NBER WORKING PAPER SERIES

EFFICIENT ESTIMATION OF A DYNAMIC ERROR-SHOCK MODEL

C. Hsiao*
P.M. Robinson**

Working Paper No. 157

National Bureau of Economic Research, Inc.
575 Technology Square
Cambridge, Massachusetts 02139

November 1976

*University of California and National Bureau of Economic Research. Research supported in part by National Science Foundation Grant SOC75-18919.

**Harvard University. Research supported in part by National Science Foundation Grant SOC75-13436.

Abstract

This paper is concerned with the estimation of the parameters in a dynamic simultaneous equation model with stationary disturbances under the assumption that the variables are subject to random measurement errors. The conditions under which the parameters are identified are stated. An asymptotically efficient frequency-domain class of instrumental variables estimators is suggested. The procedure consists of two basic steps. The first step transforms the model in such a way that the observed exogenous variables are asymptotically orthogonal to the residual terms. The second step involves an iterative procedure like that of Robinson [13].

Contents

1. Introduction

We assume the existence of an underlying economic system of the form

$$(1.1) \qquad \sum_{j=0}^{p} B_j(\xi_{t-j}-\mu_\xi) + \sum_{j=0}^{q} \Gamma_j(\chi_{t-j}-\mu_\chi) = \varepsilon_t \; , \quad t = 1,2,\ldots \; .$$

Here ξ_t, χ_t and ε_t are discrete vector-valued covariance-stationary processes of dimension G, K and G, respectively; they have mean vectors μ_ξ, μ_χ and zero, respectively, and satisfy $E\,\chi_t\varepsilon_s' = 0$, all t, s, the prime denoting transposition. The B_j and Γ_j are matrices (of dimensions $G \times G$ and $G \times K$ respectively) B_0 being nonsingular and all zeros of $\det\{\sum_{j=0}^{p} B_j z^j\}$ being outside the unit circle.

In this paper we assume in general that all the quantities in (1.1) are unknown or unobserved (except for certain elements of the B_j and Γ_j, knowledge of which will identify the model -- see below). The estimation of (1.1) in the case where one observes ξ_t and χ_t exactly for $t = 1,2,\ldots,T$, in the absence of stringent restrictions on the autocovariance structure of ε_t, has been considered by Hannan and Nicholls [8] (for the case $G = K = 1$), Hannan and Terrell [9] (for the case $p = q = 0$), Espasa and Sargan [1]. We relax this requirement by assuming that we observe

$$y_t = \xi_t + \zeta_t \; , \quad x_t = \chi_t + \eta_t \; , \quad t = 1,\ldots,T \; .$$

The ζ_t and η_t are $G \times 1$ and $K \times 1$ vector processes which satisfy $E\zeta_t = 0$, $E\zeta_t\zeta_s' = \delta_{ts}2\pi\Omega$, $E\eta_t = 0$, $E\eta_t\eta_s' = \delta_{ts}2\pi\Theta$, $E\xi_t\chi_s' = 0$, $E\zeta_t\eta_s' = 0$, $E\xi_t\eta_s' = 0$, δ_{ts} being Kronecker's delta. Clearly, μ_ξ and μ_χ are consistently and efficiently estimated by $\hat{\mu}_\xi = T^{-1}\sum_{t=1}^{T} y_t$, $\hat{\mu}_\chi = T^{-1}\sum_{t=1}^{T} x_t$ respectively. However, it will be apparent from knowledge of the classical errors-in-variables problem that standard estimators of the B_j, Γ_j, that would be asymptotically efficient if Θ were a null matrix, will not in general be consistent or efficient when Θ is non-null. We shall not be concerned with estimating Ω because under our assumptions it will not be identifiable. We do wish to estimate Θ, however, along with the B_j, Γ_j, by consistent and efficient methods. Since the η_t, like the ζ_t, play the role of white noise measurement errors, an a priori assumption that will often not be unreasonable, and will prevent too large an expansion of the parameter space over that in [8], [9], is that the elements of η_t are contemporaneously uncorrelated. Therefore, we assume throughout that Θ is a diagonal matrix. Moreover, we shall allow for the possibility that we know that one or more of the elements of χ_t is, almost surely, observed without error for all t, in which case we fix the corresponding diagonal element of Θ as zero. (We nowhere require Θ, or Ω, to be nonsingular). Therefore, we have achieved a generalization of the usual dynamic simultaneous equations model: prior information that an exogenous variable is error-free is equivalent to an exclusion constraint on a parameter. This will fit in well with our other constraints on the B_j, Γ_j, for we assume, for simplicity, that all these are also of the exclusion type (apart from a normalization and sign constraint on each equation).

Measurement errors with similar properties were considered by Goldberger [4,5], Geraci [3], Hsiao [10], etc. However, in their work, only the static case of (1.1) (the case $p = q = 0$) was considered, and the χ_t, ε_t were assumed to be white noise. The most significant difference is that we are principally concerned here with the case where χ_t is not white noise, when the identification problem may be solved by the use of lagged observable exogenous variables as instruments. We pay some attention also to the case where χ_t is white noise, when the lagged exogenous variables are useless as instruments and a basically different approach is required.

2. Identification and a Consistent Initial Estimator

When (1.1) is transformed in terms of the observables y_t, x_t, we have

$$\sum_{j=0}^{p} B_j(y_{t-j}-\mu_\xi) + \sum_{j=0}^{q} \Gamma_j(x_{t-j}-\mu_\chi) = u_t , \tag{2.1}$$

$$u_t = \varepsilon_t + \sum_{j=0}^{p} B_j\zeta_t + \sum_{j=0}^{q} \Gamma_j\eta_{t-j} . \tag{2.2}$$

We define the discrete Fourier transforms

$$w_x(\lambda) = (2\pi T)^{-1/2} \sum_{t=1}^{T} x_t e^{it\lambda}$$

$$w_y(\lambda) = (2\pi T)^{-1/2} \sum_{t=1}^{T} y_t e^{it\lambda}$$

$$w_u(\lambda) = (2\pi T)^{-1/2} \sum_{t=1}^{T} u_t e^{it\lambda} ,$$

and replace (2.1) by an asymptotic approximation to its Fourier transform,

$$B(\lambda)w_y(\lambda) + \Gamma(\lambda)w_x(\lambda) = w_u(\lambda) , \quad \lambda \neq 0 , \tag{2.3}$$

where

$$B(\lambda) = \sum_{j=0}^{p} B_j e^{ij\lambda} , \quad \Gamma(\lambda) = \sum_{j=0}^{q} \Gamma_j e^{ij\lambda}$$

omitting the terms in μ_ξ, μ_χ which are $O(T^{-1/2})$ when $\lambda \neq 0$. The transformed system can be treated as a contemporaneous model. Thus, because $\{\varepsilon_t\}$ and $\{\zeta_t\}$ are incoherent of $\{x_t\}$, but $\{\eta_t\}$ is not, $w_u(\lambda)$ can be decomposed into a sum of two orthogonal components, one of which (depending upon $\{\eta_t\}$) is correlated with $w_x(\lambda)$, and the other (depending upon $\{\varepsilon_t\}$, $\{\zeta_t\}$) is not. Therefore, the measurement errors in $\{y_t\}$ can be amalgamated with $\{\varepsilon_t\}$ to produce a composite, stationary, residual term. It would be possible to identify Ω if, for example, $\varepsilon_t = 0$, almost surely, all t, whence (1.1) is a homogeneous structural relation between χ_t, ξ_t, and u_t is a moving average sequence of order $\max(p,q)$, but we shall not do this; we prefer to allow for the presence of a stationary ε_t in (1.1), to possibly represent exogenous variables that should have been included in (1.1). We could, indeed, have omitted explicit reference to a measurement error pertaining to ξ_t, and indeed the assumption that this measurement error is white noise essentially plays no role in our results. In any case, if $\Theta = 0$, there is no problem in consistently estimating the B_j, Γ_j, but merely an efficient estimation problem caused by the moving average in ζ_t in (2.2): if ε_t is a moving average sequence of order r then the residual term is a moving average of order $\max(p,r)$, and a vector extension of the methods of Hannan and Nicholls [8] may be appropriate. We are concerned only with the case $\Theta \neq 0$, however.

For simplicity, we assume that $B(\lambda)$ and $\Gamma(\lambda)$ are relatively left prime, so that the redundancy in the specification can be eliminated (Hannan [7]) and the a priori information on the B_j, Γ_j is entirely in the form of exclusion restrictions. We also assume that

$$C_\chi(j) = E(\chi_t-\mu_\chi)(\chi_{t+j}-\mu_\chi)' \ , \quad j = 0,1,\ldots,r$$

are nonsingular and unrestricted for some $r > 0$. Furthermore, because of the requirement that the instrumental variable estimates discussed in §3 do not involve a singular matrix, we follow Fisher [2, Condition 6.2.1] in assuming that the $pG+(q+1)K$ elements of $\bar{\eta}_{t-1},\ldots,\bar{\eta}_{t-p},x_t,\ldots,x_{t-q}$ are not connected by any linear identities, where $\bar{\eta}_t = (\sum B_j L^j)^{-1}(\sum \Gamma_j L^j)x_t$, with L the lag operator. A sufficient condition for this to hold is: (a) all zeros $\det\{\sum_{j=0}^{p} B_j z^j\}$ and $\det\{\sum_{j=0}^{q} \Gamma_j z^j\}$ are outside the unit circle; (b) $\{\sum_{j=0}^{p} B_j z^j\}$, $\{\sum_{j=0}^{q} \Gamma_j z^j\}$ are relatively left prime; i.e., they have I_G as greatest common left divisor; and (c) $\operatorname{rank}(B_p,\Gamma_q) = G$. Then, by the same reasoning as in Hsiao [11], one can show that the number of excluded predetermined variables be at least as great as the number of included joint dependent variables less one is what is necessary for the identification of the i^{th} equation. If we let one element in each row of B_0 be prescribed as unity the necessary and sufficient condition to locally identify (2.1) is that at least $(G-1)$ zeros be prescribed in each row of

$$A = [\; B_0 \quad B_1 \quad \cdots \quad B_p \quad \Gamma_0 \quad \cdots \quad \Gamma_q \;]$$

and the rank of each submatrix of A obtained by taking the columns of A with prescribed zeros in a certain row is $(G-1)$.

If x_t is a white noise, i.e., $C_x(j) = 0$ for $j > 0$, we need a much stronger condition to identify the unknown parameters of the i^{th} equation of (2.1). In particular, in addition to the condition that the number of excluded predetermined variables has to be at least as great as the number of included joint dependent variables plus the number of unknown measurement error variances (associated with the included current and lagged exogenous variables) less one, we need additional conditions on the way the included current or lagged exogenous appear in the i^{th} behavioral equation. Let h_j denote the number of j^{th} lagged included exogenous variables which are not measured exactly. We arrange them in increasing order so that $h_j^i \geq h_{j'}^{i-1}$, for $i = 1,2,\ldots,q$. That is, the j^{th} lagged included exogenous variables contain more inaccurately measured variables than j'^{th} lagged variables. Then this additional necessary condition is that the number of

j^{th} lagged excluded exogenous variables be at least as great as the number of additional unknown measurement error variances which were not introduced by $h^0_{j_0},\ldots,h^{i-1}_{j,i-1}$. It seems unlikely that χ_t will be white noise, in the context of a time series model such as (1.1). Given the assumption that η_t is white noise, χ_t will be white noise if and only if x_t is white noise, and the latter question can easily be resolved by looking at the data.

Another way of stating the identification conditions is that there exists a sufficient number of instrumental variables. Thus, provided the model is identified we can apply an instrumental variables method equation by equation to obtain consistent estimates of the B_j's and Γ_j's and Θ (see below). The standard instruments will be the current or lagged exogenous variables which do not appear in the equation under consideration. Since the measurement errors among the exogenous variables are assumed to be uncorrelated, the excluded exogenous variables can be used as instruments irrespective of whether they are observed exactly.

Theoretically, all the lagged exogenous variables can be used as instruments and the addition of new instrumental variables will increase the efficiency unless the partial correlation between each variable in the relationship and the new instrumental variables is zero after the effects on the other instrumental variables have been allowed for. In practice, if the first few instruments are well chosen, there may be no great advantage in increasing the number of instruments. Sargan [15] has shown that the estimates have large biases if the number of instrumental variables becomes too large. Actually, he suggested that the number of instruments should not be greater than $T/20$.

3. Efficient Estimation

We assume that the normalization conditions on the equations (1.1) are that the diagonal elements of B_0 are all units. We write

$$B(\lambda) - I_G = B(e_p(\lambda) \otimes I_G) , \quad \Gamma(\lambda) = \Gamma(e_q(\lambda) \otimes I_K) \tag{3.1}$$
$$B = [B_0 - I_G, B_1, \ldots, B_p] , \quad \Gamma = [\Gamma_0, \Gamma_1, \ldots, \Gamma_q] ,$$
$$e_p(\lambda)' = [1, e^{i\lambda}, \ldots, e^{ip\lambda}] , \quad e_q(\lambda)' = [1, e^{i\lambda}, \ldots, e^{iq\lambda}] .$$

All prior constraints on B and Γ are zero ones, and we incorporate them in a way like that in Robinson [12]. Suppose there are G_1 zero constraints on B. Then the unconstrained ones may be written as the $G^2(p+1) \times 1$ vector $\beta = L_1 \mathrm{vec}(B)$, where L_1 is obtained from $I_{G^2(p+1)}$ by eliminating rows corresponding to zero elements. Likewise, if there are G_2 zero constraints on Γ we write the unconstrained parameters as $\gamma = L_2 \mathrm{vec}(\Gamma)$, where L_2 is obtained from $I_{GK(q+1)}$ by eliminating rows corresponding to zero elements. Also, we write $\theta = L_3 \mathrm{vec}(\Theta)$, where L_3 is obtained from the $(K-F) \times K^2$ matrix obtained from I_{K^2} by eliminating rows corresponding to the off-diagonal elements and the F, $0 \le F \le K$, <u>a priori</u> zero diagonal elements of Θ.

Since all processes are covariance stationary, we define the auto-covaraince and cross-autocovariance matrices

$$C_x(j) = E\{(x_t - Ex_t)(x_{t+j} - Ex_t)'\} , \quad C_{yx}(j) = E\{(y_t - Ey_t)(x_{t+j} - Ex_t)'\} ,$$

and we assume the existence of the spectral and cross-spectral density matrices

$$f_x(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} C_x(j)e^{-ij\lambda}\ , \quad f_{yx}(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} C_{yx}(j)e^{-ij\lambda}\ , \quad -\pi < \lambda \leq \pi\ .$$

Similar notation will be used to denote second order properties pertaining to other sequences. We note that $f_\eta(\lambda) = \Theta$, $-\pi < \lambda \leq \pi$.

Now it is known that, under fairly wide conditions (see Hannan [6, Chapter IV]) the limiting covariance between the discrete Fourier transforms of two stationary sequences is the cross-spectral density of the sequences. Thus

(3.1) $$\lim_{T\to\infty} Ew_u(\lambda)w_x(\lambda)^* = f_{ux}(\lambda) = \Gamma(\lambda)\Theta\ , \qquad \lambda \neq 0\ .$$

We therefore rewrite (2.3) as

(3.2) $$\begin{aligned} w_y(\lambda) &= (I_G - B(\lambda))w_y(\lambda) - \Gamma(\lambda)(I_K - \Theta f_x(\lambda)^{-1})w_x(\lambda) + w_{\tilde{u}}(\lambda)\ , \\ w_{\tilde{u}}(\lambda) &= w_u(\lambda) - \Gamma(\lambda)\Theta f_x(\lambda)^{-1} w_x(\lambda)\ , \end{aligned}$$

when $f_x(\lambda)$ is nonsingular, and where I_G is the $G \times G$ identity matrix. Now because of (3.1) and because

$$\lim_{T\to\infty} Ew_x(\lambda)w_x(\lambda)^* = f_x(\lambda)\ , \qquad \lambda \neq 0\ ,$$

we have

$$\lim_{T\to\infty} Ew_{\tilde{u}}(\lambda)w_x(\lambda)^* = f_{ux}(\lambda) - \Gamma(\lambda)\Theta f_x(\lambda)^{-1} f_x(\lambda) = 0\ .$$

Thus, (3.2) possesses (asymptotically) the classical property of orthogonality between the "exogenous variable" $w_x(\lambda)$ and the "residual" $w_{\tilde{u}}(\lambda)$.

We now rewrite (3.2) as

$$w_y(\lambda) = -B(e_p(\lambda) \otimes w_y(\lambda)) - \Gamma(e_q(\lambda) \otimes w_x(\lambda)) + \Gamma(e_q(\lambda) \otimes I_K)\Theta f_x(\lambda)^{-1} w_x(\lambda) + w_{\tilde{u}}(\lambda) \ . \tag{3.3}$$

We use the relation $\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$ to rewrite (3.3) as

$$w_y(\lambda) = X(\lambda)L'\delta + w_{\tilde{u}}(\lambda) \ , \tag{3.4}$$

$$\delta' = (\beta', \gamma', \theta') \ ,$$

$$X(\lambda) = [-(e_p'(\lambda) \otimes w_y'(\lambda) \otimes I_G), -(e_q'(\lambda) \otimes w_x'(\lambda) \otimes I_K), w_x'(\lambda) f_x'(\lambda)^{-1} \otimes \Gamma(\lambda)] \ ,$$

$$L = \begin{bmatrix} L_1 & 0 & 0 \\ 0 & L_2 & 0 \\ 0 & 0 & L_3 \end{bmatrix} .$$

We shall consider (3.4) for T equally-spaced values of λ over $(-\pi, \pi]$, denoted $\omega_\ell = 2\pi\ell/T$, $-T/2 < \ell \leq [T/2]$. Now it is known that under fairly general conditions (see Hannan [6, Chapter IV]) the $w_{\tilde{u}}(\omega_\ell)$ are asymptotically independent (complex) normally distributed, with zero means and covariance matrix $f_{\tilde{u}}(\omega_\ell)$. Now, $f_x(\lambda)$ and $\Gamma(\lambda)$, in $X(\lambda)$, are unknown but consistent estimation of both is possible. Thus, an asymptotically efficient instrumental variables method will be possible after we find an appropriate instrument for $w_y(\lambda)$, and a consistent estimate of

$$f_{\tilde{u}}(\lambda) = \lim_{T \to \infty} E w_{\tilde{u}}(\lambda) w_{\tilde{u}}^*(\lambda) \ .$$

The method we propose follows that of Robinson [12]. One major difference is that in Robinson [12] no measurement error was allowed for. A second major difference is that in [13] a system of <u>differential</u> equations was to be estimated. The discrete approximation used there led, in the frequency domain, to matrix polynomials in $i\lambda$ rather than, as here, in $e^{i\lambda}$. As noted in [13], the method applies to difference equation models if one replaces $i\lambda$ by $e^{i\lambda}$. A third departure from [13] lies in the identification conditions. In both cases, a fundamental feature of the identification problem is an aliasing problem connected to the exogenous variables. However, in [13] the problem is concerned with identifying a continuous-time signal from knowledge of a discrete one, whereas here it is concerned with extracting a signal in the presence of noise. A fourth difference from [13] is that here an initial consistent estimate of γ, as well as of β, is essential.

As in [13], iteration may well be desirable, and so we describe our procedure as if it were iterative, although iteration produces no improvement in efficiency. Our procedure is efficient in the sense that the limiting covariance matrix of our estimates is the same as that of maximum-likelihood estimates based on Gaussian $w_{\tilde{u}}(\lambda)$. Efficient estimates could also be obtained by a minimum-distance procedure, using a suitable metric (like that in Robinson [14]). They could also be obtained by replacing $w_y(\lambda)$ in (3.2) by its instrument, and then using a type of generalized least squares (like that in Hannan and Terrell [9]). With all these procedures, again, iteration is probably desirable, but providing they are initiated with consistent estimates, and providing the type of iterative step taken is appropriate, asymptotically efficient estimates will result after a single step. The reason we concentrate on our procedure is that it seems among the simplest to compute and to describe.

Before describing the method, some interim computations must be detailed. For an integer M, much less than T (see below) we introduce the $2M$ sets

$$\mathcal{B}_m = \{\lambda |\ \lambda_m - \tfrac{\pi}{2M} < \lambda \le \lambda_m + \tfrac{\pi}{2M};\ \lambda_m = \tfrac{\pi m}{M}\} \ , \qquad |m| \le M-1$$

$$\mathcal{B}_M = \{\lambda |\ -\pi < \lambda \le -\pi(1 - \tfrac{1}{2M}),\ \pi(1 - \tfrac{1}{2M}) < \lambda \le M\} \ ,$$

with $\lambda = 0$ omitted from $\mathcal{B}_0$. Then for all m, $-M-1 \le m \le M$, we define

$$\text{(3.5)} \qquad \begin{aligned} &\hat{f}_y(\lambda_m) = \frac{2M}{T}\sum_{\mathcal{B}_m} w_y(\omega_\ell)w_y(\omega_\ell)^* \ , \quad \hat{f}_x(\lambda_m) = \frac{2M}{T}\sum_{\mathcal{B}_m} w_x(\omega_\ell)w_x(\omega_\ell)^* \ , \\ &\hat{f}_{yx}(\lambda_m) = \frac{2M}{T}\sum_{\mathcal{B}_m} w_y(\omega_\ell)w_x(\omega_\ell)^* \ , \quad \hat{f}_{xy}(\lambda_m) = \hat{f}_{yx}(\lambda_m)^* \ , \end{aligned}$$

where the sums are over $\omega_\ell \in \mathcal{B}_m$. We assume that $2M/T$ is sufficiently small for the number of ω_ℓ in each $\mathcal{B}_m$ to be at least $\max(G,K)$. In that case, $\hat{f}_x(\lambda_m)$ will in general be nonsingular.

Now denote by $\hat{\Gamma}^{(j)}(\lambda)$, $\hat{\Theta}^{(j)}$ the estimates of $\Gamma(\lambda)$, Θ obtained on the j^{th} iterative step, with $\hat{\Gamma}^{(0)}(\lambda)$, $\hat{\Theta}^{(0)}$ the consistent ones referred to in §2 and below. Then define $\hat{\Gamma}^{(j)}$ so that $\hat{\Gamma}^{(j)}(\lambda) = \hat{\Gamma}^{(j)}(e_q(\lambda) \otimes I_K)$. Returning to (3.2), we consider the first-order Taylor approximation

$$\text{(3.6)} \qquad \begin{aligned} \Gamma(\lambda)(I_K - \Theta f_x(\lambda)^{-1}) &\simeq \hat{\Gamma}^{(j)}(\lambda)(I_K - \hat{\Theta}^{(j)} f_x(\lambda)^{-1}) \\ &\quad + \Gamma(\lambda)(I_K - \hat{\Theta}^{(j)} f_x(\lambda)^{-1}) - \hat{\Gamma}^{(j)}(\lambda)(I_K - \Theta f_x(\lambda)^{-1}) \\ &= \Gamma(\lambda)(I_K - \hat{\Theta}^{(j)} f_x(\lambda)^{-1}) + \hat{\Gamma}^{(j)}(\lambda)(\Theta - \hat{\Theta}^{(j)}) f_x(\lambda)^{-1} \ . \end{aligned}$$

We replace λ by ω_ℓ, and then replace $f_x(\omega_\ell)$ by $\hat{f}_x(\lambda_m)$, where $\omega_\ell \in \mathcal{B}_m$. Then we approximate (3.2), with $\lambda = \omega_\ell$, by

$$w_y(\omega_\ell) + \hat{\Gamma}^{(j)}(\lambda_m)\hat{\Theta}^{(j)}\hat{f}_x(\lambda_m)^{-1}w_x(\omega_\ell) = X^{(j)}(\omega_\ell,\lambda_m)L'\delta + w_{\tilde{u}}(\omega_\ell) \ , \qquad \omega_\ell \in \mathcal{B}_m \ ,$$

$$X^{(j)}(\omega,\lambda)' = \begin{bmatrix} -e_p(\omega) \otimes w_y(\omega) \otimes I_G \\ e_q(\omega) \otimes (I_K - \hat{\Theta}^{(j)}\hat{f}_x(\lambda)^{-1})w_x(\lambda) \otimes I_G \\ \hat{f}_x(\lambda)^{-1}w_x(\omega) \otimes \hat{\Gamma}^{(j)}(\lambda)' \end{bmatrix}$$

An estimate of $f_{\tilde{u}}$ is needed. We first put

$$\Delta(\lambda) = -\psi(\lambda)(I_K-\Theta f_x(\lambda)^{-1}) , \quad \psi(\lambda) = B(\lambda)^{-1}\Gamma(\lambda) .$$

The "solution" of (3.2) is thus

$$w_y(\lambda) = \Delta(\lambda)w_x(\lambda) + w_v(\lambda) , \quad w_v(\lambda) = B(\lambda)^{-1}w_{\tilde{u}}(\lambda) , \tag{3.7}$$

where $\lim_{T\to\infty} Ew_x(\lambda)w_v(\lambda)^* = 0$. We have, therefore,

$$f_{yx}(\lambda) = \Delta(\lambda)f_x(\lambda) , \tag{3.8}$$

$$\begin{aligned} f_v(\lambda) &= f_y(\lambda) - \Delta(\lambda)f_{xy}(\lambda) - f_{yx}(\lambda)\Delta(\lambda)^* + \Delta(\lambda)f_{yx}(\lambda)\Delta(\lambda)^* \\ &= f_y(\lambda) - f_{yx}(\lambda)f_x(\lambda)^{-1}f_{xy}(\lambda) \end{aligned} \tag{3.9}$$

$$f_{\tilde{u}}(\lambda) = B(\lambda)f_{\tilde{u}}(\lambda)B(\lambda)^* . \tag{3.10}$$

We thus define, using (3.9), (3.10),

$$\hat{f}_v^{(0)}(\lambda_m) = \hat{f}_y(\lambda_m) - \hat{f}_{yx}(\lambda_m)\hat{f}_x(\lambda_m)^{-1}\hat{f}_{xy}(\lambda_m) ,$$

$$\hat{f}_{\tilde{u}}^{(0)}(\lambda_m) = \hat{B}^{(0)}(\lambda_m)\hat{f}_v^{(0)}(\lambda_m)B^{(0)}(\lambda_m)^* ,$$

where $\hat{B}^{(j)}(\lambda) = \hat{B}^{(j)}(e_q(\lambda)\otimes I_K)$, $\hat{B}^{(j)}$ being the estimate of B from the j^{th} step. On later iterative steps an estimate of $\Delta(\lambda)$ that incorporates the prior constraints may be used. If $\hat{\Theta}^{(j)}$ is the estimate of Θ from the j^{th} step, define

$$\begin{aligned} \hat{\Delta}^{(j)}(\lambda) &= -\hat{\psi}^{(j)}(\lambda)(I_K-\hat{\Theta}^{(j)}\hat{f}_x(\lambda)^{-1}) , \\ \hat{\psi}^{(j)}(\lambda) &= \hat{B}^{(j)}(\lambda)^{-1}\hat{\Gamma}^{(j)}(\lambda) . \end{aligned} \tag{3.11}$$

As already noted, the $\hat{B}_j^{(0)}$, $\hat{\Gamma}_j^{(0)}$ may be one of many consistent instrumental variables estimates. A consistent $\hat{\Theta}^{(0)}$ may be found by applying minimum distance methods to (3.2), after replacing $B(\lambda)$, $\Gamma(\lambda)$, $f_x(\lambda)$ by $\hat{B}^{(0)}(\lambda)$, $\hat{\Gamma}^{(0)}(\lambda)$, $\hat{f}_x(\lambda)$. Then from (3.8), (3.10), put

$$\hat{f}_v^{(j)}(\lambda_m) = \hat{f}_y(\lambda_m) - \hat{\Delta}^{(j)}(\lambda_m)\hat{f}_{xy}(\lambda_m) - \hat{f}_{yx}(\lambda_m)\hat{\Delta}^{(j)}(\lambda_m)^* + \hat{\Delta}^{(j)}(\lambda_m)\hat{f}_x(\lambda_m)\hat{\Delta}^{(j)}(\lambda_m)^* ,$$

$$\hat{f}_{\tilde{u}}^{(j)}(\lambda_m) = \hat{B}^{(j)}(\lambda_m)\hat{f}_v^{(j)}(\lambda_m)\hat{B}^{(j)}(\lambda_m)^*$$

for $j \geq 1$.

We now discuss the instrument for $w_y(\lambda)$. From (3.7) we would like to use $\Delta(\lambda)w_x(\lambda)$. The instrument for $w_y(\omega_\ell)$ on the j^{th} step will therefore be

(3.12) $$\hat{\Delta}^{(j)}(\omega_m)w_x(\omega_\ell)$$

where $\hat{\Delta}^{(j)}(\lambda)$ is (3.11) for $j \geq 1$, and

(3.13) $$\hat{\Delta}^{(0)}(\lambda) = \hat{f}_{yx}(\lambda)\hat{f}_x(\lambda)^{-1}$$

(cf. (3.8)). As noted earlier, we could replace $w_y(\omega_\ell)$ by (3.12) in (3.6), and then use GLS like in [9]. Our procedure might be preferred in that it seems to involve one less approximation. On the other hand, the GLS approach has the advantage in that, unlike ours, it involves the inversion of a <u>symmetric</u> matrix. (Of course, our matrix <u>converges</u> to a symmetric matrix.) Both types of procedure reduce, essentially, to three stage least squares (3SLS) in the classical simultaneous equations case $p = q = K' = 0$, $f_{\tilde{u}}(\lambda)$ constant, <u>a priori</u>. As noted in Robinson [13], (3.13) is a narrow-band version of the reduced form estimate used in 3SLS.

We are now able to define our efficient estimates,

(3.14) $$\hat{\delta}^{(j+1)} = (LD^{(j)}L')^{-1}Ld^{(j)} , \quad j \geq 0 ,$$

where

$$D^{(j)} = \frac{1}{T}\sum_m \sum_{\omega_\ell \in B_m} Z^{(j)}(\omega_\ell,\lambda_m)^* X^{(j)}(\omega_\ell,\lambda_m)$$

$$d^{(j)} = \frac{1}{T}\sum_m \sum_{\omega_\ell \in B_m} Z^{(j)}(\omega_\ell,\lambda_m)^* w_y(\omega_\ell)$$

$$Z^{(j)}(\omega,\lambda)' = \begin{bmatrix} e_p(\omega)\otimes\hat{\Delta}^{(j)}(\lambda)w_x(\omega)\otimes\hat{f}_{\tilde{u}}^{(j)}(\lambda)^{-1} \\ e_q(\omega)\otimes(I_K-\hat{\Theta}^{(j)}\hat{f}_x(\lambda)^{-1})w_x(\omega)\otimes\hat{f}_{\tilde{u}}^{(j)}(\lambda)^{-1} \\ \hat{f}_x(\lambda)^{-1}w_x(\omega)\otimes\hat{\Gamma}^{(j)}(\lambda)'\hat{f}_{\tilde{u}}^{(j)}(\lambda)^{-1} \end{bmatrix}$$

To prove asymptotic properties, conditions additional to those in §§1,2 are needed. We describe these as follows. Let $\{\eta_t\}$, $\{\zeta_t\}$ be mutually independent sequences of independent, identically distributed (i.i.d.) random vectors. Let $\{\chi_t\}$, $\{\varepsilon_t\}$ be mutually independent sequences, independent also of $\{\xi_t\}$, $\{\zeta_t\}$, with representations

$$\chi_t = \sum_{j=-\infty}^{\infty} a_j \tau_{t-j}\ , \quad \varepsilon_t = \sum_{j=-\infty}^{\infty} b_j \nu_{t-j}\ , \tag{3.15}$$

where $\{\tau_t\}$, $\{\nu_t\}$ are i.i.d. sequences with finite second moments. Also, let the spectra $f_\chi(\lambda)$, $f_\varepsilon(\lambda) \in \text{Lip}\ \alpha$, $\alpha > \frac{1}{2}$ (i.e. satisfy Lipschitz conditions of order greater than one half, see Zygmund [16]). (This is slightly stronger than assuming $\sum_{-\infty}^{\infty}\|a_j\| < \infty$, $\sum_{-\infty}^{\infty}\|b_j\| < \infty$, where $\|\cdot\|$ is the Euclidean norm.) Therefore, $f_x(\lambda) \in \text{Lip}\ \alpha$, $\alpha > \frac{1}{2}$. Thus, assuming $\det\{f_\chi(\lambda)\}$ is bounded away from zero, $f_x(\lambda)^{-1} \in \text{Lip}\ \alpha$, $\alpha > \frac{1}{2}$ also, since it is a rational function of the elements of $f_\chi(\lambda)$. Therefore, $\Delta(\lambda) \in \text{Lip}\ \alpha$, $\alpha > \frac{1}{2}$. Also, from (3.1), (3.2) it may be inferred that

$$f_{\tilde{u}}(\lambda) = f_u(\lambda) - \Gamma(\lambda)\Theta f_x(\lambda)^{-1}\Theta\Gamma(\lambda)\ .$$

We note, parenthetically, that this is nonnegative definite, as is apparent on rewriting it as

$$f_{\tilde{u}}(\lambda) = f_{\varepsilon}(\lambda) + B(\lambda)\Omega B(\lambda)^{*} + \Gamma(\lambda)\left(f_{\chi}(\lambda) + f_{\chi}(\lambda)f_{x}(\lambda)^{-1}f_{\chi}(\lambda)\right)\Gamma(\lambda)^{*} .$$

Because $\hat{f}_{\tilde{u}}^{-1}$ is involved in our estimates, we must assume also that $f_{\tilde{u}}$ is positive definite. Now from the above, we may infer that $f_{u}(\lambda) \in \text{Lip}\ \alpha$, $\alpha > \frac{1}{2}$, and thus $f_{\tilde{u}}(\lambda) \in \text{Lip}\ \alpha$, $\alpha > \frac{1}{2}$. It follows from Zygmund [16] that $f_{\tilde{u}}(\lambda)$ may be written as

$$\left(\sum_{-\infty}^{\infty} c_j e^{ij\lambda}\right)\left(\sum_{-\infty}^{\infty} c_j e^{ij\lambda}\right)^{*} , \quad \sum_{-\infty}^{\infty} \| c_j \| < \infty .$$

Thus, in a mean square sense, we may take the time domain structural residual process, orthogonal to $\{x_t\}$, to be $\tilde{u}_t$, with spectrum $f_{\tilde{u}}$ and representation

$$\tilde{u}_t = \sum_{-\infty}^{\infty} c_j \rho_{t-j} , \quad \sum_{-\infty}^{\infty} \| c_j \| < \infty ,$$

with i.i.d. $\{\rho_t\}$, the last fact being inferred from the fact that $\{\tau_t\}$, $\{\nu_t\}$ are i.i.d. sequences. Further, in the solution of the system,

$$y_t = \sum_{-\infty}^{\infty} \Delta_j x_{t-j} + v_t \tag{3.16}$$

(i.e. the time domain version of (3.7)), v_t must have a representation

$$v_t = \sum_{-\infty}^{\infty} g_j \phi_{t-j} , \quad \sum_{-\infty}^{\infty} \| g_j \| < \infty ,$$

where $\{\phi_t\}$ is an i.i.d. sequence. Thus, because $\{x_t\}$, $\{v_t\}$ are strictly stationary and ergodic sequences with continuous spectra, because also

$Ex_s v_t' = 0$, all s, t, and because $\Delta(\lambda) = \sum_{-\infty}^{\infty} \Delta_j e^{ij\lambda} \in \mathrm{Lip}\,\alpha$, $\alpha > \frac{1}{2}$, we can, essentially, analyze the estimates in terms of the model (3.16), for which theorems in [6], [13] are available. A small additional condition is needed in order for the error of approximation of the Fourier transform of (3.16) by (3.12) to be asymptotically negligible in the central limit theorem. Let η_t have finite fourth moments and cross moments and let the fourth cumulant functions of all elements of $x_{t_1}, x_{t_2}, x_{t_3}, x_{t_4}$ be finite and expressible as the (trivariate) Fourier transform of a continuous function for all t_1, t_2, t_3, t_4. Then x_t also possesses the latter property.

It should be noted that the theorem below would hold under weaker conditions than those above; in particular our i.i.d. assumption could be relaxed. We have used these conditions for simplicity of exposition, and because the weaker conditions would be unfamiliar to many readers.

We define a matrix

$$D = \begin{bmatrix} D_{11} & D_{12} & D_{13} \\ D_{12}' & D_{22} & D_{23} \\ D_{13}' & D_{23}' & D_{33} \end{bmatrix},$$

where the partitions are $(p+1)G^2 : (q+1)GK : K^2$, and

$$D_{11} = \frac{1}{2\pi} \int_{-\pi}^{\pi} e_p(-\lambda) e_p'(\lambda) \otimes \Delta(\lambda) f_x(-\lambda) \Delta(\lambda)^* \otimes f_{\tilde{u}}(\lambda)^{-1} d\lambda \;,$$

$$D_{12} = \frac{1}{2\pi} \int_{-\pi}^{\pi} e_p(-\lambda) e_q'(\lambda) \otimes \Delta(\lambda) f_x(-\lambda) \otimes f_{\tilde{u}}(\lambda)^{-1} d\lambda \;,$$

$$D_{13} = -\frac{1}{2\pi} \int_{-\pi}^{\pi} e_p(-\lambda) \otimes \Delta(\lambda) \otimes f_{\tilde{u}}(\lambda)^{-1} \Gamma(\lambda) d\lambda \;,$$

$$D_{22} = \frac{1}{2\pi}\int_{-\pi}^{\pi} e_q(-\lambda)e_q'(\lambda)\otimes f_x(-\lambda)f_x^{-1}(-\lambda)f_x(-\lambda)\otimes f_{\tilde{u}}(\lambda)^{-1}d\lambda \ ,$$

$$D_{23} = -\frac{1}{2\pi}\int_{-\pi}^{\pi} e_q(-\lambda)\otimes f_x(-\lambda)f_x^{-1}(-\lambda)\otimes f_{\tilde{u}}(\lambda)^{-1}\Gamma(\lambda)d\lambda \ ,$$

$$D_{33} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f_x(-\lambda)^{-1}\otimes \Gamma(\lambda)^* f_{\tilde{u}}(\lambda)^{-1}\Gamma(\lambda)d\lambda \ .$$

Theorem. Under the above conditions, there exists some sequence $M = M(T)$ increasing as $T \to \infty$, such that $\hat{\delta}^{(j)} \to \delta$ almost surely (a.s.) and $T^{1/2}(\hat{\delta}^{(j)}-\delta)$ has a limiting multivariate normal distribution with zero means and covariance matrix $(LDL')^{-1}$, where D is strongly consistently estimated by $\hat{D}^{(j)}$, $j \geq 0$.

The proof will not be given in detail, as it resembles theorems in [8], [9], [13], [14]. In any case, the theorem need be proved only for $j = 1$. We first deal with consistency. $\hat{f}_x(\lambda_m)$, $\hat{f}_{\tilde{u}}^{(0)}(\lambda_m)$, $\hat{\Gamma}^{(0)}(\lambda_m)$, $\hat{\Delta}^{(0)}(\lambda_m)$ were replaced by $f_x(\omega_\ell)$, $f_{\tilde{u}}(\omega_\ell)$, $\Gamma(\omega_\ell)$, $\Delta(\omega_\ell)$, $\omega_\ell \in B_m$, strong consistency certainly follows. However, we note that under our conditions, including those described in §2, the initial estimates will be strongly consistent. Also, for some $M(T)$ increasing with T,

$$\hat{f}_x(\lambda) \longrightarrow f_x(\lambda) \ , \quad \hat{f}_{yx}(\lambda) \longrightarrow f_{yx}(\lambda) \ , \quad f_{\tilde{u}}(\lambda) \longrightarrow f_{\tilde{u}}(\lambda)$$

a.s. and uniformly in λ, where the band is roughly centered on, and degenerates to, λ (see [14]). Because of these results, it follows that there is an $M(T)$ such that the above replacement is possible and so $\hat{\delta}^{(1)} \to \delta$. Next we consider asymptotic normality. In this case, we can show that we may replace $\hat{f}_x(\lambda_m)$, $\hat{f}_{\tilde{u}}(\lambda_m)$, $\hat{\Gamma}^{(0)}(\lambda_m)$, $\hat{\Delta}^{(0)}(\lambda_m)$ by the a.s. limits as $T \to \infty$ but M stays fixed. Then in this situation

$T^{1/2}(\hat{\delta}^{(1)}-\delta)$ can be shown to be asymptotically multivariate normal, from [14]. On increasing M, then, the covariance matrix converges to $(LDL')^{-1}$.

We note that LDL' is essentially the limit of the information matrix based on Gaussian $w_{\tilde{u}}(\omega_\ell)$. Therefore, its nonsingularity requires local identifiability of the model. Note that, if $\psi(\lambda)$ is identifiable, Θ is identifiable by the relation

$$f_{yx}(\lambda) + \psi(\lambda)f_x(\lambda) = \psi(\lambda)\Theta \quad . \tag{3.17}$$

4. Comments

1) An alternative frequency domain approach to the problem of dealing with measurement error is that of simply eliminating from one's estimate those frequencies that seem likely to have a small signal-to-noise ratio, usually high ones (see [9], for example). This approach has the advantage over ours of making no explicit assumptions about the autocovariance structure of the measurement error, and of being somewhat easier computationally. However, the portion of the frequency band with a small signal-to-noise ratio may be rather large, and so if all these frequencies are omitted the resulting estimate may have rather large variance. Moreover, there may be no frequencies for which the signal-to-noise ratio is really large; this may be the case when, as we assume, the measurement error has uniform spectrum. On the other hand, our approach would often seem to be more efficient, for F will tend to be small relative to the number of other parameters. Its disadvantage, however, lies in the very strong assumptions about f_η, which, if invalid, might lead to serious bias. The choice would often seem to depend on whether the danger of bias in our method seems greater or less than the danger of large variances, and possibly bias

also resulting from the other one. Some clues may be available by inspecting $\hat{f}_x(\lambda_m)$. If the diagonal elements tend, say, to be very large for small λ_m but very small for large λ_m, the frequency-elimination method may be the more suitable. On the other hand, if $\hat{f}_x(\lambda_m)$ is more stable over $(-\pi,\pi]$, our approach might be preferred. The assumption that f_n be diagonal, as well as flat, may also be examined. If it is reasonable, the $\hat{f}_x(\lambda_m)$ will tend to be very well conditioned with the product of the i^{th} and j^{th} diagonal elements substantially greater than the squared modulus of the $(i,j)^{th}$ element, for all i, j. On the other hand, this phenomenon would occur also if the elements of x_t have low coherence. A more direct way of verifying the flatness and diagonality assumptions would be to use (3.17), investigating which matrices Θ approximately satisfy

$$\hat{f}_{yx}(\lambda_m) + \hat{\psi}^{(0)}(\lambda_m)\hat{f}_x(\lambda_m) = \hat{\psi}^{(0)}(\lambda_m)\Theta ,$$

for each λ_m.

2) It seems that a frequency-domain approach is particularly natural in the present case. With many time series models, an alternative efficient time domain approach is based on an autoregression specification for the residuals. A low order autoregressive specification may produce better results in moderate samples than a frequency domain one, which seems to require both M and T/M to be fairly large. Thus, Hannan and Terrell [9] consider both types of approach for estimating simultaneous equations with stationary errors. However, in our case, the effects of the measurement errors are such that in general no autoregressive transformation could possibly produce a model with white noise residuals incoherent of $\{x_t\}$.

3) Our procedure has been developed with computational considerations in mind, and in this respect it seems simpler than some other procedures

that might be used. However, the computation of the estimates still seems a formidable task, particularly if there is iteration. Nevertheless, time domain procedures, as well as seemingly rather inappropriate and inflexible (see comment 2) seem unlikely to be much easier computationally. It is true that it is a tedious task to express the components of the $\hat{b}^{(j)}$, $\hat{a}^{(j)}$ in terms of the real and imaginary parts of the summands, particularly as inverses of complex matrices are involved. However, complex arithmetic can often be carried out directly on the computer.

References

[1] Espasa, A. and J.D. Sargan, "The Spectral Estimation of Simultaneous Equation Systems with Lagged Endogenous Variables." Paper presented at the Third World Congress of the Econometric Society, Toronto, Canada, August 1975.

[2] Fisher, F.M., The Identification Problem in Econometrics. New York: McGraw-Hill, 1966.

[3] Geraci, J., "Simultaneous Equation Models with Measurement Error." Ph.D. Thesis, University of Wisconsin, 1974.

[4] Goldberger, A.S., "Structural Equation Methods in the Social Sciences." Econometrica 40 (1972), 979-1001.

[5] ______________, "Unobservable Variables in Econometrics." In Frontiers of Econometrics, P. Zarembka (ed.), New York: Academic Press, 1973.

[6] Hannan, E.J., Multiple Time Series. New York: Wiley, 1970.

[7] ___________, "The Identification Problem for Multiple Equation Systems with Moving Average Errors." Econometrica 39 (1971), 751-765.

[8] ___________ and D.F. Nicholls, "The Estimation of Mixed Regression, Autoregression, Moving Average and Distributed Lag Models." Econometrica 40 (1972), 529-547.

[9] ___________ and R.D. Terrell, "Multiple Equation Systems with Stationary Errors." Econometrica 41 (1973), 299-320.

[10] Hsiao, C., "Identification and Estimation of Simultaneous Equation Models with Measurement Error." International Economic Review 17 (1976), 319-339.

[11] ________, "Measurement Error in a Dynamic Simultaneous Equations Model with Stationary Disturbances." Paper presented at the Econometric Society annual meeting, Atlantic City, N.J., September 1976.

[12] Robinson, P.M., "Identification, Estimation and Large-Sample Theory for Regressions Containing Unobservable Variables." International Economic Review 15 (1974), 680-692.

[13] _____________, "Instrumental Variables Estimation of Differential Equations." Econometrica 44 (1976), 765-776.

[14] _____________, "Fourier Estimation of Continuous Time Models." In Statistical Inference in Continuous Time Economic Models, A.R. Bergstrom (ed.), Amsterdam: North-Holland, 1976.

[15] Sargan, J.D., "The Estimation of Economic Relationships Using Instrumental Variables." Econometrica 26 (1958), 393-415.

[16] Zygmund, A., *Trigonometric Series*, *Vol. I*. Cambridge: Cambridge University Press, 1959.