COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Post Office Box 208281
New Haven, Connecticut 06520–8281

COWLES FOUNDATION DISCUSSION PAPER NO. 1082

FULLY MODIFIED IV, GIVE AND GMM ESTIMATION WITH POSSIBLY NON–STATIONARY REGRESSORS and INSTRUMENTS

by

Yuichi Kitamura and Peter C. B. Phillips

September 1994

FULLY MODIFIED IV, GIVE AND GMM ESTIMATION WITH POSSIBLY NON-STATIONARY REGRESSORS AND INSTRUMENTS[1]

Yuichi Kitamura
Department of Economics
University of Minnesota

Peter C. B. Phillips
Cowles Foundation for Research in Economics
Yale University

First Version: November, 1992
Second Version: February, 1994
Third Version: May, 1994

[1]Kitamura thanks the Sloan Foundation for a 1992/93 Fellowship and Phillips thanks the NSF for support under Grant No. SES 9122142. Both authors thank Glena Ames for her skill and effort in keyboarding the manuscript of the paper.

ABSTRACT

This paper develops a general theory of instrumental variables (IV) estimation that allows for both I(1) and I(0) regressors and instruments. The estimation techniques involve an extension of the fully modified (FM) regression procedure that was introduced in earlier work by Phillips–Hansen (1990). FM versions of the generalized instrumental variable estimation (GIVE) method and the generalized method of moments (GMM) estimator are developed. In models with both stationary and nonstationary components, the FM–GIVE and FM–GMM techniques provide efficiency gains over FM–IV in the estimation of the stationary components of a model that has both stationary and nonstationary regressors. The paper exploits a result of Phillips (1991a) that we can apply FM techniques in models with cointegrated regressors and even in stationary regression models without losing the method's good asymptotic properties. The present paper shows how to take advantage jointly of the good asymptotic properties of FM estimators with respect to the nonstationary elements of a model and the good asymptotic properties of the GIVE and GMM estimators with respect to the stationary components. The theory applies even when there is no prior knowledge of the number of unit roots in the system or the dimension or the location of the cointegration space. An FM extension of the Sargan (1958) test for the validity of the instruments is proposed.

JEL Classification No. 211

Keywords: Cointegration; fully modified least squares; GIVE; GMM; instrument validity; long run covariance; semiparametric correction; unit roots.

1. Introduction

In models with nonstationary regressors it is generally not necessary to utilize GLS-type transformations in order to achieve asymptotically efficient estimation of the nonstationary components. Grenander and Rosenblatt (1957) established an important early result of this type showing, in effect, that OLS is asymptotically equivalent to GLS in detrending a time series that is stationary about a deterministic trend. Phillips and Park (1988) showed that this result continues to apply in the case of trends that are I(1) or stochastically nonstationary in the sense that they have an autoregressive unit root in their generating mechanism. There is a more fundamental way of thinking about this result. In models where the parameters may appear as coefficients of I(1) regressors and as part of the stationary (or I(0)) dynamics, the information matrix is asymptotically block diagonal between the I(1) and the I(0) elements, so that asymptotically optimal estimates of the I(0) parameters are not needed for optimal estimation of the I(1) coefficients. This result was shown in Phillips (1991b) and it is helpful in explaining why semiparametric approaches to estimation are efficient in models where the I(1) coefficients are treated parametrically and the I(0) components nonparametrically.

Most econometric models for time series involve both trending mechanisms and stationary (or transient) dynamic effects often without distinguishing them as separate components. In such cases, there would seem to be scope for the use of some of the estimation methodology that was developed originally for stationary systems, such as generalized instrumental variable estimation (GIVE) and generalized method of moments (GMM) estimation. The GIVE and GMM procedures are well known to deliver efficient estimates in certain stationary regression contexts — see Sargan (1958, 1959, 1988) and Hansen (1982). However, like least squares (OLS) and crude instrumental variables (IV) estimation, these methods do not have good asymptotic properties in nonstationary regression models because estimates of the nonstationary components generally suffer from second order bias problems, as explained in the papers by Phillips (1991b) and Phillips–Hansen (1990). Thus, in order to achieve optimality in the estimation of both the stationary and the nonstationary components in such models, both GIVE and GMM procedures need to be modified to adjust for the presence of nonstationarity. Since the directions in which the data are stationary are not always known *a priori* and often need to be empirically determined, the modifications to GIVE and GMM that will achieve optimality are by no means obvious.

The present paper seeks to explore this problem in detail. For the case of the OLS and crude IV methods, Phillips–Hansen (1990) showed how to construct fully modified (FM) estimators that adjust for nonstationarity and deliver asymptotically optimal estimates of the nonstationary components. The FM–OLS and FM–IV estimators in the Phillips–Hansen paper were developed for the case where the regressors form a set of full rank I(1) processes, i.e. where there are no stationary regressors and there is no cointegration in the regressor space. The present paper deals with a much more general set up. The line of approach we adopt exploits a rather fascinating result discovered recently by one of the authors (Phillips, 1991a) that we can apply the FM procedure in models with cointegrated regressors, I(1) + I(0) regressors and even models with only stationary regressors without losing the method's good asymptotic properties. We develop here FM versions of both the GIVE and the GMM procedures which are applicable in models where the regressors are possibly nonstationary but neither the number nor the location of the unit roots need to be known *a priori*. In addition, we allow for stationary and nonstationary instrumental variables and for the potential of cointegration among the instruments. The resulting estimators, FM–GIVE and FM–GMM, are therefore applicable in rather general regression models with possibly nonstationary regressors and instruments. Under conditions that broadly correspond to those in the literature on the stationary components of the regressors and the instruments, these estimators have asymptotic properties that are analogous to those of conventional GIVE and GMM with respect to the stationary components and yet are also optimal with respect to the nonstationary components. The FM–GIVE and FM–GMM estimators both have a limit theory that is normal for the stationary components and mixed normal for the nonstationary components. This means that conventional asymptotic chi-squared criteria can be used to test hypotheses about the parameters in the model. In addition, we show how to construct an FM version of the Sargan test for the validity of the instruments which is also asymptotically chi-squared. Thus, all of the essential features of the GIVE and GMM procedures which make these estimators appealing in stationary regression models are embodied in our theory for FM–GIVE and FM–GMM. Since this theory allows for stationary and nonstationary regressors and instruments, the framework is broad enough to be useful in much empirical research, while at the same time placing no requirements on the investigator to pretest the data for the presence of unit roots or cointegration.

The paper proceeds as follows. Section 2 gives a preliminary outline of the problem and explains the general idea behind FM estimators. Section 3 details the general model that will concern us, lays out some of the key assumptions and gives a lemma whose results are important in motivating the construction of our estimators. Section 4 develops a general theory for FM–IV estimators that allows for cointegrated regressors and cointegrated instrumental variables. Section 5 shows how to extend this theory to FM–GMM and FM–GIVE estimators. Section 6 gives asymptotic chi-squared tests for the validity of the instruments in GMM and GIVE estimation. Section 7 concludes the paper with a brief summary of our main formulae and results so that these are more accessible to empirical researchers. Derivations and proofs are given in a technical appendix in Section 8. A table of the main notation that we use to distinguish the variables and the estimators in the paper by the various affixes is included in Section 9.

A summary word on notation in the paper which is not explained in the table in Section 9. We use vec(A) to stack the *rows* of a matrix A into a column vector, P_A to signify the projection matrix onto the space spanned by a matrix A, and $[x]$ to denote the smallest integer $\leq x$. We use the symbols "$\xrightarrow{d}$," "$\xrightarrow{p}$," and "$\equiv$" to signify convergence in distribution, convergence in probability, and equality in distribution, respectively. The inequality "> 0" denotes positive definite (p.d.) when applied to matrices. We use $I(d)$ to signify a time series that is integrated of order d, $BM(\Omega)$ to denote a vector Brownian motion with covariance matrix Ω. We write integrals with respect to Lebesgue measure such as $\int_0^1 B(s)ds$ more simply as $\int_0^1 B$ to achieve notational economy. The symbolism $MN(0,V)$ signifies the mixture normal distribution $MN(0,V) \equiv \int_{V>0} N(0,V)dP(V)$. Finally, all limits given in this paper are taken as the sample size T tends to ∞ unless otherwise stated.

2. Some Preliminary Discussion of the Problem

In this section we present some informal arguments that use a simple model to illustrate the problems discussed in the paper. We consider the regression

$$y_t = \beta' x_t + u_{0t} \tag{1}$$

where $\{u_{0t}\}$ is a stationary time series, and $\{x_t\}$ is a vector time series which is either I(1) or I(0). In either case, we allow for endogeneity in the regressors: when x_t is I(0), some elements of x_t can be correlated with u_{0t}, and when x_t is I(1), some elements of $\Delta x_t = u_{2t}$ can be correlated with

u_{0s} for some s. For the time being, in the I(1) case we assume that x_t is a full rank I(1) process, i.e. the number of unit roots in the stochastic process x_t is equal to the dimension of x_t (and thus the elements of x_t are not cointegrated). When y_t and x_t are I(1) equation (1) is usually called a cointegrating regression.

In the I(0) case, the use of OLS generally yields an inconsistent estimator of β, and the instrumental variable (IV) method is commonly employed to deal with this problem. In order to apply IV methods successfully, we need valid instruments, and we can test the validity of the instruments that we use by following the approach of Sargan (1958, 1959). On the other hand, nowadays it is well known that OLS estimators are T-consistent in cointegrating regressions, though they do involve nuisance parameters and are not asymptotically unbiased. In the I(1) case, as in Phillips and Loretan (1991), the asymptotic distribution of the OLS estimator of β (denoted by $\widehat{\beta}$, say) is given by

$$T(\widehat{\beta}-\beta) \xrightarrow{d} \left(\int_0^1 B_1 B_1'\right)^{-1}\left(\left[\int_0^1 B_1 d(B_{0\cdot 1}+\omega_{01}\Omega_{11}^{-1}B_1)\right]+\delta\right), \tag{2}$$

where $\delta = \Sigma_{k=0}^{\infty}E(u_{0k}u_{10}), (B_0,\, B_1')' = BM(\Omega)$, $B_{0\cdot 1} = B_0 - \omega_{01}\Omega_{11}^{-1}B_1$ and Ω is the "long run variance" matrix of $u_t = (u_{0t},\, u_{1t}')'$ and is partitioned conformably. Observe that the second term in the parentheses involving the coefficient $\omega_{01}\Omega_{11}^{-1}$ and the term involving δ both induce bias, asymmetry and nuisance parameters (i.e. Ω_{11}, ω_{01}, δ) into the limit distribution.

Several ways have been proposed to resolve these problems: see Johansen (1988), Park (1992), Phillips (1991b,c), Phillips and Hansen (1990), Phillips and Loretan (1991), Saikkonen (1991) and Stock and Watson (1992). Among them, the fully modified (FM) estimator proposed by Phillips and Hansen seems to be particularly useful in practice because it enables investigators to run regressions much like least squares that yield asymptotically efficient estimates of the cointegrating coefficients. The procedure eliminates nuisance parameters in the following way. First, we modify y_t using the transformation $\widehat{y}_t^+ = y_t - \widehat{\omega}_{01}\widehat{\Omega}_{11}^{-1}\Delta x_t$ and the error in (1) also, giving $\widehat{u}_{0t}^+ = u_{0t} - \widehat{\omega}_{01}\widehat{\Omega}_{11}^{-1}\Delta x_t$. This is a correction for endogeneity. Next we construct a serial correlation correction term $\widehat{\delta}^+$, which is a consistent estimator of $\delta^+ = \Sigma_{k=0}^{\infty}E(u_{0k}^+ u_{01}')$ where $u_{0t}^+ = u_{0t} - \omega_{01}\Omega_{11}^{-1}\Delta x_t$. The FM estimator combines these two corrections in the least squares regression formula and is given by

$$\overline{\beta} = (\Sigma_1^T x_t x_t')^{-1}(\Sigma_1^T x_t \widehat{y}_t^+ - T\widehat{\delta}^+),$$

which is asymptotically-median unbiased and nuisance parameter free.

Now, what if we allow for cointegration in the regressor variables x_t? This means that there are some stationary components in x_t, and therefore, a natural strategy might be to use IV estimators for the stationary components and FM estimators for the I(1) components. If the cointegrating vectors for x_t were known, or the location of the unit roots were specified *a priori*, the stationary components and the I(1) components would be identified and the above strategy would clearly work. However, such vectors are usually unknown and need to be determined empirically unless prior economic knowledge is sharp and very informative. Moreover, a simple and important example in practice is the case where we do not know whether some individual regressors are either I(1) or I(0). If some of the regressors are I(0), we sometimes say that the regressors as a whole are "trivially cointegrated," since any vector which puts non-zero weights on the I(0) components and zero weights on the I(1) components is a cointegrating vector. In the following, we explore a methodology that allows us to deal with systems of this type that have possibly nonstationary processes without using prior information about the location of unit roots or even the full dimension of the cointegration space.

3. The Model, Conditions and a Useful Lemma

Let $\{y_t\}$ be an n-dimensional time series generated by

$$y_t = Ax_t + u_{0t}, \tag{3}$$

where A is an $n \times m$ coefficient matrix and x_t is an $m = (m_1 + m_2)$-dimensional vector of cointegrated regressors specified as follows:

$$H_1' x_t = x_{1t} = u_{1t} \; : \; m_1 \times 1 \tag{4a}$$

$$H_2' \Delta x_t = \Delta x_{2t} = u_{2t} \; : \; m_2 \times 1 \tag{4b}$$

where $H = (H_1, H_2)$ is an $m \times m$ orthogonal matrix. Using the rotations prescribed in (4), equation (3) can be rewritten as

$$y_t = A_1 x_{1t} + A_2 x_{2t} + u_{0t} \tag{3'}$$

where $A_1 = AH_1$ and $A_2 = AH_2$. We let z_t denote a q-vector of instruments driven by

$$G_1' z_t = z_{1t} = u_{z1t} \ : \ q_1 \times 1 \tag{5a}$$

$$G_2' z_t = \Delta z_{2t} = u_{z2t} \ : \ q_2 \times 1. \tag{5b}$$

We will use the notation $u_{xt} = \Delta x_t$ and $u_{zt} = \Delta z_t$. This partition of the regressors and the instruments will be instructive in the development of our theory. However, as will become clear neither our methods nor our results are contingent on the knowledge of or the nature of these partitions.

We now impose assumptions on the random variables $w_t = (u_{0t}', u_{1t}', u_{2t}', u_{z1}', u_{z2}')'$ that drive the system (3)–(4) and the instrument set z_t.

ASSUMPTION EC *(Error Condition)*

(a) $\{w_t\}_1^\infty$ *is fourth order stationary with absolutely summable fourth cumulant function;*

(b) $E(w_t) = 0$;

(c) $E|w_{it}|^\beta < \infty$ $(i = 1, ..., n + m + q)$ *for some* $4 \leq \beta < \infty$;

(d) $\{w_t\}_1^\infty$ *is either* φ*-mixing with mixing coefficients* φ_m *such that* $\Sigma_{m=1}^\infty \varphi_m^{1-1/\beta} < \infty$ *or* α*-mixing with mixing coefficients* α_m *such that* $\Sigma_{m=1}^\infty \alpha_m^{1-2/\beta} < \infty$;

(e) *The long run variance matrix of* w_t, $\Omega = \Sigma_{-\infty}^{+\infty} E(w_{t+i} w_t')$ $(= \Sigma_{-\infty}^{+\infty} \Gamma(i)$, *say) is positive definite.*

Assumptions EC(a), EC(b) and EC(c) imply that the regressor x_t is cointegrated and that each column of H_1 is its cointegrating vector. Assumption EC(e) also ensures that z_{2t} is I(1), but it excludes cointegration among the elements of z_{2t} and between x_{2t} and z_{2t}. It also excludes the possibility of "multicointegration" of y_t and x_{2t} as defined by Granger and Lee (1989). The assumption of no cointegration between x_{2t} and z_{2t} will be relaxed later on in the paper. For subsequent use, we decompose the long run covariance matrix given in (e) as follows

$$\Omega = \Sigma + \Lambda + \Lambda',$$

where $\Sigma = E(w_t w_t')$ and $\Lambda = \Sigma_{i=1}^{\infty} E(w_{t+i} w_t') = \Sigma_{i=1}^{\infty} \Gamma(i)$; and we define the "one sided long run covariance matrix"

$$\Delta = \Sigma + \Lambda = \Sigma_{i=0}^{\infty} E(w_{t+i} w_t') = \Sigma_{i=0}^{\infty} \Gamma(i).$$

Under Assumption EC, a multivariate invariance principle (IP) for $\{w_t\}$ holds, viz.

$$T^{-1/2} \sum_{i=1}^{[Tr]} w_i \xrightarrow{d} B(r) \equiv BM(\Omega), \ 0 \le r \le 1, \tag{6}$$

as shown in Phillips and Durlauf (1986). We partition B and Ω conformably with w_t as

$$B(r) = \begin{bmatrix} B_0(r) \\ B_1(r) \\ B_2(r) \\ B_{z1}(r) \\ B_{z2}(r) \end{bmatrix}, \ \Omega = \begin{bmatrix} \Omega_{00} & \Omega_{01} & \Omega_{02} & \Omega_{0z_1} & \Omega_{0z_2} \\ \Omega_{10} & \Omega_{11} & \Omega_{12} & \Omega_{1z_1} & \Omega_{1z_2} \\ \Omega_{20} & \Omega_{21} & \Omega_{22} & \Omega_{2z_1} & \Omega_{2z_2} \\ \Omega_{z_10} & \Omega_{z_11} & \Omega_{z_12} & \Omega_{z_1z_1} & \Omega_{z_1z_2} \\ \Omega_{z_20} & \Omega_{z_21} & \Omega_{z_22} & \Omega_{z_2z_1} & \Omega_{z_2z_2} \end{bmatrix},$$

and define the nq_1-vector

$$\phi_{z_1 t} = u_{0t} \otimes z_1. \tag{7}$$

We now state some additional conditions that are important for the analysis of the stationary components of the model.

ASSUMPTION IV *(Instrument Validity Conditions)*

(a) $E\phi_{z_1 t} = E(u_{0t} \otimes z_1) = 0$ *for all* t *[orthogonality condition];*

(b) $E[x_{1t} z_{1t}'] = K_{z_1}$ *is of full row rank (rank* m_1*) [relevance condition];*

(c) $E[z_{1t} z_{1t}'] = M_{z_1}$ *is nonsingular [nonsingular second moment];*

(d) $\{\phi_{z_1}\}_1^{\infty}$ *satisfies the same conditions as Assumption* EC(b), (c), (d), (e) *[regularity conditions];*

(e) $m_2 \le q_2$ *[order condition on* I(1) *instruments].*

In conventional IV estimation, we choose instruments satisfying (a) whose dimension is equal to or greater than the dimension of the regressors. In our case, we wish to maintain $m_1 \le q_1$, which is a necessary condition for part (b), but m_1 is unknown a priori. Therefore, in the above model specification the set of instruments is required to be "large enough" so that $m \leq q$ and the necessary order condition in terms of dimension is satisfied. Part (d) (given part (a)) allows for the use of a central limit theorem (CLT) with respect to $\phi_{z_1 t}$. Other sets of conditions, are possible in place of (d), of course, and are explored elsewhere, e.g. in White (1984). Part (e) is a nonstationary counterpart of part (b). Note that it suffices to impose an "order condition" here, since the sample moment matrix $T^{-2}\Sigma x_{2t}z_{2t}'$ converges in distribution to a random matrix that is of full rank almost surely as long as both x_{2t} and z_{2t} carry full rank stochastic trends. This point arises from the asymptotic theory of spurious regression and has been shown by Phillips and Hansen (1990, Lemma A3).

Define the data matrices $U_0 = [u_{01}, ..., u_{0T}]'$, $X = [x_1, ..., x_T]'$, and $Z = [z_1, ..., z_T]'$. Similarly, we write $XH = [X_1, X_2] = X_H$ and $ZG = [Z_1, Z_2] = Z_G$, where the subscripts "H" and "G" signify that rotations by H and G have been performed. We also define the $nq \times nq$ matrix $S_{z_1} = \Sigma_{i=-\infty}^{+\infty} R_{z_1}(i)$, where

$$R_{z_1}(j) = E(u_{0t}u_{0t+j}' \otimes z_{1t}z_{1j+t}').$$

Then, under mild regularity conditions such as Assumption IV (a) and (d), we have the central limit theorem (CLT)

$$T^{-1/2}U_0'Z_1 \xrightarrow{d} N(0, S_{z_1}). \tag{8}$$

Next define a sequence of $n \times m$ random matrices $\{C_T\}$, which will be used to illustrate some properties that are common to all IV estimators in this paper, by

$$C_T = (U_0'P_z - \Psi_T)X(X'P_zX)^{-1}. \tag{9}$$

The matrix C_T represents a generic form of the matrix of IV estimation errors for the parameter A. In (9), Ψ_T is an $n \times T$ random matrix of abstract correction terms. It is convenient for us now to impose the following conditions on the asymptotic behavior of Ψ_T and later we will justify them under more primitive conditions.

CONDITION CT *(Correction Term Conditions)*

(C1) $\Psi_T X_1 = o_p(\sqrt{T})$

(C2) $T^{-1}\Psi_T X_2 \xrightarrow{d} \Psi_2$, say *where* Ψ_2 *may be a random matrix.*

The following lemma is fundamentally important in our subsequent theory.

LEMMA 3.1: *Suppose Assumptions* EC, IV *and Condition* CT *hold. Then*

(a) $\sqrt{T}C_T H_1 = \sqrt{T}U_0' P_{z_1} X_1 (X_1' P_{z_1} X_1)^{-1} + o_p(1)$
$$\xrightarrow{d} N(0, J_{z_1} S_{z_1} J_{z_1}'),$$

(b) $TC_T H_2 = T(U_0' P_{z_2} - \Psi_T) X_2 (X_2' P_{z_2} X_2)^{-1} + o_p(1)$
$$\xrightarrow{d} \left(\int_0^1 dB_0 \widetilde{B}_2' + \Delta_{0z_2} \left(\int_0^1 B_{z_2} B_{z_2}' \right)^{-1} \int_0^1 B_{z_2} B_2' - \Psi_2 \right) \left(\int_0^1 \widetilde{B}_2 \widetilde{B}_2' \right)^{-1},$$

where $J_{z_1} = [I_n \otimes (K_{z_1} M_{z_1}^{-1} K_{z_1}')^{-1} K_{z_1} M_{z_1}^{-1}]$, K_{z_1} *and* M_{z_1} *are as defined in Assumption* IV (b) *and* (c) *and* $\widetilde{B}_2 = \int_0^1 B_2 B_{z_2}' (\int_0^1 B_{z_2} B_{z_2}')^{-1} B_{z_2}$.

Part (a) of Lemma 3.1 shows we can add any correction terms that satisfy (C1) and (C2) without affecting the asymptotic behavior of $C_T H_1$. Moreover, $C_T H_1$ is asymptotically equivalent to the estimation error that obtains when we apply the conventional IV estimator to a stationary regression model with the regressor vector x_{1t} and the instrumental variable vector z_{1t}. Part (b) of Lemma 3.1 shows that $C_T H_2$ has the usual asymptotics of a cointegrating IV regression with the additional term Ψ_2. In sum, if we construct a correction term Ψ_T so that it satisfies (C1) and (C2) and yields a limit matrix Ψ_2 that correctly adjusts the asymptotics in part (b), then the limit behavior of C_T and its various functionals may become nuisance parameter free and have some other good properties like asymptotic median unbiasedness and possibly even optimality. In fact, the FM–IV estimator and its variants that are proposed in this paper are designed so that their correction terms satisfy the conditions just mentioned. The lemma is helpful in understanding the key elements in and the motivation behind the construction of these estimators. We will use it frequently in the analysis that follows.

4. Estimation Theory

This section studies the estimation of the model proposed in Section 3, allowing for the regressors to be cointegrated and to be correlated with the errors. Cointegration among the instruments is also allowed for. In the construction of the estimator, we use the vector of instruments z_t, consisting of both stationary and nonstationary components.

The following formula defines the FM–IV estimator of the coefficient matrix A in (3)

$$\begin{aligned}\widetilde{A} &= (\widehat{Y}^{+\prime}Z - T\widehat{\Delta}_{0z}^{+})(Z'Z)^{-1}Z'X(X'P_zX)^{-1} \\ &= [Y'P_z - \widehat{\Omega}_{0a}\widehat{\Omega}_{aa}^{-1}U_a'P_z - T\widehat{\Delta}_{0z}^{+}(Z'Z)^{-1}Z']X(X'P_zX)^{-1} \end{aligned} \tag{10}$$

where $\widehat{Y}^{+} = Y' - \widehat{\Omega}_{0a}\widehat{\Omega}_{aa}^{-1}U_a'$, $\widehat{\Delta}_{0z}^{+} = \widehat{\Delta}_{0z} - \widehat{\Omega}_{0a}\widehat{\Omega}_{aa}^{-1}\widehat{\Delta}_{az}$. $\widehat{\Delta}_{0z}$ denotes the estimate of the one-sided long-run covariance between u_0 and u_z. We use the subscript "a" in these formula to signify elements that correspond to u_{xt} and u_{z_t} taken together. Note that in the definition (10), the second term in the square bracket is the correction term for the endogeneity of the nonstationary instrument z_2 and the regressor x_t, while the third term corrects for serial correlation.

Before studying the asymptotics of the FM–IV estimator, we will prove two lemmas which are useful in evaluating the asymptotic contribution of the correction terms in our estimators. In these lemmas, the long run covariance matrices can be estimated by the use of kernel estimators or smoothed periodogram estimators. Kernel estimators of the long run covariance and one-sided long run covariance matrices between $\{u_{at}\}$ and $\{u_{bt}\}$ take the following forms:

$$\widehat{\Omega}_{ab} = \sum_{j=-T+1}^{T-1} w(j/K)\widehat{\Gamma}_{u_a u_b}(j),\ \widehat{\Delta}_{u_a u_b} = \sum_{j=0}^{T} w(j/K)\widehat{\Gamma}_{u_a u_b}(j), \tag{11}$$

where $w(\cdot)$ is a kernel function, $\widehat{\Gamma}_{u_a u_b}(j) = T^{-1}\Sigma u_{at+j}u_{bt}'$ and K is a lag truncation or bandwidth parameter satisfying $K = o(T^{1/2})$ as $T \to \infty$. In some cases (e.g. for the quadratic spectral estimator) the kernel function $w(\cdot)$ is nonzero outside the interval $[-1,1]$ and then there is no truncation in the summation in (11). It will be helpful to be specific at this stage about the class of kernel functions that we will include in our theory. The class is given by the following:

CONDITION KL *(Kernel Condition): The kernel function* $w(\cdot) : \mathbb{R} \to [-1,1]$ *is such that* $w(x) = w(-x)$, $w(0) = 1$, $w(x)$ *is continuous at zero and continuously differentiable at all but a finite number of values of* $x \in \mathbb{R}$, $\int_{\mathbb{R}} |w(x)|dx < \infty$, *and* $\int_{\mathbb{R}} w(x)e^{ix\lambda} \geq 0$ *for all* $\lambda \in \mathbb{R}$.

Suppose $r(>0)$ is the largest integer such that

$$\lim_{u\to 0}\frac{1-w(u)}{|u|^r}<\infty. \tag{12}$$

This implies that

$$\lim_{u\to 0}\frac{dw(u)/du}{u^{r-1}}=w_{(r)}<\infty. \tag{13}$$

In fact, r is what Parzen (1957) calls the characteristic exponent of the kernel $w(\cdot)$ using the expression (12). For our purposes expression (13) turns out to be the more useful. Several well known kernels satisfy Condition KL. We will be concerned mainly with kernels whose characteristic exponent $r=2$. Among these we have the following (noting that the Tukey–Hanning does not satisfy the positivity requirement of Condition KL, which is desirable but not essential):

$$\text{Parzen}:\quad w(x)=\begin{cases}1-6x^2+6|x|^3 & \text{for } 0\le |x|\le 1/2\\ 2(1-|x|)^3 & \text{for } 1/2\le |x|\le 1\\ 0 & \text{otherwise}\end{cases}$$

$$\text{Tukey-Hanning}:\quad w(x)=\begin{cases}(1+\cos(\pi x))/2 & \text{for } |x|\le 1\\ 0 & \text{otherwise}\end{cases}$$

$$\text{Quadratic Spectral}:\quad w(x)=\frac{25}{12\pi^2x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5}-\cos(6\pi x/5)\right).$$

In practice we need to estimate the unknown sequence $\{u_{0t}\}$ to construct estimators of long run covariance matrices such as $\widehat{\Omega}_{0a}$. Conventional IV estimators and residuals defined by

$$\widehat{A}=Y'P_zX(X'P_zX)^{-1},\ \text{and}\ \widehat{u}_t=y_t-\widehat{A}x_t \tag{14}$$

can be used for this purpose. $\widehat{A}$ is consistent for A under Assumptions EC and IV, since the estimator $\widehat{A}_1=\widehat{A}H_1$ is $\sqrt{T}$-consistent for A_1 and the estimator $\widehat{A}_2=\widehat{A}H_2$ is T-consistent for A_2. It is straightforward to justify these consistency results using Lemma 3.1, since $\Psi_T=0$ in this "naive" IV regression. In finite samples there may, of course, be some advantage to using a third stage FM–IV estimator in which the estimates of long run covariance matrices like Ω_{0a} are refined by using the residuals from the second state FM–IV regression to estimate u_{0t}. This is a matter that will be explored in subsequent simulation experiments with our methods.

Finally, we assume

ASSUMPTION LR *(Long Run Covariance Matrix Estimation): Any of the Parzen, Tukey–Hanning or the Quadratic Spectral* (QS) *kernel estimators are used in the estimation of the long run covariance matrices. The covariance functions* $\Gamma_{u_0 u_\beta}(\cdot)$ *and* $\Gamma_{u_b u_\beta}(\cdot)$ *satisfy the summability condition*

$$\Sigma_{j=-\infty}^{\infty} j \|\Gamma(j)\| < \infty \tag{15}$$

where $u_b = (u_{2t}', u_{z2t}')'$ *and* $u_\beta = (u_{1t}', u_{z1t}')'$. *The parameter* K *in* (11) *grows at the rate of* T^k *for some* $k \in (1/4, 1/2)$.

The kernel estimators specified in Assumption LR are all commonly used in long run covariance matrix estimation. The summability condition (15) allows for a wide range of time series including quite general finite order stationary vector ARMA specifications for the error processes. Under stationary ARMA specifications, of course, $\Gamma(j)$ decays exponentially and (15) is automatic.

Now we postulate an additional condition concerning the unknown stationary component $\{z_{1t}\}$:

ASSUMPTION NF *(No Feedback):* $E[u_{0t+j} \otimes z_{1t}] = 0$ *for all* $j \geq 1$.

Assumption NF does not seem restrictive in empirical applications, since it holds in two situations where conventional instrumental variable methods are most frequently used. First, the assumption is trivially satisfied when all the instruments are stationary and strictly exogenous. Second, the hypothesis of rational expectations will usually entail that there is no feedback from the regressors or instruments to the errors. In typical rational expectations models components of stationary variables involving past information are orthogonal to current errors, and this fact provides the opportunity for instrumental variables estimation of rational expectation models. Research along these lines was initiated by McCallum (1979) and extended by Hansen and Singleton (1982) and many others, particularly to the estimation of rational expectations models with future expectations. In the case of nonstationary models, the rational expectations assumption imposes restrictions on the stationary linear combinations of nonstationary variables, i.e. the cointegrated variables, as pointed out in Hansen and Sargent (1991). In our model, the stationary linear combinations of nonstationary instrumental variables are denoted by z_{1t}, and their past values must be orthogonal to the current stationary error u_{0t}. This implies Assumption NF.

Define $u_{ht} = (\Delta u_{1t}', u_{2t}')' \; (= \Delta x_{ht} = H' u_{xt})$ and $z_{gt} = (\Delta z_{1t}', z_{2t}')' \; (= \Delta z_{Gt} = G' u_{zt})$, where we use the subscript "h" and "g" to denote elements corresponding to $\{\Delta u_{1t}\}$ and $\{u_{2t}\}$ ($\{\Delta u_{z1t}\}$ and $\{u_{z2t}\}$) taken together. The following lemma describes the asymptotic behavior of the component elements of the FM–IV estimator.

LEMMA 4.1: *Under Assumptions* EC, IV *and* LR

(a) $\widehat{\Omega}_{0h}\widehat{\Omega}_{hh}^{-1}\left(T^{-1}U_h'Z_G - \widehat{\Delta}_{hg}\right) = [o_p(1/\sqrt{T}) \vdots \Omega_{02}\Omega_{22}^{-1}N_T + o_p(1)]$,

(b) $(Z_G'Z_G)^{-1}Z_G'X_1 = \begin{bmatrix} (Z_1'Z_1)^{-1}Z'X_1 \\ O_p\left(\frac{1}{T}\right) \end{bmatrix}$, $(Z_G'Z_G)^{-1}Z_G'X_2 = \begin{bmatrix} O_p(1) \\ M_T \end{bmatrix}$,

where $N_T \xrightarrow{d} \int_0^1 dB_2B_{z_2}'$, *and* $M_T \xrightarrow{d} \left(\int_0^1 B_{z_2}B_{z_2}'\right)^{-1}\int_0^1 B_{z_2}B_{z_2}'$.

LEMMA 4.2: *Under Assumptions* EC, IV, LR *and* NF,

(a) $\widehat{\Delta}_{0g}(Z_G'Z_G)^{-1}Z_G'X_1 = T^{-1}\widehat{U}_0'P_{z_1}X_1 + o_p(1/\sqrt{T}) = o_p(1/\sqrt{T})$, $\Delta_{0g}(Z_G'Z_G)^{-1}Z_G'X_2 \xrightarrow{d} \Delta_{0z_2}\left(\int_0^1 B_{z_2}B_{z_2}'\right)^{-1}\int_0^1 B_{z_2}B_2'$,

(b) $\widehat{\Omega}_{0h}\widehat{\Omega}_{hh}^{-1}\left(\frac{1}{T}\Delta X_H'P_{z_G} - \widehat{\Delta}_{hg}(Z_G'Z_G)^{-1}Z_G'\right)X_1 = o_p(1/\sqrt{T})$, *and*

(c) $\widehat{\Omega}_{0h}\widehat{\Omega}_{hh}^{-1}\left(\frac{1}{T}\Delta X_H'P_{z_G} - \widehat{\Delta}_{hg}(Z_G'Z_G)^{-1}Z_G'\right)X_2 \xrightarrow{d} \Omega_{02}\Omega_{22}^{-1}\int_0^1 dB_2B_{z_2}'\left(\int_0^1 B_{z_2}B_{z_2}'\right)^{-1}\int_0^1 B_{z_2}B_2'$.

With these results in hand we now turn to study the asymptotic behavior of the FM–IV estimator $\widetilde{A}$. We rotate coordinates in $\mathbb{R}^m$ by the orthogonal matrix H that was introduced in Section 3 so that we can analyze the component matrices $\widetilde{A}_1 = \widetilde{A}H_1$ and $\widetilde{A}_2 = \widetilde{A}H_2$ separately. The asymptotic behavior of these two components is quite different as the following theorem shows.

THEOREM 4.3: *Under Assumptions* EC, IV *and* LR,

(a) $\sqrt{T}(\widetilde{A} - A)H_1 \xrightarrow{d} N(0,\, J_{z_1}S_{z_1}J_{z_1}')$

(b) $T(\widetilde{A} - A)H_2 \xrightarrow{d} MN(0,\, \Omega_{00\cdot b} \otimes \left(\int_0^1 \widetilde{B}_2\widetilde{B}_2'\right))$.

Remarks

(a) In the statement of Theorem 4.3 we use the following notation for limit processes that are adjusted for their conditional means. For the partitioned limit process $B = (B_1', B_2')$ we define the process $B_{1\cdot 2} = B_1 - \Omega_{12}\Omega_{22}^{-1}B_2 \equiv BM(\Omega_{11\cdot 2})$, which is independent of the Brownian motion B_2. We use the subscript "b" to signify elements corresponding to u_{2t} and $u_{z_{2t}}$ jointly. Note that $\widetilde{B}_2$, which appears in part (b) and which was defined in Lemma 3.1, is the projection in $L_2[0,1]^{m_2}$ of B_2 onto the subspace spanned by the elements of $I_{m_2} \otimes B_{z_2}'$.

(b) Theorem 4.3 shows some of the advantages of the FM–IV estimation procedure. The estimator $\widetilde{A}$ is $\sqrt{T}$-consistent and its limit distribution is normal in the direction of H_1, as a result of the use of valid instruments. At the same time, in the direction of H_2 the estimator is T-consistent and its limit distribution is mixed normal, symmetric and median unbiased, with nuisance parameters (other than scale) being eliminated by the FM correction terms.

Another interesting and practically important situation which violates Assumption EC (e) is one in which the I(1) instruments and the regressors are cointegrated. This case happens, for example, when the set of nonstationary instrumental variables includes lagged values of regressors. Such instruments are commonly used in instrumental variables estimation and in the estimation of rational expectation models. Regressors and instruments that are found by lagging regressors are naturally cointegrated if the regressors are stochastically nonstationary. Fortunately, this case can be treated without any changes in the above definitions and only involves a minor change in the asymptotic properties of the estimators. To illustrate, suppose that the I(1) processes $\{x_{2t}\}$ and $\{z_{2t}\}$ are jointly driven by the following cointegrated system:

$$F_1' \begin{pmatrix} x_{2t} \\ z_{2t} \end{pmatrix} = u_{c_{1t}}, \; F_2'\Delta \begin{pmatrix} x_{2t} \\ z_{2t} \end{pmatrix} = u_{c_{2t}},$$

where $\{u_{c_{1t}}\}$ and $\{u_{c_{2t}}\}$ are ℓ_1- and ℓ_2-dimensional and $F = (F_1, F_2)$ is an $\ell \times \ell$ orthogonal matrix with $\ell = \ell_1 + \ell_2 = m_2 + q_2$. We continue to require that Assumption EC holds with $w_t = (u_{0t}', u_{1t}', u_{2t}', u_{1t}', u_{2t}')'$ now replaced by $w_t = (u_{0t}', u_{1t}', u_{z_1}', u_{c_{1t}}', u_{c_{2t}}')'$. This assumption implies that each column of F_1 is a cointegrating vector of $\{(x_{2t}', z_{2t}')'\}$.

With these adjustments, part (a) of Theorem 4.1 remains valid without any changes, while part (b) holds with subscripts "b" replaced by "c_2". The latter result is a direct consequence of Lemma 4.2, though we need three coordinate rotations to achieve it; rotation by H in $\mathbb{R}^m$ to decompose $\{x_t\}$ (into $\{x_{1t}\}$ and $\{x_{2t}\}$), rotation by G in $\mathbb{R}^q$ to decompose $\{z_t\}$ (into $\{z_{1t}\}$ and $\{z_{2t}\}$), and rotation by F in $\mathbb{R}^\ell$ to decompose $\{(x_{2t}', z_{2t}')'\}$ into its I(0) and I(1) components.

5. Efficient Estimation

In the analysis of systems with cointegrated regressors in the previous section, we have shown that the FM–IV (and FM–IV/CI) estimators of the nonstationary components of the model are asymptotically median unbiased and the limit distributions are nuisance parameter free (up to scale), as a result of the "fully modified regression" methodology. However, as far as the stationary components of the model are concerned, the FM–IV procedure proposed above uses the standard IV estimation method. So there is the potential of efficiency gain with respect to the stationary components, for example by the use of a GLS-type transformation. GLS-type transformations have not been a popular tool in the recent literature of nonstationary time series analysis, since in general the effect of a GLS-type transformation asymptotically vanishes and no efficiency gain is to be expected, as shown in Phillips and Park (1988). In our model, however, both stationary and nonstationary components are included in the regressors and they are not identified *a priori.* Thus, it seems worthwhile applying GLS-type transformations to the whole model including its nonstationary components to see if an efficiency gain is realized for the coefficients of the stationary components.

In the following we suggest the use of two well known approaches to obtain an efficiency gain by data transformations. The first corresponds to the GMM procedure with optimal choice of the "distance matrix" proposed by Hansen (1982) for nonlinear estimation problems. The second is the GIVE procedure, originally proposed by Sargan (1958, 1959). Also, following Bowden and Turkington (1984), one may call the former the "IV–OLS analog" and the latter the "IV–GLS analog." The former is valid under fairly general assumptions upon the instruments, such as those that are implied by usual rational expectations (RE) models with predetermined but not exogenous instruments. The latter method can be relatively efficient over the former asymptotically, as in the case where the instruments are strictly exogenous. The limit theory of estimators of the nonstationary components of the model is not affected by either transformation.

5.1. FM–GMM (FM–IV–OLS Analog) Estimator

Here by the term "GMM" we mean a linear version of the GMM estimator with an "optimal" choice of the distance matrix. However, unlike conventional GMM, we need to deal with nonstationarity, both in the regressors and the instruments. For exposition of this case, we will use the same model as that considered in Section 4.1, i.e. the model specified as (3)–(5), with Assumptions EC and IV. To simplify our presentation in what follows we use capital script letters to represent the Kronecker products of the $n \times n$ identity matrix I_n with matrices of observations. For example, we use $\mathcal{X} = (I_n \otimes X)$, $\mathcal{Z} = (I_n \otimes Z)$, and so on.

We define the FM–GMM estimator $\tilde{A}_{\mathrm{GMM}}$ as follows:

$$\mathrm{vec}\ \tilde{A}_{\mathrm{GMM}} = \left(\mathcal{X}'\mathcal{Z}S_{zT}^{-1}\mathcal{Z}'\mathcal{X}\right)^{-1}\mathcal{X}'\mathcal{Z}S_{zT}^{-1}\ \mathrm{vec}(\check{Y}^{+\prime}Z - T\check{\Delta}_{0z}^{+}). \tag{16}$$

where the distance matrix S_{zT} (rotated by G) is partitioned as

$$G'S_{zT}G = \begin{pmatrix} S_{z_1T} & S_{z_1z_2T} \\ S_{z_2z_1T} & S_{z_2T} \end{pmatrix} \begin{matrix} nq_1 \\ nq_2 \end{matrix},$$

and each block must satisfy the following conditions

$$S_{z_1T} \xrightarrow{p} S_{z_1} = \Sigma_{i=-\infty}^{+\infty} R_{z_1}(i)\ , \tag{17a}$$

$$S_{z_1z_2T} = (S_{z_2z_1T})' = O_p(1),\ T^{-1}S_{z_2T} \xrightarrow{p} \Omega_{00} \otimes \textstyle\int_0^1 B_{z_2}B_{z_2}'. \tag{17b}$$

The notation used here is analogous in form to our earlier notation. However, the use of the affix " ˇ " in place of " ˆ " indicates that the estimate of the unknown process $\{u_{0t}\}$ is not obtained through a naive IV regression, but is instead the GMM residual $\hat{u}_{0t\mathrm{GMM}} = y_t - \hat{A}_{\mathrm{GMM}}x_t$, where

$$\mathrm{vec}\ \hat{A}_{\mathrm{GMM}} = \left(\mathcal{X}'\mathcal{Z}S_{zT}^{-1}\mathcal{Z}'\mathcal{X}\right)^{-1}\mathcal{X}'\mathcal{Z}S_{zT}^{-1}\ \mathrm{vec}(Y'Z). \tag{18}$$

In the literature, several techniques to obtain the optimal distance matrix for the GMM estimators have been proposed, and we can use these in the FM–GMM procedure. The first method uses the spectral estimator

$$S_{zT} = \frac{1}{2M}\sum_{k=-M+1}^{M-1}\hat{f}_{u_0u_0}\left(\frac{\pi k}{M}\right) \otimes \hat{f}_{zz}\left(-\frac{\pi k}{M}\right), \tag{19}$$

where $M = o(T^{1/2})$ as $T \to \infty$. In the formula (19) the spectral density estimates are of the form

$$\hat{f}_{ab}(\lambda) = \frac{1}{2\pi}\sum_{j=-T+1}^{T-1} w(j/M)\hat{\Gamma}_{ab}(j)e^{ij\lambda},$$

where the sample covariance matrix is

$$\widehat{\Gamma}_{ab}(j) = \frac{1}{T}\sum_{t=1}^{T-j} a_{t+j}b_t', \ \widehat{\Gamma}_{ab}(-j) = \widehat{\Gamma}_{ba}'(j), \ 1 \le j \le T,$$

the bandwidth parameter K is as before in (11), and the lag window $w(\cdot)$ satisfies Condition KL. By following the arguments in Hannan (1963) and Phillips (1991) we can show that the spectral estimator (19) satisfies (17a) and (17b), respectively. A second method is to estimate a VAR model for the error process $\{u_{0t}\}$ and use it to construct an estimator of the long run covariance matrix of $\{u_{0t}\}$. This approach will be pursued later in the section on the "FM–GIVE" estimator. In either method, we can substitute the estimated process $\{\widehat{u}_{0t}\}$ in (11) for the unobserved sequence $\{u_{0t}\}$, without affecting the asymptotic behavior of these estimators.

The following lemma describes the asymptotic behavior of the stationary part of the correction term in (18)

LEMMA 5.1: *Under Assumptions* EC, IV, LR *and* NF

$$\mathcal{X}_1'\mathcal{Z}S_{zT}^{-1}\ \mathrm{vec}(\check{\Delta}_{0z}) = T^{-1}\mathcal{X}_1'\mathcal{Z}_1S_{z_1}^{-1}\ \mathrm{vec}(\widehat{U}_{0\mathrm{GMM}}'Z_1) + o_p(1/\sqrt{T}) = o_p(1/\sqrt{T}).$$

Our next result gives the limit theory for the FM–GMM estimator:

THEOREM 5.2: *Under Assumptions* EC, IV, LR *and* NF

$$\sqrt{T}(\widetilde{A}_{\mathrm{GMM}} - A)H_1 \xrightarrow{d} N\left(0,\ [\mathcal{K}_{z_1}S_{z_1}^{-1}\mathcal{K}_{z_1}']^{-1}\right),$$

where $\mathcal{K}_{z_1} = [I_n \otimes K_{z_1}]$. *Further,* $\widetilde{A}_{\mathrm{GMM}}H_2$ *is asymptotically equivalent to* $\widetilde{A}H_2$, *which is the* FM–IV *estimator of* $A_2 = AH_2$.

If we compare these results with those of the FM–IV estimator given in Theorem 4.3, the advantage of the FM–GMM estimator should be clear. For the coefficient of the stationary components of the model, we obtain an efficiency gain in estimation as a result of the "optimal" choice of distance matrix. This follows from the well known inequality $[\mathcal{K}_{z_1}S_{z_1}^{-1}\mathcal{K}_{z_1}']^{-1} \le J_{z_1}S_{z_1}J_{z_1}'$ between the asymptotic covariance matrix of the two IV estimators. As far as the nonstationary components are concerned, the two estimators of these coefficients are asymptotically equivalent, because the effects of the GMM transformations of the integrated processes cancel out, just as the effects of GLS transformations cancel out in regressions with full rank integrated processes (as shown in Phillips and Park, 1988).

5.2. FM–GIVE (FM–IV–GLS Analog) Estimator

The FM–GMM estimator considered in the two last subsections is designed to incorporate an asymptotically optimal choice of the "distance matrix." Hence, we obtain an asymptotic efficiency gain over the FM–IV estimators of Section 4, at least with respect to the stationary components of the model. In the literature on IV estimation, there is extensive discussion of the choice of optimal instruments in the stationary time series context, and the generalized instrumental variable estimator (GIVE) was proposed as another approach — see Sargan (1988, Ch. 5.4) for a recent treatment. Roughly speaking, the GIVE procedure employs a GLS-type transformation to correct the data (including the instruments) for serial dependence in the equation errors. Some further efficiency gains (potentially even over GMM) may be obtained, though some additional assumptions are needed in order to justify the transformations. In the following, we show that efficient estimation of the stationary components of a possibly cointegrated nonstationary model can be achieved by the use of a fully modified version of the GIVE procedure. We shall assume strict exogeneity of the instruments in our development in this paper but in later work we will give an extension of the GIVE methodology that allows for the same set up as we have used in our GMM analysis. As one might expect from our earlier theory on IV and GMM, estimators of the nonstationary components are shown to be asymptotically invariant to the GLS transformations that underlie the GIVE procedure.

We will employ a parametric GLS transformation here, though it is probably worth pointing out that a nonparametric treatment is possible by the use of a corresponding technique in the frequency domain. (See Corbae, Ouliaris and Phillips (1991) for the form of the frequency domain GIVE estimator and an application to nonstationary time series.) In this subsection and the next, we assume that the error term $\{u_{0t}\}$ is generated by a p-th order vector autoregression (VAR).

ASSUMPTION VR: *The stochastic process* $\{u_{0t}\}$ *is generated by the* VAR(p) *model*

$$u_{0t} = -\Sigma_{r=1}^{p} C_r u_{0t-r} + \varepsilon_t$$

where $\varepsilon_t \equiv \text{iid}(0, \Sigma_\varepsilon)$. *If* $C(L) = I_n + \Sigma_{r=1}^{p} C_r L^r$, *where* L *denotes the backshift operator, then the* *roots of* $|C(L)| = 0$ *are greater than one in absolute value.*

The model can be rewritten in matrix form as

$$U_0' = -\Sigma_{r=1}^{p} C_r U_{0_{-r}}' + E',$$

where $U_{0_{-r}}'$ and E' denote observation matrices of u_{0t-r} and ε_t, respectively. Now set the first r rows of $U_{0_{-r}}'$ to be vectors of zeros (i.e., the initial values are ignored) and define the $T \times T$ matrix

$$\mathcal{L} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & 1 \\ 0 & 0 & & \cdots & 0 \end{bmatrix},$$

which is similar to the circulant matrix, but has its $(T, 1)$ element zero, not unity. Letting $C_0 = I_n$, we have $\Sigma_{r=0}^{p} C_r U_0' \mathcal{L}_T^r = E'$, or

$$\left(\Sigma_{r=0}^{p}(C_r \otimes \mathcal{L}_T^{r'})\right) \text{vec}(U_0') = \text{vec}(E').$$

Thus we have

$$W_T \, \text{vec}(U_0') \sim N(0, I_{nT}), \text{ where}$$

$$W_T = (\Sigma_\varepsilon^{-1/2} \otimes I_T)\left(\Sigma_{r=0}^{p}(C_r \otimes \mathcal{L}_T^{r'})\right) = \Sigma_{r=0}^{p} C_r^\varepsilon \otimes \mathcal{L}_T^{r'} \text{ and } C_r^\varepsilon = \Sigma_\varepsilon^{-1/2} C_r.$$

Therefore, the nonsphericality in the model is removed by the premultiplication of the stacked observation matrices by W_T. For later use, we let V_{0T} denote the $nT \times nT$ covariance matrix of $\text{vec}(U_0')$. (Then $W_T' W_T = V_{0T}^{-1}$.) In practice, we need estimates of W_T to achieve this GLS transform. To do so we first estimate $\{u_{0t}\}$ using some $\sqrt{T}$-consistent estimates of A such as the naive IV estimator $\widehat{A}$ (see (11)). Then we estimate a p-th order VAR process by OLS using $\{\widehat{u}_{0T}\}$ and plug the resulting estimates $\widehat{C}_r^\varepsilon = \widehat{\Sigma}_\varepsilon^{-1/2} \widehat{C}_r$ in the above definition of W_T, giving $\widehat{W}_T$. In the following, we use the affix "*" to indicate premultiplication by $\widehat{W}_T$, e.g. $\mathcal{A}^* = \widehat{W}_T \mathcal{A} = \widehat{W}_T(I_n \otimes A)$ and $\text{vec}(A)^* = \widehat{W}_T \, \text{vec}(A)$.

In FM–IV estimation and FM–GMM estimation, one of the key requirements for the consistency of those methods is the (contemporaneous) orthogonality of $\{u_{0t}\}$ and $\{z_{1t}\}$, viz. Assumption IV(a). This assumption holds, for instance, in rational expectation models where the instruments are predetermined but not necessarily exogenous, given a suitable choice of instruments. In the case FM–GIVE estimation, the orthogonality condition $E[C(L)u_{0t} \otimes C(L)z_{1t}] = 0$ is especially convenient. As is well known, this does not hold for predetermined (but not exogenous) instruments. However, it does hold for strictly exogenous instruments. Therefore we assume,

ASSUMPTION SE *(Strict Exogeneity): $\{z_t\}$ is strictly exogenous.*

This is a strong version of Assumption NF. As explained at the beginning of this subsection, this assumption is stronger than necessary. In order to obtain consistency and the asymptotically normal and mixed normal results, only the strict exogeneity of the *stationary* instruments $\{z_{1t}\}$ is needed. In fact, even this assumption may be relaxed to allow for lagged dependent variables as in Sargan (1988). We will continue to assume Assumption SE in this paper for the following reasons: (i) without the exogeneity of $\{z_t\}$, the definition of the FM–GIVE is much more involved, and (ii) Assumption SE is the most convenient one in the case where we will consider cointegrated instruments. Extension of our approach to accommodate more general assumptions than SE will be included in later work.

For the model (3), we define the FM–GIVE estimator as follows:

$$\begin{aligned} \text{vec}\, \widetilde{A}_{\text{GIVE}} &= (\mathcal{X}^{*\prime} P_{\mathcal{Z}}^{*} \mathcal{X}^{*})^{-1} \\ &\quad \times \mathcal{X}^{*\prime}\mathcal{Z}^{*} \left\{ (\mathcal{Z}^{*\prime}\mathcal{Z}^{*})^{-1}\mathcal{Z}^{*\prime}\text{vec}(Y')^{*} - \text{vec}[\widehat{\Omega}_{0x}\widehat{\Omega}_{xx}^{-1}(U_x'Z - T\widehat{\Delta}_{xz})(Z'Z)^{-1}] \right\}. \end{aligned} \tag{20}$$

It is possible to define the FM–GIVE estimator in other forms that are asymptotically equivalent and possibly computationally easier. We use the above definition chiefly for conceptual convenience. Note that unlike the instruments, the regressors and the dependent variables, the correction terms in the FM–GIVE formula (given by the expression inside square brackets in (20)) are not transformed, since these terms are designed to persist asymptotically only in the estimator of the coefficients $A_2 = AH_2$ corresponding to nonstationary components. As stated before, GLS transformations usually cancel out as scale effects in the estimation of nonstationary components, and there is, therefore, no need to transform the correction terms. As far as feasible GLS estimation is concerned, any consistent estimates of $\{u_{0t}\}$ may be used as in the case of FM–GMM. The following theorem gives the limit theory for the FM–GIVE estimator.

THEOREM 5.3: *Under Assumptions* EC, IV, LR, VR *and* SE,

$$\sqrt{T}(\tilde{A}_{\text{GIVE}} - A)H_1 \xrightarrow{d} N\left(0,\ \left[\mathcal{K}^v_{z_1}\mathcal{M}^{v-1}_{z_1}\mathcal{K}^{v\prime}_{z_1}\right]^{-1}\right),$$

where $\mathcal{K}^v_{z_1} = \text{plim}(T^{-1}\mathcal{K}^{*\prime}_1\mathcal{Z}^*_1) = \text{plim}(T^{-1}\mathcal{X}_1 V_{0T}^{-1}\mathcal{Z}_1)$, *and* $\mathcal{M}^v_{z_1} = \text{plim}(T^{-1}\mathcal{Z}^*_1\mathcal{Z}^*_1)$ $= \text{plim}(T^{-1}\mathcal{Z}'_1 V_{0T}^{-1}\mathcal{Z}_1)$. *Further,* $\tilde{A}_{\text{GIVE}}H_2$ *is asymptotically equivalent to* $\tilde{A}H_2$, *the* FM–IV *estimator of* $A_2 = AH_2$.

Potentially, the FM-GIVE estimator of $A_1 = AH_1$ can be more efficient than the FM–GMM estimator. In the conventional setting, namely systems in which only stationary variables appear, this is already known. White (1984), for instance, gives a detailed argument about the asymptotic relative efficiency of IV–GLS type estimators over IV–OLS type estimators and provides some sufficient conditions. In our case, if the instrumental variables $\{z_{1t}\}$ appear as variables in the reduced form equations of $\{x_t\}$, we can show the asymptotic relative efficiency of FM–GIVE over FM–GMM with respect to the estimation of A_1. The demonstration is essentially the same as that for the linear simultaneous equations model with serially dependent errors.

6. IV Validity Tests for Overidentifying Restrictions

It is well known that in IV estimation with stationary processes a test for the validity of instruments that was originally proposed by Sargan (1958, 1959) is available when the total number of orthogonality conditions exceeds the total number of unknown coefficients. Hansen (1982) extended this test to the case of the GMM estimator. The test is also known as a test of overidentifying restrictions or as an IV misspecification test. In what follows we will extend this IV testing principle to the FM–IV procedure and its various generalizations that we have studied earlier in this paper.

We strengthen Assumption NF slightly to accommodate a limit theory of our statistics. Let $\eta_t = (u'_{0t-j}, u'_{z1t+1}, u'_{z2t+1})$, $j \geq 0$ and $\mathcal{F}_t = \sigma(\eta_t, \eta_{t-1}, ...)$ be the σ-algebra generated by $\{\eta_s\}_{s=-\infty}^{t}$.

ASSUMPTION NF[2]: $(u_{0t}, \mathcal{F}_t)$ *is a martingale difference sequence.*

This assumption is stronger than Assumption NF. However, the inclusion of u_{z2t} in the orthogonality conditions probably makes little difference in practice, since if we assume rational expectation models as we referred to in Section 4.2, it is implied that both $\{u_{z1t}\}$ and $\{u_{z2t}\}$ are in the current information set and orthogonal to the future prediction errors. Of course if we assume the strict exogeneity of $\{z_t\}$, Assumption NF^2 holds trivially. In fact, under Assumption NF^2 the FM–IV (or –GMM) estimators need not be corrected for the serial correlation between $\{u_{0t}\}$ and $\{u_{z2t}\}$, thus we can remove $\widehat{\Delta}_{0z}$ and $\check{\Delta}_{0z}$ from the definitions of these estimators. Accordingly, the following replacements are possible:

$$\widehat{\Delta}^{+}_{0z} = \widehat{\Delta}_{0z} - \widehat{\Omega}_{0a}\widehat{\Omega}^{-1}_{aa}\widehat{\Delta}_{az} \rightarrow -\widehat{\Omega}_{0a}\widehat{\Omega}^{-1}_{aa}\widehat{\Delta}_{az} \text{ [in (10)]} \tag{21a}$$

$$\check{\Delta}^{+}_{0z} = \check{\Delta}_{0z} - \widehat{\Omega}_{0a}\widehat{\Omega}^{-1}_{aa}\widehat{\Delta}_{az} \rightarrow -\widehat{\Omega}_{0a}\widehat{\Omega}^{-1}_{aa}\widehat{\Delta}_{az} \text{ [in (16)].} \tag{21b}$$

First, we will consider the instrument validity test for the FM–GMM estimator. The model is taken to be the same one used in Sections 4 and 5, where the instruments are assumed to be not cointegrated among themselves. Using the definition of the FM–GMM estimator $\widetilde{A}_{\text{GMM}}$, and the residual $\{\widetilde{u}_{0\text{GMM}t}\}$ defined as

$$\widetilde{u}_{0\text{GMM}t} = y - \widetilde{A}_{\text{GMM}}x_t,$$

we now define $\{\widetilde{u}^{+}_{0t}\}$, which is corrected for endogeneity with respect to $u_{at} = (u'_{xt}, u'_{zt})'$:

$$\widetilde{u}^{+}_{0\text{GMM}t} = \widetilde{u}_{0\text{GMM}t} - \widehat{\Omega}_{0a}\widehat{\Omega}^{-1}_{aa}u_{at}.$$

(In the calculation of $\widehat{\Omega}_{0a}$, $\{\widetilde{u}_{0\text{GMM}t}\}$ may be used.) Now we define an $n \times q$ (unstandardized) score matrix Ξ

$$\Xi = \widetilde{U}'_{0\text{GMM}}Z.$$

Using the corrected residuals $\{\widetilde{u}^{+}_{0\text{GMM}t}\}$ defined above, we define the score with the endogeneity correction as

$$\Xi^{+} = \widetilde{U}^{+\prime}_{0\text{GMM}}Z.$$

Next, in parallel with fully modified estimation, we correct Ξ^{+} for serial correction terms as well, giving the "fully modified" score,

$$\Xi^{+*} = \Xi^{+} - T\check{\Delta}^{+}_{0z}. \tag{22}$$

and where $\widehat{\Delta}^{+}_{0z} = \widehat{\Delta}_{0z} - \widehat{\Omega}_{0a}\widehat{\Omega}^{-1}_{aa}\widehat{\Delta}_{az}$ as earlier.

LEMMA 6.1: *Suppose Assumptions* EC, IV, NF^2 *and* LR *hold. Then*

(a) $T^{-1/2}\Xi^{+*}G_1 = o_p(1)$,

(b) $T^{-1}\left(\widehat{\Omega}_{00\cdot a} \otimes T^{-2}Z_2'Z_2\right)^{-1/2} \mathrm{vec}(\Xi^{+*}G_2) \xrightarrow{d} MN(0,\ [I_{nq_2} - P_{D_{z_2}}])$,

where $D_{z_2 2} = \left\{I_n \otimes \left(\int_0^1 B_{z_2}B_{z_2}'\right)^{-1/2} \int_0^1 B_{z_2}B_2'\right\}$.

Next we construct test statistics for IV validity using the scores defined above

$$\zeta = \mathrm{vec}(\Xi^{+*})(\widehat{\Omega}_{00\cdot a} \otimes Z'Z)^{-1}\,\mathrm{vec}(\Xi^{+*}). \tag{23}$$

The following theorem follows directly from Lemma 6.1.

THEOREM 6.2: *Under the same conditions as those in Lemma* 6.1,

$$\zeta \xrightarrow{d} \chi^2_{n(q_2-m_2)}.$$

However, since the cointegration structure of $\{z_t\}$ is unknown, so is q_2. The limit distribution of ζ is bounded by $\chi^2_{n(q-m)}$ and this could be used to construct a bounds test. However, we now propose a way to avoid such uncertainty in the limit theory.

We suggest the use of the one-sided long-run covariance estimator $\check{\Delta}_{0z}$. Rotating coordinates in $\mathbb{R}^q$ by the orthogonal matrix G, we have

$$\begin{aligned} T\check{\Delta}_{0z}G &= [T\check{\Delta}_{0z_1} \vdots T\check{\Delta}_{0z_2}] \\ &= [\widetilde{U}_{0\mathrm{GMM}}'Z_1 + o_p(\sqrt{T}) \vdots T\Delta_{0z_2} + o_p(T)] \\ &= [\widetilde{U}_{0\mathrm{GMM}}'Z_1 + o_p(\sqrt{T}) \vdots o_p(T)]. \end{aligned} \tag{24}$$

For the second block in the matrix in the second line, see Lemma 5.1. The third equality holds by Assumption NF^2. Next, define

$$\zeta_\Delta = T\,\mathrm{vec}(\check{\Delta}_{0z})'S_{zT}^{-1}\,\mathrm{vec}(\check{\Delta}_{0z}). \tag{25}$$

We rotate the matrix Z by G every time Z appears and using (24) it is easy to establish that

$$\begin{aligned} \zeta_\Delta &= \mathrm{vec}(\check{\Delta}_{0z})\mathcal{Z}_1 S_{z_1T}^{-1}\mathcal{Z}_1'\,\mathrm{vec}(\check{\Delta}_{0z}) + o_p(1) \\ &\xrightarrow{d} \chi^2_{n(q_1-m_1)}. \end{aligned}$$

Now let

$$\zeta^* = \zeta_\Delta + \zeta. \tag{26}$$

We have,

THEOREM 6.3: Suppose Assumptions EC, IV, LR and NF^2 hold, then

$$\zeta^* \xrightarrow{d} \chi^2_{n(q-m)}.$$

Notice that the limit distribution of ζ^* does not depend upon the unknown parameter q_2 as a result of the augmentation of the statistic. Roughly speaking, the orthogonality condition $E[u_{0t} \otimes z_{1t}]$ is transferred from ζ_2 to ζ_1 by $\check{\Delta}_{0z_1}$ so that we now do not lose q_1 degrees of freedom as we did before in Theorem 6.2.

It is straightforward to extend the above results to the FM–GIVE estimation procedure. We assume strict exogeneity of the instruments as in Section 5.2. Let

$$\begin{aligned} \widetilde{u}_{0\text{GIVE}t} &= y - \widetilde{A}_{\text{GIVE}} x_t, \\ \text{vec}(\widetilde{U}'_{0\text{GIVE}})^* &= \widehat{W}_T \, \text{vec}(\widetilde{U}'_{0\text{GIVE}}). \end{aligned}$$

and define the score with respect to the stationary instruments as

$$\begin{aligned} \text{vec}(\Xi_{1\text{GIVE}}) &= \mathcal{Z}^{*\prime} \text{vec}(\widetilde{U}'_{0\text{GIVE}})^* \\ &= \mathcal{Z}' \widehat{V}_{0T}^{-1} \, \text{vec}(\widetilde{U}'_{0\text{GIVE}}). \end{aligned}$$

(We do not have a simple expression for $\Xi_{1\text{GIVE}}$ before vectorization.) Also define $\{\widetilde{u}_{0\text{GIVE}t}\}$, which is corrected for endogeneity with respect to u_{at} as:

$$\widetilde{u}^+_{0\text{GIVE}t} = \widetilde{u}_{0\text{GIVE}t} - \widehat{\Omega}_{0a} \widehat{\Omega}_{aa}^{-1} u_{at}.$$

(In the calculation of $\widehat{\Omega}_{0x}$, $\{\widetilde{u}_{0\text{GIVE}t}\}$ may be used.) Using $\{\widetilde{u}^+_{0\text{GIVE}t}\}$, we define the "fully-modified" score as:

$$\Xi^+_{\text{GIVE}} = \widetilde{U}^{+*}_{0\text{GIVE}} Z - T \widehat{\Delta}_{0z}.$$

We let

$$\zeta_{\text{GIVE}} = \text{vec}(\Xi^{+*}_{\text{GIVE}})' (\widehat{\Omega}_{00\cdot x} \otimes Z'Z)^{-1} \text{vec}(\Xi^{+*}_{\text{GIVE}}). \tag{27}$$

As a direct consequence of Theorem 6.2, we have

THEOREM 6.2′: *Under Assumptions* EC, IV, LR, NF *and* SE

$$\zeta_{\text{GIVE}} \xrightarrow{d} \chi^2_{n(q_2-m_2)}.$$

Without augmentation, the test statistic ζ_{GIVE} converges in distribution to a chi-squared random variable with $n(q_2 - m)$ degrees of freedom as in the case of the FM–GMM procedure and the uncertainty with respect to the parameter q_2 arises again.

We therefore proceed to construct augmented test statistics. First define the autocorrelation function of the transformed processes $\text{vec}(\widehat{\Gamma}_{0z^*}(j)) = \frac{1}{T}\Delta \mathcal{Z}^* \text{vec}(\widehat{U}_{0\text{GIVE}_j})^* = \frac{1}{T}\Delta \mathcal{Z}\widehat{V}_{0T}^{-1}(\widehat{U}'_{0\text{GIVE}_j})$ where $\widehat{U}_{0\text{GIVE}_j}$ is the observation matrix of the residuals $\{\widehat{u}_{0\text{GIVE}_{t+j}}\}$. Then let the estimator of the corresponding one-side long run covariance matrix be

$$\text{vec}(\widehat{\Delta}_{0z^*}) = \sum_{j=0}^{T} w(j/K)\text{vec}(\widehat{\Gamma}_{0z^*}(j)),$$

where $w(\cdot)$ is a kernel function as in the preceding sections. Given the strict exogeneity of the instruments, we have

$$\begin{aligned}
\text{vec}[\widehat{\Delta}_{0z^*}G_1] &= \text{vec}\widehat{\Delta}_{0z_1^*} = \frac{1}{T}\mathcal{Z}_1^* \ \text{vec}(\widehat{U}'_{0\text{GIVE}})^* + o_p(1/\sqrt{T}) \\
\text{vec}[\widehat{\Delta}_{0z^*}G_2] &= \text{vec}\widehat{\Delta}_{0z_2^*} \xrightarrow{p} 0.
\end{aligned}$$

We define

$$\begin{aligned}
\zeta_{\Delta\text{GIVE}} &= \text{vec}(\check{\Delta}_{0z^*})'(\mathcal{Z}^{*\prime}\mathcal{Z}^*)^{-1} \ \text{vec}(\check{\Delta}_{0z^*}) &(28)\\
&= \text{vec}(\check{\Delta}_{0z^*})' \left(\mathcal{Z}'\widehat{V}_{0T}^{-1}\mathcal{Z}\right)^{-1} \text{vec}(\check{\Delta}_{0z^*}). &(29)
\end{aligned}$$

Then using ζ_{GIVE} as defined in (27), we let

$$\zeta^*_{\text{GIVE}} = \zeta_{\Delta\text{GIVE}} + \zeta_{\text{GIVE}}. \tag{30}$$

The following theorem can be established by the same lines of argument as Theorem 6.3.

THEOREM 6.3′: *Under Assumptions* EC^2, IV^2, LR^2, NF^2 *and* SE

$$\zeta^*_{\text{GIVE}} \xrightarrow{d} \chi^2_{n(q-m)}.$$

We can conduct tests of IV validity based on ζ^*_{GIVE} in the usual fashion. Note that the degrees of freedom of ζ^*_{GIVE} in the limit are $n(q-m)$, of which nq = (the number of equations) $\times$ (the total number of instruments), and nm is the total number of unknowns. This can be interpreted as the number of overidentifying restrictions, just as in classical test statistics for IV validity.

7. A Practical Guide to Our Formulae for Empirical Implementation

In the previous sections of this paper, we developed our theory by starting with simple models and moving towards more complicated cases. This presentation of our theory is chosen chiefly for an expository purpose. As a result of this progressive approach, a wide variety of estimators and test statistics have been included in our development. Therefore, it may be useful in this final section of our paper to provide practitioners with recipes for empirical applications of our FM–IV estimators and test statistics.

We consider a multiple regression model

$$\underset{n\times 1}{y_t} = \underset{n\times m}{A}\ \underset{m\times 1}{x_t} + \underset{n\times 1}{u_{0t}},$$

just as before. Let Z_t denote a q-vector of instruments. Here we assume no knowledge about the cointegrating relationships among the regressors and the instruments (that is, within the regressors, within the instruments, and between the two). Then, our FM–GMM estimation procedure can be implemented in the following way. (The procedures inside the square brackets are optional in what follows.)

FM–GMM Procedure

STEP 1: Run the "naive" IV regression

$$\widehat{A} = Y'P_zX(X'P_zX)^{-1}$$

and calculate the residual

$$\widehat{u}_{0t} = y_t - \widehat{A}x_t.$$

STEP 2: Use $\{\widehat{u}_{0t}\}$ obtained in Step 1 to calculate S_{zT} using the formula (19). Let

$$\text{vec}\ \widehat{A}_{\text{GMM}} = \left\{(I_n \otimes X'Z)S_{zT}^{-1}(I_n \otimes Z'X)\right\}^{-1}(I_n \otimes X'Z)S_{zT}^{-1}\ \text{vec}(Y'Z)$$

and calculate the GMM residual

$$\widehat{u}_{0\mathrm{GMM}t} = y_t - \widehat{A}_{\mathrm{GMM}} x_t.$$

STEP 3: Use $\{\widehat{u}_{0\mathrm{GMM}t}\}$ obtained in Step 2 and $\{u_{at}\} = \{(\Delta x_t', \Delta z_t')'\}$ to estimate the long-run covariance matrices

$$\Omega_{0a},\ \Omega_{aa} \text{ and } \Delta^{+}_{u_0 z} = \Delta_{u_0 z} - \Omega_{0a}\Omega_{aa}^{-1}\Delta_{u_a},$$

using kernel estimators (see formula (11)) with a kernel function that satisfies the conditions stated in Assumption LR. [Also calculate S_{zT} again as in Step 2, but use $\{\widehat{u}_{0\mathrm{GMM}t}\}$ in place of $\{\widehat{u}_{0t}\}$.] Using the estimates $\widehat{\Omega}_{0a}$, $\widehat{\Omega}_{aa}$ and $\check{\Delta}^{+}_{0z}$, obtained above, construct

$$\mathrm{vec}\ \widetilde{A}_{\mathrm{GMM}} = \left\{(I_n \otimes X'Z)S_{zT}^{-1}(I_n \otimes Z'X)\right\}^{-1}(I_n \otimes X'Z)S_{zT}^{-1}\ \mathrm{vec}(\check{Y}^{+\prime}Z - T\check{\Delta}^{+}_{u_0 z})$$

where

$$\check{Y}^{+\prime} = Y' - \widehat{\Omega}_{0a}\widehat{\Omega}_{aa}^{-1}U_a'.$$

Calculate

$$\widetilde{u}_{0\mathrm{GMM}t} = y_t - \widetilde{A}_{\mathrm{GMM}} x_t.$$

This completes our FM–GMM estimation. [We can iterate this process by returning to the beginning of Step 3 and using $\{\widetilde{u}_{0\mathrm{GMM}t}\}$ in place of $\{\widehat{u}_{0\mathrm{GMM}t}\}$.]

STEP 4: Estimate $\Delta_{u_0 z}$ [and Ω_{0a} again] using the GMM residual $\{\widetilde{u}_{0\mathrm{GMM}t}\}$ obtained in Step 3 and call the estimate $\check{\Delta}_{u_0 z}$ [and $\widehat{\Omega}_{0a}$]. Calculate the corrected GMM residuals

$$\widetilde{u}^{+}_{0\mathrm{GMM}t} = \widetilde{u}_{0\mathrm{GMM}t} - \widehat{\Omega}_{0a}\widehat{\Omega}_{aa}^{-1}u_{at}$$

and its long run variance estimate

$$\widehat{\Omega}_{00\cdot a} = \widehat{\Omega}_{00} - \widehat{\Omega}_{0a}\widehat{\Omega}_{aa}^{-1}\widehat{\Omega}_{a0}.$$

[We could also calculate S_{zT} again as in Step 2, but use $\{\widetilde{u}_{0\mathrm{GMM}t}\}$ in place of $\{\widehat{u}_{0t}\}$.]

STEP 5: Construct the fully modified score matrix

$$\widetilde{\Xi}^{+*} = \widetilde{U}_{0\mathrm{GMM}}'Z - T\check{\Delta}_{u_0 z},$$

and the test statistics

$$\zeta = \text{vec}(\Xi^{+*})'(\widehat{\Omega}_{00\cdot a} \otimes Z'Z)^{-1}\text{vec}(\Xi^{+*}),$$

$$\zeta_\Delta = T\ \text{vec}(\check{\Delta}_{u_0z})'S_{zT}^{-1}\ \text{vec}(\check{\Delta}_{u_0z}), \text{ and}$$

$$\zeta^* = \zeta_\Delta + \zeta.$$

We can conduct instrument validity tests using ζ^* as an asymptotic χ^2 criterion with $n(q-m)$ degrees of freedom. We call this the FM–GMM instrument validity test. □

We may also want to use FM–GIVE when certain additional conditions hold. For instance, it is assumed here that Z_t is strictly exogenous in what follows (but again as discussed in Section 5.2 this can be relaxed). We also work under errors of the form prescribed in Assumption VR. Then the following procedure is suggested.

FM–GIVE Procedure

STEP 1′: = Step 1.

STEP 2′: Use $\{\widehat{u}_{0t}\}$ to estimate the VAR model in Assumption VR by the use of OLS. Using the estimates $\widehat{C}$ and $\widehat{\Sigma}_\varepsilon$, obtain the transformation matrix $\widehat{W}_T$ given in the formulae of Section 5.3. Let

$$\mathcal{X}^* = \widehat{W}_T(I \otimes Z),\ \mathcal{Z}^* = \widehat{W}_T(I \otimes Z),\ \text{vec}(Y')^* = \widehat{W}_T\ \text{vec}(Y').$$

STEP 3′: Construct the estimator

$$\text{vec}\ \widehat{A}_{\text{GIVE}} = (\mathcal{X}^{*\prime}P_z\mathcal{X}^*)^{-1}\mathcal{X}^{*\prime}\mathcal{Z}^*\left\{(\mathcal{Z}^{*\prime}\mathcal{Z}^*)^{-1}\mathcal{Z}^{*\prime}\ \text{vec}(Y')^*\right\},$$

and associated residual

$$\widehat{u}_{0\text{GIVE}t} = y_t - \widehat{A}_{\text{GIVE}}x_t.$$

STEP 4′: Estimate $\Omega_{0x}\Omega_{xx}$ and Δ_{xz} using $\{\widehat{u}_{0t}\}$ [or $\{\widehat{u}_{0\text{GIVE}t}\}$] and call these estimates $\widehat{\Omega}_{0x}$, $\widehat{\Omega}_{xx}$ and $\widehat{\Delta}_{xz}$. Construct the final FM–GIVE estimator

$$\begin{aligned}\text{vec}\ \widetilde{A}_{\text{GIVE}} &= (\mathcal{X}^{*\prime}P_Z\mathcal{X}^*)^{-1} \\ &\quad \times \mathcal{X}^{*\prime}\mathcal{Z}^*\left\{(\mathcal{Z}^{*\prime}\mathcal{Z}^*)^{-1}\mathcal{Z}^{*\prime}\ \text{vec}(Y')^* - \text{vec}(\widehat{\Omega}_{0x}\widehat{\Omega}_{xx}^{-1}(U_x'Z - T\widehat{\Delta}_{xz})(Z'Z)^{-1})\right\},\end{aligned}$$

and calculate the residual

$$\widetilde{u}_{0\mathrm{GIVE}t} = y_t - \widetilde{A}_{\mathrm{GIVE}} x_t.$$

[Once again we can iterate this process by returning to the beginning of Step 2′ or Step 4′ and using $\{\widetilde{u}_{0\mathrm{GIVE}t}\}$ in place of $\{\widehat{u}_{0t}\}$ [$\{\widehat{u}_{0\mathrm{GIVE}t}\}$].]

STEP 5′: Use $\{\widetilde{u}_{0\mathrm{GIVE}t}\}$ to calculate $\mathrm{vec}(\widehat{\Delta}_{0z^*})$ following the formulae in Section 6. [Also calculate $\widehat{\Omega}_{0x}$ again.] Calculate the corrected GIVE residual

$$\widetilde{u}^{*}_{0\mathrm{GIVE}t} = \widetilde{u}_{0\mathrm{GIVE}t} - \widehat{\Omega}_{0x}\widehat{\Omega}_{xx}^{-1}u_{x_t},$$

and its long run variance estimate

$$\widehat{\Omega}_{00\cdot a} = \widehat{\Omega}_{00} - \widehat{\Omega}_{0x}\widehat{\Omega}_{xx}^{-1}\widehat{\Omega}_{x0}.$$

STEP 6′: Construct the fully modified GIVE score matrix

$$\Xi^{+*}_{\mathrm{GIVE}} = U^{+*}_{0\mathrm{GIVE}} Z - T\widehat{\Delta}_{0z^*},$$

and the test statistics

$$\begin{aligned}
\zeta_{\Delta\mathrm{GIVE}} &= T\,\mathrm{vec}(\widehat{\Delta}_{0z^*})'(Z^{*\prime}Z^{*})^{-1}\mathrm{vec}(\widehat{\Delta}_{0z^*}),\\
\zeta_{\mathrm{GIVE}} &= \mathrm{vec}(\Xi^{+*}_{\mathrm{GIVE}})'(\widehat{\Omega}_{00\cdot x}\otimes Z'Z)^{-1}\mathrm{vec}(\Xi^{+*}_{\mathrm{GIVE}}), \text{ and}\\
\zeta^{*}_{\mathrm{GIVE}} &= \zeta_{\Delta\mathrm{GIVE}} + \zeta_{\mathrm{GIVE}}.
\end{aligned}$$

FM–GIVE instrument validity tests can now be conducted using $\zeta^{*}_{\mathrm{GIVE}}$ as an asymptotic $\chi^2_{n(q-m)}$ criterion. □

As shown in the above procedures, the calculation of FM–IV estimators involves the computation of correction terms at the first stage, and an IV regression at the second stage. As a result of this two stage structure, which is common to all FM estimators, it is easy to check the impact of the estimator modifications in the course of analysis. In particular, the values of the FM correction terms can be used to assess the degree of endogeneity and the extent of serial correlation in the model. Thus, these corrections provide useful information which suggest features of the model that are empirically relevant and important. In sum, while the FM(–IV) methods have many convenient theoretical properties, they also have advantages that seem to be important and useful in empirical implementation.

Finally, when we apply FM–IV methods in practice, the choice of instrumental variables is important. As far as the stationary components of instruments are concerned, the usual IV validity conditions need to be satisfied, as those discussed in Section 3. Also, the number of nonstationary components in the instruments must be at least equal to the number of the nonstationary components in the regressors. As in usual applications of IV procedures, to ensure that these conditions are met we may explore a range of possible candidates for instruments. Then we seek to employ a "large enough" number of (stationary and nonstationary) instruments so that the aforementioned "order" conditions are satisfied. The validity of these instruments can subsequently be tested using our FM–GMM validity test.

In many cases, in fact, economic theory suggests sets of IV candidate variables. For example, in many RE models, as we mentioned earlier, lagged regressors are assumed to be contained in economic agents' information sets and are therefore orthogonal to subsequent innovations that affect the outcome of agents' decision making. Such variables can then be used as instruments and Assumption NF (NF^2) is satisfied. The advantage of the use of lagged regressors as instruments is the apparent fact that the integratedness properties of the regressors and the instruments coincide if such instruments are employed. Thus, if we use a vector of lagged regressors as instruments, such a choice of IV has certain advantages. But of course we need to be careful to ensure that the instrument set is not "too large," so that it does not distort finite sample performance.

As for the choice of nonstationary instruments, artificially generated nonstationary processes could also be used as valid instruments, at least theoretically. This method exploits the spurious correlation between independent nonstationary processes (see Phillips (1986) and Phillips and Hansen (1990)). If we are short of nonstationary IV candidate variables, this method might be used. However, attention should be paid to the finite sample properties of the FM estimators if such instruments were to be used (see Hansen and Phillips (1990) for some discussion of this point). In any case, whenever we employ the FM–IV procedure, we need to investigate candidates for instruments carefully, just as we do in usual IV regressions with stationary time series. However, our theory allows us to choose instruments from a very large set of potential candidates, especially in the case of FM-GMM. In fact, this is a great advantage of IV and GMM methods in general.

8. Appendix

8.1. Proof of Lemma 3.1: Rotating coordinates in the regressor space $\mathbb{R}^m$ by the orthogonal matrix H, we have

$$C_T H = (U_0' P_z - \Psi_T) X_h (X_h' P_z X_h)^{-1}.$$

By straightforward calculation part (a) and part (b) can be established; see proof of Theorem 4.1 in Phillips (1994). Then for part (a), Assumption IV [(a), (b) and (c)] ensures the validity of the stationary instruments and the required CLT is given by (8). The usual weak convergence arguments for cointegrating regressions (see Phillips (1991b), for example) deliver part (b) of the lemma. □

8.2. Proof of Lemma 4.1: In the following, we need to calculate stochastic orders of quantities such as

$$\begin{aligned} \widehat{\Omega}_{u_0 \Delta u_1} &= \sum_{j=-K+1}^{K-1} w(j/K)\widehat{\Gamma}_{u_0 \Delta u_1}(j) = \sum_{j=-K+1}^{K-1} w(j/K)(\widehat{\Gamma}_{u_0 u_1}(j) - \widehat{\Gamma}_{u_0 u_1}(j+1)) \qquad (A.1) \\ &= -w((K-1)/K)\widehat{\Gamma}_{u_0 u_1}(K) + w((-K+1)/K)\widehat{\Gamma}_{u_0 u_1}(-K+1) \\ &\quad + \sum_{j=-K+2}^{K-1} (w(j/K) - w((j-1)/K))\widehat{\Gamma}_{u_0 u_1}(j) \\ &= F_{1T} + F_{2T} + F_{3T}, \text{ say.} \end{aligned}$$

Note that the summation will be taken from $-T+1$ to $T-1$ in the case of the quadratic spectral kernel but the argument in the rest of the proof is otherwise unaltered by the change.

We first focus on the component F_{3T} in (A.1), which is a sum of the autocovariances weighted by the first difference of the lag window $w(j/K)$. In what follows, we assume twice differentiability of $w(\cdot)$ as in Phillips (1994). By the mean value theorem

$$w(j/K) - w((j-1)/K) = K^{-1} w'(j^*/K),$$

where $j^* \in [j-1, j]$ and is defined for each j. Then

$$F_{3T}^0 = \sum_{j=-K+2}^{K-1} (w(j/K) - w((j-1)/K))\Gamma_{u_0u_1}(j) \tag{A.2}$$
$$= \frac{1}{K}\sum_{j=-K+2}^{K-1} w'(j^*/K)\Gamma_{u_0u_1}(j) + \Sigma(w(j/K) - w((j-1)/K))\Gamma_{u_0u_1}(j)$$
$$= \frac{1}{K^r}\sum_{j=-K+2}^{K-1} \frac{w'(j^*/K)}{(j^*/K)^{r-1}}(j^*/j)^{r-1}(j)^{r-1}\Gamma_{u_0u_1}(j) + \Sigma(w(j/K) - w((j-1)/K))\Gamma_{u_0u_1}(j).$$

The condition (13) implies that $w'(j^*/K)/(j^*/K)^{r-1}$ converges boundedly to $w_{(r)}$ for each fixed j. Thus F_{3T}^0 is of order $O(K^{-r})$. Also, following Hannan (1970, p. 280, Theorem 9), we have

$$\lim_{T\to\infty} KT \text{ Var[vec } F_{3T}] = \lim_{T\to\infty} KT\frac{1}{K^2}\text{Var}\left[\text{vec}\left\{\sum_{j=-K+2}^{K-1} w'(j^*/K)\widehat{\Gamma}_{u_0u_1}(j)\right\}\right]$$
$$= \lim_{T\to\infty} \frac{T}{K}\text{Var}\left[\text{vec}\left\{\sum_{j=-K+2}^{K-1} w'(j^*/K)\widehat{\Gamma}_{u_0u_1}(j)\right\}\right]$$
$$= \text{constant.}$$

Thus, combining expressions for the variance and the bias (cf. Hannan, 1970, Theorem 10, p. 283) we have

$$E[\text{vec}(F_{3T})\,\text{vec}(F_{3T})'] = O\left(\frac{1}{KT}\right) + O\left(\frac{1}{K^{2r}}\right) = O\left(\frac{1}{T^{k+1}}\right) + O\left(\frac{1}{T^{2kr}}\right)$$

where $K = O(T^k)$. Therefore

$$F_{3T} = O_p\left(T^{-(\frac{k+1}{2}\wedge rk)}\right).$$

In the following, we assume $r = 2$, and then $F_{3T} = O_p(T^{-\delta})$, with $\delta = ((k+1)/2) \wedge 2k$.

On the other hand, F_{1T} and F_{2T} are negligible since they are of order $o_p((K-1)^{-2})$. In sum, we deduce that $\widehat{\Omega}_{u_0\Delta u_1} = O_p(T^{-\delta})$.

Next we consider

$$\widehat{\Omega}_{0\Delta u_1} = \widehat{\Omega}_{\widehat{u}_0\Delta u_1} = \widehat{\Omega}_{u_0\Delta u_1} + \sum_{j=-K+1}^{K-1} w(j/K)(\widehat{A}-A)\widehat{\Gamma}_{x\Delta u_1}(j) \tag{A.3}$$
$$= \widehat{\Omega}_{u_0\Delta u_1} - w((K-1)/K)(\widehat{A}-A)\widehat{\Gamma}_{xu_1}(K) + w((-K+1)/K)(\widehat{A}-A)\widehat{\Gamma}_{xu_1}(-K+1)$$
$$+ \sum_{j=-K+2}^{K-1} (w(j/K) - w((j-1)/K))(\widehat{A}-A)\widehat{\Gamma}_{xu_1}(j)$$
$$= B_{1T} + B_{2T} + B_{3T} + B_{4T}, \text{ say.}$$

We have shown that $B_{1T} = O_p(T^{-\delta})$. B_{2T} and B_{3T} are easily seen to be of order $o_p(1/\sqrt{T})$; see (P13), (P14) in Phillips (1994). In sum, we have

$$\widehat{\Omega}_{0\Delta u_1} = O_p(T^{-\delta}) + O_p(T^{-1}) = O_p(T^{-\delta}).$$

By following a similar line of argument, we find $\widehat{\Omega}_{u_2\Delta u_1} = O_p(T^{-\delta})$.

In what follows, we also need to invert the estimator of a long run variance matrix of $I(-1)$ variables, e.g. $\widehat{\Omega}_{\Delta u_1\Delta u_1}$. As before, we find $\widehat{\Omega}_{\Delta u_1\Delta u_1} = O_p(T^{-\delta})$. By considering the terms at lag zero, we can verify that the rate of convergence of $\widehat{\Omega}_{\Delta u_1\Delta u_1}$ is no faster than $T^{-\delta}$ and indeed $\widehat{\Omega}^{-1}_{\Delta u_1\Delta u_1} = O_p(T^{2k})$. For a rigorous treatment of this point, see Phillips (1994).

Now, using the partitioned matrix inversion formula and writing $\widehat{\Omega}_{u_2u_2\cdot\Delta u_1} = \widehat{\Omega}_{u_2u_2} - \widehat{\Omega}_{u_2\Delta u_1}\widehat{\Omega}^{-1}_{\Delta u_1\Delta u_1}\widehat{\Omega}_{\Delta u_1 u_2}$ we obtain

$$\begin{aligned}
\widehat{\Omega}_{0h}\widehat{\Omega}^{-1}_{hh} &= \widehat{\Omega}_{0\Delta u_1}\big(\widehat{\Omega}^{-1}_{\Delta u_1\Delta u_1} - \widehat{\Omega}^{-1}_{\Delta u_1\Delta u_1}\widehat{\Omega}_{\Delta u_1 u_2}\widehat{\Omega}^{-1}_{u_2u_2\cdot\Delta u_1}\widehat{\Omega}_{u_2\Delta u_1}\widehat{\Omega}^{-1}_{\Delta u_1\Delta u_1} \\
&\quad \vdots - \widehat{\Omega}^{-1}_{\Delta u_1\Delta u_1}\widehat{\Omega}_{\Delta u_1 u_2}\widehat{\Omega}^{-1}_{u_2u_2\cdot\Delta u_1}\big) + \widehat{\Omega}_{0u_2}\left(-\widehat{\Omega}^{-1}_{u_2u_2\cdot\Delta u_1}\widehat{\Omega}_{\Delta u_1 u_2}\widehat{\Omega}^{-1}_{\Delta u_1\Delta u_1} \vdots \widehat{\Omega}^{-1}_{u_2u_2\cdot\Delta u_1}\right) \\
&= O_p(T^{-\delta})\Big(O_p(T^{2k}) - O_p(T^{2k})O_p(T^{-\delta})O_p(1)O_p(T^{-\delta})O_p(T^{2k}) \vdots O_p(T^{2k})O_p(T^{-\delta})\Big) \\
&\quad + (\Omega_{0u_2} + o_p(1))\Big(O_p(T^{-\delta})O_p(T^{2k}) \vdots \Omega^{-1}_{u_2u_2} + o_p(1)\Big) \\
&= \Big[O_p(T^{-\delta+2k}) \vdots \Omega_{0u_2}\Omega^{-1}_{u_2u_2} + o_p(1)\Big].
\end{aligned}$$

Next we evaluate the matrix $T^{-1}U_h' Z_G - \widehat{\Delta}_{hg}$ block by block. For the (1,1) block, we have

$$\begin{aligned}
T^{-1}\Delta U_1' Z_1 - \widehat{\Delta}_{\Delta u_1\Delta z_1} &= w((K-1)/K)\widehat{\Gamma}_{\Delta u_1 z_1}(K) - \sum_{j=1}^{K-1}(w(j/K) - w((j-1)/K))\widehat{\Gamma}_{\Delta u_1 z_1}(j) \\
&= O_p(T^{-\delta}).
\end{aligned}$$

As for the (2,1) block we can show that $T^{-1}U_2' Z_1 - \widehat{\Delta}_{u_2\Delta z_1} = O_p(T^{-\delta})$ in the same way. The (1,2) block is $T^{-1}U_1' Z_2 - \widehat{\Delta}_{\Delta u_1 u_{z_2}} = O_p(T^{-1/2})$ as in Lemma 8.1(g) of Phillips (1994).

The result for the (2,2) block is familiar from the original Phillips–Hansen (1990) study. Combining the above results, we have

$$T^{-1}U_h' Z_G - \widehat{\Delta}_{hg} = \begin{bmatrix} O_p(T^{-\delta}) & \vdots & O_p(T^{-1/2}) \\ \cdots & \cdots & \cdots \\ O_p(T^{-\delta}) & \vdots & N_T \end{bmatrix}, \text{ where } N_T \xrightarrow{d} \int_0^1 dB_2 B_{z_2}'.$$

In sum, we conclude that

$$\widehat{\Omega}_{0h}\widehat{\Omega}_{hh}^{-1}\left(\frac{1}{T}U_h'Z_G-\widehat{\Delta}_{hg}\right)=\left[O_p(T^{-\delta})+o_p(1/\sqrt{T})\ \vdots\ \Omega_{0u_2}\Omega_{u_2u_2}^{-1}N_T+O_p(T^{2k-\delta-1/2})\right].$$

Further, if $k\in(1/4,2/3)$, then $2k-\delta-1/2\leq-1/2$ and

$$\widehat{\Omega}_{0h}\widehat{\Omega}_{hh}^{-1}\left(\frac{1}{T}U_h'Z_G-\widehat{\Delta}_{hg}\right)=\left(o_p(1/\sqrt{T})\ \vdots\ \Omega_{0u_2}\Omega_{u_2u_2}^{-1}N_T+O_p(1\sqrt{T})\right).$$

This proves part (a).

Part (b) can be proved by straightforward calculation. □

8.3. Proof of Lemma 4.2: For the first equality in part (a) in the lemma it suffices to show that

$$\widehat{\Delta}_{0z_1}\left(=\ \widehat{\Delta}_{\widehat{u}_0\Delta z_1}\right)=T^{-1}\widehat{U}_0'Z_1+o_p(1/\sqrt{T}).$$

By definition,

$$\begin{aligned}
\widehat{\Delta}_{\widehat{u}_0\Delta z_1} &= \widehat{\Gamma}_{\widehat{u}_0z_1}(0)-w((K-1)/K)\widehat{\Gamma}_{\widehat{u}_1z_1}(K)+\sum_{j=1}^{K-1}(w(j/K)-w((j-1)/K))\widehat{\Gamma}_{\widehat{u}_0z_1}(j)\\
&= \widehat{\Gamma}_{\widehat{u}_0z_1}(0)-w((K-1)/K)\widehat{\Gamma}_{u_0z_1}(K)-w((K-1)/K)(\widetilde{A}-A)\widehat{\Gamma}_{xz_1}(K)\\
&\quad+\sum_{j=1}^{K-1}(w(j/K)-w(j-1)/K))\widehat{\Gamma}_{u_0z_1}(j)\\
&\quad+\sum_{j=1}^{K-1}(w(j/K)-w((j-1)/K))(\widetilde{A}-A)H\widehat{\Gamma}_{xz_1}(j)\\
&= G_{1T}+G_{2T}+G_{3T}+G_{4T}+G_{5T},\ \text{say.}
\end{aligned}$$

Note that $G_{2T}=o_p(1/\sqrt{T})$ since $w((K-1)/K)=o_p(1)$ for the truncated kernels we use in the paper, and $\widehat{\Gamma}_{u_0z_1}(K+1)=O_p(1/\sqrt{T})$ by Assumption NF. G_{3T} and G_{5T} are also of order $o_p(1/\sqrt{T})$, just as in the analysis of (A.3) in the proof of Lemma 4.1. Thus, $G_{4T}=o_p(T^{-1/2})$; see Phillips (1994), Lemma 8.1(h). In sum, the equality at the beginning of the proof is now established. This result and Lemma 4.1(b) prove the first equality in the lemma.

For the second equality, we start by using the definition of $\widehat{U}_0$, i.e.

$$Y'=\widehat{A}_1X_1'+\widehat{A}_2X_2'+\widehat{U}_0'.$$

Thus,

$$\widehat{U}_0'P_{z_1}X_1\ =\ (A_1-\widehat{A}_1)X_1'P_{z_1}X_1+(A_2-\widehat{A}_2)X_2'P_{z_1}X_1+U_0'P_{z_1}X_1$$

$$\begin{aligned} &= (A_1 - \widehat{A}_1)X_1'P_{z_1}X_1 + U_0'P_{z_1}X_1 + O_p(1) \\ &= -U_0'P_{z_1}X_1(X_1'P_{z_1}X_1)^{-1}X_1'P_{z_1}X_1 + U_0'P_{z_1}X_1 + O_p(1) \\ &= O_p(1), \end{aligned}$$

where the second equality above follows from the fact that $\widehat{A}_2$ is T-consistent. The second equality in part (a) of the lemma now follows immediately. Other results directly follow from Lemma 4.1. □

8.4. Proof of Theorem 4.3: First, following the notation used in (3.7) we define

$$\begin{aligned} \Psi_T &= \widehat{\Omega}_{0a}\widehat{\Omega}_{aa}^{-1}U_a'P_z + T\widehat{\Delta}_{0z}^{+}(Z'Z)^{-1}Z' \\ &= T\widehat{\Delta}_{0z}(Z'Z)^{-1}Z' + T\widehat{\Omega}_{0a}\widehat{\Omega}_{aa}^{-1}(T^{-1}U_a'Z - \widehat{\Delta}_{az})(Z'Z)^{-1}Z' = \Psi_{1T} + \Psi_{2T}, \text{ say.} \end{aligned}$$

Then $\Psi_1 X_1 = \widetilde{U}_0'P_{z_1}X_1 + o_p(\sqrt{T})$ by Lemma 4.2(a) and 4.2(b). Thus (C1) holds. Lemma 4.2(c) also shows that (C2) holds. Therefore we can apply Lemma 3.1, 4.1(a) and (c) and establish the required results. □

8.5. Proof of Lemma 5.1: This follows the same lines as the proof of Lemma 4.2(a) and is therefore omitted.

8.6. Proof of Theorem 5.2: Comparing the form of (9) and the estimator (16), we define Ψ_T as

$$\text{vec}\,\Psi_T = \mathcal{Z}S_{zT}^{-1}\,\text{vec}\left(\widehat{\Omega}_{0a}\widehat{\Omega}_{aa}^{-1}(U_a'Z - T\check{\Delta}_{az}) - T\check{\Delta}_{0z}\right).$$

Therefore

$$\text{vec}(\Psi_T X_1) = \mathcal{X}_1'\mathcal{Z}S_{zT}^{-1}\,\text{vec}\left(\widehat{\Omega}_{0a}\widehat{\Omega}_{aa}^{-1}(U_a'Z - T\check{\Delta}_{az})\right) - \mathcal{X}_1'\mathcal{Z}S_{zT}^{-1}\text{vec}(T\check{\Delta}_{0z}).$$

We now vectorize the result in Lemma 4.2 and replace $(\mathcal{Z}'\mathcal{Z})^{-1}$ by S_{zT}^{-1}. (This replacement does not change the stochastic orders, given (17).) Part (b) of the lemma shows that the first term in the last expression is of order $o_p(\sqrt{T})$. Therefore, by applying Lemma 5.2 to the second term, we have $\text{vec}(\Psi_T X_1) = o_p(\sqrt{T})$, and thus (C1) holds.

To establish (C2), we use part (c) of Lemma 4.2 (modified as indicated above) and (17). We have,

$$T^{-1}\mathcal{X}_2' \text{ vec}(\Psi_T) \xrightarrow{d} \left\{\Omega_{00}^{-1} \otimes \int_0^1 dB_2 B_{z_2}' \left(\int_0^1 B_{z_2} B_{z_2}'\right)^{-1}\right\} \text{vec}(\Delta_{0z_2}) + (\Omega_{00}^{-1}\Omega_{0b}\Omega_{bb}^{-1} \otimes I)\left(\int_0^1 dB_b \otimes \widetilde{B}_2\right),$$

which establishes (C2). Then, by Lemma 3.1(a) and (17a) we have

$$\begin{aligned}\sqrt{T}\text{ vec}((\widetilde{A}_{\text{GMM}} - A)H_1) &= \left(\mathcal{X}_1 \mathcal{Z}_1 S_{z_1T}^{-1} \mathcal{Z}_1' \mathcal{X}_1\right)^{-1} \mathcal{X}_1 \mathcal{Z}_1 S_{z_1T}^{-1} \text{ vec}(U_0' Z_1) + o_p(1) \\ &= \sqrt{T}\text{ vec}(\widehat{A}_{\text{GMM}} - A)H_1 + o_p(1) \\ &\xrightarrow{d} N\left(0,\ [\mathcal{K}_{z_1} S_{z_1}^{-1} \mathcal{K}_{z_1}]^{-1}\right),\end{aligned}$$

proving the first part of the theorem. For the second part, we use Lemma 3.1(b), (17b) and the limit of $T^{-1}\mathcal{X}_2$ vec(Ψ_T) obtained above. We have

$$\begin{aligned}\sqrt{T}\text{ vec}[(\widetilde{A}_{\text{GMM}} - A)H_2] &\xrightarrow{d} \left\{\Omega_{00}^{-1} \otimes \int_0^1 dB_2 B_z' \left(\int_0^1 B_z B_z'\right)^{-1} \int_0^{\cdot} B_z B_2'\right\}^{-1} \\ &\times \left\{ \left(\Omega_{00}^{-1} \otimes \int_0^1 dB_2 B_z' \left(\int_0^1 B_z B_z'\right)^{-1}\right) \left(\int_0^1 dB_b \otimes B_z + \text{vec}(\Delta_{0z}) - \text{vec}(\Delta_{0z})\right) \right. \\ &\left. - (\Omega_{00}^{-1} \otimes I)\left(\int_0^1 d\Omega_{0b}\Omega_{bb}^{-1} B_b \otimes \widetilde{B}_2\right) \right\} \\ &= \left\{I \otimes \left(\int_0^1 \widetilde{B}_2 \widetilde{B}_2'\right)^{-1}\right\} \left(\int_0^1 dB_{0\cdot b} \otimes \widetilde{B}_2\right),\end{aligned}$$

giving the required result. □

8.7. Proof of Theorem 5.3: First, note that the premultiplication by w_T does not change the order of integration of a time series. This point can be seen as follows. Take the matrix $\mathcal{X} = (I_n \otimes X)$ which frequently appears after the vectorization of the estimator. By the use of the $(nm \times nm)$ rotation matrix $\mathcal{H} = [\mathcal{H}_1 \vdots \mathcal{H}_2] = [I_n \otimes H_1 \vdots I_n \otimes H_2]$, we decompose the matrix $\mathcal{X}$ as $\mathcal{X}\mathcal{H} = [\mathcal{X}_1 \vdots \mathcal{X}_2]$, so that the first nm_1 columns are observations of stationary series, while the last nm_2 columns are observations of nonstationary series. Clearly, after premultiplication by W_T, we have $W_T\mathcal{X}\mathcal{H} = [\mathcal{X}_1^* \vdots \mathcal{X}_2^*]$, which has the same property in terms of the orders of integration in the given decomposition.

In view of the above, we can use the results in Lemma 3.1, 4.1 and Lemma 4.2 by vectorizing them and assigning the superscript "*" to each matrix and vector as necessary to signify that the transformation by W_T has been performed. Since the correction terms used in the definition of the FM–GIVE estimator (20) are the same as those in the definition of the FM–IV estimator, we know that conditions (C1) and (C2) hold. Then, by Lemma 3.1(a) we have the first result in the theorem.

Next, using the idea of the so-called Beveridge–Nelson (1981), or BN, decomposition, we observe that

$$\begin{aligned}\mathcal{X}_2^* &= \widehat{W}_T(I_n \otimes X_2) = (\hat{\Sigma}_\varepsilon^{-1/2} \otimes I_T)\left(\Sigma_{r=0}^p(\hat{C}_r \otimes \mathcal{L}_T^{r\prime})\right)(I_n \otimes X_2) = \Sigma_{r=0}^p \hat{C}_r \otimes X_{2-r} \\ &= C^\varepsilon(1) \otimes X_2 + \Xi + O_p(1),\end{aligned}$$

where $C^\varepsilon(L) = \Sigma_\varepsilon^{-1/2}C(L), \tilde{C}_r^\varepsilon = \Sigma_{s=r+1}^p C_r^\varepsilon$, and $\Xi = -\Sigma_{r=0}^p \tilde{C}_r^\varepsilon \otimes U_{2-r}$ the last of which represents the observation matrix of the stationary terms. Therefore, we can think of $C^\varepsilon(1) \otimes X_2$ as the long run "approximation" of $\mathcal{X}_2^*$. Similarly, we have $\mathcal{Z}^* \simeq C^\varepsilon(1) \otimes Z$ and $\text{vec}(U_0')^* \simeq (C^\varepsilon(1) \otimes I)\text{vec}(U_0')$. Sample covariance matrices of these transformed data matrices have the following asymptotics:

$$T^{-2}\mathcal{X}_2^{*\prime}\mathcal{Z}_2^* = C^\varepsilon(1)'C^\varepsilon(1) \otimes T^{-2}X_2'Z_2 + o_p(1) \xrightarrow{d} \Omega_{00}^{-1} \otimes \textstyle\int_0^1 B_2 B_{z_2}' \tag{A.6a}$$

$$T^{-2}\mathcal{Z}_2^{*\prime}\mathcal{Z}_2^* = C^\varepsilon(1)'C^\varepsilon(1) \otimes T^{-2}Z_2'Z_2 + o_p(1) \xrightarrow{d} \Omega_{00}^{-1} \otimes \textstyle\int_0^1 B_{z_2} B_{z_2}' \tag{A.6b}$$

$$T^{-1}\mathcal{Z}_2^{*\prime}\ \text{vec}(U_0')^* = [C^\varepsilon(1)'C^\varepsilon(1) \otimes I]\text{vec}(T^{-1}U_0'Z_2) + o_p(1) \xrightarrow{d} (\Omega_{00}^{-1} \otimes I)\left(\textstyle\int_0^1 dB_0 B_{z_2}\right). \tag{A.6c}$$

In these expressions we use the fact that $C^\varepsilon(1)'C^\varepsilon(1) = C(1)'\Sigma_\varepsilon^{-1}C(1) = \Omega_{00}^{-1}$. Notice that the asymptotics of (A.6c) do not involve a one-sided long run covariance term in view of the strict exogeneity of $\{z_{2t}\}$. By (A.6), we have

$$\begin{aligned}&\mathcal{X}_2^{*\prime}P_{\mathcal{Z}_2^*}\mathcal{X}_2^* \xrightarrow{d} \Omega_{00}^{-1} \otimes \textstyle\int_0^1 \tilde{B}_2\tilde{B}_2' \\ &\mathcal{X}_2^{*\prime}P_{\mathcal{Z}^*}\ \text{vec}(U_0')^* \xrightarrow{d} (\Omega_{00}^{-1} \otimes I)\left(\textstyle\int_0^1 dB_0 \otimes \tilde{B}_2\right),\end{aligned}$$

and

$$\mathcal{X}_2^{*\prime}\mathcal{Z}_2^*\ \text{vec}\left[\hat{\Omega}_{0x}\hat{\Omega}_{xx}^{-1}(U_x'Z_2 + T\hat{\Delta}_{xz_2})(Z_2'Z_2)^{-1}\right] \xrightarrow{d} (\Omega_{00}^{-1} \otimes I)\left(\textstyle\int_0^1 \Omega_{02}\Omega_{22}^{-1}dB_0 \otimes \tilde{B}_2\right).$$

By Lemma 3.1(b) we have

$$\begin{aligned}\text{vec}\left(T(\tilde{A}_{\text{GIVE}} - A)H_2\right) &= T(\mathcal{X}_2^{*\prime}P_{\mathcal{Z}_2^*}\mathcal{X}_2^*)^{-1} \\ &\times \mathcal{X}_2^{*\prime}\mathcal{Z}_2^*\left(P_{\mathcal{Z}_2}^*\ \text{vec}(Y')^* - \text{vec}\left[\hat{\Omega}_{0x}\hat{\Omega}_{xx}^{-1}(U_x'z_2 + T\hat{\Delta}_{xz_2})(Z_2'Z_2)^{-1}\right]\right) + o_p(1),\end{aligned}$$

and utilizing the above limit results we establish the second part of the theorem. □

8.8. Proof of Lemma 6.1:

$$
\begin{aligned}
T^{-1/2}\Xi^{+*}G_1 &= T^{-1/2}\Xi^{+}G_1 - T^{1/2}\check{\Delta}^{+}_{0z_1} \\
&= T^{-1/2}(\widetilde{U}'_{0\text{GMM}}Z_1 - T\check{\Delta}_{0z_1}) - T^{1/2}\widehat{\Omega}_{0a}\widehat{\Omega}^{-1}_{aa}[U'_aZ_1 - T\check{\Delta}_{za_1}] \\
&= o_p(1).
\end{aligned}
$$

The second term in the second line is of order $o_p(1)$ by Lemma 4.1. Under Assumption NF, which implies the (one-sided) exogeneity of $\{z_{1t}\}$, the first term is also $o_p(1)$. To see this, notice that $T^{1/2}\check{\Delta}_{0z_1} = T^{-1/2}\widetilde{U}'_{0\text{GMM}}Z_1 + o_p(1)$, which can be shown in the same way as Lemma 5.1. This proves part (a) of the lemma. For part (b),

$$
\begin{aligned}
T^{-1}\Xi^{+*}_2G_2 &= T^{-1}U'_0Z_2 - \widehat{\Omega}_{0a}\widehat{\Omega}^{-1}_{aa}T^{-1}U'_aZ_2 - \widehat{\Delta}^{+}_{0z_2} - (\widetilde{A}_{1\text{GMM}} - A_1)T^{-1}X'_1Z_2 \\
&\quad - T(\widetilde{A}_{2\text{GMM}} - A_2)T^{-1}X'_2Z_2 \\
&= (T^{-1}U'_0Z_2 - \widehat{\Delta}_{0z_2}) - \widehat{\Omega}_{0a}\widehat{\Omega}^{-1}_{aa}(T^{-1}U'_aZ_2 - \widehat{\Delta}_{az_2}) \\
&\quad - T(\widetilde{A}_{2\text{GMM}} - A_2)T^{-1}X'_2Z_2 + O_p(1/\sqrt{T}) \\
&\xrightarrow{d} \int_0^1 dB_0B'_{z_2} - \widehat{\Omega}_{0b}\widehat{\Omega}^{-1}_{bb}\int_0^1 dB_bB'_{z_2} - \int_0^1 dB_{0\cdot b}\widetilde{B}'_2\left(\int_0^1 \widetilde{B}_2\widetilde{B}'_2\right)^{-1}\int_0^1 B_2B'_{z_2} \\
&= \int_0^1 dB_{0\cdot b}B'_{z_2}\left\{I - \left(\int_0^1 B_{z_2}B'_{z_2}\right)^{-1}\int_0^1 B_{z_2}B'_2\left(\int_0^1 \widetilde{B}_2\widetilde{B}'_2\right)^{-1}\int_0^1 B_2B'_{z_2}\right\}.
\end{aligned}
$$

In the second line the stated error obtains because $(\widetilde{A}_1\text{GMM} - A_1) = O_p(1/\sqrt{T})$, while Lemma 4.1 and Theorem 5.2 establish the third line. Next we define

$$
D_{z2T} = (I_n \otimes T^{-2}Z'_2Z_2)^{-1/2}(T^{-2}Z'_2\mathcal{X}_2) \xrightarrow{d} \left\{I_n \otimes \left(\int_0^1 B_{z_2}B'_{z_2}\right)^{-1/2}\int_0^1 B_{z_2}B'_2\right\} = D_{z_2}.
$$

Then, recalling the definition $\widetilde{B}_2(r) = \int_0^1 B_2B'_{z_2}\left(\int_0^1 B_{z_2}B'_{z_2}\right)^{-1}B_{z_2}(r)$, we get

$$
\begin{aligned}
T^{-1/2}\left(\widehat{\Omega}_{00\cdot a}\otimes T^{-2}Z'_2Z_2\right)^{-1/2}\text{vec}(\Xi^{+*}_2) &\xrightarrow{d} (I_{nq} - P_{D_{z2}})\left\{\Omega^{-1/2}_{00}\otimes\left(\int_0^1 B_{z_2}B'_z\right)^{-1/2}\right\} \\
&\qquad \times \int_0^1 dB_{0\cdot b}\otimes B'_z \\
&\equiv MN(0,\ [I_{nq_2} - P_{D_{z2}}]).\quad \square
\end{aligned}
$$

8.9. Proof of Theorem 6.2: Using Lemma 6.1,

$$\begin{aligned}
\zeta &= \operatorname{vec}(\Xi^{+*})'\left(\widehat{\Omega}_{00\cdot a}\otimes Z'Z\right)^{-1}\operatorname{vec}(\Xi^{+*}) \\
&= \operatorname{vec}(\Xi^{+*}G_1)'\left(\widehat{\Omega}_{00\cdot a}\otimes Z_1'Z_1\right)^{-1}\operatorname{vec}(\Xi^{+*}G_1) \\
&\quad + \operatorname{vec}(\Xi^{+*}G_2)'\left(\widehat{\Omega}_{00\cdot a}\otimes Z_2'Z_2\right)^{-1}\operatorname{vec}(\Xi^{+*}G_2) + o_p(1) \\
&= \operatorname{vec}(\Xi^{+*}G_2)'\left(\widehat{\Omega}_{00\cdot a}\otimes Z_2'Z_2\right)^{-1}\operatorname{vec}(\Xi^{+*}G_2) + o_p(1) \\
&\xrightarrow{p} \chi^2_{n(q_2-m_2)}.
\end{aligned}$$

9. Table of Notation for Variable and Estimator Affixes

Symbol	Meaning
x_t	vector of regressors
x_{1t} $(= u_{1t})$	vector of the I(0) components of regressors
x_{2t}	vector of the I(1) components of regressors
u_{xt}	$= \Delta x_t$
u_{xt}	$= \Delta x_{2t}$
z_t	vector of instruments
z_{1t} $(= u_{z1t})$	vector of the I(0) components of instruments
z_{2t}	vector of the I(1) components of instruments
u_{zt}	$= \Delta z_t$
u_{2zt}	$= \Delta z_{2t}$
subscript "a"	elements corresponding to u_{xt} and u_{zt} jointly
subscript "b"	elements corresponding to u_{2t} and u_{z2t} jointly
superscript "+"	endogeneity correction with respect to $\{u_{at}\}$
superscript "*"	serial correlation correction
Ω_{ij}	long-run covariance of $\{u_{it}\}$ and $\{u_{jt}\}$
Δ_{ij}	one-sided long-run covariance of $\{u_{it}\}$ and $\{u_{jt}\}$
Δ^{+}_{ij}	$= \Delta_{ij} - \Omega_{ia}\Omega_{aa}^{-1}\Delta_{aj}$
affix "$\hat{}$"	estimator without FM corrections
affix "$\tilde{}$"	estimator without FM corrections

10. REFERENCES

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

Bowden, R. J. and D. A. Turkington (1984). *Instrumental Variables.* Cambridge: Cambridge University Press.

Corbae, D., S. Ouliaris and P. C. B. Phillips (1991). "A reexamination of the consumption function using frequency domain regressions," Cowles Foundation Discussion Paper No. 997.

Granger, C. W. J. and T.-H. Lee (1990). "Multicointegration," *Advances in Econometrics*, 8, 71–84.

Grenander, U. and M. Rosenblatt (1957). *Statistical Analysis of Stationary Time Series.* New York: Wiley.

Hannan, E. J. (1963). "Regression for time series." In M. Rosenblatt, ed., *Time Series Analysis.* New York: Wiley, pp. 17–37.

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

Hansen, B. E. (1990). "Estimation and inference in models of cointegration," *Advances in Econometrics*, 8, 225–248.

Hansen, L. P. (1982). "Large sample properties of generalized method of moments estimators," *Econometrica*, 50, 1029–1054.

Hansen, L. P. and T. J. Sargent (1991). *Rational Expectations Econometrics.* Boulder: Westview Press.

Hansen, L. P. and K. J. Singleton (1982). "Generalized instrumental variables estimation of nonlinear rational expectation models," *Econometrica*, 50, 1269–1286.

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

McCallum, B. T. (1979). "Topics concerning the formulation, estimation, and use of macroeconometric models with rational expectations," *American Statistical Association Proceedings of the Business and Economic Statistics Section*, 65–72.

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

Parzen, E. (1957). "On consistent estimates of the spectrum of a stationary time series," *Annals of Mathematical Statistics*, 28, 329–348.

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

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

Phillips, P. C. B. (1991b). "Optimal inference in cointegrated systems," *Econometrica*, 59, 283-306.

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. (1994). "Fully modified least squares and vector autoregressions," Cowles Foundation Discussion Paper, Yale University.

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 regression 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 (1988). "Asymptotic equivalence of ordinary least squares and generalized least squares in regressions with integrated regressors," *Journal of the American Economic Association*, 83, 111-115.

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

Sargan, J. D. (1958). "The estimation of economic relationships using instrumental variables," *Econometrica*, 26, 393-415.

Sargan, J. D. (1959). "The estimation of relationships with autocorrelated residuals by the use of the instrumental variables," *Journal of the Royal Statistical Society B*, 21, 91-105.

Sargan, J. D. (1988). *Lectures on Advanced Econometrics*. New York: Basil Blackwell.

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

White, H. (1984). *Asymptotic Theory for Econometricians*. New York: Academic Press.