17.874 Lecture Notes
Part 7: Systems of Equations

7. Systems of Equations

Many important social science problems are more structured than a single relationship or function. Markets, game theoretic models, decision theory models, and evolutionary models all have predictions about behavior that take the form of equilibria. We would like to study a single relationship, say, how price changes affect the supply of a commodity or the effect of an increase in policing on the crime rate. However, the outcomes are determined by more than one equation, making it difficult to isolate a particular effect or relationship of interest.

Specifically, least squares will usually not estimate the correct relationship between two variables when outcomes are "jointly determined" – that is, when X causes Y and Y causes X.

Some examples illustrate the range of topics in which this problem arises.

i Markets. Bargaining among buyers and sellers is represented in the aggregate by a supply curve and a demand curve. The Quantity Supplied by sellers is increasing in the Price that buyers are willing to pay, as well as other factors affecting production. The Quantity Demanded by buyers is decreasing the in price sellers are willing to take, as well as other factors affecting consumption, such as income. In a market-clearing equilibrium, Quantity Demanded equals Quantity Supplied and Prices are equal as well.

ii Social Policy and Crime. People will commit crimes to the extent that they face wages low enough such that the benefit of committing a crime times the probability of being caught and punished exceeds the wages they would receive if they did not commit a crime. Society, through public policies about the number of police, the severity of punishments, and the operation of the prison system, chooses the probability of being caught and punished as well as the wages lost when punished. The number of police on the streets, then, is a function of the underlying forces that produce crime. The crime rate is a function of the number of police.

iii Incumbency advantages and Strategic career decisions. Estimates of the incumbency advantage contrast the vote shares of incumbents and non-incumbents. If people retire because of short-term fluctuations in the (expected) vote, then incumbency might depend on the vote, as much as the vote depends on incumbency.

Mathematical models represent these relationships with sets of simultaneous equations. The dependent variable in one equation is an independent variable in another equation. Solving these equations yields the equilibrium levels of the dependent variables. In this respect the values of the dependent variables are jointly determined by the system of equations.

7.1. Structural and Reduced Forms

When two variables are jointly determined they are both dependent and independent variables, but in different equations. Usually, there are also a set of independent variables that are only on the right-hand sides of the various equations. A structural model expresses the relationships among dependent and independent variables. We call the dependent variables *endogenous* variables, because they are determined within the entire model, even if a particular model is not on the left-hand side in a particular equation. We call the independent variables *exogenous* variables.

Consider the following simple model in which there are two independent variables X_1 and X_2 and two dependent variables Y_1 and Y_2. To give the variables meanings, Y_1 might be price and Y_2 quantity in a market model. Or, Y_1 might be vote and Y_2 whether an incumbent runs.

$$y_{1i} = \alpha_{10} + \alpha_{11} y_{2i} + \alpha_{13} X_{1i} + \epsilon_{1i} \tag{1}$$

$$y_{2i} = \alpha_{20} + \alpha_{21} y_{1i} + \alpha_{23} X_{2i} + \epsilon_{2i} \tag{2}$$

With this model we can refine our terminology somewhat. For each equation, we consider which exogenous variables are in the model and which side of the equation the endogenous variables are on. If the endogenous variable is on the left-hand side we refer to it as the dependent variable in an equation. Any endogenous variables on the right-hand side are referred to as endogenous included variables. Any endogenous variables that are not included in a particular equation are called endogenous excluded. In this model there are no endogenous excluded variables in either equation.

Parallel terminology is used for the independent variables. Exogenous variables included in a given structural equation are called exogenous included variables. Exogenous variables not included in a given structural equation are called exogenous excluded variables.

For example, economists generally conjecture that income affects demand (consumption) for food directly, but weather does not affect demand directly. Hence income is an included variable in the demand curve equation, but weather is not. Similarly, weather directly affects the supply of food, but not demand directly (unless there is hoarding).

More general notation for the structural model can be derived from the above notation. Isolate the y variables on one side of the model.

$$y_{1i} - \alpha_{11} y_{2i} = \alpha_{10} + \alpha_{13} X_{1i} + \epsilon_{1i} \tag{3}$$

$$-\alpha_{21} y_{1i} + y_{2i} = \alpha_{20} + \alpha_{23} X_{2i} + \epsilon_{2i} \tag{4}$$

The matrix $\mathbf{Y}$ has columns of Y_1 and Y_2. Denote $\mathbf{\Gamma}$ as the parameters such that $\mathbf{\Gamma Y}$ returns the left-hand side of the equations above. That is,

$$\mathbf{Y\Gamma} = \begin{pmatrix} y_{11}, y_{21} \\ y_{12}, y_{22} \\ y_{13}, y_{23} \\ . \\ . \\ y_{1n}, y_{2n} \end{pmatrix} \begin{pmatrix} 1, -\alpha_{11} \\ -\alpha_{21}, 1 \end{pmatrix}$$

The matrix $\mathbf{X}$ contains columns corresponding to all of the exogenous variables in the entire system. Denote $\mathbf{B}$ as the parameters such that $\mathbf{BX}$ returns the right-hand side of the equations above.

$$\mathbf{XB} = \begin{pmatrix} 1, x_{11}, x_{21} \\ 1, x_{12}, x_{22} \\ 1, x_{13}, x_{23} \\ . \\ . \\ 1, x_{1n}, x_{2n} \end{pmatrix} \begin{pmatrix} \alpha_{10}, \alpha_{13}, 0 \\ \alpha_{20}, 0, \alpha_{23} \end{pmatrix}$$

Finally, the matrix $\mathbf{E}$ has columns equal to the two error terms.

$$\mathbf{E} = \begin{pmatrix} \epsilon_{11}, \epsilon_{21} \\ \epsilon_{12}, \epsilon_{22} \\ \epsilon_{13}, \epsilon_{23} \\ . \\ . \\ \epsilon_{1n}, \epsilon_{2n} \end{pmatrix}$$

We may write any system of equations as:

$$\mathbf{Y\Gamma} = \mathbf{XB} + \mathbf{E}$$

We may solve the systems equations in terms Y. Pre-multiplying both sides of the equation by the inverse of $\mathbf{\Gamma}$,

$$\mathbf{Y} = \mathbf{X\Gamma^{-1}B} + \mathbf{E\Gamma^{-1}} = \mathbf{X\Pi} + \mathbf{V}$$

Such a solution is called the *reduced form*. It expresses the endogenous variables in terms of the exogenous variables. Of note, even if an exogenous variable is excluded in the structural model for a specific equation, it need not be in the reduced form equation for a given endogenous variable.

We can use this express to solve the simple model above.

$$\mathbf{\Pi} = \mathbf{bf}\mathbf{B\Gamma^{-1}} = \begin{pmatrix} \alpha_{10}, \alpha_{13}, 0 \\ \alpha_{20}, 0, \alpha_{23} \end{pmatrix} \frac{\mathbf{1}}{\mathbf{1} - \alpha_{11}\alpha_{21}} \begin{pmatrix} 1, -\alpha_{11} \\ -\alpha_{21}, 1 \end{pmatrix}$$

The statistical model is further characterized by a set of assumptions about the errors. We make the usual statistical assumptions about this model for the application of the least squares estimator: $E[\mathbf{X}'\mathbf{E}] = \mathbf{0}$. We will also assume that the error in one equation is independent of the error in the other equations and that there is no autocorrelation or heterskedasticity among the errors.

The matrix of residuals has J columns corresponding to J equations and n rows, for the observations. For a given observation i, the variance covariance matrix of the errors of the J structural equations is:

$$\mathbf{\Sigma} = \mathbf{E}[\epsilon_{\mathrm{i}}\epsilon_{\mathrm{i}}']$$

The variance covariance matrix of the errors of the reduced form equations is:

$$E[(\mathbf{\Gamma}^{-1})'\epsilon_{\mathrm{i}}\epsilon_{\mathrm{i}}'\mathbf{\Gamma}^{-1}] = (\mathbf{\Gamma}^{-1})'\mathbf{\Sigma}\mathbf{\Gamma}^{-1} =$$

With these assumptions we can characterize the covariance structure of the data. Let $plim\frac{1}{n}\mathbf{X}'\mathbf{X} = \mathbf{Q}$.

$$plim\frac{1}{n}\begin{bmatrix}\mathbf{Y}_{\mathrm{j}}' \\ \mathbf{X}'\end{bmatrix}[\mathbf{Y}_{\mathrm{j}}\mathbf{X}] = plim\frac{1}{n}\begin{pmatrix}\mathbf{Y}'\mathbf{Y}, \mathbf{Y}'\mathbf{X}, \mathbf{Y}'\mathbf{V} \\ \mathbf{X}'\mathbf{Y}, \mathbf{X}'\mathbf{X}, \mathbf{X}'\mathbf{V} \\ \mathbf{V}'\mathbf{Y}, \mathbf{V}'\mathbf{X}, \mathbf{V}'\mathbf{V}\end{pmatrix} = \begin{pmatrix}\mathbf{\Pi}'\mathbf{Q}\mathbf{\Pi} + \quad, \mathbf{\Pi}'\mathbf{Q}, & \\ \mathbf{Q}'\mathbf{\Pi}, \mathbf{Q}, \mathbf{0} & \\ ' + \mathbf{\Pi}'\mathbf{Q}\mathbf{\Pi}, \mathbf{0}, & \end{pmatrix}$$

7.2. Estimation

The objective is to use the data to estimate the parameters of the behavioral (structural) relationships. Under the assumption that the exogenous variables are uncorrelated with the errors – the usual regression assumptions – we can derive unbiased estimates of the paramters of the reduced form model using least squares.

Unfortunately, these are not the parameters of interest, the α's. They are functions of the α's. An array of values of parameters of the structural equations are consistent with the estimates in the reduced form equations. That many different parameter values are consistent with the same data indicates that there are many models that are observationally equivalent. In other words, the behavioral parameters are not identified from the least squares regressions for the reduced form equations, and least squares applied to the structural equations may be biased.

Strategies.

One approach is to use least squares on the structural equation. Least squares may still be alright. Simultaneity may not be a problem, or it may be sufficiently weak that the cures are worse than the disease. Indeed, this is what we do much of the time when we estimate regression models. As shown below, however, least squares is not guaranteed to be unbiased or consistent.

Instrumental Variables estimators provide a natural procedure for estimating a single equation. If we can find other factors that affect the behavioral relationships separately, then we can use these to construct instruments. Specifically, we will use the excluded exogenous variables from a given equation to estimate the predicted values of the included endogenous variables.

There is a definite tradeoff between the efficiency of least squares and the consistency of instrumental variables estimation, as discussed earlier in this course. If the consistency gains are not sufficiently good, we may still wish to use least squares, tolerating, of course, some bias.

On top of this, there is a practical research design matter. It is really hard to find good instruments for many problems. A good instrument must satisfy the exclusion restriction and it must have substantial statistical effect on the endogenous included variable.

Other approaches have been proposed. But, they are invariably problematic. One approach to the structural equations is called a path model. Path models were commonly employed in the 1960s and 1970s, and we still see them occasionally. The correlations among the variables are calculated. To calculate the importance of any path, the correlations along that path are multiplied together.

A more sensible approach is to try to identify the coefficients from other restrictions. We may impose restrictions on the covariance matrices of the errors. If we assume that the errors of the equations are uncorrelated, then we gain additional degrees of freedom to work with in identifying the parameters.

Ordinarly Least Squares is Biased and Inconsistent.

Suppose we use OLS to estimate each equation in the model as it is specified in the original structural form. Assume that the data are mean deviated so that the intercepts are no longer parameters in the model.

Let's focus on the first equation. The OLS estimate of the parameters is:

$$\mathbf{a}_1 = \begin{pmatrix} \mathbf{y}_2{}'\mathbf{y}_2, \mathbf{y}_2{}'\mathbf{x}_1 \\ \mathbf{y}_2{}'\mathbf{x}_1, \mathbf{x}_1{}'\mathbf{x}_1 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{y}_2'\mathbf{y}_1 \\ \mathbf{x}_1'\mathbf{y}_1 \end{pmatrix} = \frac{\mathbf{1}}{\mathbf{y}_2'\mathbf{y}_2\mathbf{x}_1'\mathbf{x}_1 - \mathbf{y}_2'\mathbf{x}_1{}^2} \begin{pmatrix} (\mathbf{x}_1'\mathbf{x}_1)(\mathbf{y}_2'\mathbf{y}_1) - (\mathbf{y}_2'\mathbf{x}_1)(\mathbf{x}_1'\mathbf{y}_1) \\ (\mathbf{x}_1'\mathbf{y}_1)(\mathbf{y}_2'\mathbf{y}_2) - (\mathbf{y}_2'\mathbf{x}_1)(\mathbf{y}_2'\mathbf{y}_1) \end{pmatrix}$$

We can simplify this further. Denote the determinant as $D = \mathbf{y}_2'\mathbf{y}_2\mathbf{x}_1'\mathbf{x}_1 - \mathbf{y}_2'\mathbf{x}_1{}^2$. Substitute in the structural form equation for y_1. Then,

$$\mathbf{a}_1 = \frac{\mathbf{1}}{\mathbf{D}} \begin{pmatrix} (\mathbf{x}_1'\mathbf{x}_1)(\alpha_{11}\mathbf{y}_2'\mathbf{y}_2 + \alpha_{12}\mathbf{y}_2'\mathbf{x}_1 + \mathbf{y}_2'\epsilon_1) - (\mathbf{y}_2'\mathbf{x}_1)(\alpha_{11}\mathbf{x}_1'\mathbf{y}_2 + \alpha_{12}\mathbf{x}_1'\mathbf{x}_1 + \mathbf{x}_1'\epsilon_1) \\ (\mathbf{y}_2'\mathbf{y}_2)(\alpha_{11}\mathbf{x}_1'\mathbf{y}_2 + \alpha_{12}\mathbf{x}_1'\mathbf{x}_1 + \mathbf{x}_1'\epsilon_1) - (\mathbf{y}_2'\mathbf{x}_1)(\alpha_{11}\mathbf{y}_2'\mathbf{y}_2 + \alpha_{12}\mathbf{y}_2'\mathbf{x}_1 + \mathbf{y}_2'\epsilon_1) \end{pmatrix}$$

Collecting terms and canceling yields

$$\mathbf{a}_1 = \begin{pmatrix} \alpha_{11} + \frac{1}{\mathrm{D}}[\mathbf{x}_1'\mathbf{x}_1\mathbf{y}_2'\epsilon_1 - \mathbf{x}_1'\mathbf{y}_2\mathbf{x}_1'\epsilon_1] \\ \alpha_{12} + \frac{1}{\mathrm{D}}[\mathbf{y}_2'\mathbf{y}_2\mathbf{x}_1'\epsilon_1 - \mathbf{x}_1'\mathbf{y}_2\mathbf{y}_2'\epsilon_1] \end{pmatrix}$$

This vector does not converge to the vector of true values. Each element will be asymptotically biased. The $plim\frac{1}{n}x_1'\epsilon_1 = 0$. However, $plim\frac{1}{n}y_2'\epsilon_1 = g_2\sigma^2_{\epsilon_1}$, where g_2 is the appropriate element from $\mathbf{\Gamma}^{-1}$. The denominator also converges to a finite number, so both coefficients in equation (1) are biased when we estimate the equation using Ordinary Least Squares.

More generally, let $\mathbf{Z}_\mathrm{j} = [\mathbf{Y}_\mathrm{j}, \mathbf{X}_\mathrm{j}]$ and let $\delta_\mathrm{j} = [\gamma_\mathrm{j}, \beta_\mathrm{j}]$. Then, the jth equation is represented as

$$y_j = \mathbf{Y}_\mathrm{j}\gamma_\mathrm{j} + \mathbf{X}_\mathrm{j}\beta_\mathrm{j} + \epsilon_\mathrm{j}$$

$$y_j = \mathbf{Z}_\mathrm{j}\delta_\mathrm{j} + \epsilon_\mathrm{j}$$

The OLS estimator is

$$d_j = (\mathbf{Z}_\mathrm{j}{}'\mathbf{Z}_\mathrm{j})^{-1}\mathbf{Z}_\mathrm{j}'\mathbf{y}_\mathrm{j} = \delta_\mathrm{j} + (\mathbf{Z}_\mathrm{j}'\mathbf{Z}_\mathrm{j})^{-1}\mathbf{Z}_\mathrm{j}'\epsilon_\mathrm{j}$$

and

$$plim \frac{1}{n}\mathbf{d}_{\mathrm{j}} = \delta_{\mathrm{j}} + \begin{pmatrix} \mathbf{\Pi}'\mathbf{Q}\mathbf{\Pi} + & , \mathbf{\Pi}'\mathbf{Q} \\ \mathbf{Q}'\mathbf{\Pi}, \mathbf{Q} & \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{\Gamma}^{-1\prime}\sigma^2_{\epsilon_1} \\ 0 \end{pmatrix}$$

Instrumental Variables are Consistent.

There are a variety of instrumental variables estimators, depending on whether we implement the method one equation at a time or for all equations at once. We will focus on the single equation estimation approach – Two Stage Least Squares.

The first stage in this estimator is to compute the reduced for equations, and generate predicted values of the endogenous included variables using those. In the two-variable example above, regress y_2 on X_1 and X_2. The predicted values are $\hat{\mathbf{y}_2} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_2$. The second stage is to regress y_1 on X_1 (the included exogenous variable) and $\hat{y_2}$. The first stage purges y_2 of the component from the error term that created the bias. The first stage is consistent. It remains to be shown that the second stage is as well.

In more general notation, the two-stage least squares estimator is

$$\hat{\delta}_{j,2SLS} = (\hat{\mathbf{Z}}_{\mathrm{j}}'\mathbf{Z}_{\mathrm{j}})^{-1}\hat{\mathbf{Z}}_{\mathrm{j}}'\mathbf{y}_{\mathrm{j}},$$

where $\hat{\mathbf{Z}}_{\mathrm{j}}' = [\hat{\mathbf{Y}}_{\mathrm{j}}', \mathbf{X}_{\mathrm{j}}']$. This can be rewritten as

$$\hat{\delta}_{j,2SLS} = (\hat{\mathbf{Z}}_{\mathrm{j}}'\mathbf{Z}_{\mathrm{j}})^{-1}\hat{\mathbf{Z}}_{\mathrm{j}}'(\mathbf{Z}_{\mathrm{j}}\delta_{\mathrm{j}} + \epsilon_{\mathrm{j}}) = \delta_{\mathrm{j}} + (\hat{\mathbf{Z}}_{\mathrm{j}}'\mathbf{Z}_{\mathrm{j}})^{-1}\hat{\mathbf{Z}}_{\mathrm{j}}'\epsilon_{\mathrm{j}}$$

To show consistency we can show that the $plim \hat{\mathbf{Z}}_{\mathrm{j}}'\epsilon_{\mathrm{j}} = \mathbf{0}$. First, let's rewrite $\hat{\mathbf{Z}}_{\mathrm{j}}$. Recall, that $\mathbf{Z}_{\mathrm{j}}' = [\mathbf{Y}_{\mathrm{j}}', \mathbf{X}']$. So, $\hat{\mathbf{Z}}_{\mathrm{j}}' = [\hat{\mathbf{Y}}_{\mathrm{j}}', \mathbf{X}'] = [(\mathbf{X}\hat{\mathbf{\Pi}})', \mathbf{X}'] = [\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_{\mathrm{j}}', \mathbf{X}']$. Using this expression for $\hat{\mathbf{Z}}_{\mathrm{j}}$,

$$plim \hat{\mathbf{Z}}_{\mathrm{j}}'\epsilon_{\mathrm{j}} = \mathbf{plim} \begin{pmatrix} (\mathbf{Y}_{\mathrm{j}}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon_{\mathrm{j}} \\ \mathbf{X}'\epsilon_{\mathrm{j}} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

7.3. Inference

7.3.1. Inferences About Parameters

It is straigtforward to see that like OLS, the 2SLS estimator is a sum of random variables and, therefore, will follow a normal distribution. We can perform inferences about single parameters and sets of parameters using t-statistics and F-statistics, as with OLS. What must be determined, though, is the variance of this estimator.

Because this is an instrumental variables estimator, we can show that the estimated asymptotic variance-covariance matrix of the coefficients is of the form:

$$asym.V[\hat{\delta}_{\mathrm{j}}] = \sigma_{\mathrm{jj}}[(\mathbf{Z}_{\mathrm{j}}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Z}_{\mathrm{j}})]$$

Which can be estimated as:

$$Est.Asym.V[\hat{\delta}_{\mathrm{j}}] = \hat{\sigma}_{\mathrm{jj}}[(\hat{\mathbf{Z}}_{\mathrm{j}}'\hat{\mathbf{Z}}_{\mathrm{j}})]^{-1}$$

The definition of the variance of the estimator is:

$$V[\hat{\delta}_{\mathrm{j}}] = \mathbf{E}[(\hat{\delta}_{\mathrm{j}} - \delta_{\mathrm{j}})(\hat{\delta}_{\mathrm{j}} - \delta_{\mathrm{j}})']$$

Using the expression for the estimator as a function of δ_{j} and ϵ,

$$V[\hat{\delta}_{\mathrm{j}}] = \mathbf{E}[(\delta_{\mathrm{j}} + (\hat{\mathbf{Z}}_{\mathrm{j}}'\mathbf{Z}_{\mathrm{j}})^{-1}\hat{\mathbf{Z}}_{\mathrm{j}}'\epsilon_{\mathrm{j}} - \delta_{\mathrm{j}})(\delta_{\mathrm{j}} + (\hat{\mathbf{Z}}_{\mathrm{j}}'\mathbf{Z}_{\mathrm{j}})^{-1}\hat{\mathbf{Z}}_{\mathrm{j}}'\epsilon_{\mathrm{j}} - \delta_{\mathrm{j}})'] = \mathbf{E}[(\hat{\mathbf{Z}}_{\mathrm{j}}'\mathbf{Z}_{\mathrm{j}})^{-1}\hat{\mathbf{Z}}_{\mathrm{j}}'\epsilon_{\mathrm{j}}\epsilon_{\mathrm{j}}'\hat{\mathbf{Z}}_{\mathrm{j}}(\hat{\mathbf{Z}}_{\mathrm{j}}'\mathbf{Z}_{\mathrm{j}})^{-1}]$$

Note that $\hat{\mathbf{Z}}_{\mathrm{j}}'\mathbf{Z}_{\mathrm{j}} = \mathbf{Z}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z} = \hat{\mathbf{Z}}_{\mathrm{j}}'\hat{\mathbf{Z}}_{\mathrm{j}}$. Why?

$$\hat{\mathbf{Z}}_{\mathrm{j}}'\mathbf{Z}_{\mathrm{j}} = \begin{pmatrix} (\mathbf{Y}_{\mathrm{j}}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ \mathbf{X}' \end{pmatrix} \begin{pmatrix} \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_{\mathrm{j}} \\ \mathbf{X} \end{pmatrix}$$

$$= \begin{pmatrix} (\mathbf{Y}_{\mathrm{j}}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_{\mathrm{j}}, & (\mathbf{Y}_{\mathrm{j}}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} \\ \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_{\mathrm{j}}, & \mathbf{X}'\mathbf{X} \end{pmatrix}$$

$$= \begin{pmatrix} (\mathbf{Y}_{\mathrm{j}}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y}_{\mathrm{j}}), & (\mathbf{Y}_{\mathrm{j}}'\mathbf{X}) \\ \mathbf{X}'\mathbf{Y}_{\mathrm{j}}, & \mathbf{X}'\mathbf{X} \end{pmatrix} = \hat{\mathbf{Z}}_{\mathrm{j}}'\hat{\mathbf{Z}}_{\mathrm{j}}$$

We may use this formulation for the variance calculation.

Loosely speaking, if we take the expectation inside of the formula for the variance of $\hat{\delta}_{2SLS}$, then

$$V[\hat{\delta}_{\mathrm{j}}] = \sigma_{\mathrm{jj}}(\hat{\mathbf{Z}}_{\mathrm{j}}'\hat{\mathbf{Z}}_{\mathrm{j}})^{-1} = \sigma_{\mathrm{jj}}[(\mathbf{Z}_{\mathrm{j}}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Z}_{\mathrm{j}})]$$

This argument isn't exactly right, though, because Z_j contains ϵ. The Expected value is hard to calculate because $\hat{\delta}_{2SLS}$ is the ratio functions of random variables. A more complicated mathematical argument, using the Delta Method, shows that the expectation of the ratio is off by a small amount on the order of $1/n$. As n becomes larger, the Variance-Covariance matrix of the Two Stage Least Squares Estimator approaches the above formula.

7.3.2. 2SLS versus OLS.

When is OLS okay? Two-stage least squares must make a sufficiently good improvement in terms of bias that it is worth using this method. I say worth because 2SLS is a less efficient method.

If we compare this with the OLS estimator variance, we see that OLS is more efficient. In the IV estimator we use $\hat{\mathbf{Z}}_{\mathrm{j}}$ instead of $\mathbf{Z}_{\mathrm{j}}$ in OLS. We can see exactly how far off the asymptotic variance estimators are by comparing the *plim* of the matrices $\hat{\mathbf{Z}}_{\mathrm{j}}'\hat{\mathbf{Z}}_{\mathrm{j}}$ and $\mathbf{Z}_{\mathrm{j}}'\mathbf{Z}_{\mathrm{j}}$. As noted in section 7.1 above,

$$plim\frac{1}{n}\mathbf{Z}_{\mathrm{j}}'\mathbf{Z}_{\mathrm{j}} = \mathbf{plim}\frac{\mathbf{1}}{\mathbf{n}}\begin{bmatrix}\mathbf{Y}_{\mathrm{j}}'\\ \mathbf{X}'\end{bmatrix}[\mathbf{Y}_{\mathrm{j}}\mathbf{X}] = \begin{pmatrix}\mathbf{\Pi}'\mathbf{Q}\mathbf{\Pi}+ & ,\mathbf{\Pi}'\mathbf{Q}\\ \mathbf{Q}'\mathbf{\Pi},\mathbf{Q}\end{pmatrix}$$

Direct analysis of $\hat{Z}_j'\hat{Z}_j$ yields

$$plim\frac{1}{n}\hat{\mathbf{Z}}'\hat{\mathbf{Z}} = \mathbf{plim}\frac{\mathbf{1}}{\mathbf{n}}\begin{bmatrix}\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y},\mathbf{Y}'\mathbf{X}\\ \mathbf{X}'\mathbf{Y},\mathbf{X}'\mathbf{X}\end{bmatrix} = \begin{pmatrix}\mathbf{\Pi}'\mathbf{Q}\mathbf{\Pi},\mathbf{\Pi}'\mathbf{Q}\\ \mathbf{Q}'\mathbf{\Pi},\mathbf{Q}\end{pmatrix}$$

The first matrix corresponds to the asymptotic variance-covariance matrix of the OLS regression and the second, the asymptotic variance covariance matrix of the IV. Because the first submatrix (upper left) contains the positive definite matrix - , the inverse of this matrix will be "smaller", than the inverse of the second. That is the difference will be negative definite. They will be equal only if - $= 0$.

We may apply Hausman's specification test in this problem. Compute the difference between the vector of OLS coefficients and 2SLS estimates: $\mathbf{d} = \hat{\delta}_{\mathrm{OLS}} - \hat{\delta}_{\mathrm{2SLS}}$. The Wald statistic is:

$$W = \mathbf{d}(\mathbf{V}[\hat{\delta}_{\mathrm{2SLS}} - \mathbf{V}[\hat{\delta}_{\mathrm{OLS}}])^{-1}\mathbf{d}$$

7.3.3. Validity and Strength of Instruments

It is important to have strong first stage estimates, and for the excluded exogenous variables, as a set, to explain a significant portion of the variance in the endogenous included variables.

When you have weak instruments, say a single excluded variable with a t-statistic in the first stage only slightly above 2, then the instrumentation doesn't given you much leverage, and the model is "barely identified." While we usually discuss identification in absolute terms (identified or not), weak instruments create problems in performing test statistics and usually do not show improvement over OLS, unless the bias in OLS is severe.

One way to test for the strength of the fist stage estimates is to perform an F-test for the significance of the set of excluded exogenous variables. We expect that the p-value will be extremely small (not just barely over the hurdle for significance). It is also good to look at the partial correlation. How much of the explained variation in $\mathbf{Y}_\mathrm{j}$ is accounted for by $\mathbf{X}_\mathrm{j}^*$? You can calculate this by running a partial regression. Purge $\mathbf{Y}_\mathrm{j}$ and $\mathbf{Y}_\mathrm{j}$ of their correlation with the included exogenous variables, $\mathbf{X}_\mathrm{j}$. The regress the new $\mathbf{Y}_\mathrm{j}|\mathbf{X}_\mathrm{j}$ on $\mathbf{X}_\mathrm{j}^*|\mathbf{X}_\mathrm{J}$. The $R - square$ from this is the partial R^2 and tells you whether your first stage equations explain more than a trivial amount of the variation in the included exogenous variables.

The strength of the instrumentation is easy to verify. The validity of the instruments is extremely hard to verify. When there is one excluded exogenous variable for each included endogenous variable, the model is "exactly identified." In this case it is impossible to test for the validity of the instrument. We have used all of the degrees of freedom available just to identify the coefficients.

When the model is overidentified (i.e., there is more than one excluded exogenous variable per included endogenous variable), tests for the validity of the exclusion restriction are possible. Anderson and Rubin (1950) and Hausman (1983) propose tests based on the overidentification restriction. Assuming that at least some of the exclusions are valid, then one may regress the residuals from the second stage on all of the exogenous variables (included and excluded). The statistic nR^2 will follow a χ^2 distribution with degrees of freedom equal to the number of excluded exogenous variables in a given equation minus the number of endogenous included variables in that equation.

This is a useful test, but it does have some limitations. First, the model must be identified by a subset of the excluded variables, though not all need be valid. Second, this test has fairly low power against alternative hypotheses.

EXAMPLE: Endogenous Campaign Spending.

An extensive formal theory literature begins with the following behavioral model. (1) Two competiting campaigns spend money in order to influence voters, who decide on the basis of party, ideology, and affect toward the candidates. (2) Each candidate maximizes the probability of winning, a function of spending, subject to the cost of raising money. The equilibria to these models express Spending and Probabilities of winning in terms of the underlying cost and vote parameters.

EXAMPLE: Strategic Retirement and the Incumbency Advantage.

One may consider the incumbency advantage as consisting of (at least) two behavioral relationships. First, incumbents, based on their sense of the electoral circumstances and on their personal situations, decide whether to stand for reelection. Second, citizens vote based on the qualities of the candidates and their personal preferences and partisan leanings. Suppose there are incumbents (i) in districts (j) observed over time (t). Let I indicate whether an incumbent runs for reelection and of which party: +1 for Democratic incumbent runs, 0 for open, -1 for Republican incumbent runs. Let V denote the vote (or expected vote); X_1 is Age, and X_2 is opportunity. The opportunity to run for reelection is restricted by term limits for some offices in some years. Also, let N_j denote the partisan division of the normal vote, measured as the average division of the presidential vote in the legislative district or state.

$$I_{ijt} = \alpha_{10} + \alpha_{11} V_{ijt} + \alpha_{12} X_{1ijt} + \alpha_{13} X_{2ijt} \epsilon_{1ijt} \tag{5}$$

$$V_{ijt} = \alpha_{20} + \alpha_{21} I_{ijt} + \alpha_{22} N_j + \epsilon_{2ijt} \tag{6}$$

1. How would you estimate each of these equations? Are both equations identified? What exact regressions would you run?

2. How strong is the first stage?

3. Consider the OLS and 2SLS estimates handed out in class? Contrast the estimates (coefficients and standard errors). Was there much improvement in the IV estimates?

4. What did the Hausman test reveal?