\cdot]}u_t \longrightarrow_d B(\cdot) = BM(\Omega) \ , \quad \Omega = C(1)\Sigma_\varepsilon C(1)' \tag{4}$$

and

$$n^{-1/2}\Sigma_1^n \varphi_{t,0} \longrightarrow_d N(0,\ \Omega_{\varphi\varphi}) \ , \quad \Omega_{\varphi\varphi} = \Sigma_{j=-\infty}^{\infty} E(u_{0t}u_{0t+j}' \otimes u_{1t}u_{1t+j}') \ . \tag{5}$$

The variance matrix Σ and long-run variance matrix Ω of u_t are partitioned into cell submatrices Σ_{ij} and Ω_{ij} $(i, j = 0, 1, 2)$ conformably with u_t. We similarly partition the Brownian motion B in (4) into cell vectors B_i $(i = 0, 1, 2)$. When u_{0t} and u_{1s} are independent for all t, s we have $\Omega_{\varphi\varphi} = \Sigma_{j=-\infty}^{\infty} E(u_{0t}u_{0t+j}') \otimes E(u_{1t}u_{1t+j}')$ and when, in addition, $u_{0t} \equiv \text{iid}(0,\ \Sigma_{00})$ we have $\Omega_{\varphi\varphi} = \Sigma_{00} \otimes \Sigma_{11}$.

We will also need the one-sided long-run covariance matrices

$$\Delta = \Sigma_{j=0}^{\infty} E(u_j u_0') = \Sigma_{j=0}^{\infty}\Gamma(j) = (\Delta_{ij}) \ ,$$

and

$$\Lambda = \Sigma_{j=1}^{\infty} E(u_j u_0') = \Sigma_{j=1}^{\infty}\Gamma(j) = (\Lambda_{ij}) \ ,$$

where the cell submatrices Δ_{ij} and Λ_{ij} $(i, j = 0, 1, 2)$ again conform to the partition of the vector u_t.

Both Ω and Δ are typically estimated by kernel smoothing of the component sample autocovariances. Since u_{0t} must itself be estimated, we will use in its place in these calculations the residuals $\hat{u}_{0t} = y_t - \hat{A}x_t$ from a preliminary least squares regression on (5). Under EC(c), $\hat{A} \longrightarrow_p A$ and the replacement of u_{0t} by $\hat{u}_{0t}$ will not affect our results.

Kernel estimates of Ω and Δ have the general form

$$\hat{\Omega} = \Sigma_{j=-T+1}^{T-1} w(j/K)\hat{\Gamma}(j) \ , \quad \text{and} \quad \hat{\Delta} = \Sigma_{j=0}^{T-1} w(j/K)\hat{\Gamma}(j) \ , \tag{6}$$

where $w(\cdot)$ is a kernel function and K is a lag truncation or bandwidth parameter. Truncation in the sums given in (6) occurs when $w(j/K) = 0$ for $|j| \geq K$. The sample covariances in (6) are given by

$$\hat{\Gamma}(j) = T^{-1}\Sigma'\hat{u}_{t+j}\hat{u}_t' \ , \quad \hat{u}_t = (\hat{u}_{0t}', u_{1t}', u_{2t}')'$$

where Σ' signifies summation over $1 \le t,\ t+j \le T$. The class of admissible kernels that we employ is made explicit in

ASSUMPTION KL (*Kernel Condition*): *The kernel function* $w(\cdot) : \mathbf{R} \to [-1, 1]$ *is a twice continuously differentiable even function with:*

(a) $w(0) = 1, w'(0) = 0, w''(0) \neq 0;$

and either

(b) $w(x) = 0,\ |x| \geq 1;$

with

$$\lim_{|x|\to 1} w(x)/(|1-|x||)^2 = \text{constant} \ .$$

or

(b′) $w(x) = O(x^{-2})$, *as* $|x| \to \infty$. □

Under KL we have

$$\lim_{x\to 0} (1 - w(x))/x^2 = -(1/2)w''(0) \ ,$$

and thus Parzen's characteristic exponent of the kernel $w(x)$ is $r = 2$. Under KL with (a) and (b) come the commonly used Parzen and Tukey-Hanning kernels and under KL with (a) and (b′) comes the Bartlett-Priestley or quadratic spectral kernel (e.g. see Priestley, 1981, p. 463). Assumption KL is similar to kernel conditions employed in earlier econometric work (see Andrews, 1991; Phillips, 1992; Kitamura-Phillips, 1992) but is somewhat more restrictive. The explicit exponent ($r = 2$ and $w''(0) \neq 0$), truncation (KL(b)) and derivative requirements in KL are helpful in achieving explicit formulae in some of our asymptotic developments. They could be relaxed at the cost of greater complexity in some of our proofs and with some changes in our final formulae and convergence rates.

It will also be useful to be explicit about the bandwidth expansion rate and we assume the following

ASSUMPTION BW (*Bandwidth expansion rate*) *The bandwidth parameter K in the kernel estimates* (6) *has an expansion rate of the form*

BW(i) $K = O(T^k)$ *for some* $k \in (1/4, 2/3)$ (*i.e.* $K^{3/2}/T \to 0$ *and* $K^4/T \to \infty$ *as* $T \to \infty$).

Some of our results require other bandwidth expansion rates which we designate as

BW(ii) $K = O(T^k)$ *for some* $k \in (1/4, 1/3)$

BW(iii) $K = O(T^k)$ *for some* $k \in (1/4, 1)$

BW(iv) $K = O(T^k)$ *for some* $k \in (0, 1)$. □

As noted in Phillips (1992), conditions like BW(i)-(iii) rule out the "optimal" growth rate $K = O(T^{1/5})$ that applies when minimizing the asymptotic mean squared error of kernel estimates such as $\hat{\Omega}$ with kernels that satisfy KL. However, since our objective is estimation of the model (3) and estimation of Ω and Δ arise only incidentally in this process, it is perhaps not too surprising that BW is not fully compatible with the "optimal" estimation of these nuisance parameters. The reason for Assumption BW and the role of the exponent k that appears in BW(i)-(iv) will become clear in our later analysis.

We now define $u_{ht} = (\Delta u_{1t}', u_{2t}')'$ ($= \Delta x_{ht} = H'\Delta x_t = H'u_{xt}$, say) using the subscript "h" to signify that elements corresponding to Δu_{1t} and u_{2t}, which occur after use of the rotation H, are taken together. In a similar way, we define the long-run covariance matrices Ω_{0h}, Ω_{hh}, Δ_{0h}, Δ_{hh} and their kernel estimates in terms of the autocovariances and sample autocovariances of u_{ht}. We observe that the leading submatrix of Ω_{hh} corresponding to the difference Δu_{1t}, viz. $\Omega_{\Delta u_1 \Delta u_1}$, is a zero matrix, since Δu_{1t} is an I(-1) process and has zero long-run variance. The following lemma describes the limit behavior of the component submatrices of these long-run covariance matrices more precisely.

3.1 LEMMA: *Under Assumptions* EC, KL *and* BW, *the following hold:*

(a) $K^2\hat{\Omega}_{\Delta u_1\Delta u_1} \longrightarrow_p -w''(0)\Omega_{11}$;

(b) $\hat{\Omega}_{u_0\Delta u_1} = K^{-2}w''(0)\Phi_{01} + O_p(1/\sqrt{KT})$, *where* $\Phi_{01} = \Sigma_{j=-\infty}^{\infty}(j-1/2)\Gamma_{u_0u_1}(j)$, *and*
$\hat{\Omega}_{u_1\Delta u_1} = K^{-2}w''(0)\Phi_{21} + O_p(1/\sqrt{KT})$, *where* $\Phi_{21} = \Sigma_{j=-\infty}^{\infty}(j-1/2)\Gamma_{u_2u_1}(j)$;

(c) $\hat{\Omega}_{0\Delta u_1} := \hat{\Omega}_{\hat{u}_0\Delta u_1} = \hat{\Omega}_{u_0\Delta u_1} + O_p(1/T)$;

(d) $\hat{\Omega}_{0h}\hat{\Omega}_{hh}^{-1} = \left[-(\Phi_{01} - \Omega_{02}\Omega_{22}^{-1}\Phi_{21})\Omega_{11}^{-1} + O_p((K^3/T)^{1/2}) + o_p((K^3/T)^{1/2}) \;\vdots\; \Omega_{02}\Omega_{22}^{-1} + o_p(1)\right]$;

(e) $K^2[T^{-1}\Delta u_1'u_1 - \hat{\Delta}_{\Delta u_1\Delta u_1}] \longrightarrow_p w''(0)\{\Delta_{11} - (1/2)\Sigma_{11}\}$;

(f) $T^{-1}u_2'u_1 - \hat{\Delta}_{u_2\Delta u_1} = K^{-2}w''(0)\Psi_{21} + O_p(1/\sqrt{KT})$, *where* $\Psi_{21} = \Sigma_{j=1}^{\infty}(j-1/2)\Gamma_{u_2u_1}(j)$;

(g) $T^{-1}\Delta u_1'X_2 - \hat{\Delta}_{\Delta u_1u_2} = T^{-1}u_{1T}x_{2T}' + K^{-2}w''(0)\Psi_{12} + O_p(1/\sqrt{KT})$, *where*
$\Psi_{12} = \Sigma_{j=0}^{\infty}(j+1/2)\Gamma_{u_1u_2}(j)$;

(h) $\hat{\Delta}_{0\Delta u_1} := \hat{\Delta}_{\hat{u}_0\Delta u_1} = O_p(1/\sqrt{KT})$;

(i) $\hat{\Delta}_{0u_2} := \hat{\Delta}_{\hat{u}_0u_2} = \Delta_{02} + O_p((K/T)^{1/2})$;

(j) $T^{-1}U_2'X_2 - \hat{\Delta}_{u_2u_2} = N_{22T} \longrightarrow_d \int_0^1 dB_2B_2'$;

(k) $T^{-1}U_0'X_2 - \hat{\Delta}_{0u_2} = N_{02T} \longrightarrow_d \int_0^1 dB_0B_2'$;

(ℓ) $T^{-2}X_2'X_2 \longrightarrow_d \int_0^1 B_2B_2'$. □

3.2 REMARKS

(a) Result (a) shows that $\hat{\Omega}_{\Delta u_1\Delta u_1} = O_p(K^{-2})$, giving the rate at which $\hat{\Omega}_{\Delta u_1\Delta u_1}$ converges to the zero matrix in the limit. Note that one consequence of the explicit representation of the limit of $K^2\hat{\Omega}_{\Delta u_1\Delta u_1}$ is that we can describe the behavior of its inverse, viz.

$$K^{-2}\hat{\Omega}_{\Delta u_1\Delta u_1}^{-1} \longrightarrow_p -(1/w''(0))\Omega_{11}^{-1} .$$

(b) Results (b) and (c) show that $\hat{\Omega}_{0\Delta u_1}$ also converges to a zero matrix, but at a rate that may differ from that of $\hat{\Omega}_{\Delta u_1\Delta u_1}$ depending on the expansion rate of K as $T \to \infty$. In particular, if $K = O(T^k)$ (with $k > 1/4$ as in Assumptions BW(i)-(iii)) we get

$$K^2\hat{\Omega}_{0\Delta u_1} = w''(0)\Phi_{01} + O_p(T^{2k-(k+1)/2}) \rightarrow_p w''(0)\Phi_{01} \ , \quad \text{for} \quad k < 1/3$$

and

$$\hat{\Delta}_{0\Delta u_1} = O_p(1/\sqrt{KT}) = O_p(T^{-(k+1)/2}) \ , \quad \text{for} \quad k > 1/3 \ .$$

Thus, for $K = O(T^k)$ with $1/4 < k < 1/3$ the rate of convergence of $\hat{\Omega}_{0\Delta u_1}$ to zero is the same as that of $\hat{\Omega}_{\Delta u_1 \Delta u_1}$. But for $K = O(T^k)$ with $k > 1/3$ the convergence rate of $\hat{\Omega}_{0\Delta u_1}$ to zero is slower than that of $\hat{\Omega}_{\Delta u_1 \Delta u_1}$. This difference and the way in which it depends on the expansion rate of K is important. As we will see, it affects the order of magnitude of terms that appear in the expansion of the estimation errors of the stationary component in the model (3′).

(c) From result (d) of the lemma we see that the first block submatrix of $\hat{\Omega}_{0h}\hat{\Omega}_{hh}^{-1}$ has elements that are of order $O((K^3/T)^{1/2})$. If $K = O(T^k)$ with $k > 1/3$ then these terms dominate and the elements of this submatrix diverge as $T \to \infty$. However, when $1/4 < k < 1/3$ we have

$$\hat{\Omega}_{0h}\hat{\Omega}_{hh}^{-1} \rightarrow_p [-(\Phi_{01} - \Omega_{02}\Omega_{22}^{-1}\Phi_2)\Omega_{11}^{-1} \ \vdots \ \Omega_{02}\Omega_{22}^{-1}]$$

and this matrix is well behaved as $T \to \infty$. Thus, even though some elements of $\hat{\Omega}_{hh}^{-1}$ diverge as $T \to \infty$ (corresponding to the fact that some elements of $u_{ht} = (\Delta u_{1t}, u_{2t})$ are I(-1) processes with a null long-run covariance matrix) the matrix product $\hat{\Omega}_{0h}\hat{\Omega}_{hh}^{-1}$ has a finite probability limit, at least when $K = O(T^k)$ and $1/4 < k < 1/3$.

(d) Remarks similar to (a) and (b) above apply also to the results (e), (f) and (g) for the correction terms that involve one-sided long-run covariance matrix estimates. These remarks indicate that the bandwidth expansion rate has an important role to play in the asymptotics of the FM estimator when there are stationary components in the estimated model, like x_{1t} in (3′).

(e) Combining the results in Lemma 3.1 we obtain expressions for the asymptotic behavior of the component elements (or correction terms) that appear in the FM estimator that is defined in equation (7) below. We give these expressions in the next lemma.

3.3 LEMMA: *Under Assumptions* EC, KL *and* BW *we have:*

(a) $\hat{\Omega}_{0h}\hat{\Omega}_{hh}^{-1}[T^{-1}U_k'X_h - \hat{\Delta}_{hh}]$

$$= [O_p(K^{-2}) + O_p(1/\sqrt{KT}) \;\vdots\; \Omega_{02}\Omega_{22}^{-1}N_{22T} + O_p(T^{-1/2}) + O_p(K^{3/2}/T) + o_p(1)]$$

where $N_{22T} \longrightarrow_d \int_0^1 dB_2B_2'$;

(b) $T^{1/2}\hat{\Omega}_{0h}\hat{\Omega}_{hh}^{-1}[T^{-1}U_k'X_1 - \hat{\Delta}_{h\Delta u_1}] = O_p(K^{-2}T^{1/2}) + O_p(K^{-1/2})$;

(c) $T^{1/2}[T^{-1}U_0'X_1 - \hat{\Delta}_{0\Delta u_1}] = T^{-1/2}U_0'X_1 + O_p(K^{-1/2}) \longrightarrow_d N(0, \Omega_{\varphi\varphi})$. □

3.4 REMARKS

(a) The partition in the matrix that appears in part (a) of Lemma 3.2 corresponds to the separation of the FM correction terms into those that relate to the stationary and nonstationary coefficients, respectively. Part (b) gives the stationary coefficient correction more explicitly (and when it is scaled by $T^{1/2}$, as it is in the analysis of the limit distribution of the FM estimates of the stationary coefficients). The correction term in this case has magnitude of order $O_p(K^{-2}T^{1/2}) + O_p(K^{-1/2})$ which is $o_p(1)$ when the bandwidth expansion rate $K = O(T^k)$ satisfies $k > 1/4$. Part (c) shows that the FM correction term for serial correlation also has no effect asymptotically and is $O_p(K^{-1/2})$. Both these results indicate that, at least for the *stationary* coefficients, the faster the bandwidth expansion rate $K = O(T^k)$, the closer the FM estimates will be to the OLS estimates which under Assumption EC(c) are consistent.

(b) The second submatrix in the partition that appears in part (a) relates to the FM endogeneity correction for the nonstationary coefficients. For the endogeneity correction to work we want this matrix to be $O_p(1)$ and to be as close to its dominating term, viz. $\Omega_{02}\Omega_{22}^{-1}N_{22T}$, as possible. Note that the error in this case involves a term of order $O_p(K^{3/2}/T)$. Thus the correction term operates satisfactorily provided $K = O(T^k)$ with $0 < k < 2/3$. In this case, therefore, we do not want the bandwidth to grow too fast with T.

(c) Combining the effects of the error terms for the stationary and the nonstationary coefficients we see that the correction terms work satisfactorily provided the bandwidth expansion rate $K = O(T^k)$ satisfies $1/4 < k < 2/3$, i.e. the rate BW(i) given in Assumption BW. □

3.5. THE FM-OLS ESTIMATOR

Lemma 3.3 makes it easy to derive a limit theory for regression estimators that depend on long-run covariance matrix estimates like the FM-OLS estimator. The FM estimator given in (7) below is constructed by making corrections for endogeneity and for serial correlation to the least squares estimator $\hat{A} = Y'X(X'X)^{-1}$ in (3). The endogeneity correction is achieved by modifying the variable y_t in (3) with the transformation

$$y_t^+ = y_t - \hat{\Omega}_{0x}\hat{\Omega}_{xx}^{-1}\Delta x_t \ .$$

In this transformation $\hat{\Omega}_{0x}$ and $\hat{\Omega}_{xx}$ are kernel estimates of the long-run covariances, $\Omega_{0x} = \text{lrcov}(u_{0t}, \Delta x_t)$ and $\Omega_{xx} = \text{lrcov}(\Delta x_t, \Delta x_t)$. The serial correlation correction term has the form

$$\hat{\Delta}_{0x}^+ = \hat{\Delta}_{0x} - \hat{\Omega}_{0x}\hat{\Omega}_{xx}^{-1}\hat{\Delta}_{xx} \ ,$$

where $\hat{\Delta}_{0x}$ and $\hat{\Delta}_{xx}$ are kernel estimates of the one-sided long-run covariance matrices $\Delta_{0x} = \text{lrcov}_+(u_{0t}, \Delta x_t)$ and $\Delta_{xx} = \text{lrcov}_+(\Delta x_t, \Delta x_t)$. Combining these two corrections we have the FM-OLS regression formula

(7) $$\hat{A}^+ = (Y^{+\prime}X - T\hat{\Delta}_{0x}^+)(X'X)^{-1} \ .$$

In deriving a limit theory for $\hat{A}^+$ we need to pay attention not only to the sample moment matrices of the data and their orders of magnitude (which in turn depend on the directions of stationarity and nonstationarity in the regressors), but also to the behavior of the kernel estimates $\hat{\Delta}_{0x}$, $\hat{\Delta}_{xx}$, $\hat{\Omega}_{0x}$ and $\hat{\Omega}_{xx}$ that appear in the correction terms of $\hat{A}^+$. The latter is especially important in the present case because the presence of stationary components (*viz.*, x_{1t}) in the regressors x_t means that the kernel estimator $\hat{\Omega}_{xx}$ tends to a singular limit due to the fact that $\Omega_{x_1x_1} = H_1'\Omega_{xx}H_1 = 0$. Lemmas 3.1 and 3.3 enable us to take this singularity into account in the asymptotic analysis and determine the impact it has on the asymptotic behavior of the estimator $\hat{A}^+$ in both stationary and nonstationary directions. In this regard, the bandwidth expansion rate of K is especially important in determining the error rate of convergence and this is why the results given in these lemmas are stated with several explicit error orders of magnitude, which arise from possibly different sources but which all need to be monitored to ensure that the given limit theory applies.

Thus, with these preliminary results in hand, we can proceed to derive the limit theory for the FM-OLS estimator $\hat{A}^+$. It is helpful in formulating our asymptotic theory to consider the component submatrices $A_1 = AH_1$ and $A_2 = AH_2$ in the model (3′) that correspond to the stationary and nonstationary elements of the regressors. We have:

3.6. THEOREM: *Under Assumptions* EC, KL *and* BW

(a) $\sqrt{T}(\hat{A}^+ - A)H_1 \to_d N(0,\ (I \otimes \Sigma_{11}^{-1})\Omega_{\varphi\varphi}(I \otimes \Sigma_{11}^{-1}))$,

(b) $T(\hat{A}^+ - A)H_2 \to_d \left(\int_0^1 dB_{0\cdot2}B_2'\right)\left(\int_0^1 B_2B_2'\right)^{-1}$,

where $B_{0\cdot2} = B_0 - \Omega_{02}\Omega_{22}^{-1}B_2 = BM(\Omega_{00\cdot2})$ *and* $\Omega_{00\cdot2} = \Omega_{00} - \Omega_{02}\Omega_{22}^{-1}\Omega_{02}$. *Part* (a) *holds for the bandwidth expansion rate* BW(iii), *i.e.* $K = O(T^k)$ *with* $1/4 < k < 1$. *The bandwidth expansion rate required for part* (b) *to hold is* $0 < k < 2/3$. *Parts* (a) *and* (b) *both hold when* $K = O(T^k)$ *and* $1/4 < k < 2/3$, *i.e. under* BW(i). □

3.7. COROLLARY (*Stationary regressor case*): *When* $m_2 = 0$ *in model* (3′) *and under Assumptions* EC, KL *and* BW *with bandwidth expansion rate* $K = O(T^k)$ *for* $1/4 < k < 1$ *we have*

$$\sqrt{T}(\hat{A}^+ - A) \to_d N(0,(I \otimes \Sigma_{11}^{-1})\Omega_{\varphi\varphi}(I \otimes \Sigma_{11}^{-1}))\ . \quad \square$$

3.8. COROLLARY (*Full rank integrated regressor case*): *When* $m_1 = 0$ *in model* (3′) *and under Assumptions* EC, KL *and* BW *with bandwidth expansion rate* $K = O(T^k)$ *for* $0 < k < 1$ *we have*

$$T(\hat{A}^+ - A) \to_d \left(\int_0^1 dB_{0\cdot2}B_2'\right)\left(\int_0^1 B_2B_2'\right)^{-1}\ . \quad \square$$

3.9. REMARKS

(a) Corollary 3.7 shows that the FM estimator $\hat{A}^+$ is consistent and has the same limit distribution as the OLS estimator $\hat{A}$ in the case where $\hat{A}$ is itself consistent, i.e. under Assumption EC(c). Note that EC(c) allows the equation error u_{0t} to be serially dependent and in this event the estimator $\hat{A}$ (and hence $\hat{A}^+$) is not necessarily efficient. However, efficient GLS-type extensions of $\hat{A}^+$ can be constructed along the lines of the FM-GIVE estimator developed in Kitamura and Phillips (1992). They will not be explored in this paper.

(b) Let $\eta_t = (u_{0t}', u_{1t+1}')'$ and $\mathcal{F}_{\eta t} = \sigma(\eta_t, \eta_{t-1}, \ldots)$ be the σ-algebra generated by $(\eta_j)_{-\infty}^t$. The condition

EC(c'): $(u_{0t}, \mathcal{F}_{\eta t})$ *is a martingale difference sequence* (mds)

ensures that $E(u_{0t+j}\, u_{1t}') = 0$ for all $j \geq 0$ and hence EC(c) holds. Moreover, under EC(c') we have

$$E(u_{0t} u_{0t+j} \otimes u_{1t} u_{1t+j}') = \begin{cases} 0 & \text{for all } \; j \neq 0 \\ \Sigma_{00} \otimes \Sigma_{11} & \text{for } \quad j = 0 \end{cases}$$

and therefore $\Omega_{\varphi\varphi} = \Sigma_{00} \otimes \Sigma_{11}$. In this case, the asymptotics

$$(8) \qquad \sqrt{T}(\hat{A}^+ - A) \longrightarrow_d N(0,\, \Sigma_{00} \otimes \Sigma_{11}^{-1})$$

correspond to those of the usual multivariate linear regression model with mds errors.

(c) One case where condition EC(c') is especially relevant occurs when there are lagged dependent variables in the regressor set. Suppose some linear combinations of the dependent variable y_t in (3) are stationary and are also independent of future realizations of the equation error u_{0t} which are pure innovations. If the stationary variables x_{1t} in the transformed system (3') include these variables in lagged form, then EC(c') holds and we get the limit theory given in (8). This situation arises in stationary autoregressions and will be examined further in the next section of the paper.

(d) As it stands Theorem 3.6 says nothing about possible dependence between the limit distributions of the stationary and nonstationary components given in parts (a) and (b) of the theorem. It turns out that with a slight strengthening of condition EC(c′) we can establish that these distributions are independent. Let $\eta_t = (u'_{0t}, u'_{1t+1}, u'_{2t+1})'$ and $\mathcal{F}_{\eta t} = \sigma(\eta_t, \eta_{t-1}, \ldots)$ be the σ-algebra generated by $(\eta_j))_{-\infty}^{t}$. This enlarges the σ-algebra used in condition EC(c′) in Remark (b) above. The condition

EC(c″): *$(u_{0t}, \mathcal{F}_{\eta t})$ is a martingale difference sequence with $E(u_{0t}u'_{0t}|\mathcal{F}_{\eta-1}) = \Sigma_{00}$* a.s.

is stronger than EC(c′) and ensures that, in addition, $E(u_{0t+j}u'_{2t}) = 0$ for all $j \geq 0$. As the proof of Theorem 3.6 makes clear, the limit distribution in (a) depends on that of $T^{-1/2}U'_0X_1 = T^{-1/2}\Sigma_1^T u_{0t}x'_{1t} = T^{-1/2}\Sigma_1^T u_{0t}u'_{1t}$. The limit distribution in (b) depends on that of $T^{-1}U'_0X_2$, $T^{-1}U'_2X_2$ and $T^{-2}X'_2X_2$, which in turn depend on the limit of the process $T^{-1/2}\Sigma_1^{[T\cdot]}(u'_{0t}, u'_{2t})'$. Under EC(c″) we have

$$E(u_{0t} \otimes u_{0t} \otimes u_{1t}) = E[I \otimes I \otimes u_{1t}][E\{(u_{0t} \otimes u_{0t} \otimes 1)|\mathcal{F}_{\eta t-1}\}] = 0 ,$$

and

$$E(u_{0t} \otimes u_{1t} \otimes u_{2t}) = E\{E[u_{0t} \otimes u_{1t} \otimes u_{2t}|\mathcal{F}_{\eta t-1}]\} = 0 ,$$

so that the limit distributions of $T^{-1/2}\Sigma_1^T u_{0t}u'_{1t}$ and $T^{-1/2}\Sigma_1^{[T\cdot]}(u'_{0t}, u'_{2t})'$ are uncorrelated and, being Gaussian, are therefore independent. The functionals of these limit processes that appear in parts (a) and (b) of Theorem 3.6 are therefore also independent. Hence, under condition EC(c″), $\sqrt{T}(\hat{A}_1^+ - A_1)$ and $\sqrt{T}(\hat{A}_2^+ - A_2)$ are independent in the limit. An important case where condition EC(c″) holds is the vector autoregressive model with some unit roots and this will be our subject of analysis in Section 5.

(e) The limit theory for the nonstationary coefficients that is given in Theorem 3.6(b) and Corollary 3.8 applies without making any condition like EC(c) or EC(c′) on the stationary components of the system. This limit theory corresponds to that of the optimal estimator obtained by maximum likelihood under Gaussian errors which was derived in Phillips (1991). Thus, even if EC(c) does not hold and the OLS and FM-OLS estimators of the stationary components are inconsistent, the FM-OLS estimator of the nonstationary component is still an optimal estimator. This is because we still have a negligible contribution from the I(0) component in the I(1) asymptotics. In particular,

$$T^{-1}U_0'U_1 - \hat{\Delta}_{0\Delta U_1} = O_p(K^{-2}) + O_p(1/\sqrt{KT})$$

and

$$T^{-1}\Delta U_1 U_1 - \hat{\Delta}_{\Delta U_1 \Delta U_1} = O_p(K^{-2})$$

as in the proofs of Lemma 3.1(e) and (f). Hence, referring to the proof of Theorem 3.6, the first term in (P31) -- which carries the effects of the estimation of the stationary components on the asymptotics for the nonstationary coefficients -- is $o_p(1)$ as $T \to \infty$ and can therefore be neglected.

(f) From Theorem 3.6 we get the (potentially degenerate) asymptotics for the full coefficient matrix $\hat{A}^+$, viz.

$$\sqrt{T}(\hat{A}^+ - A) = \sqrt{(}\hat{A}^+ - A)HH' = \sqrt{T}(\hat{A}_1^+ - A_1)H_1' + \sqrt{T}(\hat{A}_2^+ - A_2)H_2'$$

(9) $$\longrightarrow_d N(0,\ (I \otimes H_1\Sigma_{11}^{-1})\Omega_{\varphi\varphi}(I \otimes \Sigma_{11}^{-1}H_1'))$$

(10) $$= N(0,\ \Sigma_{00} \otimes H_1\Sigma_{11}^{-1}H_1')\ ,$$

the last line holding under EC(c′).

(g) When EC(c′) holds we can construct a consistent estimate of the covariance matrix $\Sigma_{00} \otimes H_1\Sigma_{11}^{-1}H_1'$ of the limit distribution (10) directly from the matrix $\hat{\Sigma}_{00} \otimes T(X'X)^{-1}$. This is because

(11) $$T(X'X)^{-1} \longrightarrow_p H_1\Sigma_{11}^{-1}H_1'$$

(see Phillips, 1988, p. 95) and since $\hat{A}, \hat{A}^+ \longrightarrow_p A$,

$$\hat{\Sigma}_{00} = T^{-1}\Sigma_1^T\hat{u}_{0t}\hat{u}_{0t}' = T^{-1}\Sigma_1^T u_{0t}u_{0t}' + o_p(1) \longrightarrow_p \Sigma_{00}\ .$$

The covariance matrix in (9) can also be consistently estimated. We may use the matrix

$$[I \otimes T(X'X)^{-1}]\hat{\Omega}_{\hat{\varphi}_x\hat{\varphi}_x}[I \otimes T(X'X)^{-1}]\ ,$$

where $\hat{\Omega}_{\hat{\varphi}_x\hat{\varphi}_x}$ is the kernel estimate

$$\hat{\Omega}_{\hat{\varphi}_x\hat{\varphi}_x} = \sum_{j=-K+1}^{K-1} w(j/K)\hat{\Gamma}_{\hat{\varphi}_x\hat{\varphi}_x}(j) ,$$

and $\hat{\varphi}_{xt} = \hat{u}_{0t} \otimes x_t$. Noting from (11) that $T(X'X)^{-1} = H_1\Sigma_{11}^{-1}H_1' + o_p(1)$ and $H_1'x_t = x_{1t} = u_{1t}$, we have $(I \otimes H_1')\hat{\varphi}_{xt} = \hat{u}_{0t} \otimes u_{1t} = \hat{\varphi}_t$ and so

(12) $$(I \otimes H_1')\hat{\Omega}_{\hat{\varphi}_x\hat{\varphi}_x}(I \otimes H_1) = \hat{\Omega}_{\hat{\varphi}\hat{\varphi}} \rightarrow_p \Omega_{\varphi\varphi} .$$

Combining (11) and (12) we obtain

(13) $$[I \otimes T(X'X)^{-1}]\hat{\Omega}_{\hat{\varphi}_x\hat{\varphi}_x}[I \otimes T(X'X)^{-1}] \rightarrow_p (I \otimes H_1\Sigma_{11}^{-1})\Omega_{\varphi\varphi}(I \otimes \Sigma_{11}^{-1}H_1') .$$

(h) Results (9) and (13) suggest that inference about A can be performed using the asymptotic approximation

(14) $$\sqrt{T}(\hat{A}^+ - A) \sim N\left(0, [I \otimes T(X'X)^{-1}]\hat{\Omega}_{\hat{\varphi}_x\hat{\varphi}_x}[I \otimes T(X'X)^{-1}]\right) .$$

Suppose we wish to test the restrictions

$$\mathcal{H}_0 : R \text{ vec } A = r , \quad R \ (q \times nm) \text{ of rank } q .$$

A natural test statistic is the Wald statistic

(15) $$W_\varphi^+ = T(R \text{ vec } \hat{A}^+ - r)'\left\{R[I \otimes T(X'X)^{-1}]\hat{\Omega}_{\hat{\varphi}_x\hat{\varphi}_x}[I \otimes T(X'X)^{-1}]R'\right\}^{-1}(R \text{ vec } \hat{A}^+ - r) .$$

In view of (9) and (13) and provided the following rank condition holds

(RK_φ) $$\operatorname{rank}[R\{(I \otimes H_1\Sigma_{11}^{-1})\Omega_{\varphi\varphi}(I \otimes \Sigma_{11}^{-1}H_1')\}R'] = q ,$$

we have

(16) $$W_\varphi^+ \rightarrow_d \chi_q^2 , \quad \text{as} \quad T \rightarrow \infty$$

and so conventional chi-squared asymptotics apply.

(i) When Assumption EC(c′) holds, the limit distribution (10) applies and we can use the asymptotic approximation

$$(14') \qquad \sqrt{T}(\hat{A}^{+}-A) \sim N(0,\ \hat{\Sigma}_{00} \otimes T(X'X)^{-1})\ .$$

To test $\mathscr{H}_0$, the natural statistic in this case is

$$W_{00}^{+} = T(R\ \text{vec}\,\hat{A}^{+} - r)'\left[R\{\hat{\Sigma}_{00} \otimes T(X'X)^{-1}\}R'\right]^{-1}(R\ \text{vec}\,\hat{A}^{+} - r)$$

and if

$$(\text{RK}) \qquad \text{rank}[R\{\Sigma_{00} \otimes H_1\Sigma_{11}^{-1}H_1'\}R'] = q$$

we have $W_{00}^{+} \rightarrow_d \chi_q^2$ as in (16).

(j) We now consider the interesting case where the rank condition (RK) fails. This occurs when the restriction matrix R isolates some of the nonstationary coefficients. Thus, suppose $R = R_1 \otimes R_2'$ and the hypothesis $\mathscr{H}_0$ has the form

$$(17) \qquad \mathscr{H}_0 : \underset{q_1\times n}{R_1} A \underset{m\times q_2}{R_2} = \underline{R}\ , \quad \text{vec}\ \underline{R} = r\ ,$$

where R_1 and R_2 are of rank q_1 and q_2, respectively. If $R_2'H_1$ is of deficient row rank, then (RK) fails. In this case we may write

$$(18) \qquad R_2 = [\underset{q_{21}}{R_{21}} \ \vdots\ \underset{q_{22}}{R_{22}}] = [H_1, H_2]\begin{bmatrix} S_{20} & S_{h1} & \vdots & 0 \\ 0 & S_{h2} & \vdots & S_{22} \end{bmatrix}$$
$$= [H_1S_{20},\ H_1S_{h1} + H_2S_{h2}\ \vdots\ H_2S_{22}]$$

for some matrices S_{20}, S_{h1}, S_{h2} and S_{22}. Without loss of generality (and by rotating the restrictions (17), if necessary) we may assume that the matrix S_{h1} has full column rank. The hypotheses about A that correspond to the columns R_{22} of R_2 relate solely to the nonstationary coefficients in A, i.e. to $A_2 = AH_2$, because $R_1AR_{22} = R_1AH_2S_{22} = R_1A_2S_{22}$. Now $R_{22}'H_1 = 0$ and then we have

$$R\{\Sigma_{00} \otimes H_1\Sigma_{11}^{-1}H_1'\}R' = R_1\Sigma_{00}R_1' \otimes \begin{bmatrix} R_{21}'H_1 \\ 0 \end{bmatrix}\Sigma_{11}^{-1}[H_1'R_{21}, 0] ,$$

which has rank $q_1q_{21} < q_1(q_{21} + q_{22}) = q$. What is the limit distribution of the statistic W_{00}^{+} in this case when Condition RK fails? The following theorem provides the answer.

3.10. THEOREM: *Under Assumptions* EC, EC(c'), KL *and* BW *the Wald statistic* W_{00}^{+} *for testing the restrictions* H_1' : $R_1AR_2 = R$ *has a limit distribution which is a mixture of* χ^2 *variates. In particular, when* R_2 *has the form given in* (18) *we have*

$$W_{00}^{+} \to_d \Sigma_{i=1}^{q_1}\chi_{q_{21}}^2(i) + \Sigma_{j=1}^{q_1}d_j\,\chi_{q_{22}}^2(j) = \chi_{q_1q_{21}}^2 + \Sigma_{j=1}^{q_1}d_j\,\chi_{q_{22}}^2(j) , \tag{19}$$

where $\chi_{q_{21}}^2(i) = \mathrm{iid}(\chi_{q_{21}}^2)$, $\chi_{q_{22}}^2(j) = \mathrm{iid}(\chi_{q_{22}}^2)$ and $\chi_{q_{21}}^2(i)$ and $\chi_{q_{22}}^2(j)$ *are independent for all i and j. The coefficients* d_j *in* (19) *are the latent roots of the matrix* $(R_1\Omega_{00\cdot2}R_1')(R_1\Sigma_{00}R_1')^{-1}$. □

3.11. REMARKS

(a) Under EC(c'), $\Omega_{00\cdot2} = \Omega_{00} - \Omega_{02}\Omega_{22}^{-1}\Omega_{20} = \Sigma_{00} - \Omega_{02}\Omega_{22}^{-1}\Omega_{20} < \Sigma_{00}$. Thus $(R_1\Omega_{00\cdot2}R_1')^{1/2}(R_1\Sigma_{00}R_1')^{-1}(R_1\Omega_{00\cdot2}R_1')^{1/2} \le I$ and therefore the latent roots d_j ($j = 1, ..., q_1$) that appear in (19) as weights satisfy $0 < d_j \le 1$. It follows that in the limit (19) is bounded above by the variate $\chi_{q_1q_{21}}^2 + \Sigma_{j=1}^{q_1}\chi_{q_{22}}^2(j) = \chi_{q_1q_{21}}^2 + \chi_{q_1q_{22}}^2 = \chi_{q_1q_2}^2$. Tests of conservative size (asymptotically) can therefore always we constructed for W_{00}^{+} using the $\chi_{q_1q_2}^2$ distribution.

(b) Now suppose we construct the Wald statistic using the variance matrix estimator $\hat{\Omega}_{00\cdot\Delta x} = \hat{\Omega}_{00} - \hat{\Omega}_{0x}\hat{\Omega}_{xx}\hat{\Omega}_{x0} = \hat{\Omega}_{00} - \hat{\Omega}_{0h}\hat{\Omega}_{hn}^{-1}\hat{\Omega}_{h0}$ constructed from the long-run variance and covariance matrices of $\hat{u}_{0t}$ and Δx_t. Since $\hat{\Omega}_{00\cdot\Delta x} \to_p \Omega_{00} - \Omega_{02}\Omega_{22}^{-1}\Omega_{20} = \Omega_{00\cdot2}$, we obtain in the same way as Theorem 3.10 and under the same conditions the limit result

$$W_{00x}^{+} = T(R\ \mathrm{vec}\ \hat{A}^{+} - r)'\left[R\{\hat{\Omega}_{00\cdot\Delta x} \otimes T(X'X)^{-1}\}R'\right]^{-1}(R\ \mathrm{vec}\ \hat{A}^{+} - r)$$

$$\to_d \Sigma_{i=1}^{q_1}(1/d_i)\chi_{q_{21}}^2(i) + \chi_{q_1q_{22}}^2 .$$

It follows that in the limit $W^{+}_{00\cdot x}$ is bounded below by the limit distribution $\chi^2_{q_1 q_2}$. An asymptotically liberal test of the hypothesis H_0' can therefore always be constructed using $W^{+}_{00\cdot x}$.

(c) Note that $d_i = 1$ $(i = 1, ..., q_1)$ when $\Sigma_{00} = \Omega_{00\cdot 2}$, i.e., when $\Omega_{02} = 0$ or when u_{0t} and $u_{2t} = \Delta x_{2t}$ have long-run zero covariance. Observe also that when there are no nonstationary components (i.e., $x_t = I(0)$) we have $\hat{\Omega}_{00\cdot\Delta x} \to_p \Omega_{00} = \Sigma_{00}$ under EC(c′) and then both W^{+}_{00}, $W^{+}_{00x} \to_d \chi^2_{q_1 q_2}$ in the limit. When there are no stationary components in the model we have $\Omega_{00\cdot\Delta x} \to_p \Omega_{00\cdot 2}$ and again $W^{+}_{00x} \to_d \chi^2_{q_1 q_2}$. Thus, $W^{+}_{00\cdot x}$ has the desirable property of being asymptotically $\chi^2_{q_1 q_2}$ in both extreme cases (stationary regressors only or full rank nonstationary regressors). It will be interesting to explore the finite sample performance of W^{+}_{00} and W^{+}_{00x} in intermediate cases where there are both stationary and nonstationary components to the regressors.

3.12. EXTENSIONS TO MODELS WITH DETERMINISTIC REGRESSORS

The main results given earlier in this section continue to hold (with some modifications to the formulae) when there are deterministic regressors in the system (3) and when the regressors x_t may have deterministic components. The limit theory for the FM-OLS estimator and associated Wald tests can be developed as in Theorems 3.6 and 3.10. These generalizations are not difficult and we will therefore only illustrate what is involved here. For example, suppose the model (3) is replaced by

$$(3'') \qquad y_t = Ax_t + \Pi k_t + u_{0t} = \Phi z_t + u_{0t}\ , \quad \text{say}$$

where k_t is a p-vector of deterministic regressors and the vector x_t can be decomposed into I(0), I(1) and deterministic components as

$$x_t = H_1 x_{1t} + H_2 x_{2t} + Fk_t\ ,$$

for some $m \times p$ matrix F.

The regressors k_t will usually involve polynomials in time, in which case we can write

$$k_t = \left(t^{s_1}, t^{s_2}, ..., t^{s_p}\right)'\ , \quad 0 \le s_1 < s_2 < \cdots < s_p\ ,$$

for some integers s_i ($i = 1, ..., p$). Note that s_1 may be zero and we therefore allow for the presence of an intercept in (3″), a possibility which seems to be excluded in work by Hansen (1992) on FM cointegrating regressions with deterministic trends. For such regressors we use the weight matrix $\delta_T = \mathrm{diag}(T^{s_1}, ..., T^{s_p})$ and then

$$\delta_T^{-1} k_{[Tr]} \rightarrow k(r) = \left(r^{s_1}, ..., r^{s_p}\right)'$$

uniformly in $r \in [0, 1]$. The limit functions $k(r)$ are linearly independent in $L_2[0, 1]$ and $\int_0^1 kk' > 0$.

The FM-OLS estimator of Φ in (3″) is

$$\hat{\Phi}^+ = [\hat{A}^+ \ \vdots\ \hat{\Pi}^+] = (Y^{+\prime}Z - [T\hat{\Delta}_{0x}^+ \ \vdots\ 0])(Z'Z)^{-1} ,$$

which is an augmented version of (7) and a formula that was given originally in Phillips and Hansen (1990). But in the above expression the long run covariance estimates that arise in $\hat{\Delta}_{0x}^+ = \hat{\Delta}_{0x} - \hat{\Omega}_{0x}\hat{\Omega}_{xx}^{-1}\hat{\Delta}_{xx}$ are based on $(\hat{u}_{0t}', \hat{u}_{xt}')$, where $\hat{u}_{0t} = y_t - \hat{A}x_t - \hat{\pi}k_t$ is a first stage OLS residual and $\hat{u}_{xt} = \Delta\hat{u}_{kt}$ wherein $\hat{u}_{kt} = x_t - \hat{F}k_t$ is the residual from the OLS regression of x_t on k_t. We remark that if k_t involves an intercept as its lead component then the corresponding column of $\hat{F}$ is inconsistent (and, in fact, diverges) when $m_2 \geq 1$. However, this component of $\hat{F}$ is eliminated by the difference transformation $\hat{u}_{xt} = \Delta\hat{u}_{kt}$ and the remaining columns of $\hat{F}$ are consistent since $s_i \geq 1$ ($i > 1$) and the regressors k_{it} ($i > 1$) dominate the stochastic trend and stationary components of x_t. Thus, $\hat{u}_{xt} = H_1\Delta x_{1t} + H_2\Delta x_{2t} + (F-\hat{F})\Delta k_t = H_1\Delta x_{1t} + H_2\Delta x_{2t} + o_p(1)$ and therefore the correction terms work in the same way as those in regressions with no deterministic trends.

The limit theory for the components of the FM-OLS estimator $\hat{\Phi}^+$ can be deduced in much the same way as Theorem 3.6. But some care needs to be taken over the extra partitioning in Φ corresponding to the I(0) and I(1) components. Again, we will just provide the basic approach here.

In making the construction it is useful to employ a composite weight matrix of the form

$$D_T = \begin{bmatrix} [H_1 \ \vdots \ H_2 T^{1/2}] & F\delta_T \\ 0 & \delta_T \end{bmatrix}.$$

Then

$$D_T^{-1} = \begin{bmatrix} \begin{bmatrix} H_1' \\ T^{-1/2}H_2' \end{bmatrix} & -\begin{bmatrix} H_1' \\ T^{-1/2}H_2' \end{bmatrix} F \\ 0 & \delta_T^{-1} \end{bmatrix}, \quad \text{and} \quad D_T^{-1} z_t = \begin{bmatrix} x_{1t} \\ T^{-1/2}x_{2t} \\ \delta_T^{-1} k_t \end{bmatrix},$$

which reorganizes and suitably weights the components of the regressors z_t. Note that for some fixed $r > 0$

$$D_T^{-1} z_{[Tr]} \longrightarrow_d (x_{1\infty}', B_2(r)', k(r)')' = (x_{1\infty}, J(r)') \ , \quad \text{say,}$$

giving the limit processes that correspond to these standardized regressors. The limit theory for $\hat{\Phi}^+$ is now

$$\sqrt{T}(\hat{\Phi}^+ - \Phi)D_T = \left[\sqrt{T}(\hat{A}^+ - A)H_1 \ \vdots \ T(\hat{A}^+ - A)H_2 \ \vdots \ \sqrt{T}\left\{(\hat{\Pi}^+ - \Pi) + (\hat{A}^+ - A)F\right\}\delta_T\right]$$

$$\longrightarrow_d \left[N(0,(I \otimes \Sigma_{11}^{-1})\Omega_{\varphi\varphi}(I \otimes \Sigma_{11}^{-1}) \ \vdots \ \int_0^1 dB_{0\cdot 2}J'\left(\int_0^1 JJ'\right)^{-1}\right],$$

which extends Theorem 3.6 to allow for deterministic trends. The component of this limit distribution corresponding to the stationary part of x_t is identical to part (a) of Theorem 3.6, where there are no deterministic regressors. The component that corresponds to the nonstationary part of x_t differs from part (b) of Theorem 3.6 in that it involves the deterministic function $k(r)$ as part of the limit function $J(r)$. The coefficients of the nonstationary part of x_t and the deterministic regressors k_t in (3″) are taken together in the limit variate $(\int dB_{0\cdot 2}J)(\int JJ')^{-1}$. However, like part (b) of Theorem 3.6 this limit variate is mixed normal and this limit distribution facilitates statistical inference in the same way as before.

For instance, if we wish to test $\mathcal{H}_0$ the natural Wald test is

$$W_K^+ = T(R \ \text{vec} \ \hat{A}^+ - r)'\left[R\left\{\Sigma_{00} \otimes T(X'Q_K X)^{-1}\right\}R'\right]^{-1}(R \ \text{vec} \ \hat{A}^+ - r) \ ,$$

where Q_K is the orthogonal projection matrix onto $\mathcal{R}(K)^{\perp}$ and K is the matrix of observations of k_t. It is easy to show that $T(X'Q_KX)^{-1} \rightarrow_p H_1\Sigma_{11}^{-1}H_1'$ and then $W_K^{*} \rightarrow_d \chi_q^2$, provided condition RK holds. If RK fails and the hypothesis is of the form $\mathcal{H}_0$ given in (17) then W_K^{+} has the same limit as (19) and Theorem 3.10 applies.

In addition to these extensions of our theory, we can also consider the case where the regression equation (3″) does not include all of the deterministic regressors k_t. Again, closely related results are obtained. As in the case above, the limit theory for the FM-OLS estimator of the coefficients of the nonstationary part of x_t and the included deterministic regressors must be taken together but the limit distribution is still mixed normal. In consequence, Wald statistics that are formed in the usual way have limit chi-squared or mixed chi-squared distribution, just as in Theorem 3.10.

Finally, we remark that extensions of our theory to include deterministic regressors with breaking trends may also be accommodated. In this case the corresponding limit functions will involve some simple cadlag functions, as in Park (1992). The other aspects of our limit theory for $\hat{\Phi}^+$ go through as before, as does the limit theory for the associated Wald tests.

4. APPLICATION TO AUTOREGRESSION

4.1. FM-OLS IN THE STATIONARY AND NONSTATIONARY AR(1)

An important application of the theory of FM regression is to the simple AR(1) model

(20) $$y_t = \alpha y_{t-1} + u_t\ , \quad t = 1, 2, \ldots, T$$

where the initialization y_0 at $t = 0$ can be any random variable including a constant.

The case of a unit root $\alpha = 1$ in (20) was dealt with by the author in recent work (1992), where it was shown that even for general stationary errors u_t in (20) the FM estimator $\hat{\alpha}^+$ is hyperconsistent. Under BW with a bandwidth expansion rate $K = O(T^k)$ with $1/4 < k < 1/2$ the author showed that $\hat{\alpha}^+$ is $T^{3/2}$-consistent for $\alpha = 1$. The intuition behind this hyperconsistency result is as follows. Since the dependent variable y_t and the regressor y_{t-1} in (20) are cointegrated and y_{t-1} is a full rank I(1) process when $\alpha = 1$, we can expect the limit theory of Corollary

3.8 to apply. But note that $u_{0t} = u_t$ and $\Delta x_t = \Delta y_{t-1} = u_{t-1}$ so that the corresponding Brownian motions, i.e. $B_0(r)$ and $B_2(r)$, are identical. Moreover, if ω^2 is the long-run variance of u_t then $\Omega_{02} = \Omega_{22} = \omega^2$ also, because the long-run covariance of u_t and $\Delta y_{t-1} = u_{t-1}$ is the same as the long-run variance of u_t. Let $B(r) = B_0(r) = B_2(r)$, say. Then, in the present case we have

$$B_{0\cdot 2}(r) = B_0(r) - \Omega_{02}\Omega_{22}^{-1}B_2(r) = B_{0\cdot 0}(r) = B(r) - B(r) = 0 ,$$

and so the limit theory from Corollary 3.8, viz.

$$T(\hat{\alpha}^{+} - 1) = \left(\int_0^1 dB_{0\cdot 0}B\right)\left(\int_0^1 B^2(r)\right)^{-1} = 0 \tag{21}$$

is degenerate. Thus, $\hat{\alpha}^{+} = 1 + o_p(T^{-1})$ and $\hat{\alpha}^{+}$ is hyperconsistent.

Next, consider the stationary case of (20) with $|\alpha| < 1$ and suppose $u_t \equiv \text{iid}(0, \sigma^2)$. In this model the OLS estimator $\hat{\alpha}$ is consistent and its limit theory is given by $\sqrt{T}(\hat{\alpha} - \alpha) \to_d N(0, 1-\alpha^2)$. The limit theory for the FM-OLS estimator $\hat{\alpha}^{+}$ in this case is covered by Corollary 3.6. Since the error u_t is an mds the limiting distribution of $\sqrt{T}(\hat{\alpha}^{+} - \alpha)$ is given by formula (8) with $\Sigma_{00} = \sigma^2$ and $\Sigma_{11} = \text{var}(y_{t-1}) = \sigma^2/(1-\alpha^2)$. Hence, we have

$$\sqrt{T}(\hat{\alpha}^{+} - \alpha) \to_d N(0,\ 1-\alpha^2) \tag{22}$$

and $\hat{\alpha}^{+}$ is asymptotically equivalent to $\hat{\alpha}$ in the stationary case. Since $\hat{\alpha}^{+}$ is hyperconsistent at $\alpha = 1$, this confirms that the OLS estimator has an infinite deficiency relative to $\hat{\alpha}^{+}$ at $\alpha = 1$, while (at least asymptotically) $\hat{\alpha}^{+}$ is not penalized in the stationary region.

The bandwidth expansion rate for $K = O(T^k)$ under which both (21) and (22) hold is given by the intersection of the regions that apply separately for Corollaries 3.7 and 3.8, i.e. $\{1/4 < k < 1\} \cap \{0 < k < 1\} = \{1/4 < k < 1\}$. Thus, there is a common bandwidth rate for which $\hat{\alpha}^{+}$ and $\hat{\alpha}$ are asymptotically equivalent when $|\alpha| < 1$ and for which $\hat{\alpha}^{+}$ dominates $\hat{\alpha}$ when $\alpha = 1$. Under these conditions $\hat{\alpha}^{+}$ is certainly the preferred estimator in large samples.

- - - - - - - - - - - - - -

Figures 1-5 about here

- - - - - - - - - - - - - -

Figures 1-3(a & b) graph the sampling densities obtained from 10,000 Monte Carlo replications of the standardized and centered estimators $\sqrt{T}(\hat{\alpha} - \alpha)$, $\sqrt{T}(\hat{\alpha}^{+} - \alpha)$ from the model

(20) with u_t = iid $N(0, 1)$ and $y_0 = 0$. The figures show the estimated densities (based on a normal kernel with a plug in adaptive bandwidth) of these estimators for the parameter values α = 0.4, 0.8, 0.90 and T = 50 (Figure a), 100 (Figure b). The case where $\alpha = 1$ is shown in Figure 4(a & b) and here the estimators are scaled as $T(\hat{\alpha}-1)$ and $T(\hat{\alpha}^+-1)$ for comparative purposes. For all these cases the FM estimator $\hat{\alpha}^+$ is computed using the quadratic spectral kernel and the bandwidth employed was $K = T^{3/4}$ for each value of α. As is apparent from the proofs in Phillips (1992), with this bandwidth choice $\hat{\alpha}^+$ is hyperconsistent at $\alpha = 1$ (just as shown in (21) above) but does not achieve the $T^{3/2}$ consistency rate. In fact, when $K = O(T^{3/4})$ the rate of consistency is $T^{5/4}$ at $\alpha = 1$. Issues of bandwidth selection in autoregressions are presently being explored by the author and will be reported in another paper. The present computations are designed to illustrate the gains that are achievable even with rather mechanical rules.

TABLE 1

Bias and RMSE of OLS and FM-OLS Estimators of α in the AR(1) Model (20) with T = 100 (10,000 replications)

	Mean Bias		Median Bias		RMSE	
α	OLS	FM-OLS	OLS	FM-OLS	OLS	FM-OLS
1.00	-0.017	-0.013	-0.008	-0.004	0.035	0.031
0.96	-0.018	-0.010	-0.009	-0.001	0.044	0.043
0.90	-0.017	-0.009	-0.007	-0.000	0.053	0.055
0.80	-0.016	-0.011	-0.008	-0.004	0.067	0.069
0.70	-0.014	-0.011	-0.007	-0.004	0.075	0.078
0.60	-0.011	-0.010	-0.006	-0.005	0.082	0.084
0.50	-0.011	-0.010	-0.004	-0.004	0.089	0.091
0.40	-0.008	-0.009	-0.006	-0.007	0.092	0.094
0.30	-0.004	-0.005	-0.000	-0.001	0.096	0.097
0.20	-0.005	-0.005	-0.005	-0.005	0.097	0.099
0.10	-0.002	-0.004	-0.001	-0.002	0.099	0.099
0.00	-0.001	-0.002	-0.002	-0.003	0.100	0.100

Figures 1-3 show that the sampling distributions of $\hat{\alpha}$ and $\hat{\alpha}^+$ are very close for α = 0.4, and quite close for α = 0.8. At α = 0.8 and more so at α = 0.90 the distribution of $\hat{\alpha}^+$ is shifted to the right and appears to be somewhat less asymmetric than that of $\hat{\alpha}$. The characteristic long left hand tail of the distribution of $\hat{\alpha}$ is also evident in the distribution of $\hat{\alpha}^+$. It is interesting to note that for $\alpha = 0.80, 0.90$ the distribution of $\hat{\alpha}^+$ is *less* biased than that of $\hat{\alpha}$ in terms of both of the median and the mean. Table 1 gives these and other summary statistics from the Monte Carlo simulation for values of $\alpha \in [0, 1]$. As far as central location is concerned, the estimator $\hat{\alpha}^+$ has very good performance characteristics over the whole region [0, 1] and actually reduces the bias of the OLS estimator for $\alpha \in [0.5, 1]$. This bias reduction comes at a slight increase in dispersion, which is reflected in the RMSE statistics in Table 1, for $\alpha \in [0, 0.9]$.

4.2. THE NONSTATIONARY AR(1) WITH DETERMINISTIC TRENDS

Our theory for the stationary and nonstationary AR(1) model (20) can be extended to AR models with deterministic trends in much the same way as our analysis of Section 3.12. Since some of the results are of special interest we will briefly comment on them and illustrate their derivation here. Later work by the author will provide a more complete derivation and discussion.

We will take as our example the AR(1) + Tr(p) model

$$(20)' \qquad y_t = \alpha y_{t-1} + \beta' k_t + u_{0t} = \varphi' z_t + u_{0t}\,, \quad \text{say}$$

where $\alpha = 1$, $k_t = (1, t, ..., t^p)$ and $\beta = (\beta_0, \beta_1, ..., \beta_p)$ with $\beta_p = 0$. Then we can write

$$y_{t-1} = (y_0 + \Sigma_1^{t-1} u_{0j}) + \beta' \Sigma_1^{t-1} k_j = y_{t-1}^0 + \pi' k_t\,, \quad \text{say}$$

where $y_t^0 = I(1)$.

The FM-OLS estimator of φ in (20′) is

$$\hat{\varphi}^+ = [\hat{\alpha}^+, \hat{\beta}^{+\prime}] = (Y^{+\prime}Z - [T\hat{\Delta}_{0y}^+ \,\vdots\, 0])(Z'Z)^{-1}\,.$$

In this formula $\hat{\Delta}_{0y}^+ = \hat{\Delta}_{0y} - \hat{\Omega}_{0y}\hat{\Omega}_{yy}^{-1}\hat{\Delta}_{yy}$ and the long-run covariances that appear in this expression are calculated from $(\hat{u}_{0t}, \hat{u}_{yt})$ where $\hat{u}_{0t} = y_t - \hat{\alpha} y_{t-1} - \hat{\beta}' k_t$ and $\hat{u}_{yt} = \Delta \hat{u}_{kt}$ with $\hat{u}_{kt} = y_{t-1} - \bar{\pi} k_t$ (from the OLS regression of y_{t-1} on k_t).

As in Section 3.12 we employ a composite weight matrix to separate out the I(1) and deterministic regressors in (20′) and their associated coefficients. Here we use

$$D_T = \begin{bmatrix} 1 & \pi' \\ 0 & I \end{bmatrix} \begin{bmatrix} T^{1/2} & 0 \\ 0 & \delta_T \end{bmatrix} = Ed_T \text{ , say}$$

$$= \begin{bmatrix} T^{1/2} & \pi'\delta_T \\ 0 & \delta_T \end{bmatrix} = \begin{bmatrix} T^{-1/2} & -T^{-1/2}\pi' \\ 0 & \delta_T^{-1} \end{bmatrix}^{-1} .$$

with $\delta_T = \text{diag}(1, ..., T^p)$. Note that (20′) has the following form in transformed coefficients and regressors

$$(20'') \quad y_t = \varphi' E E^{-1} z_t + u_{0t} = \alpha y^0_{t-1} + \underline{\beta}' k_t + u_{0t}$$
$$= \underline{\varphi}' \underline{z}_t + u_{0t} \text{ , say}$$

where $\underline{\beta}' = \beta' + \alpha\pi' = \beta' + \pi'$ and $\underline{z}_t = (y^0_{t-1}, k_t')'$. With these transformations we have

$$d_T^{-1} \underline{z}_{[Tr]} \to_d \begin{bmatrix} B_0(r) \\ k(t) \end{bmatrix} = J(r) \text{ , say}$$

and the limit theory for $\hat{\varphi}^+$ and $\hat{\varphi}$ now follows in a straightforward way.

Note that $\Omega_{0y} = \Omega_{yy} = \omega^2 = \text{lrvar}(u_{0t})$ and hence $B_{0\cdot 0}(r) = B_0(r) - \Omega_{0y}\Omega_{yy}^{-1}B_y(r) = 0$. (Here the affix "y" signifies the use of $\hat{u}_{yt} = \Delta\hat{u}_{kt} = \Delta y_{t-1} - \hat{\pi}'\Delta k_t = u_{0t-1} + o_p(1)$). For the OLS estimator $\hat{\varphi}$ we have the limit

$$\sqrt{T} D_T'(\hat{\varphi} - \varphi) = \begin{bmatrix} T(\hat{\alpha} - \alpha) \\ \sqrt{T}\delta_T(\hat{\underline{\beta}} - \underline{\beta}) \end{bmatrix} \to_d \left(\int_0^1 JJ'\right)^{-1} \left(\int_0^1 J dB_0 + \begin{bmatrix} \Delta_{0y} \\ 0 \end{bmatrix}\right) ;$$

whereas for the FM-OLS estimator $\hat{\varphi}^+$ we have the limit

$$\sqrt{T} D_T'(\hat{\varphi}^+ - \varphi) = \begin{bmatrix} T(\hat{\alpha}^+ - \alpha) \\ \sqrt{T}\delta_T(\hat{\underline{\beta}}^+ - \underline{\beta}) \end{bmatrix} \to_d \left(\int_0^1 JJ'\right)^{-1} \left(\int_0^1 J dB_{0\cdot 0}\right) = 0 ,$$

showing that the estimator $\hat{\varphi}^+$ is hyperconsistent. The most interesting feature of this result is that $\underline{\beta}^+$ is hyperconsistent as well as $\hat{\alpha}^+$. Thus, the FM-OLS procedure accelerates the rate of convergence of the OLS estimates of the deterministic coefficients in (20″) as well as that of the unit root $\alpha = 1$. Note that we have established this result for general stationary errors u_{0t} in (20″).

The hyperconsistency of the FM-OLS estimates of the deterministic coefficients in (20″) may seem surprising. It is, however, explained by the fact that if we knew $\alpha = 1$ then we know that $y_{t-1} = y_{t-1}^0 + \pi' k_t$ and we can write equation (20′) as

$$y_t - y_{t-1}^0 - \Omega_{0y}\Omega_{yy}^{-1}\Delta y_{t-1}^0 = \underline{\beta}' k_t + u_{0t} - u_{0t-1} ,$$

that is as

$$(20''') \qquad \underline{y}_t = \underline{\beta}' k_t + e_t , \quad \text{with} \quad e_t = u_{0t} - u_{0t-1} .$$

Now OLS regression on (20‴) yields an estimator of $\underline{\beta}$ that is $O(T^2)$ consistent, as distinct from the $T^{3/2}$-consistent OLS estimator $\hat{\underline{\beta}}$. Of course, we cannot presume that $\alpha = 1$ is known, but since $\hat{\alpha}^+$ is hyperconsistent for $\alpha = 1$ there is a spillover effect of this accelerated convergence on the estimation of $\underline{\beta}$ by FM-OLS that operates in the same way as when $\alpha = 1$ is known, i.e. by effectively transforming the error in equation (20″) to an I(−1) process just as in (20‴).

5. FM VECTOR AUTOREGRESSION WITH SOME UNIT ROOTS

In this section we will consider the use of FM-OLS regression in VAR models where there are possibly some unit roots and some cointegrating relations. The model we will adopt is similar to that of Johansen (1988) in that we will allow the levels coefficient matrix (in a VAR in differences) to be of reduced rank, but our approach is different in that we do not employ reduced rank regression. Thus, our procedure will be an alternative to unrestricted levels VAR estimation.

The n-vector time series $\underline{y}_t$ is assumed to be generated by the following k'th order VAR model

$$(23) \qquad \underline{y}_t = \underline{J}(L)\underline{y}_{t-1} + \underline{\varepsilon}_t , \quad t = 1, 2, ..., T$$

where $\underline{J}(L) = \Sigma_{i=1}^k \underline{J}_i L^{i-1}$. The system (23) is initialized at $t = -k+1, ..., 0$ and since our asymptotics do not depend on the initial values $\{y_{-k+1}, \ldots, y_0\}$ we can let them be any random vectors including constants. However, it is sometimes convenient to set the initial conditions so that the I(0) component of (23) is stationary (rather than asymptotically stationary) and we will proceed as if this has been done. We define

$$\underline{J}^*(L) = \Sigma_{i=1}^{k-1} \underline{J}_i^* L^{i-1} \ , \ \text{ with } \ \underline{J}_i^* = -\Sigma_{h=i+1}^{k} \underline{J}_h$$

$$\underline{A} = J(1) \ ,$$

and then (23) can be written as

$$(24) \qquad \underline{y}_t = \underline{J}^*(L)\Delta\underline{y}_{t-1} + \underline{A}\,\underline{y}_{t-1} + \underline{\varepsilon}_t \ ,$$

or in the equivalent error correction model (ECM) format

$$(25) \qquad \Delta\underline{y}_t = J^*(L)\Delta\underline{y}_{t-1} + (\underline{A} - I)\underline{y}_{t-1} + \underline{\varepsilon}_t \ .$$

To fix ideas in what follows we need to be more specific about (23), its allowable roots, the dimension of the cointegration space and the form of the cointegrating coefficients. The following assumption is convenient for this purpose.

ASSUMPTION VAR (*Vector Autoregression*)

(a) $\underline{\varepsilon}_t$ *satisfies Assumption* EC(b), *i.e. is* iid *with zero mean, variance matrix* $\underline{\Sigma}_{\varepsilon\varepsilon} > 0$ *and finite fourth cumulants.*

(b) *The determinantal equation* $|I_n - \underline{J}(L)L| = 0$ *has roots on or outside the unit circle, i.e.* $|L| \geq 1$.

(c) $\underline{A} = I + \alpha\beta'$ *where* α *and* β *are* $n \times r$ *matrices of full column rank* r, $0 \leq r \leq n$. *(If* $r = 0$ *then* $A = I$; *if* $r = n$ *then* β *has rank* n *and* $\beta'\underline{y}_t$ *and hence* $\underline{y}_t$ *are (asymptotically) stationary).*

(d) $\alpha_\perp'(\underline{J}^*(1) - I_n)\beta_\perp$ *is nonsingular, where* $\alpha_\perp$ *and* $\beta_\perp$ *are* $n \times (n-r)$ *matrices of full column rank such that* $\alpha_\perp'\alpha = 0 = \beta_\perp'\beta$. *(If* $r = 0$ *then we take* $\alpha_\perp = I_n = \beta_\perp$*).* □

Under Assumption VAR, $\underline{y}_t$ has r cointegrating vectors (the columns of β) and $n-r$ unit roots. Condition VAR(d) ensures that the Granger representation theorem applies, so that $\Delta\underline{y}_t$

is stationary, $\beta'\underline{y}_t$ is stationary and $\underline{y}_t$ is an I(1) process when $r < n$. These conditions are now standard in the study of VAR's with some unit roots and are discussed more fully elsewhere, e.g. Johansen (1988, 1991) and Toda and Phillips (1991).

Our attention will focus on unrestricted estimation of the system (24), where the regressors have both stationary and nonstationary components but the dimension $(n-r)$ of the latter is unknown *a priori*. In studying this problem it is helpful to transform the system so that it conforms to our analysis in Section 3 of the paper. We can do so without loss of generality in the following way.

First let the columns of β be orthonormal. (This can be achieved with no loss of generality, and no issues of identification of individual cointegrating relations will arise in our work, so we need not be concerned with the problems raised in Park (1990) and Phillips and Park (1991).) Construct the orthogonal matrix $H = [\beta, \beta_{\perp}] = [H_1, H_2]$, say and define $y_t = H'\underline{y}_t$. The system (24) transforms to

$$(24') \qquad y_t = J^*(L)\Delta y_{t-1} + Ay_{t-1} + \varepsilon_t$$

where the transformed coefficients are

$$(26) \qquad A = H'\underline{A}H \ , \ \ J^*(L) = H'\underline{J}^*(L)H \ , \ \ \varepsilon_t = H'\underline{\varepsilon}_t \ , \ \ \Sigma_{\varepsilon\varepsilon} = H'\underline{\Sigma}_{\varepsilon\varepsilon}H \ .$$

We emphasize that H is unknown but that the asymptotic properties of regression estimators in (24) can be studied via the properties of the corresponding estimators in (24′) by simply reversing the transformations given in (26). For example, if $\hat{A}$ is the unrestricted OLS estimator of A in (24′) then $\hat{\underline{A}} = H\hat{A}H'$ where $\hat{\underline{A}}$ is the OLS estimator of $\underline{A}$ in (24), and so on.

We partition y_t according to the partition of H as

$$(27) \qquad y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} H_1'\underline{y}_t \\ H_2'\underline{y}_t \end{bmatrix} = \begin{bmatrix} I(0) \\ I(1) \end{bmatrix} \begin{matrix} r \\ n-r \end{matrix} \ .$$

Note that the matrix A in (24′) has the specific partitioned form

$$(28) \qquad A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} I_r + \beta'\alpha & 0 \\ \beta_\perp'\alpha & I_{n-r} \end{bmatrix}.$$

The $r \times n\text{-}r$ zero submatrix A_{12} in (28) delivers $r(n-r)$ restrictions on the matrix A. These restrictions on A correspond to the reduced rank (or cointegration) restrictions on the matrix $\underline{A}-I = \alpha\beta'$. Observe that there are $2nr$ parameters in the matrix product $\alpha\beta'$ but only $nr + r(n-r) = 2nr - r^2$ identified parameters. We can, of course, choose to write the cointegrating matrix β' as $\beta' = [I_r, \beta]$ leading to $r(n-r)$ identified parameters in the submatrix B. These parameters together with the nr "factor loading" parameters in the matrix α produce the $2nr - r^2$ identified parameters of the $\alpha\beta'$ matrix product. The $r(n-r)$ zero matrix A_{12} in (28) on the other hand is clearly identified as a submatrix of the coefficient matrix A in the system (24'). As such it can be regarded as the parameterization in (24') of the identified components of the cointegrating matrix β' in the original system (24) with $\underline{A} = I + \alpha\beta'$.

Notice, in addition, from (28) that the submatrix A_{22} has the special form $A_{22} = I_{n-r}$. Here the coefficient matrix A_{22} embodies the $n-r$ unit roots that occur in the original system (24) and relates these unit roots specifically to the subsystem of (24') that corresponds to the generating mechanism for the I(1) process y_{2t}.

Define $z_t = (\Delta y_{t-1}', \ldots, \Delta y_{t-k+1}')'$ and $J = [J_1^*, \ldots, J_{k-1}^*]$. Then (24') can be written more simply as

$$(29) \qquad y_t = Jz_t + Ay_{t-1} + \varepsilon_t$$

or, in partitioned form, as

$$(30a) \qquad y_{1t} = J_1 z_t + A_{11} y_{1t-1} + A_{12} y_{2t-1} + \varepsilon_{1t}$$

$$(30b) \qquad y_{2t} = J_2 z_t + A_{21} y_{1t-1} + A_{22} y_{2t-1} + \varepsilon_{2t}\ .$$

Using the explicit form of $A_{12} = 0$ and $A_{22} = I$ from (28), the true form of this system is

$$(31a) \qquad y_{1t} = J_1 z_t + A_{11} y_{1t-1} + \varepsilon_{1t}$$

(31b) $y_{2t} = y_{2t-1} + u_{2t}\,, \quad u_{2t} = \varepsilon_{2t} + J_2 z_t + A_{21} y_{1t-1}\,.$

In (31a) we can arrange initial conditions so that the variables y_{1t} and z_t are stationary. Hence, $A_{12} = 0$ in (30a) necessarily, otherwise the regression would be spurious. In (31b) y_{2t} is I(1), there are $n-r$ unit roots in the equation and the error u_{2t} is stationary.

We will need the long-run covariance matrix of $\eta_t = (\varepsilon_t', u_{2t}')'$ in the theory that follows and we accordingly introduce the matrix

$$\text{(32)} \qquad \mathrm{lrvar}(\eta_t) = \Omega_{\eta\eta} = \begin{bmatrix} \Omega_{\varepsilon\varepsilon} & \Omega_{\varepsilon 2} \\ \Omega_{2\varepsilon} & \Omega_{22} \end{bmatrix} = \begin{bmatrix} \Sigma_{\varepsilon_1\varepsilon_1} & \Sigma_{\varepsilon_1\varepsilon_2} & \Omega_{\varepsilon_1 2} \\ \Sigma_{\varepsilon_2\varepsilon_1} & \Sigma_{\varepsilon_2\varepsilon_2} & \Omega_{\varepsilon_2 2} \\ \Omega_{2\varepsilon_1} & \Omega_{2\varepsilon_2} & \Omega_{22} \end{bmatrix},$$

partitioning the final matrix above conformably with $(\varepsilon_t' u_{2t}') = (\varepsilon_{1t}', \varepsilon_{2t}', u_{2t}')$. With this notation in hand, we define the conditional long-run variance matrices

$$\text{(33)} \qquad \Omega_{\varepsilon\varepsilon\cdot 2} = \Sigma_{\varepsilon\varepsilon} - \Omega_{\varepsilon 2}\Omega_{22}^{-1}\Omega_{2\varepsilon}\,, \quad \Omega_{\varepsilon_1\varepsilon_1\cdot 2} = \Sigma_{\varepsilon_1\varepsilon_1} - \Omega_{\varepsilon_1 2}\Omega_{22}^{-1}\Omega_{2\varepsilon_1}\,.$$

Observe that in (32) and in the formulae just given we use the fact that ε_t is iid under Assumption VAR(a) and therefore $\Omega_{\varepsilon\varepsilon} = \Sigma_{\varepsilon\varepsilon}$.

We now estimate (29) by FM regression. Write (29) in matrix form as

$$\text{(29')} \qquad Y' = JZ' + AY_{-1}' + E' = FX' + E'$$

and let $Q_Z = I - Z(Z'Z)^{-1}Z'$ and $\Delta Y_{-1}' = Y_{-1}' - Y_{-2}'$. The FM regression estimator of F in (29') is

$$\hat{F}^+ = [\hat{J}^+ \,\vdots\, \hat{A}^+] = [Y'Z \,\vdots\, Y^{+\prime}Y_{-1} - T\hat{\Delta}_{\varepsilon\Delta y}^+](X'X)^{-1}$$

$$\text{(34)} \qquad = [Y'Z \,\vdots\, Y'Y_{-1} - T\hat{\Delta}_{\varepsilon\Delta y} - \hat{\Omega}_{\varepsilon y}\hat{\Omega}_{yy}^{-1}(\Delta Y_{-1}'Y_{-1} - T\hat{\Delta}_{\Delta y\Delta y})](X'X)^{-1}\,.$$

In these formulae $\hat{\Omega}_{\varepsilon y}$, $\hat{\Omega}_{yy}$ are kernel estimates of the long-run covariance matrices of ($\hat{\varepsilon}_t = y_t - \hat{F}x_t$, Δy_{t-1}) and Δy_{t-1}, respectively. Similarly, $\hat{\Delta}_{\varepsilon\Delta y}$ and $\hat{\Delta}_{\Delta y\Delta y}$ are kernel estimates of the one-sided long-run covariance matrices of ($\hat{\varepsilon}_t = y_t - \hat{F}x_t$, Δy_{t-1}) and Δy_{t-1}, respectively.

Note that in constructing $\hat{F}^+$ we use the endogeneity correction that involves the use of Y^+ only where it is needed, i.e. with respect to the levels regressors Y_{-1} in (29′). The regressors z_t in (29) are lagged differences Δy_{t-i} (i = 1, ..., k-1) which are known to be I(0) and therefore correction with respect to the estimation of their coefficient matrix J is known to be unnecessary.

In addition, under Assumption VAR(a) the error ε_t in (29) is a martingale difference and it is therefore not necessary to make a serial correlation correction with respect to the term $E'Y_{-1}$. More specifically, under VAR(a) we know that $\Delta_{\varepsilon\Delta y} = \Sigma_{j=0}^{\infty}E(\varepsilon_j\Delta y_{-1}) = 0$ and, hence, we can exclude the term $T\hat{\Delta}_{\varepsilon\Delta y}$ in (34) with no affect on the asymptotics. (We did not mention the fact earlier, but this could also be done in FM estimation of the single equation AR(1) as studied in Section 4.) Although the limit distribution is unaffected by the inclusion or exclusion of $T\hat{\Delta}_{\varepsilon\Delta y}$, there may be some advantage arising from reduced variance in small samples from excluding the term. This gives us the following adjusted formula for $\hat{F}^+$

$$(30') \qquad \hat{F}^+ = [Y'Z \;\vdots\; Y'Y_{-1} - \hat{\Omega}_{\varepsilon y}\hat{\Omega}_{yy}^{-1}(\Delta Y_{-1}'Y_{-1} - T\hat{\Delta}_{\Delta y\Delta y})](X'X)^{-1} .$$

A further partitioning of (29′) is useful in the development of our asymptotic theory. This is because some elements of Y_{-1}' are stationary (corresponding to y_{1t-1}) and some are nonstationary (the elements of y_{2t-1}). We therefore write (29′) as

$$(29'') \qquad Y' = F_1X_1' + F_2X_2' + E' ,$$

where $x_{1t}' = (z_t', y_{1t-1}')$ is the composite vector of stationary regressors and $x_{2t} = y_{2t-1}$ is the vector of full rank nonstationary regressors. In this form, (29″) corresponds with the earlier model (3′) of Section 3 and we can therefore avail ourselves of the earlier theory that relates to this model more readily.

The limit distribution of $\hat{F}^+$ is given as follows.

5.1. THEOREM (*FM-VAR Limit Theory*): *Under Assumptions* KL, BW *and* VAR

(a) $\sqrt{T}(\hat{F}_1^+ - F_1) \to_d N(0, \Sigma_{\varepsilon\varepsilon} \otimes \Sigma_{11}^{-1})$ *where* $\Sigma_{11} = E(x_{1t}x_{1t}')$; *and*

(b) $T(\hat{F}_2^+ - F_2) \to_d (\int_0^1 dB_{\varepsilon\cdot 2}B_2')\left(\int_0^1 B_2B_2'\right)^{-1}$ *where* $B_{\varepsilon\cdot 2} = B_\varepsilon - \Omega_{\varepsilon 2}\Omega_{22}^{-1}B_2 \equiv BM(\Omega_{\varepsilon\varepsilon\cdot 2})$ *and*

$\Omega_{\varepsilon\varepsilon\cdot 2} = \Omega_{\varepsilon\varepsilon} - \Omega_{\varepsilon 2}\Omega_{22}^{-1}\Omega_{2\varepsilon} = \Sigma_{\varepsilon\varepsilon} - \Omega_{\varepsilon 2}\Omega_{22}^{-1}\Omega_{2\varepsilon}$. *The bandwidth expansion rates under which* (a) *and* (b) *hold are the same as those given in Theorem* 3.5. *In particular, both* (a) *and* (b) *hold when the bandwidth* $K = O(T^k)$ *and* $1/4 < k < 2/3$, *i.e.* BW(i).

The limit distributions given in parts (a) *and* (b) *above are statistically independent.* □

5.2. COROLLARY (*Stationary* VAR *case*): *When* $r = n$ *and under Assumptions* VAR, KL, *and* BW *with bandwidth expansion rate* $K = O(T^k)$ *for* $1/4 < k < 1$ *we have*

$$\sqrt{T}(\hat{F}^{+} - F) \to_d N(0,\ \Sigma_{\varepsilon\varepsilon} \otimes \Sigma_{xx}^{-1})$$

where $\Sigma_{xx} = E(x_t x_t')$.

5.3. COROLLARY (VAR *with n unit roots*): *When* $r = 0$ *and under Assumptions* VAR, KL *and* BW *with bandwidth expansion rate* $K = O(T^k)$ *for* $0 < k < 1$ *we have*

$$\sqrt{T}(\hat{F}^{+} - F) = \sqrt{T}(\hat{J}^{+} - J) \to_d N(0,\ \Sigma_{\varepsilon\varepsilon} \otimes \Sigma_{11}^{-1})$$

and

$$T(\hat{F}_2^{+} - I_n) = T(\hat{A}^{+} - I) \to_p 0$$

i.e. $\hat{F}_2^{+}$ *is hyperconsistent for* I_n.

5.4. REMARKS

(a) Theorem 5.1 shows that the limit theory for the FM regression estimator $\hat{F}^{+}$ is normal and mixed normal. Note that in the case of part (b) of Theorem 5.1 we have

$$\left(\int_0^1 dB_{\varepsilon\cdot 2}B_2'\right)\left(\int_0^1 B_2B_2'\right)^{-1/2} \equiv N(0,\ \Omega_{\varepsilon\varepsilon\cdot 2} \otimes I)$$

and then

$$T(\hat{F}_2^{+} - F_2) \to_d \int_{G=\int_0^1 B_2B_2'>0} N(0,\ \Omega_{\varepsilon\varepsilon\cdot 2} \otimes G^{-1})dP(G)\ .$$

Of special significance is the fact that a submatrix of F_2 involves the $n-r$ unit roots of the system. Thus, from (28) we have

$$F_2 = \begin{bmatrix} 0 \\ I_{n-r} \end{bmatrix} = \begin{bmatrix} F_{21} \\ F_{22} \end{bmatrix}, \quad \text{say.}$$

In consequence, part (b) of Theorem 5.1 can be decomposed into the following two parts

$$(35) \qquad T\hat{F}^{+}_{21} \rightarrow_d \left(\int_0^1 dB_{\varepsilon_1 \cdot 2} B_2'\right)\left(\int_0^1 B_2 B_2'\right)^{-1},$$

and

$$(36) \qquad T(\hat{F}^{+}_{22} - I) \rightarrow_d \left(\int_0^1 dB_{\varepsilon_2 \cdot 2} B_2'\right)\left(\int_0^1 B_2 B_2'\right)^{-1}.$$

The latter result (36) shows the rather remarkable outcome that in FM-VAR estimation when there are some unit roots in the system, there are no limiting distributions of the unit root (or matrix unit root) type! All the limit theory is normal or mixed normal irrespective of the number of unit roots or dimension of the cointegrating space (provided $r > 0$).

(b) When $r = 0$, there are no cointegrating vectors in the system and the nonstationary part of the system is a full set of unit roots of dimension n. In this case Corollary 5.3 applies and we have hyperconsistent estimation of all of the unit roots in the system by FM regression. This gives a matrix generalization of the earlier result by the author (1992) on hyperconsistent estimation of a unit root in a single equation model with one unit root. Interestingly, the presence of the stationary component z_t in the model (29) does not interfere with this hyperconsistency. As shown in the author's (1992) paper in the single equation case, the precise rate of hyperconsistency depends on the bandwidth expansion rate. The arguments given in that paper can be extended to the present case, but we will not do so here (to help keep the present paper at a manageable length).

(c) The mixed normal limit given in (35) for the submatrix F_{21} relates to the cointegrating space restrictions. As explained in the discussion following (28) the submatrix F_{21} (which is the same as the submatrix A_{12} in (28)) has true value zero and in the transformed system (see equations (29) and (31a)) this can be regarded as a parameterization of the identified components of the cointegrating matrix β'. In other words, when β' is a cointegrating matrix $y_{1t} = \beta' y_t$ is stationary and equation (31a) for y_{1t} involves only stationary variables, because $F_{21} = 0$ (equivalently, $A_{12} = 0$ in (28)) eliminates the nonstationary variables $y_{2t} = \beta_\perp' y_t$ from this equation. Loosely speaking, therefore, we can regard the limit distribution of $T\hat{F}_{21}^{+}$ given in (35) as relating to the errors of estimation of the identifiable components of the cointegrating matrix. The following simple example taken from Section 2 will help to illustrate. Suppose $\beta' = (I_r + BB')^{-1/2}[I_r, -B]$ for some $r \times n-r$ matrix B and the original system (23) is

$$\underline{y}_t = \begin{bmatrix} 0 & B \\ 0 & I_{n-r} \end{bmatrix} \underline{y}_{t-1} + \underline{\varepsilon}_t \,, \quad \underline{A} = \begin{bmatrix} 0 & B \\ 0 & I_{n-r} \end{bmatrix}.$$

The first subsystem of this equation is the cointegrating relation

$$\underline{y}_{1t} = B\underline{y}_{2t-1} + \underline{\varepsilon}_{1t} \,, \tag{37}$$

and the second is the I(1) relation

$$\underline{y}_{2t} = \underline{y}_{2t-1} + \underline{\varepsilon}_{2t} \,. \tag{38}$$

We now transform this system using the orthogonal matrix

$$H = [\beta \; \vdots \; \beta_\perp] = \left[\begin{bmatrix} I \\ -B' \end{bmatrix} (I + BB')^{-1/2} \;\; \vdots \;\; \begin{bmatrix} B \\ I \end{bmatrix} (I + B'B)^{-1/2} \right].$$

We obtain, following (24) and (26), the new system

$$y_t = H'\underline{A}Hy_{t-1} + \varepsilon_t = Ay_{t-1} + \varepsilon_t \,, \tag{39}$$

with

$$A = \begin{bmatrix} 0 & 0 \\ (I + B'B)^{1/2}B'(I + BB')^{-1/2} & I_{n-r} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ A_{21} & I \end{bmatrix}.$$

Explicitly,

(37′) $\quad y_{1t} = \qquad\qquad \varepsilon_{1t}$

(38′) $\quad y_{2t} = A_{21}y_{1t-1} + y_{2t-1} + \varepsilon_{2t}$.

The cointegrating relation (37) is replaced in the transformed system by the stationary relation (37′). What was, in (37), the matrix of identified cointegrating coefficients (*viz.* B) is replaced in (37′) by the zero coefficient matrix for the nonstationary variable y_{2t-1}. The I(1) relation (38) is replaced in (38′) with a system of full $(n-r)$ unit roots and some additional stationary inputs (*viz.* $A_{12}y_{1t-1}$).

(d) The explicit form (38′) helps to explain why the FM-VAR estimates of the unit root coefficient matrix $F_{22} = I_{n-r}$ have a mixed normal limit distribution rather than the conventional matrix unit root distribution. The latter would arise if we ran the regression of y_{2t} on y_{2t-1} giving the estimate $F_{22}^{*} = (Y_2'Y_{2,-1})(Y_{2,-1}'Y_{2,-1})^{-1}$, which has the limit theory

(40) $$T(F_{22}^{*} - I_{n-r}) \to_d \left(\int_0^1 dB_2B_2'\right)\left(\int_0^1 B_2B_2'\right)^{-1} .$$

What happens in the case of the FM-VAR estimator $\hat{F}_{22}^{+}$ is that the coefficient of y_{2t-1} in (38′) is treated as a cointegrating coefficient matrix and because of the endogeneity correction in the FM procedure the FM estimation errors depend on the "endogeneity corrected" errors from this equation, *viz.* $\varepsilon_{2t}^{+} = \varepsilon_{2t} - \Omega_{\varepsilon_2u_2}\Omega_{u_2u_2}^{-1}u_{2t}$ where $u_{2t} = \varepsilon_{2t} + A_{21}y_{1t-1}$. Thus, *because* of the presence of the stationary component $A_{21}y_{1t-1}$ in (38′) $\mathrm{var}(\varepsilon_{2t}^{+}) > 0$, and ε_{2t}^{+} has long-run zero covariance with u_{2t}. Consequently, the limit Brownian motion $B_{\varepsilon_2^{+}}(r) = B_{\varepsilon_2\cdot u_2}(r)$ that arises from partial sums of ε_{2t}^{+} is independent of the Brownian motion $B_{u_2}(r)$ that arises from y_{2t-1}, i.e. from partial sums of u_{2t}. In contrast to (40), the limit distribution of the FM estimator is $\left(\int_0^1 dB_{\varepsilon_2\cdot 2}B_2'\right)\left(\int_0^1 B_2B_2'\right)^{-1}$ and the independence of $B_{\varepsilon_2\cdot 2}$ and B_2 ensures that this limit distribution is mixed normal.

(e) The explanation of the mixed normal limit distribution for the unit roots estimator $\hat{F}_{22}^{+}$ just given in Remark (d) also applies to subsystem estimation of unit roots. Thus, suppose we treat (38′) as a subsystem of (39) but estimate (38′) independently. The limit theory for the FM estimator of the unit roots matrix $F_{22} = I$ is the same and is mixed normal, again because of the presence of the stationary component y_{1t-1} in this regression. When this additional stationary component is not present in the regression the FM estimator is hyperconsistent because in this case $\varepsilon_{2t} = u_{2t}$, and so $\varepsilon^+_{2t} = 0$ a.s., which leads to the fact that $T(\hat{F}^+_{22} - I) \to_p 0$, just as in Corollary 5.3.

(f) The limit theory given in Theorem 5.1 can be compared with that of the OLS estimator $\hat{F} = [\hat{F}_1 \vdots \hat{F}_2] = Y'X(X'X)^{-1}$. We have

5.5. THEOREM (*Levels* VAR *Limit Theory*): *Under Assumption* VAR *the limit theory for the* OLS *regression estimator* $\hat{F} = [F_1 \vdots F_2]$ *is*

(a) $\sqrt{T}(\hat{F}_1 - F_1) \to_d N(0, \Sigma_{\varepsilon\varepsilon} \otimes \Sigma_{11}^{-1})$,

(b) $T(\hat{F}_2 - F_2) \to_d (\int_0^1 dB_\varepsilon B_2')\left(\int_0^1 B_2 B_2'\right)^{-1}$

$$= (\int_0^1 dB_{\varepsilon\cdot 2} B_2')\left(\int_0^1 B_2 B_2'\right)^{-1} + \Omega_{\varepsilon 2}\Omega_{22}^{-1}\int_0^1 dB_2 B_2'\left(\int_0^1 B_2 B_2'\right)^{-1} . \tag{41}$$

5.6. REMARK

Note that the limit theory for the stationary component F_1 in Theorem 5.5(a) is identical to that of the FM estimator. The limit theory of $\hat{F}_2$ given in (b) has two components. The first is identical to the limit theory for the FM-VAR estimator $\hat{F}^+_2$. The second is a matrix unit root distribution whose overall importance depends on the magnitude of the coefficient matrix $\Omega_{\varepsilon 2}\Omega_{22}^{-1}$. Note that from (31b) we have the representation $u_{2t} = \varepsilon_{2t} + J_2 z_t + A_{21} y_{1t-1}$, so that u_{2t} involves ε_{2t} as one of its components. Consequently, $\Omega_{\varepsilon 2}$ will be nonzero. Indeed, when there are no additional stationary elements in equation (31b) (i.e. when $J_2 = 0, A_{21} = 0$) we have $u_{2t} = \varepsilon_{2t}$. In this case, $\Omega_{\varepsilon_2 2}\Omega_{22}^{-1} = \Omega_{\varepsilon_2\varepsilon_2}\Omega_{\varepsilon_2\varepsilon_2}^{-1} = I_{n-r}$ and only the second component of (41) is retained in the sub block corresponding to F_{22} because $B_{\varepsilon_2\cdot 2} = 0$ a.s. When this occurs, the limit distribution of the levels VAR estimator $\hat{F}_{22}$ is the matrix unit root distribution, i.e.

$$T(\hat{F}_{22} - I) \to_d (\int_0^1 dB_2 B_2')(\int_0^1 B_2 B_2') .$$

This is precisely the case when the FM-VAR estimator $\hat{F}^+_{22}$ is hyperconsistent for $F_{22} = I$ and therefore when $\hat{F}^+$ dominates $\hat{F}_{22}$ by virtue of its faster rate of convergence. □

Finally in this section we will consider the limit theory for the FM estimator in the original coordinate system. Recall that in the original VAR coordinates (see equations (23) and (24)) we have $\underline{y}_t = Hy_t$. Using the matrix H to transform (24′), and hence (29), back to the original coordinates we obtain

$$\underline{y}_t = HJ(I_{k-1} \otimes H')\underline{z}_t + HAH'\underline{y}_{t-1} + \underline{\varepsilon}_t = \underline{J}\underline{z}_t + \underline{A}\underline{y}_{t-1} + \underline{\varepsilon}_t \tag{29'}$$

$$= \underline{F}\underline{x}_t + \underline{\varepsilon}_t\,, \quad \text{with } \underline{F} = [\underline{J}, \underline{A}] = HF(I_k \otimes H')\,. \tag{42}$$

The FM-VAR estimator of $\underline{F}$ is

$$\underline{\hat{F}}^+ = H\hat{F}^+(I_k \otimes H') = H[\hat{J}^+, \hat{A}^+](I_k \otimes H')\,. \tag{43}$$

Using this representation and the limit theory for $\hat{F}^+$ given in Theorem 5.1, we obtain the (potentially degenerate) asymptotics for the matrix $\underline{\hat{F}}^+$, *viz.*

5.7. THEOREM (FM-VAR *Limit Theory in original coordinates*): *Under the conditions of Theorem* 5.1

(a) $\sqrt{T}(\underline{\hat{F}}^+ - \underline{F}) \to_d N(0, \underline{\Sigma}_{\varepsilon\varepsilon} \otimes G\Sigma_{11}^{-1}G')$ *where* $\underline{\Sigma}_{\varepsilon\varepsilon} = \Sigma_{\underline{\varepsilon}\underline{\varepsilon}}$ *and*

$$G = \begin{bmatrix} I_{k-1} \otimes H & 0 \\ 0 & \beta \end{bmatrix} nk \times n(k-1) + r\,.$$

Alternatively,

(a′) $\sqrt{T}(\underline{\hat{F}}^+ - \underline{F})G \to_d N(0, \underline{\Sigma}_{\varepsilon\varepsilon} \otimes \Sigma_{11}^{-1})$; *and*

(b) $T(\underline{\hat{F}}^+ - \underline{F})G_\perp \to_d \left(\int_0^1 dB_{\varepsilon\cdot 2}B_2'\right)\left(\int_0^1 B_2B_2'\right)^{-1}$, *where* $G_\perp' = [0 \;\vdots\; \beta_\perp'](n-r) \times nk$. □

5.8. REMARK

The limit theory for the OLS levels VAR estimator $\underline{\hat{F}}$ is obtained in the same way as (a′) and (b′) of Theorem 5.6 using the results of Theorem 5.5. For this estimator we have:

$$\sqrt{T}(\underline{\hat{F}} - \underline{F})G \to_d N(0, \underline{\Sigma}_{\varepsilon\varepsilon} \otimes \Sigma_{11}^{-1})\,, \tag{44}$$

and

(45) $$T(\hat{\underline{F}}-\underline{F})G_{\perp} \to_d \left(\int_0^1 dB_{\underline{\varepsilon}}B_2'\right)\left(\int_0^1 B_2B_2'\right)^{-1} .$$

So, in stationary directions, $\hat{\underline{F}}$ is asymptotically equivalent to the FM estimator $\hat{\underline{F}}^+$. But the estimators differ in nonstationary directions, where the rate of convergence is $O(T)$. The limit theory for the FM-VAR estimator in nonstationary directions is mixed normal. This involves: (i) the identified components of the cointegrating matrix, where the limit theory of the FM estimator corresponds to that of the optimal estimator (see Phillips, 1991); and (ii) the matrix of unit roots in the system, where the limit theory of the FM estimator is again mixed normal and, when the system has a full set of unit roots, is actually hyperconsistent. The levels VAR estimator $\hat{F}$ is $O(T)$ consistent in nonstationary directions, but involves: (i) second-order bias (i.e. simultaneous equations bias) effects in the estimation of the identified components of the cointegrating matrix; and (ii) a composite of a matrix unit root distribution and a mixed normal in the estimation of the system's unit roots. The bias effects and matrix unit root distribution arise because of the dependence of the two Brownian motions $B_{\underline{\varepsilon}}$ and B_2 that appear in (45) and were discussed earlier in Remark 5.6. Asymptotic theory therefore clearly favors the FM-VAR estimator because of its better properties in nonstationary directions.

6. HYPOTHESIS TESTING IN FM-VAR REGRESSION

For testing purposes we use the VAR model (24) in original coordinates and write this for convenience in condensed format as we have done earlier in (42), to repeat here:

(42) $$\underline{y}_t = \underline{F}\underline{x}_t + \underline{\varepsilon}_t \ , \quad \underline{F} = [\underline{J}, \underline{A}] \ , \quad \underline{\varepsilon}_t = \text{iid}(0, \underline{\Sigma}_{\varepsilon\varepsilon}) \ .$$

Suppose we wish to test restrictions such as

(46) $$\mathcal{H}_0 : R\ \text{vec}(\underline{F}) = r \ , \quad R(q \times n^2k) \ \text{ of rank } \ q \ .$$

When R has the Kronecker structure $R = R_1 \otimes R_2'$ then $\mathcal{H}_0$ has the simpler form

(47) $$\mathcal{H}_0' : R_1\underline{F}R_2 = \underline{R}$$

for some suitable matrix $\underline{R}$. This set up corresponds to the framework used for our analysis of hypothesis testing in Section 3 -- see the earlier Remarks 3.8(h) and (i). A special case of $\mathscr{H}_0'$ that arises in VAR modelling that is of particular importance in practice is the case of causality restrictions. In the notation of equation (24) the hypothesis that the subvector $\underline{y}_{3t}$ $(n_3 \times 1)$ has no Granger-causal effect on the subvector $\underline{y}_{1t}$ $(n_1 \times 1)$ would be formulated as

(48) $$\mathscr{H}_0' : J_{i13}^* = 0 \quad (i = 1, ..., k-1) \ , \ \underline{A}_{13} = 0 \ .$$

In (47) this would correspond to the following settings of the restriction matrices

(49) $$R_1 = [I_{n_1} \ \vdots \ 0] \ , \ R_2 = I_k \otimes \begin{bmatrix} 0 \\ I_{n_3} \end{bmatrix} , \ \underline{R} = 0 \ .$$

For unrestricted levels VAR estimation of (42) Wald tests of the causality restrictions (48) have been used extensively in past empirical research. An asymptotic theory for such tests that accommodates nonstationary data has recently been developed for trivariate systems in Sims, Stock and Watson (1991) and in full generality by Toda and Phillips (1991). These authors show that when the VAR system has some unit roots and some cointegrating relations the asymptotic theory of Wald tests of (48) involves nuisance parameters and nonstandard distributions that make a valid asymptotic basis for inference very awkward. Toda and Phillips (1991, Theorem 1) show that the form of the limit distribution depends on the rank of a certain submatrix of the cointegrating matrix. But the cointegrating matrix is estimated only indirectly in levels VAR estimation, and since, as we have discussed earlier in Remark 5.8, the limit theory for these VAR estimates of the cointegrating matrix involve nonstandard distributions and nuisance parameters, it is not possible to provide an asymptotic theory that justifies the general use of VAR regressions for causality testing at least in correctly specified models.

On the other hand, we can artificially augment the correct order of the VAR so that normal asymptotics obtain with respect to the coefficient matrices up to the correct lag order (much as $\hat{F}_1$ is asymptotically normal in Theorem 5.5(a)) and then asymptotic chi-squared tests of causality restrictions can be applied to the submatrix of the coefficients up to the correct order. This idea was explored in some recent work by Toda and Yamamoto (1993) and relates to a similar suggestion made by Choi (1993) for avoiding nonstandard distributions in scalar unit root tests. The method is interesting but does involve the inefficiency, which may be costly in terms of the method's power properties, of having to estimate coefficient matrices for surplus lags.

The alternative approach we explore is to use Wald tests based on the FM-VAR regression estimator. From Theorem 5.7(a) we have $\sqrt{T}(\hat{\underline{F}}^{+}-\underline{F}) \to_d N(0, \underline{\Sigma}_{\varepsilon\varepsilon} \otimes G\Sigma_{11}^{-1}G')$, and since $T(\underline{X}'\underline{X})^{-1} \to_p G\Sigma_{11}^{-1}G'$ we consider using the asymptotic approximation

$$(50) \qquad \sqrt{T}(\hat{\underline{F}}^{+}-\underline{F}) \sim N(0, \hat{\underline{\Sigma}}_{\varepsilon\varepsilon} \otimes T(\underline{X}'\underline{X})^{-1})$$

just as in (14'). To test $\mathscr{H}_0'$ the natural Wald statistic is then

$$(51) \qquad W_F^{+} = T(R \text{ vec } \hat{F}^{+} - r)'\left[R\{\hat{\underline{\Sigma}}_{\varepsilon\varepsilon} \otimes T(\underline{X}'\underline{X})^{-1}\}R\right]^{-1}(R \text{ vec } \hat{F}^{+} - F) \ .$$

When

$$(\text{RK}_G) \qquad \text{rank}[R\{\underline{\Sigma}_{\varepsilon\varepsilon} \otimes G\Sigma_{11}^{-1}G'\}R'] = q$$

we have $W_F^{+} \to_d \chi_q^2$, and standard limit theory applies.

When (RK_G) fails the situation is different. We follow the analysis in Section 3 of this case, now in a VAR setting. (RK_G) fails when the restrictions in $\mathscr{H}_0$ relate to some of the nonstationary coefficients. We therefore focus again on the case of $\mathscr{H}_0'$ where $R = R_1 \otimes R_2'$ and $R_2 = \text{diag}[R_{2J}, R_{2A}]$ is $nk_1 \times (q_J + q_A)$, so that the restrictions can be written out explicitly as $R_1\underline{F}R_2 = R_1[\underline{J} \ \vdots \ \underline{A}]R_2 = \underline{R}$, or

$$\mathscr{H}_0'' : R_1\underline{J}R_{2J} = \underline{R}_J \quad \text{and} \quad R_1\underline{A}R_{2A} = \underline{R}_A \ .$$

Next suppose that R_{2A} has the form

$$(52) \qquad R_{2A} = [\underset{q_{21}}{R_{21}} \ \vdots \ \underset{q_{22}}{R_{22}}] = [H_1, H_2]\begin{bmatrix} S_{20} & S_{h1} & \vdots & 0 \\ 0 & S_{h2} & \vdots & S_{22} \end{bmatrix} = [H_1S_{20}, H_1S_{h1} + H_2S_{h2} \ \vdots \ H_2S_{22}] \ ,$$

with $q_A = q_{21} + q_{22}$, and for some matrices S_{20}, S_{h1}, S_{h2} and S_{22}, just as in (18). The hypotheses about $\underline{A}$ that correspond to the columns R_{22} of R_{2A} in $\mathscr{H}_0''$ relate to nonstationary coefficients. Observe that

$$(53)\qquad R_2'G = \begin{bmatrix} R_{2J}'(I_{k-1}\otimes H) & 0 \\ 0 & R_{21}'\beta \\ 0 & R_{22}'\beta \end{bmatrix} = \begin{bmatrix} R_{2J}'(I_{k-1}\otimes H) & 0 \\ 0 & R_{21}'\beta \\ 0 & 0 \end{bmatrix},$$

which is of deficient row rank and therefore condition (RK_G) fails. The situation is entirely analogous to the one studied in Remark (i) of Section 3. We have the following analogue of Theorem 3.9 for the VAR case.

6.1. THEOREM (FM-VAR *Wald test asymptotics*): *Under Assumptions* KL, BW *and* VAR *with* $0 < r < n$, *the Wald statistic* W_F^+ *for testing the hypothesis* $\mathcal{H}_0'' : R_1 \underline{F} R_2 = \underline{R}$ *has a limit distribution that is a mixture of* χ^2 *variates. In particular, when* R_2 *has the form* $R_2 = \mathrm{diag}(R_{2J}, R_{2A})$ *whose dimension is* $nk_1 \times (q_J + q_A)$ *and where* R_{2A} *is given by* (52), *we have the limit*

$$(54)\qquad W_F^+ \longrightarrow_d \chi^2_{q_1 q_J} + \Sigma_{i=1}^{q_1}\chi^2_{q_{21}}(i) + \Sigma_{j=1}^{q_1} d_j \chi^2_{q_{22}}(j) = \chi^2_{q_1(q_J+q_{21})} + \Sigma_{j=1}^{q_1} d_j \chi^2_{q_{22}}(j)$$

where $\chi^2_{q_{22}}(j) \equiv \mathrm{iid}(\chi^2_{q_{22}})$ $(j = 1, \ldots, q_1)$ *and are independent of the* $\chi^2_{q_1(q_J+q_{21})}$ *member of the last equation of* (54). *The coefficients* d_j $(j = 1, \ldots, q_1)$ *that appear in* (54) *are the latent roots of the matrix* $(R_1 \underline{\Omega}_{\varepsilon\varepsilon\cdot 2} R_1)(R_1 \underline{\Sigma}_{\varepsilon\varepsilon} R_1')^{-1}$. □

6.2. REMARKS

(a) Since $\underline{\Omega}_{\varepsilon\varepsilon\cdot 2} = \underline{\Sigma}_{\varepsilon\varepsilon} - \Omega_{\varepsilon 2}\Omega_{22}^{-1}\Omega_{2\varepsilon} < \underline{\Sigma}_{\varepsilon\varepsilon}$, the latent roots d_j $(j = 1, \ldots, q_1)$ in (54) satisfy $0 < d_j \le 1$, just as in Theorem 3.9. Hence the earlier Remarks 3.10(a), (b) and (c) are also relevant here. In particular, tests that are conservative asymptotically can always be constructed using a $\chi^2_{q_1(q_J+q_A)}$ limit distribution as this is an upper bound for (54). Similarly, asymptotically liberal tests can be constructed using the $\chi^2_{q_1(q_J+q_A)}$ limit distribution when the Wald test statistic W_F^+ using the error variance matrix estimate $\underline{\hat{\Omega}}_{\varepsilon\varepsilon\cdot 2} = \underline{\hat{\Sigma}}_{\varepsilon\varepsilon} - \hat{\Omega}_{\underline{\varepsilon}2}\hat{\Omega}_{22}^{-1}\hat{\Omega}_{2\underline{\varepsilon}}$ in place of $\underline{\hat{\Sigma}}_{\varepsilon\varepsilon}$ in formula (51).

(b) It will be of interest to explore how close the actual size of the tests suggested in Remark (a) are in relation to the nominal size of the bounding variate $\chi^2_{q_1(q_J+q_A)}$ in finite sample simulations. This approach could also be usefully compared with the sequential testing procedure suggested in Toda and Phillips (1993) and the lag augmented regression procedure of Toda and Yamamoto (1993) that was mentioned earlier. An extensive Monte Carlo study designed to assess the relative merits of these three approaches would be useful. It is outside the scope of the present paper.

(c) The case $r = 0$ is rather special. In this case there are no cointegrating vectors and the limit theory of Corollary 5.3 for the FM-VAR estimator $\hat{F}^{+}$ applies. Obviously, in this case a VAR in differences could be run. But since the fact that there is a full set of n unit roots in the system is unknown in general we do need to consider the effects on $\hat{F}^{+}$ and related tests. From Corollary 5.3 we know that $\hat{\underline{F}}_2^{+} = \underline{A}^{+}$ is hyperconsistent for the unit root matrix I_n. In this case, also, tests based on the statistic W_F^{+} and a χ_q^2 $(q = q_1(q_J + q_A))$ limit theory will be conservative, as the following theorem shows.

6.3. THEOREM: *Under the Assumptions of Theorem* 6.1 *but with* $r = 0$ (*so that there is a full set* $\underline{A} = I_n$ *of n unit roots in the system* (42)) *the limit distribution of the Wald statistic* W_F^{+} *for testing* $\mathcal{H}_0'' : R_1 \underline{F} R_2 = \underline{R}$ *is given by*

(55) $$W_F^{+} \longrightarrow_d \chi^2_{q_1 q_J},$$

where q_J = $\text{rank}(R_{2J})$ *and* R_{2J} *is the leading submatrix of the restriction matrix* R_2 $= \text{diag}(R_{2J}, R_{2A})$. □

6.4. REMARK

When $q_J = 0$ we have $W_F^{+} \longrightarrow_p 0$ in place of (55). This follows directly from the hyperconsistency of $\hat{A}^{+}$. In this case we would accept the null hypothesis with probability tending to unity as $T \to \infty$ (i.e. the actual size of a test based on a χ_q^2 $(q = q_1 q_A)$ limit would tend to zero as $T \to \infty$). Use of a more efficient estimator, like $\hat{A}^{+}$ in this case, therefore does not always lead to a better test. The estimator also needs to be efficient under the alternative and the correct size of the test must be employed. When the number of unit roots in the system is unknown, as we assume in this paper, the size of a test based on W_F^{+} will inevitably be conservative in large samples, as we have seen. How this conservatism affects the power of the test in finite samples can be investigated by simulation.

7. CONCLUSION

This paper has developed a general asymptotic theory for time series regression using the principle of fully modified least squares. While the method was originally developed for estimating cointegrated systems, where it delivers optimal estimators of the identified components of a cointegrating matrix under Gaussian assumptions, we have shown that FM-OLS has some attractive features as a general method of estimation in a wider class of time series models. Our main results are as follows.

(i) FM-OLS is applicable in models with either full rank or cointegrated I(1) regressors. In such cases, the limit theory for FM estimates of the stationary components of the regressors is equivalent to that of OLS, while the FM estimates of the nonstationary components retain their optimality properties (i.e. they are asymptotically equivalent to the maximum likelihood estimates of the cointegrating matrix). When the OLS estimates of the stationary components are optimal then this property is shared by the FM-OLS estimator.

(ii) FM-OLS is applicable even in models with stationary regressors and in this case has the same limit theory as OLS. A case of special importance in practice is the stationary autoregression. For this model FM-OLS and OLS have the same asymptotic distribution. In finite samples, simulations indicate that the FM-OLS estimator of the coefficient in a stationary AR(1) is less biased than that of OLS, especially for larger values of the autoregressive coefficient.

(iii) FM-OLS is hyperconsistent with a convergence rate that exceeds $O(T)$ for a unit root in autoregression. This result, which was shown by the author in an earlier paper (1992) is shown here to apply to higher order autoregressions with a unit root and to autoregressive models with a unit root and deterministic trends. A surprising result in the case of the latter model is that FM-OLS is hyperconsistent for the coefficients of the deterministic regressors.

(iv) FM-VAR (fully modified vector autoregression) estimation also has good properties. For the case of a VAR with a full set of unit roots the FM-VAR estimator is hyperconsistent for all elements of the unit root matrix I_n. This includes diagonal and off diagonal elements. When there are stationary components in the VAR, the corresponding FM estimates of these coefficients have the same asymptotic distribution as the (levels VAR) OLS estimates.

(v) In VAR models with some unit roots and cointegrated variables, the FM-VAR estimator has some remarkable asymptotic properties. First, estimates of the identified components of the cointegrating matrix have a mixed normal limit theory, just like that of the optimal estimator in Phillips (1991a) or the reduced rank regression estimator in Johansen (1988). Second, and this is the most surprising, the FM-VAR estimates of the unit roots in the system also have mixed normal limits. This means that the limit theory for the FM-VAR estimator is normal for the stationary components of the VAR and mixed normal for the nonstationary components. In other words, there are *no* unit root limit distributions or matrix unit root limit distributions in FM-VAR estimation. Correspondingly, the FM-VAR estimates of the stationary and nonstationary components of a VAR are all asymptotically median unbiased. This gives the FM-VAR procedure a distinct advantage over OLS levels VAR estimation, where the estimates of the cointegrating vectors suffer in the limit from a second order simultaneous equations bias and estimates of the unit roots in the system have a limit theory that involves unit root distributions.

(vi) The normal and mixed normal limit theory for FM time series regression estimates helps to simplify inference. Wald statistics are shown to have a limit distribution that is a linear combination of independent chi-squared variates when the hypothesis under test involves both stationary and nonstationary coefficients. If q is the total number of restrictions then the χ_q^2 distribution is shown to be an upper bounding variate and therefore the usual χ^2 critical values can be used to construct tests that have conservative size. This avoids problems of pre-tests, nuisance parameters, overfitting and nonstandard limit distributions that arise in other approaches. The theory is applicable to VAR models and causality testing in VAR's.

8. PROOFS

To simplify the presentation of our arguments it will be convenient to assume in our proofs that we are working with long-run covariance matrix estimates that satisfy Assumption KL(a) and (b). This leads to estimates of the form

$$\hat{\Omega} = \Sigma_{j=-K+1}^{K-1} w(j/K)\hat{\Gamma}(j) \ , \quad \text{and} \quad \hat{\Delta} = \Sigma_{j=0}^{K-1} w(j/K)\hat{\Gamma}(j) \ ,$$

which correspond to (10) when the lag kernel is truncated as in KL(b), i.e. $w(x) = 0$ for $|x| > 1$. Thus, the proofs given below for Lemmas 3.1 and 3.3 hold directly under KL(a) and (b) and therefore hold for the Parzen and Tukey-Hanning kernels for example. The results stated also apply for untruncated kernel estimates that satisfy Assumption KL(b′), like QS kernel estimates, but the proofs need some modification to deal with the fact that the sums in (6) are not truncated.

To illustrate the type of modification needed we look at the proof of part (a) of Lemma 3.1 given below. In this proof we have expression (P1) whose second and third terms now become

$$\text{(P0)} \qquad w((T-1)/K)\hat{\Gamma}_{u_1\Delta u_1}(T-1) - w((-T+1)/K)\hat{\Gamma}_{u_1\Delta u_1}(-T) \ .$$

We need (P0) to be $o_p(K^{-2})$ for the remaining arguments in the proof to hold. To show this we observe that

$$\hat{\Gamma}(T-1),\ \hat{\Gamma}(-T) = O_p(T^{-1}) \ ,$$

and

$$w((T-1)/K),\ w((-T+1)/K) = O((K/T)^2) \ ,$$

in view of KL(b′). Combining these expressions we deduce that (P0) is $O_p(K^2/T^3)$ as $T \to \infty$. Thus, (P0) will be $o_p(1/K^2)$ as required if $K^4/T^3 \to 0$, i.e. for $K = O(T^k)$ with $0 < k < 3/4$. This is certainly true under the bandwidth condition BW(i) of Assumption BW whereby $K = O(T^k)$ with k satisfying $1/4 < k < 2/3$. Similar modifications elsewhere in the proofs of Lemmas 3.1 and 3.3 help to establish the stated results under KL(b′).

8.1. PROOF OF LEMMA 3.1

Part (a): By definition

$$\hat{\Omega}_{\Delta u_1 \Delta u_1} = \Sigma_{j=-K+1}^{K-1} w(j/K)\hat{\Gamma}_{\Delta u_1 \Delta u_1}(j) = \Sigma_{j=-K+1}^{K-1} w(j/K)(T^{-1}\Sigma' u_{1t+j}\Delta u_{1t}' - T^{-1}\Sigma' u_{1t-1+j}\Delta u_{1t}')$$

$$= \Sigma_{j=-K+1}^{K-1} w(j/K)[\hat{\Gamma}_{u_1 \Delta u_1}(j) - \hat{\Gamma}_{u_1 \Delta u_1}(j-1)]$$

$$\text{(P1)} \quad = \Sigma_{j=-K+1}^{K-2}[w(j/K) - w((j+1)/K)]\hat{\Gamma}_{u_1 \Delta u_1}(j) + w((K-1)/K)\hat{\Gamma}_{u_1 \Delta u_1}(K-1)$$

$$- w((-K+1)/K)\hat{\Gamma}_{u_1 \Delta u_1}(-K) .$$

Under EC(a) we have

$$\Sigma_{j=0}^{\infty} j^a |\Gamma(j)| \le |\Sigma_\varepsilon| \Sigma_{j=0}^{\infty} j^a \Sigma_{k=0}^{\infty} |C_k||C_{k+j}| = |\Sigma_\varepsilon| \Sigma_{k=0}^{\infty} |C_k| \Sigma_{j=0}^{\infty} j^a |C_{k+j}|$$

$$\text{(P2)} \quad \le |\Sigma_\varepsilon| (\Sigma_{k=0}^{\infty} |C_k|)(\Sigma_{s=0}^{\infty} s^a |C_s|) < \infty .$$

Hence $\Gamma(K) = E(u_{t+K}u_t') = o(1/K^{1+a})$ as $K \to \infty$. Further, $\text{var}(\hat{\Gamma}(K)) = O(T^{-1})$, as shown for example in Hannan (1970, p. 212). Therefore, since $a > 1$ we have

$$\text{(P3)} \quad \hat{\Gamma}(K-1),\ \hat{\Gamma}(-K) = O_p(T^{-1/2}) + o(K^{-2}) = O_p(T^{-1/2})$$

under BW. Moreover, KL implies that

$$\text{(P4)} \quad w((K-1)/K),\ w((-K+1)/K) = O(K^{-2}) ,$$

so that the second and third terms of (P1) are $o_p(K^{-2})$.

This leaves us with the first term of (P1) which we write as

$$\Sigma_{j=-K+1}^{K-2}[w(j/K) - w((j+1)/K)]\hat{\Gamma}_{u_1 \Delta u_1}(j)$$

$$\text{(P5)} \quad = (\Sigma_{\mathscr{B}_*} + \Sigma_{\mathscr{B}^*})[w(j/K) - w((j+1)/K)]\hat{\Gamma}_{u_1 \Delta u_1}(j)$$

where $\mathscr{B}_* = \{j : |j| \le K^*\}$ and $\mathscr{B}^* = \{j : |j| > K^*, -K+1 \le j \le K-2\}$ for some $K^* = K^b$ with $0 < b < 1$. Under KL we can use the following Taylor development for $w((j+1)/K)$ when $|j| \le K^*$ and $K \to \infty$:

The first sum in (P5) is then

$$w((j+1)/K) - w(j/K) = w'(j/K)(1/K) + (1/2)w''(0)(1/K^2)(1+o(1))$$
$$= w''(0)(j/K^2)(1+o(1)) + (1/2)w''(0)(1/K^2)(1+o(1)) .$$

$$-K^2 w''(0)\{\sum_{\mathscr{B}_*} j\hat{\Gamma}_{u_1\Delta u_1}(j) + (1/2)\sum_{\mathscr{B}_*}\hat{\Gamma}_{u_1\Delta u_1}(j)\}[1+o(1)] .$$

The mean of the term in braces is

$$\sum_{|j|\le K^*} j\Gamma_{u_1\Delta u_1}(j) + (1/2)\sum_{|j|\le K^*}\Gamma_{u_1\Delta u_1}(j)$$
$$\to \Sigma_{j=-\infty}^{\infty} j\Gamma_{u_1\Delta u_1}(j) + (1/2)\Sigma_{j=-\infty}^{\infty}\Gamma_{u_1\Delta u_1}(j)$$
$$= \Sigma_{j=-\infty}^{\infty} j\Gamma_{u_1\Delta u_1}(j) = \Sigma_{j=-\infty}^{\infty} j\Gamma_{u_1u_1}(j) - \Sigma_{j=-\infty}^{\infty} j\Gamma_{u_1u_1}(j+1)$$

(P6) $$= \Sigma_{j=-\infty}^{\infty}\Gamma_{u_1u_1}(j) = \Omega_{11} .$$

The second sum in (P5) is

$$\sum_{\mathscr{B}^*}[w(j/K) - w((j+1)/K)]\hat{\Gamma}_{u_1\Delta u_1}(j) = K^{-1}\sum_{\mathscr{B}^*} w'(\theta_j)\hat{\Gamma}_{u_1\Delta u_1}(j)$$

where $j/K < \theta_j < (j+1)/K$. This expression has mean given by

$$K^{-1}\Sigma_{\mathscr{B}^*} w'(\theta_j)\Gamma_{u_1\Delta u_1}(j) ,$$

whose modulus is dominated above by

$$(\sup_{|j|\le K}|w'(\theta_j)|)K^{-1}\Sigma_{|j|>K^*}|\Gamma_{u_1\Delta u_1}(j)|$$
$$\le \text{constant } K^{-1}\Sigma_{|j|>K^*}\Sigma_{s=0}^{\infty}|C_s||C_{s+j}|$$
$$\le \text{constant } K^{-1}K^{*-a}\Sigma_{|j|>K^*}\Sigma_{s=0}^{\infty}(s+j)^a|C_s||C_{s+j}|$$
$$\le \text{constant } K^{-1}K^{-ab}\Sigma_{s=0}^{\infty}|C_s|\Sigma_{r=0}^{\infty}r^a|C_r|$$
$$= O(K^{-1-ab}) .$$

Note from EC(a) that $a > 1$. We may therefore select $K^* = K^b$ with $0 < b < 1$ in such a way that $ab > 1$ (i.e. choose b so that $1/a < b < 1$). Then the mean of the second sum in (P5) has order $o(K^{-2})$ as $K \to \infty$ and therefore the mean of (P5) is dominated by the first term, as we have seen in (P6).

Next we consider the variance matrix of (P5). We start by writing

$$\Sigma_{j=-K+2}^{K-2}[w(j/K) - w((j+1)/K)]\hat{\Gamma}_{u_1\Delta u_1}(j) = -K^{-1}\Sigma_{j=-K+2}^{K-2}w'(j/K)\hat{\Gamma}_{u_1\Delta u_1}(j)[1 + O(1/K)]$$

$$= -K^{-1}s_{j=-K+2}^{K-2}[w'(j/K)\hat{\Gamma}_{u_1u_1}(j) - w'(j/K)\hat{\Gamma}_{u_1u_1}(j+1)][1 + O(1/K)]$$

$$= -K^{-1}\{\Sigma_{j=-K+3}^{K-2}[w'(j/K) - w'((j-1)/K)]\hat{\Gamma}_{u_1u_1}(j) - w'((K-2)/K)\hat{\Gamma}_{u_1u_1}(K-1)$$

$$+ w'((-K+2)/K)\hat{\Gamma}_{u_1u_1}(-K+2)\}[1 + O(1/K)]$$

(P7) $$= -K^{-2}\Sigma_{j=-K+3}^{K-2}w''((j-1)/K)\hat{\Gamma}_{u_1u_1}(j) + o_p(1/K^2) ,$$

using the fact that under KL

$$w'((K-2)/K),\ w'((-K+2)/K) = O(1/K) ,$$

and $\hat{\Gamma}_{u_1u_1}(K-1)$, $\hat{\Gamma}_{u_1u_1}(-K+2) = O_p(T^{-1/2})$, as in (P3) above. We now consider the variance matrix of the leading term of (P7). Following Hannan (1970, p. 280, Theorem 9), we have

(P8) $$\lim_{T\to\infty} K^3T\ \mathrm{Var}\left[\mathrm{vec}\{K^{-2}\Sigma_{j=-K+3}^{K-2}w''((j-1)/K)\hat{\Gamma}_{u_1u_1}(j)\}\right]$$

$$= \lim_{T\to\infty} K^3TK^{-4}\ \mathrm{Var}\left[\mathrm{vec}\{\Sigma_{j=-K+3}^{K-2}w''((j-1)/K)\hat{\Gamma}_{u_1u_1}(j)\}\right]$$

$$= \lim_{T\to\infty} \frac{T}{K}\mathrm{Var}\left[\mathrm{vec}\{\Sigma_{j=-K+3}^{K-2}w''((j-1)/K)\hat{\Gamma}_{u_1u_1}(j)\}\right]$$

$$= \text{constant} .$$

Hence the variance of the dominant terms of (P7) and (P5) is $O(1/TK^3)$.

Leaving aside terms that are $o_p(K^{-2})$ we deduce from (P1), (P6) and (P8) that for $\varepsilon > 0$

$$P[|\hat{\Omega}_{\Delta u_1\Delta u_1} - \{-K^{-2}w''(0)\}\Omega_{11}| > \varepsilon] < \mathrm{tr}\{\mathrm{Var}[\mathrm{vec}(\hat{\Omega}_{\Delta u_1\Delta u_1})]\}/\varepsilon^2 = \text{constant}/\varepsilon^2K^3T \text{ as } T\to\infty.$$

Now set $\delta = K^2\varepsilon$ and we have

$$P\left[|K^2\hat{\Omega}_{\Delta u_1\Delta u_1} - \{-w''(0)\}\Omega_{11}| > \delta\right] = O(K/T) \to 0 , \quad \text{as} \quad T\to\infty .$$

Letting δ be arbitrarily small we have the required result for part (a), *viz.*

$$K^2\hat{\Omega}_{\Delta u_1\Delta u_1} \to_p -w''(0)\Omega_{11} .$$

Part (b): Part (b) of the lemma is proved in a similar way. We have

$$\hat{\Omega}_{u_0\Delta u_1} = \Sigma_{j=-K+1}^{K-1} w(j/K)\hat{\Gamma}_{u_0\Delta u_1}(j)$$

$$= \Sigma_{j=-K+1}^{K-1}[w(j/K)\hat{\Gamma}_{u_0\mu_1}(j) - w(j/K)\hat{\Gamma}_{u_0\mu_1}(j+1)]$$

$$= \Sigma_{j=-K+2}^{K-1}[w(j/K) - w((j-1)/K)]\hat{\Gamma}_{u_0\mu_1}(j) - w((K-1)/K)\hat{\Gamma}_{u_0\mu_1}(K)$$

$$+ w((-K+1)/K)\hat{\Gamma}_{u_0\mu_1}(-K+1)$$

(P9) $$= \Sigma_{j=-K+2}^{K-1}[w(j/K) - w((j-1)/K)]\hat{\Gamma}_{u_0\mu_1}(j) + o_p(K^{-2})$$

as in the analysis following (P1). We write the first term of (P9) as

(P10) $$(\Sigma_{\mathscr{B}_*} + \Sigma_{\mathscr{B}^*})[w(j/K) - w((j-1)/K)]\hat{\Gamma}_{u_0\mu_1}(j)$$

and using the Taylor development of $w(j/K) - w((j-1)/K)$ the first sum in (P10) is

(P11) $$K^{-2}w''(0)\{\Sigma_{|j|\le K_*}(j-1)\hat{\Gamma}_{u_0\mu_1}(j) + (1/2)\Sigma_{|j|\le K^*}\hat{\Gamma}_{u_0\mu_1}(j)\}[1 + o(1)] .$$

The mean of the term in braces in (P11) is

$$\Sigma_{|j|\le K^*}(j-1)\Gamma_{u_0\mu_1}(j) + (1/2)\Sigma_{|j|\le K^*}\Gamma_{u_0\mu_1}(j) - \Sigma_{j=-\infty}^{\infty}(j-1/2)\Gamma_{u_0\mu_1}(j) .$$

The second sum in (P10) is

$$K^{-1}\Sigma_{\mathscr{B}^*}w'((j-1)/K)\hat{\Gamma}_{u_0\mu_1}(j)[1 + O(1/K)] ,$$

whose mean is given by

(P12) $$K^{-1}\Sigma_{\mathscr{B}^*}w'((j-1)/K)\Gamma_{u_0\mu_1}(j)[1 + O(1/K)] .$$

The modulus of (P12) is dominated above by

$$(\sup_x|w'(x)|)K^{-1}\Sigma_{|j|>K^*}|\Gamma_{u_0\mu_1}(j)|(1 + O(1/K)) = O(K^{-1-ab})$$

as in the proof of part (a), and for $1/a < b < 1$ this expression is therefore $o(K^{-2})$ as $K, T \to \infty$.

It follows that the mean of $\hat{\Omega}_{u_0\Delta u_1}$ is dominated by the first term of (P9) which is $O(K^{-2})$, as in (P11). In particular,

$$K^2E(\hat{\Omega}_{u_0\Delta u_1}) \to w''(0)\Sigma_{j=-\infty}^{\infty}(j-1/2)\Gamma_{u_0u_1}(j) .$$

Next we consider the variance matrix of the leading term in (P9), i.e.

$$\Sigma_{j=-K+2}^{K-1}[w(j/K) - w((j-1)/K)]\hat{\Gamma}_{u_0u_1}(j) = K^{-1}\Sigma_{j=-K+2}^{K-1}w'((j-1)/K)\hat{\Gamma}_{u_0u_1}(j)[1+O(K^{-1})] .$$

As in the analysis of (P8) above, we now have

$$\lim_{T\to\infty} KT \ \mathrm{Var}\Big[\mathrm{vec}\{K^{-1}\Sigma_{j=-K+2}^{K-1}w'((j-1)/K)\hat{\Gamma}_{u_0u_1}(j)\}\Big]$$

$$= \lim_{T\to\infty}\frac{T}{K}\ \mathrm{Var}\Big[\mathrm{vec}\{\Sigma_{j=-K+2}^{K-1}w'((j-1)/K)\hat{\Gamma}_{u_0u_1}(j)\}\Big]$$

$$= \text{constant} .$$

Thus, the variance matrix of the dominant term in (P9) is $O(1/KT)$. We deduce that

$$\sqrt{KT}\,|\hat{\Omega}_{u_0\Delta u_1} - \{K^{-2}w''(0)\Sigma_{j=-\infty}^{\infty}(j-1/2)\Gamma_{u_0u_1}(j)\}| = O_p(1) ,$$

i.e.

$$\hat{\Omega}_{u_0\Delta u_1} = K^{-2}w''(0)\Sigma_{j=-\infty}^{\infty}(j-1/2)\Gamma_{u_0u_1}(j) + O_p(1/\sqrt{KT}) ,$$

as required for the first expression in part (b). The second expression, for the limit behavior of $\hat{\Omega}_{u_2\Delta u_1}$, is proved in precisely the same way.

Part (c): To prove part (c) we need to show that

$$\hat{\Omega}_{0\Delta u_1} := \hat{\Omega}_{\hat{u}_0\Delta u_1} = \hat{\Omega}_{u_0\Delta u_1} + O_p(1/T) .$$

Now

$$\hat{\Omega}_{\hat{u}_0\Delta u_1} = \hat{\Omega}_{u_0\Delta u_1} + \Sigma_{j=-K+1}^{K-1}w(j/K)(\hat{A}-A)\hat{\Gamma}_{x\Delta u_1}(j)$$

$$\text{(P13)}\quad = \hat{\Omega}_{u_0\Delta u_1} - w((K-1)/K)(\hat{A}-A)\hat{\Gamma}_{xu_1}(K) + w((-K+1)/K)(\hat{A}-A)\hat{\Gamma}_{xu_1}(-K+1)$$

$$+ \Sigma_{j=-K+2}^{K-1}[w(j/K) - w((j-1)/K)](\hat{A}-A)\hat{\Gamma}_{xu_1}(j) .$$

The second and third terms of (P13) are $o_p(T^{-1})$ because $w((K-1)/K = O(K^{-2})$, $\hat{A}-A = O_p(1/\sqrt{T})$ and $\hat{\Gamma}_{xu_1}(K) = O_p(1)$. Hence

since $K = O(T^{1/4+G})$ for some $G > 0$. The third term of (P13) is

$$w((K-1)/K)(\hat{A}-A)\hat{\Gamma}_{xu_1}(K) = O_p(K^{-2}T^{-1/2}) = o_p(T^{-1})$$

$$\Sigma_{j=-K+2}^{K-1}[w(j/K) - w((j-1)/K)](\hat{A}-A)\hat{\Gamma}_{xu_1}(j)$$

$$\text{(P14)} \quad = \Sigma_{j=-K+2}^{K-1}[w(j/K) - w((j-1)/K)](\hat{A}_1 - A_1)\hat{\Gamma}_{x_1u_1}(j)$$

$$+ \Sigma_{j=-K+2}^{K-1}[w(j/K) - w((j-1)/K)](\hat{A}_2 - A_2)\hat{\Gamma}_{x_2u_1}(j) \ .$$

The first sum in (P14) can be decomposed as in (P10) and (P11). Using the fact that $\hat{A}_1 - A_1 = O_p(T^{-1/2})$, we find that the first term of (P14) is $O_p(1/\sqrt{KT})O_p(1/\sqrt{T}) = o_p(T^{-1})$. For the second term of (P14) we note that $\hat{A}_2 - A_2 = O_p(T^{-1})$ and

$$\Sigma_{j=-K+2}^{K-1}[w(j/K) - w((j-1)/K)]\hat{\Gamma}_{x_2u_1}(j) = K^{-1}\Sigma_{j=-K+2}^{K-1}w'(\theta_j)\hat{\Gamma}_{x_2u_1}(j) = O_p(1) \ ,$$

as in the proof of Theorem 3.1 of Phillips (1991, pp. 432-433). Thus (P14) is at most $O_p(T^{-1})$ and part (c) of the lemma follows.

Part (d): We prove part (d) of the lemma by using the partitioned inversion of Ω_{hh}^{-1}, which yields

$$\hat{\Omega}_{oh}\hat{\Omega}_{hh}^{-1} = \hat{\Omega}_{0\Delta u_1}\Big[\hat{\Omega}_{\Delta u_1\Delta u_1}^{-1} - \hat{\Omega}_{\Delta u_1\Delta u_1}^{-1}\hat{\Omega}_{\Delta u_1u_2}\hat{\Omega}_{u_2u_2\cdot\Delta u_1}^{-1}\hat{\Omega}_{u_2\Delta u_1}\hat{\Omega}_{\Delta u_1\Delta u_1}^{-1}$$

$$\vdots \ -\hat{\Omega}_{\Delta u_1\Delta u_1}^{-1}\hat{\Omega}_{\Delta u_1u_2}\hat{\Omega}_{u_2u_2\cdot\Delta u_1}^{-1}\Big] + \hat{\Omega}_{ou_2}\Big[-\hat{\Omega}_{u_2u_2\cdot\Delta u_1}^{-1}\hat{\Omega}_{\Delta u_1u_2}\hat{\Omega}_{\Delta u_1\Delta u_1}^{-1} \ \vdots \ \hat{\Omega}_{u_2u_2\cdot\Delta u_1}\Big]$$

$$= [X_{01} \ \vdots \ X_{02}] \ , \quad \text{say}$$

where $\hat{\Omega}_{u_2u_2\cdot\Delta u_1} = \hat{\Omega}_{u_2u_2} - \hat{\Omega}_{u_2\Delta u_1}\hat{\Omega}_{\Delta u_1\Delta u_1}^{-1}\hat{\Omega}_{\Delta u_1u_2}$. Using parts (a)-(c) of the lemma we find that

$$\hat{\Omega}_{u_2u_2\cdot\Delta u_1} = \hat{\Omega}_{u_2u_2} + O_p(K^{-2}) + O_p(K/T) = \hat{\Omega}_{u_2u_2} + o_p(1) \longrightarrow_p \Omega_{22} > 0 \ ,$$

$$X_{01} = \hat{\Omega}_{0\Delta u_1}\hat{\Omega}^{-1}_{\Delta u_1\Delta u_1} - \hat{\Omega}_{0\Delta u_1}\hat{\Omega}^{-1}_{\Delta u_1\Delta u_1}\hat{\Omega}_{\Delta u_1 u_2}\hat{\Omega}^{-1}_{u_2u_2\cdot\Delta u_1}\hat{\Omega}_{u_2\Delta u_1}\hat{\Omega}_{\Delta u_1\Delta u_1}$$

$$- \hat{\Omega}_{0u_2}\hat{\Omega}^{-1}_{u_2u_2\cdot\Delta u_1}\hat{\Omega}_{u_2\Delta u_1}\hat{\Omega}_{\Delta u_1\Delta u_1}$$

$$= [-\Phi_{01} + O_p(K^{3/2}/T^{1/2})]\Omega_{11}^{-1} - [O_p(K^{-2}) + O_p(K/T)][\Omega_{22} + o_p(1)]^{-1}[-\Phi_{21} + O_p(K^{3/2}/T^{1/2})]\Omega_{11}^{-1}$$

$$- [\Omega_{02} + o_p(1)][\Omega_{22} + o_p(1)]^{-1}[-\Phi_{21} + O_p(K^{3/2}/T^{1/2})]\Omega_{11}^{-1}$$

$$= -[\Phi_{01} - \Omega_{02}\Omega_{22}^{-1}\Phi_{21}]\Omega_{11}^{-1} + O_p(K^{3/2}/T^{1/2}) + o_p(K^{3/2}/T^{1/2}) ,$$

and

$$X_{02} = -\hat{\Omega}_{0\Delta u_1}\hat{\Omega}^{-1}_{\Delta u_1\Delta u_1}\hat{\Omega}_{\Delta u_1 u_2}\hat{\Omega}^{-1}_{u_2u_2\cdot\Delta u_1} + \hat{\Omega}_{0u_2}\hat{\Omega}^{-1}_{u_2u_2\cdot\Delta u_1}$$

$$= -[\Phi_{01} + O_p(K^2/\sqrt{KT})][-\Omega_{11} + o_p(1)]^{-1}[O_p(K^{-2}) + O_p(1/\sqrt{KT})] + \Omega_{02}\Omega_{22}^{-1} + o_p(1)$$

$$= \Omega_{02}\Omega_{22}^{-1} + o_p(1) .$$

This establishes part (d).

Part (e): To prove part (e) we consider

$$T^{-1}\Delta U_1'U_1 - \hat{\Delta}_{\Delta u_1\Delta u_1} = T^{-1}\Delta U_1'U_1 - \Sigma_{j=0}^{K-1} w(j/K)\hat{\Gamma}_{\Delta u_1\Delta u_1}(j)$$

$$= T^{-1}\Delta U_1'U_1 - \Sigma_{j=0}^{K-1} w(j/K)[\hat{\Gamma}_{\Delta u_1 u_1}(j) - \hat{\Gamma}_{\Delta u_1 u_1}(j+1)]$$

$$= -\Sigma_{j=1}^{K-1}[w(j/K) - w((j-1)/K)]\hat{\Gamma}_{\Delta u_1 u_1}(j) - w((K-1)/K)\hat{\Gamma}_{\Delta u_1 u_1}(K)$$

$$\text{(P15)} \quad = -\Sigma_{j=1}^{K-1}[w(j/K) - w((j-1)/K)]\hat{\Gamma}_{\Delta u_1 u_1}(j) + o_p(K^{-2})$$

using (P3) and (P4). As in the analysis that follows (P9) we rewrite the first term of (P15) as

$$\text{(P16)} \quad -\left(\Sigma_{j=1}^{K^*} + \Sigma_{j=K^*+1}^{K-1}\right)[w(j/K) - w((j-1)/K)]\hat{\Gamma}_{\Delta u_1 u_1}(j) .$$

Upon the expansion of $w(j/K) - w((j-1)/K)$, the first summation in (P16) becomes

$$-K^{-2}w''(0)\Sigma_{j=1}^{K^*}\{(j-1) + 1/2\}\hat{\Gamma}_{\Delta u_1 u_1}(j)$$

whose mean is

$$-K^{-2}w''(0)\Sigma_{j=1}^{K^*}(j-1/2)\Gamma_{\Delta u_1 u_1}(j) = -K^{-2}w''(0)\Sigma_{j=1}^{K^*}\{(j-1/2)\Gamma_{u_1u_1}(j) - (j-1/2)\Gamma_{u_1u_1}(j-1)\}$$

$$= -K^{-2}w''(0)\left\{\Sigma_{j=1}^{K^*-1}\{(j-1/2) - (j+1/2)\}\Gamma_{u_1u_1}(j) + (K^*-1/2)\Gamma_{u_1u_1}(K^*) - (1/2)\Gamma_{u_1u_1}(0)\right\}$$

$$= K^{-2}w''(0)\left\{\Sigma_{j=1}^{K^*-1}\Gamma_{u_1u_1}(j) + (1/2)\Gamma_{u_1u_1}(0)\right\} + o_p(K^{-2}) .$$

Thus

$$K^2E\left\{-K^{-2}w''(0)\Sigma_{j=1}^{K^*}(j-1/2)\hat{\Gamma}_{\Delta u_1u_1}(j)\right\} - w''(0)\left\{\Sigma_{j=1}^{\infty}\Gamma_{u_1u_1}(j) + (1/2)\Gamma_{u_1u_1}(0)\right\}$$

$$= w''(0)\{\Delta_{11} - (1/2)\Sigma_{11}\} . \tag{P17}$$

The second sum in (P16) is

$$-\Sigma_{j=K^*+1}^{K-1}[w(j/K) - w((j-1)/K)]\hat{\Gamma}_{\Delta u_1u_1}(j)$$

whose mean is given by

$$-K^{-1}\Sigma_{j=K^*+1}^{K-1}w'((j-1)/K)\Gamma_{\Delta u_1u_1}(j)[1+O(K^{-1})] . \tag{P18}$$

The modulus of (P18) is dominated above by

$$(\sup_x|w'(x)|)K^{-1}\Sigma_{j>K^*}|\Gamma_{\Delta u_1u_1}(j)|(1+O(1/K)) = O(K^{-1-ab}) = o(K^{-2})$$

for $1/a < b < 1$, as in the proof of part (a). It follows from (P15)-(P19) that

$$K^2E[T^{-1}\Delta U_1'U_1 - \hat{\Delta}_{\Delta u_1\Delta u_1}] - w''(0)\{\Delta_{11} - (1/2)\Sigma_{11}\} .$$

The variance of the dominant term of (P15) may now be shown to be $O(1/TK^3)$, just as in the proof of part (a), and it follows that

$$K^2[T^{-1}\Delta U_1'U_1 - \hat{\Delta}_{\Delta u_1\Delta u_1}] \to_p w''(0)\{\Delta_{11} - (1/2)\Sigma_{11}\}$$

as required.

Part (f): To prove part (f) we proceed in the same way as the proof of part (b) the only difference being that the sums are one-sided rather than two-sided. The mean of $T^{-1}U_2'U_1 - \hat{\Delta}_{u_2\Delta u_1}$ is $O(K^{-2})$ and satisfies

$$K^2E[T^{-1}U_2'U_1 - \hat{\Delta}_{u_2\Delta u_1}] - w''(0)\Sigma_{j=1}^{\infty}(j-1/2)\Gamma_{u_2u_1}(j) = w''(0)\Psi_{21} .$$

The variance matrix is of $O(1/KT)$ and hence

$$\sqrt{KT}\,\|\{T^{-1}U_2'U_1 - \hat{\Delta}_{u_2\Delta u_1}\} - \{K^{-2}w''(0)\Psi_{21}\}\| = O_p(1)$$

giving the stated result for part (f).

Part (g): To prove (g) we first note that

$$T^{-1}\Delta U_1'X_2 - \hat{\Delta}_{\Delta u_1u_2} = T^{-1}U_1'X_2 - T^{-1}U_{1-1}'X_2 - \hat{\Delta}_{\Delta u_1u_2}$$

$$= T^{-1}u_{1T}x_{2T}' - T^{-1}U_{1-1}'U_2 - \hat{\Delta}_{\Delta u_1u_2} . \tag{P18}$$

Next, observe that

$$\hat{\Delta}_{\Delta u_1u_2} + T^{-1}U_{1-1}'U_2 = \Sigma_{j=0}^{K-1}w(j/K)\hat{\Gamma}_{\Delta u_1u_2}(j) + \hat{\Gamma}_{u_1u_2}(-1)$$

$$= \Sigma_{j=0}^{K-1}w(j/K)[\hat{\Gamma}_{u_1u_2}(j) - \hat{\Gamma}_{u_1u_2}(j-1)] + \hat{\Gamma}_{u_1u_2}(-1)$$

$$= \Sigma_{j=0}^{K-2}[w(j/K) - w((j+1)K)]\hat{\Gamma}_{u_1u_2}(j) + w((K-1)/K)\hat{\Gamma}_{u_1u_2}(K-1)$$

$$= \Sigma_{j=0}^{K-2}[w(j/K) - w((j+1)/K)]\hat{\Gamma}_{u_1u_2}(j) + o_p(K^{-2}) . \tag{P19}$$

We now proceed as in the proof of part (b) but with a one-sided sum. We find that, for the mean, we have

$$K^2\Sigma_{j=0}^{K-2}[w(j/K) - w((j+1)/K)]\Gamma_{u_1u_2}(j) - w''(0)\Sigma_{j=0}^{\infty}(j+1/2)\Gamma_{u_1u_2}(j)$$

and the variance matrix of the first term of (P19) is $O(1/KT)$. Hence (P19) is

$$K^{-2}w''(0)\Psi_{12} + O_p(1/\sqrt{KT})$$

and combining this with (P18) we get the stated result for part (g).

Part (h): To prove part (h) we write

$$\hat{\Delta}_{\hat{u}_0\Delta u_1} = \Sigma_{j=0}^{K-1} w(j/K)\hat{\Gamma}_{\hat{u}_0\Delta u_1}(j) = \Sigma_{j=0}^{K-1} w(j/K)[\hat{\Gamma}_{\hat{u}_0 u_1}(j) - \hat{\Gamma}_{\hat{u}_0 u_1}(j+1)]$$

$$= \Sigma_{j=1}^{K-1}[w(j/K) - w((j-1)/K)]\hat{\Gamma}_{\hat{u}_0 u_1}(j) + \hat{\Gamma}_{\hat{u}_0 u_1}(0) - w((K-1)/K)\hat{\Gamma}_{\hat{u}_0 u_1}(K)$$

$$\text{(P20)} \qquad = \Sigma_{j=1}^{K-1}[w(j/K) - w((j-1)/K)]\hat{\Gamma}_{\hat{u}_0 u_1}(j) + O_p(K^{-2}T^{-1/2})$$

since $\hat{\Gamma}_{\hat{u}_0 u_1}(0) = T^{-1}\hat{U}_0'U_1 = T^{-1}\hat{U}_0'X_1 = 0$ by least squares orthogonality, $w((K-1)/K) = O(K^{-2})$ and $\hat{\Gamma}_{\hat{u}_0 u_1}(K) = O_p(T^{-1/2})$. Now

$$\hat{\Gamma}_{\hat{u}_0 u_1}(j) = \hat{\Gamma}_{u_0 u_1}(j) + (A - \hat{A})\hat{\Gamma}_{x u_1}(j)$$

and (P20) becomes

$$\text{(P21)} \quad \Sigma_{j=1}^{K}[w(j/K) - w((j-1)/K)]\hat{\Gamma}_{u_0 u_1}(j) + (A - \hat{A})\Sigma_{j=1}^{K}[w(j/K) - w((j-1)/K)]\hat{\Gamma}_{x u_1}(j) + O_p(K^{-2}T^{-1/2}) .$$

The first term of (P21) has mean zero because

$$\Gamma_{u_0 u_1}(j) = 0 \quad \text{for all} \quad j \geq 0$$

in view of EC(c). The variance of the first term of (P21) is $O(1/KT)$, just as shown in part (b). Hence,

$$\text{(P22)} \quad \Sigma_{j=1}^{K}[w(j/K) - w((j-1)/K)]\hat{\Gamma}_{u_0 u_1}(j) = O_p(1/\sqrt{KT}) .$$

Next, the second term of (P21) can be shown to be $O_p(T^{-1})$ just as (P14) in the proof of part (c). Thus, combining (P20) and (P22) we have $\hat{\Delta}_{\hat{u}_0\Delta u_1} = O_p(1/\sqrt{KT})$, as required for part (h).

Part (i): To prove (i) we write

$$\hat{\Delta}_{\hat{u}_0 u_2} = \Sigma_{j=0}^{K-1} w(j/K)\hat{\Gamma}_{\hat{u}_0 u_2}(j) = \Sigma_{j=0}^{K-1} w(j/K)\hat{\Gamma}_{u_0 u_2}(j) + (A - \hat{A})\Sigma_{j=0}^{K-1} w(j/K)\hat{\Gamma}_{x u_2}(j) .$$

The first term is

$$\Sigma_{j=0}^{K-1} w(j/K)\hat{\Gamma}_{u_0 u_2}(j) = \Delta_{02} + O_p((K/T)^{1/2})$$

since its mean is

$$\Sigma_{j=0}^{K-1} w(j/K)\Gamma_{u_0\mu_2}(j) - \Sigma_{j=0}^{\infty}\Gamma_{u_0\mu_2}(j) = \Delta_{02} ,$$

and its variance matrix satisfies

$$\lim_{T\to\infty} \frac{T}{K} \mathrm{Var}\left[\mathrm{vec}\{\Sigma_{j=0}^{K-1} w(j/K)\hat{\Gamma}_{u_0\mu_2}(j)\}\right] = \text{constant} ,$$

as in the proof of part (b). This establishes (i).

Parts (j)-(ℓ): Parts (j)-(ℓ) follow directly from the weak convergence theory for sample covariances developed in Phillips-Durlauf (1986), Phillips (1988) and Park-Phillips (1988).

8.2. PROOF OF LEMMA 3.3: Using parts (d), (e), (f), (g) and (j) of Lemma 3.1, we have

$$\hat{\Omega}_{oh}\hat{\Omega}_{hh}^{-1}[T^{-1}U_h'X_h - \hat{\Delta}_{hh}]$$

$$= \left[-(\Phi_{01} - \Omega_{02}\Omega_{22}^{-1}\Phi_{21})\Omega_{11}^{-1} + O_p\left((K^3/T)^{1/2}\right) \ \vdots \ \Omega_{02}\Omega_{22}^{-1} + o_p(1)\right]$$

$$\times \begin{bmatrix} T^{-1}\Delta U_1'U_1 - \hat{\Delta}_{\Delta u_1 \Delta u_1} & T^{-1}\Delta U_1'X_2 - \hat{\Delta}_{\Delta u_1 u_2} \\ \hline T^{-1}U_2'U_1 - \hat{\Delta}_{u_2\Delta u_1} & T^{-1}U_2'X_2 - \hat{\Delta}_{u_2 u_2} \end{bmatrix}$$

$$= \left[-(\Phi_{01} - \Omega_{02}\Omega_{22}^{-1}\Phi_{21})\Omega_{11}^{-1} + O_p\left((K^3/T)^{1/2}\right) \ \vdots \ \Omega_{02}\Omega_{22}^{-1} + o_p(1)\right]$$

$$\times \begin{bmatrix} O_p(K^{-2}) & O_p(T^{-1/2}) \\ \hline O_p(K^{-2}) + O_p(1/\sqrt{KT}) & N_{22T} \end{bmatrix}$$

$$= \left[O_p(K^{-2}) + O_p(1/\sqrt{KT}) \ \vdots \ \Omega_{02}\Omega_{22}^{-1}N_T + O_p(T^{-1/2}) + O_p(K^{3/2}/T) + o_p(1)\right] \quad \text{(P23)}$$

where $N_{22T} \to_d \int_0^1 dB_2B_2'$. This proves part (a). Part (b) follows directly from the first block submatrix of (P23) after scaling by $T^{1/2}$.

Part (c) follows from Lemma 3.1(h) and CLT (5). Thus,

$$\text{(P24)}\quad T^{1/2}[T^{-1}U_0'X_1 - \hat{\Delta}_{0\Delta u_1}] = T^{-1/2}U_0'U_1 - T^{1/2}\hat{\Delta}_{0\Delta u_1} = T^{-1/2}U_0'U_1 + O_p(1/\sqrt{K})$$
$$\longrightarrow_d N(0,\ \Omega_{\varphi\varphi})\ .$$

8.3. PROOF OF THEOREM 3.6: We write the FM-OLS estimation error as

$$\text{(P25)}\quad \hat{A}^+ - A = (U_0^{+\prime}X - T\hat{\Delta}_{0x}^+)(X'X)^{-1}$$

where $U_0^{+\prime} = U_0' - \hat{\Omega}_{0x}\hat{\Omega}_{xx}^{-1}\Delta X'$. Transforming this system by H and partitioning into stationary and nonstationary coefficients, we have

$$\text{(P26)}\quad [(\hat{A}^+ - A)H_1 \ \vdots\ (\hat{A}^+ - A)H_2] = (\hat{A}^+ - A)H = (U_0^{+\prime}X - T\hat{\Delta}_{0x}^+)H(H'X'XH)^{-1}H'H\ .$$

Note that by partitioned inversion

$$(H'X'XH)^{-1}H'H_1 = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1}\begin{bmatrix} I \\ 0 \end{bmatrix} = \begin{bmatrix} (X_1'Q_2X_1)^{-1} \\ -(X_2'Q_1X_2)^{-1}X_2'X_1(X_1'X_1)^{-1} \end{bmatrix},$$

and

$$\text{(P27)}\quad (H'X'XH)^{-1}H'H_2 = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1}\begin{bmatrix} 0 \\ I \end{bmatrix} = \begin{bmatrix} -(X_1'X_1)^{-1}X_1'X_2(X_2'Q_1X_2)^{-1} \\ (X_2'Q_1X_2)^{-1} \end{bmatrix},$$

where $Q_i = I - X_i(X_i'X_i)^{-1}X_i'$ $(i = 1,2)$. It follows that

$$\sqrt{T}(\hat{A}^{+}-A)H_1 = \sqrt{T}[T^{-1}U_0'X - \hat{\Omega}_{0x}\hat{\Omega}_{xx}^{-1}(T^{-1}\Delta X'X) - \hat{\Delta}_{0x}^{+}] \times H\begin{bmatrix} (T^{-1}X_1'Q_2X_1)^{-1} \\ -T^{-1}(T^{-2}X_2'Q_1X_2)^{-1}(T^{-1}X_2'X_1)^{-1}(T^{-1}X_1'X_1)^{-1}\end{bmatrix}$$

$$= \sqrt{T}[T^{-1}U_0'X - \hat{\Omega}_{0x}\hat{\Omega}_{xx}^{-1}(T^{-1}\Delta X'X) - \hat{\Delta}_{0x}^{+}][H_1(T^{-1}X_1'X_1)^{-1} + O_p(T^{-1})]$$

$$= \sqrt{T}[T^{-1}U_0'X_1 - \hat{\Omega}_{0x}H(H'\hat{\Omega}_{xx}H)^{-1}H'(T^{-1}\Delta X'X_1) - (\Delta_{0x} - \hat{\Omega}_{0x}\hat{\Omega}_{xx}^{-1}\hat{\Delta}_{xx})H_1](T^{-1}X_1'X_1)^{-1} + O_p(T^{-1/2})$$

$$\text{(P28)} \quad = \sqrt{T}[\{T^{-1}U_0'X_1 - \hat{\Delta}_{0\Delta u_1}\} - \hat{\Omega}_{0h}\hat{\Omega}_{hh}^{-1}\{T^{-1}U_h'X_1 - \hat{\Delta}_{h\Delta u_1}\}](T^{-1}X_1'X_1)^{-1} + O_p(T^{-1/2})$$

$$= [\{T^{-1/2}U_0'X_1 + O_p(K^{-1/2})\} - \{O_p(K^{-2}T^{1/2}) + O_p(K^{-1/2})\}](T^{-1}X_1'X_1)^{-1} + O_p(T^{-1/2}) ,$$

by virtue of Lemma 3.2(b) and (c). Thus

$$\text{(P29)} \quad \sqrt{T}(\hat{A}^{+}-A)H_1 = (T^{-1/2}U_0'X_1)(T^{-1}X_1'X_1)^{-1} + O_p(K^{-1/2}) + O_p(K^{-2}T^{1/2}) \longrightarrow_d N\left(0,\ (I \otimes \Sigma_{11}^{-1})\Omega_{\varphi\varphi}(I \otimes \Sigma_{11}^{-1})\right) ,$$

as required for part (a). From (P29) we also see that the stated result holds for a bandwidth expansion rate $K = O(T^k)$ with $1/4 < k < 1$.

Next, using (P26) and (P27) we have

(P30) $$T(\hat{A}^+ - A)H_2 = [T^{-1}U_0'X - \hat{\Omega}_{0x}\hat{\Omega}_{xx}^{-1}(T^{-1}\Delta X'X) - \hat{\Delta}_{0x}^+]$$

$$\times H\begin{bmatrix} -(T^{-1}X_1'X_1)^{-1}(T^{-1}X_1'X_2)(T^{-2}X_2'Q_1X_2)^{-1} \\ (T^{-2}X_2'Q_1X_2)^{-1} \end{bmatrix}$$

$$= -[T^{-1}U_0'X - \hat{\Omega}_{0x}\hat{\Omega}_{xx}^{-1}(T^{-1}\Delta X'X)^{-1} - \hat{\Delta}_{0x}^+]H_1(T^{-1}X_1'X_1)^{-1}(T^{-1}X_1'X_2)(T^{-2}X_2'Q_1X_2)^{-1}$$

$$+ [T^{-1}U_0'X - \hat{\Omega}_{0x}\hat{\Omega}_{xx}^{-1}(T^{-1}\Delta X'X) - \hat{\Delta}_{0x}^+]H_2(T^{-2}X_2'Q_1X_2)^{-1}$$

(P31) $$= -[(T^{-1}U_0'U_1 - \hat{\Delta}_{0\Delta u_1}) - \hat{\Omega}_{0h}\hat{\Omega}_{hh}^{-1}(T^{-1}U_h'U_1 - \hat{\Delta}_{h\Delta u_1})](T^{-1}X_1'X_1)^{-1}(T^{-1}X_1'X_2)$$

$$\times (T^{-2}X_2'Q_1X_2)^{-1} + [(T^{-1}U_0'X_2 - \hat{\Delta}_{0u_2}) - \hat{\Omega}_{0h}\hat{\Omega}_{hh}^{-1}\{(T^{-1}U_h'X_2) - \hat{\Delta}_{hu_2}\}](T^{-2}X_2'Q_1X_2)^{-1}$$

(P32) $$= -[\{O_p(T^{-1/2}) + O_p((KT)^{-1/2})\} - O_p((KT)^{-1/2})][O_p(1)]$$

$$+ [\{N_{02T} + o_p(1)\} - \{\Omega_{02}\Omega_{22}^{-1}(N_{22T} + o_p(1)) + O_p(K^{3/2}/T)\}](T^{-2}X_2'Q_1X_2)^{-1}$$

$$\to_d [\int_0^1 dB_0B_2' - \Omega_{02}\Omega_{22}^{-1}\int_0^1 dB_2B_2'\Big]\Big[\int_0^1 B_2B_2'\Big]^{-1}$$

$$= [\int_0^1 dB_{0\cdot2}B_2'\Big]\Big[\int_0^1 B_2B_2'\Big]^{-1}$$

as required for part (b). From (P32) we see that the stated result holds for a bandwidth expansion rate $K = O(T^k)$ with $0 < k < 2/3$.

Parts (a) and (b) of the theorem hold simultaneously when the bandwidth expansion rate is $K = O(T^k)$ with $1/4 < k < 2/3$.

8.4. PROOF OF COROLLARY 3.7: We work from the proof of Theorem 3.3. Since there is no second block in (P26) and no need for a rotation of the regressor space we have from (P28)

provided $K = O(T^k)$ for $1/4 < k < 1$, as stated.

8.5. PROOF OF COROLLARY 3.8: This follows from the original analysis in Phillips-Hansen (1990). We can deduce the result directly from (P31) noting that when $m_1 = 0$ there is no first term in this expression and then

$$\sqrt{T}(\hat{A}^{+}-A) = \sqrt{T}[\{T^{-1}U_0'X_1 - \hat{\Delta}_{0\Delta u_1}\} - \hat{\Omega}_{0h}\hat{\Omega}_{hh}^{-1}\{T^{-1}U_h'X_1 - \hat{\Delta}_{h\Delta u_1}\}](T^{-1}X_1'X_1)^{-1}$$

$$= \sqrt{T}[\{T^{-1}U_0'X_1 - \hat{\Delta}_{h\Delta u_1}\} - \hat{\Omega}_{0\Delta u_1}\hat{\Omega}_{\Delta u_1\Delta u_1}^{-1}\{T^{-1}\Delta U_1'U_1 - \hat{\Delta}_{\Delta u_1\Delta u_1}\}](T^{-1}X_1'X_1)^{-1}$$

$$= [\{T^{-1/2}U_0'X_1 + O_p(K^{-1/2})\} - T^{-1/2}\{-\Phi_{01}\Omega_{11}^{-1} + O_p((K^3/T)^{1/2})\}\{O_p(K^{-2})\}]$$

$$\text{(P33)} \quad = (T^{-1/2}U_0'X_1)(T^{-1}X_1'X_1)^{-1} + O_p(T^{1/2}K^{-2}) + O_p(K^{-1/2})$$

$$\longrightarrow_d N\left(0,\ (I \otimes \Sigma_{11}^{-1})\Omega_{\varphi\varphi}(I \otimes \Sigma_{11}^{-1})\right),$$

$$T(\hat{A}^{+}-A) = [(T^{-1}U_0'X_2 - \hat{\Delta}_{0u_2}) - \hat{\Omega}_{0h}\hat{\Omega}_{hh}^{-1}\{T^{-1}U_h'X_2 - \hat{\Delta}_{hu_2}\}](T^{-2}X_2'X_2)^{-1}$$

$$= [(T^{-1}U_0'X_2 - \hat{\Delta}_{0u_2}) - \hat{\Omega}_{02}\hat{\Omega}_{22}^{-1}\{T^{-1}U_2'X_2 - \hat{\Delta}_{u_2u_2}\}](T^{-2}X_2'X_2)^{-1}$$

$$= [\{N_{02T} + o_p(1)\} - \{\Omega_{02}\Omega_{22}^{-1} + o_p(1)\}\{N_{22T} + o_p(1)\}](T^{-2}X_2'X_2)^{-1}$$

$$\to \left(\int_0^1 dB_{0\cdot 2}B_2'\right)\left(\int_0^1 B_2B_2'\right)^{-1}$$

as required. The only restriction on K for this result is $K \to \infty$, $T/K \to \infty$ as $T \to \infty$. This applies when $K = O(T^k)$ and $0 < k < 1$.

8.6. PROOF OF THEOREM 3.10: Under the null hypothesis $\mathcal{H}_0$ the Wald statistic W^+ has the form

$$W_{00}^{+} = T\left\{(R_1 \otimes R_2')\mathrm{vec}(\hat{A}^{+}-A)\right\}'\left\{R_1\hat{\Sigma}_{00}R_1' \otimes R_2'T(X'X)^{-1}R_2\right\}^{-1}(R_1 \otimes R_2')\mathrm{vec}(\hat{A}^{+}-A)$$

$$= T\ \mathrm{tr}\left\{(R_1\hat{\Sigma}_{00}R_1')^{-1}[R_1(\hat{A}^{+}-A)R_2][R_2'T(X'X)^{-1}R_2]^{-1}[R_1(\hat{A}^{+}-A)R_2]'\right\}.$$

Since $R_2 = [H_1S_{21}, H_2S_{22}]$ we have

$$\text{(P34)} \quad R_2'(X'X)^{-1}R_2 = \begin{bmatrix} S_{21}'H_1'(X'X)^{-1}H_1S_{21} & S_{21}'H_1'(X'X)^{-1}H_2S_{22} \\ S_{22}'H_2'(X'X)^{-1}H_1S_{21} & S_{22}'H_2'(X'X)^{-1}H_2S_{22} \end{bmatrix}$$

and

$$\text{(P35)} \quad R_1(\hat{A}^+ - A)R_2 = R_1[(\hat{A}_1^+ - A_1)S_{21} \vdots (\hat{A}_2^+ - A_2)S_{22}] .$$

Set $D_T = \text{diag}[I_{q_{21}}, T^{1/2}I_{q_{22}}]$ and then we can write

$$\text{(P36)} \quad W_{00}^+ = \text{tr}\left\{(R_1\hat{\Sigma}_{00}R_1')^{-1}[R_1T^{1/2}(\hat{A}^+ - A)R_2]D_T\left[D_TR_2'T(X'X)^{-1}R_2D_T\right]^{-1}D_T[R_1T^{1/2}(\hat{A}^+ - A)R_2]'\right\} .$$

Now

$$T(X'X)^{-1} = TH(H'X'XH)^{-1}H' = TH\begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}H'$$

$$= TH\begin{bmatrix} (X_1'Q_2X_1)^{-1} & -(X_1'Q_2X_1)^{-1}X_1'X_2(X_2'X_2)^{-1} \\ -(X_2'Q_1X_2)^{-1}X_2'X_1(X_1'X_1)^{-1} & (X_2'Q_1X_2)^{-1} \end{bmatrix}H'$$

and

$$TH_1'(X'X)^{-1}H_1 = T(X_1'Q_2X_1)^{-1} \to_p \Sigma_{11}^{-1} ,$$

$$T^{3/2}H_1'(X'X)^{-1}H_2 = -T^{-1/2}(T^{-1}X_1'Q_2X_1)^{-1}(T^{-1}X_1'X_2)(T^{-2}X_2'X_2)^{-1} = O_p(T^{-1/2}) ,$$

$$T^2H_2'(X'X)^{-1}H_2 = (T^{-2}X_2'Q_1X_2)^{-1} \to_d \left(\int_0^1 B_2B_2'\right)^{-1} .$$

Hence,

$$D_TR_2'T(X'X)^{-1}R_2D_T \to_d \begin{bmatrix} S_{21}'\Sigma_{11}^{-1}S_{21} & 0 \\ 0 & S_{22}'\left(\int_0^1 B_2B_2'\right)^{-1}S_{22} \end{bmatrix} .$$

Also $R_1'\hat{\Sigma}_{00}R_1 \to_p R_1'\Sigma_{00}R_1$ and

$$[R_1T^{1/2}(\hat{A}^+ - A)R_2]D_T = [R_1T^{1/2}(\hat{A}_1^+ - A_1)S_{21} \vdots R_1T(\hat{A}_2^+ - A_2)S_{22}]$$

$$\to_d \left[N(0,\ R_1\Sigma_{00}R_1 \otimes S_{21}'\Sigma_{11}^{-1}S_{21}) \vdots R_1\left(\int_0^1 dB_{0\cdot 2}B_2'\right)\left(\int_0^1 B_2B_2'\right)^{-1}S_{22}\right] := [Z_1 \vdots Z_2] .$$

Combining these limits we have

$$W^{+}_{00} \rightarrow_d \operatorname{tr}\Big\{(R_1\Sigma_{00}R_1')^{-1}\Big[Z_1\Big(S_{21}'\Sigma_{11}^{-1}S_{21}\Big)^{-1}Z_1' + Z_2S_{22}'\Big(\int_0^1 B_2B_2'\Big)^{-1}S_{22}Z_2'\Big]\Big\}$$

$$= \operatorname{tr}\Big\{(R_1\Sigma_{00}R_1')^{-1}Z_1\Big(S_{21}'\Sigma_{11}^{-1}S_{21}\Big)^{-1}Z_1'\Big\} + \operatorname{tr}\Big\{(R_1\Sigma_{00}R_1')^{-1}Z_2S_{22}'\Big(\int_0^1 B_2B_2'\Big)^{-1}S_{22}Z_2'\Big\}$$

$$= \operatorname{tr}(V_1V_1') + \operatorname{tr}\Big\{(R_1\Omega_{00\cdot 2}R_1')^{1/2}(R_1\Sigma_{11}R_1')^{-1}(R_1\Omega_{00\cdot 2}R_1')^{1/2}V_2V_2'\Big\} ,$$

where

$$V_1 = (R_1\Sigma_{00}R_1')^{-1/2}Z_1(S_{21}'\Sigma_{11}^{-1}S_{21})^{-1/2} \equiv N(0, I_{q_1q_{21}}) ,$$

and

$$V_2 = (R_1\Omega_{00\cdot 2}R_1')^{-1/2}Z_2$$

$$= (R_1\Omega_{00\cdot 2}R_1')^{-1/2}\Big(R_1\int_0^1 dB_{0\cdot 2}B_2'\Big)\Big(\int_0^1 B_2B_2'\Big)^{-1}S_{22}\Big\{S_{22}'\Big(\int_0^1 B_2B_2'\Big)^{-1}S_{22}\Big\}^{-1/2} \equiv N(0, I_{q_1q_{22}}) .$$

Now let C be an orthogonal matrix for which

$$C'(R_1\Omega_{00\cdot 2}R_1')^{1/2}(R_1\Sigma_{00}R_1')^{-1}(R_1\Omega_{00\cdot 2}R_1')^{1/2}C = D = \operatorname{diag}(d_1, \ldots, d_{q_1}) ,$$

and let $\bar{V}_2 = C'V_2 \equiv N(0, I_{q_1q_{22}})$. Then

$$W^{+}_{00} \sim \operatorname{tr}(V_1V_1') + \operatorname{tr}(D\bar{V}_2\bar{V}_2') \equiv \chi^2_{q_1q_{21}} + \Sigma_{i=1}^{q_1} d_i\chi^2_{q_{22}}(i) ,$$

where $\chi^2_{q_{22}}(i) \equiv \text{iid}(\chi^2_{q_{22}})$, $i = 1, \ldots, q_{22}$. Thus, the limit distribution of W^{+} is a linear combination of χ^2 variates and the stated result follows.

8.7. PROOF OF THEOREM 5.1: The matrix X is partitioned into stationary and nonstationary components as

$$X = [Z, Y_{-1}] = [Z, Y_{1,-1} \vdots Y_{2,-1}] = [Z, V \vdots Y_{2,-1}] = [X_1 \vdots X_2] .$$

Using this partition and the formula for $\hat{F}^{+}$ given in (30′) we have in an obvious subscript notation

(P37) $\hat{F}^{+} - F = [E'Z,\ E'V - \hat{\Omega}_{\varepsilon y}\hat{\Omega}_{yy}^{-1}(\Delta Y_{-1}'V - T\hat{\Delta}_{\Delta y \Delta v})$

$$\vdots\ E'X_2 - \hat{\Omega}_{\varepsilon y}\hat{\Omega}_{yy}^{-1}(\Delta Y_{-1}'X_2 - T\hat{\Delta}_{\Delta y \Delta x_2})](X'X)^{-1}\ .$$

Now, partitioning the inverse of $X'X$ we obtain

$$\sqrt{T}(\hat{F}_1^{+} - F_1) = \sqrt{T}[T^{-1}E'Z,\ T^{-1}E'V - \hat{\Omega}_{\varepsilon y}\hat{\Omega}_{yy}^{-1}(T^{-1}\Delta Y_{-1}'V - \hat{\Delta}_{\Delta y \Delta v})](T^{-1}X_1'Q_2X_1)^{-1}$$

$$- T^{-1/2}[T^{-1}E'X_2 - \hat{\Omega}_{\varepsilon y}\hat{\Omega}_{yy}^{-1}(T^{-1}\Delta Y_{-1}'X_2 - \hat{\Delta}_{\Delta y \Delta x_2})](T^{-2}X_2'X_2)^{-1}(T^{-1}X_2'X_1)(T^{-1}X_1'Q_2X_1)^{-1})$$

$$= [T^{-1/2}E'X_1 + O_p(K^{-2}T^{1/2}) + O_p(K^{-1/2})](T^{-1}X_1'X_1)^{-1} + O_p(T^{-1/2})\ ,$$

where the error orders of magnitude follow from Lemma 3.1 much as in the proof of the first part of Theorem 3.5. We deduce that

(P38) $\sqrt{T}(\hat{F}_1^{+} - F_1) = (T^{-1/2}E'X_1)(T^{-1}X_1'X_1)^{-1} + o_p(1) \longrightarrow_d N(0,\ \Sigma_{\varepsilon\varepsilon} \otimes \Sigma_{11}^{-1})\ ,$

where $\Sigma_{11} = E(x_{1t}x_{1t}')$ and is positive definite, as shown in Lemma 1(iii) of Toda and Phillips (1991).

Next we consider the second block of (P37), i.e.

$$\sqrt{T}(\hat{F}_2^{+} - F_2) = -[T^{-1}E'X_1 + O_p(K^{-2}) + O_p(K^{-1/2}T^{-1/2})](T^{-1}X_1'Q_2X_1)^{-1}(T^{-1}X_1'X_2)(T^{-2}X_2'X_2)^{-1}$$

$$+ [T^{-1}E'X_2 - \hat{\Omega}_{\varepsilon y}\hat{\Omega}_{yy}^{-1}(T^{-1}\Delta Y_{-1}'X_2 - \hat{\Delta}_{\Delta y \Delta x_2})](T^{-2}X_2'Q_1X_2)^{-1}$$

$$= O_p(T^{-1/2}) + [T^{-1}E'X_2 - \Omega_{\varepsilon 2}\Omega_{22}^{-1}(T^{-1}\Delta X_2'X_2 - \hat{\Delta}_{\Delta x_2 \Delta x_2})][T^{-2}X_2'X_2 + O_p(T^{-1})]^{-1}$$

using Lemma 3.1. Now

$$T^{-1}E'X_2 \longrightarrow_d \int_0^1 dB_\varepsilon B_2'\ ,\quad T^{-2}X_2'X_2 \longrightarrow_d \int_0^1 B_2B_2'\ ,$$

$$T^{-1}\Delta X_2'X_2 \longrightarrow_d \int_0^1 dB_2B_2' + \Delta_{u_2u_2}\ ,\quad \text{and}\quad \hat{\Delta}_{\Delta x_2 \Delta x_2} \longrightarrow_p \Delta_{u_2u_2}\ .$$

Hence

(P39) $T(\hat{F}_2^{+} - F_2) \longrightarrow_d \left(\int_0^1 dB_{\varepsilon\cdot 2}B_2'\right)\left(\int_0^1 B_2B_2'\right)^{-1}\ ,$

where $B_{\varepsilon\cdot 2} = B_\varepsilon - \Omega_{\varepsilon 2}\Omega_{22}^{-1}B_2 = BM(\Omega_{\varepsilon\varepsilon\cdot 2})$ with $\Omega_{\varepsilon\varepsilon\cdot 2} = \Omega_{\varepsilon\varepsilon} - \Omega_{\varepsilon 2}\Omega_{22}^{-1}\Omega_{2\varepsilon}$ $= \Sigma_{\varepsilon\varepsilon} - \Omega_{\varepsilon 2}\Omega_{22}^{-1}\Omega_{2\varepsilon}$.

Again, the error orders of magnitude in these derivations follow as in the proof of Theorem 3.5. Consequently, the bandwidth expansion rates under which (P38) and (P39) hold are the same as those given in Theorem 3.6 for the stationary and nonstationary components. The stated result follows directly.

8.9. PROOF OF COROLLARY 5.2: This follows directly from Theorem 5.1 because the submatrix F_2 is null when $r = n$.

8.10. PROOF OF COROLLARY 5.3: When $r = 0$, $F_2 = I_n$ and $A = I_n$ in (24′). We then have the model

$$[I - J^*(L)L]\Delta y_t = \varepsilon_t$$

or

$$\text{(P40)} \quad y_t = y_{t-1} + u_t\,, \quad \text{with} \quad u_t = [I - J^*(L)L]^{-1}\varepsilon_t\,.$$

In this case the subscript "2" that appears in our various formulae, like the limit theory in part (b) of Theorem 5.1 refers to the entire vector u_t or ε_t as the case may be. From (P40), the long-run covariance matrix of u_t is

$$\Omega_{uu} = C\Omega_{\varepsilon\varepsilon}C' = C\Sigma_{\varepsilon\varepsilon}C'\,, \quad C = [I - J^*(1)]^{-1}\,.$$

Let $B_u(r) = BM(\Omega_{uu})$ and $B_\varepsilon(r) = BM(\Omega_{\varepsilon\varepsilon})$ be the limits of the partial sum process $T^{-1/2}\Sigma_1^{[Tr]}u_t$ and $T^{-1/2}\Sigma_1^{[Tr]}\varepsilon_t$. Then, necessarily, $B_u(r) = CB_\varepsilon(r)$ (e.g. see Phillips and Solo (1992)), and

$$\begin{aligned} B_{\varepsilon u}(r) &= B_\varepsilon(r) - \Omega_{\varepsilon u}\Omega_{uu}^{-1}B_u(r) = B_\varepsilon(r) - \Omega_{\varepsilon\varepsilon}C'(C\Omega_{\varepsilon\varepsilon}C')^{-1}CB_\varepsilon(r) \\ &= B_\varepsilon(r) - B_\varepsilon(r) = 0 \quad \text{a.s.} \end{aligned}$$

Hence, the limit distribution given in Theorem 5.1(b) for this case where $r = 0$ is

and thus $T(\hat{F}_2^+ - I_n) \longrightarrow_p 0$.

$$\left(\int_0^1 dB_{\varepsilon u}B_u'\right)\left(\int_0^1 B_u B_u'\right)^{-1} = 0\ , \quad \text{a.s.}$$

8.11. PROOF OF THEOREM 5.5: The error in the levels VAR estimator is $\hat{F}-F = E'X(X'X)^{-1}$. Partitioned regression yields:

$$\sqrt{T}(\hat{F}_1 - F_1) = (T^{-1/2}E'Q_2X_1)(T^{-1}X_1'Q_2X_1)^{-1} \to_d N(0,\ \Sigma_{\varepsilon\varepsilon} \otimes \Sigma_{11}^{-1})$$

giving part (a); and

$$T(\hat{F}_2 - F_2) = (T^{-1}E'Q_1X_2)(T^{-2}X_2'Q_1X_2)^{-1} \to_d \left(\int_0^1 dB_\varepsilon B_2'\right)\left(\int_0^1 B_2B_2'\right)^{-1}\ .$$

Using the decomposition $B_\varepsilon = B_{\varepsilon\cdot 2} + \Omega_{\varepsilon 2}\Omega_{22}^{-1}B_2$ (from Phillips, 1989, Lemma 3.1) we get the stated result for part (b).

8.12. PROOF OF THEOREM 5.7: From (42) and (43) we have

$$\hat{\underline{F}}^+ - \underline{F} = H(\hat{F}^+ - F)(I_k \otimes H')\ .$$

We now partition $\hat{F}^+ - F$ on the right side of this equation as $\hat{F}^+ - F = [\hat{F}_1^+ - F_1 \ \vdots\ \hat{F}_2^+ - F_2]$ with the corresponding partition of $I_k \otimes H'$, viz.

$$I_k \otimes H' = \begin{bmatrix} I_{k-1} \otimes H' & 0 \\ 0 & \beta' \\ \hdashline 0 & \beta_\perp' \end{bmatrix} = \begin{bmatrix} G' \\ \hdashline G_\perp' \end{bmatrix}.$$

Note that

$$\sqrt{T}(\hat{\underline{F}}^+ - \underline{F}) = H[\sqrt{T}(\hat{F}_1^+ - F_1)\ \vdots\ O_p(T^{-1/2})]\begin{bmatrix} G' \\ \hdashline G_\perp' \end{bmatrix} \to_d N(0,\ H\Sigma_{\varepsilon\varepsilon}H' \otimes G\Sigma_{11}^{-1}G')$$

using part (a) of Theorem 5.1. Observing that $\underline{\Sigma}_{\varepsilon\varepsilon} = \Sigma_{\underline{\varepsilon}\underline{\varepsilon}} = H\Sigma_{\varepsilon\varepsilon}H'$ gives the stated result (a), and (a′) follows immediately.

To prove part (b) we write

$$T(\underline{\hat{F}}^+ - \underline{F})G_\perp = H[T(\hat{F}_1^+ - F_1) \,\vdots\, T(\hat{F}_2^+ - F_2)]G'G_\perp$$

$$= H[T(\hat{F}_1^+ - F_1) \,\vdots\, T(\hat{F}_2^+ - F_2)]\begin{bmatrix} 0 \\ 0 \\ I_{n-r} \end{bmatrix} = HT(\hat{F}_2^+ - F_2)$$

$$\to_d H\left(\int_0^1 dB_{\varepsilon\cdot 2}B_2'\right)\left(\int_0^1 B_2B_2'\right)^{-1} = \left(\int_0^1 dB_{\underline{\varepsilon}\cdot 2}B_2'\right)\left(\int_0^1 B_2B_2'\right)^{-1}$$

giving the required result.

8.13. PROOF OF THEOREM 6.1: The proof is essentially the same as the proof of Theorem 3.10. The additional $\chi^2_{q_1 q_J}$ term that appears in the limit (54) of W_F^+ comes from the quadratic form associated with the restrictions $R_1 \underline{J} R_{2J} = \underline{R}_J$ in $\mathscr{H}_0''$ that relate to the known stationary coefficients $\underline{J}$ in the model (42). The remaining components in (54) arise precisely in the same manner as those in Theorem 3.10.

8.14. PROOF OF THEOREM 6.3: When $r = 0$ we have $\underline{F} = [\underline{J} \,\vdots\, \underline{A}] = [F_1 \,\vdots\, F_2]$. From Theorem 5.7 we have

(P41) $\sqrt{T}(\underline{\hat{F}}^+ - \underline{F}) \to_d N(0, \underline{\Sigma}_{\varepsilon\varepsilon} \otimes G_H \Sigma_{11}^{-1} G_H')$

where

$$G_H = \begin{bmatrix} I_{k-1} \otimes H \\ 0 \end{bmatrix} nk \times n(k-1)$$

and

(P42) $T(\underline{\hat{A}}^+ - I_n) \to_p 0$.

Next the test statistic is

$$W_F^+ = T\ \mathrm{tr}\left\{(R_1\hat{\underline{\Sigma}}_{\varepsilon\varepsilon}R_1')^{-1}[R_1(\hat{\underline{F}}^+ - \underline{F})R_2][R_2'T(\underline{X}'\underline{X})^{-1}R_2]^{-1}[R_1(\hat{F}_1^+ - \underline{F})R_2]'\right\}$$

$$\text{(P43)} \quad = \mathrm{tr}\Big\{(R_1\hat{\underline{\Sigma}}_{\varepsilon\varepsilon}R_1')^{-1}[R_1T^{1/2}(\hat{\underline{J}}^+ - \underline{J})R_{2J} \ \vdots\ R_1T(\hat{\underline{A}}^+ - I)R_{2A}]$$

$$[D_TR_2'T(\underline{X}'\underline{X})^{-1}R_2D_T]^{-1}[R_1T^{1/2}(\hat{\underline{J}}^+ - \underline{J})R_{2J} \ \vdots\ R_1T(\hat{\underline{A}}^+ - I)R_{2A}]'\Big\}$$

where $D_T = \mathrm{diag}(I_{q_J}, T^{1/2}I_{q_A})$. Now, writing $X = [\underline{Z}, \underline{Y}_{-1}]$ and performing a partitioned inversion of $(X'X)^{-1}$ we have, in a conventional notation,

$$D_TR_2'T(\underline{X}'\underline{X})^{-1}R_2D_T$$

$$= D_TR_2'\begin{bmatrix} (T^{-1}\underline{Z}'Q_1\underline{Z})^{-1} & -T^{-1}(\underline{Z}'Q_1\underline{Z})^{-1}(T^{-1}Z'\underline{Y}_{-1})(T^{-2}\underline{Y}_{-1}'\underline{Y}_{-1})^{-2} \\ -T^{-1}(T^{-2}\underline{Y}_{-1}'\underline{Y}_{-1})^{-1}(T^{-1}\underline{Y}_{-1}'\underline{Z})(T^{-1}\underline{Z}'Q_1\underline{Z})^{-1} & T^{-1}(T^{-2}\underline{Y}_{-1}'Q_Z\underline{Y}_{-1})^{-1} \end{bmatrix}R_2D_T$$

$$= \begin{bmatrix} R_{2J}'(T^{-1}\underline{Z}'Q_1\underline{Z})^{-1}R_{2J} & O_p(T^{-1/2}) \\ O_p(T^{-1/2}) & R_{2A}'(T^{-2}\underline{Y}_{-1}'Q_{\underline{Z}}\underline{Y}_{-1})^{-1}R_{2A} \end{bmatrix}$$

$$\text{(P44)} \quad = \begin{bmatrix} R_{2J}'(T^{-1}\underline{Z}'\underline{Z})^{-1}R_{2J} & 0 \\ 0 & R_{2A}'(T^{-2}\underline{Y}_{-1}'\underline{Y}_{-1})^{-1}R_{2A} \end{bmatrix} + o_p(1)\ .$$

Inverting (P44) and using the fact that $T(\hat{\underline{A}}^+ - I_n) = o_p(1)$ from (P42), we deduce that

$$W_F^+ = \mathrm{tr}\left\{(R_1\hat{\underline{\Sigma}}_{\varepsilon\varepsilon}R_1')^{-1}[R_1T^{1/2}(\hat{\underline{J}}^+ - \underline{J})R_{2J}][R_{2J}'(T^{-1}\underline{Z}'\underline{Z})^{-1}R_{2J}]^{-1}[R_1T^{1/2}(\hat{\underline{J}}^+ - \underline{J})R_{2J}]'\right\} + o_p(1)\ .$$

Finally, from (P41) we have

$$\sqrt{T}(\hat{\underline{F}}_1^+ - \underline{F}_1) = \sqrt{T}(\hat{\underline{J}}^+ - \underline{J}) \longrightarrow_d N\left(0,\ \underline{\Sigma}_{\varepsilon\varepsilon} \otimes (I_{k-1} \otimes H)\Sigma_{11}^{-1}(I_{k-1} \otimes H)'\right)$$

and

$$T(\underline{Z}'\underline{Z})^{-1} = T\left[(I_{k-1} \otimes H)Z'Z(I_{k-1} \otimes H')\right]^{-1} \longrightarrow_p (I_{k-1} \otimes H)\Sigma_{11}^{-1}(I_{k-1} \otimes H)\ .$$

Thus, $W_F^+ \sim \chi^2_{q_1 q_J}$, as stated.

9. REFERENCES

Andrews, D. W. K. (1991). "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, 59, 817-858.

Cappuccio, N. and D. Lubian (1992). "The relationships among some estimators of the cointegrating coefficient: Theory and Monte Carlo evidence," mimeographed, University of Padova.

Choi, I. (1993). "Asymptotic normality of least squares estimates for higher order autoregressive integrated processes with some applications," *Econometric Theory* (forthcoming).

Engle, R. F. and C. W. J. Granger (1987). "Cointegration and error-correction: Representation, estimation and testing," *Econometrica*, 55, 251-276.

Hannan, E. J. (1970). *Multiple Time Series*. New York: John Wiley & Sons.

Hansen, B. E. and P. C. B. Phillips (1991). "Estimation and inference in models of cointegration: A simulation study," *Advances in Econometrics*, 8, 225-248.

Hargreaves, C. H. (1993). "A review of methods of estimating cointegrating relationships," in C. Hargreaves (ed.), *Nonstationary Time Series Analysis and Cointegrating*. Oxford: Oxford University Press (forthcoming).

Hendry, D. F. (1993). *Econometrics: Alchemy or Science*. Oxford: Blackwell.

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

Kitamura, Y. and P. C. B. Phillips (1992). "Fully modified IV, GIVE and GMM estimation with possibly nonstationary regressors and instruments," Yale University, mimeographed.

Park, J. Y. (1990). "Maximum likelihood estimation of simultaneous cointegrated models," Institute of Economics, Aarhus University, No. 1990-18.

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

Park, J. Y. and P. C. B. Phillips (1988). "Statistical inference in regressions with integrated processes: Part 1," *Econometric Theory*, 4, 468-497.

Park, J. Y. and P. C. B. Phillips (1989). "Statistical inference in regressions with integrated processes: Part 2," *Econometric Theory*, 5, 95-131.

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

Phillips, P. C. B. (1987). "Time series regression with a unit root," *Econometrica*, 55, 277-301.

Phillips, P. C. B. (1988). "Multiple regression with integrated processes," in N. U. Prabhu (ed.), *Statistical Inference from Statistical Processes, Contemporary Mathematics*, 80, 79-106.

Phillips, P. C. B. (1991a). "Optimal inference in cointegrated systems," *Econometrica*, 59, 238-306.

Phillips, P. C. B. (1991b). "Time series regression with cointegrated regressors," mimeographed notes, Yale University.

Phillips, P. C. B. (1991c). "Spectral regression for cointegrated time series," in W. Barnett, J. Powell and G. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Economics and Statistics*. New York: Cambridge University Press.

Phillips, P. C. B. (1992a). "Hyperconsistent estimation of a unit root in time series regression," Cowles Foundation Discussion Paper No. 1040.

Phillips, P. C. B. (1992b). "Simultaneous equations bias in level VAR estimation," *Econometric Theory*, 8, 307. [Solution in Vol. 9, 324-326.]

Phillips, P. C. B. and S. N. Durlauf (1986). "Multiple time series with integrated variables," *Review of Economic Studies*, 53, 473-496.

Phillips, P. C. B. and B. E. Hansen (1990). "Statistical inference in instrumental variables regressions with I(1) processes," *Review of Economic Studies*, 57, 99-125.

Phillips, P. C. B. and M. Loretan (1991). "Estimating long-run economic equilibria," *Review of Economic Studies*, 59, 407-436.

Phillips, P. C. B. and J. Y. Park (1991). "Unidentified components in reduced rank regression estimation of ECM's," Cowles Foundation Discussion Paper No. 1003, Yale University.

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

Priestley, M. B. (1981). *Spectral Analysis and Time Series*, Volumes I and II. New York: Academic Press.

Rau, H-H (1992). "Estimation and inference of linear regression models with cointegrated regressors," Yale University doctoral dissertation.

Sims, C. A., J. H. Stock and M. W. Watson (1990). "Inference in linear time series models with some unit roots," *Econometrica*, 58, 113-144.

Toda, H. and P. C. B. Phillips (1991). "Vector autoregressions and causality," Cowles Foundation Discussion Paper No. 977.

Toda, H. and P. C. B. Phillips (1993). Vector autoregression and causality: A theoretical overview and simulation study," *Econometric Reviews* (forthcoming).

Toda, H. Y. and T. Yamamoto (1993). "Statistical inference in vector autoregressions with possibly integrated processes," mimeographed, Tsukuba University.

Figure 1a: T = 50, α = 0.90

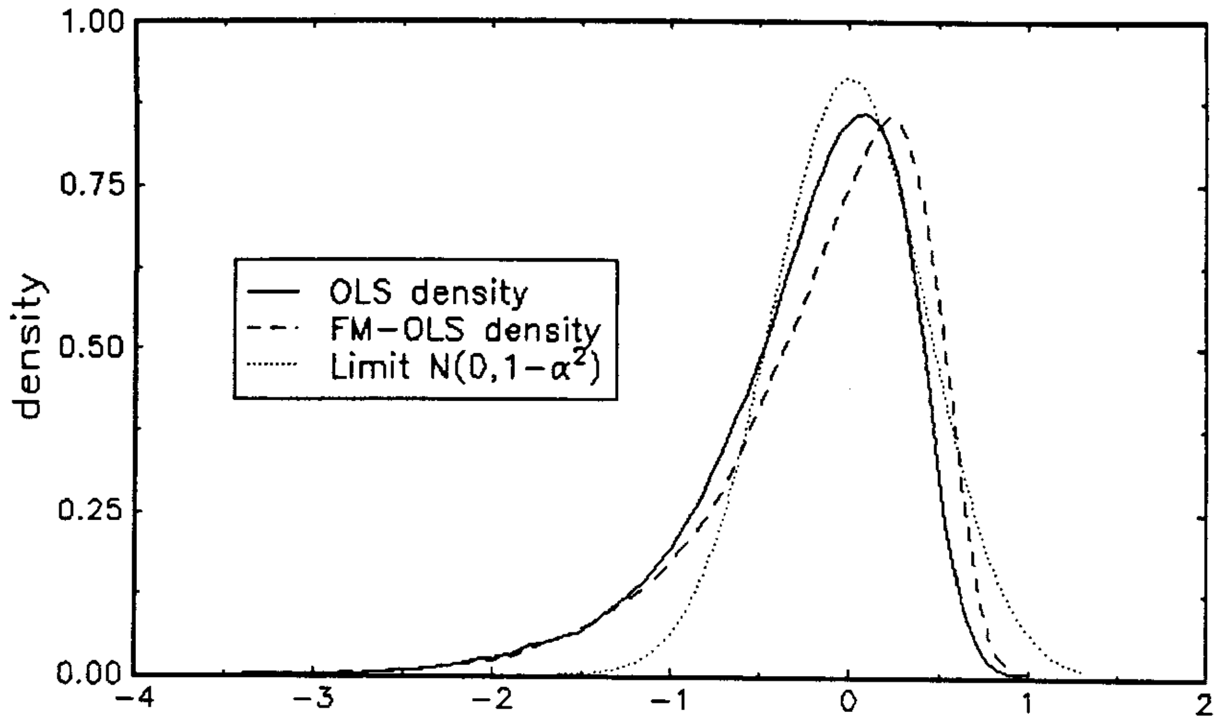

Figure 1b: T = 100, α = 0.90

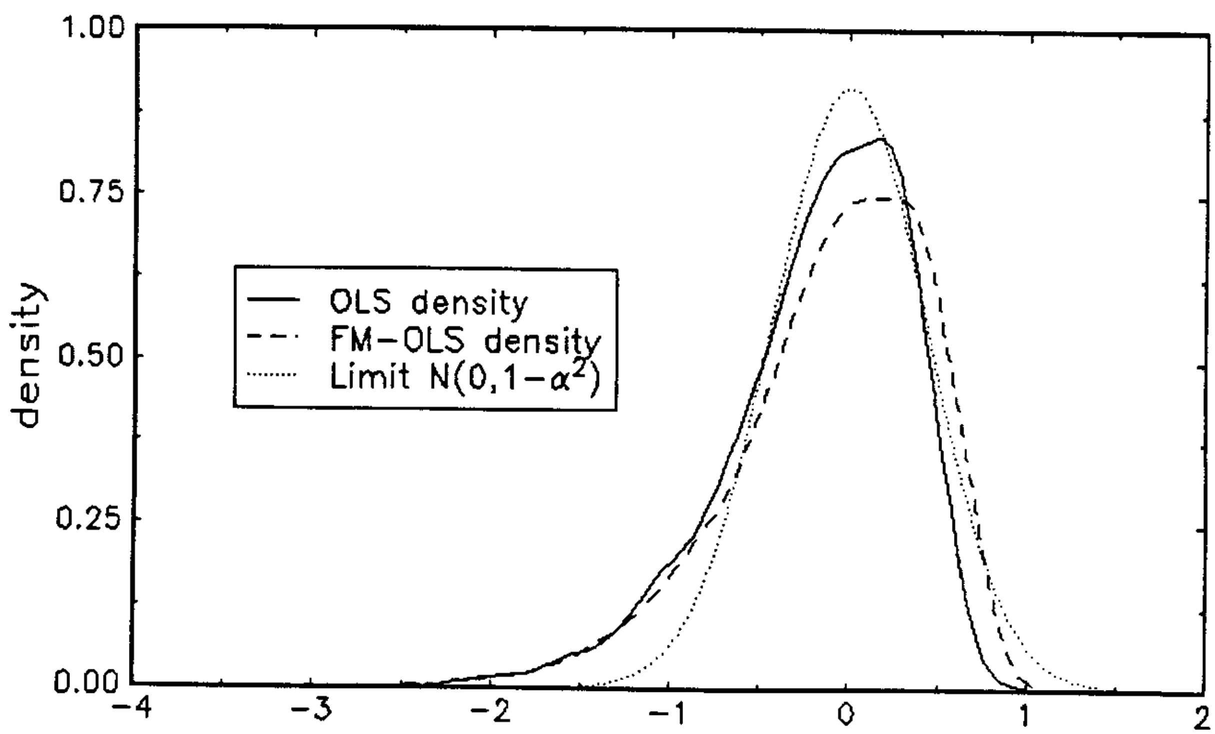

Figure 2a: T = 50, α = 0.80

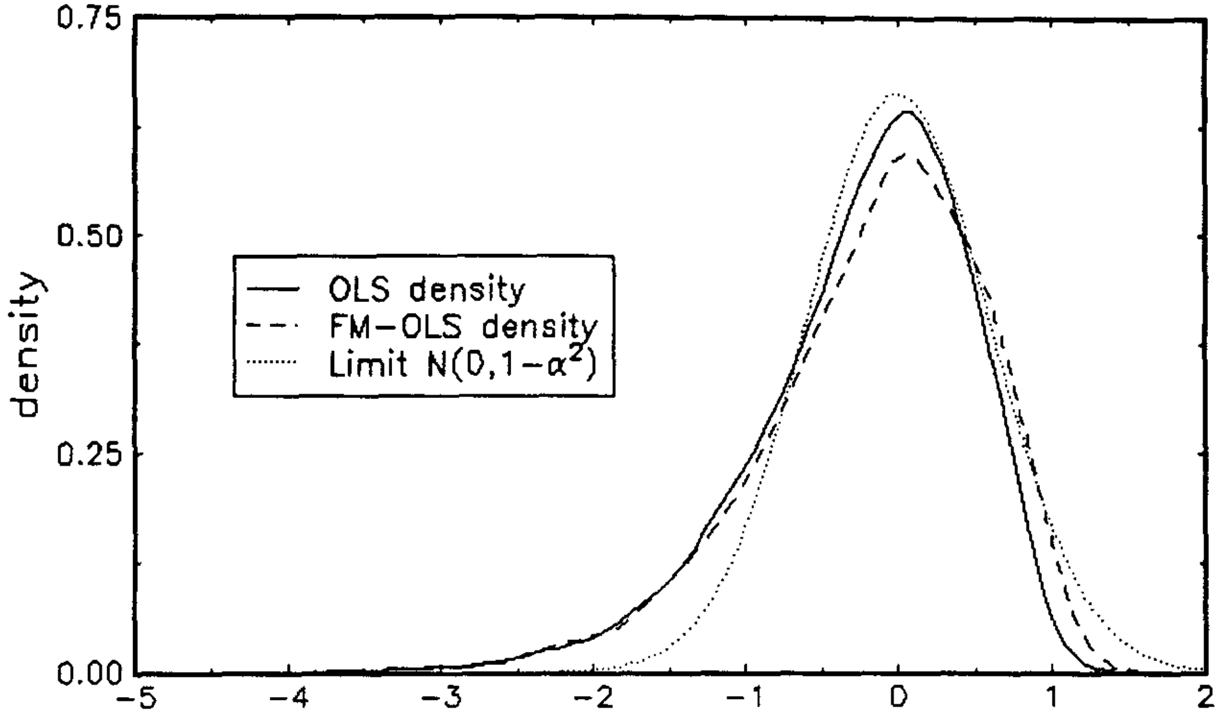

Figure 2b: T = 100, α = 0.80

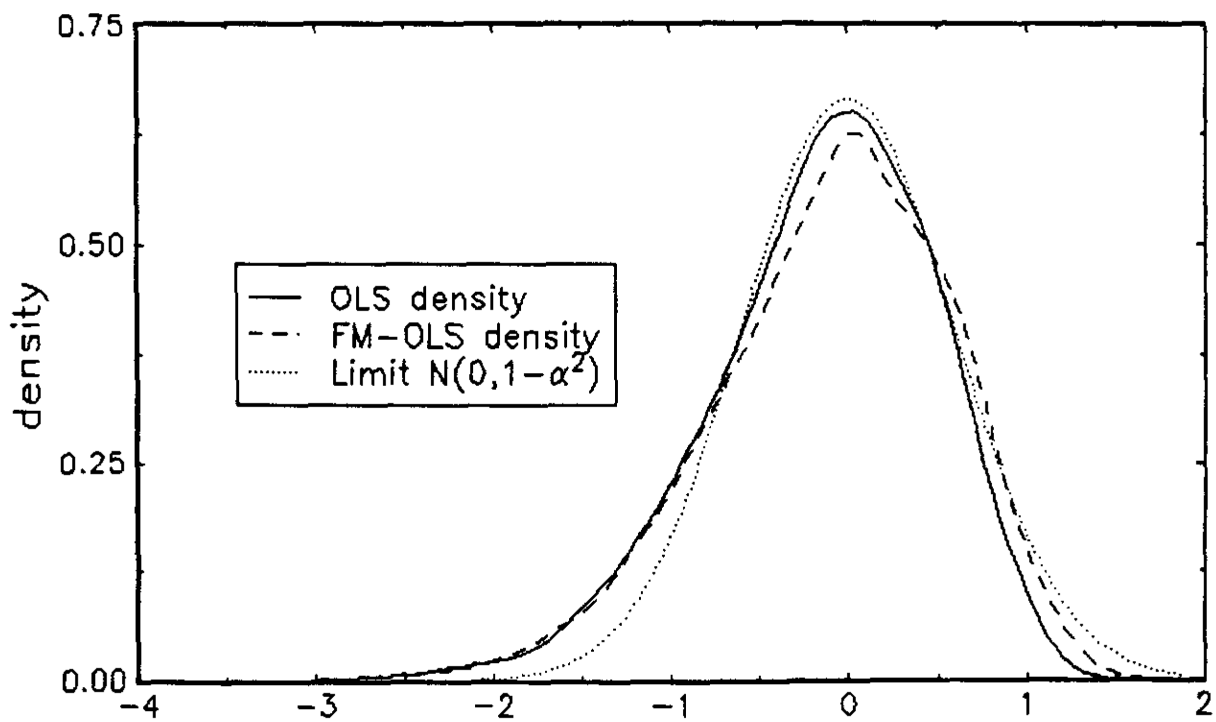

Figure 3a: T = 50, α = 0.40

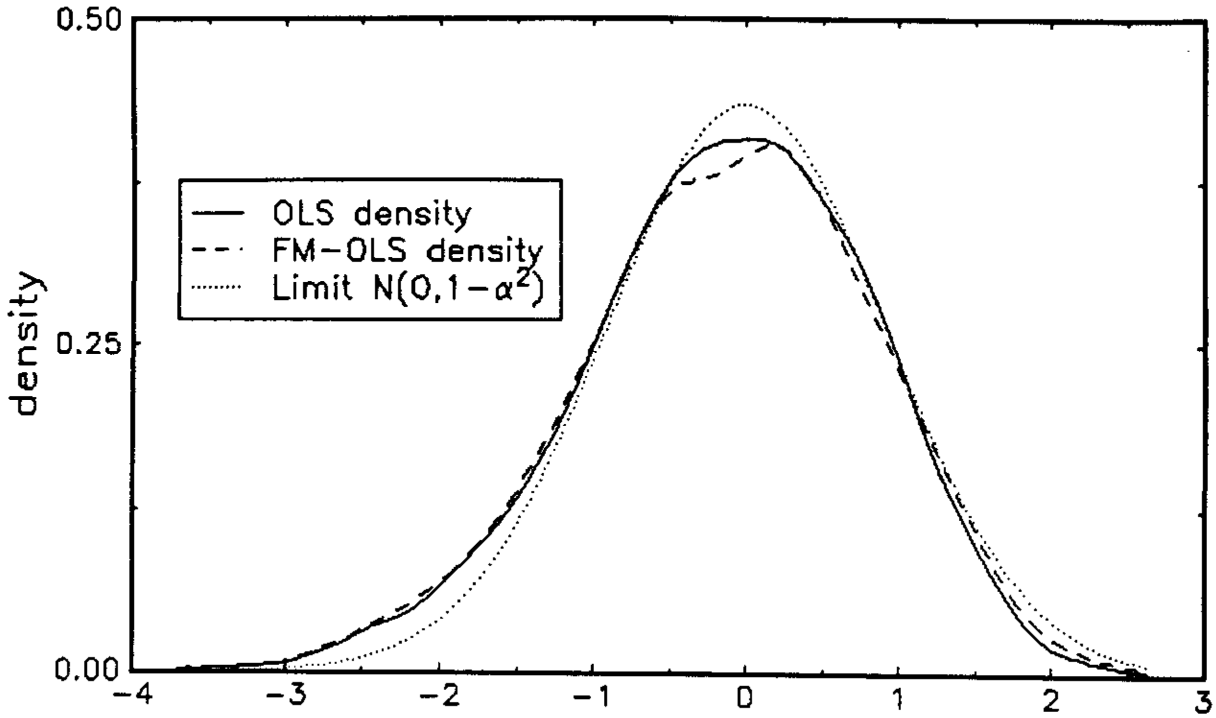

Figure 3b: T = 100, α = 0.40

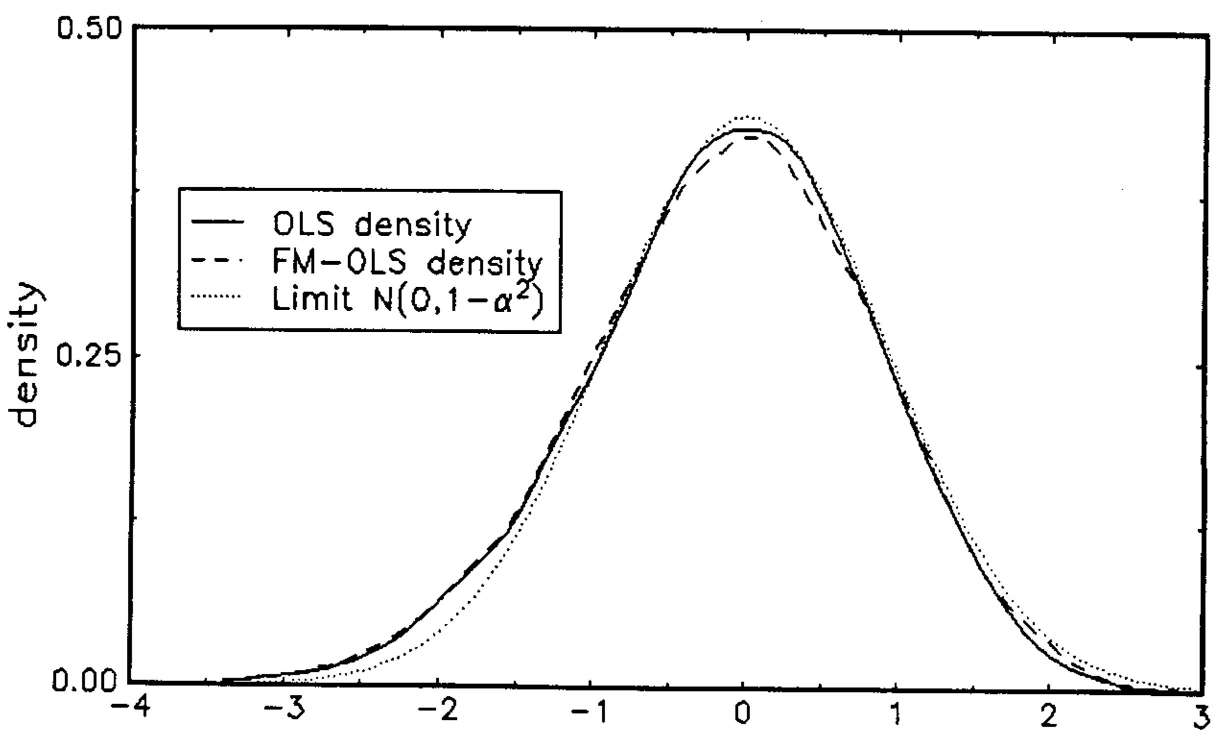

Figure 4a: T = 50, α = 1.00

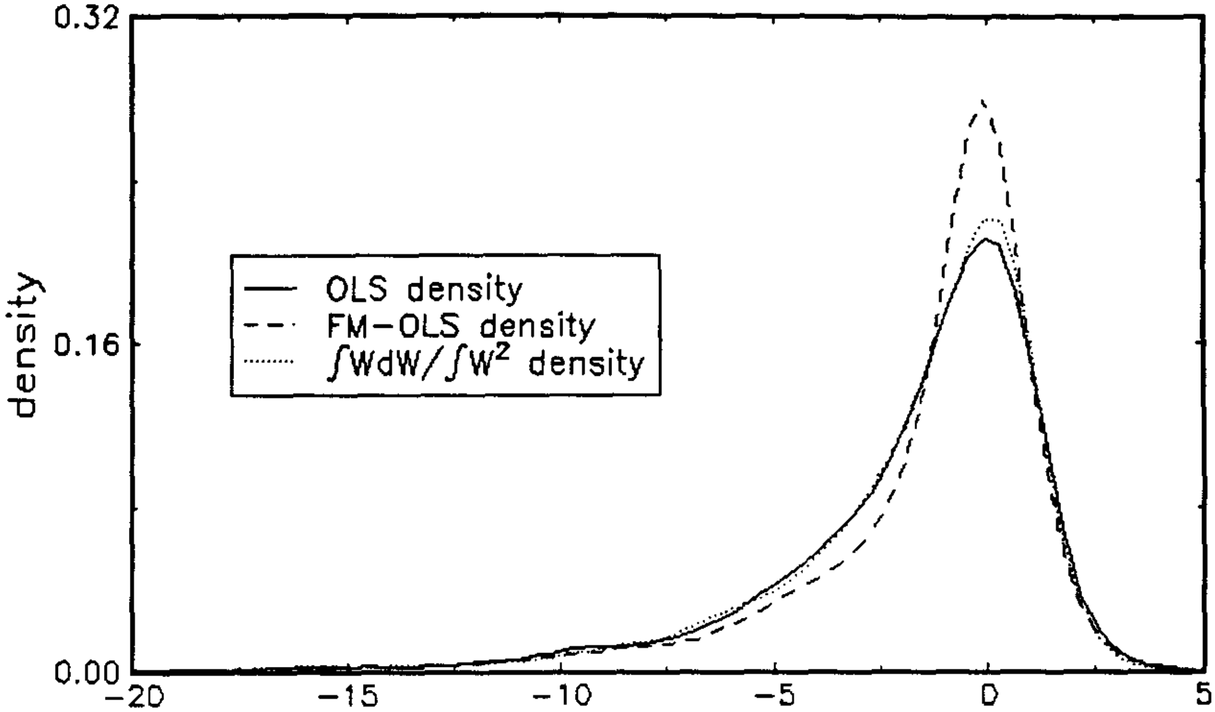

Figure 4b: T = 100, α = 1.00