Chapter 11
Simultaneous Equations Models

Chapter Contents

Simultaneous Equations Models

- We will consider econometric models for data that are jointly determined by two or more economic relations
  - These simultaneous equations models differ from those previously studied because in each model there are two or more dependent variables rather than just one
  - Simultaneous equations models also differ from most of the econometric models we have considered so far, because they consist of a set of equations

11.1 A Supply and Demand Model

- A very simple supply and demand model might look like:
- (11.1) Demand: $Q = \alpha_1 P + \alpha_2 X + e_d$
- (11.2) Supply: $Q = \beta_1 P + e_s$
- It takes two equations to describe the supply and demand equilibrium
- The two equilibrium values, for price and quantity, P^* and Q^*, respectively, are determined at the same time

- In this model the variables P and Q are called **endogenous** variables because their values are determined within the system we have created

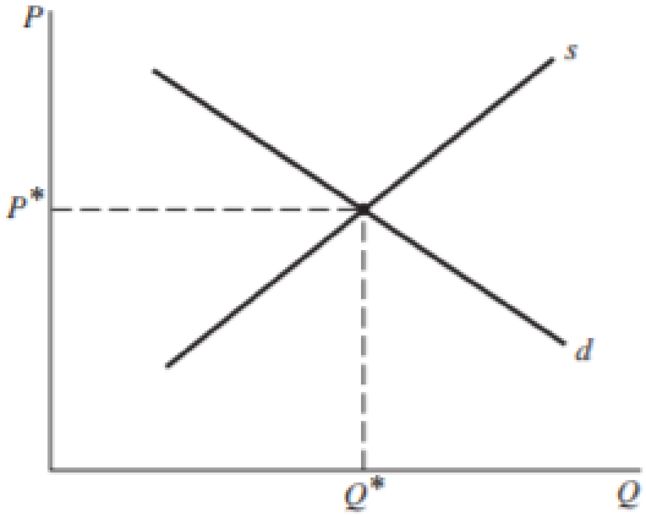

FIGURE 11.1 Supply and demand equilibrium.

11.1 A Supply and Demand Model 3 of 6

- The endogenous variables *P* and *Q* are <u>dependent variables</u> and both are random variables
- The income variable *X* has a value that is determined <u>outside this system</u>
  - Such variables are said to be **exogenous**, and these variables are treated like usual "*x*" explanatory variables

11.1 A Supply and Demand Model 4 of 6

- We refer to all the values of X_i as X, where X =(X_1 , X_2 ,…, X_N) . Then
- (11.3) $E(e_{di}|X) = 0, \quad E(e_{si}|X) = 0$
- Any value of the exogenous variable X_j is uncorrelated with the error terms in the demand and supply equations $\text{cov}(e_{di}, X_j) = 0$ and $\text{cov}(e_{si}, X_j) = 0$
- The error terms in the demand and supply equations are assumed to be homoscedastic, $\text{var}(e_{di}|X) = \sigma_d^2$, and $\text{var}(e_{si}|X) = \sigma_s^2$

11.1 A Supply and Demand Model 5 of 6

- An "influence diagram" is a graphical representation of relationships between model components
- The fact that *P* is an endogenous variable on the right-hand side of the supply and demand equations means that we have an explanatory variable that is random
  - This is contrary to the usual assumption of "fixed explanatory variables"

11.1 A Supply and Demand Model 6 of 6

- The problem is that the endogenous regressor *P* also **contemporaneously correlated** with the random errors in the demand and supply equations

- When an explanatory variable is contemporaneously correlated with the regression error term then the OLS estimator is biased and inconsistent

11.1 Influence Diagrams

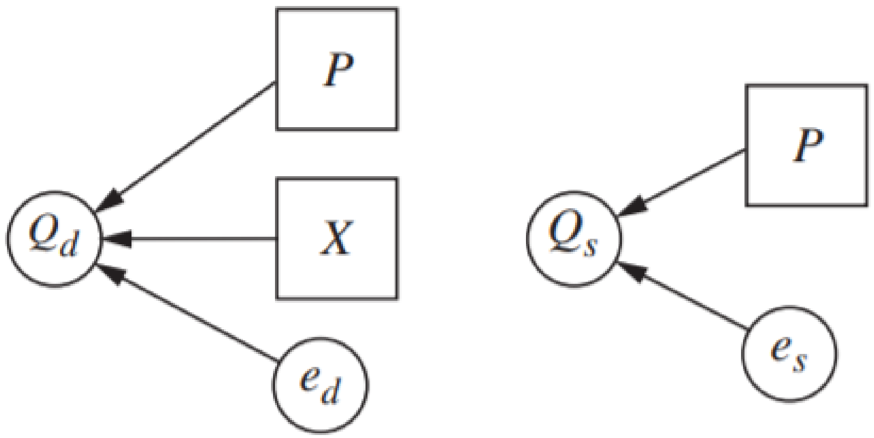

FIGURE 11.2 Influence diagrams for two regression models.

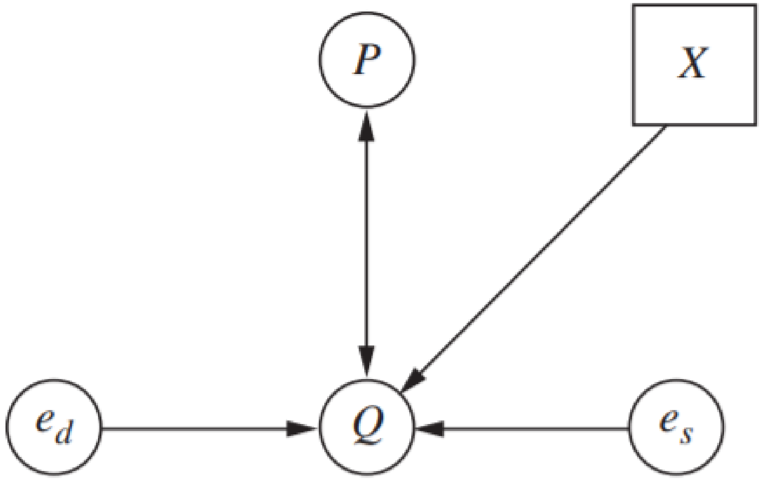

FIGURE 11.3 Influence diagram for a simultaneous equations model.

11.2 The Reduced-Form Equations 1 of 4

- The two structural equations (11.1) and (11.2) can be solved to express the endogenous variables *P* and *Q* as functions of the exogenous variable *X*
  - This reformulation of the model is called the **reduced form** of the structural equation system
- The reduced form is very important in its own right, and also helps us understand the structural equation system

11.2 The Reduced-Form Equations 2 of 4

- To solve for P, set Q in the demand and supply equations to be equal:

$$\beta_1 P_i + e_{si} = \alpha_1 P_i + \alpha_2 X_i + e_{di}$$

- Solve for P_i:
- (11.4) $P_i = \dfrac{\alpha_2}{(\beta_1 - \alpha_1)} X_i + \dfrac{e_{di} - e_{si}}{(\beta_1 - \alpha_1)} = \pi_1 X_i + v_{1i}$
- To solve for Q_i , substitute the value of P_i in (11.4) into either the demand or supply equation

11.2 The Reduced-Form Equations

- The supply equation is simpler, so substitute P_i into (11.2) and simplify:

- (11.5)

$$Q_i = \beta_1 P_i + e_{si} = \beta_1 \left[\frac{\alpha_2}{(\beta_1 - \alpha_1)} X + \frac{e_{di} - e_{si}}{(\beta_1 - \alpha_1)} \right] + e_{si}$$

$$= \frac{\beta_1 \alpha_2}{(\beta_1 - \alpha_1)} X + \frac{\beta_1 e_{di} - \alpha_1 e_{si}}{(\beta_1 - \alpha_1)} = \pi_2 X_i + v_{2i}$$

- The parameters π_1 and π_2 in equations (11.4) and (11.5) are called **reduced-form parameters**. The error terms v_{1i} and v_{2i} are called **reduced-form errors**

11.2 The Reduced-Form Equations 4 of 4

- The reduced-form equations (11.4) and (11.5) have an endogenous variable on the left-hand side and exogenous variables, and a random error term, on the right-hand side
- The terms reduced-form equation and first-stage equation are interchangeable
- The reduced-form equations are important for economic analysis
- These equations relate the equilibrium values of the endogenous variables to the exogenous variables

11.3 The Failure of Least Squares Estimation

- The least squares estimator of parameters in a structural simultaneous equation is biased and inconsistent because of the correlation between the random error and the endogenous variables on the right-hand side of the equation

- The least squares estimator of β_1 will understate the true parameter value in this model, because of the negative contemporaneous correlation between the endogenous variable P_i and the error term e_{si}

11.3.1 Proving the Failure of OLS 1 of 2

- First obtain the conditional covariance between P_{i} and e_{si}

$$\operatorname{cov}(P_i e_{si}|X) = E[P_i - E(P_i|X)][e_{si} - E(e_{si}|X)|X]$$

- The OLS estimator of the supply equation (11.2) is $b_1 = \dfrac{\sum P_i Q_i}{\sum P_i^2}$

- Substitute for Q from the reduced-form equation (11.5) and simplify

- The expected value of the least squares estimator is:

11.3.1 Proving the Failure of OLS 2 of 2

- Expected value of the sum is sum of expected values

$$E(b_1|X) = \beta_1 + \sum \left[E\left(\frac{P_i e_{si}}{\sum P_i^2}\right)|X \right]$$

- Expected value terms in the sum are not zero $\neq \beta_1$

- All we can really conclude is that the least squares estimator is biased, because e_{si} and P_i are contemporaneously correlated

11.4 The Identification Problem 1 of 3

- In the supply and demand model given by equations (11.1) and (11.2):
- The parameters of the demand equation, α_1 and α_2, *cannot* be consistently estimated by *any* estimation method
- The slope of the supply equation, β_1, *can* be consistently estimated
- It is the *absence* of variables in one equation that are *present* in another equation that makes parameter estimation possible

Figure 11.4 The Effect of Changing Income

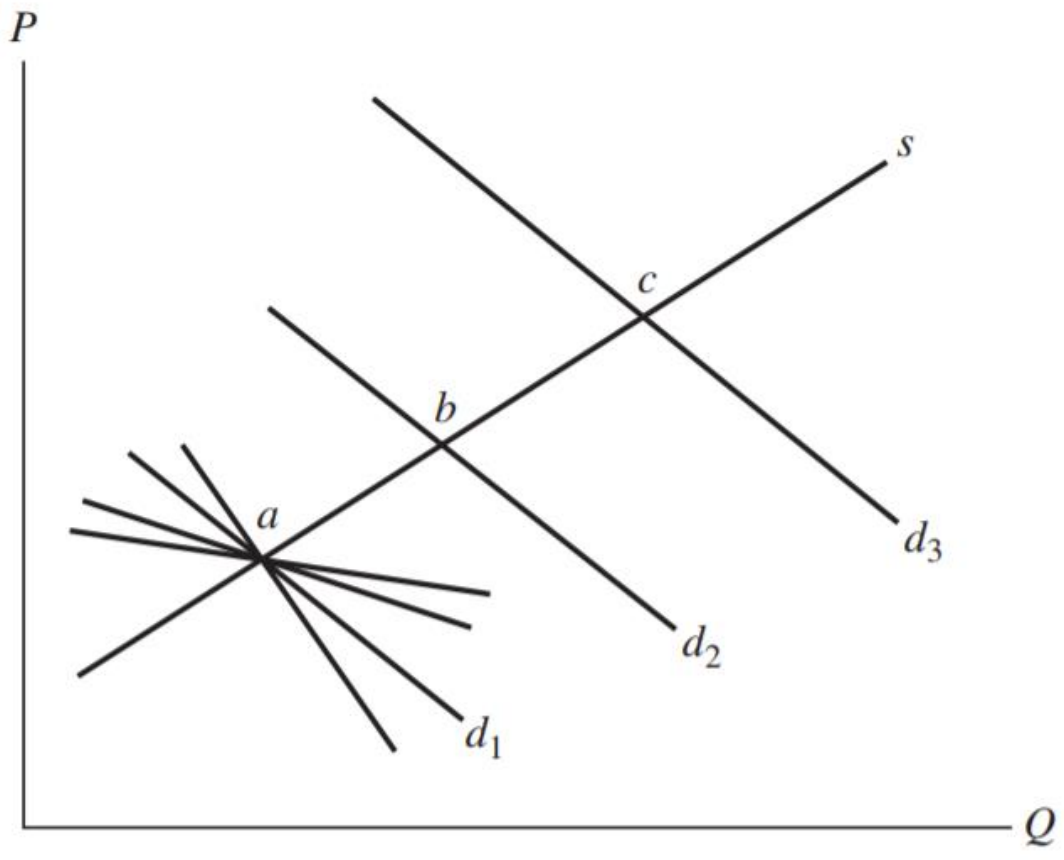

FIGURE 11.4 The effect of changing income.

11.4 The Identification Problem 2 of 3

- A general rule, which is called a **necessary condition** for *identification* of an equation, is:
- **A NECESSARY CONDITION FOR IDENTIFICATION**: In a system of M simultaneous equations, which jointly determine the values of M endogenous variables, at least M - 1 variables must be absent from an equation for estimation of its parameters to be possible

11.4 The Identification Problem 3 of 3

- When estimation of an equation's parameters is possible, then the equation is said to be *identified*, and its parameters can be estimated consistently.
- If fewer than M - 1variables are omitted from an equation, then it is said to be *unidentified*, and its parameters cannot be consistently estimated
- The identification condition must be checked before trying to estimate an equation
- If an equation is not identified, then changing the model must be considered before it is estimated

11.5 Two-Stage Least Squares Estimation 1 of 4

- The most widely used method for estimating the parameters of an identified structural equation is called **two-stage least squares**
- This is often abbreviated as 2SLS
- The name comes from the fact that it can be calculated using two OLS regressions
- Recall that we should not apply the usual OLS procedure to estimate β_1 in this equation because the endogenous variable Pi on the right-hand side of the equation is contemporaneously correlated with the error term e_{si}

11.5 Two-Stage Least Squares Estimation 2 of 4

- The reduced-form model is: (11.16) $P_i = E(P_i|X_i) + v_{1i}$
- Suppose we know $E(P_i|X_i)$
  - Then through substitution:
  - (11.7) $Q_i = \beta_1[E(P_i|X_i) + v_{1i}] + e_{si} = \beta_1 E(P_i|X_i) + (\beta_1 v_{1i} + e_{si})$
- We can *estimate* $(P_i|X_i)$ using $\hat{\pi}_1$ from the reduced-form equation for P_i
- A consistent estimator for $E(P_i|X_i)$ is: $\hat{P}_i = \hat{\pi}_1 X_i$

11.5 Two-Stage Least Squares Estimation 3 of 4

- Then:
  - (11.8) $Q = \beta_1 \hat{P}_i + e_{*i}$
- Estimating equation (11.8) by least squares generates the **two-stage least squares** estimator of β_1
  - which is consistent and asymptotically normal
- Because the two-stage least squares estimator is consistent it converges to the true value in large samples

11.5 Two-Stage Least Squares Estimation 4 of 4

- The two stages of the estimation procedure are:

1. Least squares estimation of the reduced-form equation for P_i and the calculation of its predicted value $\hat{P}_i$
2. Least squares estimation of the structural equation in which the right-hand-side endogenous variable is replaced by its predicted value $\hat{P}_i$

11.5.1 The General Two-Stage Least Squares Estimation Procedure 1 of 4

- Suppose the first structural equation in a system of *M=3* simultaneous equations is:
- (11.9) $y_{1i} = \alpha_2 y_{i2} + \alpha_3 y_{i3} + \beta_1 x_{i1} + \beta_2 x_{i2} + e_{i1}$
- If this equation is identified, then its parameters can be estimated in the two steps:

1. Estimate the parameters of the reduced-form equations

$$y_{i2} = \pi_{12}x_{i1} + \pi_{22}x_{i2} + \cdots + \pi_{K2}x_{iK} + v_{i2}$$
$$y_{i3} = \pi_{13}x_{i1} + \pi_{23}x_{i2} + \cdots + \pi_{K3}x_{iK} + v_{i3}$$

11.5.1 The General Two-Stage Least Squares Estimation Procedure 2 of 4

- Obtain the predicted values
  - (11.10) $$\hat{y}_{i2} = \hat{\pi}_{12}x_{i1} + \hat{\pi}_{22}x_{i2} + \cdots + \hat{\pi}_{K2}x_{iK}$$ $$\hat{y}_{i3} = \hat{\pi}_{13}x_{i1} + \hat{\pi}_{23}x_{i2} + \cdots + \hat{\pi}_{K3}x_{iK}$$

2. Replace the endogenous variables, y_{i2} and y_{i3}, on the right-hand side of the structural (11.9) by their predicted values from (11.10):

$$y_{i1} = \alpha_2\hat{y}_{i2} + \alpha_3\hat{y}_{i3} + \beta_1x_{i1} + \beta_2x_{i2} + e_{i1}^*$$

- Estimate the parameters of this equation by least squares

11.5.1 The General Two-Stage Least Squares Estimation Procedure 3 of 4

- In practice, we should always use software designed for 2SLS or IVs estimation. It will correctly carry out the calculations of the 2SLS estimates and their standard errors
- There are M = 3 equations so M − 1 = 2 variables must be omitted from each equation
- The alternative description of the condition for identification is that the number of omitted exogenous variables, K_1^*, must be greater than, or equal to, the number of included, right-hand side, endogenous variables

11.5.1 The General Two-Stage Least Squares Estimation Procedure 4 of 4

- Keep in mind that

1. **Two-stage least squares** and **IVs estimation** are identical
2. **IVs**, or just **instruments**, are exogenous variables that do not appear in the equation. Instruments are **excluded exogenous variables**
3. **The reduced-form equations** in simultaneous equations modeling are the **first-stage equations** in IVs, two-stage least squares, estimation

11.5.2 The Properties of the Two-Stage Least Squares Estimator 1 of 2

- The properties of the two-stage least squares estimator are as follows:
  - The 2SLS estimator is a biased estimator, but it is consistent
  - In large samples the 2SLS estimator is approximately normally distributed
  - The variances and covariances of the *2SLS* estimator are unknown in small samples, but for large samples we have expressions for them that we can use as approximations

11.5.2 The Properties of the Two-Stage Least Squares Estimator 2 of 2

- If you obtain *2SLS* estimates by applying two least squares regressions using ordinary least squares regression software, the standard errors and *t*-values reported in the second regression are not correct for the *2SLS* estimator

EXAMPLE 11.1 Supply and Demand for Truffles

- Consider a supply and demand model for truffles:
- (11.11) Demand: $Q_i = \alpha_1 + \alpha_2 P_i + \alpha_3 PS_i + \alpha_4 DI_i + e_{di}$
- (11.12) Supply: $Q_i = \beta_1 + \beta_2 P_i + \beta_3 PF_i + e_{si}$
- In the demand equation Q is the quantity of truffles traded in a particular French marketplace, indexed by I
- P is the market price of truffles, PS is the market price of a substitute for real truffles
- PF, the price of a factor of production

EXAMPLE 11.1 Supply and Demand for Truffles: Identification

- The rule for identifying an equation is:
  - In a system of *M* equations at least *M* - 1 variables must be omitted from each equation in order for it to be identified
  - In the demand equation the variable *PF* is not included; thus the necessary *M* – 1 = 1 variable is omitted
  - In the supply equation both *PS* and *DI* are absent; more than enough to satisfy the identification condition

EXAMPLE 11.1 Supply and Demand for Truffles: The reduced-form equations

- The reduced-form equations express each endogenous variable, P and Q, in terms of the exogenous variables PS, DI, PF, and the intercept, plus an error term
- They are:

$$Q_i = \pi_{11} + \pi_{21} PS_i + \pi_{31} DI_i + \pi_{41} PF_i + v_{i1}$$
$$P_i = \pi_{12} + \pi_{22} PS_i + \pi_{32} DI_i + \pi_{42} PF_i + v_{i2}$$

- We can estimate these equations by OLS since the right-hand side variables are exogenous and contemporaneously uncorrelated with the random errors vi1 and vi2

Table 11.1 Representative Truffle Data

TABLE 11.1 **Representative Truffle Data**

OBS	*P*	*Q*	*PS*	*DI*	*PF*
1	29.64	19.89	19.97	2.103	10.52
2	40.23	13.04	18.04	2.043	19.67
3	34.71	19.61	22.36	1.870	13.74
4	41.43	17.13	20.87	1.525	17.95
5	53.37	22.55	19.79	2.709	13.71
Summary Statistics					
Mean	62.72	18.46	22.02	3.53	22.75
Std. Dev.	18.72	4.61	4.08	1.04	5.33

Table 11.2a Reduced Form for Quantity of Truffles (Q)

TABLE 11.2a Reduced Form for Quantity of Truffles (Q)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.8951	3.2434	2.4342	0.0221
PS	0.6564	0.1425	4.6051	0.0001
DI	2.1672	0.7005	3.0938	0.0047
PF	−0.5070	0.1213	−4.1809	0.0003

Table 11.2b Reduced Form for Price of Truffles (P)

TABLE 11.2b Reduced Form for Price of Truffles (P)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	−32.5124	7.9842	−4.0721	0.0004
PS	1.7081	0.3509	4.8682	0.0000
DI	7.6025	1.7243	4.4089	0.0002
PF	1.3539	0.2985	4.5356	0.0001

EXAMPLE 11.1 Supply and Demand for Truffles: The Structural Equations

- A From Table 11.2b we have:

$$\hat{P}_i = \hat{\pi}_{12} + \hat{\pi}_{22}PS_i + \hat{\pi}_{32}DI_i + \hat{\pi}_{42}PF_i$$
$$= -32.512 + 1.708PS_i + 7.603DI_i + 1.354PF_i$$

TABLE 11.3a 2SLS Estimates for Truffle Demand

Variable	Coefficient	Std. Error	*t*-Statistic	Prob.
C	−4.2795	5.5439	−0.7719	0.4471
P	−0.3745	0.1648	−2.2729	0.0315
PS	1.2960	0.3552	3.6488	0.0012
DI	5.0140	2.2836	2.1957	0.0372

Table11.3b 2SLS Estimates for Truffle Supply

TABLE 11.3b **2SLS Estimates for Truffle Supply**

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	20.0328	1.2231	16.3785	0.0000
P	0.3380	0.0249	13.5629	0.0000
PF	−1.0009	0.0825	−12.1281	0.0000

EXAMPLE 11.2 Supply and Demand at the Fulton Fish Market 1 of 8

- If supply is fixed, with a vertical supply curve, then price is demand-determined
- Higher demand leads to higher prices but no increase in the quantity supplied
- If this is true, then the feedback between prices and quantities is eliminated
- Such models are said to be **recursive** and the demand equation can be estimated by ordinary least squares rather than the more complicated two-stage least squares procedure

EXAMPLE 11.2 Supply and Demand at the Fulton Fish Market 2 of 8

- A key point is that "simultaneity" does not require that events occur at a simultaneous moment in time
  - Specify the demand equation for this market as:
  - (11.13) $$\ln(QUAN_t) = \alpha_1 + \alpha_2 \ln(PRICE_t) + \alpha_3 MON_t + \alpha_4 TUE_t + \alpha_5 WED_t + \alpha_6 THU_t + e_t^d$$
  - $QUAN_t$ is the quantity sold, in pounds
  - $PRICE_t$ is the average daily price per pound, α_2 is the price elasticity of demand

EXAMPLE 11.2 Supply and Demand at the Fulton Fish Market 3 of 8

- The supply equation is:
- (11.14) $\ln(QUAN_t) = \beta_1 + \beta_2 \ln(PRICE_t) + \beta_3 STORMY_t + e_{st}$
  - β_2 is the price elasticity of supply
- The necessary condition for an equation to be identified is that in this system of M = 2 equations, it must be true that at least M – 1 = 1 variable must be omitted from each equation

EXAMPLE 11.2 Supply and Demand at the Fulton Fish Market 4 of 8

- In the demand equation the weather variable *STORMY* is omitted, and it does appear in the supply equation
- In the supply equation, the four daily indicator variables that are included in the demand equation are omitted
- Thus the demand equation shifts daily, while the supply remains fixed
  - since the supply equation does not contain the daily indicator variable

EXAMPLE 11.2 Supply and Demand at the Fulton Fish Market 5 of 8

- The reduced-form equations specify each endogenous variable as a function of all exogenous variables:
- (11.15) $$\ln(QUAN_t) = \pi_{11} + \pi_{21}MON_t + \pi_{31}TUE_t + \pi_{41}WED_t + \pi_{51}THU_t + \pi_{61}STORMY_t + v_{t1}$$
- (11.16) $$\ln(PRICE_t) = \pi_{12} + \pi_{22}MON_t + \pi_{32}TUE_t + \pi_{42}WED_t + \pi_{52}THU_t + \pi_{62}STORMY_t + v_{t2}$$

Table11.4a Reduced Form for ln(Quantity) Fish

TABLE 11.4a **Reduced Form for ln(Quantity) Fish**

Variable	Coefficient	Std. Error	*t*-Statistic	Prob.
C	8.8101	0.1470	59.9225	0.0000
STORMY	−0.3878	0.1437	−2.6979	0.0081
MON	0.1010	0.2065	0.4891	0.6258
TUE	−0.4847	0.2011	−2.4097	0.0177
WED	−0.5531	0.2058	−2.6876	0.0084
THU	0.0537	0.2010	0.2671	0.7899

Table11.4b Reduced Form for ln(Price) Fish

TABLE 11.4b Reduced Form for ln(Price) Fish

Variable	Coefficient	Std. Error	*t*-Statistic	Prob.
C	−0.2717	0.0764	−3.5569	0.0006
STORMY	0.3464	0.0747	4.6387	0.0000
MON	−0.1129	0.1073	−1.0525	0.2950
TUE	−0.0411	0.1045	−0.3937	0.6946
WED	−0.0118	0.1069	−0.1106	0.9122
THU	0.0496	0.1045	0.4753	0.6356

EXAMPLE 11.2 Supply and Demand at the Fulton Fish Market 6 of 8

- The key reduced-form equation is (11.16) for ln(*PRICE*):
- To identify the supply curve, the daily indicator variables must be jointly significant
- To identify the demand curve, the variable *STORMY* must be statistically significant
  - The supply has a significant shift variable, so that we can reliably estimate the demand equation
  - The two-stage least squares estimator performs very poorly if the shift variables are not strongly significant

EXAMPLE 11.2 Supply and Demand at the Fulton Fish Market 7 of 8

- To implement two-stage least squares, we take the predicted value from the reduced-form regression and include it in the structural equations in place of the right-hand-side endogenous variable:

$$\ln(\widehat{PRICE_t}) = \hat{\pi}_{12} + \hat{\pi}_{22}MON_t + \hat{\pi}_{32}TUE_t + \hat{\pi}_{42}WED_t + \hat{\pi}_{52}THU_t + \hat{\pi}_{62}STORMY_t$$

EXAMPLE 11.2 Supply and Demand at the Fulton Fish Market 8 of 8

- If the coefficients on the daily indicator variables are all identically zero, then:

$$\ln(\widehat{PRICE_t}) = \hat{\pi}_{12} + \hat{\pi}_{62} STORMY_t$$

- If we replace ln(*PRICE*) in the supply equation (Eq. 11.14) with this predicted value, there will be exact collinearity between $\ln(\widehat{PRICE})$ and the variable *STORMY*, which is already in the supply equation

- Two-stage least squares will fail

TABLE 11.5 2SLS Estimates for Fish Demand

TABLE 11.5 **2SLS Estimates for Fish Demand**

Variable	Coefficient	Std. Error	*t*-Statistic	Prob.
C	8.5059	0.1662	51.1890	0.0000
ln(*PRICE*)	−1.1194	0.4286	−2.6115	0.0103
MON	−0.0254	0.2148	−0.1183	0.9061
TUE	−0.5308	0.2080	−2.5518	0.0122
WED	−0.5664	0.2128	−2.6620	0.0090
THU	0.1093	0.2088	0.5233	0.6018

EXAMPLE 11.3 Klein's Model I

- The model has three equations:
  - there are eight endogenous variables and eight exogenous variables
- The first equation is a consumption function, in which aggregate consumption in year t
- CN_t is related to total wages earned by all worker
- W_t Total wages are divided into wages of workers earned in the private sector + public sector

EXAMPLE 11.3 Klein's Model I 2 of 5

- In addition, consumption expenditures are related to nonwage income (profits) in the current year, P_t, which are endogenous, and profits from the previous year
- The consumption function is
- (11.17) $$CN_t = \alpha_1 + \alpha_2(W_{1t} + W_{2t}) + \alpha_3 P_t + \alpha_4 P_{t-1} + e_{1t}$$
- In the consumption equation, W_{1t} and P_t are endogenous and contemporaneously correlated with the random error e_t
- wages in the public sector, W_{2t}, are set by public authority and are assumed exogenous and uncorrelated with the current period random error e_{1t}

EXAMPLE 11.3 Klein's Model I 3 of 5

- The second equation in the model is the investment equation. Net investment, It , is specified to be a function of current and lagged profits, P_t , and P_{t-1}
  - as well as the capital stock at the end of the previous year, K_{t-1}
  - This lagged variable is predetermined and treated as exogenous
- The investment equation is:
- (11.18) $I_t = \beta_1 + \beta_2 P_t + \beta_3 P_{t-1} + \beta_4 K_{t-1} + e_{2t}$

EXAMPLE 11.3 Klein's Model I

- There is an equation for wages in the private sector, W_{1t}
- Let $E_t = CN_t + I_t + (G_t - W_{2t})$
- G_t is government spending
- Consumption and investment are endogenous and government spending and public sector wages are exogenous
- The sum, E_t , total national product minus public sector wages, is endogenous
- Wages are taken to be related to Et and the predetermined variable E_{t-1}, plus a time trend variable

EXAMPLE 11.3 Klein's Model I

- The wage equation is
- (11.19) $W_{1t} = \gamma_1 + \gamma_2 E_t + \gamma_3 E_{t-1} + \gamma_4 TIME_t + e_{3t}$
- In total, there are eight exogenous and predetermined variables, which can be used as Ivs
- Another exogenous variable is the constant term, the "intercept" variable in each equation, $X_{it} \equiv 1$
- The predetermined variables are lagged profits, P_{t-1}, the lagged capital stock, K_{t-1}, and the lagged total national product minus public sector wages, E_{t-1}

Key Words

- contemporaneous correlation
- endogenous variables
- exogenous variables
- first-stage equation
- identification
- instruments
- instrumental variables (IV) estimator
- predetermined variables
- reduced-form equation
- reduced-form errors
- reduced-form parameters
- simultaneous equations
- structural parameters
- two-stage least squares

Copyright