Exceptions to Ordinary Least Squares
Lecture XXXI

I. Heteroscedasticity

A. Using the derivation of the variance of ordinary least squares estimator

$$\hat{\beta} = \beta + (X'X)^{-1} X'\varepsilon \Rightarrow V(\hat{\beta}) = (X'X)^{-1} X'\varepsilon\varepsilon' X (X'X)^{-1}$$

$$V(\hat{\beta}) = (X'X)^{-1} X'SX (X'X)^{-1} \ni : S = E(\varepsilon\varepsilon')$$

under the Gauss-Markov assumptions $S = E(\varepsilon\varepsilon') = \sigma^2 I_{T\times T}$.

B. However, if we assume that $S = E(\varepsilon\varepsilon') \neq \sigma^2 I_{T\times T}$ the ordinary least squares estimator is still unbiased, but is no longer efficient. In this case, we use the generalized least squares estimator

$$\tilde{\beta} = (X'AX)^{-1}(X'Ay)$$

1. The variance of this estimator is then

$$\begin{aligned}
\tilde{\beta} &= (X'AX)^{-1}(X'AX\beta + X'A\varepsilon) \\
&= (X'AX)^{-1}(X'AX)\beta + (X'AX)^{-1} X'A\varepsilon \\
&= \beta + (X'AX)^{-1} X'A\varepsilon \\
V(\tilde{\beta} - \beta) &= (X'AX)^{-1} X'A\varepsilon\varepsilon'A'X (X'AX)^{-1} \\
&= (X'AX)^{-1} X'ASA'X (X'AX)^{-1}
\end{aligned}$$

2. Setting $A = S^{-1}$

$$\begin{aligned}
V(\tilde{\beta} - \beta) &= (X'AX)^{-1} X'A'X (X'AX)^{-1} \ni : A' = A \\
&= (X'AX)^{-1}
\end{aligned}$$

C. Seemingly Unrelated Regressions

1. One of the uses of generalized least squares is the estimation of simultaneous systems of equations without endogeneity.

   a) Derived input demand equations derived from cost minimization implies relationship between the parameters

$$\begin{aligned}
x_1 &= \alpha_1 + A_{11} w_1 + A_{12} w_2 + \Gamma_{11} y + \varepsilon_1 \\
x_2 &= \alpha_2 + A_{21} w_1 + A_{22} w_2 + \Gamma_{21} y + \varepsilon_2
\end{aligned}$$

where x_1 and x_2 are input levels, w_1 and w_2 are the respective input prices, y is the level of output, and α_1, α_2, A_{11}, A_{12}, A_{21}, A_{22}, Γ_{11} and Γ_{21} are estimated parameters.

b) Both relationships can be estimated simultaneously by forming the regression matrices as

$$\begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1n} \\ x_{21} \\ x_{22} \\ \vdots \\ x_{2n} \end{bmatrix} = \begin{bmatrix} 1 & w_{11} & w_{21} & y_1 & 0 & 0 & 0 & 0 \\ 1 & w_{12} & w_{22} & y_2 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & w_{1n} & w_{2n} & y_n & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & w_{11} & w_{21} & y_1 \\ 0 & 0 & 0 & 0 & 1 & w_{12} & w_{22} & y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 1 & w_{1n} & w_{2n} & y_n \end{bmatrix} \begin{bmatrix} \alpha_1 \\ A_{11} \\ A_{12} \\ \Gamma_{11} \\ \alpha_2 \\ A_{21} \\ A_{22} \\ \Gamma_{21} \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1n} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{2n} \end{bmatrix}$$

c) It would be tempting to conclude that this formulation implies that the input demand system requires generalized least squares estimation. Specifically, using a two-step methodology, we can estimate the parameter vector using ordinary least squares. The ordinary least squares coefficients could then be used to estimate the variance for each equation. This variance could then be used to estimate the A matrix

$$A = \begin{bmatrix} 1/s_1^2 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1/s_1^2 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/s_1^2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1/s_2^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1/s_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1/s_2^2 \end{bmatrix}$$

d) However, the separable nature the estimation implies that there is no change in efficiency. To introduce changes in efficiency, we need to impose the restriction that $A_{12} = A_{21}$. Imposing this restriction on the matrix of independent variables implies

$$\begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1n} \\ x_{21} \\ x_{22} \\ \vdots \\ x_{2n} \end{bmatrix} = \begin{bmatrix} 1 & w_{11} & w_{21} & y_1 & 0 & 0 & 0 \\ 1 & w_{12} & w_{22} & y_2 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & w_{1n} & w_{2n} & y_n & 0 & 0 & 0 \\ 0 & 0 & w_{11} & 0 & 1 & w_{21} & y_1 \\ 0 & 0 & w_{12} & 0 & 1 & w_{22} & y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & w_{1n} & 0 & 1 & w_{2n} & y_n \end{bmatrix} \begin{bmatrix} \alpha_1 \\ A_{11} \\ A_{12} \\ \Gamma_{11} \\ \alpha_2 \\ A_{22} \\ \Gamma_{21} \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1n} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{2n} \end{bmatrix}$$

e) In the later case, generalized least squares will yield efficiency gains.

II. Two Stage Least Squares and Instrumental Variables

A. The forgoing example does not involve dependency between equations. For example, assume that the supply and demand curves for a given market can be written as

$$q^s = -3.0 + 4p_1 - p_2 + \varepsilon_1$$
$$q^d = 10 - p_1 + 2y + \varepsilon_2$$

Solving this two equation systems yields

$$p_1 = \frac{13}{5} + \frac{1}{5}p_2 + \frac{2}{5}y - \varepsilon_1 + \varepsilon_2$$

1. Ignoring the problem of simultaneity, the supply equation can be estimated as

$$q^s = \alpha_0 + \alpha_1 p_1 + \alpha_2 p_2 + \varepsilon_1$$

2. The results for this simple estimation are presented in Table 1:

Table 1. Ordinary Least Squares

	Ordinary Least Squares
α_0	1.2273
	(1.4657)
α_1	3.3280
	(0.1930)
α_2	-0.7181
	(0.1566)

3. Obviously, these results are not close to the true values: Why?
4. The basic problem is a simultaneous equation bias. Substituting the solution of p_1 into the estimated equation yields

$$q^s = \alpha_0 + \alpha_1\left(\frac{13}{5} + \frac{1}{5}p_2 + \frac{2}{5}y - \varepsilon_1 + \varepsilon_2\right) + \alpha_2 p_2 + \varepsilon_1$$

Substituting $\tilde{p}_1 = \frac{13}{5} + \frac{1}{5}p_2 + \frac{2}{5}y + \varepsilon_2 \Rightarrow p_1 = \tilde{p}_1 - \varepsilon_1$ we note that the x matrix is now correlated with the residual vector. Specifically

$$E\ \ p_1\varepsilon_1\ \ = -\sigma_1^2 \neq 0$$

B. Two Stage Least Squares

1. The first approach developed by Theil was to estimate the reduced form of the price model and then use this estimated value in the regression.
2. In this example

$$\hat{p}_1 = \gamma_0 + \gamma_1 p_2 + \gamma_3 y + \nu$$

yielding:

Table 2. First Stage Estimation

	Ordinary Least Squares
γ_0	2.65762
	(0.23262)
γ_1	0.15061
	(0.03497)
γ_2	0.40602
	(0.01863)

3. Next, we estimate the supply equation using the

$$q^s = \tilde{\alpha}_0 + \tilde{\alpha}_1 \hat{p}_1 + \tilde{\alpha}_2 p_2 + \varepsilon_1$$

Table 3. Two Stage Least Squares Estimates

	Two Stage Least Squares
$\tilde{\alpha}_0$	-3.27703
	(0.41171)
$\tilde{\alpha}_1$	3.95311
	(0.05456)
$\tilde{\alpha}_2$	-0.74746
	(0.04128)

4. In the same way estimating the demand equation as

$$q^d = \tilde{\beta}_0 + \tilde{\beta}_1 \hat{p}_1 + \tilde{\beta}_2 y + \varepsilon_2$$

Table 4. Two Stage Least Squares Estimator for Demand

	Two Stage Least Squares
$\tilde{\beta}_0$	9.9121
	(0.9118)
$\tilde{\beta}_1$	-1.0096
	(0.2761)
$\tilde{\beta}_2$	2.0150
	(0.1113)

C. Generalized Instrumental Variables

1. The alternative would be to use variables as instruments to remove the correlation between endogenous variables. In this case, we assume that

$$y = X\beta + \varepsilon$$

a) Under the endogeneity assumption

$$\frac{1}{N} X'\varepsilon \nrightarrow 0$$

b) But, we have a set of instrument (Z) which are correlated with the residuals and imperfectly correlated with X .

2. The generalized instrumental variable solution is

$$\beta_{IV} = \left(X'P_Z X\right)^{-1}\left(X'P_Z y\right)$$

where $P_Z = Z\left(Z'Z\right)^{-1}Z'$.

a) In the current case, we use $Z = \begin{bmatrix} 1 & p_2 & y \end{bmatrix}$ yielding

$$\beta_{IV} = \begin{bmatrix} -3.2770 \\ 3.9531 \\ -0.7475 \end{bmatrix}$$