

Section 7: Panel Data & Endogenous Regressors

Jeffrey Greenbaum

March 2, 2007

Contents

1	Section Preamble	1
2	Panel Data	2
2.1	Fixed Effects	3
2.2	Random Effects	4
2.3	2004 Exam, 1C	5
3	OLS problems in endogeneity	6
4	Instrumental Variables	9
5	Estimation	11
5.1	Just-Identified Case: IV Estimation	11
5.1.1	Asymptotics for the IV estimator	11

1 Section Preamble

Today we complete our discussion on the generalized regression model and relax the final hypothesis about linear expectation. Specifically, we begin by considering panel data models. Most empirical microeconomics papers estimate panel data models, and panel data models are certainly an active area of research in econometrics. In addition to the sheer importance of becoming familiar with panel data models, random effects models, a class of panel data models, can be seen as a special case of the generalized regression model. Accordingly this model fits well with the theme of relaxing the variance-covariance hypothesis.

Then, we will return to the classical linear model and begin discussing endogenous regressors, the final topic of Professor Powell's part of 240B. The final hypothesis to relax is the linear expectations hypothesis: $E(y|X) = X\beta, \Rightarrow E(\varepsilon) = 0$. This hypothesis implies that $E(X'\varepsilon) = 0$:

$$E(X'\varepsilon) = E(E(X'\varepsilon|X)) = E(X'E(\varepsilon|X)) = E(X'0) = 0$$

As a result, $E(X'\varepsilon) \neq 0 \Rightarrow E(\varepsilon|X) \neq 0$.

Per usual, we ask the two questions associated with relaxing an hypothesis:

1. What happens to the classical model if we relax $E(X'\varepsilon) = 0$? As we will show, β is no longer identified because it cannot be written as a function of population moments. Moreover, $\hat{\beta}_{OLS}$ is no longer unbiased nor consistent. As with generalized least squares, an inconsistent estimator is incredibly problematic because we want to get closer to the true parameter if we collect more data. Clive Granger, a Nobel Laureate econometrician, once remarked, "If you can't get it right as n goes to infinity, you shouldn't be in this business."

2. How can we solve this problem?

We can need to find an instrument. An appropriate instrumental variable will allow us to identify β and construct an estimator that is unbiased, consistent, and asymptotically normal. We conclude that we have a good instrument if our instrument, Z , is [highly] correlated with X and uncorrelated with ε . For identification, we require that Z contains at least as many variables as we seeks to instrument in X . In some models, we can deduce a valid instrument from our data. However, in most applications, it is necessary to collect more data to find an instrument. As is seen in the empirical literature, an economist must often demonstrate that $Cov(Z, \varepsilon) = 0$ by showing that the instrument is not correlated with hypothetical components of the error term.

2 Panel Data

Panel data models are those in which we have data about a set of individuals over a set of time periods. We say that the panel is balanced if there is data about the same group of individuals for each time period in the sample. Although this may seem to resemble SUR in the case that we are tracing a group of individuals over time, in panel data, we first stack over time period for each individual and then over individuals.

The general framework for our panel data models is:

$$y_{it} = x'_{it}\beta + \alpha_i + \varepsilon_{it}$$

where we assume $E(\varepsilon_{it}|X) = E(\varepsilon_{it}) = 0$, $Var(\varepsilon_{it}) = \sigma_\varepsilon^2$ and $Cov(\varepsilon_{it}, \varepsilon_{js}) = 0$ if $i \neq j$. i denotes the individual where we have N individual, and t denotes time where we have T time periods. Stacking observations for each individual over time and then across individuals yields:

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

where y is a $NT \times 1$ vector, X is a $NT \times k$ matrix, D is a $NT \times N$ matrix with T $N \times N$ vertically stacked identity matrices. As Professor Powell proved in lecture, X does not include an intercept because if it did, $[X, D]$ would not be full column rank.

2.1 Fixed Effects

α_i is our individual-level fixed effects, which are unobservable characteristics that are time invariant for each individual. Accordingly, our error term, ε_{it} , includes all individual-year shocks if t measures year, in addition to individual-invariant shocks for each year. Moreover, we allow for an arbitrary relationship between α_i and x to exist. Note that we could also include time fixed effects or only time fixed effects. The only requirement is that we must leave some shocks in the error term, so including both individual and year fixed effects leaves the individual-time shocks. These shocks are often not accounted for because it is more sensible to motivate the individual or year fixed effects.

Hypothetically, we can say that $\alpha_i = z_i^* \delta$ where z_i^* are the collection of unobserved variables. We do not necessarily care about δ or in fact know all of the variables that belong in z , but we want our estimator to account for these characteristics. For example, in the returns to education regression, an individual fixed effect would include characteristics such as ability, motivation, and beauty. By unobserved, we mean that they are unobserved to the econometrician, or in other words, we do not have reliable data to measure these relevant variables.

The *fixed effects* (FE) or *within* (W) or *least squares dummy variable* (DV) estimator for β can be obtained by partitioned regression, which comes from the Frish-Waugh Theorem. As a refresher on partitioned regression, see Professor Powell's first set of lecture notes. It is useful when the econometrician wants to calculate point estimates for a set of variables in the design matrix. However, the econometrician is not directly interested the remaining set of variables. In our application, the second set of variables are the fixed effects that are relevant for properly specifying the model but not are directly meaningful because we do not observe any of them. Accordingly, applying the expression of the Frish-Waugh Theorem,

$$\hat{\beta}_{FE} = (\tilde{X}' \tilde{X})^{-1} \tilde{X}' \tilde{y}$$

where $\tilde{X} = (I_{NT} - D(D'D)^{-1}D')X$ and $\tilde{y} = (I_{NT} - D(D'D)^{-1}D')y$ which are the residuals of the regression of X on D and Y on D respectively.

Note that \tilde{X}' is,

$$\begin{pmatrix} X_1 - l_T(T^{-1} \sum_{t=1}^T x_{1t}) \\ \cdot \\ \cdot \\ X_N - l_T(T^{-1} \sum_{t=1}^T x_{Nt}) \end{pmatrix} = \begin{pmatrix} X_1 - l_T x_{1.} \\ \cdot \\ \cdot \\ X_N - l_T x_{N.} \end{pmatrix}$$

Writing these expressions in summation notation yields:

$$\hat{\beta}_{DV} = \hat{\beta}_{FE} = \hat{\beta}_W = \left[\sum_{i=1}^N \sum_{t=1}^T (x_{it} - x_i)(x_{it} - x_i) \right]^{-1} \sum_{i=1}^N \sum_{t=1}^T (x_{it} - x_i)(y_{it} - y_i)$$

As Professor Powell presented in lecture, these two equivalent estimators are manipulations such at the individual fixed effects drop from the regression.

Note that the difference-in-differences framework can be viewed as a special case of the fixed effects model. In the baseline case, we have two groups, control and treatment, and two time periods of data, pre-treatment and post-treatment. We allow for there to be individual and time fixed effects. We take first-differences and then run the regression. In doing so, note that individual fixed effects drop because they are constant for all individuals in both periods. Also note that with only one control, the presence of being in the treatment group, this variable reduces to 0 for the control, 1 for treatment. The least squares estimator that comes from this framework is the difference between treatment and control of the difference in y over each time period for both groups.

2.2 Random Effects

Notice that the fixed effects model fails to identify any components of β that correspond to regressors that constant over time for a given individual. For this model to yield a consistent estimator, α_i must be uncorrelated with x_{it} . Accordingly, treating the α 's as random variables, we assume the following in a random effects model:

- $y_{it} = x'_{it}\beta + \alpha_i + \epsilon_i$
- α_i is independent of ϵ_{it}
- α_i is independent of x_{it} and
- $E(\alpha_i) = \alpha, Var(\alpha_i) = \sigma_\alpha^2, Cov(\alpha_i, \alpha_j) = 0$ if $i \neq j$.

We can then rewrite the model as:

$$\begin{aligned} y_{it} &= x'_{it}\beta + \alpha_i + \epsilon_i \\ &= x'_{it}\beta + \alpha + u_{it} \end{aligned}$$

where $u_{it} = \epsilon_{it} + (\alpha_i - \alpha)$ and $E(u_{it}) = 0, Var(u_{it}) = \sigma_\epsilon^2 + \sigma_\alpha^2, Cov(u_{it}, u_{js}) = 0$ if $i \neq j$, and $Cov(u_{it}, u_{is}) = \sigma_\alpha^2$.

Stacking the model we have,

$$y = X\beta + \alpha l_{NT} + u$$

which produces a non-spherical variance-covariance matrix for each individual:

$$Var(u_i) = \begin{pmatrix} \sigma_\epsilon^2 + \sigma_\alpha^2 & \sigma_\alpha^2 & \dots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\epsilon^2 + \sigma_\alpha^2 & \dots & \sigma_\alpha^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \vdots & \dots & \sigma_\epsilon^2 + \sigma_\alpha^2 \end{pmatrix}_{T \times T}$$

and

$$Var(u) = \sigma_\epsilon^2 I_{NT} + \sigma_\alpha^2 (I_N \otimes l_T l_T')$$

The least squares estimate of the RE model can be found using Frisch-Waugh theorem again:

$$\hat{\beta}_{LS} = (X^*{}' X^*)^{-1} X^*{}' y^*$$

where $X^* = (I_{NT} - l_{NT}(l'_{NT}l_{NT})^{-1}l'_{NT})X$ and $y^* = (I_{NT} - l_{NT}(l'_{NT}l_{NT})^{-1}l'_{NT})y$ which are the residuals of the regression of X on l_{NT} and Y on l_{NT} respectively.

Expanding this estimator gives the the following representation in summation:

$$\hat{\beta}_{LS} = \left[\sum_{i=1}^N \sum_{t=1}^T (x_{it} - x_{..})(x_{it} - x_{..})' \right]^{-1} \sum_{i=1}^N \sum_{t=1}^T (x_{it} - x_{..})(y_{it} - y_{..})$$

where $x_{..}$ is the *grand mean*, i.e. the average of x_{it} over i and t . This estimator is unbiased and consistent but inefficient though.

We know that GLS is efficient relative OLS. We call it the GLS Random Effects Estimator, which is given by:

$$(\hat{\beta}_{GLS}, \hat{\alpha}_{GLS})' = (Z'\Omega^{-1}(\theta)Z)^{-1}Z'\Omega^{-1}(\theta)y$$

where $X = [l_{NT}X]$, $\Omega(\theta) = I_{NT} + \theta(I_N \otimes l_T l_T')$ and $\theta = \sigma_\alpha^2 / \sigma_\epsilon^2$

It can be shown that the GLS or RE estimator is a matrix-weighted average between the *within* and the *between* groups estimators:

$$\hat{\beta}_{RE} = A(w_0)\hat{\beta}_{FE} + [I_K - A(w_0)]\hat{\beta}_B$$

where $\hat{\beta}_B$ is the between estimator that captures variation only between groups since there is none within groups:

$$\hat{\beta}_B = \left[\sum_{i=1}^N (x_{i.} - x_{..})(x_{i.} - x_{..})' \right]^{-1} \sum_{i=1}^N (x_{i.} - x_{..})(y_{i.} - y_{..})$$

As $T \rightarrow \infty$ and N is fixed, it can be proved that $A(w_0) \rightarrow I_K$, hence FE and RE are asymptotically equivalent. See section 24.9 for more detail.

It should be clear that we have the usual problems with hypothesis testing. Fixed effects models can be relaxed so that they are written with variance-covariance matrices that take a pure heteroskedastic form. In that case, we would want to use heteroskedastic-robust consistent standard errors. Similarly, if we do not know the elements of the variance-covariance matrix for random effects, then we must construct a feasible estimator. Professor Powell presented a feasible estimator in his lecture.

2.3 2004 Exam, 1C

Note that this question is the only test question ever asked on panel data. In fact, Professor Powell acknowledges that "this is a tricky problem" and that he initially had an incorrect answer in mind when making up the question.

Question: For a balanced panel data regression model with random individual effects, $y_{it} = x'_{it}\beta + \alpha_i + \varepsilon_{it}$ (where the α_i are independent of ε_{it} , and all error terms have mean zero, constant variance,

and are serially independent across i and t), suppose that only the number of time periods T tends to infinity, while the number of individuals N stays fixed. The the "fixed effect" estimator for β will be consistent as $T \rightarrow \infty$, but the "random effects" GLS estimator is infeasible, since the joint covariance matrix of the error terms is not consistently estimable.

Answer: False. It is true that the joint covariance matrix of the error terms is not consistently estimable - specifically σ_α^2 isn't estimable because there are only N realizations of α_i available in the sample, and N is fixed - but this does not mean GLS is either "infeasible" or inconsistent. It is also true that the "fixed effect" estimator β_{FE} is consistent; as in Ruud's text, the FGLS estimator can be written as a matrix-weighted average of the fixed-effect and "between" estimators, where the latter is inconsistent (being based upon only N time averages). However, inspection of the weight matrices for the FGLS estimator reveals that the weight on the "between" estimator goes to zero, and the corresponding weight on "fixed effects" goes to the identity matrix, as $T \rightarrow \infty$. Moreover, it can be shown that FGLS and "fixed effects" are asymptotically equivalent,

$$\sqrt{T}(\hat{\beta}_{FGLS} - \hat{\beta}_{RE}) \rightarrow_p 0$$

under the usual conditions on the regressors, etc. So, at least as $T \rightarrow \infty$, FGLS behaves just like the fixed effect estimator for β , and is consistent.

3 OLS problems in endogeneity

As previously mentioned, without the linear expectations hypothesis, β is no longer identified. Moreover, $\hat{\beta}_{OLS}$ is no longer unbiased nor consistent. We now show these properties:

- Identification:

$$\begin{aligned} y &= x'\beta + \varepsilon \\ \Rightarrow xy &= xx'\beta + x\varepsilon \\ \Rightarrow E(xy) &= E(xx')\beta + E(x\varepsilon) \\ \Rightarrow \beta &= E(xx')^{-1}E(xy) - E(xx')^{-1}E(x\varepsilon) \end{aligned}$$

Thus, β is no longer identified because it cannot be written as a function of population moments since the ε are not observed.

- Bias:

$$E(\hat{\beta}|X) = E((X'X)^{-1}X'y|X) = (X'X)^{-1}X'E(y|X) = \beta + (X'X)^{-1}X'E(\varepsilon|X)$$

and from what we saw in the Section Preamble, $E(\varepsilon|X) \neq 0$. As a result, we should expect a bias in the estimation by OLS in general.

- Inconsistency: even worse,

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

so $\hat{\beta} \rightarrow_p \beta + \left(\text{plim} \frac{X'X}{n}\right)^{-1} \text{plim} \frac{X'\varepsilon}{n}$. Using the same method as used to show the classical least squares estimator converges to β , we can show that $\text{plim} \frac{X'\varepsilon}{n} = E(x\varepsilon)$, which is now not necessarily zero. As a result, in general the OLS estimator does not converge to β .

In social sciences research, you should often be suspicious of whether this assumption is satisfied. We can only be certain that we have avoided this problem in a laboratory experiment where we can isolate the experiment from all external influence. The scientist manipulates the control and treatment group and controls all conditions of the experimental environment. In this case, we could randomly assign subjects to either the control or treatment group, do the treatment to the one group and not to the other, and then assess the differences. If all the other factors are the same for both groups, then any difference in the outcomes of the two groups is due to the effect of the treatment.

For instance, a biologist can put two groups of bacteria in the same environment conditions, and change, let's say, oxygen levels in one of the groups. If more bacteria grow in the group with more oxygen, then we can conclude that a certain amount of oxygen causes a certain amount of growth. By repeating the same experiment many times, we could be sure of the results because of consistency.

The problem in the social sciences literature is that you cannot ethically do this to human beings. For starters, the research cannot involve humans in an experiment without telling them so. The research cannot tell a subject about the control and treatment, but the subject must agree to not know. Nonetheless, the subject may suspect what the control and treatment groups are and modify people's actions. Known as the Hawthorne Effect, laboratory experiments would not be accurate if subjects adjusted their behavior. At least in developed countries, citizens and funding agencies often do not always like the idea of randomly helping some individuals and not other individuals.

A physician can get really close to experimental results in a double blind experiment. If the treatment, let's say pills with medicine, instead of placebo, is assigned randomly, and neither the patients, nor the physician know who is taking what, then we can trust that the results are on average accurate. Of course people are not the same, do not live in the same environment. However, since they were randomly assigned to each group, we can expect the groups to be the same on average if we have a sufficiently large sample, with the only difference being the medicine intake.

If we performed a regression:

$$y_i = x_i\beta + \varepsilon_i$$

where x is a dummy for having taken the medicine, instead of the placebo, and y being a certain health exam result, then if everybody took whatever it is that they were supposed to take, we can trust that β really is the effect of the treatment in the health exam result. If of course we think that some observable characteristics also affect the outcome, then these should be explicitly controlled for.

In social sciences, the treatment is seldom random. The classic example is Josh Angrist's AER (1990) paper on the long-term affect on one's income of having served in the Vietnam War. In this case, the treatment is going to war while the control is not going, but of course, the assignment is not random. If we find that people that went to war are poorer in the future, it can be because of war, but it can also be because the people who went to war were less prepared to the labor market in the first place. Why might they have enlisted? Some of them could have done so because they could not find another job, or because the payment in the army is better than what they would get otherwise. Also, there is considerable government support for veterans, which influences, and possibly the decision to enlist.

Consider the regression:

$$y_i = x_i\beta + \varepsilon_i$$

where y_i is income, x_i is a binary variable of having gone to war.

Is x_i uncorrelated with ε_i ? Hardly, because there are plenty of things that are correlated with earnings that are also correlated with the decision of enlisting. Notice that the error is everything that affects earnings (y_i) other than having served in the war (x_i). A person's earnings are of course not only determined by having been in a war. They are determined by other things such as education, ability, experience, personal networks, appearance, race, and so forth. All of these things appear in the residuals of the earnings after we took out the effect of war. Now, are these correlated with x_i , having served in war? Quite likely. For instance, a less educated person will earn less in average than a more educated one, and also a less educated person is more likely to enlist, because the army is a good job that pays better than working in say pizza delivery and earns more respect. In this case, $\hat{\beta}_{OLS}$ will not only reflect the exact effect of going to war in earnings, but also the effect of education in earnings **through** going to war as seen by the identifiability math.

However, we could control for education, health, parents' education, work experience, and all other relevant observable characteristics that are relevant to both income and serving in the war. The regression would be:

$$y_i = \alpha_0 + \alpha_1 w_{1i} + \dots + \alpha_p w_{pi} + \beta x_i + \varepsilon_i$$

where the each w is one of the control variables. If we do OLS, by partitioned regression, $\hat{\beta}_{OLS}$ would be the effect of going to war on income, controlling for the effect of each w . However, some variable could still be missing from the data set, or worse, unobservable. That is, there could still be something in individuals that enlist that makes them earn systematically less in the future. It could be just a psychological characteristic. We would still have endogeneity, and we wouldn't be able to solve it even if we had all of the information that can be collected. How can we trust the results?

Professor Powell will discuss some classical situations that present endogeneity, and some of them have a standard solution. Some of them, however, do not, and require a creative solution:

1. Lagged dependent variable and Serial correlation
2. Omitted Variables
3. Measurement error
4. Simultaneous Equations

All of the solutions require finding an instrumental variable, and performing what we call and Instrumental Variable Regression (IV), a Two Stage Least Squares Regression (2SLS), or a Generalized Method of Moments regression (GMM) as appropriate.

4 Instrumental Variables

The solution to such problem is change the way we estimate β so that first it is identifiable. For this we will need to have an "instrument", a variable which we will refer to as z . Observe that we will later talk about the relevance of the dimension of z .

We need to satisfy two conditions about z to consider it a valid instrument:

1. z must be uncorrelated with ε : $E(z\varepsilon) = 0$.
2. z must be correlated to x , and preferably, this correlation will be as high as possible: $E(zx) \neq 0$. This means that if we regressed x on z , and any other control variable:

$$x_i = \alpha_0 + \alpha_1 w_{1i} + \dots + \alpha_p w_{pi} + \gamma z_i + v_i \quad (1)$$

where the coefficient γ will be different from 0. In this regression, we assume the classical linear assumptions wherein $E(v_i) = 0$, $Var(v_i) = \sigma_v^2$, and full row rank regressors.

If z satisfies these conditions, then it will solve the identification problem. That is, we can express β as a function of population moments:

$$\begin{aligned} y &= x'\beta + \varepsilon \\ zy &= zx'\beta + z\varepsilon \\ E(zy) &= E(zx')\beta + E(z\varepsilon) \\ \Rightarrow E(zx')\beta &= E(zy) \end{aligned} \quad (2)$$

which entirely defines β as a function of population moments.

However, finding an instrument is not usually easy. Omitted variables, for example, requires the researcher to be creative. We cannot always mathematically deduce or confirm an instrument because we do not observe the error terms. Often in such papers, the contribution is the instrumental

variable strategy. A significant part of the paper is devoted to motivating the instrument, and then defending that it is not correlated with possible components of the error term. The latter is known as the robustness checks. If someone has any doubt about the lack of correlation then the instrument is no longer valid. Moreover, we want an instrument that is not weakly correlated with regressors though this condition is straightforward to check from a least squares regression.

Let's return to Angrist (1990), about the effects of war on future earnings. In the Vietnam era, there was a moment when the government needed people to enlist, and voluntary enlistment was not covering the demand for soldiers. The government instituted compulsory enlistment, which at that time meant going to war for sure, provided one passed certain physical and mental requirements. All of the men in a certain age range were randomly given a number. Then the army would call starting from the lowest number and going up. Observe that the number was absolutely random, so a person being called did not depend on any social, psychological or economic characteristic, in such a way that those who stayed were on average similar to the ones who ended up going to war¹.

Angrist exploits the person's number in the lottery draft as an instrument. Let's see if it satisfies the requirements of an instrument.

- $E(z_i \varepsilon_i) = 0$? The lottery number is randomly assigned, so it must be independent of any unobservable that could influence earnings. At least, we will assume this to be the case for all practical purposes.
- $E(z_i x'_i) \neq 0$? Although the correlation is not perfect, there is a huge increase in the probability of going to war if one's lottery number is low, so x and z are definitely correlated.

Instrumental variables has provoked very interesting discussions amongst empirical researchers, at least in labor economics and economics history, as well as in econometrics. If the instrumental variable is not valid, usually, $E(Z'X) \neq 0$ then the researcher has not solved the problems that arise with OLS without the linear expectations hypothesis, that is the lack of identification, consistency, and unbiasedness. In other words, the researcher does not have a reliable estimate for β . Angrist's paper is often cited as the most convincing application of the instrumental variables strategy because there is no situation in which the outcomes are randomized than a randomized lottery. Nevertheless, some are not convinced that $E(z_i x'_i) = 0$. Without going into too many of the details, some economists are more open to the use of instruments that require creativity. In this circle, there have been some strong debates whether particularly influential instrumental variable contributions were valid. Even in papers that have been published in top journals, the choice and definition of some instruments continue to be doubted. And even if the instrument is valid,

¹I exaggerate to make the point. There were some problems with the lottery, and people's decisions to enlist, since those who had low numbers knew that, and tended to enlist voluntarily before being called. However, even if the lottery number did not absolutely determine one's going or not to war, a lower number certainly meant a higher probability of going to war in average. Angrist discusses all of this in the paper, but here we just want to understand the nature of an instrument, so I'm taking some permissions.

inference may not be as convincing if the instrument is considered to be weak, which has been an area of research amongst econometricians. Although this discussion may sound cynical, it is too emphasize the difficulty in asserting causality in empirical economic research. At least in labor, some alternative research designs, such as regression discontinuity design, have been developed and recently somewhat widely applied, to assert causality at least locally in a population.

Though you are not responsible for any of the papers that use instrumental variables or understanding its role in the profession, hopefully it is clear how prevalent endogeneity is, how serious it is, and how difficult it is to convincingly address. For these reasons, it is a highly important topic, and as we will shortly see, quite a bit of math that we can do that is within the scope of the course to understand solutions to the problem. It should thus not be surprising that instrumental variables has appeared on every recent exam.

5 Estimation

5.1 Just-Identified Case: IV Estimation

We say that are in the just-identified case if the dimension of z is the same as the dimension of x so that we have as many instrumental variables as we have endogenous regressors. That is, $rank(E(z_i x_i')) = K$, Equation (2) transforms to:

$$\beta = E(zx')^{-1}E(zy)$$

and we start off with estimating it by using the sample moments:

$$\hat{\beta}_{IV} = \left[\frac{1}{n} \sum_{i=1}^n z_i x_i' \right]^{-1} \frac{1}{n} \sum_{i=1}^n z_i y_i = (Z'X)^{-1} Z'Y$$

5.1.1 Asymptotics for the IV estimator

We now return to our asymptotic calculations and modify them for our new estimator. We want to demonstrate that our estimator is consistent that we can use hypothesis testing, at least in large samples. These two reasons are the main motivation for why asymptotics is so important to this course. We are interested in these two asymptotic properties for each estimator we have discussed in Econ 240B, although for some the math is beyond the scope of the course.

We start by computing $\hat{\beta} - \beta$:

$$\hat{\beta}_{IV} = (Z'X)^{-1} Z'(X\beta + \varepsilon) = \beta + (Z'X)^{-1} Z'\varepsilon$$

As a result,

$$\hat{\beta}_{IV} - \beta = (Z'X)^{-1} Z'\varepsilon = \left(\frac{1}{n} \sum_{i=1}^n z_i x_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^n z_i \varepsilon_i$$

As usual, we treat each part separately so that we can ultimately apply Slutsky's Theorem:

$$\overline{\frac{1}{n} \sum_{i=1}^n z_i x_i}:$$

Since we still have random sampling, the x_i and z_i are iid, and thus so is $z_i x_i$. Also, assuming that the second moments are finite, the weak law of large number states that:

$$\frac{1}{n} \sum_{i=1}^n z_i x_i \longrightarrow_p E(zx')$$

$$\overline{\frac{1}{n} \sum_{i=1}^n z_i \varepsilon_i}:$$

Since the ε_i are iid², so are the $z_i \varepsilon_i$. Also, assuming $E(z_i^2 \varepsilon_i^2)$ finite³, we can apply the weak law of large numbers and the central limit theorem to obtain:

$$\frac{1}{n} \sum_{i=1}^n z_i \varepsilon_i \longrightarrow_p E(z\varepsilon) = 0$$

and

$$\sqrt{n} \frac{1}{n} \sum_{i=1}^n z_i x_i \longrightarrow_d N(0, E(z^2 \varepsilon^2))$$

Convergence in probability implies convergence in distribution. Apply that to $\frac{1}{n} \sum_{i=1}^n z_i \varepsilon_i \longrightarrow_p E(z\varepsilon) = 0$. As a result of the Continuous Mapping Theorem and then Slutsky's Theorem,

$$\left(\frac{1}{n} \sum_{i=1}^n z_i x_i \right)^{-1} \frac{1}{n} \sum_{i=1}^n z_i \varepsilon_i \longrightarrow_p 0 \Rightarrow \hat{\beta}_{IV} \longrightarrow_p \beta$$

and

$$\sqrt{n}(\hat{\beta} - \beta) \longrightarrow_d N(0, E(zx')^{-1} E(z^2 \varepsilon^2) E(xz')^{-1})$$

Thus a pivotal statistic is for hypothesis testing:

$$\frac{\sqrt{n}(\hat{\beta} - \beta)}{\left(\left(\frac{1}{n} \sum_{i=1}^n z_i x_i' \right)^{-1} \hat{V} \left(\frac{1}{n} \sum_{i=1}^n x_i z_i' \right)^{-1} \right)^{-1/2}} \longrightarrow_d N(0, I_K)$$

where \hat{V} is a consistent estimator of the $plim \left(\frac{Z' \Omega Z}{n} \right)$, and $\Omega = E(\varepsilon \varepsilon')$, so the middle term is: $s^2 I_n$ in the homoskedasticity case, Eicker-White ($Diag(y_i - x_i \hat{\beta}_T V)$) with heteroskedasticity, and Newey-West in the serial correlation case.

Next week we will discuss the over-identified case and GMM estimation.

²Observe that we don't have heteroskedasticity here. We're only dropping the non-endogeneity hypothesis, nothing else.

³We only assumed the z_i and the ε_i to be uncorrelated. If we want to go further and assume independence, then $E(z_i^2 \varepsilon_i^2) = E(z_i^2) E(\varepsilon_i^2) = \sigma^2 E(z_i^2)$