VI. Simultaneous Equation Models

1. An introduction to some problems associated with simultaneous equation models

2. Basic Notation and alternative representations
- **a. Structural representation**
- **b. Reduced form representation**
- **c. Final form or transfer function representation**

3. Identification
- **a. Identification as a mapping (3 examples)**
- **b. Examples**

4. Estimation of reduced form parameters
- **a. Least squares no restrictions (LSNR)**
- **b. Estimators derived from structural estimators**
- **c. SURE–generally the same as LSNR**

5. Estimation of Structural Models
- **a. Notation**
- **b. Single equation methods: OLS, IV, 2SLS, LIML, k-class, GMM, IV, other**
- **c. Simultaneous equation methods: FIML and 3SLS**

6. Statistical inference
- **a. Identification**
- **b. Reduced form coefficients**
- **c. Structural parameters**

7. Exercises

James B. McDonald
Brigham Young University
1/2013

VI. SIMULTANEOUS EQUATION MODELS

1. INTRODUCTION TO SIMULTANEOUS EQUATION MODELS

There are several problems encountered with structural equations in simultaneous equations models which are not generally associated with single equation models. These include (1) the identification problem, (2) inconsistency of ordinary least squares (OLS) estimators of structural parameters, (3) questions about the interpretation of structural parameters, and (4) the validity of the OLS "t statistics" associated with structural coefficients.

In order to introduce these problems, we review two important papers. The paper on identification by E. J. Working [1927, QJE] is considered in the first section. The work of Haavelmo [1947, JASA] dealing with alternative methods of estimating the marginal propensity to consume is described in the second section. The third section contains a brief summary.

a. STRUCTURAL AND REDUCED FORM REPRESENTATIONS, IDENTIFICATION, AND INTERPRETATIONS OF COEFFICIENTS

Consider the problem of estimating the impact of an increase in the price of crude oil upon the equilibrium price and quantity of gasoline. The corresponding increase in the equilibrium price of gasoline will depend upon several factors including the slope of the demand curve.

This is illustrated in the following figure:

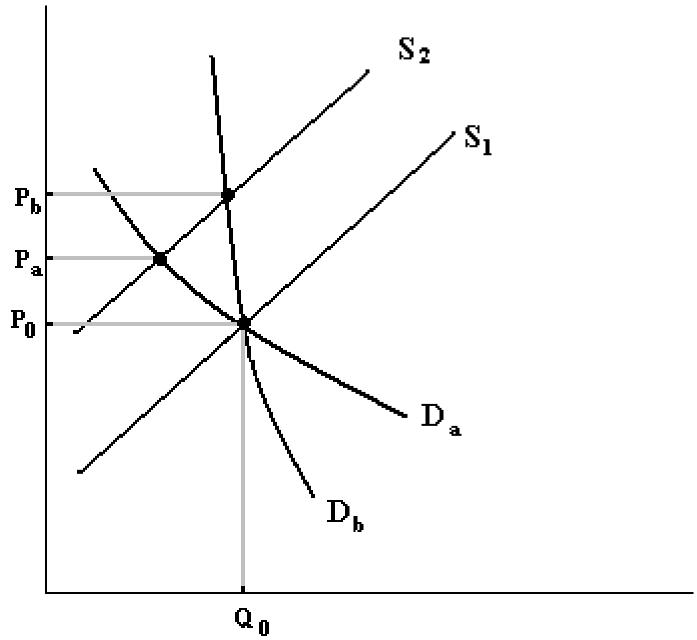


Figure 1

Assume that (Q_0, P_0) denotes the original equilibrium. Assume that the increase in the price of crude oil results in the supply curve shifting from S_1 to S_2. The associated change in P depends upon the relevant demand schedule with the more inelastic curve being associated with the larger price increases. This example clearly indicates the importance of estimating the slope of the demand schedule in order to make predictions about the impact of changes in factor price upon equilibrium price.

Estimation of the slope of the demand curve might begin by collecting observations on (P, Q) which might appear as in Figure 2.

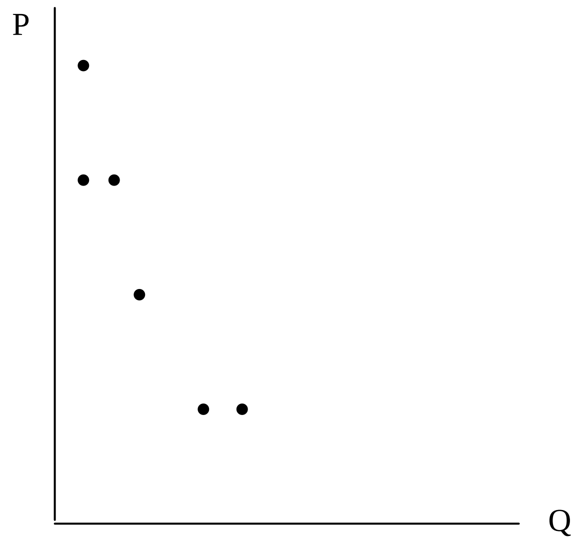


Figure 2

The reader would probably be tempted to draw a line through the points or perform a least squares estimation on $p = \beta_1 - \beta_2 Q$ in order to estimate the demand schedule. But how would we estimate the demand curve if a plot of P and Q appeared as in Figure 3 rather than as in Figure 2?

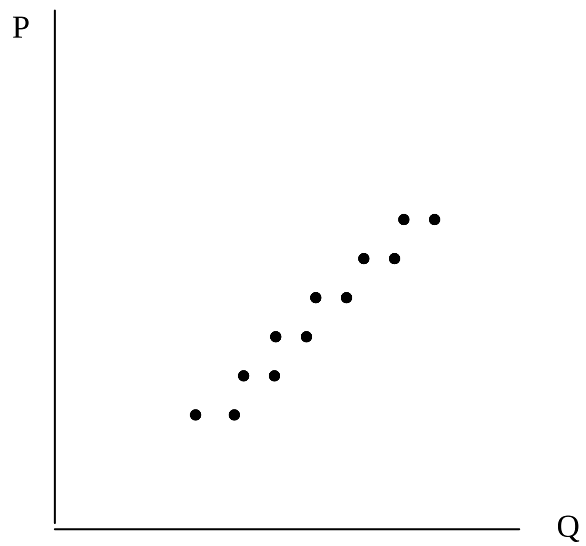


Figure 3

The data in Figure 3 appears to define a supply curve rather than a demand curve. Alternatively, how could a demand curve be estimated if the data appear as in Figure 4?

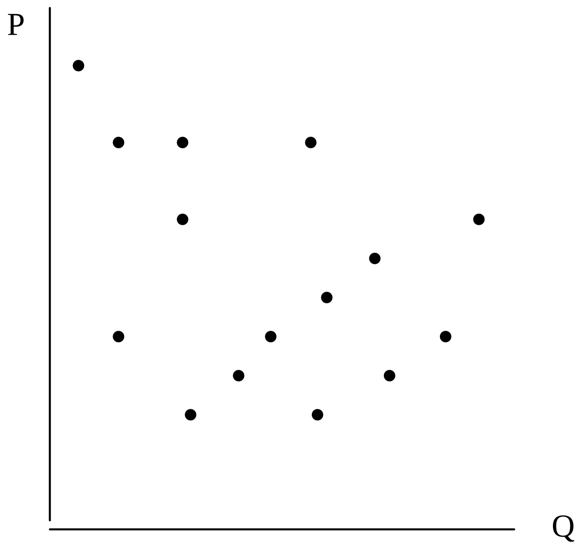


Figure 4

In order to answer this question, we need to recall that equilibrium price and quantity are determined by supply <u>and</u> demand factors and not by supply <u>or</u> demand alone. The observations depicted in Figure 2 could have been generated by either of the following scenarios:

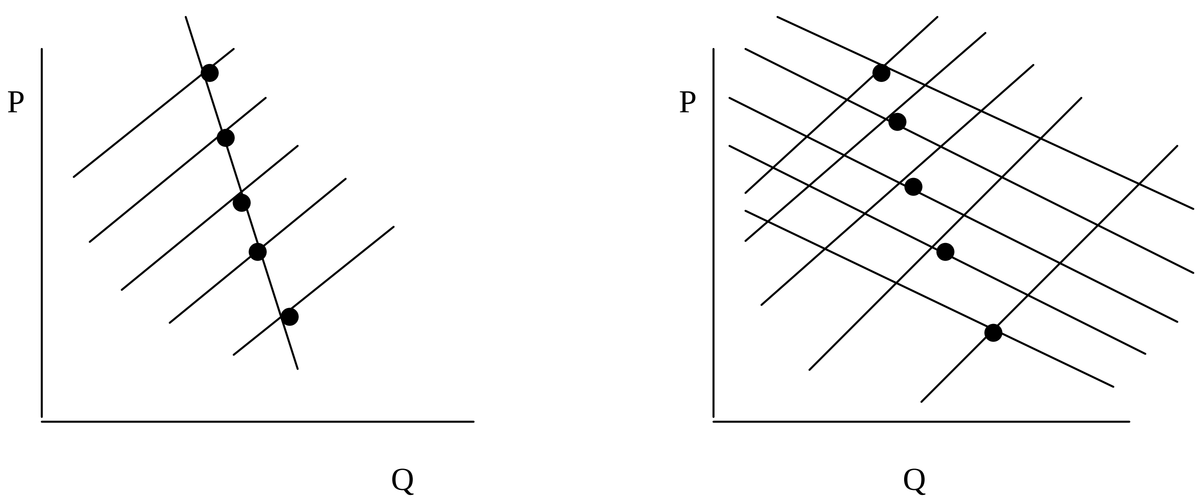


Figure 5

If the demand curve is stable and the supply curve shifts, then the demand curve is "traced out." If both curves shift, fitting a relationship to the observed (Q, P) would not correspond to the underlying demand curve(s). Similarly, Figure 3 could correspond to a relatively stable supply curve and a shifting demand curve or both curves shifting. Figure 4 would appear to correspond to both curves shifting.

Consider the following model:

(1.1) Demand: $Q = \gamma_{11} - \beta_{12}P + \gamma_{12}Y + \varepsilon_{1t}$

(1.2) Supply: $Q = \gamma_{21} + \beta_{22}P - \gamma_{23}FC + \varepsilon_{2t}$

or equivalently,

$$\begin{bmatrix} -1 & -\beta_{12} \\ -1 & \beta_{22} \end{bmatrix} \begin{bmatrix} Q_t \\ P_t \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} & 0 \\ \gamma_{21} & 0 & -\gamma_{23} \end{bmatrix} \begin{bmatrix} 1 \\ Y_t \\ FC_t \end{bmatrix} + \begin{bmatrix} \varepsilon_{t1} \\ \varepsilon_{t2} \end{bmatrix} = 0 .$$

Equations (1.1) and (1.2) will be referred to as the <u>structural model</u> with Q and P as endogenous (dependent) variables and income (Y) and factor costs (crude oil, FC) as exogenous (independent) variables. In order to draw a demand curve or supply curve using (Q, P) as coordinates, Y and FC must be fixed at some arbitrary level.

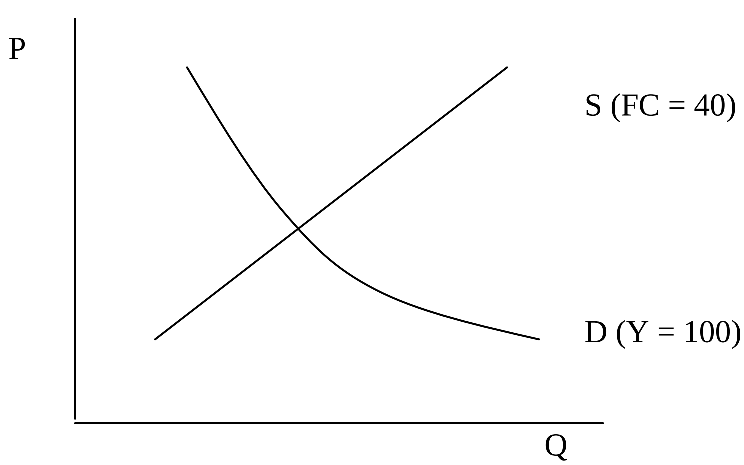


Figure 6

A change in factor costs (income fixed) will shift the supply curve and “trace” the depicted demand curve and a change in income (factor costs fixed) will shift the demand curve and “trace” the depicted supply curve, et cet. paribus. It is interesting to observe that by including factor costs (FC) in the supply equation and not the demand equation we are able to "identify" the demand equation. Similarly, by including income (Y) in the demand equation and not in the supply equation we are able to "identify" the supply equation. Hence, one way of "identifying" a structural equation is by excluding variables from the equation we want to estimate, which are included in other structural equations. These excluded variables are referred to as instrumental variables. This is the general approach to the identification problem developed by E. J. Working [1927]. A more formal development will be considered later.

We note from Figure 6 that for each level of factor costs and income there is a corresponding equilibrium price and quantity determined by the intersection of the supply and demand curves. If we solve the structural model for the explicit relationship between (P, Q) and FC and Y we obtain

$$\begin{bmatrix} Q_t \\ P_t \end{bmatrix} = -\begin{bmatrix} -1 & -\beta_{12} \\ -1 & \beta_{22} \end{bmatrix}^{-1} \left\{ \begin{bmatrix} \gamma_{11} & \gamma_{12} & 0 \\ \gamma_{21} & 0 & -\gamma_{23} \end{bmatrix} \begin{bmatrix} 1 \\ Y_t \\ FC_t \end{bmatrix} + \begin{bmatrix} \varepsilon_{t1} \\ \varepsilon_{t2} \end{bmatrix} \right\}$$

$$= \left(\frac{1}{\beta_{12}+\beta_{22}} \right) \begin{bmatrix} \beta_{22} & \beta_{12} \\ 1 & -1 \end{bmatrix} \left\{ \begin{bmatrix} \gamma_{11} & \gamma_{12} & 0 \\ \gamma_{21} & 0 & -\gamma_{23} \end{bmatrix} \begin{bmatrix} 1 \\ Y_t \\ FC_t \end{bmatrix} + \begin{bmatrix} \varepsilon_{t1} \\ \varepsilon_{t2} \end{bmatrix} \right\}$$

$$= \left(\frac{1}{\beta_{12}+\beta_{22}} \right) \begin{bmatrix} \beta_{22}\gamma_{11}+\beta_{12}\gamma_{21} & \beta_{22}\gamma_{12} & -\beta_{12}\gamma_{23} \\ \gamma_{11}-\gamma_{21} & \gamma_{12} & \gamma_{23} \end{bmatrix} \begin{bmatrix} 1 \\ Y_t \\ FC_t \end{bmatrix}$$

$$(1.3a\text{-}b) \qquad + \begin{bmatrix} \dfrac{\beta_{22}\varepsilon_{t1}+\beta_{12}\varepsilon_{t2}}{\beta_{12}+\beta_{22}} \\ \\ \dfrac{\varepsilon_{t1}-\varepsilon_{t2}}{\beta_{12}+\beta_{22}} \end{bmatrix}$$

$$= \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \end{bmatrix} \begin{bmatrix} 1 \\ Y_t \\ FC_t \end{bmatrix} + \begin{bmatrix} \eta_{t1} \\ \eta_{t2} \end{bmatrix}$$

Note: $$\frac{\partial Q}{\partial Y} = \frac{\beta_{22}\gamma_{12}}{\beta_{12}+\beta_{22}} = \pi_{12} > 0, \quad \frac{\partial Q}{\partial FC} = \frac{-\beta_{12}\gamma_{23}}{\beta_{12}+\beta_{22}} = \pi_{13} < 0$$

$$\frac{\partial P}{\partial Y} = \frac{\gamma_{12}}{\beta_{12}+\beta_{22}} = \pi_{22} > 0, \quad \frac{\partial P}{\partial FC} = \frac{\gamma_{23}}{\beta_{12}+\beta_{22}} = \pi_{23} > 0$$

Equations (1.3a, b) are referred to as the reduced form equations for Q and P corresponding to the structural model defined by (1.1) and (1.2). Note that each reduced form equation expresses the equilibrium value (P or Q) as a function of the exogenous variables FC and Y.

In order to determine the impact of an increase in the price of crude oil upon the price of gasoline, we employ the reduced form representation, i.e.,

$$\frac{\partial P}{\partial FC} = \frac{\gamma_{23}}{\beta_{12}+\beta_{22}} = \pi_{23} > 0$$

which takes into account the slopes of the supply and demand curves as well as how far the supply curve shifts in response to an increase in the price of crude oil. The equilibrium quantity would also change according to

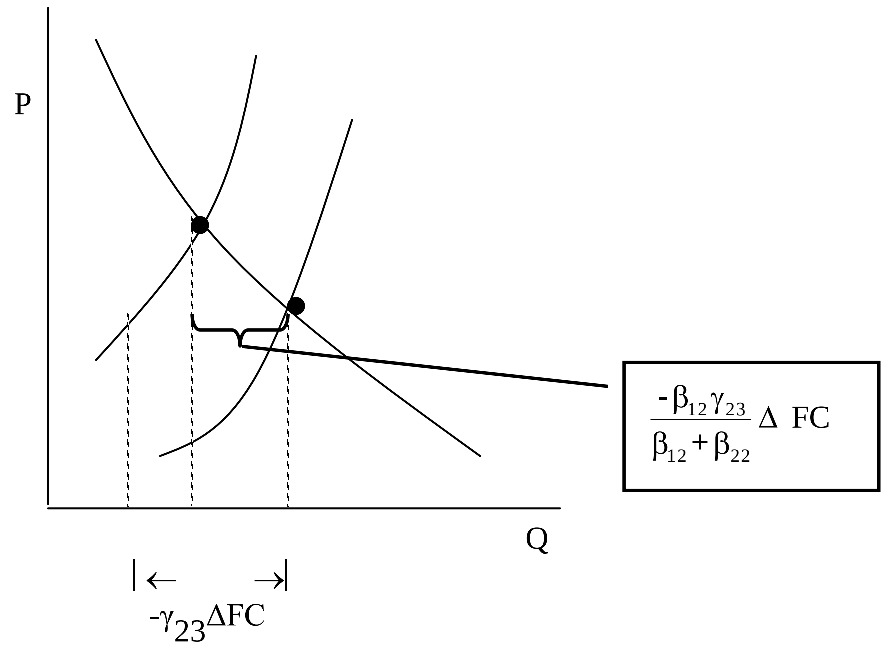
$$\frac{\partial Q}{\partial FC} = \frac{-\beta_{12}\gamma_{23}}{\beta_{12}+\beta_{22}} = \pi_{13} < 0.$$

The reader might wonder why

$$\frac{\partial Q^s}{\partial FC} = -\gamma_{23} < 0$$

doesn't characterize the change in equilibrium quantity.

The following figure will illustrate why the reduced form provides the necessary information.



Differentiating the supply equation with respect to FC yields $(-\gamma_{23})$ and assumes that P is fixed and hence merely represents the horizontal shift of the supply curve and not the change in equilibrium quantity. The reduced form equation for Q expresses the equilibrium quantity as a function of FC and Y and takes account of the increase in equilibrium price associated with an increase in factor costs.

To summarize, the reduced form coefficients represent the change in equilibrium values corresponding to changes in the predetermined or exogenous variables, i.e., the reduced form coefficients are the multipliers. The structural coefficients represent slopes or shifts of structural schedules in response to changes in predetermined or exogenous variables.

EXERCISE:

1. The Asymptotic Bias of the OLS estimator of the slope for the demand curve is given by

$$\frac{(\beta_{22}+\beta_{12})\,\sigma^2_{\varepsilon 1}}{\sigma^2_{\varepsilon 1}+\sigma^2_{\varepsilon 2}+\gamma^2_{23}(1-\mathrm{COR}^2(Y, FC))}$$

   where COR(Y, FC) = correlation between Y and FC.

   (a) Mathematically analyze the impact of increases in ${\sigma_{\varepsilon 2}}^2$, ${\gamma_{23}}^2$, and COR(Y, FC) upon the asymptotic bias of $\hat{\beta}_{12}$.

   (b) Graphically analyze the impact of increases in ${\sigma_{\varepsilon 2}}^2$, ${\gamma_{23}}^2$, and COR(Y, FC) upon the "bias of β_{12}."

b. INCONSISTENCY OF ORDINARY LEAST SQUARES, ALTERNATIVE ESTIMATORS, AND STATISTICAL INFERENCE

Haavelmo [1947] considered the following simple macro model:

(2.1) $C_t = \alpha + \beta Y_t + \varepsilon_t$

(2.2) $Y_t = C_t + Z_t$

where Y_t, C_t, and Z_t ($Z \equiv Y - C$) respectively denote income, consumption and nonconsumption expenditure.

The <u>reduced form representation</u> corresponding to (2.1) and (2.2) is given by

(2.3) $C_t = \pi_{11} + \pi_{12} Z_t + \eta_t$

(2.4) $Y_t = \pi_{21} + \pi_{22} Z_t + \eta_t$

where (2.5a-e) $\eta_t = \varepsilon_t/(1-\beta)$

$\pi_{11} = \alpha/(1-\beta)$

$\pi_{12} = \beta/(1-\beta)$

$\pi_{21} = \alpha/(1-\beta)$

$\pi_{22} = 1/(1-\beta)$

Note that π_{12} and π_{22} correspond to the multipliers discussed in simple macroeconomics models. Haavelmo's analysis of the simple model defined by (2.1) and (2.2) pointed out many problems which are also associated with larger econometric models. For this reason we will consider this model in detail.

Estimation. Past experience might suggest that the OLS estimator of β would have desirable statistical properties if ε_t in (2.1) is not characterized by autocorrelation or heteroskedasticity. The OLS estimator of β in (2.1) is defined by

$$(2.6)\ \hat{\beta} = \frac{\sum(\mathrm{Y}-\overline{\mathrm{Y}})(\mathrm{C}-\overline{\mathrm{C}})}{\sum(\mathrm{Y}-\overline{\mathrm{Y}})^2} = \frac{Cov(Y,C)}{Var(Y)}$$

but from (2.3) and (2.4), we see that

$$(2.7)\ \mathrm{C}-\overline{\mathrm{C}} = \pi_{12}(\mathrm{Z}-\overline{\mathrm{Z}}) + \frac{\varepsilon-\overline{\varepsilon}}{1-\beta}$$

$$= \frac{\beta}{1-\beta}(\mathrm{Z}-\overline{\mathrm{Z}}) + \frac{\varepsilon-\overline{\varepsilon}}{1-\beta}$$

and

$$(2.8)\ \mathrm{Y}-\overline{\mathrm{Y}} = \pi_{22}(\mathrm{Z}-\overline{\mathrm{Z}}) + \frac{\varepsilon-\overline{\varepsilon}}{1-\beta}$$

$$= \frac{1}{1-\beta}(\mathrm{Z}-\overline{\mathrm{Z}}) + \frac{\varepsilon-\overline{\varepsilon}}{1-\beta};$$

hence, after substituting (2.7) and (2.8) into (2.6), we can write

$$(2.9)\ \hat{\beta} = \frac{\sum\left\{\frac{(\mathrm{Z}-\overline{\mathrm{Z}})}{(1-\beta)} + \frac{(\varepsilon-\overline{\varepsilon})}{1-\beta}\right\}\left\{\frac{\beta}{1-\beta}(\mathrm{Z}-\overline{\mathrm{Z}}) + \frac{(\varepsilon-\overline{\varepsilon})}{1-\beta}\right\}}{\sum\left\{\frac{(\mathrm{Z}-\overline{\mathrm{Z}})}{(1-\beta)} + \frac{(\varepsilon-\overline{\varepsilon})}{1-\beta}\right\}^2}$$

$$\hat{\beta} = \frac{\sum\left\{\frac{\beta}{(1-\beta)^2}(\mathrm{Z}-\overline{\mathrm{Z}})^2 + \frac{(1+\beta)(\mathrm{Z}-\overline{\mathrm{Z}})(\varepsilon-\overline{\varepsilon})}{(1-\beta)^2} + \frac{(\varepsilon-\overline{\varepsilon})^2}{(1-\beta)^2}\right\}}{\sum\left\{\frac{(\mathrm{Z}-\overline{\mathrm{Z}})^2}{(1-\beta)^2} + 2\frac{(\varepsilon-\overline{\varepsilon})(\mathrm{Z}-\overline{\mathrm{Z}})}{(1-\beta)^2} + \frac{(\varepsilon-\overline{\varepsilon})^2}{(1-\beta)^2}\right\}}$$

$$= \frac{\beta\sum(\mathrm{Z}-\overline{\mathrm{Z}})^2/\mathrm{N} + (1+\beta)\sum(\mathrm{Z}-\overline{\mathrm{Z}})(\varepsilon-\overline{\varepsilon})/\mathrm{N} + \sum(\varepsilon-\overline{\varepsilon})^2/\mathrm{N}}{\sum\left\{(\mathrm{Z}-\overline{\mathrm{Z}})^2/\mathrm{N} + \sum(\varepsilon-\overline{\varepsilon})(\mathrm{Z}-\overline{\mathrm{Z}})/\mathrm{N} + \sum(\varepsilon-\overline{\varepsilon})^2/\mathrm{N}\right\}}.$$

Assuming that:

$$\sum_{t=1}^{N}(\mathrm{Z}-\overline{\mathrm{Z}})^2/\mathrm{N} \to \sigma_Z^2 \qquad \text{as } \mathrm{N}\to\infty,$$

$$\sum_{t=1}^{N}(\mathrm{Z}-\overline{\mathrm{Z}})(\varepsilon-\overline{\varepsilon})/\mathrm{N} \to 0 \qquad \text{as } \mathrm{N}\to\infty, \text{ and}$$

$$\sum_{t=1}^{N}(\varepsilon - \bar{\varepsilon})^2/N \to \sigma^2 \qquad \text{as } N \to \infty$$

$$(2.10) \qquad \hat{\beta} \to \frac{\beta\sigma_Z^2 + \sigma^2}{\sigma_Z^2 + \sigma^2}.$$

$$= \beta + \frac{\sigma^2(1-\beta)}{\sigma_Z^2 + \sigma^2}$$

as $N \to \infty$. Hence, we see from (2.10) that $\hat{\beta}$ is an inconsistent estimator of β with asymptotic bias equal to the second term in (2.10)

$$\frac{\sigma^2(1-\beta)}{\sigma_Z^2 + \sigma^2}.$$

This may seem like a surprising result in light of the apparent simplicity of the consumption function. It may not be obvious which of the assumptions

(A.1) ε_t distributed normally

(A.2) $E(\varepsilon_t) = 0$ for all t

(A.3) $Var(\varepsilon_t) = \sigma^2$ for all t

(A.4) $E(\varepsilon_t \varepsilon_s) = 0$ for $t \neq s$

(A.5) Y_t and ε_t are independent.

are violated. But upon closer inspection (hint: see (2.4)) we note that

$$E(Y_t \varepsilon_t) = E\left[\left(\pi_{21} + \pi_{22} Z_t + \frac{\varepsilon_t}{1-\beta}\right)(\varepsilon_t)\right]$$

$$= E(\varepsilon_t^2)/(1-\beta)$$

$$= \sigma^2/(1-\beta) \neq 0;$$

hence, (A.5) is violated and OLS estimators of the structural parameters α and β are biased and inconsistent. In fact, this is typically the case when OLS is used to estimate structural relationships which include endogenous variables on the right hand side of the structural equation. Right hand side endogenous variables are commonly referred to as ***endogenous regressors***.

As another example, the asymptotic bias of the OLS estimator of β_{12} in (1.1) is given by

(2.11) $$\frac{(\beta_{22} + \beta_{12})\sigma_{\varepsilon 1}^2}{\sigma_{\varepsilon 1}^2 + \sigma_{\varepsilon 2}^2 + \gamma_{23}^2(1-\mathrm{Corr}^2(Y, FC))}.$$

How can we obtain consistent estimators of the unknown structural parameters?

Two stage least squares or an appropriate application of instrumental variables estimation provides a solution. It is instructive to consider an alternative estimator first. Recall that the ordinary least squares estimators of the reduced form equations (referred to as least squares no restrictions, LSNR) will yield <u>unbiased</u> and <u>consistent</u> estimators of the π_{ij}'s which will be denoted by $\hat{\pi}_{ij}$. This observation provides the basis for obtaining consistent estimators of α and β in the Haavelmo model. From (2.5 c,e) we note that

$$\beta = \pi_{12}/\pi_{22}$$

hence, a consistent estimator of β can be obtained from

(2.12) $$\beta^* = \hat{\pi}_{12}/\hat{\pi}_{22}$$

where $$\hat{\pi}_{12} = \frac{\sum(C-\bar{C})(Z-\bar{Z})}{\sum(Z-\bar{Z})^2}$$

$$\hat{\pi}_{22} = \frac{\sum(Y-\bar{Y})(Z-\bar{Z})}{\sum(Z-\bar{Z})^2}$$

or

(2.13) $$\beta^* = \frac{\sum(C-\bar{C})(Z-\bar{Z})}{\sum(Y-\bar{Y})(Z-\bar{Z})}$$

In order to verify the consistency of β* in (2.13) we replace $(C-\bar{C})$ and $(Y-\bar{Y})$ in (2.12) by (2.7) and (2.8) to obtain

(2.14) $$\beta^* = \frac{\sum\left\{\left[\frac{\beta}{(1-\beta)}(Z-\bar{Z}) + \frac{\varepsilon-\bar{\varepsilon}}{1-\beta}\right]\left[Z-\bar{Z}\right]\right\}}{\sum\left\{\left[\frac{1}{1-\beta}(Z-\bar{Z}) + \frac{(\varepsilon-\bar{\varepsilon})}{1-\beta}\right]\left[Z-\bar{Z}\right]\right\}}$$

$$= \frac{\beta\sum(Z-\bar{Z})^2/N + \sum(\varepsilon-\bar{\varepsilon})(Z-\bar{Z})/N}{\left\{\sum(Z-\bar{Z})^2/N + \sum(\varepsilon-\bar{\varepsilon})(Z-\bar{Z})/N\right\}}$$

Now as $N \to \infty$

$$\beta^* \to \beta;$$

hence, β* is a consistent estimator and is obtained by obtaining consistent estimators of the reduced form (LSNR) and then deducing corresponding estimates of structural coefficients. This general method is referred to as <u>indirect least squares</u> (ILS), but is not applicable for all structural models.

The consistent estimator β* can also be obtained by replacing the dependent variable on the right hand side of (2.1) by its predicted value (from the reduced form)

$$\hat{Y} = \hat{\pi}_{21} + \hat{\pi}_{22}Z$$

or $$\hat{Y} - \bar{Y} = \hat{\pi}_{22}(Z-\bar{Z})$$

and then applying least squares to the resultant expression. More explicitly,

$$\text{(2.15 a-e)} \qquad \beta^* = \frac{\sum(\hat{Y}-\overline{Y})(C-\overline{C})}{\sum(\hat{Y}-\overline{Y})^2}$$

$$= \frac{\hat{\pi}_{22}}{\hat{\pi}_{22}^2}\frac{\sum(Z-\overline{Z})(C-\overline{C})}{\sum(Z-\overline{Z})^2}$$

$$= \frac{1}{\hat{\pi}_{22}}\frac{\sum(Z-\overline{Z})(C-\overline{C})}{\sum(Z-\overline{Z})^2}$$

$$= \left\{\frac{\sum(Z-\overline{Z})^2}{\sum(Y-\overline{Y})(Z-\overline{Z})}\right\}\left\{\frac{\sum(Z-\overline{Z})(C-\overline{C})}{\sum(Z-\overline{Z})^2}\right\}$$

$$= \frac{\sum(Z-\overline{Z})(C-\overline{C})}{\sum(Y-\overline{Y})(Z-\overline{Z})}$$

which corresponds to (2.13). Compare (2.15 a) with (2.6) and note that the only difference is that $\hat{Y}$ (predicted value) replaces Y in (2.6). The structural estimator, obtained by applying least squares to the structural equation which has been modified by replacing the right hand dependent variables by their reduced form predictions is referred to as two stage least squares (2SLS). 2SLS yields consistent estimators, and is applicable even when indirect least squares is not. Another way of looking at the alternative estimator is obtained by comparing (2.6) and (2.15e). Here we see that the difference is that the right hand side dependent variable Y in (2.6) is replaced by Z (an instrumental variable) which is correlated with Y, but not with C; hence, these estimators are sometimes referred to as instrumental variables estimators.

A numerical example: the Haavelmo data set, (Haavelmo.dat).

Using the data provided by Haavelmo, the regular OLS estimates of the consumption function given by

$$\hat{C}_{OLS} = 84.01 + .732Y$$

$$\left(s_{\hat{\beta}_i}\right) \quad (14.55) \quad (.030)$$

$$R^2 = .971$$
$$s^2 = 58.21.$$

The corresponding 2SLS estimates of the consumption function are given by

$$\hat{C}_{2SLS} = \underset{(17.8)}{113.1} + \underset{(.037)}{.672Y}$$
$$s^2 = 71.29.$$

The LSNR estimates of the reduced form equations are given by

$$\hat{C} = \underset{(16.48)}{344.70} + \underset{(.341)}{2.048Z}$$

$$R^2 = .668$$

$$\hat{Y} = \underset{(16.48)}{344.70} + \underset{(.341)}{3.048Z}$$

$$R^2 = .668$$

The reader should verify that the indirect least squares estimators are equal to the 2SLS. However, except for pedagogical examples, the reader will apply 2SLS or instrumental variables estimation directly and not use the two step procedure. Also, the two step procedure yields incorrect standard errors.

CONFIDENCE INTERVALS. In determining confidence intervals for structural parameters, the reader might be inclined to use the results associated with the OLS or 2SLS estimates of the structural equation under consideration. As an example of this we compute "95% confidence intervals for β (the MPC)."

(a) Based upon OLS: (t = 2.101)

$$\hat{\beta}_{OLS} \pm ts_{\hat{\beta}}$$

$$= (.732 \pm 2.101(.0299))$$

$$= (.669, .795)$$

(b) Based upon 2SLS

$$\hat{\beta}_{2SLS} \pm ts_{\hat{\beta}}$$

$$= (.672 \pm 2.101(.0368))$$

$$= (.594, .748)$$

These confidence intervals are very different and one might ask which if either is appropriate. As it turns out, neither is completely satisfactory since

$$\frac{\hat{\beta} - \beta}{s_{\hat{\beta}}}$$

is not exactly distributed as a t-statistic where $\hat{\beta}$ is obtained from the technique of OLS or 2SLS.

One way in which we can determine which (if either) of the previous confidence intervals is closest is to note that

$$\frac{\hat{\pi}_{ij} - \pi_{ij}}{s_{\hat{\pi}_{ij}}} \sim t(n-2);$$

hence,

$$1-\alpha = \Pr[-t_{\alpha/2} \le \frac{\hat{\pi}_{22} - \pi_{22}}{s_{\hat{\pi}_{22}}} \le t_{\alpha/2}]$$

$$= \Pr[\hat{\pi}_{22} - t_{\alpha/2}\, s_{\hat{\pi}_{22}} \le \pi_{22} \le \hat{\pi}_{22} + t_{\alpha/2} s_{\hat{\pi}_{22}}]$$

$$= \Pr[\hat{\pi}_{22} - t_{\alpha/2}\, s_{\hat{\pi}_{22}} \le \frac{1}{1-\beta} \le \hat{\pi}_{22} + t_{\alpha/2} s_{\hat{\pi}_{22}}]$$

$$= \Pr[1 - \frac{1}{\hat{\pi}_{22} - t_{\alpha/2}\, s_{\hat{\pi}_{22}}} \le \beta \le 1 - \frac{1}{\hat{\pi}_{22} + s_{\hat{\pi}_{22}} t_{\alpha/2}}].$$

Making the appropriate substitutions we obtain

$$(.57, .73)$$

which is much closer to the results obtained using two least squares than from OLS. One might be inclined to conjecture that a reason for the poor performance of OLS confidence intervals is due to the asymptotic bias of OLS estimator,

$$\frac{\sigma^2(1-\beta)}{\sigma^2+\sigma^2}.$$

It might be instructive to estimate the asymptotic bias. Doing so we obtain for OLS estimates of $\sigma^2(s^2=58.2)$, $\beta(\hat{\beta}=.732)$, σ_z^2 (285.55); hence asymptotic bias ($\hat{\beta}_{OLS}$) = .0454; for 2SLS estimates of $\sigma^2(s^2=71.29)$, $\beta(\hat{\beta}=.672)$, σ_z^2 (285.55), asymptotic bias ($\hat{\beta}_{OLS}$) = .0655. Note that the difference between the OLS and 2SLS is (.732 - .672 = .06).

PREDICTIONS. In order to make predictions, one should use the reduced form representation.

c. A BRIEF OVERVIEW

The mathematical formulation of an economic model is generally referred to as the structural representation. The structural equations in the structural representation will often include endogenous regressors (endogenous variables on the right hand side) as well as exogenous variables.

The reduced form representation corresponding to the structural representation is characterized by separate equations which express each dependent variable as a function of the exogenous variables. The reduced form provides explicit expressions for the equilibrium values of the dependent variables in the model, conditional on an arbitrary, but given, set of values for the exogenous variables. The reduced form coefficients can be interpreted as "multipliers" and yield comparative static results. The reduced form representation is usually the form used for obtaining forecasts from econometric models.

After the econometrician is satisfied that a given econometric model is consistent with relevant economic theory, it is important that each structural equation be identified. Identification should be checked even before attempting to estimate the model. A necessary condition (order condition) for identification is that the number of exogenous (predetermined) variables excluded (K_2) from a structural equation is at least as large as the number of endogenous regressors (one less that the number of endogenous variables in the equation being checked (G_Δ)),

$$\boxed{K_2 \geq G_\Delta - 1}$$

Stated a little differently, the number of instrumental variables must be at least as large as the number of endogenous regressors. This condition must be satisfied for each structural equation. The values for K_2 and G_Δ may vary from one equation to another. Identities do not contain unknown parameters and need not be checked for identification.

OLS estimates of parameters in **structural models** are typically biased and inconsistent and have unreliable t-statistics. This problem is due to non-zero correlation between the error term and the endogenous regressors on the right hand side of the equation. Two stage least squares estimators (2SLS) provide biased, but consistent estimators. They can also be viewed as instrumental variables estimators.

The **Stata command** for 2SLS is

```
ivregress 2sls y1 (Y2 Y3=X1 X2) X1
```

where Y = endogenous variables (y1 on lhs, y2 and y3 on the rhs),

X1 = exogenous variables in structural equation being estimated,

X2=Z = exogenous variables in the model, but excluded from the structural equation being estimated. The variables in X2 are often called instruments. An alternative form for the two stage estimators is given by

```
ivregress 2sls y1 (Y2 Y3=X2) X1
```

"ivreg" can be used in place of the command "ivregress". 2sls is the default and the corresponding command can be written more compactly as "ivreg y1 (y2 y3=X2) X1.

Example 1: See the problem set for some sample data

Demand: $Q = \gamma_{11} - \beta_{12}P + \gamma_{12}\, Y + \varepsilon_{1t}$

Supply: $Q = \gamma_{21} + \beta_{22}P - \gamma_{23}\, FC + \varepsilon_{2t}$

ENDOGENOUS VARIABLES: Q, P

EXOGENOUS VARIABLES: Y, FC

(a) Identification

(1) Demand $K_2 = 1$ FC is in the supply model, but not in the demand equation

$G_\Delta - 1 = 2 - 1 = 1$ One endogenous regressor (P) in the demand equation

(2) Supply $K_2 = 1$ Y is in the demand model, but not in the supply equation

$G_\Delta - 1 = 2 - 1 = 1$ One endogenous regressor (P) in the supply equation

Therefore $K_2 \geq G_\Delta - 1$ is satisfied for the supply and demand equation.

(b) 2SLS estimation of the structural parameters (STATA commands)

(1) Demand

ivregress 2sls Q (P = FC) Y or ivregress 2sls Q (P=Y FC) Y

(2) Supply

Ivregress 2sls Q (P = Y) FC or ivregress 2sls Q (P=Y FC) FC

(c) Estimation of the reduced form (STATA commands)

(1) Q Equation

reg Q Y FC

(2) P Equation

reg P Y FC

Example 2. Consider the Haavelmo model and data:

$$C_t = \alpha + \beta Y_t + \varepsilon_t$$

$$Y_t = C_t + Z_t$$

(a) Identification

The exogenous variable Z is not included in the consumption function, but is in the identity.

(b) 2SLS estimation of the structural parameters (STATA commands)

```
ivregress 2sls c (Y=Z)
```

(c) Estimation of the reduced form parmaters (STATA commands)

```
reg c z

reg y z
```

The data used by Haavelmo is given

Y	C	Z
433	394	39
483	423	60
479	437	42
486	434	52
494	447	47
498	447	51
511	466	45
534	474	60
478	439	39
440	399	41
372	350	22
381	364	17
419	392	27
449	416	33
511	463	48
520	469	51
477	444	33
517	471	46
548	494	54
629	529	100

References

Haavelmo, T. "Methods of Measuring the Marginal Propensity to Consume," Journal of American Statistical Association, 42(1947):105-122.

Working, E. "What Do Statistical Demand Curves Show?" Quarterly Journal of Economics, 41(1926):212-235.