Example 1. Market for Ph.D. economists.

q = log number employed,
w = log wage rate,
s = log college enrollment
m = log median wage of lawyers.

***demand* in year t**

(1) $$q_t = \beta_{11} + \beta_{12}s_t + \beta_{13}w_t + \varepsilon_{1t} ;$$

***supply* in year t**

(2) $$q_t = \beta_{21} + \beta_{22}m_t + \beta_{23}w_t + \beta_{24}q_{t-1} + \varepsilon_{2t} ;$$

***structural simultaneous equations system*.**
college enrollments s_t and lawyer salaries m_t are *exogenous*. (1) and (2) are a *complete* system for the determination in market equilibrium of the two *endogenous* or dependent variables q_t and w_t.

Fig. 1. Demand & Supply of Economis

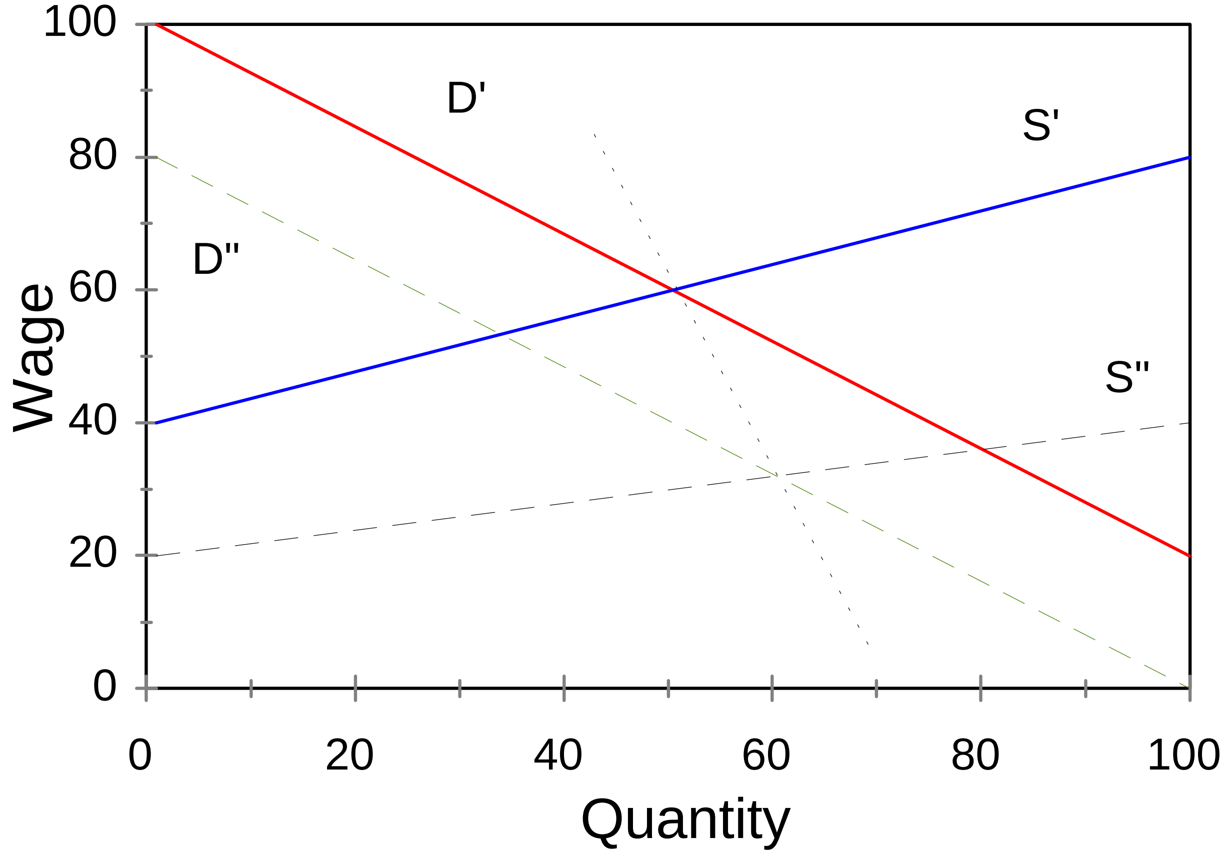

Figure 1 shows the demand and supply curves corresponding to (1) and (2), with w and q determined by market equilibrium. Two years are shown, with solid curves in the first year and dashed curves in the second. The equilibrium wage and quantity are determined by the condition that the market clear.

Suppose you are interested in the demand equation, and have data on the variables appearing in (1) and (2). How could you obtain good statistical estimates of the demand equation parameters?

Think of the "experiment" run by Nature, versus the experiment that you would ideally like to carry out to form the estimates.

If both the demand and supply curves shift between periods due to random disturbances, then the locus of equilibria will be a scatter of points (in this case, two) which will not in general lie along either the demand curve or the supply curve. In the case illustrated, the dotted line which passes through the two observed equilibria has a slope substantially different than the demand curve. If the disturbances mostly shift the demand curve and leave the supply curve unchanged, then the equilibria will tend to map out the supply curve. Only if the disturbances mostly shift the supply curve and leave the demand curve unchanged will the equilibria tend to map out the demand curve.

Consequences.

An OLS fit of equation (1) will produce a line like the dotted line in the figure that is a poor estimate of the demand curve. Only when most of the shifts over time are coming in the supply curve so that the equilibria lie along the demand curve will least squares give satisfactory results.

Exogenous variables shift the demand and supply curve in ways that can be estimated. In particular, the variable m that appears in the supply curve but not the demand curve shifts the supply curve, so that the locus of w,q pairs swept out when only m changes lies along the demand curve.

The ideal experiment you would like to run in order to estimate the slope of the demand curve is to vary m, holding all other things constant. Put another way, you need to find a statistical analysis that mimics the ideal experiment by isolating the partial impact of the variable m on both q and w.

The structural system (1) and (2) can be solved for q_t and w_t as functions of the remaining variables

$$\text{(3)} \qquad w_t = \frac{\beta_{11} - \beta_{21} + \beta_{12}s_t - \beta_{22}m_t - \beta_{24}q_{t-1} + \varepsilon_{1t} - \varepsilon_{2t}}{\beta_{23} - \beta_{13}}$$

$$\text{(4)} \qquad q_t = \frac{\beta_{11}\beta_{23} - \beta_{21}\beta_{13} + \beta_{23}\beta_{12}s_t - \beta_{13}\beta_{22}m_t - \beta_{13}\beta_{24}q_{t-1}}{\beta_{23} - \beta_{13}} + \frac{\beta_{23}\varepsilon_{1t} - \beta_{13}\varepsilon_{2t}}{\beta_{23} - \beta_{13}}$$

Equations (3) and (4) are called the *reduced form*. For this solution to exist, we need $\beta_{23} - \beta_{13}$ non-zero. This will certainly be the case when the elasticity of supply β_{23} is positive and the elasticity of demand β_{13} is negative. Hereafter, assume that the true $\beta_{23} - \beta_{13} > 0$.

Equations (3) and (4) constitute a system of regression equations, which could be rewritten in the stacked form

$$
(5) \quad \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_T \\ q_1 \\ q_2 \\ \vdots \\ q_T \end{bmatrix} = \begin{bmatrix} 1 & s_1 & m_1 & q_0 & 0 & 0 & 0 & 0 \\ 1 & s_2 & m_2 & q_1 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & s_T & m_T & q_{T-1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & s_1 & m_1 & q_0 \\ 0 & 0 & 0 & 0 & 1 & s_2 & m_2 & q_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 1 & s_T & m_T & q_{T-1} \end{bmatrix} \begin{bmatrix} \pi_{11} \\ \pi_{12} \\ \pi_{13} \\ \pi_{14} \\ \pi_{21} \\ \pi_{22} \\ \pi_{23} \\ \pi_{24} \end{bmatrix} + \begin{bmatrix} \nu_{11} \\ \nu_{12} \\ \vdots \\ \nu_{1T} \\ \nu_{21} \\ \nu_{22} \\ \vdots \\ \nu_{2T} \end{bmatrix} ,
$$

or

$$
\mathbf{y} = \mathbf{Z}\pi + \nu ,
$$

where the π's are the combinations of behavioral coefficients, and the ν's are the combinations of disturbances, that appear in (3) and (4). The system (5) can be estimated by GLS. In general, the disturbances in (5) are correlated and heteroskedastic across the two equations. However, exactly the same explanatory variables appear in each of the two equations. If the disturbances are uncorrelated across time, so that $E\nu_{it}\nu_{js} = \sigma_{ij}\delta_{ts}$, or $E\nu\nu' = I_T \otimes \Sigma$, then GLS using this covariance structure collapses to OLS, the seemingly unrelated regression case.

Suppose you are interested in estimating the parameters of the behavioral demand equation (1). For OLS applied to (1) to be consistent, it is necessary that the disturbance ε_{1t} be uncorrelated with the right-hand-side variables, which are s_t and w_t. This condition is met for s_t, provided it is indeed exogenous. However, from (3), an increase in ε_{1t} increases w_t, other things being equal, and in (1) this results in a positive correlation of the RHS variable w_t and the disturbance ε_{1t}.

Instrumental variables estimation is one alternative for the estimation of (1). In this case, one needs to introduce at least as many instrumental variables as there are RHS variables in (1), and these variables need to be uncorrelated with ε_{1t} and fully correlated with the RHS variables. The list of instruments should include the exogenous variables in (1), which are the constant, 1, and s_t. Other candidate instruments are the exogenous and predetermined variables elsewhere in the system, m_t and q_{t-1}.

Will IV work? In general, to have enough instruments, there must be at least as many predetermined variables excluded from (1) and appearing elsewhere in the system as there are endogenous variables on the RHS of (1). When this is true, (1) is said to satisfy the *order condition* for identification. In the example, there is one RHS endogenous variable, w_t, and two excluded exogenous and predetermined variables, m_t and q_{t-1}, so the order condition is satisfied. If there are enough instruments, then from the general theory of IV estimation, the most efficient IV estimator is obtained by first projecting the RHS variables on the space spanned by the instruments, and then using these projections as instruments. In other words, the best combinations of instruments are obtained by regressing each RHS variable in (1) on the instruments 1, s_t, m_t, and q_{t-1}, and then using the fitted values from these regressions as instruments. But the reduced form equation (3) is exactly this regression. Therefore, the best IV estimator is obtained by first estimating the reduced form equations (3) and (4) by OLS and retrieving fitted values, and then estimating (1) by OLS after replacing RHS endogenous variables by their fitted values from the reduced form. For this to yield instruments that are fully correlated with the RHS variables, it must be true that at least one of the variables m_t and q_{t-1} truly enters the reduced form, which will happen if at least one of the coefficients β_{22} or β_{24} is nonzero. This is called the *rank* condition for identification.

2. STRUCTURAL AND REDUCED FORMS

In general a behavioral or structural simultaneous equations system can be written

$$(6) \qquad y_t'B + z_t'\Gamma = \varepsilon_t',$$

where $y_t' = (y_{1t}, .,y_{Nt})$ is a 1×N vector of the endogenous variables, B is a N×N array of coefficients, $z_t' = (z_{n1}, .,z_{Mt})$ is a 1×M vector of predetermined variables, Γ is a M×N array of coefficients, and ε_t' is a 1×N vector of disturbances. Let Σ denote the N×N covariance matrix of ε_t. The reduced form for this system is

$$(7) \qquad y_t' = z_t'\Pi + \nu_t',$$

where $\Pi = -\Gamma B^{-1}$ and $\nu_t' = \varepsilon_t'B^{-1}$, so that the covariance matrix of ν_t is $\Omega = B'^{-1}\Sigma B^{-1}$. Obviously, for (6) to be a well-defined system that determines y_t, it is necessary that B be non-singular.

3. IDENTIFICATION

Some restrictions must be imposed on the coefficient arrays B and Γ, and possibly on the covariance matrix Σ, if the remaining coefficients are to be estimated consistently. First, post-multiplying (6) by a nonsingular diagonal matrix leaves the reduced form solution (7) unchanged, so that all versions of (6) that are rescaled in this way are observationally equivalent. Then, for estimation of (6) it is necessary to have a scaling normalization for each equation. B, Γ, and Σ contain N(N-1) + NM + N(N+1)/2 parameters, excluding the N parameters set by the scaling normalizations and taking into account the symmetry of Σ. However, Π and Ω contain only NM + N(N+1)/2 parameters. Therefore, an additional N(N-1) restrictions on parameters are necessary to determine the remaining structural parameters from the reduced form parameters.

It is traditional to define *order* and *rank* conditions for identification. These come from the structure of the B and Γ matrices and the conditions $\Pi B + \Gamma = 0$ and $B'\Omega B = \Sigma$ relating the reduced form coefficients to the structural parameters. But it is simpler to think of identification in terms of the possibility for IV estimation: *An equation (with associated restrictions) is identified if and only if there exists a consistent IV estimator for the parameters in the equation; i.e., if there are sufficient instruments for the RHS endogenous variables that are fully correlated with these variables.*

Even covariance matrix restrictions can be used in constructing instruments. For example, if you know that the disturbance in an equation you are trying to estimate is uncorrelated with the disturbance in another equation, then you can use a consistently estimated residual from the second equation as an instrument. If you are not embarrassed to let a computer do your thinking, you can even leave identification to be checked numerically: an equation is identified if and only if you can find an IV estimator for the equation that empirically has finite variances.

Exercise 1. Show that the condition above requiring N(N-1) restrictions on parameters will hold if the *order condition*, introduced in the example of the market for economists, holds for each equation. In the general case, the order condition for an equation states that the number of excluded predetermined (including strictly exogenous) variables is at least as great as the number of included RHS endogenous variables. Add the number of excluded RHS endogenous variables to each side of this inequality, and sum over equations to get the result.

4. 2SLS

For discussions of estimators for simultaneous equations systems, it is convenient to have available the systems (6) and (7) stacked two different ways. First, one can stack (6) and (7) vertically by observation to get

(8) $$YB + Z\Gamma = \varepsilon$$

and

(9) $$Y = Z\Pi + \nu,$$

where Y, ε, and ν are T×N and Z is T×K. With this stacking, one has $E\varepsilon'\varepsilon/T = \Sigma$ and $E\nu'\nu/T = B^{-1}\Sigma B'^{-1}$. Note that post-multiplying (8) by a non-singular diagonal matrix leaves the reduced form unchanged; hence this modification is observationally equivalent. Then, we can choose any convenient diagonal matrix as a normalization. In particular, we can renumber the equations and rescale them so that the dependent variable y_{nt} appears with a coefficient of one in the n-th equation. This is equivalent to saying that we can write B = I - A, where A is a matrix with zeros down the diagonal, and that the behavioral system (8) can be written

(10) $$Y = YA - Z\Gamma + \varepsilon \equiv [Y \mid Z]\begin{bmatrix} A \\ -\Gamma \end{bmatrix} \equiv XC + \varepsilon.$$

In this setup, Y and ε are T×N, X is T×(N+K), and C is (N+K)×N.

Restrictions that exclude some variables from some equations will force some of the parameters in C to be zero. Rewrite the n-th equation from (10), taking these restrictions into account, as

(11) $$y_n = Y_n A_n - Z_n \Gamma_n + \varepsilon_n \equiv X_n C_n + \varepsilon_n,$$

where this equation includes M_n endogenous variables and K_n predetermined variables on the RHS. Then, y_n is $T\times 1$, Y_n is $T\times M_n$, and Z_n is $T\times K_n$, and X_n is $T\times(M_n+K_n)$.

A second method of stacking which is more convenient for empirical work is to write down all the observations for the first equation, followed by all the observations for the second equation, etc. This amounts to starting from (11), and stacking the T observations for the first equation, followed by the T observations for the second equation, etc. Since the C_n differ across equations, the stacked system looks like

$$\text{(12)} \qquad \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & X_N \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix} \equiv \mathbf{Xc} + \mathbf{e}.$$

Note that X in (12) is <u>not</u> the same as X in (10); X is NT×J, where $J = J_1 + . + J_N$ and $J_n = M_n + K_n$ is the number of RHS variables in the n-th equation. The system (12) has the appearance of a system of regression equations. Because of RHS endogenous variables, OLS will not be consistent, so that we have to turn to IV methods. In addition, there are GLS issues due to the correlation of disturbances across equations.

Suppose you are interested in estimating a single equation from the system, say

$$y_1 = Y_1A_1 - Z_1\Gamma_1 + \varepsilon_1 \equiv X_1c_1 + \varepsilon_1.$$

The IV method states that if you can find instruments W that are uncorrelated with ε_1 and fully correlated with X_1, then the best IV estimator,

$$\hat{c}_1 = [X_1'W(W'W)^{-1}W'X_1]^{-1}X_1'W(W'W)^{-1}W'y_1$$

is consistent. But the potential instruments for this problem are $Z = [Z_1 \mid Z_{-1}]$, where Z_{-1} denotes the predetermined variables that are in Z, but not in Z_1. The *order* condition for identification of this equation is that the number of variables in Z_{-1} be at least as large as the number of variables in Y_1, or *the number of excluded predetermined must be as large as the number of included RHS endogenous*. The *rank* condition is that X_1' W be of maximum rank. For consistency, you need to have $X_1'W/T$ converging in probability to a matrix of maximum rank.

Exercise 2. Show that the rank condition implies the order condition. Show in the supply and demand for economists that the order condition can be satisfied, but the rank condition can fail, so that the order condition is necessary but not sufficient for the rank condition.

The best IV estimator can be written $\hat{c}_1 = [X_{1e}{}'X_{1e}]^{-1}X_{1e}{}'y_1$, where $X_{1e} = W(W'W)^{-1}W'X_1$ is the array of fitted values from an OLS regression of X_1 on the instruments W = Z; i.e., the reduced form regression. Then, the estimator has a *two-stage OLS* (2SLS) interpretation:

> **(1) Estimate the reduced form by OLS, and retrieve the fitted values of the endogenous variables.**
>
> **(2) Replace endogenous variables in a behavioral equation by their fitted values from the reduced form, and apply OLS.**

Recall from the general IV method that the procedure above done by conventional OLS programs will not produce consistent standard errors. Correct standard errors can be obtained by first calculating residuals from the 2SLS estimators in the original behavioral model, $u_1 = y_1 - X_1\hat{c}_{2SLS}$, estimating $\hat{\sigma}^2 = u_1{}'u_1/(T\text{-}K_1)$, and then estimating $V_e(\hat{c}_{2SLS}) = \hat{\sigma}^2[X_1{}'X_1]^{-1}$.

5. 3SLS

The 2SLS method does not exploit the correlation of the disturbances across equations. You saw in the case of systems of regression equations that using FGLS to account for such correlations improved efficiency. This will also be true here. To motivate an estimator, write out all the moment conditions available for estimation of each equation of the system:

$$\textbf{(13)}\qquad \begin{bmatrix} Z'y_1 \\ Z'y_2 \\ \vdots \\ Z'y_N \end{bmatrix} = \begin{bmatrix} Z'X_1 & 0 & \dots & 0 \\ 0 & Z'X_2 \dots & & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & Z'X_N \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{bmatrix} + \begin{bmatrix} Z'\varepsilon_1 \\ Z'\varepsilon_2 \\ \vdots \\ Z'\varepsilon_N \end{bmatrix}$$

or

$$(\mathbf{I}_N \otimes \mathbf{Z}')\mathbf{y} \equiv [(\mathbf{I}_N \otimes \mathbf{Z}')\mathbf{X}]\mathbf{c} + (\mathbf{I}_N \otimes \mathbf{Z}')\boldsymbol{\varepsilon}.$$

The disturbances in the NK×1 system (13) have the covariance matrix $\Sigma \otimes (\mathbf{Z}'\mathbf{Z})$. Then, by analogy to GLS, the best estimator for the parameters should be

$$\hat{\mathbf{c}}_{3SLS} = \{X'(I_N \otimes Z)(\Sigma^{-1} \otimes (Z'Z)^{-}\ \mathbf{X}'(\mathbf{I}_N \otimes \mathbf{Z})(\Sigma^{-1} \otimes (\mathbf{Z}'\mathbf{Z})^{-1})(\mathbf{I}_N \otimes \mathbf{Z}')\mathbf{y}$$

$$= \{X'(\Sigma^{-1} \otimes (Z(Z'Z)^{-1}Z'))X\}^{-1}\ \mathbf{X}'(\Sigma^{-1} \otimes (\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1})\mathbf{Z}'))\mathbf{y}\ .$$

This estimator can be obtained in three OLS stages, hence its name:

(1-2) Do 2SLS on each equation of the system, and retrieve the residuals calculated at the 2SLS estimators and the original (not the fitted) RHS variables.

(3) Estimate Σ from the residuals just calculated, and then do FGLS regression of y on X using the GLS weighting matrix $\Sigma^{-1}\otimes(Z(Z'Z)^{-1})Z')$.

The large-sample approximation to the covariance matrix for $\hat{c}_{3SLS}$ is, from the usual GLS theory,

$$\textbf{(15)} \quad V(\hat{c}_{3SLS}) = \left\{X'(\Sigma^{-1}\otimes(Z(Z'Z)^{-1}Z'))X\right\}^{-1} .$$

The FGLS third stage for the 3SLS estimator can be done conveniently by a OLS on transformed data. Let L be a lower triangular Cholesky factor of Σ_e^{-1} and Q be a lower triangular Cholesky factor of $(Z(Z'Z)^{-1})Z'$. Then $(L\otimes Q)(L\otimes Q)' = \Sigma_e^{-1}\otimes(Z(Z'Z)^{-1})Z')$. Transform $(L\otimes Q)y = (L\otimes Q)Xc + \eta$ and apply OLS to this system to get the 3SLS estimators.

The main advantage of 3SLS over 2SLS is a gain in asymptotic efficiency. The main disadvantage is that the estimators for a single equation are potentially less robust, since they will be inconsistent if the IV assumptions that Z is predetermined fail in any equation, not just a particular one of interest.

6. TESTING FOR OVER-IDENTIFYING RESTRICTIONS

Consider an equation $y = X\beta + u$ from a system of simultaneous equations, and let W denote the array of instruments (exogenous and predetermined variables) in the system. Let $X^* = P_W X$ denote the fitted values of X obtained from OLS estimation of the reduced form; where $P_W = W(W'W)^- W'$ is the projection operator onto the space spanned by W. The equation is *over-identified* if the number of instruments W exceeds the number of right-hand-side variables X. From Chapter 3, the GMM test statistic for over-identification is the minimum in β of

$$2nQ_n(\beta) = u'P_W u/\sigma^2 = u'P_{X^*} u/\sigma^2 + u'(P_W - P_{X^*})u/\sigma^2,$$

where $u = y - X\beta$. One has $u'(P_W - P_{X^*})u = y'(P_W - P_{X^*})y$, and at the minimum in β, $u'P_{X^*}u = 0$, so that

$$2nQ_n = y'(P_W - P_{X^*})y/\sigma^2.$$

Under H_o, this statistic is asymptotically chi-squared distributed with degrees of freedom equal to the difference in ranks of W and X^*. This statistic is the difference in the sum of squared residuals from the 2SLS regression of y on X and the sum of squared residuals from the reduced form regression of y on W, normalized by σ^2.

A computationally convenient equivalent form is

$$2nQ_n = \|\hat{y}_W - \hat{y}_{X*}\|^2/\sigma^2,$$

the sum of squares of the difference between the reduced form fitted values and the 2SLS fitted values of y, normalized by σ^2. Finally, $2nQ_n = y'Q_{X*}P_WQ_{X*}y/\sigma^2 = nR^2/\sigma^2$, where R^2 is the multiple correlation coefficient from regressing the 2SLS residuals on all the instruments; this result follows from the equivalent formulas for the projection onto the subspace of W orthogonal to the subspace spanned by X^*. This test statistic does *not* have a version that can be written as a quadratic form with the wings containing a difference of coefficient estimates from the 2SLS and reduced form regressions. Note that if the equation is *just identified*, with the number of proper instruments excluded from the equation exactly equal to the number of right-hand-side included endogenous variables, then there are no over-identifying restrictions and the test has no power. However, when the number of proper instruments exceeds the minimum for just identification, this test amounts to a test that all the exclusions of the instruments from the structural equation are valid.

7. TIME-SERIES APPLICATIONS OF SIMULTANEOUS EQUATIONS MODELS

The example of the market for economists that introduced this chapter was a time- series model that involved lagged dependent variables. In the example, we assumed away serial correlation, but in general serial correlation will be as issue to be dealt with in applications of simultaneous equations models to time series. The setup (6) for a linear simultaneous equations model can be expanded to make dependence on lagged dependent variables explicit:

(16) $$y_t'B + y_{t-1}'\Lambda + z_t'\Gamma = \varepsilon_t'.$$

Recall that the variables y_{t-1} and z_t in this model are *predetermined* if they are uncorrelated with the disturbance ε_t, and *strongly predetermined* if ε_t is statistically independent of y_{t-1} and z_t. In this model, the strictly exogenous variables z_t may include lags (and, if it makes economic sense, leads). It is not restrictive to write the model as a first-order lag in y_t, as higher-order lags can be incorporated by including lagged values of the dependent variables as additional components of y_t, with identities added to the system of equations to link the variables at different lags. (This was done in Chapter 5 in discussing the stability of vector autoregressions.)

The reduced form for the system (16), also called the *final form* in time series applications, is

$$\text{(17)} \qquad y_t' = y_{t-1}'\Theta + z_t'\Pi + \nu_t',$$

where $\Theta = -\Lambda B^{-1}$, $\Pi = -\Gamma B^{-1}$, and $\nu_t' = \varepsilon_t' B^{-1}$, so that the covariance matrix of ν_t is $\Omega = B'^{-1}\Sigma B^{-1}$. Identification of the model requires that B be nonsingular, and that there be exclusion and/or covariance restrictions that satisfy a rank condition. Stability of the model requires that the characteristic roots of Θ all be less than one in modulus. If one started with a stable structural model that had disturbances that were serially correlated with an autoregressive structure, then with suitable partial differencing the model could be rewritten in the form (17), the disturbances ν_t would be innovations that are independent across t, and the explanatory variables in (17) would be strongly predetermined. Further, the dynamics of the system would be dominated by the largest modulus characteristic root of Θ. In this stable case, estimation of the model can proceed in the manner already discussed: Estimate the reduced form, use fitted values of y_t (along with z_t and y_{t-1}) as instruments to obtain 2SLS estimates of each equation in (17), and finally use fitted covariances from these equations (calculated at the 2SLS estimates) to carry out 3SLS.

If the final form (17) is not stable, and in particular Λ has one or more unit roots, then the statistical properties of 2SLS or 3SLS estimates are quite different: some estimates may converge in asymptotic distribution at rate T rather than the customary $T^{1/2}$, and the asymptotic distribution may not be normal. Consequently, one must be careful in conducting statistical inference using these estimates. There is an extensive literature on analysis of systems containing unit roots; see the chapter by Jim Stock in the *Handbook of Econometrics IV*. When a system is known to contain a unit root, then it may be possible to transform to a stable system by appropriate differencing.

8. NONLINEAR SIMULTANEOUS EQUATIONS MODELS

In principle, dependent variables may be simultaneously determined within a system of equations that is nonlinear in variables and parameters. One might, for example, consider a system

(18) $$F_i(y_{1t},y_{2t},...,y_{Nt};z_{it},\theta) = \varepsilon_{it},\ i = 1,...,N$$

for the determination of $(y_{1t},y_{2t},...,y_{Nt})$ that depends on a K×1 vector of parameters θ, vectors of exogenous variables z_{it}, and disturbances ε_{it}. Such systems might arise naturally out of economic theory. For example, consumer or firm optimization may be characterized by first-order conditions that are functions of dependent decision variables and exogenous variables describing the economic environment of choice, with the ε_{it} appearing due to errors in optimization by the economic agents, arising perhaps because ex post realizations differ from ex ante expectations, or due to approximation errors by the analyst. For many plausible economic models, linearity of the system (18) in variables and parameters would be the exception rather than the rule, with the common linear specification justifiable only as an approximation.

The nonlinear system (18) is well-determined if it has a unique solution for the dependent variables, for every possible configuration of the z's and ε's, and for all θ's in a specified domain. If it is well-determined, then it has a reduced form

$$(19)\qquad y_{it} = f_i(z_{1t},z_{2t},\ldots,z_{Nt},\varepsilon_{1t},\varepsilon_{2t},\ldots,\varepsilon_{Nt},\theta),\ i = 1,\ldots,N.$$

This reduced form can also be written

$$(20)\qquad y_{it} = h_i(z_{1t},z_{2t},\ldots,z_{Nt},\theta) + u_{it},\ i = 1,\ldots,N$$

where

$$h_i(z_{1t},z_{2t},\ldots,z_{Nt},\theta) = E\{f_i(z_{1t},z_{2t},\ldots,z_{Nt},\varepsilon_{1t},\varepsilon_{2t},\ldots,\varepsilon_{Nt},\theta)\,|\,z_t\},$$

and u_{it} is the disturbance with conditional mean zero that makes (20) hold. (20) is a system of nonlinear equations, and the treatment in Chap. 5 can also be applied to estimate the structural parameters from this reduced form. (The specification (20) guarantees that the reduced form disturbances have conditional expectation zero; but the additional assumption that u's are statistically independent of z's, or even that they are homoskedastic, is rarely justifiable from economic theory. Then statistical analysis based on this assumption may be invalid and misleading for many application.)

Recall that in Chapter 4, estimation of a nonlinear equation with contaminated explanatory variables was discussed, a best nonlinear 2SLS (BN2SLS) estimator was defined, and practical approximations to the BN2SLS were discussed. The equations in (18) would correspond directly to this structure if in equation i, one had

(21) $$F_i(y_{1t},y_{2t},\ldots,y_{Nt};z_{it},\theta) = y_{it} - h(y_{1t},\ldots,y_{i-1,t},y_{i+1,t},\ldots,y_{Nt},z_{it},\theta),$$

Absent this normalization, some other normalization is needed for identification in F_i, either on the scale of the dependence of F_i on one variable, or in the scale of ε_{it}. This is no different in spirit than the normalizations needed in a linear simultaneous equations specification. Given an identifying normalization, it is possible to proceed in essentially the same way as in Chapter 4. Make a first-order Taylor's expansion of (18) about an initial parameter vector θ_o to obtain

(22) $$F_i(y_{1t},y_{2t},\ldots,y_{Nt};z_{it},\theta_o) \approx -\sum_{k=1}^{K} \frac{\partial F_i(y_{1t},y_{2t},\ldots,y_{Nt};z_{it},\theta_o)}{\partial\theta_k}\cdot(\theta_k-\theta_{ok}) + \varepsilon_{it}.$$

Treat the expressions $x_{itk} = -\partial F_i(y_{1t},y_{2t},\ldots,y_{Nt};z_{it},\theta_o)/\partial\theta_k$ as contaminated explanatory variables, and the expectations of x_{ikt} given $z_{1t},\ldots,z_{Nt}$ as the ideal best instruments.

Approximate these best instruments by regressing the x_{itk} on suitable functions of the z's, as in Chapter 4, and then estimate (22) by this approximation to best 2SLS. Starting from an initial guess for the parameters, iterate this process to convergence, using the estimated coefficients from (22) to update the parameter estimates. The left-hand-side of (22) is the dependent variable in these 2SLS regressions, with the imposed normalization guaranteeing that the system is identified. This procedure can be carried out for the entire system (22) at one time, rather than equation by equation. This will provide nonlinear 2SLS estimates of all the parameters of the system. These will not in general be best system estimates because they do not take into account the covariances of the ε's across equations. Then, a final step is to apply 3SLS to (22), using the previous 2SLS estimates to obtain the feasible GLS transformation. The procedure just described is what the LSQ command in TSP does when applied to a system of nonlinear equations without normalization, with instrumental variables specified.

When the nonlinear reduced form (20) can be obtained as an analytic or computable model, it is possible to apply nonlinear least squares methods directly, either equation by equation as N2SLS or for the system as N3SLS. This estimation procedure is described in Chapter 5. One caution is that while the disturbances u_{it} in (20) have conditional mean zero by construction, economic theory will rarely imply that they are, in addition, homoskedastic, and the large sample statistical theory needs to be reworked when heteroskedasticity of unknown form is present. Just as in linear models, consistency is generally not at issue, but standard errors will typically not be estimated consistently. At minimum, one should be cautious and use robust standard error estimates that are consistent under heteroskedasticity of unknown form.