

On the cost of approximate specification in simultaneous equations estimation (1961)

1 T. C. Liu's objections to simultaneous equations methods

In recent years, there has been growing discussion of the techniques and problems of simultaneous equation estimation.¹ Most of this discussion has been concerned primarily with the techniques of estimation in the overidentified case. An important exception to this, however, is the work of T. C. Liu.

In a 1955 article² Liu advanced a disturbing argument which he has vigorously maintained ever since.³ This argument is disturbing because its premises apparently cannot be doubted and because its conclusions, if accepted, imply that the hope of structural estimation by any techniques whatsoever is forlorn indeed.

It is perhaps best to emphasize immediately that this chapter *only* takes issue with the very general and most disturbing form of Liu's position. The fact that I argue that such a *general* position is largely untenable should not be taken to mean that I think the problems to which Liu calls attention are not important in many particular instances. Some of these Liu has pointed out.

Liu's argument is as follows. Econometric models are only approximations to reality. Of necessity, they abstract from the real world by limiting the number of variables included in each equation and in the system as a whole. In fact, however, there are always more variables "really" in each

I am indebted to Paul A. Samuelson for suggesting this study. This chapter was written while I was a member of the University of Chicago Department of Economics and was supported in part by the Ford Foundation Econometrics Workshop of that department.

equation – either variables included in other equations of the model or variables excluded altogether therefrom – than are assumed in the practical approximations. Furthermore, the equations stated in the model are not a complete set of equations. In fact, other outside equations always exist, either relating variables in the model alone or relating them to unincluded variables.

Now, Liu goes on to argue, these points have serious consequences which are of three types. First, the usual form of *a priori* restriction used for identification – the restriction that some parameters in a structural equation are zero, that is, that some variables included elsewhere in the system do not enter into that equation – is likely to be incorrect. Such restrictions, at best, will only hold approximately, and approximately is not good enough. If the variables in question really belong in the equation being studied, then that equation is underidentified and its parameters cannot be estimated by any reasonable technique.

Second, the existence of other variables in the equation that are not included elsewhere in the system, together with the existence of unincluded equations relating those variables to the included ones or relating the included variables to each other, means that the number of equations in the system has been understated so that the necessary order condition for identifiability – which runs in terms of the number of *a priori* restrictions and the number of equations⁴ – is not satisfied in fact. Finally, few variables are truly exogenous to the complete system; to count only the explicitly stated equations is to treat endogenous variables as exogenous. Once again it follows that identification and, *a fortiori*, overidentification is the unusual case and underidentification the usual one.

Therefore, says Liu, the current emphasis on techniques of estimation in overidentified systems is entirely misplaced. Structural estimation is generally not possible in simultaneous systems, and only reduced forms can be obtained. Furthermore, unrestricted least squares estimation of the reduced form is the appropriate method, for restricting the reduced form by the *a priori* restrictions not only adds no more information, it is positively harmful, as it adds *misinformation*. Forecasting should therefore always be done with the unrestricted reduced form which, as is well known, has smallest error variance for the sample observations, as there is no reason to expect mistakenly oversimplified and restricted models to do as well or better.

Now, this is a powerful and disturbing argument. There can be no doubt that econometric models do in fact abstract from reality in the way described by Liu and thus that *a priori* restrictions which exclude variables from equations are frequently misspecified. Further, there can be no doubt that Liu is largely correct in stating that the treatment of certain variables as exogenous is also likely to be misspecification. The

economic system even considered as a whole is embedded in a far larger socio-physical framework. If Liu is right in his conclusions, then there must be serious doubt as to the possibility of structural estimation.

Nor do two counter-observations which might be advanced here seem very helpful in a direct way. The first of these is that it has now been shown that *a priori* restrictions of a far more general kind than those considered by Liu may be used for identification.⁵ The usual case is certainly the one which Liu considers, and even the use of inequalities (which are common restrictions) will not directly help very much.

The second observation is that variables need not be exogenous to the system as a whole to be exogenous with respect to some subset of equations. It is therefore not sufficient to argue that few variables are exogenous to the complete system; one must also hold that there are excluded equations specifically providing some "feedback" effect (however weak) from endogenous to assumed exogenous variables or connecting the latter with variables influencing the error terms. Otherwise, the system may be what we shall define below as "block recursive." However, since the existence of such excluded equations cannot be denied in general – especially equations connecting the variables in the error terms of the original system and the exogenous variables – this remark is also not directly helpful save as a reservation to Liu's argument in special cases.

While these counter-arguments are not of much help directly, however, they do provide aid in an indirect way, for they indicate that the problem may not be quite so dichotomous as Liu makes it. Thus, in the next section, we use the above-mentioned results on general restrictions to show that the proper question is not whether certain parameters assumed to be zero are in fact so; the issue is rather whether they are in some sense sufficiently small – whether the restriction that the corresponding variables do not appear is or is not a good approximation. We show that the problem is not the discontinuous one of just identification or overidentification if the restrictions hold exactly and underidentification if they do not, but rather one of diminishing estimation inconsistency as the restrictions are better and better approximations.

Similarly, in later sections, we go on to consider the problems of omitted variables and of exogenous variables and show that the issue is not whether a variable assumed exogenous is "really" so, or whether omitted variables "really" have zero coefficients, but rather whether these things are sufficiently so in an approximate sense. We thus justify the practice of breaking down a complete system into parts by assuming certain variables exogenous that are only approximately so, and by assuming certain variables absent that truly appear with very small coefficients.

These results have several consequences. First, of course, structural

estimation is seen to be entirely possible in general, so that discussion and criticism must be directed toward the goodness or badness of the approximate assumptions in a particular case and not toward the truth or falsity thereof.

Second (and this is a somewhat different way of looking at the matter), we may say that whereas, hitherto, estimation techniques required the knowledge or assumption that certain things (parameters, covariances, and so forth) were zero, the results here presented allow estimation with negligible inconsistency provided only that such things are known (or assumed) to be small. Since the latter kind of knowledge is easier to come by than the former and the assumption involved is likely to be held with greater confidence, this is a reassuring and perhaps not insignificant result.

Third, even if it is considered obvious that small errors have small consequences,⁶ it is always of interest to know exactly where errors must be small and thus to clarify assumptions.

Finally, an objection to simultaneous equation estimation somewhat better known than that of Liu is the position of H. Wold⁷ which states that the real world is not truly simultaneous at all, but that causation is unilateral and that true systems are always recursive. Part of the Liu objection, on the other hand, can be crudely put as the argument that the real world is always more simultaneous than we think – that no economic model ever fully states the true simultaneity – and therefore that estimation is impossible since the world is truly underidentified. It would seem at first glance as though no middle ground between these two positions is possible, that once the possibility of simultaneous causation is admitted, one cannot stop short of an explicit statement of the total set of equations explaining the socio-physical universe. We shall see, however, that this is not the case; a position intermediate between Wold and Liu is indeed possible, for systems – and highly plausible systems at that – do exist in which simultaneity is present but not completely overriding. Estimation of the usual simultaneous type is thus entirely possible.⁸

2 Misspecification in the *a priori* restrictions

We first consider in isolation the problem of the misspecification of the *a priori* restrictions. Throughout this section it will be assumed that the system is correctly specified save for the *a priori* restrictions. We shall remove this assumption below.

Consider the system of equations

$$u(t) = Ax(t) \tag{1.1}$$

where $u(t)$ is an $(m + 1)$ -dimensional column vector of disturbances, A is

an $(m+1) \times (n+1)$ matrix of coefficients with first row A_1 , and $x(t)$ is an $(n+1)$ -dimensional column vector of variables. Without loss of generality we may renumber the variables so that

$$x(t) = \begin{bmatrix} y(t) \\ z(t) \end{bmatrix} \quad (1.2)$$

where $y(t)$ is an $(m+1)$ -dimensional column vector of endogenous variables and $z(t)$ is an $(n-m)$ -dimensional column vector of exogenous variables. Then A may be partitioned accordingly into

$$A = [B \mid G] \quad (1.3)$$

where B is a square nonsingular matrix of rank $m+1$. We further assume that there are no linear identities connecting the exogenous variables.

Now, let ϕ' be a matrix with K columns and $n+1$ rows, where $n \geq K \geq m$. Each column of ϕ' is to contain precisely one unit element and have all other elements zero.⁹ Impose (incorrectly) the *a priori* restrictions on A_1

$$A_1 \phi' = 0, \quad (1.4)$$

and assume that the rank of the matrix $A\phi'$ is m , so that the necessary and sufficient conditions for unique identifiability of A_1 are satisfied if (1.4) holds. Select some endogenous variable not excluded from the first equation of (1.1) by (1.4) – of course, such must exist – and normalize by setting the corresponding element of A_1 equal to unity.¹⁰ Without loss of generality, we may assume that it is the first variable which is thus treated. Extend the definition of ϕ' by adding to it a new first column which has a one at the top and zeros elsewhere. ϕ' now has $K+1$ columns, and (1.4) is replaced by

$$A_1 \phi' = (1, 0, \dots, 0). \quad (1.4')$$

It is easy to see that the new $A\phi'$ must have rank $m+1$.

We now assume that (1.4) and hence (1.4') are misspecified so that the variables excluded from the first equation of (1.1) are really in that equation, so that (1.4) is only approximately true at best. Let the *true a priori* restrictions corresponding to (1.4') be

$$A_1 \phi' = (1 \quad \bar{\eta}) \quad (1.5)$$

where $\bar{\eta}$ is a K -dimensional row vector.

It will be convenient at this stage to rewrite the first equation of (1.1). Let $y_1(t)$ be the first variable. Let $y^1(t)$ be the vector of endogenous variables whose coefficients in the first equation are not specified by (1.5); similarly, let $z^1(t)$ be the vector of exogenous variables whose coefficients in the first equation are not specified by (1.5). Let B_1^1 and G_1^1 be the

8 Block-recursive systems

corresponding vectors of true coefficients. Finally, let $w(t)$ be the vector of variables (other than $y_1(t)$) whose coefficients are specified by (1.5). Then the first equation of (1.1) can be written

$$y_1(t) + B_1^1 y^1(t) + G_1^1 z^1(t) + \bar{\eta} w(t) = u_1(t). \quad (1.6)$$

Now, let $a_{1j}(\eta, \bar{\eta})$ be the probability limit – if it exists – of the estimate of A_{1j} obtained by imposing

$$A_1 \phi' = (1 \quad \eta) \quad (1.7)$$

when (1.5) is true and using one of the standard estimation techniques.¹¹ Thus, $a_{1j}(0, \bar{\eta})$ is the probability limit – if it exists – of the estimate of A_{1j} obtained by imposing (1.4') and using one of the standard techniques. Similarly, $a_{1j}(\bar{\eta}, \bar{\eta})$ is the probability limit – if it exists – of the estimate of A_{1j} obtained by imposing the true restrictions.

In what follows, we restrict our attention to estimators of Theil's k -class.¹² This class includes two-stage least squares and limited-information maximum likelihood as well as other estimators. Similar theorems could obviously be proved for estimators not in the k -class, save that for full-information maximum likelihood misspecifications elsewhere in the system must be assumed to go to zero.

Now, choose a set of T values for $u(t)$ and for those predetermined variables which are not lagged endogenous variables; further, choose a set of initial values for those predetermined variables which are lagged endogenous variables. The values of the endogenous variables are then completely determined by the matrix A . We shall assume that all asymptotic variances and covariances (between variables or between variables and residuals) are finite.

Next, fix all rows of A other than the first and all elements of B_1^1 and G_1^1 . Consider an infinite sequence of equation systems differing only in $\bar{\eta}$, such that $\bar{\eta}$ converges to the zero vector as we move along the sequence. We shall assume that there exists a neighborhood of the zero vector such that all systems in the sequence for which $\bar{\eta}$ lies within that neighborhood are stable.

This assumption requires some discussion. We shall show that it is a sufficient condition for $a_{1j}(0, \bar{\eta})$ to approach A_{1j} as a limit as $\bar{\eta}$ goes to zero, and hence that it implies that simultaneous equation estimators are only negligibly inconsistent for all elements of $\bar{\eta}$ sufficiently small (that is, for (1.4') a good enough approximation to (1.5)). The assumption may not be necessary, however; indeed, it seems quite clear that the theorem in question holds provided that there exists a neighborhood of zero for which consistent estimation is possible under correct specification (provided that identifiability is present). Whether there exist unstable

systems for which the usual estimators are not consistent under correct specification is as yet an open question save in special cases. Our assumption should be regarded as a declaration that we are not here concerned with answering this question, but that we are dealing with only one problem at a time.¹³

Theorem 1.1:

$$\lim_{\bar{\eta} \rightarrow 0} |a_{1j}(0, \bar{\eta}) - A_{1j}| = 0 \text{ for all } j = 1, \dots, n+1.$$

Proof: It is shown in Fisher [3] (p. 444, Theorem 13) that a necessary and sufficient condition for the identifiability of A_1 under (1.5) is that the rank of $A\phi'$ be $m+1$. We have already observed that this is the case for $\bar{\eta} = 0$; it must therefore also be the case for $\bar{\eta}$ sufficiently close to zero, for the elements of $A\phi'$ are either constant or continuous functions of $\bar{\eta}$ and the value of a determinant is a continuous function of its elements. Hence, for $\bar{\eta}$ sufficiently close to zero, A_1 is identifiable under (1.5).

Now, the theorem is trivially true for those elements of A_1 which are also elements of $\bar{\eta}$. We may therefore restrict our attention to the elements of B_1^1 and G_1^1 .

Following Theil,¹⁴ define a new endogenous variable $q(t)$ as

$$q(t) \equiv y_1(t) + \bar{\eta}w(t) \quad (1.8)$$

and rewrite (1.6) as

$$q(t) + B_1^1 y^1(t) + G_1^1 z^1(t) = u_1(t). \quad (1.9)$$

We call the system of equations formed by (1.8) and (1.1) with the first equation of the latter set replaced by (1.9) the *auxiliary* system to (1.1). It is evident that (1.1) and its auxiliary system are equivalent and that the inhomogeneous restrictions (1.5) corresponding to (1.1) have been replaced with equivalent restrictions of the usual zero-coefficient type in the auxiliary system. Further, observe that the variables that are included in $y^1(t)$ and $z^1(t)$ do not depend on the value of $\bar{\eta}$ or on statements about that value. That is, the auxiliary equation (1.9) differs from the original equation (1.6) with the misspecified restriction $\bar{\eta} = 0$ imposed only in that $q(t)$ and not $y_1(t)$ appears.

Now consider an estimator of the k -class. Let $a_1^1(\eta, \bar{\eta})$ be the column vector of those $a_{1j}(\eta, \bar{\eta})$ corresponding to the elements of B_1^1 and G_1^1 . Let V be a matrix of observations on the residuals from the least squares estimates of the reduced-form equations explaining the elements of $y^1(t)$. Let all other capital letters denote the observation matrices of the

corresponding lower-case variables. Then the typical k -class estimator of (1.9) is given by

$$\begin{bmatrix} (Y^1)' Y^1 - kV' V & (Y^1)' Z^1 \\ (Z^1)' Y^1 & (Z^1)' Z^1 \end{bmatrix}^{-1} \begin{bmatrix} (Y^1 - kV)' Q \\ (Z^1)' Q \end{bmatrix}. \quad (1.10)$$

Letting H be the matrix whose inverse is taken, this may be rewritten

$$(TH^{-1}) \left(1/T \begin{bmatrix} (Y^1 - kV)' Q \\ (Z^1)' Q \end{bmatrix} \right) \quad (1.11)$$

and, under the stability and identifiability assumptions made, the probability limit of (TH^{-1}) exists for $\bar{\eta}$ sufficiently close to zero, as does the probability limit of the second factor. Further, the probability limit of the product is $a_1^1(\bar{\eta}, \bar{\eta})$, and

$$a_1^1(\bar{\eta}, \bar{\eta}) = \begin{bmatrix} (B_1^1)' \\ (G_1^1)' \end{bmatrix} \quad (1.12)$$

provided that the probability limit of k is one.

There are now two cases to consider. The first case is that in which the probability limit is independent of η – that is, independent of our misspecification. This is the case of two-stage least squares (k identically equal to one) and of certain other members of the k -class.¹⁵ The case in which the probability limit of k does depend on $\bar{\eta}$ is exemplified by limited-information maximum likelihood, and we shall restrict our attention in the second case to this example.

Suppose now that the probability limit of k is independent of η . Form the k -class estimator of (1.6) on the assumption that $\bar{\eta} = 0$, that is, under misspecification. This estimator can be written as

$$(TH^{-1}) \left(1/T \begin{bmatrix} (Y^1 - kV)' Y_1 \\ (Z^1)' Y_1 \end{bmatrix} \right). \quad (1.13)$$

The crucial point here is that H is unaltered by the misspecification; hence the probability limit of (TH^{-1}) exists as does the probability limit of the second factor which involves only asymptotic moment matrices. The probability limit of the product is $a_1^1(0, \bar{\eta})$, and we have

$$a_1^1(\bar{\eta}, \bar{\eta}) - a_1^1(0, \bar{\eta}) = \text{plim} \left\{ (TH^{-1}) \left(1/T \begin{bmatrix} (Y^1 - kV)' W \\ (Z^1)' W \end{bmatrix} \right) (\bar{\eta}') \right\}, \quad (1.14)$$

and this clearly approaches zero as a limit as $\bar{\eta}$ goes to zero, thus proving the theorem in this case in view of (1.12).

The case of limited-information maximum likelihood is similar, save that here k is not independent of η , since it is equal to $1 + \rho$ where ρ is the smallest root of the determinantal equation

$$|M_1 - (1 + \rho)M| = 0 \quad (1.15)$$

where M_1 and M are different for the auxiliary and the misspecified systems. For the auxiliary system, they are the moment matrices of the estimated residuals in the least squares regressions of $q(t)$ and the elements of $y^1(t)$ on the elements of $z^1(t)$ and on all the exogenous variables respectively. For the misspecified system, on the other hand, they are the same with $y_1(t)$ in place of $q(t)$. Of course, the two coincide for $\bar{\eta} = 0$.

The probability limits of M_1 and M and hence of ρ and k exist in either case. It is clear, moreover, that the two cases differ only in the first rows and columns of M_1 and M . As $\bar{\eta}$ approaches zero, these first rows and columns approach each other, in view of (1.8), as do the probability limits of the corresponding k s. For $\bar{\eta}$ sufficiently close to zero, therefore, the probability limit of k in the misspecified system will be sufficiently close to that in the auxiliary system (unity) to insure the existence of the probability limit of (TH^{-1}) when H is computed using the k corresponding to the misspecified system. The proof now proceeds as before.

Some remarks are now in order. First, it is clear that the theorem could easily be generalized to the case where $\bar{\eta}$ approaches η for given η . We should then merely have $(\bar{\eta} - \eta)'$ in place of $\bar{\eta}'$ on the extreme right of (1.14). Moreover, the same generalization shows that the theorem remains true if we allow η to approach $\bar{\eta}$ for fixed $\bar{\eta}$, that is, allow our statement to approach the truth rather than (as above) allowing the truth to approach our statement. (We have adopted the latter course because of the context of the problem.)

Second, we have now shown that it is not the case that the use of *a priori* restrictions which only hold approximately necessarily leads to the abandonment of simultaneous equation methods of estimation. It is true that such use leads to inconsistency in the estimates, but, provided the approximations involved are good enough, such inconsistencies will be negligible.

What is involved in deciding whether particular approximations are "good enough" is an interesting but complex and difficult question. It has been answered by Theil¹⁶ for the case of a single equation with misspecified *a priori* restrictions. The answer for simultaneous equations is far more difficult to obtain and is a fit subject for further work, especially in view of the fact (brought out in the above proof) that

12 *Block-recursive systems*

different estimators may have different sensitivities to this kind of misspecification. We shall return to this below. Here we are concerned only to show that some “good enough” approximation does exist, so that simultaneous equation methods need not be discarded simply because restrictions are only approximate. It may indeed be true that many or all *a priori* restrictions actually used are not “good enough” approximations (in the sense that they lead to inconsistencies too large to be tolerated); however, this must be decided on a case-by-case basis, and no general *a priori* argument can be made to this effect.

Third, it follows from our theorem and discussion that Liu is wrong in claiming that the unrestricted least squares estimates of the reduced form should be used for prediction because the use of *a priori* restrictions adds only misinformation. The reduced-form matrix is a continuous transformation of the coefficient matrix of the system; it follows that as inconsistencies in the estimates of the elements of the latter go to zero, so do inconsistencies in the estimates of the reduced-form coefficients. The use of *a priori* restrictions which are approximate thus leads to negligible inconsistencies in such estimates also – for good enough approximations. It follows that the restricted estimates of the reduced form obtained from structural equation estimates converge more rapidly to probability limits that differ slightly or negligibly from the true reduced-form coefficients than the unrestricted least squares reduced-form estimates converge to the true reduced-form parameters. Here is a case in which a slightly or negligibly inconsistent estimator is more efficient than a consistent estimator. In applying the restrictions in the estimation of the reduced form, one trades consistency for efficiency. Such a trade of precise accuracy for convenient closeness is always the price of approximate assumptions. Provided the approximations are close enough, the efficiency properties of simultaneous equation estimators will more than compensate for their inconsistency. Here again, there is no general *a priori* argument that approximations will not be “close enough;” this can only be decided in particular cases.

3 **Block-recursive systems: omitted variables**

In the last section the effect of misspecification in the *a priori* restrictions alone was considered. Throughout this section, unless otherwise stated, we assume that the *a priori* restrictions are correctly specified and turn to the next type of approximate misspecification considered by Liu.

It will be convenient to alter our notation somewhat and to define some new concepts. A triangular matrix is a square matrix with zeros everywhere below the main diagonal. As is well known, an equation system whose matrix is triangular is a recursive system; its equations may

be treated singly or sequentially rather than simultaneously and may be estimated by least squares provided that the disturbances in the equations are all mutually uncorrelated. The equations of the system form a unilateral causal chain.¹⁷

We now generalize these concepts in the following way. For any matrix M , let $r(M)$ be the number of rows of M and $c(M)$ be the number of columns. Consider a square matrix

$$W = \begin{bmatrix} R^1 & S^1 & & & \\ 0^{21} & R^2 & S^2 & & \\ 0^{31} & 0^{32} & R^3 & S^3 & \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0^{N1} & 0^{N2} & 0^{N3} & \dots & R^N \end{bmatrix} \quad (1.16)$$

where the R^i are nonsingular (and hence square) matrices and the S^i are matrices which may or may not be zero. The 0^{ij} are zero matrices; clearly,

$$\begin{aligned} r(0^{ij}) &= r(R^i), \\ c(0^{ij}) &= c(R^j) = r(R^j) \end{aligned} \quad (i, j = 1, \dots, N). \quad (1.17)$$

Such a matrix is said to be *block triangular*; it is triangular in blocks rather than in single elements.¹⁸ The reason for imposing nonsingularity on the R^i is one of convenience and will be apparent shortly.

Observe that each of the submatrices

$$R^N, \begin{bmatrix} R^{N-1} & S^{N-1} \\ 0^{NN-1} & R^N \end{bmatrix}, \begin{bmatrix} R^{N-2} & S^{N-2} & \\ 0^{N-1N-2} & R^{N-1} & S^{N-1} \\ 0^{NN-2} & 0^{NN-1} & R^N \end{bmatrix} \dots, W$$

is square and nonsingular and is block triangular (the last property holding in an empty sense for R^N which has only one block).

Now consider the equation system

$$u = Ax. \quad (1.18)$$

(We have dropped the time argument for convenience; it is to be understood.) As in the last section, partition A and x to correspond to endogenous and exogenous variables, the endogenous variables coming first. Thus rewrite (1.18) as

$$u = [B \ G] \begin{bmatrix} y \\ z \end{bmatrix} \quad (1.19)$$

14 Block-recursive systems

Now assume that B is block triangular with N blocks and partition u and y into N corresponding blocks, thus:

$$u = \begin{bmatrix} u^1 \\ u^2 \\ u^3 \\ \vdots \\ u^N \end{bmatrix} \quad y = \begin{bmatrix} y^1 \\ y^2 \\ y^3 \\ \vdots \\ y^N \end{bmatrix} \quad (1.20)$$

We call an equation such as (1.18), in which the part of the coefficient matrix corresponding to the endogenous variables is block triangular, *block recursive*. Any equation system (block recursive or not) which has a square coefficient matrix (in our notation, where G is null) will be called *self-contained*.

Block-recursive systems have the property that – given the exogenous variables, if any – the variables in y^N are (stochastically) determined solely by the equations corresponding to B^N ; the variables in y^{N-1} are then determined by the exogenous variables, the variables in y^N , and the equations corresponding to B^{N-1} ; and so forth. Accordingly it is clear that if all the equations that involve the variables have been included in the system and the system itself has been correctly specified, the parameters of the subset of equations with u^j as left-hand member (say the j th subset) may be estimated with regard only for the equations in that subset and without regard for the existence of the remaining equations.¹⁹ The variables in any y^i may clearly be considered exogenous to the j th subset of equations provided that $i > j$, or to any union of such subsets as may be seen by solving (1.19) for y in terms of z and u . It is clear that ordinary recursive systems are special cases of block-recursive systems with the B^i single rows and the 0^{ij} single elements.

The analysis of causation for self-contained block-recursive systems has been given by H. Simon.²⁰ He argues (considering more general systems as well) that the variables in y^N can plausibly be regarded as “causing” the variables in y^{N-1} in such a case; the variables in y^N and y^{N-1} can be regarded as “causing” the variables in y^{N-2} ; and so forth, a concept of causality which Strotz and Wold have aptly termed “vector causality.”²¹ The dynamic properties of block-recursive systems are obviously closely related to those of unilaterally coupled systems considered by R. Goodwin;²² indeed, block-recursive systems are the generalization of unilaterally coupled systems to more than two markets.

We shall have something to say about such dynamic properties in a later section.

We turn now to the second issue of specification error raised by Liu, that of the incorrect or only approximately correct omission of relevant variables from the whole system of equations being studied. As observed above, Liu argues that such omitted variables are likely to be connected with the included variables by equations other than those explicitly included. It will be convenient both now and later to frame our discussion (which is quite general) in terms of block-recursive systems.

Consider the system of equations explicitly under investigation:

$$u^2 = [D^{21} B^2 G^2] \begin{bmatrix} x^1 \\ y^2 \\ z^2 \end{bmatrix} \quad (1.21)$$

where B^2 is square and nonsingular; G^2 and D^{21} are rectangular; and the reason for the superscripts will appear shortly. In (1.21), y^2 is a vector of explicitly endogenous variables, and z^2 is a vector of explicitly assumed exogenous variables. x^1 , on the other hand, is a vector of variables assumed absent from the system. In other words, D^{21} is incorrectly assumed to be a zero matrix.

In what follows we shall assume that the specification of z^2 as a vector of exogenous variables is correct, leaving the alternative case to a later section. We further suppose that sufficient *a priori* restrictions of the usual type have been imposed on the first equation of the system to identify it when D^{21} is in fact zero, that is (where ϕ' is, as before, the transposed matrix of the coefficients of the restrictions), we suppose that the rank of the matrix $[B^2 G^2] \phi'$ is $r(B^2) - 1$. For the present we continue to assume that all such *a priori* restrictions are true ones so that the problems studied in the last section do not arise.

We now consider all equations relating the variables of x^1 to the included variables of y^2 and z^2 . We do this in the following way: consider equations explaining each of the elements of x^1 (y^2 is explained by (1.21) and z^2 is exogenous). If such equations involve variables not in (1.21) expand the definition of x^1 and hence of D^{21} to include all such variables. Continue this process either until there are as many new independent equations as there are variables in the expanded x^1 , or until all variables left unexplained in the expanded x^1 are exogenous to the entire system. (There may be very many equations in the result.) Now partition x^1 into a vector of variables endogenous to the complete system, y^1 , and a vector (possibly null) of variables exogenous to the complete system, z^1 .

Correspondingly, partition D^{21} into E^{21} and F^{21} . The complete system of which (1.21) is a part is thus

$$\begin{bmatrix} u^1 \\ u^2 \end{bmatrix} = \left[\begin{array}{c|c|c} B^1 & H^1 & G^1 \\ \hline E^{21} & F^{21} & B^2 \mid G^2 \end{array} \right] \begin{bmatrix} y^1 \\ z^1 \\ y^2 \\ z^2 \end{bmatrix} \quad (1.22)$$

where B^1 is square and corresponds to y^1 . We shall assume B^1 to be nonsingular for the present. It is easily seen that $r(E^{21}) = r(B^2)$ and $c(E^{21}) = r(B^1)$. It is thus clear (if you like, reshuffle columns to put all exogenous variables together) that the assumption that $E^{21} = 0$ is the assumption that the system (1.22) is block recursive so that the subsystem (1.21) can be estimated in isolation. Of course, the assumption that $F^{21} = 0$ is the additional assumption that all exogenous variables that appear in (1.22) have been explicitly included in (1.21).

Let n be the total number of variables and m be the total number of equations in (1.22). Let

$$A = [D^{21} B^2 G^2] = [E^{21} F^{21} B^2 G^2] \quad (1.23)$$

and consider the problem of estimating the first row of A , as before, A_1 . Let the probability limit of the estimate of A_{1j} obtained by assuming that $D^{21} = 0$ and applying a standard technique²³ be $\alpha_{1j}(0, D^{21})$. As in the previous section, choose an indefinitely large set of values for u^1 and u^2 and for those elements of z^1 and z^2 that are not lagged endogenous variables; choose a set of initial values for all lagged endogenous variables in z^1 and z^2 . Now fix B^1, H^1, G^1, B^2 and G^2 , and consider a collection of systems differing only in D^{21} . Consider an infinite sequence of such systems such that D^{21} approaches the zero matrix. We prove that the inconsistency involved in assuming D^{21} zero – that is, in omitting x^1 from the equations – goes to zero as D^{21} goes to zero, as the approximation involved gets better and better.

Theorem 1.2:

$$\lim_{D^{21} \rightarrow 0} |\alpha_{1j}(0, D^{21}) - A_{1j}| = 0 \text{ for all } j = 1, \dots, n.$$

Proof: We prove the theorem by showing it to be a special case of Theorem 1.1. Estimation of A_1 on the assumption that $D_1^{21} = 0$ is estimation applying as many new (though incorrect) *a priori* restrictions as there are elements in D_1^{21} , that is, as many restrictions as there are variables in x^1 . By assumption, however, there are already available at least $r(B^2) - 1$ restrictions on the other elements of A_1 ; hence the

necessary order condition for the identifiability of A_1 considered as a part of (1.22) rather than just of (1.21) is certainly satisfied when $D_1^{21} = 0$ since

$$\begin{aligned} r(B^2) - 1 + c(D^{21}) &\geq r(B^2) - 1 + c(E^{21}) \\ &= r(B^2) + r(B^1) - 1 = m - 1. \end{aligned} \quad (1.24)$$

If F^{21} is not null, there are more restrictions than necessary (as there may also be if there were more than $r(B^2) - 1$ original restrictions).

We must now show that the necessary and sufficient rank condition is also satisfied when $D_1^{21} = 0$. Arrange the restrictions $D_1^{21} = 0$ so that the coefficients thereof form a unit matrix with $c(D^{21})$ rows and columns. The rank condition will be satisfied if the rank of the matrix

$$\begin{aligned} &\left[\begin{array}{c|c|c} B^1 & H^1 & G^1 \\ \hline E^{21} & F^{21} & B^2 & G^2 \end{array} \right] \left[\begin{array}{c|c} I & 0 \\ \hline 0 & \phi' \end{array} \right] \\ &= \left[\begin{array}{c|c|c} B^1 & H^1 & G^1 \phi' \\ \hline E^{21} & F^{21} & (B^2 G^2) \phi' \end{array} \right] \end{aligned} \quad (1.25)$$

is $m - 1$.

To see that this is the case for D^{21} sufficiently close to zero, strike out the row of the above product matrix that corresponds to A_1 , denoting the result by the use of lower-case instead of upper-case letters. Further, select $r(B^2) - 1$ independent columns from ϕ' and assume (for convenience only) that these are the only columns in ϕ' . Consider the following submatrix of the above matrix:

$$\left[\begin{array}{c|c} B^1 & G^1 \phi' \\ \hline e^{21} & (b^2 g^2) \phi' \end{array} \right] = W(e^{21}), \text{ say.} \quad (1.26)$$

This is $(m - 1) \times (m - 1)$. Now consider $W(0)$, and suppose that it were singular. Then a nonzero row vector, say λ , would exist such that $\lambda W(0) = 0$. Partition λ into $(\lambda^1 \lambda^2)$ corresponding to the two rows of blocks in $W(0)$. Singularity of $W(0)$ then implies that

$$(\lambda^1 \lambda^2) \left[\begin{array}{c|c} B^1 & G^1 \phi' \\ \hline 0 & (b^2 g^2) \phi' \end{array} \right] = 0 \quad (1.27)$$

and hence that

$$\lambda^1 B^1 + \lambda^2 0 = \lambda^1 B^1 = 0 \quad (1.28)$$

which is impossible unless $\lambda^1 = 0$, since B^1 was assumed nonsingular. It thus follows that

$$\lambda^1 G^1 \phi' + \lambda^2 (b^2 g^2) \phi' = \lambda^2 (b^2 g^2) \phi' = 0. \quad (1.29)$$

This is impossible, however, unless $\lambda^2 = 0$, for $(b^2 g^2) \phi'$ is also nonsingular by assumption. Hence $\lambda = 0$ and $W(0)$ must be nonsingular. It now follows immediately that $W(e^{21})$ must also be nonsingular for D^{21} (and hence e^{21}) sufficiently close to zero, for the determinant of a matrix is a continuous function of its elements. Therefore the product matrix in (1.25) has rank $m - 1$ for D^{21} sufficiently close to zero, and the rank condition is satisfied.

This being the case, estimation of A_1 on the assumption that $D_1^{21} = 0$ is estimation which involves incorrect *a priori* restrictions; however, for the remaining rows of D^{21} sufficiently close to zero, identification of A_1 is present when the restrictions are actually correct. *This is just the situation considered in Theorem 1.1, however;*²⁴ hence as *all* rows of D^{21} (including the first) approach zero, $\alpha_{1j}(0, D^{21})$ and A_{1j} approach each other for all j , by Theorem 1.1.

It is thus the case that, when the specification that the variables in x^1 do not appear in the system (1.21) is a "good enough" approximation, the inconsistency involved in applying it will be negligible in estimating any equation of (1.21) which is otherwise identified. (Once again, it is far more difficult precisely to define "good enough.") Moreover, a more important result than this is readily available.

The theorem just proved dealt with the estimation of one equation on the assumption that *all* specification errors approached zero – that not only were the omitted variables nearly absent from the equation in question, but they were nearly absent from the system (1.21) – that is, that *all* rows of D^{21} approached zero, not just the first. It is easy to show, however, that the negligibility of inconsistencies in the estimation of A_1 almost never depends on the closeness to zero of the last $r(B^2) - 1$ rows of D^{21} . To see this, observe that the only use made of such closeness in the proof of the theorem was to show that $W(e^{21})$ was nonsingular, and hence that the rank condition was satisfied for the last rows of D^{21} close enough to zero. *The necessary order condition is always satisfied regardless of the value of D^{21}* , in view of (1.24). Furthermore, the closeness to zero of the last rows of D^{21} is a sufficient, but by no means a necessary condition for the rank condition to hold; $W(e^{21})$ can be nonsingular for other configurations. Indeed, since a determinant is a linear function of any element and of the elements in any row, and since $W(0)$ is known to be nonsingular, $W(e^{21})$ must be nonsingular for all points in the space of the

elements of e^{21} , save for a set of measure zero. Hence the rank condition is satisfied almost everywhere in the space of the elements of the last $r(B^2) - 1$ rows of D^{21} , and we have the following theorem.

Theorem 1.3 (Limited Effects of Other Errors Theorem):

$$\lim_{D_1^{21} \rightarrow 0} |\alpha_{1j}(0, D^{21}) - A_{1j}| = 0 \text{ for all } j = 1, \dots, n$$

almost always, regardless of the last $r(B^2) - 1$ rows of D^{21} .²⁵

This is an important theorem, for it shows that the mistaken omission of nonnegligible variables from a given structural equation generally affects only the estimates of the parameters of that equation and not the estimates of parameters of other equations in the system. Specification errors of this type are thus of limited effect, and the estimates of the parameters of a given equation can be judged with regard only for the question of the goodness of approximation in the specification of that equation.

Some further remarks are now in order. If F^{21} and H^1 are not null, that is, if some of the omitted variables are exogenous to (1.22), then it can easily happen that even when e^{21} is such as to make $W(e^{21})$ singular, the rank condition holds, so that the preceding theorem is slightly stronger. Second, the assumption that B^1 is nonsingular seems an innocuous one. It can be discarded, however, and replaced with the statement that $W(0)$ (and hence $W(e^{21})$ for small e^{21}) will be nonsingular almost everywhere in the space of the elements of B^1 . As before, the rank condition can hold even with singularity of $W(e^{21})$ if there are omitted exogenous variables so that H^1 is not null.

Finally, we observe that the proofs of Theorems 1.2 and 1.3 do not depend on the truth or falsity of the original *a priori* restrictions. If those restrictions only hold approximately, it suffices to observe that the rank of $(b^2 g^2) \phi'$ will be $r(B^2) - 1$ almost everywhere in the space of the elements of $(b^2 g^2)$ and must be $r(B^2) - 1$ if the original *a priori* restrictions are close enough approximations (this follows as in the proof of Theorem 1.1). Hence we may treat all *a priori* restrictions – the original ones and the restrictions that $D_1^{21} = 0$ – together. Let $(B_1^2 G_1^2) \phi' = \bar{\eta}$ as in the preceding section. Denote by $\hat{a}_{1j}(0, \bar{\eta}; 0, D^{21})$ the probability limit of the estimate of A_{1j} obtained on the assumption that $D^{21} = 0$ and $(B_1^2 G_1^2) \phi' = 0$. Consider the same sequence of systems as described before Theorem 1.2, save that B_1^2 and G_1^2 are only fixed in those elements not determined by the *a priori* restrictions. We have the following.

Corollary to Theorems 1.1 and 1.2:

$$\lim_{\substack{\bar{\eta} \rightarrow 0 \\ D^{21} \rightarrow 0}} |\bar{a}_{1j}(0, \bar{\eta}; 0, D^{21}) - A_{1j}| = 0 \text{ for all } j = 1, \dots, n;$$

Corollary to Theorems 1.1 and 1.3:

$$\lim_{\substack{\bar{\eta} \rightarrow 0 \\ D^{21} \rightarrow 0}} |\bar{a}_{1j}(0, \bar{\eta}; 0, D^{21}) - A_{1j}| = 0$$

for all $j = 1, \dots, n$ almost always, regardless of the last $r(B^2) - 1$ rows of D^{21} .

Thus as approximations of both kinds so far considered get better and better, the inconsistency of estimates goes to zero. It follows that for good enough approximations – both in the *a priori* restrictions and in the omission of variables – such inconsistencies will be negligible. The effects of the assumption of block recursiveness and of approximate *a priori* restrictions are thus independent in this regard.

All this has been on the assumption that all variables assumed exogenous really are. We shall investigate the consequences of misspecification in this area after further discussion in the next section of the effects of omitted variables.

4 First generalization of Proximity Theorem

In the last section, we established conditions under which the inconsistency of the structural parameter estimates is negligible, even though variables are mistakenly or approximately excluded from the equation system studied. Those conditions were essentially that the coefficients of the excluded variables in the true system be very close to zero. In this section, we ask a somewhat different question. Suppose that the true coefficients of the omitted variables are not very close to zero. Under what conditions can the coefficients of the included variables be estimated with negligible inconsistency? In other words, let $J^2 = [B^2 G^2]$ and rewrite (1.21) as

$$u^2 = [D^{21} J^2] \begin{bmatrix} x^1 \\ x^2 \end{bmatrix}. \quad (1.30)$$

Continue to assume that the rank of $J^2 \phi'$ is $r(J^2) - 1$, where ϕ' is the coefficient matrix of true linear homogeneous *a priori* restrictions, $J_1^2 \phi' = 0$. We wish to ascertain the conditions under which J_1^2 , the first

row of J^2 , can be estimated with negligible inconsistency when D^{21} is assumed zero but is in fact *not* very close to zero.²⁶

Now, define a new residual vector v as

$$v \equiv u^2 - D^{21}x^1 = J^2x^2. \quad (1.31)$$

Estimation of J^2 on the assumption that $D^{21} = 0$ is thus seen to be estimation of (1.31) on the assumption that the elements of v are distributed independently of the exogenous variables z^2 . Leaving the interpretation of what follows in terms of a complete system such as (1.22) to later discussion, suppose that in fact v is related to x^2 by

$$v = Px^2 + u^2 + \mu = Ly^2 + Qz^2 + u^2 + \mu, \quad (1.32)$$

that is, the relations between the excluded and included variables are such that $D^{21}x^2 = -Px^2 + \mu$. We continue to assume that the true residuals—the elements of $(u^2 + \mu)$ —are in fact distributed independently of the elements of z^2 . It is easy to see, however, that this cannot generally be the case with the elements of v unless some rows of P are zero, since eliminating y^2 from (1.31) and (1.32) gives

$$v = [I - L(B^2)^{-1}]^{-1}[Q - L(B^2)^{-1}G^2]z^2 + [I - L(B^2)^{-1}]^{-1}(u^2 + \mu). \quad (1.33)$$

An element of v can thus only be independent of the elements of z^2 if some row of the matrix $[Q - L(B^2)^{-1}G^2]$ is zero. This will be the case if the corresponding row of $P = [L \ Q]$ is zero; otherwise, it can only happen (on a set of measure zero) if the effects of z^2 on v directly are exactly cancelled by the indirect effects through y^2 .²⁷

As before, choose an indefinitely large number of values for u^2 and μ and for such elements of z^2 as are not lagged endogenous variables, choosing initial conditions for the remaining elements of z^2 . Fix J^2 and consider systems differing only in P . Let $j_1(0, P)$ be the vector of the probability limits of the estimates of J_1^2 obtained for given P by wrongly assuming that $P = 0$ and applying one of the standard simultaneous equation techniques.²⁸ *Note that this may, but need not be, the mistaken assumption that no variables have been omitted* (the assumption of the last section). We shall discuss this below. We prove the following theorem.

Theorem 1.4 (First Generalized Proximity Theorem):

- (a) $\lim_{P \rightarrow 0} j_1(0, P) = \lim_{P \rightarrow 0} (J_1^2 - P_1) = J_1^2;$
- (b) $\lim_{P_1 \rightarrow 0} j_1(0, P) = \lim_{P_1 \rightarrow 0} (J_1^2 - P_1) = J_1^2$

almost always, regardless of the last $r(P) - 1$ rows of P .

Proof: Solve (1.31) and (1.32) for $u^2 + \mu$, obtaining

$$u^2 + \mu = (J^2 - P)x^2. \quad (1.34)$$

This system of equations is in the same form as (1.31), and $(u^2 + \mu)$ meets the assumptions made about (1.31) – its elements are distributed independently of z^2 . Hence some linear combination of the rows of (1.31) will be estimated. It remains to be shown that as P goes to zero, that estimate approaches $J_1^2 - P_1$. Consider the vector $(J_1^2 - P_1)\phi'$. As P approaches zero, this must approach $J_1^2\phi' = 0$ by assumption. Therefore, as P approaches zero, $J_1^2 - P_1$ comes closer and closer to satisfying the *a priori* restrictions. Furthermore, as P approaches zero, the matrix $(J^2 - P)\phi'$ approaches $J^2\phi'$. The determinant of a matrix, however, is a continuous function of its elements; hence, for P close enough to zero, the rank of $(J^2 - P)\phi'$ and of $J^2\phi'$ will be the same, namely $r(J^2) - 1$. It now follows that $j_1(0, P)$ is an estimate satisfying all the conditions of Theorem 1.1, for P sufficiently close to zero, and statement (a) of the theorem has been proved.

Statement (b) now follows as in the last section by observing that the rows of P other than the first only enter in the above proof in consideration of the rank of $(J^2 - P)\phi'$.

Before going on to discuss the reason for naming the theorem as we have, some remarks are in order regarding its consequences. Let us return to the notation of the last section (of (1.22) in particular). Here, however, we make one change. Previously we agreed to expand x^1 until all its elements were either explained or exogenous; now we must expand it until all its elements are either explained or distributed independently of z^2 . This is not the same thing, for functions of exogenous variables are themselves exogenous. We suppress all such independent variables (and also H^1 and F^{21}) by writing them as functions of x^1 with zero coefficients, and thus including them in y^1 . From (1.22)

$$y^1 = -(B^1)^{-1}G^1x^2 + (B^1)^{-1}u^1. \quad (1.35)$$

Hence

$$\begin{aligned} v &= u^2 - D^{21}x^1 = u^2 - E^{21}y^1 \\ &= E^{21}(B^1)^{-1}G^1x^2 + u^2 - E^{21}(B^1)^{-1}u^1 \end{aligned} \quad (1.36)$$

and this is precisely (1.32) with

$$\begin{aligned} P &= E^{21}(B^1)^{-1}G^1, \\ \mu &= -E^{21}(B^1)^{-1}u^1 \end{aligned} \quad (1.37)$$

It follows that the conditions of the second part of the theorem will be

satisfied if the first row of $E^{21}(B^1)^{-1}G^1$ goes to zero; hence inconsistencies will be negligible in the estimation of J_1^2 if all elements in that first row are sufficiently close to zero. Let us ask what this means.

In the first place, observe that the row in question will be near zero if the first row of E^{21} is near zero. This is scarcely surprising. If that is the case, then clearly inconsistencies in the estimation of E_1^{21} on the assumption that the latter is zero will be negligible. Combined with the results of Theorem 1.4, this gives us Theorems 1.2 and 1.3 as special cases of (a) and (b), respectively.²⁹

Secondly, the row in question will be near zero if G^1 and thus $(B^1)^{-1}G^1$ is near zero. It is the second matrix which is of most interest. This is a quasi-reduced-form matrix showing the full effects of the included variables on the omitted ones. Clearly, if all such effects are near zero, inconsistencies will be negligible.³⁰

Finally, we can make a general statement covering all cases. Inconsistencies will be negligible if to each nonnegligible element of E_1^{21} , say E_{1k}^{21} , there corresponds a negligible row of $(B^1)^{-1}G^1$, the k th row. In other words, inconsistencies will be negligible provided that the only omitted variables that enter into the first equation with nonnegligible coefficients are those variables on which the influence of the included variables is slight.

We have called Theorem 1.4 a generalization of the Proximity Theorem. The latter theorem is due to Wold³¹ and applies to a single equation. It states that the least squares regression coefficients of a single equation will be nearly unbiased either if the residuals from that equation are small or if they are nearly uncorrelated with the explanatory variables. The two effects reinforce each other. In the case of omitted variables, where only one equation is involved (that is, $r(J^2) = 1$), it is easy to see that this is essentially the same as our theorem with unbiasedness substituted for consistency.³² Clearly, in this case, the single row of P will be negligible if and only if the zero-order least squares regression coefficients of $D^{21}x^1$ on the included variables x^2 are all zero. These regression coefficients, however, are the products of correlation coefficients and the ratio of the variance of $D^{21}x^1$ to the variance of the variable in question. P will therefore be close to zero if either all the correlations are near zero or the variance of $D^{21}x^1$ is near zero; and the two effects reinforce each other (they multiply). Hence our theorem essentially reduces to the Proximity Theorem in the single equation case. This is not the only generalization of the Proximity Theorem to the simultaneous equations case, however. Another such will be given in the next section.³³

We may emphasize that the regression coefficients and not simply the correlations of the residual with the explanatory variables are important here.³⁴ Even if all such correlations were perfect, if the effects on the

residual of moving the explanatory variables were very small, relative to the direct effects of such movement on the dependent variable, the inconsistency involved would be very slight. A similar statement holds for the simultaneous equation case.³⁵

Finally, we observe that Theorem 1.4 and Theorem 1.1 can hold at the same time, that is, all inconsistencies in the estimation of J^2 will go to zero as P_1 goes to zero and as the *a priori* restrictions become better approximations. Thus such inconsistencies will be negligible if P_1 is close to zero and all *a priori* restrictions are good enough approximations.

5 Second generalization of Proximity Theorem

Thus far we have assumed that all variables assumed to be exogenous actually are so. The time has now come to remove that restriction. Until further notice, we shall assume that all *a priori* restrictions are correctly specified and that no variables have been omitted from the system. More general cases will be considered below.

Consider the system of equations

$$u^1 = A^1 x = \begin{bmatrix} B^1 & C^1 \end{bmatrix} \begin{bmatrix} x^1 \\ x^2 \end{bmatrix} \quad (1.38)$$

where, as usual, B^1 is assumed nonsingular. Here x^1 is assumed endogenous, and x^2 is (incorrectly) assumed exogenous. We further assume that, were those assumptions correct, A^1 would be identified, that is, that there exist *a priori* restrictions, $A^1 \phi' = 0$ and that the rank of $A^1 \phi'$ is $r(A^1) - 1$.

Now, suppose that in reality (1.38) is part of a larger system:

$$\begin{bmatrix} u^1 \\ u^2 \end{bmatrix} = \begin{bmatrix} B^1 & G^1 \\ E^{21} & B^2 \end{bmatrix} \begin{bmatrix} x^1 \\ x^2 \end{bmatrix}. \quad (1.39)$$

For simplicity, we assume that B^2 is square and nonsingular, that is, that no additional variables are needed to explain x^2 . In short, we assume that (1.39) is self-contained. More complex and realistic cases will be considered below.

We have already investigated the effect of assuming that $E^{21} = 0$ on the estimates of the second block of equations in (1.39). We now consider the effect of the same assumption on the estimates of the first block. Note once more that this assumption is that of block recursiveness. Should we prove that the effects of the assumption are here negligible for E^{21} sufficiently close to zero, we shall be in a position to speak strongly about the effects of separating total systems for partial analysis.

Let the coefficient matrix of (1.39) be A and rewrite (1.39) as

$$u = Ax. \quad (1.40)$$

Then (assuming A to be nonsingular, which it must be for E^{21} sufficiently close to zero),

$$x = A^{-1}u. \quad (1.41)$$

However,

$$A^{-1} = \left[\begin{array}{c|c} (B^1 - G^1(B^2)^{-1}E^{21})^{-1} & -(B^1 - G^1(B^2)^{-1}E^{21})^{-1}G^1(B^2)^{-1} \\ \hline -(B^2 - E^{21}(B^1)^{-1}G^1)^{-1}E^{21}(B^1)^{-1} & (B^2 - E^{21}(B^1)^{-1}G^1)^{-1} \end{array} \right] \quad (1.42)$$

which reduces when $E^{21} = 0$ to

$$A^{-1} = \left[\begin{array}{c|c} (B^1)^{-1} & -(B^1)^{-1}G^1(B^2)^{-1} \\ \hline 0 & (B^2)^{-1} \end{array} \right] \quad (1.43)$$

so that, in the latter case, x^2 depends only on u^2 and not on u^1 . It follows that only when $E^{21} = 0$ is the assumption that x^2 consists of variables exogenous to (1.38) correct.

Now, as before, choose an indefinitely large set of values for the elements of u . Fix B^1 , G^1 , and B^2 . Consider a sequence of systems differing only in E^{21} and allow the latter matrix to go to zero. Note that (in view of (1.42)) the values of x^1 and x^2 , and hence all estimates, are continuous functions of the elements of E^{21} . Clearly, we can prove a theorem analogous to Theorem 1.1 if we can show that all estimates are continuous functions of the *assumed* value of E^{21} also.

This, however, is not hard to do. It will prove instructive to transform the problem slightly. Denote the true asymptotic variance-covariance matrix³⁶ of the elements of u^1 by $\bar{M}_{u^1u^1}$ and the asymptotic covariance matrix of the elements of x^2 and u^1 by $\bar{M}_{x^2u^1}$, the rows corresponding to the elements of x^2 and the columns to the elements of u^1 . Let Q be the matrix in the left-hand lower corner of A^{-1} in (1.42). Then, from (1.41) and the remarks in note 19 above,

$$\bar{M}_{x^2u^1} = Q\bar{M}_{u^1u^1}. \quad (1.44)$$

Furthermore, observe that, given $\bar{M}_{x^2u^1}$ and $\bar{M}_{u^1u^1}$, Q is determined by

(1.44) (we assume $\bar{M}_{u^1 u^1}$ to be nonsingular; we are choosing u^1). Given B^1 , G^1 , and B^2 , it is easy to show that E^{21} is determined by Q . It follows that, with everything else fixed, E^{21} – and hence the values of the elements of x – is determined as a continuous function of $\bar{M}_{x^2 u^1}$; further, that function is such that E^{21} is zero if and only if $\bar{M}_{x^2 u^1}$ is zero.

The assumption that the elements of x^2 are exogenous with respect to (1.38) is precisely the assumption that $\bar{M}_{x^2 u^1} = 0$. It is clear, however, that this is but a limiting assumption – one out of an infinite number that might be made. Given *any* value assumed for $\bar{M}_{x^2 u^1}$, and a method of estimation consistent for $\bar{M}_{x^2 u^1} = 0$, and making an appropriate stability assumption, estimates of A_1^1 will exist. Indeed, the latter estimates generally depend only on the first column of $\bar{M}_{x^2 u^1}$. If the true value of $\bar{M}_{x^2 u^1}$ and the assumed value coincide, such estimates will be consistent. This is most clearly seen in the case of the method of instrumental variables, where the covariances in question appear explicitly. Since limited-information maximum likelihood and two-stage least squares estimators are equivalent to the method of instrumental variables for appropriate choices of instruments,³⁷ the same properties obviously hold for them. Such properties also hold for the full-information maximum likelihood method.

Let the assumed value of $\bar{M}_{x^2 u^1}$ be $M_{x^2 u^1}$. Let the vector of the probability limits of the estimates of the elements of A_1^1 obtained by making that assumption and applying one of the standard methods³⁸ be $a_1^1(M_{x^2 u^1}, \bar{M}_{x^2 u^1})$. It is clear that each element of the latter vector is a continuous function of both (matrix) arguments. The following theorem now follows exactly as in the proof of Theorem 1.1.

Theorem 1.5 (Second Generalized Proximity Theorem):

$$\lim_{\bar{M}_{x^2 u^1} \rightarrow 0} a_1^1(0, \bar{M}_{x^2 u^1}) = A_1^1.$$

Further, since $\bar{M}_{x^2 u^1}$ and E^{21} are in a one-to-one relation and so are the assumed values thereof, we have a corollary (where the notation is obvious).

Corollary to Theorem 1.5:

$$\lim_{E^{21} \rightarrow 0} \alpha_1^1(0, E^{21}) = \lim_{E^{21} \rightarrow 0} a_1^1(0, \bar{M}_{x^2 u^1}) = A_1^1.$$

It thus follows that inconsistencies in the estimation of A_1^1 will be negligible for E^{21} close enough to zero – for the variables in x^2 “nearly” exogenous to (1.38).

We have called Theorem 1.5 another generalization of the Proximity

Theorem discussed in the last section. That it is so may be seen as follows. In the case of a single equation, $\bar{M}_{x^2u^1}$ consists of a single row, the vector of covariances between the single residual and the assumed exogenous variables. Any such covariance, however, may be written as the product of the corresponding correlation coefficient, the standard error of the residual, and the standard error of the variable in question. Hence $\bar{M}_{x^2u^1}$ will go to zero in this case as both the correlations between the independent variables and the residual and the variance of the residual go to zero, the two effects reinforcing each other.³⁹

We may now remove the assumption that (1.39) is self-contained. To begin with, it is clear that no problem whatsoever is created by the appearance of variables exogenous to (1.38) in the equations explaining x^2 ; we shall not include them explicitly to avoid unduly complicating our notation.

A problem does arise, however, when variables appear in the equations explaining x^2 which are not truly exogenous to (1.38). Call the vector of all such variables x^0 ; it is to be thought of as expanded to include variables explaining x^2 , variables explaining such variables, and so forth, until either a self-contained system is reached or all remaining unexplained variables are truly exogenous to (1.38). (As just stated, we may assume the first of these alternatives for simplicity.) Suppose that, instead of (1.39), (1.38) is a subsystem of

$$\begin{bmatrix} u^0 \\ u^1 \\ u^2 \end{bmatrix} = \begin{bmatrix} B^0 & E^{01} & E^{02} \\ 0^{10} & B^1 & G^1 \\ E^{20} & E^{21} & B^2 \end{bmatrix} \begin{bmatrix} x^0 \\ x^1 \\ x^2 \end{bmatrix} \quad (1.45)$$

where B^0 is assumed nonsingular. Note the appearance of 0^{10} ; we are still assuming that there are no omitted variables in (1.38) and that A_1^1 is identified under true *a priori* restrictions.

By an analysis similar to that just given it is not hard to show that all inconsistencies in estimating A_1^1 on the assumption that the elements of x^2 are exogenous will go to zero as both the matrices E^{21} and $E^{20}(B^0)^{-1}E^{01}$ go to zero. The interpretation of this condition is as follows. E^{21} , as before, represents the direct effects of x^1 on x^2 . Negligible inconsistency clearly requires that these effects be close to zero. The matrix $E^{20}(B^0)^{-1}E^{01}$, however, represents the indirect effects of x^1 on x^2 by way of the effects of x^1 on x^0 and x^0 on x^2 . For negligible inconsistency, these effects must be close to zero also.⁴⁰ The latter condition may be interpreted further.

Assume that $E^{21} = 0$. If $E^{20} = 0$, (1.45) is block recursive as it stands, and we have (1.39), the case already considered. In this case, we already know that $E^{21} = 0$ is a sufficient condition for consistency of all estimates. Now suppose that $E^{20} \neq 0$ but that $E^{01} = 0$. In this case, the

system is block recursive also, but x^0 and x^2 must be written in the same block. This is the case mentioned above, where the additional variables explaining x^2 are exogenous to (1.38). Finally, $E^{20}(B^0)^{-1}E^{01}$ can be near zero, even if E^{20} and E^{01} both contain nonnegligible elements. The full condition for this is that the j th column of E^{20} can be nonnegligible only if the j th row of E^{01} is negligible. In other words, any variable in x^0 which has nonnegligible influence on the variables in x^2 must be influenced only negligibly by all variables in x^1 .⁴¹

We observe that, as usual, nothing in the above analysis would be affected if the *a priori* restrictions held only approximately. We conclude that inconsistencies will be near zero if all *a priori* restrictions are sufficiently close approximations *and* the above conditions hold.

We may now remove the last restriction and consider the general case where there are variables omitted from (1.38). Without loss of generality, we may expand x^0 to include such variables and all variables explaining them and so forth until either a self-contained system is reached or all unexplained variables are exogenous to (1.38). For simplicity, we continue to assume the first of these alternatives.⁴² x^0 now contains all variables needed to explain x^1 and x^2 and all variables needed to explain such variables, and so forth. Replacing 0^{10} in (1.45) by E^{10} , we have the *perfectly general system*⁴³

$$\begin{bmatrix} u^0 \\ u^1 \\ u^2 \end{bmatrix} = \begin{bmatrix} B^0 & E^{01} & E^{02} \\ E^{10} & B^1 & G^1 \\ E^{20} & E^{21} & B^2 \end{bmatrix} \begin{bmatrix} x^0 \\ x^1 \\ x^2 \end{bmatrix}. \quad (1.46)$$

We continue to assume the B^i nonsingular; this is the general case.

It is easy to show that the conditions given for the near exogeneity of x^2 to the second block of equations in (1.46) and hence for negligible inconsistencies are unchanged. They are the same as those just discussed in the case (1.45) (of course, the magnitudes of the inconsistencies involved may be different; the point is that the limit – zero – is the same in both cases). Further, as observed above, all inconsistencies will go to zero as both those conditions are satisfied *and* as all *a priori* restrictions are better approximations. It was shown in Section 3 above, however, that estimation of the first equation of the second block of (1.46) on the assumption that $E_1^{10} = 0$, that is, on the erroneous omission of x^0 , is in general equivalent to estimation with erroneous *a priori* restrictions, and hence that as E_1^{10} goes to zero so do the effects of this misspecification. It follows that all our theorems hold concurrently (a similar statement holds for Theorem 1.4). Thus, inconsistencies in the estimation of the first row of the second block of equations in (1.46) will be negligible if all the following conditions hold.

1. All *a priori* restrictions are close approximations.

2. E_1^{10} is close to zero; that is, omitted variables have small coefficients (this is sufficient save on a space of measure zero).⁴⁴
3. E^{21} is near zero; that is, the endogenous variables have negligible direct effects on the assumed exogenous variables.
4. $E^{20}(B^0)^{-1}E^{01}$ is close to zero; that is, the endogenous variables have negligible indirect effects on the assumed exogenous variables.

If, in addition, the other rows of E^{10} are near zero, so that all coefficients of omitted variables in the block of equations investigated are small, then conditions 2–4 imply that the system is nearly block recursive either as it stands or as rewritten with the first and third blocks of equations grouped together. In either case, analysis of the second block of equations is possible without explicit reference to the total system of which it is a part. *It is the existence of situations such as this which permits estimation of partial economic models*, indeed, which permits estimation of general economic models which are in turn embedded in models of the socio-physical universe. Liu's objections to simultaneous equation estimation are thus not generally damning. They cannot be taken as general objections, but as highly important criticisms which must be considered case by case in the light of how good the approximations involved are, not in the light of whether such approximations are exactly true or not.

Having said this much, however, we must go on to inquire whether the four conditions named above are easily fulfilled in practice. It seems clear that the first three are likely to be; any reasonably attentive investigator will pay fairly close attention to them. The fourth condition, on the other hand, is not so easily disposed of. It is very easy to overlook the possibility that the endogenous variables may have sizeable effects on the assumed exogenous variables by way of effects on a third set, which set may be exceedingly large. Therefore this is the point to which close attention must be paid in practice.

Finally, we may draw a further conclusion. Liu's objections, though not valid as stated, point to a redirection of some current work. The theorems in this paper speak only of consistency. They have no direct bearing on the variance of estimators nor on the small-sample properties thereof. We have seen that different estimators can have different probability limits under misspecifications of the types discussed here. It is likely that their asymptotic variances are different and almost certain that their small sample properties are different. Even if the approximations involved in such misspecifications are very good (as they must be for negligible inconsistencies) one estimator may do better than another. That is, one estimator may be less sensitive than another to errors of this kind. The choice among estimators may therefore depend on their behavior under approximative misspecification, even though negligible

inconsistencies are assured for all. To put it slightly differently, the goodness of a "good enough" approximation may depend, particularly in small samples, on the estimator used. Since the situations envisaged by Liu arise commonly (as he convincingly argues) it follows that, even though his specific conclusions are wrong, it is imperative that work be done along the lines indicated. In particular, a good deal of future work in Monte Carlo experiments should be directed towards discovering the small-sample properties of various estimators under the types of approximate misspecification pointed out by Liu and discussed here.

6 Conjectures on dynamics

A good deal of this paper has been concerned with estimation problems in equation systems that are close to being block recursive. We have seen that such closeness is a sufficient condition to permit estimation of the usual kind to take place despite Liu's objections. Since such closeness is also a necessary condition, in general, it seems clear that a large part of the profession implicitly believes the world to be constructed in this way. Indeed, it is interesting that, considered from this point of view, Liu's objections to simultaneous equation estimation lie at the other extreme from the better-known objections of Wold. Wold has insisted that the real world is recursive in the ordinary sense. Liu's objections, if they are to hold, must imply that the real world is not generally recursive in any sense. As just observed, the middle ground that the real world is nearly block recursive is implicitly held by those practicing simultaneous equation methods.

If the real world is nearly block recursive, however, it becomes interesting to consider the dynamics of block-recursive systems. A first step in this direction is the consideration of the dynamics of a first-degree difference or differential equation whose matrix is block triangular. Explicitly, consider the equation system

$$x(t) \equiv \begin{bmatrix} x^1(t) \\ x^2(t) \\ \vdots \\ x^N(t) \end{bmatrix} = \begin{bmatrix} B^1 & G^1 & & \\ 0^{21} & B^2 & G^2 & \\ \vdots & \vdots & \vdots & \vdots \\ 0^{N1} & 0^{N2} & 0^{N3} & \vdots \end{bmatrix} \begin{bmatrix} x^1(t-1) \\ x^2(t-1) \\ \vdots \\ x^N(t-1) \end{bmatrix} \equiv Ax(t-1) \quad (1.47)$$

where the B^i are nonsingular. Clearly, the subsystems obtained by dropping the first k rows and columns of the matrices and the correspond-

ing x^i have the property that their dynamic properties may be investigated without regard for the remainder of the system.

Now consider matrices that are nearly block triangular, that is, matrices such as

$$A(\eta) = A + \eta D \quad (1.48)$$

where η is a positive scalar and D is any matrix with finite elements. We are interested in the dynamic properties of systems with such matrices for small η .

Now, a special case of (1.47) arises when all the G^i are zero. In this case, A is completely decomposable, and $A(\eta)$ is nearly completely decomposable for small η . For nearly completely decomposable matrices, Simon and Ando [13] have proved an important theorem. Crudely interpreted, this theorem states that equilibrium in a system with a nearly completely decomposable matrix (for small η) is approached in three stages. During the first stage the variables within each $x^i(t)$ adjust to reach equilibrium positions that are related to each other according to the limiting completely decomposable system – to the B^i . Following this, in the second stage, maintaining such equilibrium at all times, the variables adjust *en bloc* – the variables with each $x^i(t)$ moving together – until at the final stage, as in any linear system, the variables approach the rate of growth corresponding to the largest latent root. These properties permit aggregation of the variables within each $x^i(t)$ during and after the second stage.

It is clear that nearly completely decomposable matrices are a special case of nearly block-triangular matrices. Indeed, the latter are far more apt to arise in practice. It follows that all our theorems about estimation inconsistency apply to nearly completely decomposable matrices. Further, it seems clear that the problem of estimation inconsistency and the problem of equilibrium adjustment are related here. Both the Simon–Ando theorem and the theorems proved in this paper concern the strength of feedback effects. It seems likely that this relationship is not accidental. We therefore offer the following conjectures for further investigation.

1. The Simon–Ando theorem holds, *mutatis mutandis*, for nearly block-triangular and not just for nearly completely decomposable systems.⁴⁵
2. If the first conjecture is true, there exists a direct proof thereof which depends on the theorems on estimation here proved.
3. If the first two conjectures are true, a general theorem exists concerning the relationship between negligible inconsistency for negligible misspecification on the one hand and the equilibrium properties of closely related systems on the other.

Since block-triangular systems may be fairly generally encountered, the first conjecture alone is an important one. If linkages are far stronger one way than another between related markets, for example, this conjecture, if true, would provide valuable information about the dynamic properties of the whole system of markets considered and would allow easier handling of the solution of the system.

Notes

1. See, for example, Sargan [11] and Theil [17], Chapter 6.
2. Liu [8], pp. 436–7, 464–6.
3. For example, in his contribution to the Roundtable Discussion of Simultaneous Estimation Techniques at the Chicago Meeting of the Econometric Society, December 1958 published as [10]. Liu has also given his detailed argument in [9]. A partial but still inadequate reply along the lines of the present chapter is given in Valavanis [19], pp. 130–2.
4. See Koopmans, Rubin and Leipnik [6], pp. 79, 81–2. Actually, Liu generally states the order condition in terms of the number of endogenous variables included in and the number of exogenous variables excluded from a given equation, which is an equivalent statement.
5. See Fisher [3].
6. That the obvious is not always true in this area has been shown by R. H. Strotz who has recently provided us with an example of a case where the limit of the maximum likelihood estimator as specification error goes to zero is *not* the maximum likelihood estimator of the limit. See [15], but cf. p. xiii, above.
7. [20], Chapter 2, for example.
8. Although, of course, this does not remove the possibility that either Liu or Wold is correct; the real world may be constructed as one of them describes it. The point is that it need not be so constructed and that it is very plausible that it should not.
9. In fact, ϕ' can be any constant matrix so that all remarks and theorems apply *mutatis mutandis* to any consistent set of misspecified linear homogeneous restrictions. The case given in the text is the usual one of excluding variables from equations and is thus of greatest interest. Linear but inhomogeneous restrictions could obviously be handled with but a slight change in notation. Nonlinear restrictions present certain formal complexities which would unduly obscure the discussion, but the basic theorem is always the same. For a general treatment of the last two types of restrictions see Fisher [3].
10. By Fisher [3], pp. 436–7, Lemmas 4 and 5, this is always possible no matter what ϕ' is, so long as the rank condition holds.
11. It is shown in what follows that two different estimation techniques (in particular, two-stage least squares and limited information) for both of which $a_{ij}(0, 0) = A_{ij}$ (in general, $a_{ij}(\bar{\eta}, \bar{\eta}) = A_{ij}$) may have different $a_{ij}(\eta, \bar{\eta})$ for $\eta \neq \bar{\eta}$. We shall discuss this below; however, it would unduly complicate the notation to take explicit account of it here.
12. [17], pp. 227–9, 334–6.
13. I am indebted to H. Uzawa for calling this problem to my attention by

pointing out the incompleteness of an earlier proof that ignored it and to J. D. Sargan for helpful discussion of the present proof. The stability assumption is implicitly made in all later theorems.

In general, the proof given will be valid for the nonstable consistent estimator case if T can be replaced by some $f(T)$ such that the required probability limits exist and are nonzero. We should ordinarily expect this, provided that the estimator involved is consistent under correct specification for $\bar{\eta}$ in some neighborhood of the zero vector.

14. [17], pp. 341–2.
15. For example, where k depends only on the number of variables in $y^1(t)$ and $z^1(t)$ and on the total number of exogenous variables in the system. See [17], p. 229.
16. [17], pp. 331–3.
17. See H. Wold in association with L. Juréen [20], pp. 14, 49–53, and elsewhere.
18. The term is evidently a natural one. Since first using it, it has been pointed out to me by Jerome Rothenberg that “block triangularity” has already been suggested by Walter Jacobs and used by George Dantzig to describe rather similar matrices occurring in linear programming (see Dantzig [2], p. 176). Save for the requirement that the R^i be nonsingular, my block triangularity property is the same as the well-known property of decomposability (see Solow [14]), a fact that will be of interest below.
19. This assumes that the elements of any u^i are independent of the elements of any u^j for $j \neq i$. Actually, since we shall be explicitly investigating the effects of all omitted variables, we could well assume that all elements of u are independent of each other. This is an innocuous assumption since any dependence here can always be expressed in terms of some variable wrongly omitted from two or more equations (and thus present in the error term). Of course, this assumption is only meant to apply to cases where explicit account is taken of all omitted variables. Aside from being innocuous, however, this strong assumption is unnecessary; all that is needed for the results is the assumption in the first sentence of this note for block-recursive systems and for systems supposed to be block recursive. A similar problem arises with ordinary recursive systems, and a similar assumption is always made in practice – although without the explicit sufficient justification given here, since omitted variables are not excluded from the error term. See Wold and Juréen [20], pp. 52–3. Of course, the theorems given below provide justification for this practice when the specification of no omitted variables is approximately correct.
20. [12]. The question of whether any system is really part of a self-contained system is one which need not concern us here.
21. [16], pp. 421–2.
22. [4]. See also Solow [14].
23. A similar remark to that made in note 11 above applies here.
24. It is necessary to add that in the present case V in (1.10) refers to residuals from regressions on the elements of z^2 only, rather than on all exogenous variables. A similar remark applies to the interpretation of M_1 and M in (1.15). This amounts to using the elements of z^2 as instrumental variables, and the remainder of the proof follows as before.
25. This is a special case of the general remark made in Koopmans, Rubin and Leipnik [6], p. 83.
26. The question may be raised of why estimation of J_1^2 is desirable in these

circumstances. Since we shall show below that the estimates obtained are estimates of the gross effect of the variables in x^2 (both their direct effect and their effect through the variables in x^1), are not such estimates to be desired for all purposes? The answer here is the same as in the case of reduced form versus structural estimation (indeed, the question is a special case of that question): measurement of such gross effects may indeed be helpful for prediction, but in the event of a partial structural break that disturbs the relation between the included and the (mistakenly) omitted variables, the use of gross estimates leads to serious error, whereas true structural knowledge can still be used. An example would be the imposition of a tax on one of the variables.

27. Should this exceptional case occur, in particular should it be the first element of v that is distributed independently of z^2 , it may happen that the first equation of (1.31) is underidentified unless there are more than enough *a priori* restrictions to begin with. This is so because the first row of (1.32) will be indistinguishable in form from the first row of (1.31). (This does not otherwise arise because of the dependence of the elements of v and of z^2 .) The exceptional case is of no importance; one could never know that it was present in any practical problem. The possibility of its existence does not substantially affect our discussion, save as a curiosity, since we are concerned with the more interesting case where rows of P are near zero. The possible existence of the exceptional case, however, does mean that we are stating sufficient but not always necessary conditions for negligible inconsistency.
28. A similar remark to that given in note 11 above applies here.
29. Special cases that were worth proving separately, nonetheless, because of the way in which the omission of variables enters as providing just sufficient *a priori* restrictions.
30. Note that this is equivalent to the statement that the system can be put in block-recursive form (possibly with three blocks) by an appropriate renumbering of equations and variables.
31. [20], p. 189, Theorem 12.1.3. Also see [20], pp. 37–8; more precise results have been obtained by Wold and Faxér [21].
32. We can only deal with consistency in general because simultaneous equation estimators are not generally unbiased under correct specification.
33. It is interesting to observe that taken together the results of the two sections imply that the conditions of the Proximity Theorem imply that the system being studied is nearly recursive or block recursive, with the equation being estimated forming one block. It follows that a defense of least squares based on the Proximity Theorem differs in degree but not fundamentally in kind from the defense that the real world is recursive. (An exception to this occurs when there are no omitted variables and the variance of the true structural residual is near zero.)
34. Although useful results may be obtained in the single-equation case using the correlation coefficients alone; see Harberger [5].
35. A clear statement of the theorem in the single-equation case in terms of omitted variables and regression coefficients has been given by Lintner [7], Chapter II. As usual in this area of analysis Theil has proved a general theorem (for a single equation) of which this is a special case ([18], p. 43).
36. By the strong assumption in note 19 above, this is a diagonal matrix. Equation (1.44) below follows from the weak assumption in note 19; it allows us to forget u^2 in computing covariances, which is all that is necessary.

37. See Sargan [11], p. 393, and Theil [17], p. 336.
38. A remark similar to that made in note 11 above applies here.
39. It is truly the variance of the residual and not just its standard error that is important here since, other residual variances remaining equal, the variance of the independent variable involved is a linear function of the variance of the residual in question, by the remarks in note 19 above.
40. Actually, inconsistencies can be negligible if the indirect and direct effects just cancel out, that is, if $(E^{21} - E^{20}(B^0)^{-1}E^{01}) = 0$. This could only happen by accident if the conditions named in the text are not satisfied; moreover, one could never tell whether it were nearly the case in practice. We thus disregard it.
41. Note that the full condition also implies that the system is nearly block recursive.
42. The additional remarks necessary if the second alternative holds are given in Section 3 above.
43. Perfectly general, save for the self-contained feature just discussed.
44. Alternatively, if only estimates of B_1^1 and G_1^1 are required (greatly inconsistent estimates of E_1^{10} are acceptable), we might have:
 - 2'. The first row of the matrix $E^{10}(B^0)^{-1}[E^{01} E^{02}]$ is near zero, that is, any omitted variable with a nonnegligible coefficient is only negligibly influenced by the included variables.
45. Since this was written, Professor Ando and I have proved that this conjecture is indeed correct. See A. Ando and F. M. Fisher, "Near-decomposability, partition and aggregation, and the relevance of stability discussions," and the other papers in Ando, Fisher, and Simon [1].

References

- [1] Ando, A., F. M. Fisher, and H. A. Simon, *Essays on the Structure of Social Science Models*, Cambridge, MA: MIT Press, 1963.
- [2] Dantzig, G. B., "Upper bounds, secondary constraints, and block triangularity in linear programming," *Econometrica*, 23 (2) (April 1955), 174-83.
- [3] Fisher, F. M., "Generalization of the rank and order conditions for identifiability," *Econometrica*, 27 (3) (July 1959), 431-47.
- [4] Goodwin, R. M., "Dynamical coupling with especial reference to markets having production lags," *Econometrica*, 15 (3) (July 1947) 181-204.
- [5] Harberger, A. C., "On the estimation of economic parameters," Cowles Commission Discussion Paper: Economics No. 2088, 1953 (unpublished).
- [6] Koopmans, T. C., H. Rubin, and R. B. Leipnik, "Measuring the equation systems of dynamic economics," Chapter II in T. C. Koopmans (ed.), *Statistical Inference in Dynamic Economic Models* (Cowles Commission Monograph 10), New York: John Wiley, 1950.
- [7] Lintner, J., "Dividends, earnings, leverage, stock prices and the supply of capital to corporations," *Review of Economics and Statistics*, 44 (3) (August 1962), 243-69.
- [8] Liu, T. C., "A simple forecasting model for the U.S. economy," *International Monetary Fund Staff Papers* (August 1955), 434-66.
- [9] Liu, T. C., "Structural estimation and forecasting: a critique of the Cowles Commission method," *Tsing Hua Journal* (issue in honor of Franklin Ho) (June 1963).

- [10] Liu, T. C., "Underidentification, structural estimation, and forecasting," *Econometrica*, 28 (4) (October 1960), 855-65.
- [11] Sargan, J. D., "The estimation of economic relationships using instrumental variables," *Econometrica*, 26 (3) (July 1958), 393-415.
- [12] Simon, H. A., "Causal ordering and identifiability," Chapter 3 in W. C. Hood and T. C. Koopmans (eds), *Studies in Econometric Method* (Cowles Commission Monograph 14), New York: John Wiley, 1953; reprinted as Chapter 1 in H. A. Simon, *Models of Man*, New York: John Wiley, 1957.
- [13] Simon, H. A., and A. Ando, "Aggregation of variables in dynamic systems," *Econometrica*, 29 (2) (April 1961), 111-38.
- [14] Solow, R. M., "On the structure of linear models," *Econometrica*, 20 (1) (January 1952), 29-46.
- [15] Strotz, R. H., "Interdependence as a specification error," *Econometrica*, 28 (2) (April 1960), 428-42.
- [16] Strotz, R. H., and H. Wold, "Recursive vs. Nonrecursive systems: an attempt at synthesis," *Econometrica*, 28 (2) (April 1960), 417-27.
- [17] Theil, H., *Economic Forecasts and Policy*, Amsterdam: North-Holland, 1958.
- [18] Theil, H., "Specification errors and the estimation of economic relationships," *Review of the International Statistical Institute*, 25 (1957), 41-51.
- [19] Valavanis, S., *Econometrics*, New York: McGraw-Hill, 1959.
- [20] Wold, H., in association with L. Juréen, *Demand Analysis*, New York: John Wiley, 1953.
- [21] Wold, H., and P. Faxér, "On the specification error in regression analysis," *Annals of Mathematical Statistics*, 28 (1) (March 1957), 265-7.