Panel data analysis: non-continuous variables, spatial modeling, and instrumental variables
Day 5.

David Pupovac

So far…

- **Fixed effects**

$$y_{i,t} = \alpha_i + x_{i,t}\beta + \varepsilon_{i,t}$$

- **Random effects**

$$y_{i,t} = \alpha + x_{i,t}\beta + \nu_i + \varepsilon_{i,t}$$

Swamy-Arora estimator, Hausman-Taylor estimator, random coefficients estimator

- **Dynamics**
- We distinguished between estimators for:

TSCS data:

in case of stationary variables:

$y_{it} = \beta_1 x_{it} + \beta_2 x_{it-1} + \nu_{it}$ finite distributed lag model

$y_{it} = \beta x_{it} + \phi y_{it-1} + \nu_{it}$ lagged dependent variable model

$y_{it} = \beta x_{it} + \theta y_{it-1} + \gamma x_{it-1} + \nu_{it}$ autoregressive distributed lag model

in case of non-stationary variables:

Use first differenced variables in regression or error correction model

CSTS data:

Arellano-Bond estimator and Arellano-Bover/Blundell-Bond

What will we do today?

- Non-continuous dependent variables
- Spatial modeling (a little bit)
- Instrumental variables

TSCS/CSTS data and non-continuous dependent variables

- Up until now, we assumed that our response (dependent) variable is continuous

- But what can we do if our dependent variable is not continuous

- The common cases are:
  - Binary: $y \in \{0,1\}$
  - Multinomial: $y \in \{0,1,2,3,4 \ldots \mathrm{k}\}$
  - Count: $y \in \{0,1,2,3,4 \ldots .\}$
  - Censored: (for instance) $y \in \{y^*: y^* \geq 0\}$

- In most cases, we are dealing with these response variables via framework of generalized linear model

Generalized linear model

- Generalized linear models employ a “link function” which defines the relationship between the systematic component of the data and the outcome (dependent)

- The basic philosophy is to employ a function to link the normal theory environment with Gauss-Markov assumptions, to another environment with wide class of outcome variables

- The theory of the generalized linear model is based upon the **exponential family of distributions**

- Fisher (1934) develop idea that PDF/PMFs are special cases of more a general exponential family. The exponential family refers to a method in which all the terms in the expression for PDF/PMF are moved into the exponent. So, parameterization specifications are restricted to cases that can be transformed to this exponential family form.

(PDF/PMF: probability density function/probability mass function)

Canonical form and canonical link

- The canonical link is used to generalize the linear model by connecting the linear additive component to the non-normal outcome variable

Distribution		*Canonical Link:* $\theta = g(\mu)$
Poisson		$\log(\mu)$
Binomial	*logit link:*	$\log\left(\frac{\mu}{1-\mu}\right)$
	probit link:	$\Phi^{-1}(\mu)$
	clog log link:	$\log(-\log(1-\mu))$
Normal		μ
Gamma		$-\frac{1}{\mu}$
Negative binomial		$\log(1-\mu)$

Example: logistic regression

- Rather than modeling response Y directly, logistic regression models the probability that Y belongs to a particular category $p(Y) = Pr(Y = 1|X)$

$$p(Y) = \frac{e^{\beta_0+\beta_1 X}}{1 + e^{\beta_0+\beta_1 X}}$$

- After a bit of manipulation, we find that:

$$\frac{p(Y)}{1 - p(Y)} = e^{\beta_0+\beta_1 X}$$

- The quantity $p(Y)/[1 - p(Y)]$ is called the odds, and can take on any value between 0 and ∞. Values of the odds close to 0 and ∞ indicate very low and very high probabilities. By taking the logarithm of both sides, we arrive at

$$log\left(\frac{p(Y)}{1 - p(Y)}\right) = \beta_0 + \beta_1 X$$

- The left-hand side is called the log-odds or logit (so, interpretation is that increasing X by one unit changes the log odds by β_1)

Incidental parameter problem

- Panel analysis methods can be extended to non-continuous variables.

- However, keep in mind that distinction between fixed and random effect as well as TSCS and CSTS data will have consequences in this respect.

- Fixed effects are subject to Neyman and Scott (1948) incidental parameter problem, where the number of parameters in the model is unbounded as N tends to infinity.

- **In other words, you cannot use fixed effects on tobit, logit, probit and other non-linear models because of incidental parameter problem.**

Incidental parameter problem (cont.)

- **Namely, in the case of fixed effects, as the number of observations (clusters) becomes large, estimators fail to converge on consistent estimators**.

- On the other hand, the parameters may be biased and standard errors may be incorrect.

- **The incidental parameters problem only occurs if the lengths of the panel is small or fixed. It is not a problem if T of the panel increases with the sample size N**

Incidental parameter problem and logit

- The incidental parameter problem can be addressed by conditional maximum likelihood estimator

  (Stata's ***xtlogit, fe*** (or ***clogit***) estimates the maximum likelihood conditional on the sum of the outcomes so that the incidental parameter problem is taken care of)

- However, besides the case of binary dependent variable, in most cases you will not be able to estimate fixed effects model.

- Let’s now briefly discuss some other types of dependent variables:

Methods for dealing with non-continuous dependent variables

- Complementary log-log model (***cloglog***) is used for binary dependent variables when outcome is a rare event

- Probit model (***probit***) is a popular specification for both ordinal or a binary response variables. It is assumed that the underlying latent variable model is normally distributed. The model is fairly interchangeable with logit especially in case of larger N

- Poisson model (***poisson***) is used for count dependent variables

- Tobit model (***tobit***) is a special case of a more general model incorporating what is called sample selection. In these models there is a second equation, called the selection equation, which determines whether an observation makes it into the sample- *tobit model = probit + truncated regression*. If no observations are censored, the Tobit model is the same as an OLS regression

Censored dependent and logistic transformation

- One way to deal with censored dependent variable is to transform it. If your variable is limited $y \geq 0$, logarithmic transformation will transform it into a variable $-\infty < ln(y) < +\infty$.

$$ln(y) = \beta_0 + \beta_1 ln(x_1) + \varepsilon$$

- You can interpret β_1 as a percentage increase in dependent that is associated with one percentage increase in independent variable (e.g. if $\beta_1 = 0.33$ the increase is 33%). If x_1 is not log transformed you can interpret β_1 as a percentage increase in dependent given a unit change in independent variable and vice versa.

- Furthermore, if we take exponent we get:

$$y = e^{\beta_0 + \beta_1 \log x_1 + \beta_2 \log x_2}$$

- Given that: $e^{\beta_1 \log x_1} = e^{\log x_1^{\beta_1}}$

$$y = e^{\beta_0 + \log x_1^{\beta_1} + \log x_2^{\beta_2}}$$

- Given that: $e^{\log x_1^{\beta_1}} = x_1^{\beta_1}$

$$\boldsymbol{y = e^{\beta_0} \times x_1^{\beta_1} \times x_2^{\beta_2}}$$

- Therefore β_1 and β_2 reflect the non-linearity of the model.

Dynamics and non-continuous dependent variables

- What we talked about in the previous class applies to binary dependent variables and does not generalize to more complicated limited dependent variables .

- The norm is to ignore all time series issues and just do logit or probit.

(for more discussion on binary dependent variable and dynamics read Beck, Katz, Tucker (1998) Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable)

Spatial modeling

- What is spatial autocorrelation: **an expectation that effects of neighboring states would spill over into each other, creating a sort of correlation across space**. Spatial econometricians model two types of spatial dependence.

  – Frist, the errors of nearby units may be correlated with the degree of correlation being a function of nearness (Beck (2001) considers this a bit strange idea as it implies that there will be no effect of unit 1 on unit 2 if variable is measures, but in the case of unmeasured variable - now in error term - i will have an effect of unit 1 on unit 2).

  – Second, spatial dependence may be addressed by a spatial lag. Spatial lag is weighted sum of the dependent variable of all other units with weight being proportional to the nearness. In cross-sectional data this may be very problematic, but in TSCS this is not such a problem.

Spatial autocorelation and dynamics

- Anselin (1988) identifies several possible extensions of the spatial model to dynamic regressions:
  - A “time-space recursive model” specifies dependence that is purely autoregressive with respect to neighbors in the previous period
  - A “time-space simultaneous” model specifies that the spatial dependence is with respect to neighbors in the current period
  - A “time-space dynamic model” specifies that autoregression depends on neighbors in both the current and the last period

What will we do today?

- We already spoke about instrumental variable estimator
  - Hausman-Taylor estimator (correlation of errors and independents)
  - Anderson-Hsiao estimator (dynamics and panel data models)
- This topic is assuming many statistical concepts and methods, such as:
  - simultaneous equations,
  - structural equation modeling,
  - problem of identification,
  - types of restrictions
  - types of estimators…
- There is no possibility to cover all these topics today, although we will have a comprehensive introduction to this field.

OLS assumptions

- Correct model specification (linear relation, no omitted variables, no irrelevant variables, no outliers, normality)
- Types of variables: all independent variables are quantitative or dichotomous and the dependent variable is quantitative, continuous, and unbounded
- All independent variables have nonzero variance
- No multicollinearity
- Constant error variance (spherical disturbances)

$$E[\varepsilon^2|X] = \sigma^2 \ \forall \ j$$
$$E[\varepsilon_i \varepsilon_j] = 0 \ \forall \ i \neq j$$

- Strict exogeneity

$$E[\varepsilon] = 0$$
$$E[\varepsilon|X] = 0 \ \forall \ i, j$$
$$E\left[\varepsilon_i x_{jk}\right] = 0 \ \forall \ i, j$$

Exogeneity – basics

- Up to now we almost always assumed (although not explicitly) that independent variables are **exogenous (i.e. not endogenous)**.

- What does it mean if variables are exogenous?

- **The exogenous variables are those that are determined outside the system while those that are determined inside the system are endogenous variables.**

- **An exogenous change is one that comes from outside the model and is unexplained by the model**

Exogeneity $E[\varepsilon_i x_{kj}] = 0$

Exogeneity means that mean of error is zero

$$E[\varepsilon_i | x_i] = 0 \ \ \forall \ i$$

- **Weak exogeneity** - the errors should have conditional mean zero (contemporaneous exogeneity)

$$E[\varepsilon_t | x_s] = 0 \ \ \ \forall \ s$$

- **Strict exogeneity** - the errors should have mean zero for all x, that is even in case if $t \neq s$

- In general, the assumption that the errors have mean zero implies that the regressors are orthogonal to the error term for all observations:

$$E[\varepsilon_i x_{kj}] = 0 \ \forall \ i, j$$

- **In other words, the problem of endogeneity arises from *correlation* of regressors (independent variables) with the errors.**

Most important consequence of endogeneity is bias of OLS estimates

- However, the problematic aspect of endogeneity is that it is a conceptual problem. What does this mean?

OLS and endogeneity

- The property of an OLS estimator is $X'\epsilon = 0$. For every column x_k of X, $x_k^T \epsilon = 0$. In other words, each regressor has zero sample correlation with the residuals, ϵ.

$$\begin{bmatrix} X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ X_{k1} & X_{k2} & \dots & X_{kn} \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ \vdots \\ e_n \end{bmatrix} = \begin{bmatrix} X_{11} \times e_1 + X_{12} \times e_2 + \dots + X_{1n} \times e_n \\ X_{21} \times e_1 + X_{22} \times e_2 + \dots + X_{2n} \times e_n \\ \vdots \\ \vdots \\ X_{k1} \times e_1 + X_{k2} \times e_2 + \dots + X_{kn} \times e_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix}$$

- The least squares criterion ensures that regression residuals will always be completely uncorrelated with all independent variables. **Note that this does not mean that in population X is uncorrelated with the disturbances ε; we have to assume this.**

Causes of endogeneity

- Endogeneity is caused by:
  - Omitted variable
  - Measurement error in independent variable
  - Reciprocal/reverse/feed-forward causation

(we will particularly discuss the last source)

Omitted variable

- Assume we have a model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

- Assume we have omitted an explanatory variable z_i which is correlated with x_i and which is expected to explain y_i.

- In this case the effect of z_i will be contained in error term, ε_i. Thus, x_i will be correlated with ε_i.

- However, we also know if relevant variable is not in the model, (in this case z_i), we will have substantially different (biased) estimate for other indicators (in this case x_i).

Measurement error in independent variables

- When the measurement error is in the dependent variable, the zero conditional mean assumption is not violated. In contrast, when the measurement error is in independent variables, the problem of endogeneity arises.
- If X_i is measured with errors. We observe $\tilde{X}_i = X_i + \varepsilon_i$ instead of X_i.

$$Y_i = \beta_0 + \beta_1 \tilde{X}_i - \beta_1 \varepsilon_i + u_i$$
$$u_i - \beta_1 \varepsilon_i = v_i$$
$$\beta_0 + \beta_1 \tilde{X}_i + v_i$$

- It can be seen that the problem of endogeneity occurs:

$$E(\tilde{X}_i, v_i) = E\big((X_i + \varepsilon_i)(u_i - \beta_1 \varepsilon_i)\big) =$$
$$-\beta_1 Var(\varepsilon_i) \neq 0$$

In other words, terms $X_i u_i = X_i \beta_1 \varepsilon_i = \varepsilon_i u_i = 0$, but $-\beta_1 \varepsilon_i \varepsilon_i \neq 0$

- This will lead to **attenuation bias** (biased OLS estimate towards zero).

Simultaneity

- One situation in which the assumption that the error term is uncorrelated with each of the independent variables in a regression equation is guaranteed to be violated is when there is simultaneity.

- Simultaneity arises when one or more of the independent variables, is jointly determined with the dependent variable, typically through an equilibrium mechanism. Suppose that the equilibrium relation between X and Y is expressed by the following simultaneous equations:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$
$$x_i = \alpha_0 + \alpha_1 y_i + v_i$$

- If we try to estimate any of the equations by substituting any of independent variables (e.g. x) as expressed in their equation form (e.g. $\alpha_0 + \alpha_1 y_i + v_i$), this will lead to simultaneity bias.

The problem of endogeneity

- The problem of endogeneity comes from assumption that change in y_i is due to change in x_i. In general we would like to estimate:

$$\beta_1 = \frac{\Delta y_i}{\Delta x_i}$$

- However, if x_i and u_i are correlated, the changes in y_i will be due to u_i

$$\frac{\Delta y_x + \Delta y_u}{\Delta x_i} = \beta_1 + \frac{\Delta y_u}{\Delta x_i}$$

- Thus, the change in y_i, will consist of two parts, a change x_i in and a change in u_i. But we are only interested in relation of y_i and x_i, so effect will bias our estimate of β_1.

Structural form

- How do we go about solving this. Assume that we hypothesize that variables t_i and z_i are explaining y_i and x_i, respectively. The structural form is formulated as:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 t_i + w_{1i}$$
$$x_i = \alpha_0 + \alpha_1 y_i + \alpha_2 z_i + w_{2i}$$

with:

$$w_{j,i} = N\left(0, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}\right)$$

- The **simultaneity** in the model, with respect explaining y_i and x_i, results from the fact that each variable depends on the contemporaneous value of the other variables in the model.

- We refer to the variables y_i and x_i as the endogenous variables. Variables t_i and z_i are considered exogenous variables

- Thus, endogenous variables act both as dependent and independent variables, while exogenous variables act purely as independent variables.

Simultaneity bias

- To be able to estimate this system of equations, we must substitute one of the equations into the other, because at least one of the equations is necessary to determine the other. So, for y_i, we get:

$$y_i = \beta_0 + \beta_1(\boldsymbol{\alpha_0 + \alpha_1 y_i + \alpha_2 z_i + w_{2i}}) + \beta_2 t_i + w_{1i}$$
$$y_i = \beta_0 + \beta_1\alpha_0 + \beta_1\alpha_1 y_i + \beta_1\alpha_2 z_i + \beta_1 w_{2i} + \beta_2 t_i + w_{1i}$$

- Moving $\boldsymbol{\beta_1\alpha_1 y_i}$ to the left side, we get:

$$y_i(1 - \beta_1\alpha_1) = \beta_0 + \beta_1\alpha_0 + \beta_1\alpha_2 z_i + \beta_2 t_i + \beta_1 w_{2i} + w_{1i}$$

- Dividing with $(1 - \beta_1\alpha_1)$, we get:

$$y_i = \frac{\beta_0 + \beta_1\alpha_0}{1 - \beta_1\alpha_1} + \underbrace{\frac{\beta_1\alpha_2}{1 - \beta_1\alpha_1}}_{\boldsymbol{\Pi_{11}}} z_i + \underbrace{\frac{\beta_2}{1 - \beta_1\alpha_1}}_{\boldsymbol{\Pi_{11}}} t_i + \underbrace{\frac{\beta_1 w_{2i} + w_{1i}}{1 - \beta_1\alpha_1}}_{\boldsymbol{\varepsilon}}$$

$$y_{1,t} = \boldsymbol{\Pi_{11}}\ z_i + \boldsymbol{\Pi_{11}}\ t_i + \boldsymbol{\varepsilon}$$

Simultaneity bias

- So, the result of our substitution of one of the equations into the other was

$$y_i = \frac{\beta_0 + \beta_1\alpha_0}{1-\beta_1\alpha_1} + \frac{\beta_1\alpha_2}{1-\beta_1\alpha_1} z_i + \frac{\beta_2}{1-\beta_1\alpha_1} t_i + \frac{\beta_1 w_{2i} + w_{1i}}{1-\beta_1\alpha_1}$$

(Analogous transformation can be performed for x_i)

- Obviously, due to $\frac{\beta_1 w_{2i} + w_{1i}}{1-\beta_1\alpha_1}$, y_i depends on w_{2i}. Nevertheless, w_{2i} is residual from equation for x_i, and thus correlated to x_i

- If one perform OLS to the original structural equation

$$\boldsymbol{y_i} = \beta_0 + \beta_1 \boldsymbol{x_i} + \beta_2 t_i + w_{1i}$$

it will lead to called simultaneity bias because $w_{1,i}$ incorporates $w_{2,i}$ and:

$$Cov(x_i, w_1) \neq 0$$

(Analogous problem is inherent to x_i as $Cov(y_i, w_2) \neq 0$)

Reduced form (cont.)

- As we saw earlier we can reduce the structural form to:

$$y_i = \Pi_{11} t_i + \Pi_{12} z_i + \varepsilon_{1,i}$$
$$x_i = \Pi_{21} t_i + \Pi_{22} z_i + \varepsilon_{2,i}$$

- The reduced form express the dependent variables **solely in terms of exogenous variables**. This equation can be estimated **consistently** by OLS as neither t_i or z_i are correlated to residual. However, the reduced form parameters are not parameters of interest. We are really interested in the structural model coefficients:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 t_i + w_{1i}$$
$$x_i = \alpha_0 + \alpha_1 y_i + \alpha_2 z_i + w_{2i}$$

which must then be derived from the reduced form

Structural and reduced form in matrix notation

- If you have G exogenous variables and K endogenous, variables you will have G equations

$$\beta_{11}y_1 + \cdots + \beta_{1G}y_G + \gamma_{11}x_1 \ldots + \gamma_{1K}x_K = u_1$$
$$\beta_{11}y_1 + \cdots + \beta_{2G}y_G + \gamma_{21}x_1 \ldots + \gamma_{2K}x_K = u_2$$
$$\ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots$$
$$\beta_{G1}y_1 + \cdots + \beta_{GG}y_G + \gamma_{G1}x_1 \ldots + \gamma_{GK}x_K = u_G$$

- In matrix form structural model is: $\boldsymbol{By + \Gamma x = u}$
- Assumption of model:
  - B is nonsingular
  - X is non-stochastic (or at least uncorrelated to u)
  - E(u) = 0
  - D(u) = Σ - dispersion of u is positive definite matrix – a symmetric matrix, but it is not diagonal (that would imply that the errors uncorrelated)
- **Since B is nonsingular G*G matrix, we can invert it.**

$$\boldsymbol{y = B^{-1}\Gamma x + B^{-1}u}$$

- Which means that we get

$$\boldsymbol{y = \Pi x + v}$$

- Where $\boldsymbol{\Pi = -B^{-1}\Gamma}$ comprises reduced form parameters, which is a G * K matrix

Single equation vs. system estimation methods

- There are various (sometimes equivalent) methods of estimating these equations. The theoretical interest of an analyst is often in just a single equation, however, all equations must be considered. On the other hand, we may seek to estimate whole system of equations. Thus, we can distinguish between:

- **Single equation estimators:**
  - Indirect least squares
  - Instrumental variables (IV) estimator
  - Two-stage least squares (2SLS) estimator
  - Limited information maximum likelihood
  - Generalized method of moments (GMM) estimator
- Today we will only discuss single equation estimators ***(IV and 2SLS, in particular)***. These methods can be applied equation-by-equation but then you are not accounting for possible correlation between the residuals.

- **System estimators:**
  - Three-stage least squares (3SLS) estimator
  - Iterated three-stage least squares (I3SLS) estimator
  - Full information maximum likelihood

- The major advantage of system estimation is that it uses more information, and therefore results in more precise parameter estimates. The major disadvantages are that it requires more data and is sensitive to model specification errors.

Instrumental variables

- Assume for now that we only have cross-sectional data.

- Instrumental variables estimators are concerned with single equation. Assume that we want predict y_i and that an endogenous variable x_i predicts y_i:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

- The idea of an instrumental variable is to find instrument z_i where changes in z_i are associated with changes in x_i, but are not associated with change in errors in u_i. So, conditions for instruments are:

Instrument exogeneity $Cov(z_i, u_i) = 0$
Instrument relevance $Cov(z_i, x_i) \neq 0$

- The idea is that z_i is causing x_i which in turn is causing y_i – in other words, we assume that instrument z_i can influence y_i only through x_i. So, we can specify equation:

$$x_i = \delta_0 + \delta z_i + \varepsilon_i$$

Instrumental variables - estimation

- In matrix notation, OLS estimation is based on following expressions:

$$\hat{\beta}_{OLS} = (X^T X)^{-1} X^T y$$
$$y - X\hat{\beta}_{OLS} = (I - X(X^T X)^{-1} X^T) y$$

- Assume that X is a $n \times (1 + k)$ vector of constants and a independent endogenous variables, Z is a $n \times (1 + k)$ vector of constants and instrumental variables and y is $n \times 1$ vector of observations on the dependent variable

- Then, instrumental variables estimation is based on applying the following formula to the data: $\hat{\beta}_{IV} = (Z^T X)^{-1} Z^T y$

- Note that Z does not completely replace X. We are not replacing regressors, but rather we simply premultiply with Z

Instrumental variables - derivation

- How do we formally derive an explicit form of instrumental variable estimator for a bivariate model? We take covariance of both sides of equation $y_i = \beta_0 + \beta_1 x_i + u_i$ with z_i.

$$Cov(z_i, y_i) = Cov(z_i, \beta_0 + \beta_1 x_i + u_i) =$$
$$= Cov(z_i, \beta_0) + \beta_1 Cov(z_i, x_i) + Cov(z_i, u_i)$$

- As we know that covariance of constant and a variable is zero $Cov(z_i, \beta_0) = 0$, and as we assumed that there is no covariance of a (good) instrument and and error, $Cov(z_i, u_i) = 0$, we get:

$$Cov(z_i, y_i) = \beta_1 Cov(z_i, x_i)$$

$\beta_1 = \frac{Cov(z_i, y_i)}{Cov(z_i, x_i)}$ and this is an IV estimator.

Instrumental variables (cont.)

- Under the maintained assumptions, the IV estimator is consistent as $N \to \infty$. The IV estimator can have a substantial bias in small samples and thus large samples are preferred.

- The IV estimator is not necessarily asymptotically efficient. This is because an endogenous variable can have more than one relevant instrumental variable.

- In this case, you can do one of two things. 1) Use as your instrumental variable the exogenous variable that is most highly correlated with the endogenous variable. 2) Use as your instrumental variable the linear combination of candidate exogenous variables most highly correlated with the endogenous variable.

- However, in this case, you are more likely to use 2SLS (we will discuss this next)

Bad and weak instruments

- **Bad instrument** is an instrument that violates assumption of $Cov(z_i, u_i) = 0$. Larger the covariance the worse is the bias. In this case the bias of IV estimator will be worse than bias of OLS.

- **Weak instrument** satisfies $Cov(z_i, u_i) = 0$. However in this case $Cov(z_i, x_i) \to 0$, that is, there is weak correlation between endogenous variable and the instrument - effectively being a zero.

  - The bias of IV estimator is approximately $\frac{Cov(z_i, \varepsilon_i)}{Var(\varepsilon_i)} \times \frac{1}{1+F}$. F in this expression is the F statistics produce by running "instrumental" (auxiliary) equation: $x_i = \delta z_i + \varepsilon_i$. So, if F statistics is 0, meaning that $\delta = 0$ (or in other words there is no correlation between z_i and x_i) bias is $\frac{Cov(z_i, \varepsilon_i)}{Var(\varepsilon_i)}$. Thus, in this case, it can be shown that the bias of IV estimator will be equal to the bias of OLS estimator.

Two-stage Least Squares (2SLS) estimator

- The 2SLS estimator is a generalization of the IV estimator and is the most common instrumental-variables estimator.

- 2SLS addresses the case in which the number of instruments exceeds the number of parameters to be estimated. Here there is no unique solution. 2SLS would be used to boil available instruments down into the single instrument needed. However, 2SLS reduces to the simple IV estimator if the equation is exactly identified.

- 2SLS estimator consists of two steps:

1. **step:** Regress each right-hand side endogenous variable in the equation to be estimated on all exogenous variables using the OLS estimator. Calculate the fitted values for each of these endogenous variables.

2. **step:** In the equation to be estimated, replace each endogenous right-hand side variable by its fitted value variable. Estimate the equation using the OLS estimator.

2SLS – estimation

- Remember that OLS estimates were based on: $\hat{\beta}_{OLS} = (X^T X)^{-1} X^T y$

$$\hat{\beta}_{2SLS} = (X^T P X)^{-1} X^T P y$$

$$\textit{where: } P = Z(Z^T Z)^{-1} Z^T$$

(thus, we are premultiplying both X and y by matrix P. Note that Z is now a $T \times I$ matrix, where I is the number of instruments (identifying and other))

- Similarly to OLS, if the error term has constant variance and the errors are uncorrelated, then the variance-covariance matrix of estimates is:

$$Var\_cor\left(\hat{\beta}_{2SLS}\right) = \sigma^2 (X^T P X)^{-1}$$

- The estimated variance-covariance matrix replaces unknown σ^2 with the estimate $\sigma^2 = RSS/n$ (where RSS is residual sum of squares)
  - Remember that OLS variance-covariance matrix is:

$$Var\_cor\left(\hat{\beta}_{OLS}\right) = \sigma^2 (X^T X)^{-1}$$

Instrumental variables and two-stage least squares for panel data models

- There are various estimators for fitting panel-data models with endogenous variables.
- These estimators are generalizations of simple panel-data estimators we have discussed in the first classes. Stata offers estimators for:
  - Fixed effects
  - Between effects
  - Random effects
  - First difference estimator (we talked about this one yesterday)
- If Y are endogenous and X are exogenous variables, the formal panel data would be

$$y_{it} = Y_{it}\gamma - X_{it}\beta + \mu_i + \nu_{it}$$

- If $Z_{it} = [Y_{it}X_{it}]$ then:

$$y_{it} = Z_{it}\delta + \mu_i + \nu_{it}$$

2SLS fixed (within) effects panel data model

- The within estimator (in Stata: FE2SLS) fits the model after sweeping out the μ_i by removing the panel-level means from each variable. The basis of the model is the within transformation. The within transform of a variable is:

$$\tilde{y}_{it} = y_{it} - \bar{y}_i + \bar{y}$$

where, if T_i is the number of observations on panel i, n is the number of panels, and N is the total number of observations, the within transform of a variable:

$$\bar{y}_i = \frac{1}{T_i}\sum_{t=1}^{T_i} y_{it} \qquad \bar{y} = \frac{1}{N}\sum_{i=1}^{n}\sum_{t=1}^{T_i} y_{it}$$

- Thus, the within transform of: $y_{it} = Z_{it}\delta + \mu_i + \nu_{it}$ is:

$$\tilde{y}_{it} = \tilde{Z}_{it}\delta + \tilde{\nu}_{it}$$

- With the μ_i gone, the within 2SLS estimator can be obtained from a two-stage least-squares regression

2SLS between effects panel data model

- Between estimator (in Stata: BE2SLS) models the panel averages. In other words, the between transform is based on:

$$\bar{y}_i = \frac{1}{T_i} \sum_{t=1}^{T_i} y_{it}$$

- After passing $y_{it} = Z_{it}\delta + \mu_i + \nu_{it}$ through the between transform, we get:

$$\bar{y}_i = \alpha + \bar{Z}_i\delta + \mu_i + \bar{\nu}_i$$

- We implement 2SLS on this expression.

2SLS random effects panel data model

- The two available random-effects estimators, treat the μ_i as random variables. Again, we assume that the μ_i are uncorrelated with the other covariates.

- Stata offers two estimation methods:

1. Balestra and Varadharajan-Krishnakumar method uses the exogenous variables after they have been passed through the feasible GLS transform - G2SLS

2. Baltagi's method uses the instruments $\tilde{X}_{it}$ and $\bar{X}_{it}$, where $\tilde{X}_{it}$ is constructed by passing each of the variables X_{it} through the within transform and $\bar{X}_{it}$ is constructed by passing each variable through the between transform - EC2SLS

2SLS random effects panel data model

- As we are dealing with the one-way RE framework, there are two variance components to estimate: the variance of the μ_i and the variance of the ν_{it} (these are our usual ν_i and ε_{it} but we had to change notation)

- Stata offers two choices to estimate these components:

  1. a Swamy-Arora method (applicable to unbalanced panels)
  2. a consistent estimators (Baltagi and Chang (2000)) which makes small-sample adjustments

2SLS first-differenced panel data model

- First-differenced estimator (in Stata: FD2SLS) removes the μ_i by fitting the model in first differences. Consequently, the model is:

$$y_{it} - y_{ti-1} = (Z_{it} - Z_{it-1})\delta + \nu_{it} - \nu_{it-1}$$

- Estimates are based on standard two-stage least-squares regression of Δy_{it} on ΔZ_{it} with instruments ΔX_{it}

- The first-differenced estimator is typically used when μ_i are not truly fixed-effects and when the model contains a lagged dependent (endogenous) variable

- We have already talked about this model yesterday when we talked about Anderson–Hsiao estimator