Lecture 8
Instrumental Variables

1

CLM: New Assumptions

• Last lecture, we presented a new set of assumptions for the CLM:
(**A1**) DGP: $\mathbf{y} = \mathbf{X}\beta + \boldsymbol{\varepsilon}$.
(**A2'**) **X** stochastic, but E[$\mathbf{X}'\boldsymbol{\varepsilon}$]= 0 and E[$\boldsymbol{\varepsilon}$]=**0.**
(**A3**) Var[$\boldsymbol{\varepsilon}|\mathbf{X}$] = $\sigma^2\,\mathbf{I}_T$
(**A4'**) plim ($\mathbf{X}'\mathbf{X}/T$) = **Q** (p.d. matrix with finite elements, rank= k)

• We studied the large sample properties of OLS:
- **b** and s^2 are consistent
- $\mathbf{b} \xrightarrow{a} N(\boldsymbol{\beta}, (\sigma^2/T)\,\mathbf{Q}^{-1})$
- *t-tests* asymptotically N(0,1), Wald tests asymptotically $\chi^2_{\text{rank}(\mathbf{S}\text{T})}$ and *F-tests* asymptotically $\chi^2_{\text{rank}(\text{Var}[\mathbf{m}])}$.
- Small sample behavior may be understood by simulations and/or bootstrapping.

The IV Problem

• We start with our CLM:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \qquad \text{(DGP)}$$

- Let's pre-multiplying the DGP by $\mathbf{X'}$

$$\mathbf{X'}\,\mathbf{y} = \mathbf{X'}\,\mathbf{X}\boldsymbol{\beta} + \mathbf{X'}\,\boldsymbol{\varepsilon}.$$

- We can interpret $\mathbf{b}$ as the solution obtained by first approximating $\mathbf{X'}\boldsymbol{\varepsilon}$ by zero, and then solving the k equations in k unknowns

$$\mathbf{X'y} = \mathbf{X'X}\,\mathbf{b} \qquad \text{(normal equations).}$$

Note: What makes $\mathbf{b}$ consistent when $\mathbf{X'}\boldsymbol{\varepsilon}/T \xrightarrow{p} \mathbf{0}$ is that approximating ($\mathbf{X'}\boldsymbol{\varepsilon}/T$) by $\mathbf{0}$ is reasonably accurate in large samples.

• Now, we challenge this approximation. We relax the assumption that $\{x_i, \varepsilon_i\}$ is a sequence of independent observations. That is,

$$\text{plim}\,(\mathbf{X'}\boldsymbol{\varepsilon}/T) \neq \mathbf{0}. \qquad \Rightarrow \text{This is the IV Problem!}$$

The IV Problem: OLS is Inconsistent

• A correlation between $\mathbf{X}$ & $\boldsymbol{\varepsilon}$ is not rare in economics, especially in corporate finance, where endogeneity is pervasive.

• *Endogenous in econometrics*: A variable is correlated with the error term.

• Q: What is the implication of the violation of plim($\mathbf{X'}\boldsymbol{\varepsilon}/T$) = 0?
From the asymptotic CLM version, we keep (**A1**), (**A3**), and (**A4'**):
(**A1**) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.
(**A3**) $\text{Var}[\boldsymbol{\varepsilon}|\mathbf{X}] = \sigma^2\,\mathbf{I}_T$
(**A4'**) $\text{plim}\,(\mathbf{X'X}/T) = \mathbf{Q}$

• Now, we assume (**A2''**) $\text{plim}(\mathbf{X'}\boldsymbol{\varepsilon}/T) \neq 0$.

• Then, $\text{plim}\,\mathbf{b} = \text{plim}\,\boldsymbol{\beta} + \text{plim}\,(\mathbf{X'X}/T)^{-1}\,\text{plim}\,(\mathbf{X'}\boldsymbol{\varepsilon}/T)$

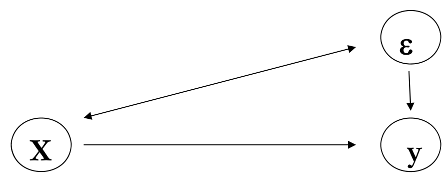
The IV Problem: OLS is Inconsistent

- $$\text{plim } \mathbf{b} = \text{plim } \boldsymbol{\beta} + \text{plim } (\mathbf{X'X}/T)^{-1} \text{ plim } (\mathbf{X'}\boldsymbol{\varepsilon}/T)$$
  $$= \boldsymbol{\beta} + \mathbf{Q}^{-1} \text{ plim } (\mathbf{X'}\boldsymbol{\varepsilon}/T) \neq \boldsymbol{\beta}$$

Under the new assumption, **b** is not a consistent estimator of $\boldsymbol{\beta}$.

Note: For finite samples, we could have challenged assumption (**A2**) $E[\boldsymbol{\varepsilon}|\mathbf{X}] = 0$. Then, $\text{Cov}(\mathbf{X},\boldsymbol{\varepsilon}) \neq 0 \Rightarrow E[\mathbf{b}|\mathbf{X}] \neq \boldsymbol{\beta}$.

- Diagram with $\text{Cov}(\mathbf{X},\boldsymbol{\varepsilon}) \neq 0$



The IV Problem: Structural Model

- **y** and **X** are both endogenous. Suppose, we also model **X** as a function of some exogenous variable **Z**. Then, the model becomes a *structural model* (everything is modeled):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$
$$\mathbf{X} = \mathbf{Z}\,\boldsymbol{\Pi} + \mathbf{V}$$

where **V** & $\boldsymbol{\varepsilon}$ are correlated.

The researcher is not interested in estimating the whole structural model, it is interested on the first equation: the impact of **X** on **y**.

Now, we can rewrite the inconsistency as

$$\text{plim } \mathbf{b} = \boldsymbol{\beta} + \mathbf{Q}^{-1} \text{ plim } ((\boldsymbol{\Pi}'\,\mathbf{Z}'\boldsymbol{\varepsilon} + \mathbf{V}'\boldsymbol{\varepsilon})/T)$$
$$= \boldsymbol{\beta} + \mathbf{Q}^{-1}\,\boldsymbol{\Pi}' \text{ plim } (\mathbf{Z}'\boldsymbol{\varepsilon}/T) + \mathbf{Q}^{-1} \text{ plim } (\mathbf{V}'\boldsymbol{\varepsilon}/T)$$

$\Rightarrow$ OLS inconsistency depends on relation between **Z** & $\boldsymbol{\varepsilon}$ and **V** & $\boldsymbol{\varepsilon}$.

The IV Problem: Example 1

• Suppose we want to study the relation between a firm's CEO's compensation (**y**) and a CEO's network (**x**).

Usually, a linear regression model is used, relating **y** and **x**, with additional "*control variables*" (**W**) controlling for other features that make one CEO's compensation different from another. The term **ε** represents the effects of individual variation that have not been controlled for with **W** or **x**.

The model is:

$$\mathbf{y} = \mathbf{x}\beta + \mathbf{W}\gamma + \boldsymbol{\varepsilon}$$

If a CEO's network is influenced by the CEO's natural skills, we have a problem: **y** and **x** are both endogenous –i.e., influenced by the unobserved CEO's skills, say **S**.

The IV Problem: Example 1

• **y** and **x** are both influenced by an unobservable variable. Then,

$$\text{Cov}(\mathbf{x}, \boldsymbol{\varepsilon}) \neq 0 \qquad (\Rightarrow \text{by LLN, plim } (\mathbf{X}'\boldsymbol{\varepsilon}/T) \neq 0)$$

• It looks like an *omitted variable problem.* Assuming linearity, it can be solved by adding as a control variable "CEO's skills," **S**:

$$\mathbf{y} = \mathbf{x}\beta + \mathbf{W}\gamma + \mathbf{S}\boldsymbol{\theta} + \boldsymbol{\eta}$$

However, **S** is unobservable.

<u>Note</u>: **x** is endogenous. It needs a model! Say, it depends on **Z**:

$$\mathbf{x} = \mathbf{Z}\,\pi + \mathbf{v} \qquad (\text{where } \boldsymbol{\sigma_{\varepsilon V}} \text{ measures the endogeneity of } \mathbf{x}.)$$

• Recall: Endogeneity occurs when a variable, **X**, is correlated with **ε**.

The IV Problem: Example 2

• Suppose we want to study the effect of military service (**x**) on earnings (**y**). We use a linear model, adding some control variables (**W**), controlling for other features that affect **y**:

$$\mathbf{y} = \mathbf{x}\,\beta + \mathbf{W}\gamma + \boldsymbol{\varepsilon}$$

β would measure "the causal effect" we would get if **x** were randomly assigned. But, there is *selection bias* by both individuals and military recruiters.

That is, **x** is not randomly assigned: Unobserved factors that affect **y**, also affect **x** ⇒ $\text{Cov}(\mathbf{x},\boldsymbol{\varepsilon}) \neq 0$.

.

The IV Problem: Example 3

• In this example, we introduce *measurement error* in **X**. That is, DGP:

$\mathbf{y} = \beta\mathbf{x}^* + \boldsymbol{\varepsilon}$ $\qquad\qquad \boldsymbol{\varepsilon} \sim iid\ D(\mathbf{0}, \sigma_\varepsilon^2)$

$\mathbf{x} = \mathbf{x}^* + \mathbf{u}$ $\qquad\qquad \mathbf{u} \sim iid\ D(\mathbf{0}, \sigma_u^2)$ -no correlation to $\boldsymbol{\varepsilon}$

We are interested in **x***, and its marginal effect β, but we observe/measure **x**, which measures **x*** with error (**u**).

All of the CLM assumptions apply. Then,

$$\mathbf{y} = \beta(\mathbf{x} - \mathbf{u}) + \boldsymbol{\varepsilon} = \beta\mathbf{x} + \boldsymbol{\varepsilon} - \beta\mathbf{u} = \beta\mathbf{x} + \mathbf{w}$$

$$E[\mathbf{x}'\mathbf{w}] = E[(\mathbf{x}^*+\mathbf{u})'(\boldsymbol{\varepsilon} - \beta\mathbf{u})] = -\beta\sigma_u^2 \neq 0 \quad \& \text{ plim } (\mathbf{X}'\mathbf{w}/T) \neq 0$$

⇒ CLM assumptions violated ⇒ OLS inconsistent!

The IV Problem: Example 4

• Simple supply and demand model for some good, where quantity (Q) and price (P) are endogenous variables –i.e., determined by the model. In equilibrium $Q_S = Q_d = Q$. We have a *simultaneous equation model* (SEM):

$$Q = \alpha_1 P + \alpha_2 Y + \varepsilon_d$$
$$Q = \beta_1 P + \varepsilon_S$$

where Y is income, considered exogenous, and ε_S and ε_d are the error terms.

Suppose we are interested in estimating β_1. An OLS regression with $\mathbf{X}$ = P will not work, since Cov($\mathbf{P}$, $\boldsymbol{\varepsilon}_S$) ≠ 0.

The IV Problem: Usual Cases

• Q: When might an explanatory variable (a regressor) be correlated with the error term?
- Omitted variables
- Selection bias
- Measurement error
- Simultaneous equations
- Misspecification
- Correlated shocks across linked equations
- Model has a lagged dependent variable and a serially correlated error term

Instrumental Variables: New CLM Assumptions

• New Framework:

(**A1**) DGP: $\mathbf{y} = \mathbf{X}\,\beta + \boldsymbol{\varepsilon}$.

(**A2''**) plim $(\mathbf{X'}\boldsymbol{\varepsilon}/T) \neq 0$

(**A3**) $\text{Var}[\boldsymbol{\varepsilon}|\mathbf{X}] = \sigma^2\,\mathbf{I}_T$

(**A4'**) plim $(\mathbf{X'X}/T) = \mathbf{Q}$ (p.d. matrix, with rank k)

$\Rightarrow$ **b** is not a consistent estimator of $\boldsymbol{\beta}$.

• Q: How can we construct a consistent estimator of $\boldsymbol{\beta}$?

We will assume that there exists a set of l variables, **Z** such that

(1) plim $(\mathbf{Z'X}/T) \neq \mathbf{0}$ (*relevant condition*)

(2) plim $(\mathbf{Z'}\boldsymbol{\varepsilon}/T) = \mathbf{0}$ (*valid condition*)

• The variables in **Z** are called *instrumental variables* (IVs). In general, not all the **X** will be correlated with error $\boldsymbol{\varepsilon}$.

Instrumental Variables: Endogeneity

• We can also write the new framework, emphasizing endogeneity, as:

(**A1**) DGP: $\mathbf{y} = \mathbf{Y}\,\boldsymbol{\beta} + \mathbf{U}\,\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$.

(**A2''**) plim $(\mathbf{Y'}\,\boldsymbol{\varepsilon}/T) \neq 0$ (**Y**: "problem," *endogenous*, variables)

(**A2''**) plim $(\mathbf{U'}\,\boldsymbol{\varepsilon}/T) = 0$ (**U**: *clean* variables)

(**A3**) $\text{Var}[\boldsymbol{\varepsilon}|\mathbf{Y},\mathbf{U}] = \sigma_\varepsilon^2\,\mathbf{I}_T$

(**A4**) **Y** and **U** have full column rank. Say k_x and k_u.

• We have **Z**, a matrix of l "*excluded instruments*" –the IVs. The IVs have no impact on **y** except through **Y**. We relate **Y** to **Z** linearly by:

$$\mathbf{Y} = \mathbf{Z}\,\boldsymbol{\Pi} + \mathbf{U}\,\boldsymbol{\Phi} + \mathbf{V} \qquad – \mathbf{V} \sim D(\mathbf{0}, \sigma_V^2\,\mathbf{I}_T)$$

<u>Note</u>: When the number k_x of "endogenous" variables is greater than one, we have a system of multiple equations. The estimation of this equation is called "*first stage*."

Instrumental Variables: Endogeneity

• Concentrating on two equations (and let **X**=**Y**):

(**A1**) $\mathbf{y} = \mathbf{X}\,\boldsymbol{\beta} + \mathbf{U}\,\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$ (called *structural equation)*

$\mathbf{X} = \mathbf{Z}\,\boldsymbol{\Pi} + \mathbf{U}\,\boldsymbol{\Phi} + \mathbf{V}$ (*first stage* regression)

Replacing the second equation in (**A1**):

$$\mathbf{y} = (\mathbf{Z}\,\boldsymbol{\Pi} + \mathbf{U}\,\boldsymbol{\Phi} + \mathbf{V})\,\boldsymbol{\beta} + \mathbf{U}\,\boldsymbol{\gamma} + \boldsymbol{\varepsilon} = \mathbf{Z}\,\boldsymbol{\Pi}\,\boldsymbol{\beta} + \mathbf{U}\,\boldsymbol{\varphi} + \boldsymbol{\xi}$$

This equation is called *reduced form*, where

$$\boldsymbol{\varphi} = \boldsymbol{\Phi}\,\boldsymbol{\beta} + \boldsymbol{\gamma}$$
$$\boldsymbol{\xi} = \mathbf{V}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

<u>Note</u>: Usually, **V** and **ε** are N(0, $\boldsymbol{\sigma_{JJ} I}$). But, they can be correlated.

• In this lecture, the parameter of interest is **β**. OLS cannot estimate it. But OLS works on the reduced form to consistently estimate **Γ**=**Πβ**.

Instrumental Variables: Notation

• Model:

$\mathbf{y} = \mathbf{X}\,\boldsymbol{\beta} + \mathbf{U}\,\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$ - *structural equation*

$\mathbf{X} = \mathbf{Z}\,\boldsymbol{\Pi} + \mathbf{U}\,\boldsymbol{\Phi} + \mathbf{V}$ - *first stage equation*

$\mathbf{y} = \hat{X}\boldsymbol{\beta} + \mathbf{U}\,\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$ - *second stage equation*

$\mathbf{y} = \mathbf{Z}\,\boldsymbol{\Gamma} + \mathbf{U}\,\boldsymbol{\varphi} + \boldsymbol{\xi}$ - *reduced form*

• Variables

y, **X**: endogenous variables –i.e., correlated with **ε**.

U & **Z**: exogenous variables –i.e., uncorrelated with **ε**.

U: included instruments, clean variables ("*controls"*)

Z: excluded instruments, IVs –i.e., satisfies the *relevant condition* and the *valid condition*, also referred as *exclusion restriction*. (Excluded = not included in the structural equation.)

Instrumental Variables: Notation

• Parameters

β: Structural parameter, usually the parameter of interest

Π: 1st-stage parameter. It captures the strength of the IV, **Z**. If **Π**≈0, not very powerful.

Γ: Reduced form parameter. It can show the potential of **Z** as instrument.

• Equations

Structural equation: Theory dictates this relation: it relates **y** and **X** (or **Y**). It measures the causal effect of **X** on **y**, **β**; but the effect is blurred by endogeneity.

First stage: Regression of **X** on the instrument, **Z** (it measures a causal effect from **Z** to **X**).

Reduced form: Regression of **y** on the instrument is called the reduced form (it measures the direct causal effect from **Z** to **y**).

Instrumental Variables: Assumptions

• New assumption: we have l IVs, **Z**, such that

$$\text{plim}(\mathbf{Z'X}/T) \neq \mathbf{0} \quad \text{but} \quad \text{plim}(\mathbf{Z'}\boldsymbol{\varepsilon}/T) = \mathbf{0}$$

• Then, we state assumptions to construct an alternative (to OLS) consistent estimator of **β**.

Assumptions:

$\{x_i, z_i, \varepsilon_i\}$ is a sequence of RVs, with:

$E[\mathbf{X'X}] = \boldsymbol{Q}_{xx}$ (pd and finite) (LLN $\Rightarrow$ $\text{plim}(\mathbf{X'X}/T) = \boldsymbol{Q}_{xx}$)

$E[\mathbf{Z'Z}] = \boldsymbol{Q}_{zz}$ (finite) (LLN $\Rightarrow$ $\text{plim}(\mathbf{Z'Z}/T) = \boldsymbol{Q}_{zz}$)

$E[\mathbf{Z'X}] = \boldsymbol{Q}_{zx}$ (pd and finite) (LLN $\Rightarrow$ $\text{plim}(\mathbf{Z'X}/T) = \boldsymbol{Q}_{zx}$)

$E[\mathbf{Z'}\boldsymbol{\varepsilon}] = \mathbf{0}$ (LLN $\Rightarrow$ $\text{plim}(\mathbf{Z'}\boldsymbol{\varepsilon}/T) = \mathbf{0}$)

Instrumental Variables: Estimation

• To construct a new estimator, we start by pre-multiplying the DGP by **W'Z'**, where **W** $l \times k$ weighting matrix that we choose:

$$\mathbf{W'Z'y} = \mathbf{W'Z'}(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}) = \mathbf{W'Z'X}\boldsymbol{\beta} + \mathbf{W'Z'}\boldsymbol{\varepsilon}$$

⇒ **W** helps to create a $k \times k$ square matrix, needed for inversion.

• Following the same idea as in OLS, we get a system of equations:

$$\mathbf{W'Z'X}\,\mathbf{b}_{IV} = \mathbf{W'Z'y}$$

• We have two cases where estimation is possible:

- **Case 1:** $l = k$ -i.e., number of instruments = number of regressors.
- **Case 2:** $l > k$ -i.e., number of instruments > number of regressors.

The second case is the usual situation. We can throw l-k instruments, but throwing away information is never optimal.

IV Estimation

• **Case 1:** $l = k$ -i.e., number of instruments = number of regressors.

To get the IV estimator, we start from the system of equations:

$$\mathbf{W'Z'X}\,\mathbf{b}_{IV} = \mathbf{W'Z'y}$$

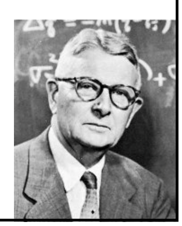
- dim(**Z**) = dim(**X**): Txk ⇒ **Z'X** is a kxk pd matrix
- In this case, **W** is irrelevant, say, **W**=**I**. Then,

$$\mathbf{b}_{IV} = (\mathbf{Z'X})^{-1}\mathbf{Z'y}$$

Note: Let **Z** = **X**. Then,

$$\mathbf{b}_{IV} = \mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'y}$$

That is, under the usual assumptions, **b** is an IV estimator with **X** as its own instrument.



Sewall G. Wright (1889 – 1988, USA)

IV Estimators: Properties – Consistency

• Properties of $\mathbf{b}_{IV}$

(1) Consistent

$$\mathbf{b}_{IV} = (\mathbf{Z'X})^{-1}\mathbf{Z'y} = (\mathbf{Z'X})^{-1}\mathbf{Z'}(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon})$$
$$= (\mathbf{Z'X}/T)^{-1} (\mathbf{Z'X}/T)\, \boldsymbol{\beta} + (\mathbf{Z'X}/T)^{-1}\mathbf{Z'}\boldsymbol{\varepsilon}/T$$

Under assumptions:

$$\text{plim}(\mathbf{b}_{IV}) = \boldsymbol{Q}_{zx}^{-1}\, \boldsymbol{Q}_{zx}\, \boldsymbol{\beta} + \boldsymbol{Q}_{zx}^{-1}\, \text{plim}(\mathbf{Z'}\boldsymbol{\varepsilon}/T)$$
$$= \boldsymbol{\beta} + \boldsymbol{Q}_{zx}^{-1}\, \text{plim}(\mathbf{Z'}\boldsymbol{\varepsilon}/T) \xrightarrow{p} \boldsymbol{\beta}$$

<u>Note</u>:

- Under the context of Lecture 7 –i.e., (**A2'**) $\text{plim}(\mathbf{X'}\boldsymbol{\varepsilon}/T) = \mathbf{0}$–, **b** is consistent. But, $\mathbf{b}_{IV}$ is also consistent (though, not efficient)!

- Under the context of this Lecture –i.e., (**A2'**) $\text{plim}(\mathbf{X'}\boldsymbol{\varepsilon}/T) \neq \mathbf{0}$–, *only* the IV estimator is consistent, **b** is not.

IV Estimators: Properties – Asy. Normality

• Properties of $\mathbf{b}_{IV}$

(2) Asymptotic normality

$$\sqrt{T}\,(\mathbf{b}_{IV} - \boldsymbol{\beta}) = \sqrt{T}\,(\mathbf{Z'X})^{-1}\mathbf{Z'}\boldsymbol{\varepsilon}$$
$$= (\mathbf{Z'X}/T)^{-1}\sqrt{T}\,(\mathbf{Z'}\boldsymbol{\varepsilon}/T)$$

Using the Lindberg-Feller CLT

$$\sqrt{T}\,(\mathbf{Z'}\boldsymbol{\varepsilon}/T) \xrightarrow{d} N(0, \sigma^2 \boldsymbol{Q}_{zz})$$

Then,

$$\sqrt{T}\,(\mathbf{b}_{IV} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \sigma^2 \boldsymbol{Q}_{zx}^{-1} \boldsymbol{Q}_{zz} \boldsymbol{Q}_{xz}^{-1})$$

IV Estimators: Asympotic Var[b_{IV}]

• Properties of $\hat{\sigma}^2$, under IV estimation: Consistency

- We define $\hat{\sigma}^2$: $\hat{\sigma}^2 = \frac{1}{T}\sum_{i=1}^{T} e_{IV}^2 = \frac{1}{T}\sum_{i=1}^{T}(y_i - x' b_{IV})^2$

where $\mathbf{e}_{IV} = \mathbf{y} - \mathbf{X}\,\mathbf{b}_{IV} = \mathbf{y} - \mathbf{X}(\mathbf{Z'X})^{-1}\mathbf{Z'y} = [\mathbf{I} - \mathbf{X}(\mathbf{Z'X})^{-1}\mathbf{Z'}]\mathbf{y} = \mathbf{M_{zx}}\,\mathbf{y}$

$\mathbf{e}_{IV} = [\mathbf{I} - \mathbf{X}(\mathbf{Z'X})^{-1}\mathbf{Z'}] * (\mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = [\mathbf{X}\,\boldsymbol{\beta} - \mathbf{X}(\mathbf{Z'X})^{-1}\mathbf{Z'X}\,\boldsymbol{\beta}] + \mathbf{M_{zx}}\,\boldsymbol{\varepsilon}$

- Then,

$\hat{\sigma}^2 = \mathbf{e}_{IV}'\mathbf{e}_{IV}/T = \boldsymbol{\varepsilon}'\mathbf{M_{zx}}'\mathbf{M_{zx}}\boldsymbol{\varepsilon}/T$

$= \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}/T - 2\,\boldsymbol{\varepsilon}'\mathbf{X}\,(\mathbf{Z'X})^{-1}\mathbf{Z'}\boldsymbol{\varepsilon}/T + \boldsymbol{\varepsilon}'\mathbf{Z}\,(\mathbf{Z'X})^{-1}\mathbf{X'X}(\mathbf{Z'X})^{-1}\mathbf{Z'}\boldsymbol{\varepsilon}/T$

$\Rightarrow \text{plim}\,\hat{\sigma}^2 = \text{plim}(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}/T) - 2\,\text{plim}[(\boldsymbol{\varepsilon}'\mathbf{X}/T)\,(\mathbf{Z'X}/T)^{-1}\,(\mathbf{Z'}\boldsymbol{\varepsilon}/T)] +$
$+ \text{plim}(\boldsymbol{\varepsilon}'\mathbf{Z}\,(\mathbf{Z'X})^{-1}\mathbf{X'X}(\mathbf{Z'X})^{-1}\mathbf{Z'}\boldsymbol{\varepsilon}/T) = \sigma^2$

Est Asy. Var[$\mathbf{b}_{IV}$] $= E[(\mathbf{Z'X})^{-1}\,\mathbf{Z'}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{Z}\,(\mathbf{Z'X})^{-1}] = \hat{\sigma}^2\,(\mathbf{Z'X})^{-1}\,\mathbf{Z'Z}(\mathbf{Z'X})^{-1}$

IV Estimators: Example

Simplest case: Linear model, two endogenous variables, one IV.

$\mathbf{y_1} = \mathbf{y_2}\,\beta + \boldsymbol{\varepsilon}$ $\qquad - \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_{\varepsilon\varepsilon})$

$\mathbf{y_2} = \mathbf{z}\,\pi + \mathbf{v}$ $\qquad - \mathbf{v} \sim N(\mathbf{0}, \sigma_{VV})$

with reduced form:

$\mathbf{y_1} = \mathbf{z}\,\pi\,\beta + \mathbf{v}\,\beta + \boldsymbol{\varepsilon} = \mathbf{z}\,\gamma + \boldsymbol{\xi}.$

The parameter of interest is β ($= \gamma/\pi$).

• We estimate β with IV: $b_{IV} = \frac{\frac{1}{T}\sum_i^T (y_{1,i} - \overline{y_1})(z_i - \bar{z})}{\frac{1}{T}\sum_i^T (y_{2,i} - \overline{y_2})(z_i - \bar{z})}$

<u>Note</u>: With a reasonably large T both numerator and denominator are well approximated by Normals and if $\pi \neq 0$, as T gets large, then the ratio will eventually be well approximated by a normal distribution.

IV Estimators: Example

• To analyze the bias,

$$\mathbf{b}_{IV} = (\mathbf{z'}\,\mathbf{y_2})^{-1}\,\mathbf{z'}\,\mathbf{y_1} = \beta + (\mathbf{z'}\,\mathbf{y_2})^{-1}\,\mathbf{z'}\,\boldsymbol{\varepsilon}$$

$$\text{plim}(\mathbf{b}_{IV}) - \beta = \text{plim}(\mathbf{z'}\boldsymbol{\varepsilon}/T)/\text{plim}(\mathbf{z'}\mathbf{y_2}/T) \xrightarrow{p} \text{Cov}(\mathbf{z}, \boldsymbol{\varepsilon})/\text{Cov}(\mathbf{z}, \mathbf{y_2})$$

• When $\text{Cov}(\mathbf{z}, \boldsymbol{\varepsilon}) \neq 0$, IV estimation is inconsistent.

• If $\text{Cov}(\mathbf{z}, \boldsymbol{\varepsilon})$ is small, but $\pi \approx 0$, the inconsistency can get large ($\pi \approx 0$)

$\Rightarrow \text{Cov}(\mathbf{z}, \mathbf{y_2}) = \text{Cov}(\mathbf{z}, (\mathbf{z}\pi + \mathbf{v})) = \pi\ \text{Var}(\mathbf{z}) + \text{Cov}(\mathbf{z}, \mathbf{v}) = \pi\ \text{Var}(\mathbf{z}) \approx 0$

• When $\pi = 0 \Rightarrow \text{Cov}(\mathbf{z}, \mathbf{y_2}) = 0$, thus, the IV estimator is not defined. When $\pi = 0$, the instrument provides no information. It is an *irrelevant instrument*.

IV Estimators: Example

• When π is small, we say **z** is a *weak instrument*. It provides information, but, as we will see later, not enough.

• Note that even when $\pi=0$, in finite samples, the sample analogue to $\text{Cov}(\mathbf{z}, \mathbf{y_2}) \neq 0$. Not very useful fact, the sampling variation in $\text{Cov}(\mathbf{z}, \mathbf{y_2})$ is not helpful to estimate β.

Note: The weak instrument literature is concerned with testing H_0:$\beta = \beta_0$ when π is too close to 0. As we will see later, the normal approximation to the ratio will not be accurate.

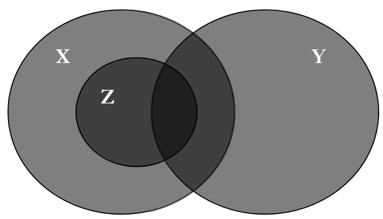
IV Estimators: Weak and Strong Instruments

• We assume that there exists a set of *l* variables, **Z** such that

(1) plim $(\mathbf{Z'X}/T) \neq \mathbf{0}$ (*relevant condition*)

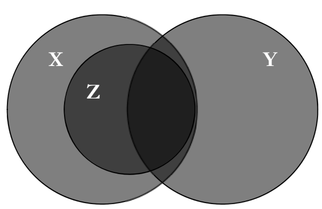
(2) plim $(\mathbf{Z'}\boldsymbol{\varepsilon}/T) = \mathbf{0}$ (*valid condition* –or *exclusion restriction*)

• We are going to use the variation in **Z**, which is uncorrelated with **ε**, to explain the variation of **y**. Condition (1) allows to do this. Suppose the relation between **Z**, **X** and **y** is given in the following diagram:



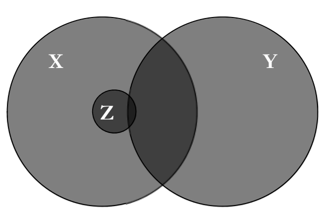
Now, not all the variation in **X** is used. Only the portion of **X** which is "explained" by **Z** can be used to explain **y**.

IV Estimators: Weak and Strong Instruments



• Best situation : A lot of **X** is explained by **Z**, and most of the overlap between **X** and **Y** is accounted for.

⇒ **Z** is a *strong* IV.



Usual situation : Not a lof of **X** is explained by **Z**, or what is explained does not overlap much with **Y**.

⇒ **Z** is a *weak* IV.

IV Estimators: 2SLS (2-Stage Least Squares)

• **Case 2:** $l > k$ -i.e., number of instruments > number of regressors.

- This is the usual case. We can throw l-k instruments, but throwing away information is never optimal.

- The IV normal equations are an l x k system of equations:

$$\mathbf{Z'y} = \mathbf{Z'X}\boldsymbol{\beta} + \mathbf{Z'}\boldsymbol{\varepsilon}$$

Note: We cannot approximate all the $\mathbf{Z'}\boldsymbol{\varepsilon}$ by **0** simultaneously. There will be at least l-k non-zero residuals. (Similar setup to a regression!)

- From the IV normal equations $\Rightarrow$ $\mathbf{W'Z'X}\, \mathbf{b}_{IV} = \mathbf{W'Z'y}$

- We define a different IV estimator

- Let $\mathbf{ZW} = \mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z'X} = \mathbf{P_Z X} = Z\hat{\Pi} = \hat{X}$

- Then, $\mathbf{X'P_Z X}\, \mathbf{b}_{IV} = \mathbf{X'P_Z y}$

$$b_{IV} = (X'P_Z X)^{-1} X'P_Z y = (X'P_Z P_Z X)^{-1} X'P_Z P_Z y = (\hat{X}'\hat{X})^{-1}\hat{X}'\hat{y}$$

IV Estimators: 2SLS: Properties

• It is easy to derive properties for $\mathbf{b}_{IV}$:

$$b_{IV} = \beta + (X'P_Z X)^{-1} X'P_Z \varepsilon$$

(1) $\mathbf{b}_{IV}$ is consistent

(2) $\mathbf{b}_{IV}$ is asymptotically normal.

- This is estimator is also called GIVE (*Generalized IV estimator*)

Note: In general, we think of $\mathbf{X} = \mathbf{Z\Pi} + \mathbf{V}$, where $\mathbf{V} \sim N(\mathbf{0}, \boldsymbol{\sigma_{VV}} \boldsymbol{I})$.
In this case, we add the assumption: plim($\mathbf{Z'V}/T$) = **0**.

• Interpretations of $\mathbf{b}_{IV}$

$b_{IV} = b_{2SLS} = (\hat{X}'\hat{X})^{-1}\hat{X}'y$ This is the 2SLS interpretation

$b_{IV} = (\hat{X}'X)^{-1}\hat{X}'y$ This is the usual IV $Z = \hat{X}$

IV Estimators: 2SLS - Interpretation

• Interpretation of $\mathbf{b}_{IV}$ as a 2SLS regression -Theil (1953).

$$b_{2SLS} = (\hat{X}'\hat{X})^{-1}\hat{X}'y$$

- First stage, an OLS regression of **X** on **Z**. Get fitted values $\hat{X}$.
- Second stage, another OLS regression of **y** on $\hat{X}$. Get $\mathbf{b}_{IV} = \mathbf{b}_{2SLS}$.

Notes:

- In the 1st stage, any variable in **X** that is also in **Z** will achieve a perfect fit (these **X** are *clean*), so that this variable is carried over without modification to the second stage.
- In the 2nd stage, under the usual linear model for **X**: $\mathbf{X} = \mathbf{Z\Pi} + \mathbf{V}$,

$$y = X\beta + \varepsilon = \hat{X}\beta + \{\varepsilon + (X - \hat{X})\beta\}$$

The second component of the error term is a source of finite sample bias, but not inconsistency.

IV Estimators: 2SLS - Interpretation

• In the simplest case with one explanatory variable and one instrument –i.e., $\mathbf{x} = \mathbf{z}\,\pi + \mathbf{v}$– we get the simple IV estimator:

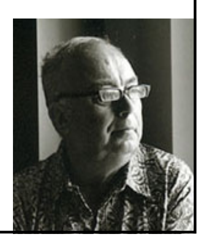
$$b_{2SLS} = (\hat{X}'X)^{-1}\hat{X}'y = (\hat{\pi}'Z'X)^{-1}\hat{\pi}'Z'y = (ZX)^{-1}Z'y$$

• The 2SLS estimator can be interpreted as a member of the family of GMM estimators.

• In this case the moment is E[**Z'ε**] and GMM selects β to minimize the weighted quadratic distance:

$$\boldsymbol{\varepsilon}'\mathbf{Z}\,\mathbf{W}_T\,\mathbf{Z}'\boldsymbol{\varepsilon}$$

where $\mathbf{W}_T$ is a weight matrix.



Henri Theil (1924-2000, Netherlands)

IV Estimators: 2SLS – Simultaneous Equations

• To check the factors that affect the behavior of IV, let's go back to a simultaneous equation setting:

$$\mathbf{y_1} = \mathbf{Y}\ \boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad – \boldsymbol{\varepsilon} \sim \text{N}(\mathbf{0}, \boldsymbol{\sigma_{\varepsilon\varepsilon}} \boldsymbol{I})$$

$$\mathbf{Y} = \mathbf{Z}\ \boldsymbol{\Pi} + \mathbf{V} \qquad – \mathbf{V} \sim \text{N}(\mathbf{0}, \boldsymbol{\sigma_{VV}} \boldsymbol{I})$$

Then,

$$\mathbf{b}_{2SLS} = [\mathbf{Y}'\ \mathbf{P_z}\ \mathbf{Y}]^{-1}\ \mathbf{Y}'\mathbf{P_z}\ \mathbf{y_1}$$

$$= [(\boldsymbol{\Pi}'\mathbf{Z}'+\mathbf{V}')\ \mathbf{P_z}\ (\mathbf{Z}\boldsymbol{\Pi}+\mathbf{V})]^{-1}\ (\boldsymbol{\Pi}'\mathbf{Z}'+\mathbf{V}')\ \mathbf{P_z}\ (\mathbf{Y}\ \beta + \boldsymbol{\varepsilon})$$

$$\mathbf{b}_{2SLS} - \boldsymbol{\beta} = [\boldsymbol{\Pi}'\mathbf{Z}'\mathbf{Z}\boldsymbol{\Pi} + \mathbf{V}'\mathbf{P_z}\mathbf{V} + \boldsymbol{\Pi}'\mathbf{Z}'\mathbf{V} + \mathbf{V}'\mathbf{Z}\boldsymbol{\Pi}]^{-1}\ (\boldsymbol{\Pi}'\mathbf{Z}'\boldsymbol{\varepsilon} + \mathbf{V}'\mathbf{P_z}\boldsymbol{\varepsilon})$$

The parameter $\lambda = \boldsymbol{\Pi}'\mathbf{Z}'\mathbf{Z}\boldsymbol{\Pi}/\sigma_{VV}$ is called the *concentration parameter*.

• The bias depends on the behavior of $\mathbf{Z}'\boldsymbol{\varepsilon}$ (correlation between $\mathbf{Z}$ & $\boldsymbol{\varepsilon}$), $\mathbf{V}'\mathbf{Z}$ (exogeneity of $\mathbf{Z}$), and $\mathbf{Z}\boldsymbol{\Pi}$ (correlation between $\mathbf{Z}$ &$\mathbf{Y}$).

IV Estimators: 2SLS - Example

• Simplest case: Two endogenous variable, one IV.

$$\mathbf{y_1} = \mathbf{y_2}\ \beta + \boldsymbol{\varepsilon} \qquad – \boldsymbol{\varepsilon} \sim \text{N}(\mathbf{0}, \sigma_{\varepsilon\varepsilon})$$

$$\mathbf{y_2} = \mathbf{z}\ \pi + \mathbf{v} \qquad – \mathbf{v} \sim \text{N}(\mathbf{0}, \sigma_{VV})$$

• The 2SLS bias term ($\mathbf{P_z} = \mathbf{zz}'/\Sigma_i z_i^2$)

$$b_{2SLS} – \beta = [\pi^2\ \mathbf{z}'\mathbf{z} + \mathbf{v}'\mathbf{P_z}\mathbf{v} + 2\ \pi\ \mathbf{z}'\mathbf{v}]^{-1}\ (\mathbf{z}'\boldsymbol{\varepsilon} + \mathbf{v}'\mathbf{P_z}\boldsymbol{\varepsilon})$$

We call $\lambda = (\pi^2\ \Sigma_i z_i^2)/\sigma_{VV}$ the *concentration parameter*.

If $\mathbf{z}$ is uncorrelated with $\mathbf{v}$ –i.e., exogenous–, then:

$$b_{2SLS} – \beta = [\pi^2\ \mathbf{z}'\mathbf{z}]^{-1}\ (\mathbf{z}'\boldsymbol{\varepsilon})$$

• When Cov($\mathbf{z}$, $\boldsymbol{\varepsilon}$) ≠ 0, 2SLS is inconsistent. If λ is close to 0 the bias term will get larger (λ ≈ 0 when π ≈ 0 –i.e., Cov($\mathbf{z}$, $\mathbf{y_2}$) ≈ 0).

IV Estimators: 2SLS - Example

• Subtle point: Even if Cov(**z**, **ε**) = 0, in small samples $\mathbf{b}_{2SLS}$ can be misleading (biased with downward biased SEs).

Problems can be serious when $\pi \approx 0$ and/or l is large relative to k.

IV Estimators: Identification

• **Case 3:** $l < k$ -i.e., number of instruments < number of regressors.

- We cannot estimate **β**. We do not have enough information in **Z** to estimate **β**.
- This is the *identification problem*. This is the case where we need to rethink the estimation strategy.
- When we can estimate **β**, we say the model is *identified*. This happens when $l \geq k$.

Note : When $l \geq k$, we have two cases:

-When $l = k$, we say the model is *just identified*.

-When $l > k$, we say the model is *over-identified*.

Asymptotic Covariance Matrix for 2SLS (Greene)

• The asymptotic variance for the IV and 2SLS is given by:

$$E[(b_{IV}-\beta)(b_{IV}-\beta)'] = \sigma_\varepsilon^2 (Z'X)^{-1}(Z'Z)(X'Z)^{-1}$$
$$E[(b_{2SLS}-\beta)(b_{2SLS}-\beta)'] = \sigma_\varepsilon^2 (\hat{X}'X)^{-1}(\hat{X}'\hat{X})(X'\hat{X}')^{-1}$$
$$= \sigma_\varepsilon^2 (\hat{X}'\hat{X})^{-1}$$

• To estimate Asy Var[$\mathbf{b}_{2SLS}$] we need to estimate σ_ε^2:

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{T}\sum_{i=1}^{T}(y_i - x_i b_{2SLS})^2$$

• Do not use the inconsistent estimator:

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{T}\sum_{i=1}^{T}(y_i - \hat{x}_i b_{2SLS})^2$$

Asymptotic Covariance Matrix for 2SLS (Greene)

• A little bit of algebra relates the asymptotic variances of $\mathbf{b}_{IV}$ & $\mathbf{b}_{OLS}$:

$$V[(b_{IV}] = \sigma_\varepsilon^2 (Z'X)^{-1}(Z'Z)(X'Z)^{-1}$$
$$= \sigma_\varepsilon^2 (X'X)^{-1}(X'X)(Z'X)^{-1}(Z'Z)(X'Z)^{-1}$$
$$= V[(b_{OLS}][(X'X)(Z'X)^{-1}](Z'Z)(X'Z)^{-1}$$
$$= V[(b_{OLS}][\Phi^{-1}\hat{\Pi}^{-1}]$$

where we assume $\mathbf{X} = \mathbf{Z\Pi} + \mathbf{V}$ and $\mathbf{\Phi}$ is the coefficient in the reverse first stage regression.

Two things to notice:

- As $\mathbf{Z} \longrightarrow \mathbf{X}$, V[$\mathbf{b}_{IV}$] $\longrightarrow$ V[$\mathbf{b}_{OLS}$].
- As Cov(**Z**,**X**) gets smaller –i.e., **Z** becomes a *weak instrument*-, V[$\mathbf{b}_{IV}$] gets larger.

Asymptotic Covariance Matrix for 2SLS (Greene)

• This relation between Cov(**Z, X**) and estimation uncertainty also applies to the 2SLS estimators:

$$V[(b_{2SLS}] = \sigma_\varepsilon^2 (\hat{X}'\hat{X})^{-1}$$
$$= \sigma_\varepsilon^2 (\hat{\Pi}'Z'Z\hat{\Pi})^{-1}$$

$\Rightarrow$ Weak instruments create big uncertainty about $\mathbf{b}_{2SLS}$.

2SLS Has Larger Variance than OLS (Greene)

A comparison to OLS

Asy.Var[2SLS]=$\sigma^2(\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}$

Neglecting the inconsistency,

Asy.Var[LS] $=\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$

(This is the variance of LS around its mean, not $\boldsymbol{\beta}$)

Asy.Var[2SLS] $\geq$ Asy.Var[LS] in the matrix sense.

Compare inverses:

$\{\text{Asy.Var[LS]}\}^{-1} - \{\text{Asy.Var[2SLS]}\}^{-1} = (1/\sigma^2)[\mathbf{X}'\mathbf{X} - \hat{\mathbf{X}}'\hat{\mathbf{X}}]$

$= (1/\sigma^2)[\mathbf{X}'\mathbf{X} - \mathbf{X}'(\mathbf{I} - \mathbf{M}_Z)\mathbf{X}] = (1/\sigma^2)[\mathbf{X}'\mathbf{M}_Z\mathbf{X}]$

This matrix is nonnegative definite. (Not positive definite as it might have some rows and columns which are zero.)

Implication for "precision" of 2SLS.

The problem of "Weak Instruments"

Asymptotic Efficiency (Greene)

• The variance is larger than that of OLS. (A large sample type of Gauss-Markov result is at work.)

(1) OLS is inconsistent.

(2) Mean squared error is uncertain:

MSE[estimator | $\boldsymbol{\beta}$] = Variance + square of bias.

For a long time, IV was thought to be "*the cure*" to the biases introduced by OLS. But, in terms of MSE, IV may be better or worse. It depends on the data: **X**, **Z** and $\boldsymbol{\varepsilon}$.

A Popular Misconception (Greene)

• A popular misconception. If only one variable in **X** is correlated with $\boldsymbol{\varepsilon}$, the other coefficients are consistently estimated. False.

Suppose only the first variable is correlated with $\boldsymbol{\varepsilon}$

Under the assumptions, $\text{plim}(\mathbf{X}'\boldsymbol{\varepsilon}/n) = \begin{pmatrix} \sigma_{1\varepsilon} \\ 0 \\ \cdots \\ . \end{pmatrix}$. Then

$$\text{plim } \mathbf{b} - \boldsymbol{\beta} = \text{plim}(\mathbf{X}'\mathbf{X}/n)^{-1} \begin{pmatrix} \sigma_{1\varepsilon} \\ 0 \\ \cdots \\ . \end{pmatrix} = \sigma_{1\varepsilon} \begin{pmatrix} q^{11} \\ q^{21} \\ \cdots \\ q^{K1} \end{pmatrix}$$

$= \sigma_{1\varepsilon}$ times the first column of $\mathbf{Q}^{-1}$

The problem is "smeared" over the other coefficients.

Small sample properties of IV

• Now, we do not have the condition E[**ε**|**X**] = 0, we cannot get simple expressions for the moments of $\mathbf{b}_{2SLS}$:

$$\mathbf{b}_{2SLS} = \beta + [\mathbf{Y'}\,\mathbf{P_z}\,\mathbf{Y}]^{-1}\,\mathbf{Y'}\,\mathbf{P_z}\,\boldsymbol{\varepsilon}$$

by first taking expectations of conditioned on **X** and **Z**. The bias:

$$\mathbf{b}_{2SLS} - \beta = [\mathbf{Y'}\,\mathbf{P_z}\,\mathbf{Y}]^{-1}\,\mathbf{Y'}\,\mathbf{P_z}\,\boldsymbol{\varepsilon}$$

• We cannot say that $\mathbf{b}_{2SLS}$ is unbiased (even when Cov(**z, ε**) = 0!), or that it has the Var[$\mathbf{b}_{2SLS}$] equal to its Asy Var[$\mathbf{b}_{2SLS}$].

• Also, recall that the 1st stage introduces a source of finite sample bias: the estimation of **Π**.

• In fact, $\mathbf{b}_{2SLS}$ can have very bad small-sample properties.

Small sample properties of IV - Simulation

• To study the behavior of $\mathbf{b}_{IV}$, for small *T*, we set up a simple Monte Carlo experiment using a model appropriate to the context.

• Recall the asymptotic distribution of $\mathbf{b}_{IV}$

$$\sqrt{T}\left(b_{IV} - \beta\right) \xrightarrow{d} N\left(0, \frac{\sigma_\varepsilon^2}{\sigma_X^2} \times \frac{1}{\rho_{XZ}^2}\right)$$

• We will see that the small sample behavior of $\mathbf{b}_{IV}$ will depend on the nature of the model, the correlation between ***X*** and ***ε***, and the correlation between ***X*** and ***Z***.

16

Small sample properties of IV – Simulation

• We start with a simple linear model:

$$Y = \beta_1 + \beta_2 X + \varepsilon$$
$$X = \pi_1 Z + \pi_2 U + \varepsilon$$

with observations on ***Z***, ***U***, and ***ε*** are drawn independently from a N(0,1). We think of ***Z*** and ***U*** as variables and of ***ε*** as the error term in the model. π_1 and π_2 are constants.

• By construction, ***X*** is not independent of ***ε***. OLS is inconsistent and its standard errors and tests will be invalid.

• ***Z*** is correlated with ***X***, but independent of ***ε***. It serves as an instrument. (***U*** is included to provide some variation in ***X***, not connected with either ***Z*** or ***ε***.)

17

Small sample properties of IV – Simulation

• To start the simulation, we set:

$$\beta_1 = 10,\ \beta_2 = 5,\ \pi_1 = 0.5, \text{ and } \pi_2 = 2.0.$$

• That is,

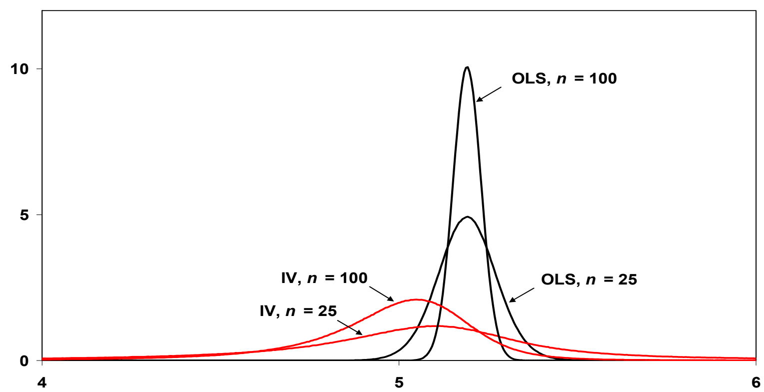
$$Y = 10 + 5X + \varepsilon \qquad \varepsilon \sim iid\ N(0,1)$$
$$X = 0.5\,Z + 2.0\,U + \varepsilon \qquad Z \sim iid\ N(0,1); \quad U \sim iid\ N(0,1)$$

• It is easy to check that plim $b_{2,\text{OLS}}$ = 5.19 (=5+1/(.5²+2²+1)). Of course, plim $b_{2,\text{IV}}$ = 5.00. We draw *n*=25, 100 & 3,200. We do 1 million simulations.

Sample Size	$b_{2,\text{OLS}}$ (SE[$b_{2,\text{OLS}}$])	$b_{2,\text{IV}}$ (SE[$b_{2,\text{IV}}$])	MSEs
n = 25	5.190 (0.080)	4.998 (0.137)	.055 - .019
n = 100	5.191 (0.040)	5.000 (0.054)	.038 - .003
n = 3200	5.191 (0.007)	5.000 (0.009)	.036 - .0001

20

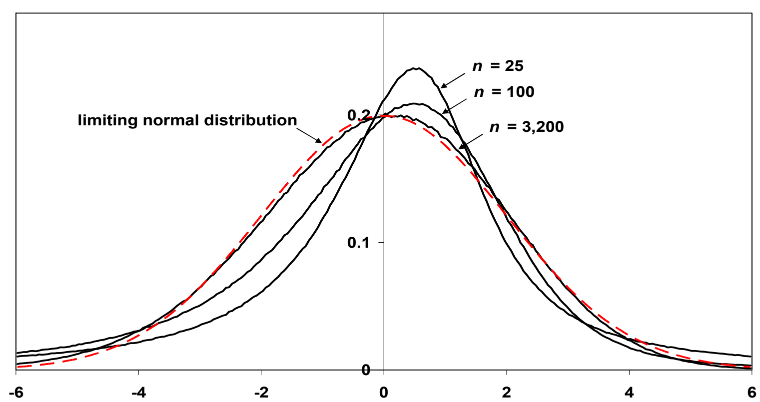
Small sample properties of IV – Simulation



• $b_{2,IV}$ has a greater variance than $b_{2,OLS}$. For small samples (say, $n = 25$ or 100) OLS may be better in terms of MSE. But, as n grows, $b_{2,IV}$ and $b_{2,OLS}$ tend to their plims ($b_{2,IV}$ more slowly than $b_{2,OLS}$, because it has a larger variance).

21

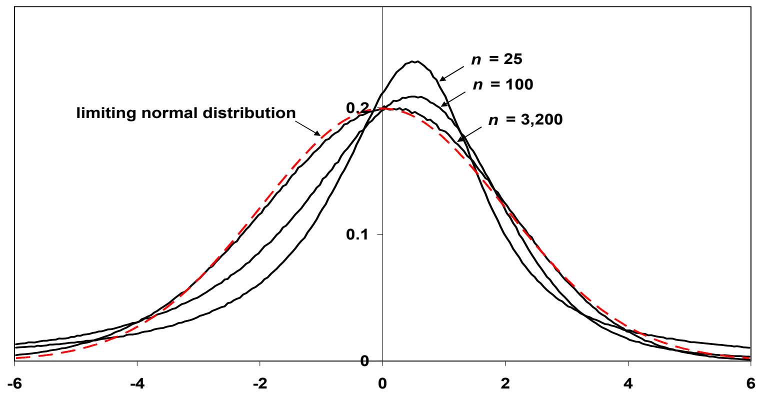
Small sample properties of IV – Simulation



• We have the distribution of $\sqrt{n}\,(b_{2,IV} - b_2)$ for $n = 25$, 100, and 3,200. It also shows, as the dashed red line, the limiting normal distribution. For $n = 3{,}200$ is very close to the limiting distribution. Inference would be OK with samples of this magnitude.

24

Small sample properties of IV – Simulation



• For n=25, 100, the tail are *too fat*. Inference would give rise to excess instances of Type I error (under rejection). The distortion for small sample sizes is partly attributable to the low ρ_{xz}=corr(**X**,**Z**)=0.22 (=.5/sqrt(5.25)) (or *weak instruments;* common in IV estimation).

24

Small sample properties of IV – Simulation

• To check the effect of ρ_{xz} on the estimation, we lower π_1 to 0.1, which brings ρ_{xz} = corr(**X**,**Z**) = 0.0045 (=.1/sqrt(5.01)) and we increase π_1 to 4, with ρ_{xz} = 0.8729 =(4/sqrt(21)).

	Empirical Type I Error		
Sample Size	ρ = .8729	ρ = .2182	ρ = .0045
n=25	0.0698	0.0717	0.0752
n = 100	0.0610	0.0628	0.0628
n = 3200	0.0511	0.0508	0.0527

• Some size problem for small n. Low ρ_{xz} slightly increases the size problem.

24

Cornwell and Rupert Data (Greene)

Cornwell and Rupert Returns to Schooling Data, 595 Individuals, 7 Years
Variables in the file are

LWAGE = log of wage = **dependent variable in regressions**

EXP = work experience
WKS = weeks worked
OCC = occupation, 1 if blue collar,
IND = 1 if manufacturing industry
SOUTH = 1 if resides in south
SMSA = 1 if resides in a city (SMSA)
MS = 1 if married
FEM = 1 if female
UNION = 1 if wage set by union contract
ED = years of education
BLK = 1 if individual is black

These data were analyzed in Cornwell, C. and Rupert, P., "Efficient Estimation with Panel Data: An Empirical Comparison of Instrumental Variable Estimators," Journal of Applied Econometrics, 3, 1988, pp. 149-155. See Baltagi, page 122 for further analysis. The data were downloaded from the website for Baltagi's text.

Application: Wage Equation (Greene)

- Are earnings affected by education? In a linear regression, we expect the education coefficient to be positive (and significant, if human capital theory is correct).

- Linear regression model:

  logWage = $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

  X = one, exp, occ, ed (education), wks

  - We expect Wks -weeks worked- to be endogenous

  - Instruments: **Z** = **one, exp, occ, ed, ind, south, smsa, ms, fem**

- Q: How do we know when a variable is exogenous?

Estimated Wage Equation (Greene)

```
+----------------------------------------------------+
| Ordinary    least squares regression               |
+----------------------------------------------------+
+--------+-------------+--------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+-------------+--------------+--------+--------+----------+
|Constant|    5.30277***       .07406       71.605   .0000          |
|EXP     |     .01294***       .00058       22.393   .0000   19.8538|
|OCC     |    -.08511***       .01575       -5.403   .0000    .51116|
|ED      |     .06694***       .00288       23.204   .0000   12.8454|
|WKS     |     .00641***       .00120        5.330   .0000   46.8115|
+--------+---------------------------------------------------------+
+----------------------------------------------------+
| Two stage   least squares regression               |
+----------------------------------------------------+
+------------------------------------------------------------------+
|Instrumental Variables:                                           |
|ONE       EXP       OCC       ED        IND       SOUTH     SMSA   |
|MS        FEM                                                     |
+------------------------------------------------------------------+
|Constant|   -6.60400***      1.81742       -3.634   .0003          |
|EXP     |     .01735***       .00205        8.457   .0000   19.8538|
|OCC     |    -.04375          .05325        -.822   .4113    .51116|
|ED      |     .07840***       .00984        7.968   .0000   12.8454|
|WKS     |     .25530***       .03785        6.745   .0000   46.8115|
+--------+---------------------------------------------------------+
```

Endogeneity Test (Hausman)

	Exogenous	Endogenous
OLS	Consistent, Efficient	Inconsistent
2SLS	Consistent, Inefficient	Consistent

• Base a test on $\mathbf{d} = \mathbf{b}_{2SLS} - \mathbf{b}_{OLS}$
 - We can use a Wald statistic: $\mathbf{d}'[\text{Var}(\mathbf{d})]^{-1}\mathbf{d}$

Note: Under H_0 (plim $(\mathbf{X}'\varepsilon/T) = 0$) $\mathbf{b}_{OLS} = \mathbf{b}_{2SLS} = \mathbf{b}$
 - Also, under H_0: $\text{Var}[\mathbf{b}_{2SLS}] = \mathbf{V}_{2SLS} > \text{Var}[\mathbf{b}_{OLS}] = \mathbf{V}_{OLS}$
 $\Rightarrow$ Under H_0, one estimator is efficient, the other one is not.

• Q: What to use for Var(**d**)?
 - Hausman (1978): $\mathbf{V} = \text{Var}(\mathbf{d}) = \mathbf{V}_{2SLS} - \mathbf{V}_{OLS}$

$$H = (\mathbf{b}_{2SLS} - \mathbf{b}_{OLS})'[\mathbf{V}_{2SLS} - \mathbf{V}_{OLS}]^{-1}(\mathbf{b}_{2SLS} - \mathbf{b}_{OLS}) \xrightarrow{d} \chi^2_{\text{rank}(\mathbf{V})}$$

Endogeneity Test (Hausman)

Q: What to use for Var($\mathbf{d}$)?

- Hausman (1978): $\mathbf{V}$ = Var($\mathbf{d}$) = $\mathbf{V}_{2SLS} - \mathbf{V}_{OLS}$

$$H = (\mathbf{b}_{2SLS} - \mathbf{b}_{OLS})'[\mathbf{V}_{2SLS} - \mathbf{V}_{OLS}]^{-1}(\mathbf{b}_{2SLS} - \mathbf{b}_{OLS})$$

• Hausman gets Var($\mathbf{d}$) by using the following result:

"The covariance between an efficient estimator ($\mathbf{b}_E$) *and its difference from an inefficient estimator* ($\mathbf{b}_E - \mathbf{b}_I$) *is zero."* That is,

$$\text{Cov}(\mathbf{b}_E, \mathbf{b}_E - \mathbf{b}_I) = \text{Cov}(\mathbf{b}_E, \mathbf{b}_E) - \text{Cov}(\mathbf{b}_E, \mathbf{b}_I)$$

$$= \text{Var}(\mathbf{b}_E) - \text{Cov}(\mathbf{b}_E, \mathbf{b}_I) = 0$$

$$\Rightarrow \text{Var}(\mathbf{b}_E) = \text{Cov}(\mathbf{b}_E, \mathbf{b}_I)$$

• Hausman's case: aVar($\mathbf{b}_{OLS}$) = aCov ($\mathbf{b}_{OLS}$, $\mathbf{b}_{2SLS}$)

Then, aVar($\mathbf{d}$) = aVar($\mathbf{b}_{OLS}$) + aVar($\mathbf{b}_{2SLS}$) - 2 aCov ($\mathbf{b}_{OLS}$, $\mathbf{b}_{2SLS}$)

= aVar($\mathbf{b}_{2SLS}$) – aVar($\mathbf{b}_{OLS}$)

Endogeneity Test (Hausman)

• $H = (\mathbf{b}_{2SLS} - \mathbf{b}_{OLS})'\mathbf{V}^{-1}(\mathbf{b}_{2SLS} - \mathbf{b}_{OLS})$

where $\mathbf{V} = \mathbf{V}_{2SLS} - \mathbf{V}_{OLS}$.

• There are different variations of H, depending on which estimator of $\mathbf{V}$ is used. Using $\mathbf{V}[\mathbf{b}_{OLS}]$ and $\mathbf{V}[\mathbf{b}_{2SLS}]$ can create problems in small sample ($\mathbf{V}$ may not be pd).

• There are a couple of solutions to this problem, for example, imposing a common estimate of σ. If we use s^2, the OLS estimator, we have Durbin's (1954) version of the test.

Endogeneity Test: The Wu Test

• The Hausman test has some computation issues.

• Simplification: The Wu test.

• Consider a regression **y** = **X**β + ε, an array of proper instruments **Z**, and an array of instruments **W** that includes **Z** plus other variables that may be either clean or contaminated.

• Wu test for H_0: **X** is clean. Setup:

(1) Regress **X** on **Z** (first stage). Keep fitted values $\hat{X} = \mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z'X}$

(2) Using **W** as instruments, do a 2SLS regression of **y** on **X**, keep RSS_1.

(3) Do a 2SLS regression of **y** on **X** and a subset of *m* columns of $\hat{X}$ that are linearly independent of **X**. Keep RSS_2.

(4) Do an *F*-test: $F = [(RSS_1 - RSS_2)/m]/[RSS_2/(T-k)]$.

Endogeneity Test: The Wu Test

• Under H_0: **X** is clean, the *F* statistic has an approximate $F_{m,T-k}$ distribution.

• The test can be interpreted as a test for whether the *m* auxiliary variables from $\hat{X}$ should be omitted from the regression.

• When a subset of $\hat{X}$ of maximum possible rank is chosen, this statistic turns out to be asymptotically equivalent to the Hausman test statistic.

Note: If **W** contains **X**, then the 2SLS in the second and third steps reduces to OLS.

Endogeneity Test: The Wu Test

Note: If **W** contains **X**, then the 2SLS in the second and third steps reduces to OLS.

Davidson and MacKinnon (1993) point out that the DWH test really tests whether possible endogeneity of the right-hand-side variables not contained in the instruments makes any difference to the coefficient estimates.

• These types of exogeneity tests are usually known as DWH (Durbin, Wu, Hausman) tests.

Endogeneity Test: Augmented DWH Test

• Davidson and MacKinnon (1993) suggest an augmented regression test (DWH test), by including the residuals of each endogenous right-hand side variable.

• Model: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}\boldsymbol{\gamma} + \boldsymbol{\varepsilon},$ we suspect **X** is endogenous.

• Steps for augmented regression DWH test:

1. Regress **x** on IV (**Z**) and **U**:

$$\mathbf{x} = \mathbf{Z}\,\boldsymbol{\Pi} + \mathbf{U}\,\boldsymbol{\varphi} + \boldsymbol{\upsilon} \quad \Rightarrow \text{save residuals } \mathbf{v_x}$$

2. Do an augmented regression: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}\boldsymbol{\gamma} + \mathbf{v_x}\,\delta + \boldsymbol{\varepsilon}$
3. Do a *t-test* of δ. If the estimate of δ, say **d,** is significantly different from zero, then OLS is not consistent.

Endogeneity Test: Augmented DWH Test

Intuition: Since each instrument, $\mathbf{Z}$, is uncorrelated with $\boldsymbol{\varepsilon}$, $\mathbf{x}$ is uncorrelated with $\boldsymbol{\varepsilon}$ only if $\mathbf{v}_\mathbf{x}$ is uncorrelated with $\boldsymbol{\varepsilon}$. Then, the DWH tests becomes

$$H_0: E[\mathbf{v}_\mathbf{x}\, \boldsymbol{\varepsilon}] = 0.$$

• This is the most popular version of the DWH test.

Implication of DWH: Reject $H_0 \Rightarrow$ OLS is inconsistent. IV results should be preferred (the rest of lecture puts some breaks to this implication!)

Wu Test (Greene)

```
+----------------------------------------------------+
| Ordinary    least squares regression                |
| LHS=LWAGE    Mean                  =    6.676346    |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
|Constant|   -6.60400***       .50833      -12.992   .0000            |
|EXP     |     .01735***       .00057       30.235   .0000     19.8538|
|OCC     |    -.04375***       .01489       -2.937   .0033      .51116|
|ED      |     .07840***       .00275       28.489   .0000     12.8454|
|WKS     |     .00355***       .00114        3.120   .0018     46.8115|
|WKSHAT  |     .25176***       .01065       23.646   .0000     46.8115|
+--------+------------------------------------------------------------+
| Note: ***, **, * = Significance at 1%, 5%, 10% level.               |
+---------------------------------------------------------------------+

--> Calc     ; list ; Wutest = b(kreg)^2 / Varb(kreg,kreg) $
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 WUTEST  =     559.119128        (=23.646^2)      => OLS is inconsistent!
```

Measurement Error

• DGP: $\mathbf{y}^* = \beta\mathbf{x}^* + \boldsymbol{\varepsilon}$ - $\boldsymbol{\varepsilon} \sim iid\, D(\mathbf{0}, \sigma_\varepsilon^2)$
- all of the CLM assumptions apply.

Problem: $\mathbf{x}^*$, $\mathbf{y}^*$ are not observed or measured correctly. **x, y** are observed:

$\mathbf{x} = \mathbf{x}^* + \mathbf{u}$ $\mathbf{u} \sim iid\, D(\mathbf{0}, \sigma_u^2)$ -no correlation to $\boldsymbol{\varepsilon}$,**v**

$\mathbf{y} = \mathbf{y}^* + \mathbf{v}$ $\mathbf{v} \sim iid\, D(\mathbf{0}, \sigma_v^2)$ -no correlation to $\boldsymbol{\varepsilon}$,**u**

• Let's consider two cases:

- **CASE 1** - Only $\mathbf{x}^*$ is measured with error ($\mathbf{y}=\mathbf{y}^*$).

- **CASE 2** - Only $\mathbf{y}^*$ is measured with error ($\mathbf{x}=\mathbf{x}^*$).

Measurement Error

CASE 1 - Only $\mathbf{x}^*$ is measured with error.

$$\mathbf{y} = \mathbf{y}^* = \beta\mathbf{x}^* + \boldsymbol{\varepsilon}$$
$$\mathbf{y} = \beta(\mathbf{x} - \mathbf{u}) + \boldsymbol{\varepsilon} = \beta\mathbf{x} + \boldsymbol{\varepsilon} - \beta\mathbf{u} = \beta\mathbf{x} + \mathbf{w}$$
$$E[\mathbf{x'w}] = E[(\mathbf{x}^* + \mathbf{u})'(\boldsymbol{\varepsilon} - \beta\mathbf{u})] = -\beta\sigma_u^2 \neq 0$$

$\Rightarrow$ OLS biased & inconsistent. We need IV!

• Typical IV solution: Find another noisy measure of $\mathbf{x}^*$, say $\mathbf{z}$:

$\mathbf{z} = \mathbf{x}^* + \boldsymbol{\eta}$ $\boldsymbol{\eta} \sim iid\, D(\mathbf{0}, \sigma_w^2)$ -no correlation to $\boldsymbol{\varepsilon}$, **v**, **u**

Check IV conditions:

- $Cov(\mathbf{z}, \boldsymbol{\varepsilon}) = Cov(\mathbf{x}^* + \boldsymbol{\eta}, \boldsymbol{\varepsilon}) = 0$
- $Cov(\mathbf{z}, \mathbf{x}) = Cov(\mathbf{x}^* + \boldsymbol{\eta}, \mathbf{x}^* + \mathbf{u}) = Var(\mathbf{x}^*) \neq 0$

Then,

$$\mathbf{b}_{IV} = Cov(\mathbf{z}, \mathbf{x})^{-1}\, Cov(\mathbf{z}, \mathbf{y}) = Var(\mathbf{x}^*)^{-1}\, Cov(\mathbf{z}, \mathbf{y})$$

$\Rightarrow$ IV removes the variance in noise.

Measurement Error

• Q: What happens when OLS is used –i.e., we regress **y** on **x**?

A: Least squares attenuation:

$$\text{plim } b = \frac{\text{cov}(x,y)}{\text{var}(x)} = \frac{\text{cov}(x^* + u, \beta x^* + \varepsilon)}{\text{var}(x^* + u)}$$

$$= \frac{\beta \text{ var}(x^*)}{\text{var}(x^*) + \text{var}(u)} < \beta$$

• Q: Why is OLS attenuated?

$\mathbf{y} = \beta\mathbf{x}^* + \boldsymbol{\varepsilon}$

$\mathbf{x} = \mathbf{x}^* + \mathbf{u}$

$\mathbf{y} = \beta\mathbf{x} + (\boldsymbol{\varepsilon} - \beta\mathbf{u}) = \beta\mathbf{x} + \mathbf{v}$, $\quad$ $\text{cov}(\mathbf{x}, \mathbf{v}) = -\beta \text{ var}(\sigma_u^2)$

Some of the variation in **x** is not associated with variation in **y**. The effect of variation in **x** on **y** is dampened by the measurement error.

Measurement Error

CASE 2 - Only **y*** is measured with error.

$\mathbf{y}^* = \mathbf{y} - \mathbf{v} = \beta\mathbf{x}^* + \boldsymbol{\varepsilon} = \beta\mathbf{x} + \boldsymbol{\varepsilon}$

$\Rightarrow \quad \mathbf{y} = \beta\mathbf{x} + \boldsymbol{\varepsilon} + \mathbf{v} = \beta\mathbf{x} + (\boldsymbol{\varepsilon} + \mathbf{v})$

• Q: What happens when **y** is regressed on **x**?

A: Nothing! We have our usual OLS problem since $\boldsymbol{\varepsilon}$ and **v** are independent of each other and $\mathbf{x}^*$. CLM assumptions are not violated!

• Q: Is measurement error in finance/economics a problem?

A: Yes! In surveys and forms, mistakes are common. Most relevant problem: often, economic theories deal with unobservables (**x***).

<u>Famous unobservables</u>: Market portfolio, innovation, growth opportunities, potential output, target debt-equity ratio, business cycles, worker's skills.

Measurement Error: Proxy Variables

• Often, economic theories deal with unobservables ($\mathbf{x}^*$). To test these theories, practitioners use a *proxy* ($\mathbf{x}$), instead of $\mathbf{x}^*$.

A proxy is a variable that has a "close" relation (usually, linear) with the unobservable:

$\mathbf{x} = \delta\, \mathbf{x}^* + \mathbf{u}$ (typical measurement error problem!)

Example: The CAPM: $E[\mathbf{R}_i - \mathbf{R}_f] = \beta_i\, E[\mathbf{R}_{MP} - \mathbf{R}_f]$

The market portfolio (MP) is unobservable. According to Roll's (1977) critique, this makes the CAPM untestable!

In practice, we proxy it by a representative stock market index:

$\mathbf{R}_{Index} = \delta\, \mathbf{R}_{MP} + \mathbf{u}$

Measurement Error: Proxy Variables

Example: Testing the CAPM I.

(1) CAPM regression:

$\mathbf{R}_i - \mathbf{R}_f = \alpha_i + \beta_i\, (\mathbf{R}_{MP} - \mathbf{R}_f) + \boldsymbol{\varepsilon}$

H_0: α_i=0 (α_i is the pricing error. Jensen's alpha.)

(2) MP unobservable. Proxy: S&P 500 stock market index

$\mathbf{R}_{SP500} = \eta\, \mathbf{R}_{MP} + \mathbf{u}$ $\Rightarrow \mathbf{R}_{MP} = \theta\, \mathbf{R}_{SP500} + \mathbf{u'}$

(3) Working CAPM regression

$\mathbf{R}_i - \mathbf{R}_f = \alpha_i + \beta_i\, [(\theta\, \mathbf{R}_{SP500} + \mathbf{u'}) - \mathbf{R}_f] + \boldsymbol{\varepsilon}$

$= \alpha_i + \beta_i\theta\, \mathbf{R}_{SP500} - \beta_i\, \mathbf{R}_f + \boldsymbol{\xi}$ ($\boldsymbol{\xi} = \beta_i\, \mathbf{u'} + \boldsymbol{\varepsilon}$)

Or, $\mathbf{R}_i = \alpha_i + \delta_i\, \mathbf{R}_f + \gamma_i\, \mathbf{R}_{SP500} + \boldsymbol{\xi}$

where $\gamma_i = \beta_i\theta$ and $\delta_i = 1 - \beta_i$

$\Rightarrow \alpha_i$ can be estimated directly, but β_i cannot be estimated directly! H_0 can be tested. (In general, *smearing* complicates the estimation.)

Measurement Error: Proxy Variables

- $\mathbf{R}_i = \alpha_i + \delta_i\, \mathbf{R}_f + \gamma_i\, \mathbf{R}_{SP500} + \xi$ $\qquad (\xi = \beta_i\, \mathbf{u'} + \boldsymbol{\varepsilon})$

(4) Usually, $\mathbf{R}_f$ is assumed constant

$\mathbf{R}_i = \alpha_i^* + \gamma_i\, \mathbf{R}_{SP500} + \xi$

where $\alpha_i^* = \alpha_i + \delta_i\, \mathbf{R}_f$

$$\mathbf{R}_i = \alpha_i^* + \gamma_i\, \mathbf{R}_{SP500} + \xi$$

We can do a regression to estimate α_i^* and γ_i. H_0 can be tested.

But since $\gamma_i = \beta_i\theta \quad \Rightarrow \beta_i$ cannot be estimated!

Note: It is common to just work with "excess returns" directly. In this case, the proxy would be:

$$\mathbf{R}_{SP500} - \mathbf{R}_f = \eta\, (\mathbf{R}_{MP} - \mathbf{R}_f) + \mathbf{u}$$

Measurement Error: Proxy Variables

Example: Testing the CAPM II. We extend the CAPM (APT style):

(1) CAPM regression with more explanatory variables (**W**):

$$\mathbf{R}_i - \mathbf{R}_f = \alpha_i + \beta_i\, (\mathbf{R}_{MP} - \mathbf{R}_f) + \psi_i\, \mathbf{W} + \boldsymbol{\varepsilon}$$

$$H_0\text{: } \psi_i = 0$$

(2) MP unobservable. Proxy: S&P 500 stock market index

$$\mathbf{R}_{SP500} = \eta\, \mathbf{R}_{MP} + \mathbf{u} \qquad \Rightarrow \mathbf{R}_{MP} = \theta\, \mathbf{R}_{SP500} + \mathbf{u}^*$$

(3) Working CAPM extended regression

$$\mathbf{R}_i = \alpha_i + \delta_i\, \mathbf{R}_f + \gamma_i\, \mathbf{R}_{SP500} + \psi_i\, \mathbf{W} + \xi \qquad (\xi = \beta_i\, \mathbf{u}^* + \boldsymbol{\varepsilon})$$

OLS estimates α_i, δ_i $(=1 - \beta_i)$, $\gamma_i(= \theta\, \beta_i)$, ψ_i (but, not β_i directly!). H_0 can be tested. (In general, *smearing* complicates the estimation.)

Note: Assuming a constant $\mathbf{R}_f$, we get estimates α_i^*, γ_i, ψ_i.

Measurement Error: Smearing Again (Greene)

Multiple regression: $y = \beta_1 x_1^* + \beta_2 x_2^* + \varepsilon$
x_1^* is measured with error; $x_1 = x_1^* + u$
x_2 is measured with out error.
The regression is estimated by least squares
Popular myth #1. b_1 is biased downward, b_2 consistent.
Popular myth #2. All coefficients are biased toward zero.
Result for the simplest case. Let
$\sigma_{ij} = \text{cov}(x_i^*, x_j^*), i, j = 1,2$ (2x2 covariance matrix)
σ^{ij} = ijth element of the inverse of the covariance matrix
$\theta^2 = \text{var}(u)$
For the least squares estimators:

$$\text{plim } b_1 = \beta_1 \left(\frac{1}{1 + \theta^2 \sigma^{11}} \right), \quad \text{plim } b_2 = \beta_2 - \beta_1 \left(\frac{\theta^2 \sigma^{12}}{1 + \theta^2 \sigma^{11}} \right)$$

The effect is called "smearing."

Measurement Error: Twinsville

• Q: Does education affects earnings?

A: We expect two people with similar natural abilities but different levels of education to be differently paid. To estimate returns-to-schooling, economists often use a linear regression model relating log earnings (**y**) to years of education (**x***), with additional control variables (**U**). The error term represents the effects of person-to-person variation that have not been controlled for. The DGP:

$$\mathbf{y} = \beta \mathbf{x}^* + \mathbf{U}\gamma + \varepsilon$$

• We expect two people with similar natural abilities:

⇒ More education, more earnings. We expect $\beta > 0$.

• <u>Problem</u>: **x*** is usually self-reported, and often reported with error.

Measurement Error: Twinsville

• Linear model: $\mathbf{y} = \beta\mathbf{x} + \mathbf{U}\gamma + \varepsilon$

• H_0: $\beta = 0$.

• We do not observe **x***, we observe self-reported **x**. We need to find an instrument to estimate the model.

• Famous application from the econ literature: Ashenfelter/Kreuger (AER,1994): A wage equation for twins that includes two measures of **x**: each twin reports their own and their twin's schooling.

• The data suggests that between 8% and 12% of the measured variance in schooling levels is error.

• <u>Instrument</u>: Reported schooling by the twin.

Measurement Error: Twinsville

TABLE 1–DESCRIPTIVE STATISTICS

Variable	Means (standard deviations in parentheses)		
	Identical twins[a]	Fraternal twins[a]	Population[b]
Self-reported education	14.11 (2.16)	13.72 (2.01)	13.14 (2.73)
Sibling-reported education	14.02 (2.14)	13.41 (2.07)	—
Hourly wage	\$13.31 (11.19)	\$12.07 (5.40)	\$11.10 (7.41)
Age	36.56 (10.36)	35.59 (8.29)	38.91 (12.53)
White	0.94 (0.24)	0.93 (0.25)	0.87 (0.34)
Female	0.54 (0.50)	0.48 (0.50)	0.45 (0.50)

Measurement Error: Twinsville

THE AMERICAN ECONOMIC REVIEW DECEMBER 1994

TABLE 3—ORDINARY LEAST-SQUARES (OLS), GENERALIZED LEAST-SQUARES (GLS), INSTRUMENTAL-VARIABLES (IV), AND FIXED-EFFECTS ESTIMATES OF LOG WAGE EQUATIONS FOR IDENTICAL TWINS[a]

Variable	OLS (i)	GLS (ii)	GLS (iii)	IV[a] (iv)	First difference (v)	First difference by IV (vi)
Own education	0.084 (0.014)	0.087 (0.015)	0.088 (0.015)	0.116 (0.030)	0.092 (0.024)	0.167 (0.043)
Sibling's education	—	—	−0.007 (0.015)	−0.037 (0.029)	—	—
Age	0.088 (0.019)	0.090 (0.023)	0.090 (0.023)	0.088 (0.019)	—	—
Age squared (÷100)	−0.087 (0.023)	−0.089 (0.028)	−0.090 (0.029)	−0.087 (0.024)	—	—
Male	0.204 (0.063)	0.204 (0.077)	0.206 (0.077)	0.206 (0.064)	—	—
White	−0.410 (0.127)	−0.417 (0.143)	−0.424 (0.144)	−0.428 (0.128)	—	—
Sample size:	298	298	298	298	149	149
R^2:	0.260	0.219	0.219	—	0.092	—

Notes: Each equation also includes an intercept term. Numbers in parentheses are estimated standard errors.
[a]Own education and sibling's education are instrumented for using each sibling's report of the other sibling's education as instruments.

Omitted Variables: IV Conditions (Again)

• The omitted variables problem is, probably, the most popular IV application in microeconomics and Corporate Finance.

• Typical omitted variables situation: In the CEO compensation model, we want to test the causal effect of networking on compensation, but a CEO's unobserved skills blurs the causality, since Cov($\mathbf{x}$, $\boldsymbol{\varepsilon}$) ≠ 0.

• Recall that IV estimators are consistent if the instruments, $\mathbf{Z}$, used are both valid and relevant/informative. That is, we look for $\mathbf{Z}$ such that

(1) Cov($\mathbf{X}$, $\mathbf{Z}$) ≠ $\mathbf{0}$ - *relevance condition*

(2) Cov($\mathbf{Z}$, $\boldsymbol{\varepsilon}$) = $\mathbf{0}$ - *valid condition (exclusion restriction)*

In an omitted variables problem, we can think of (2) as broken into two parts: (a) $\mathbf{Z}$ is uncorrelated to $\boldsymbol{\varepsilon}$, & (b) only affects $\mathbf{y}$ through $\mathbf{X}$.

Omitted Variables: IV Conditions (Again)

• (2) ⇒ **Z** is not only uncorrelated to **ε**, but only affects **y** through **X** – after all, it is excluded from structural equation!

$$\mathbf{Z} \rightarrow \mathbf{X} \rightarrow \mathbf{y}$$

From the 2nd part: Once I know the effect of **Z** on **X**, I can throw **Z**.

• Historically, the emphasis has been on the *valid* (*exogeneity*) *condition*, Cov(**Z**, **ε**)=**0**.

• But, the past 25 years added an additional source of concern: the Cov(**X**, **Z**) may not be high enough. That is, $\hat{\boldsymbol{X}}$ (from the first stage) may not be very informative about **X**:

$$\mathbf{X} = \mathbf{Z\Pi} + \mathbf{U}\boldsymbol{\delta} + \mathbf{V} \qquad – \mathbf{V} \sim N(0, \sigma_V^2\mathbf{I})$$

Omitted Variables: Finding Good IVs

• Back to CEO compensation model. We need IVs, **Z**, such that

(1) Explain the variation in networking –i.e., Cov(**x**, **Z**) ≠ **0**

(2) Do not directly affect CEO compensation –i.e., Cov(**Z**, **ε**) = **0**.

• It is not difficult to find a **Z** that meets (2), the valid condition. Many variables are not correlated with **ε**, the error term from the CEO compensation structural equation.

Examples: Potential IVs, uncorrelated with **ε**.
Earthquakes in New Zealand; past debt of Denton, TX; asteroids hitting the Atlantic the year of the CEO's birth; number of letters on the name of CEO's high school.

• We like these potential IVs; they look random or orthogonal to a CEO compensation model (unrelated to **ε**). They can be safely excluded from the structural equation. That is, they meet Cov(**Z**, **ε**)=**0**.

Omitted Variables: Finding Good IVs

• But, it is dubious the effect of these IVs on networking, **x**. The relevant condition is likely not met.

Note: Deaton (2010) calls the variables in the examples *external*, since they are not determined by the model. They may not be exogenous.

• The key is to find a **Z** correlated with **X** –i.e., the relevant condition– uncorrelated with $\boldsymbol{\varepsilon}$ (and with the omitted variable, unobserved skills.)

• Starting with Angrist (1990) and Angrist and Krueger (1991) (for us, A&K), who study the effect on earnings of civilian work experience and schooling, respectively, there has been an emphasis on using a **Z** that can be defined by a *natural experiment* when the IV problem is caused by omitted variables.

• Usually, a natural experiment is exogenous to a structural model. Like the previous external examples, the exclusion condition is met.

Omitted Variables: Finding Good IVs

• The key is to find a natural experiment (defining **Z**) that is correlated with **X** and has no direct effect on **y** –the impact on **y** is through **X**.

Z is an exogenous event $\Rightarrow$ resulting values of **X** induced by **Z** may be considered *randomized* –a key feature in lab/medical experiments.

• In a lab experiment (the *gold standard* in experimental sciences), a researcher randomly assigns a treatment to a group, creating two groups: treated and not-treated or control. Then, the researcher studies the effect of the treatment on, say, the group's health.

The key feature is the randomization of the treatment. The lab researcher needs to show that the two groups are comparable along all dimensions relevant for the outcome variable (age, gender, previous health, etc.) except the one involving the treatment.

Omitted Variables: Finding Good IVs

• Recall that in the CEO compensation model, we want to test the causal effect of networking on compensation, but an omitted variable – the CEO's unobserved skills– creates endogeneity.

• A solution to the omitted variables problem is to assign networking (**x**) randomly: we have two similar groups of CEOs (with similar skills!) and randomly we assign them values (say, large network & small network).

• Of course, this randomized experiment is not possible.

• But, suppose we have a natural event, **Z**, unrelated to CEO compensation, which randomly assigns networking, **x,** to two groups. Then, we can test causality, without the omitted variable problem.

Omitted Variables: Natural Experiments

• We need to find a natural event, **Z**, unrelated to **y**, which randomly assigns **X** to two groups.

• We use **Z** to identify causality. This is why natural experiments are popular in economics & finance (especially, in Corporate Finance).

• We define *natural experiments* as historical (exogenous) episodes that provide observable, *quasi-random* variation in treatment subject to a plausible identifying assumption.

• "Natural" points out that us (the researchers) did not design the episode to be analyzed, but can use it to identify causal relationships.

Omitted Variables: Natural Experiments

• Steps of a natural experiment:

(1) Experiment defines an IV: $\mathbf{Z}_i$=1 (*i* treated), $\mathbf{Z}_i$=0 (*i* control).

(2) Identify two groups:

– treated (all *i* with $\mathbf{Z}_i$=1) with observations: $\mathbf{y}(1)$, $\mathbf{X}(1)$

– control (all *i* with $\mathbf{Z}_i$=0) with observations: $\mathbf{y}(0)$, $\mathbf{X}(0)$

(3) We analyze differences between ($\mathbf{y}(1)$, $\mathbf{X}(1)$) & ($\mathbf{y}(0)$, $\mathbf{X}(0)$).

• Remarks: These steps will be treated like a lab experiment if we show that the treatment is in fact randomly assigned. We need to show that two groups are comparable except for the treatment.

• This is the key for the experiment to be valid. We need to convince the audience that the we have a *quasi-random* treatment.

Finding Good IVs: Natural Experiments

Example: There are significant persistent differences in development ($\mathbf{y}$) among similar cities. One explanation: Location ($\mathbf{x}$); proximity to other cities matter. Cities close to another city enjoy externalities (say, transportation and school networks). We want to test this hypothesis.

I would like to estimate a model: $\mathbf{y} = \beta\mathbf{x} + \mathbf{U}\gamma + \varepsilon$ (but location, $\mathbf{x}$, is also endogenous. OLS will not work!)

- Ideal experiment: Identify 2 similar cities and remove a city next to it.

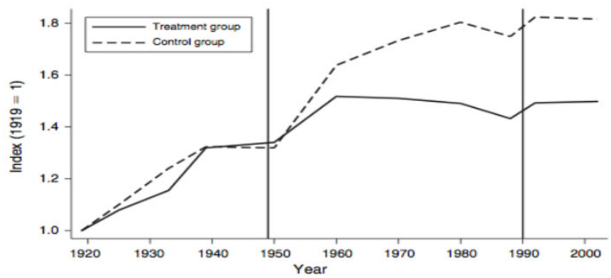
- Natural experiment: German division in 1949.

Cities close to the border lost connection to the cities on the other side of the border. It looks like randomly removing a city! –from Redding & Sturm (2008, AER).

Finding Good IVs: Natural Experiments

• Now, we need to convince the reader that the German division provides a legitimate quasi-random treatment:

We need to show that the partition was exogenous (not based on development, **y** –i.e., unrelated to the structural model!)



- We need to show that there was no cofounding treatment (something else happening along with the partition). For example, after the partition, a city close to the border may be in fear of war.

Finding Good IVs: Natural Experiments

• In the context of Natural Experiments, a good instrument, **Z**, should satisfy:

(1) Explain the variation in **x** –i.e., Cov(**x**, **Z**) ≠ **0**

(2) Do not directly affect **y** –i.e., Cov(**Z**, **ε**) = **0**.

(3) As good as randomly assigned.

• Only condition (1) is the only one we can directly check, through the first-stage regression, where we get $\hat{\boldsymbol{X}}$. Given that **ε** is unobservable, the legitimacy of (2) is usually left for theory or common sense. A researcher should also convince the audience about the validity of (3).

• Finding a **Z** that meets all requirements is not easy.

Finding an Instrument: A&K (1991)

• Back to the question: Does education, **x**, affect earnings, **y**?

We use the same linear model: $\mathbf{y} = \beta\mathbf{x} + \mathbf{U}\gamma + \boldsymbol{\varepsilon}$

We expect $\beta > 0$. But, since Cov($\mathbf{x}$, $\boldsymbol{\varepsilon}$) $\neq \mathbf{0}$; we know that OLS **b** is inconsistent. We need IV estimation.

Note: In general, **U** does not capture much of the variation of **y**.

• Angrist and Krueger's (1991, QJE) Natural experiment: Find an exogenous historical event that creates variation in schooling.

Exogenous event: Compulsory schooling laws according to age, not years of schooling completed.

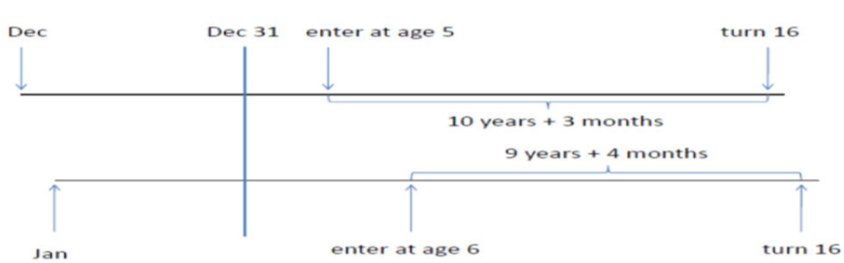
Finding an Instrument: A&K (1991)

• Years of schooling vary by quarter of birth (QOB=**z**):

– In the U.S., it is legal to drop out at 16.

– Someone born in Q1 is a little younger and can drop out with less schooling, than someone born in Q4 $\Rightarrow$ Cov($\mathbf{z}$, $\mathbf{x}$) $\neq$ 0.

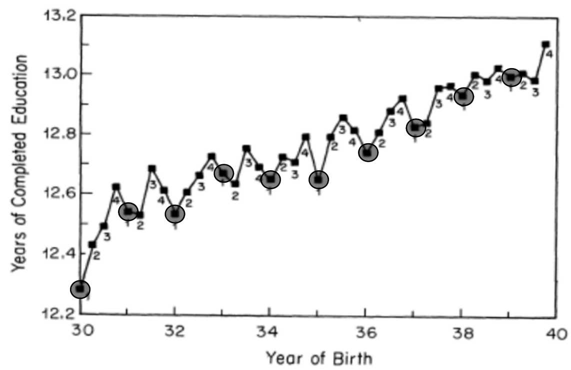
• QOB can be treated as a source of *exogeneity* in schooling, unrelated to individual ability $\Rightarrow$ Cov($\mathbf{z}$, $\boldsymbol{\varepsilon}$) = 0.



• That is, **Z** should affect earnings *only* through its effect on *schooling*.

Finding an Instrument: A&K (1991)

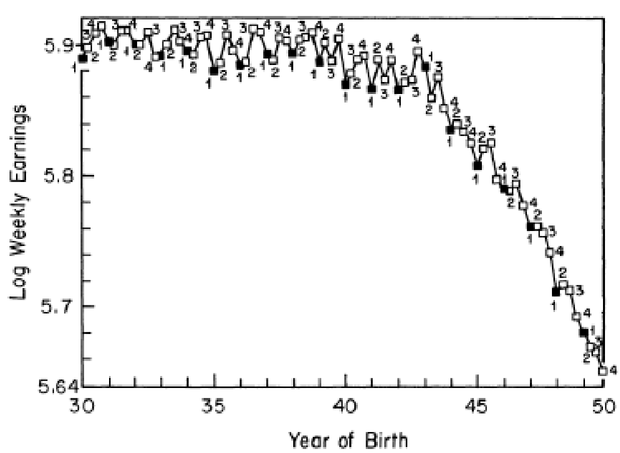
• The data for the 1930-39 cohort show that men born earlier in the year have lower schooling. QOB can be an instrument $\Rightarrow$ there is a first stage: $\mathbf{x} = \pi\, \mathbf{z} + \mathbf{D}\gamma + \mathbf{v}$ ($\mathbf{Z}$: Dummy variable for QOB)



Source: Angrist and Krueger (1991), Figure I

Finding an Instrument: A&K (1991)

- In a reduced form, we can see a relation between earnings and QOB: $\mathbf{y} = \mathbf{Z}\,\boldsymbol{\Gamma} + \mathbf{U}\,\varphi + \xi$



Finding an Instrument: A&K (1991)

- People born in Q1 do obtain less schooling
  - But pay close attention to the scale of the y-axis.
  - Mean differences between Q1 and Q4 are small (in education, the difference is only 0.151 and in log earnings 0.014).

- Thus, we need large T since $R^2_{X,Z}$ will be very small
  - A&K had over 300,000 observations for the 1930-39 cohort

- Final 2SLS model interacted QOB with year of birth (30), state of birth (150):
  - OLS: $\mathbf{b}_{OLS}$ = **.0628** (s.e. = .0003) (large $T \Rightarrow$ small SE's).
  - 2SLS: $\mathbf{b}_{2SLS}$ = **.0811** (s.e. = .0109)
  - Var[$\mathbf{b}_{IV}$] > Var[$\mathbf{b}_{OLS}$], as expected. (But, maybe too large?)

Finding an Instrument: A&K (1991)

TABLE IV
OLS AND TSLS ESTIMATES OF THE RETURN TO EDUCATION FOR MEN BORN 1920–1929: 1970 CENSUS[a]

Independent variable	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS	(5) OLS	(6) TSLS	(7) OLS	(8) TSLS
Years of education	0.0802	0.0769	0.0802	0.1310	0.0701	0.0669	0.0701	0.1007
	(0.0004)	(0.0150)	(0.0004)	(0.0334)	(0.0004)	(0.0151)	(0.0004)	(0.0334)
Race (1 = black)	—	—	—	—	0.2980	−0.3055	−0.2980	−0.2271
					(0.0043)	(0.0353)	(0.0043)	(0.0776)
SMSA (1 = center city)	—	—	—	—	0.1343	0.1362	0.1343	0.1163
					(0.0026)	(0.0092)	(0.0026)	(0.0198)
Married (1 = married)	—	—	—	—	0.2928	0.2941	0.2928	0.2804
					(0.0037)	(0.0072)	(0.0037)	(0.0141)
9 Year-of-birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8 Region of residence dummies	No	No	No	No	Yes	Yes	Yes	Yes
Age	—	—	0.1446	0.1409	—	—	0.1162	0.1170
			(0.0676)	(0.0704)			(0.0652)	(0.0662)
Age-squared	—	—	−0.0015	−0.0014	—	—	−0.0013	−0.0012
			(0.0007)	(0.0008)			(0.0007)	(0.0007)
χ^2 [dof]	—	36.0 [29]	—	25.6 [27]	—	34.2 [29]	—	28.8 [27]

a. Standard errors are in parentheses. Sample size is 247,199. Instruments are a full set of quarter-of-birth times year-of-birth interactions. The sample consists of males born in the United States. The sample is drawn from the State, County, and Neighborhoods 1 percent samples of the 1970 Census (15 percent form). The dependent variable is the log of weekly earnings. Age and age-squared are measured in quarters of years. Each equation also includes an intercept.

- OLS estimate does not appear to be badly biased. But…

IV with a Dummy Variable

• QOB is a dummy IV. It is the *treatment* in the natural experiment. In the simplest model, with one dummy IV, the IV estimator becomes:

$$b_{IV} = \frac{Cov(Z,y)}{Cov(Z,x)} = \frac{\text{indirect effect of z on y}}{\text{indirect effect of z on x}} = \frac{E[y \mid Z=1] - E[y \mid Z=0]}{E[x \mid Z=1] - E[x \mid Z=0]}$$

• This is the *Wald (1944) estimator* (also, *grouping estimator*): A ratio of differences in (group) means (**y** & **x**) in treated and control groups.

• To get the above result, recall

Cov[**Z**,**Y**] = E[**ZY**] – E[**Y**] E[**Z**]

= E[**Y**|**Z**=1] P[**Z**=1] – E[**Y**] P[**Z**=1]

<u>Derivation trick</u>: Y = Y [Z + (1 – Z)]. Then, take expectations on the last term and some algebra delivers:

Cov[**Z**,**Y**] = {E[**Y**|**Z**=1] – E[**Y**|**Z**=0]} P[**Z**=1] P[**Z**=0]

IV with a Dummy Variable

• Similar work for Cov(**X**, **Y**) gets the result: b_{IV} = Wald Estimator.

• Interpretation of Wald Estimator, as a ratio of slopes:

- First stage: $\mathbf{x} = \pi_1 + \pi_2\ \mathbf{z} + \mathbf{v}$
- Reduced form: $\mathbf{y} = \gamma_1 + \gamma_2\ \mathbf{z} + \boldsymbol{\xi}$

Taking conditional expectations on **Z** above and simple algebra:

π_2 = E[**x**|**Z**=1] – E[**x**|**Z**=0]

γ_2 = E[**y**|**Z**=1] – E[**y**|**Z**=0]

Then, $$b_{IV} = \frac{E[y \mid Z=1] - E[y \mid Z=0]}{E[x \mid Z=1] - E[x \mid Z=0]} = \frac{\gamma_2}{\pi_2}$$

• The Wald estimator is known as *local IV* or *local average treatment effect,* LATE (under some assumptions, $\mathbf{b}_{IV}$ = E[**y**(1) – **y**(0)|*compliers*]).

IV with a Dummy Variable

• Application to A&K (1991):

Panel B: Wald Estimates for 1980 Census—Men Born 1930–1939

	(1) Born in 1st quarter of year	(2) Born in 2nd, 3rd, or 4th quarter of year	(3) Difference (std. error) (1) − (2)
ln (wkly. wage)	5.8916	5.9027	−0.01110 (0.00274)
Education	12.6881	12.7969	−0.1088 (0.0132)
Wald est. of return to education			0.1020 (0.0239)
OLS return to education			0.0709 (0.0003)

b_{IV} = (5.90271 – 5.8916)/(12.7969 – 12.6881) = **0.1021**

Interpretation: The Wald estimator measures the effect of an extra year of schooling on those (dropout) students for whom an earlier birth –i.e., **Z** changes from 0 to 1– would have been forced to complete an extra year of schooling before dropping out.

IV with a Dummy Variable

• The above result can be extended to IV with multiple dummy instruments. For example, J categories; say, 4 QOB: Q1, Q2, Q3, Q4.

From the structural equation: $y_i = \beta x_i + \varepsilon_i$

$$\Rightarrow E[y_i \mid Z_i] = \beta E[x_i \mid Z_i]$$

• We replace the expectations (say, $E[y_i \mid Z_i=j]$) with sample analogs ($\bar{y}_j$). Then, in this case, the IV estimator is the same as the coefficient from a regression of J group means between **Y** and **X** , weighted by the size of the groups.

Finite Sample Problems

• IV estimators are consistent if the instruments, **Z**, used are both valid and relevant/informative, but they may be subject to significant finite sample biases.

• We now look at two distinct sources of finite sample bias:

– The use of IVs that are only weakly related to the endogenous variable(s), resulting in "weak identification" of the parameters of interest. This is the *weak instruments* problem.

– The use of "too many" instruments relative to the available sample size. This is the *overidentification* problem.

Weak Instruments: Definition and Implications

• The explanatory power of **Z** may not be enough to allow inference on β. In this case, we say **Z** is a *weak* instrument.

Definition: Weak Instrument
IVs are *weak* if the mean component of **X** that depends on the IVs – $\mathbf{Z\Pi}$– is small relative to the variability of **X**, or equivalently, to the variability of the error **V**.

• Implications:

– Gleser and Hwang (1987) and Dufour (1997) show that CIs and tests based on *t-tests* and *F* (Wald) tests are not robust to weak IVs.

– The concern is not just theoretical: Numerical studies show that coverage rates of conventional 2SLS CIs can be very poor when IVs are weak, even if *T* is large.

Weak Instruments: Detection (Greene)

• Usual detection of weak IVs: Check if H_0 (weak instruments): $\mathbf{\Pi} = \mathbf{0}$:
- – Test H_0: $\mathbf{\Pi} = \mathbf{0}$ with a standard *F-test* on **Z** in the 1st stage regression.
- – Rule of thumb: For a single endogenous regressor, Staiger and Stock (1997) suggest that 1st stage $F < 10$ is cause for concern.
- – Low partial-$R^2_{X,Z}$ –exogenous variable **U** is partialed out; see, Shea (1997).
- – Large Var[$\mathbf{b}_{IV}$].

Note: There is a theoretical problem when, under a H_0, we have unidentified parameters. Under H_0: $\mathbf{\Pi} = \mathbf{0}$, $\boldsymbol{\beta}$ is not identified.

Weak Instruments: A&K (1991)

• True story: The graduate labor class at the University of Michigan does replication exercises. Two students, Regina Baker and David Jaeger replicated the results in Angrist and Krueger (1991).

• They and their professor, John Bound, notice two things:

(1) The results are imprecise and unstable when the controls and instrument sets change.

(2) The results become precise and stable only when the 1st stage F tests reject H_0: $\mathbf{\Pi} = \mathbf{0}$ –i.e., when instruments are not weak.

Note: Consider the first stage: $\mathbf{X} = \mathbf{Z\Pi} + \boldsymbol{\xi}$.

Even if $\mathbf{\Pi} = \mathbf{0}$ in the DGP, as the number of instruments increases the R^2 of the first stage regression in the sample can only increase.

Weak Instruments: A&K (1991)

Bound et al. 1995: Table 1

Table 1. Estimated Effect of Completed Years of Education on Men's Log Weekly Earnings (standard errors of coefficients in parentheses)

	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV
Coefficient	.063 (.000)	.142 (.033)	.063 (.000)	.081 (.016)	.063 (.000)	.060 (.029)
F (excluded instruments)		13.486		4.747		1.613
Partial R^2 (excluded instruments, ×100)		.012		.043		.014
F (overidentification)		.932		.775		.725
Age Control Variables						
Age, Age^2	x	x			x	x
9 Year of birth dummies			x	x	x	x
Excluded Instruments						
Quarter of birth		x		x		x
Quarter of birth × year of birth				x		x
Number of excluded instruments		3		30		28

NOTE: Calculated from the 5% Public-Use Sample of the 1980 U.S. Census for men born 1930–1939. Sample size is 329,509. All specifications include Race (1 = black), SMSA (1 = central city), Married (1 = married, living with spouse), and 8 Regional dummies as control variables. F (first stage) and partial R^2 are for the instruments in the first stage of IV estimation. F (overidentification) is that suggested by Basmann (1960).

Note: From BJB (1995, JASA). Different instruments deliver different 1st stage F-stats. In only (2) there is a significant F-stat!

Weak Instruments: A&K (1991)

Bound et al. 1995: Table 2

Table 2. Estimated Effect of Completed Years of Education on Men's Log Weekly Earnings, Controlling for State of Birth (standard errors of coefficients in parentheses)

	(1) OLS	(2) IV	(3) OLS	(4) IV
Coefficient	.063 (.000)	.083 (.009)	.063 (.000)	.081 (.011)
F (excluded instruments)		2.428		1.869
Partial R^2 (excluded instruments, ×100)		.133		.101
F (overidentification)		.919		.917
Age Control Variables				
Age, Age^2			x	x
9 Year of birth dummies	x	x	x	x
Excluded Instruments				
Quarter of birth		x		x
Quarter of birth × year of birth		x		x
Quarter of birth × state of birth		x		x
Number of excluded instruments		180		178

NOTE: Calculated from the 5% Public-Use Sample of the 1980 U.S. Census for men born 1930–1939. Sample size is 329,509. All specifications include Race (1 = black), SMSA (1 = central city), Married (1 = married, living with spouse), 8 Regional dummies, and 50 State of Birth dummies as control variables. F (first stage) and partial R^2 are for the instruments in the first stage of IV estimation. F (overidentification) is that suggested by Basmann (1960).

Note: As the number of IVs increase, $\mathbf{b}_{2SLS}$ gets closer to $\mathbf{b}_{OLS}$.

Weak Instruments: A&K (1991)

• BJB suspected the presence of irrelevant IV. Then, they estimated the IV coefficient with a randomly assigned Z so that π=0 by construction. They reproduced the OLS estimate.

⇒ BJB's suggestion: look at the 1st stage F-stat.

Table 3. Estimated Effect of Completed Years of Education on Men's Log Weekly Earnings, Using Simulated Quarter of Birth (500 replications)

Table (column)	1 (4)	1 (6)	2 (2)	2 (4)
Estimated Coefficient				
Mean	.062	.061	.060	.060
Standard deviation of mean	.038	.039	.015	.015
5th percentile	−.001	−.002	.034	.035
Median	.061	.061	.060	.060
95th percentile	.119	.127	.083	.082
Estimated Standard Error				
Mean	.037	.039	.015	.015

NOTE: Calculated from the 5% Public-Use Sample of the 1980 U.S. Census for men born 1930–1939. Sample size is 329,509.

Weak Instruments: A&K (1991)

• QOB looked promising as an IV for education. What went wrong?

• Potential problems with QOB as an IV:

(1) Correlation between QOB and schooling is weak

- Small Cov(**X**, **Z**) introduces finite-sample bias, which will be exacerbated with the inclusion of many IV's.

(2) QOB may not be completely exogenous

- Recall that even small Cov(**Z**, $\boldsymbol{\varepsilon}$) will cause inconsistency, and this will be exacerbated when Cov(**X**, **Z**) is small.

• QOB qualifies as a *weak instrument* that may be correlated with unobserved determinants of wages (e.g., family income).

Weak Instruments: LIML

• There are alternative estimators, which have better small sample properties than 2SLS with weak instruments. One popular choice is LIML (*limited information maximum likelihood*), where we assume joint normality for the reduced form errors, $\boldsymbol{u}' = (\boldsymbol{\xi}, \mathbf{V})'$.

• Limited Information? We estimate only one equation, say, the first one $\mathbf{y}_j$, in a set of simultaneous equations:

$$\mathbf{y}_j = \mathbf{Y}_j\, \boldsymbol{\beta}_j + \mathbf{Z}_j\, \boldsymbol{\gamma}_j + \boldsymbol{\varepsilon}_j \qquad j = 1, 2, ..., k \text{ (structural equations)}$$

where $\mathbf{Y}_j$ is a matrix of k_1 included endogenous variables $\mathbf{Z}_j$ is a matrix of *included instruments*. There are $l_2 = l - l_1$ excluded instruments.

• Suppose we are interested in estimating the first equation $\mathbf{y}_1$ ($\mathbf{Y}_j$=$\mathbf{X}_j$):

$$\mathbf{y}_1 = \mathbf{X}_1\, \boldsymbol{\beta}_1 + \mathbf{Z}_1\, \boldsymbol{\gamma}_1 + \boldsymbol{\varepsilon}_1$$
$$\mathbf{X}_1 = \mathbf{Z}_1\, \boldsymbol{\Gamma}_{11} + \mathbf{Z}_2\, \boldsymbol{\Gamma}_{21} + \mathbf{V}_1$$

Weak Instruments: LIML

• Define $\boldsymbol{\theta} = (\boldsymbol{\beta}_1, \boldsymbol{\gamma}_1, \boldsymbol{\Gamma}_{11}, \boldsymbol{\Gamma}_{21})$ and

$$\mathbf{Y}_i = \begin{bmatrix} \boldsymbol{y}_{1,i} \\ \boldsymbol{X}_{1,i} \end{bmatrix}, \quad \mathbf{Z}_i = \begin{bmatrix} \boldsymbol{Z}_{1,i} \\ \boldsymbol{Z}_{2,i} \end{bmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 0 \\ -\boldsymbol{\beta}_1 & I_{k_1} \end{pmatrix}, \quad \boldsymbol{\Gamma} = \begin{pmatrix} \gamma_1 & \boldsymbol{\Gamma}_{11} \\ 0 & \boldsymbol{\Gamma}_{21} \end{pmatrix}$$

• Assume $\boldsymbol{u_1}' = (\boldsymbol{\xi}_1, \mathbf{V}_1)' \sim N(0, \boldsymbol{\Sigma})$. Then, the average log-likelihood:

$$L(\boldsymbol{\theta}, \boldsymbol{\Sigma}) = -\frac{k_1+1}{2}\log(2\pi) - \frac{1}{2}\log(|\boldsymbol{\Sigma}|) - \frac{1}{2T}\sum_{i=1}^{T}(\mathbf{B}'\mathbf{Y}_i - \boldsymbol{\Gamma}'\mathbf{Z}_i)'\,\boldsymbol{\Sigma}^{-1}(\mathbf{B}'\mathbf{Y}_i - \boldsymbol{\Gamma}'\mathbf{Z}_i)$$

<u>Note</u>: The Jacobian of the transformation from $\mathbf{Y}_i$ to $\boldsymbol{u_1}$ is $\mathbf{B}$ whose determinant is 1.

After a lot of algebra, the log of the concentrated log-likelihood with respect to $(\boldsymbol{\Sigma}, \boldsymbol{\beta}_1, \boldsymbol{\Gamma}_{11}, \boldsymbol{\Gamma}_{21})$ is:

$$L(\boldsymbol{\theta}, \boldsymbol{\Sigma}) = -\frac{k_1+1}{2}\log(2\pi) - \frac{1}{2}\log\kappa(\boldsymbol{\beta}_1) - \frac{1}{2}\log|Y'M_Z Y|$$

Weak Instruments: LIML

• After a lot of algebra, the log of the concentrated log-likelihood with respect to ($\boldsymbol{\Sigma}$, $\boldsymbol{\beta}_1$, $\boldsymbol{\Gamma}_{11}$, $\boldsymbol{\Gamma}_{21}$):

$$L(\boldsymbol{\theta}, \boldsymbol{\Sigma}) = -\frac{k_1+1}{2}\log(2\pi) - \frac{1}{2}\log \kappa(\boldsymbol{\beta}_1) - \frac{1}{2}\log |Y' M_Z Y|$$

where

$$\kappa(\boldsymbol{\beta}_1) = \frac{\delta' Y' M_1 Y \delta}{\delta' Y' M_Z Y \delta}$$

where $\delta = (1, -\boldsymbol{\beta}_1)$, $\mathbf{M}_1 = \mathbf{I}_T - \mathbf{P}_{\mathbf{Z}1}$.

• Maximizing $L(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ is equivalent to minimizing $\kappa(\boldsymbol{\beta}_1)$. Thus, sometimes LIML estimators are called *least variance ratio* estimator.

• $\hat{\kappa} = \kappa(\boldsymbol{\beta}_1) \geq 1$, since span($\mathbf{Z}_1$) $\subset$ span($\mathbf{Z}$) and the numerator of $\kappa(\boldsymbol{\beta}_1)$ cannot be smaller than the denominator for any δ. For any equation just identified, $\hat{\kappa} = 1$.

Weak Instruments: LIML

• Thanks to the special form of **B** and no exclusion restrictions in the endogenous variable regression, there is a closed-form solution to the LIML estimator (see Greene's textbook for details).

Let $\Upsilon' = [\mathbf{Z}_1\ \mathbf{X}_1]$. Then,

$$(\gamma_1, \boldsymbol{\beta}_1)' = [\Upsilon'(\mathbf{I}_T - \hat{\kappa}\mathbf{M}_\mathbf{Z})\ \Upsilon]^{-1}\ \Upsilon'(\mathbf{I}_T - \hat{\kappa}\mathbf{M}_\mathbf{Z})\ \mathbf{y_1} \qquad \textbf{(*)}$$

where $\hat{\kappa}$ is the is the smallest characteristic root of $\mathbf{W}_1\mathbf{W}^{-1}$, with:

$$\mathbf{W}_1 = \mathbf{Y}'\mathbf{M}_1\mathbf{Y} \quad \text{and} \quad \mathbf{W} = \mathbf{Y}'\mathbf{M}_Z\mathbf{Y}.$$

• The LIML estimator (*) is a *K-class estimator* (Theil (1961)). Note:

⇒ 2SLS estimator is a K-class estimator with $\hat{\kappa}$= 1,

⇒ OLS estimator is a K-class estimator with $\hat{\kappa}$= 0.

⇒ LIML = 2SLS when the equation is just identified.

Weak Instruments: LIML

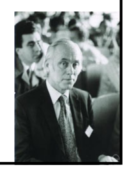
• This estimator is proposed by Anderson and Rubin (1949, 1950). It is the ML counterpart of the 2SLS.

• Under some assumptions, LIML and 2SLS have the same asymptotic distribution. But, in finite samples, they can differ.

• It turns out that LIML is a linear combination of the OLS and 2SLS estimates (with the weights depending on the data), and the weights happen to be such that they approximately eliminate the 2SLS bias.

<u>Note</u>: In the presence of "*many instruments*" (using group asymptotics) 2SLS is inconsistent, but LIML still is consistent.

• More in Lecture 16, in the context of SEM.



T. W. Anderson (1918 – 2016, USA)

Weak Instruments: LIML

TABLE 4.6.2
Alternative IV estimates of the economic returns to schooling

	(1)	(2)	(3)	(4)	(5)	(6)
2SLS	.105	.435	.089	.076	.093	.091
	(.020)	(.450)	(.016)	(.029)	(.009)	(.011)
LIML	.106	.539	.093	.081	.106	.110
	(.020)	(.627)	(.018)	(.041)	(.012)	(.015)
F-statistic (excluded instruments)	32.27	.42	4.91	1.61	2.58	1.97
Controls						
Year of birth	✓	✓	✓	✓	✓	✓
State of birth					✓	✓
Age, age squared		✓		✓		✓
Excluded instruments						
Quarter-of-birth dummies	✓	✓				
Quarter of birth*year of birth			✓	✓	✓	✓
Quarter of birth*state of birth					✓	✓
Number of excluded instruments	3	2	30	28	180	178

Notes: The table compares 2SLS and LIML estimates using alternative sets of instruments and controls. The age and age squared variables measure age in quarters. The OLS estimate corresponding to the models reported in columns 1–4 is .071; the OLS estimate corresponding to the models reported in columns 5 and 6 is .067. Data are from the Angrist and Krueger (1991) 1980 census sample. The sample size is 329,509. Standard errors are reported in parentheses.

Weak Instruments: Finance application

Example: The consumption CAPM.

After (many) assumptions, excess returns for a risky asset are a (linear) function of the covariance of the asset's returns with consumption growth:

$$E_t [r_{t+1}] - r_f = \gamma\sigma_{r\Delta} - \sigma_r^2/2$$

where $r_{t+1} = \ln(1+R_{t+1})$ - R_{t+1}: return on a risky asset.

$\sigma_r^2 = Var[\ln(1+R_{t+1})] = Var(r_{t+1})$

$\sigma_\Delta^2 = Var[\ln(c_{t+1}) - \ln(c_t)]$

$\sigma_{r\Delta} = Cov[\ln(c_{t+1}) - \ln(c_t), r_{t+1}]$

γ = Risk aversion coefficient from a CRRA utility function.

- The C-CAPM is easy to test using linear regressions.

- There is also a non-linear version of the C-CAPM.

Weak Instruments: Finance application

- In both linear and nonlinear versions of the model, IVs are weak -- see Neeley, Roy, and Whiteman (2001), and Yogo (2004).

- In the linear model in Yogo (2004):

X (endogenous variable): consumption growth

Z (the IVs): twice lagged nominal interest rates, inflation, consumption growth, and log dividend-price ratio.

- But, log consumption is close to a random walk, consumption growth is difficult to predict. This leads to the IVs being weak.

⇒ Yogo (2004) finds *F-statistics* for H_0: $\mathbf{\Pi} = 0$ in the 1st stage regression that lie between 0.17 and 3.53 for different countries.

Weak Instruments: Remedies (Greene)

- Symptom: The *relevance condition*, plim($\mathbf{Z'X}/T$) $\neq \mathbf{0}$, is close to being violated.
- Remedy:

– Not much – most of the discussion is about the condition, not what to do about it.

– Pick your best instrument and report just-identified results.

– Use LIML? Requires a normality assumption. Probably, not too restrictive.

Weak Instruments: Testing

• Irrelevant IVs –i.e., $\mathbf{\Pi}$=0– and weak IVs bias the IV estimation. Under Weak IVs, conventional asymptotics fail (see, Staiger and Stock (1997).

• Small simulation (replications = 2,000).

- Simple case: one endogenous variable, one IV. Parameters

$\mathbf{Z}, \boldsymbol{\varepsilon}, \mathbf{V} \sim N(0, \boldsymbol{\Sigma})$. Set unit variances, but Cov($\boldsymbol{\varepsilon}, \mathbf{V}$)=$\rho$

β=1; l=1 & 5; T = 100 & 1,000; ρ = .99 & .30

- Compute t =($\mathbf{b}_{2SLS}$ – 1)/SE($\mathbf{b}_{2SLS}$)

- We determine empirical size of 5% *t-test* (check $|t_{2SLS}|$>1.96)

- We study 3 cases:

1) Strong instruments (when $l=1$, π = 1; when $l=5$, π' = [1 1 0 0 0])

2) Weak instruments (when $l=1$, π = .1; when $l=5$, π' = [.1 .1 0 0 0])

3) Irrelevant instruments $\boldsymbol{\pi} \approx 0$ (approximated by .0001)

Weak Instruments: Testing

	Empirical size of 5% *t-test* (b_{2SLS})			
	T = 100		*T = 100*	
Quality of IV	*l*=1 (ρ=.99)	*l*=1 (ρ=.30)	*l*=5(ρ=.99)	*l*=5 (ρ=.30)
Strong	.065 (0.99)	.044 (1.00)	.088 (1.02)	.051 (1.01)
Weak	.195 (1.31)	.006 (2.26)	.852 (1.70)	.057 (1.22)
Irrelevant	.633 (2.01)	.001 (1.40)	.995 (1.99)	.045 (1.28)

	Empirical size of 5% *t-test* (b_{2SLS})			
	T = 1,000		*T = 1,000*	
Quality of IV	*l*=1 (ρ=.99)	*l*=1 (ρ=.30)	*l*=5(ρ=.99)	*l*=5 (ρ=.30)
Strong	.051 (1.00)	.002 (1.00)	.049 (1.00)	.050 (1.00)
Weak	.093 (0.79)	.002 (0.93)	.257 (1.13)	.059 (1.05)
Irrelevant	.631 (2.01)	.004 (0.60)	.995 (1.99)	.043 (1.31)

Weak Instruments: Testing

• In the presence of weak IVs, the usual tests have size problems. They are also not asymptotically pivotal: the distribution depends on nuisance parameters (ρ, $\mathbf{\Pi}$) that cannot be consistently estimated.

• Anderson and Rubin (1949) propose a test of H_0: $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, the AR stat, an *F*-test, that has good properties under the usual situations encountered under IV estimation.

• Intuition of AR test.

- Subtract from model $\mathbf{Y}\boldsymbol{\beta}_0$:

$$\mathbf{y} - \mathbf{Y}\boldsymbol{\beta}_0 = \mathbf{Y}\,\boldsymbol{\beta} - \mathbf{Y}\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon} = \mathbf{Y}\,(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \boldsymbol{\varepsilon}$$

- Substitute 1st stage:

$$\begin{aligned} \mathbf{y} - \mathbf{Y}\boldsymbol{\beta}_0 &= (\mathbf{Z\Pi} + \mathbf{V})\,(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \boldsymbol{\varepsilon} \\ &= \mathbf{Z\Pi}\,(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \mathbf{V}\,(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \boldsymbol{\varepsilon} \\ &= \mathbf{Z\Phi} + \mathbf{W} \end{aligned}$$

Weak Instruments: Testing

$$\mathbf{y} - \mathbf{Y}\boldsymbol{\beta}_0 = \mathbf{Z}\boldsymbol{\Phi} + \mathbf{W}$$

where

$$\boldsymbol{\Phi} = \boldsymbol{\Pi}\,(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$$
$$\mathbf{W} = \mathbf{V}\,(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \boldsymbol{\varepsilon}$$

Now, we can estimate $\boldsymbol{\Phi}$ using OLS, since $\mathbf{Z}$ is uncorrelated with $\mathbf{W}$.

Note that testing H_0: $\boldsymbol{\Phi} = \mathbf{0}$ $\Rightarrow H_0$: $\boldsymbol{\beta} = \boldsymbol{\beta}_0$.

• The AR stat is the usual F-stat for testing $\boldsymbol{\Phi} = \mathbf{0}$. Under the usual assumptions (fixed regressors and normal errors), the AR stat follows the usual F distribution.

• Under weak instruments –see Staiger and Stock (1997):

$$l \cdot \text{AR} = l \cdot F(\boldsymbol{\Phi}=\mathbf{0}) \rightarrow \chi^2(l).$$

Weak Instruments: Testing

<u>Note</u>:

- Note that under H_0: $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, the F-stat does not depend on $\boldsymbol{\Pi}$.
- The AR test is a joint test. It tests the joint hypothesis $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ and $\mathbf{Z}$ is uncorrelated with $\boldsymbol{\varepsilon}$.

• It turns out that the power of the AR test is not very good when $l>1$. The AR test tests whether $\mathbf{Z}$ enters the ($\mathbf{y} - \mathbf{Y}\boldsymbol{\beta}_0$) equation. The AR test sacrifices power: It ignores the restriction $\boldsymbol{\Phi} = \boldsymbol{\Pi}\,(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$.

• Low power leads to very wide CIs based on such tests. Kleibergen (2002) and Moreira (2001) propose an LM test whose H_0 rate is robust to weak IVs. (LM test first estimates $\boldsymbol{\Pi}$ under H_0: $\boldsymbol{\beta} = \boldsymbol{\beta}_0$.)

Weak Instruments: Testing

• There is an interesting literature on constructing CI under garbage instruments. That is, CI that are robust to the presence of weak and irrelevant instruments. See Staiger and Stock (1997) and Kleibergen (2002, 2003).

Example: It is possible to invert the AR stat to get a weak instrument robust CI interval for β:

$$CI_\alpha = \{\beta_0 : l \cdot AR \leq \chi^2_\alpha(l)\}$$

Weak Instruments: Pre-testing

• If one uses an *F-test* to detect weak IVs as a pre-test procedure, then the usual pre-testing issues arise for subsequent inference –see Hall, Rudebusch, and Wilcox (1996).

• In practice, researchers do tend to inspect and report the strength of the first stage. Tests and CI will not have the appropriate nominal size.

• Chioda and Jansson (2006) propose a similar statistic to the AR-stat to build a C.I.. It has a *non-standard distribution* conditional on the 1st stage *F*-stat. The C.I. are wider than the ones that do not condition on the 1st stage.

2SLS Bias with Many Instruments

• Let's go back to $\mathbf{b}_{2SLS} = (\mathbf{Y'}\,\mathbf{P_z}\,\mathbf{Y})^{-1}\,\mathbf{Y'P_z}\,\mathbf{y_1}$
and look at the bias: $E[\mathbf{b}_{2SLS} - \beta] = E[(\mathbf{Y'}\,\mathbf{P_z}\,\mathbf{Y})^{-1}\,(\mathbf{\Pi'Z'} + \mathbf{V'})\,\mathbf{P_z}\,\boldsymbol{\varepsilon}]$.

This expectation is hard to evaluate because the expectation operator does not pass through the inverse $(\mathbf{Y'}\,\mathbf{P_z}\,\mathbf{Y})^{-1}$, a nonlinear function.

• <u>New Tool</u>: *Group asymptotics* (or "*many instruments asymptotics")*. We use an asymptotic argument but, now, we allow *l* (number of instruments, "the group") to grow at the same rate as the sample size, *T*.

Group asymptotics assumes a condition:

$$\lim_{T\to\infty} \frac{l(T)}{T} = \alpha$$

If $\alpha = 0$, then standard aysmptotics (Lecture 7) applies –i.e., *l* is fixed.

2SLS Bias with Many Instruments

Intuition:

Lots of instruments ⇒ "instruments weak."

The number of first-stage parameters (the π_j's) grows with *l*. Many parameters become "*incidental*," creating consistency problems.

• Bekker (1994) and Newey and Smith (2004) show that GMM-type approaches to estimating structural parameters using instrumental variables, which include IV and 2SLS, may have substantial bias when *l* is not small relative to *T*.

• While 2SLS estimator is inconsistent in this many instruments environment, other estimators, including LIML & jackknife IV (JIVE), remain consistent and asymptotically normal.

2SLS Bias with Many Instruments

• Group asymptotics gives us something like an expectation, but we can take these expectations through non-linear functions:

$$E[\mathbf{b}_{2SLS} - \beta] \approx E[(\mathbf{Y}'\,\mathbf{P_z}\,\mathbf{Y})^{-1}]\,E[(\mathbf{\Pi}'\mathbf{Z}'\,\boldsymbol{\varepsilon})] + E[(\mathbf{Y}'\,\mathbf{P_z}\,\mathbf{Y})^{-1}]\,E[\mathbf{V}'\mathbf{P_z}\,\boldsymbol{\varepsilon}]$$

Since $E[(\mathbf{\Pi}'\mathbf{Z}'\,\boldsymbol{\varepsilon})] = 0$, $\quad E[\mathbf{b}_{2SLS} - \beta] \approx E[(\mathbf{Y}'\,\mathbf{P_z}\,\mathbf{Y})^{-1}]\,E[\mathbf{V}'\mathbf{P_z}\,\boldsymbol{\varepsilon}]$

• Substituting in the first stage:

$$E[\mathbf{b}_{2SLS} - \beta] \approx E[(\mathbf{\Pi}'\mathbf{Z}' + \mathbf{V}')'\,\mathbf{P_z}\,(\mathbf{Z\Pi} + \mathbf{V})]^{-1}\,E[\mathbf{V}'\mathbf{P_z}\,\boldsymbol{\varepsilon}]$$
$$= E[[\mathbf{\Pi}'\mathbf{Z}'\,\mathbf{Z\Pi}] + E[\mathbf{V}'\mathbf{P_z}\mathbf{V}]]^{-1}\,E[\mathbf{V}'\mathbf{P_z}\,\boldsymbol{\varepsilon}]$$

• Recall properties of trace (*tr*):

- *tr*: linear operator, goes through E[.], invariant to cyclic permutations
- $\mathbf{V}'\mathbf{P_z}\mathbf{V}$ is a scalar, $\Rightarrow \mathbf{V}'\mathbf{P_z}\mathbf{V} = tr(\mathbf{V}'\mathbf{P_z}\mathbf{V})$
- $tr(\mathbf{P_z}) = \text{rank}(\mathbf{P_z}) = l$ since $\mathbf{P}_{\mathbf{z}\text{ is}}$ is and idempotent matrix.

2SLS Bias with Many Instruments

• Then,

$$E[\mathbf{V}'\mathbf{P_z}\mathbf{V}] \approx E[tr(\mathbf{V}'\mathbf{P_z}\mathbf{V})] = E[tr(\mathbf{P_z}\mathbf{V}\mathbf{V}')] = tr(\mathbf{P_z}\,E[(\mathbf{V}\mathbf{V}')]$$
$$= tr(\mathbf{P_z}\,\boldsymbol{\sigma_{VV}}\boldsymbol{I}) = \boldsymbol{\sigma_{VV}}\,tr(\mathbf{P_z}) = \boldsymbol{\sigma_{VV}}\,l.$$

• Similar results applies to $E[\mathbf{V}'\mathbf{P_z}\,\boldsymbol{\varepsilon}] = \boldsymbol{\sigma_{\varepsilon V}}\,l.$

• Then, $E[\mathbf{b}_{2SLS} - \beta] \approx \boldsymbol{\sigma_{\varepsilon V}}\,l\,E[(\mathbf{\Pi}'\mathbf{Z}'\,\mathbf{Z\Pi}) + \boldsymbol{\sigma_{VV}}\,l]^{-1}$

$$= \boldsymbol{\sigma_{\varepsilon V}}/\boldsymbol{\sigma_{VV}}\;E[(\mathbf{\Pi}'\mathbf{Z}'\,\mathbf{Z\Pi})/(\boldsymbol{\sigma_{VV}}\,l) + 1]^{-1}$$

Note that $F = [E[(\mathbf{\Pi}'\mathbf{Z}'\,\mathbf{Z\Pi})]/l]/[\boldsymbol{\sigma_{VV}}]$ is the population F-statistic for H_0: $\mathbf{\Pi} = 0$ in the 1st stage regression. Thus,

$$E[\mathbf{b}_{2SLS} - \beta] \approx \boldsymbol{\sigma_{\varepsilon V}}/\boldsymbol{\sigma_{VV}}\;[1/(F+1)]$$

2SLS Bias with Many Instruments

• Thus,

$$E[\mathbf{b}_{2SLS} - \beta] \approx \sigma_{\varepsilon V}/\sigma_{VV}\ [1/(F+1)]$$

• Suppose the 1st stage coefficients, **Π**, are zero. Then, $F = 0$

$$\Rightarrow E[\mathbf{b}_{2SLS} - \beta] \approx \sigma_{\varepsilon V}/\sigma_{VV} = \sigma_{\varepsilon V}/\sigma_{YY} \quad \text{(OLS bias!)}$$

Intuition: If $\mathbf{\Pi} = 0$, then any variation in $\hat{X}$ in the sample comes from **V**. The variation in $\hat{X}$ is not different from the variation in **X**.

Note: This bias can affect tests, for example, the Hausman test.

2SLS Bias with Many Instruments

• From above:

$$F = E[\mathbf{\Pi' Z' Z \Pi}]/(\sigma_{VV}\, l)$$
$$E[\mathbf{b}_{2SLS} - \beta] \approx \sigma_{\varepsilon V}/\sigma_{VV}\ [1/(F+1)]$$

Remarks:

- If $\mathbf{\Pi} \neq 0$, but *F-stat* is small, then 2SLS will be biased towards OLS.
- *Weak instruments*: Instruments with small *F-stat*.
- The weak instrument bias tends to get worse as we add more (weak) instruments (by adding IVs with no explanatory power, the only thing changing in F is l!). These irrelevant IVs are referred as "*garbage instruments.*"
- As we add IVs, $\hat{X}$ gets closer to **X**. Then, 2SLS becomes OLS.
- If the IVs are very relevant ($F \longrightarrow \infty$), the IV bias goes to 0.

"*One good instrument is better than 50 garbage instruments*."

Excessive Overidentification

• Situation: l is much larger than k. Possible "*overidentification.*"

• Extreme case: Suppose $l = T$. In this case, $\mathbf{Z}$ is a square matrix:

$$\mathbf{b}_{2SLS} = [\mathbf{W'Z'X}]^{-1}\,\mathbf{W'Z'y} = [\mathbf{X'Z(Z'Z)}^{-1}\mathbf{Z'X}]^{-1}\,\mathbf{X'Z(Z'Z)}^{-1}\mathbf{Z'y}$$
$$= [\mathbf{X'Z\,Z}^{-1}\,\mathbf{Z'}^{-1}\,\mathbf{Z'X}]^{-1}\,\mathbf{X'Z\,Z}^{-1}\,\mathbf{Z'}^{-1}\,\mathbf{Z'y} = [\mathbf{X'X}]^{-1}\,\mathbf{X'y} = \mathbf{b}$$

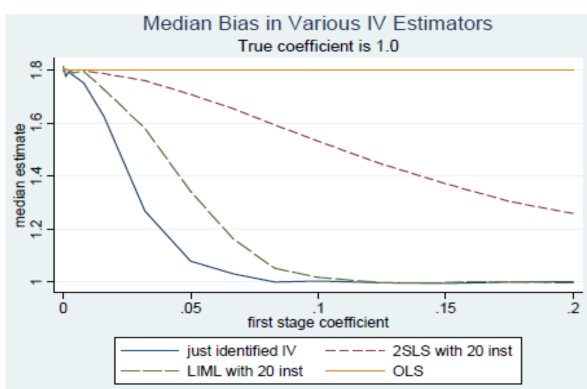
Since $\mathbf{b}$ is inconsistent and biased when $E[\boldsymbol{\varepsilon}|\mathbf{X}]\neq 0$, then so is $\mathbf{b}_{2SLS}$.

• While nobody will set $l=T$, a similar finite sample bias occurs in less extreme cases. In general, as $l \to T$, we see that $\mathbf{b}_{2SLS} \to \mathbf{b}_{OLS}$.

Note: For the IV asymptotic theory to be a good approximation, T must be much larger than l (say, $T - l > 40$ & grow linearly with T.)

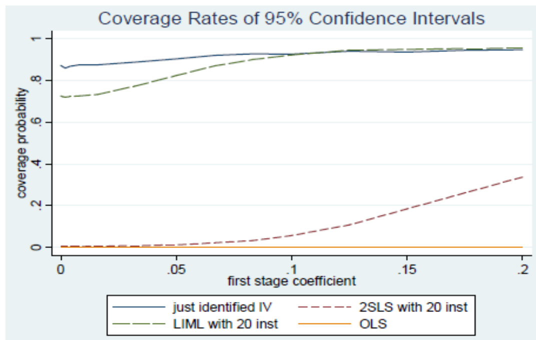
Excessive Overidentification

• Angrist and Pischke (2009) report that "just-identified 2SLS is *approximately unbiased*." They report a simulation with weak IVs, using OLS, just-identified IV and 2SLS with 20 IVs (and LIML too):



Excessive Overidentification

• Angrist and Pischke (2009) also report coverage rates of 95% C.I. (The coverage rate is the probability that a C.I. includes the true parameter.) Coverage rates for OLS and 2SLS are poor.



• Using lots of instruments can bias estimation (too many weak instruments) and cause innacurate asymptotic approximations.

Excessive Overidentification (Greene)

• Obvious symptom: **Z** has many more columns than **X**

– 1st stage of 2SLS almost reproduces **X**

– 2d stage of 2SLS becomes OLS, which is biased.

• Detection:

– Visual – there is no test.

– Check $\mathbf{b}_{2SLS}$ and $\mathbf{b}_{OLS}$. It they are similar, check that this is not a result of "too many IVs."

• Remedy:

– Fewer instruments? (Several methodological problems with this idea). Donald and Newey (2001) consider this option.

– *Jackknife* estimation –see Ackerberg and Devereux (2009).

Instrument Exogeneity: Detection & Remedies

• Symptom: The *valid condition*, plim($\mathbf{Z'}\varepsilon/T$)=$\mathbf{0}$, is close to being violated. Since the errors are not observed, very difficult to check.

• The valid condition is an exclusion restriction on the model:

$$\mathbf{y} = \mathbf{Y}\,\boldsymbol{\beta} + \mathbf{Z}\,\boldsymbol{\Theta} + \boldsymbol{\varepsilon}$$
$$\mathbf{Y} = \mathbf{Z}\,\boldsymbol{\Pi} + \mathbf{V}$$

$\Rightarrow$ The exclusion restriction imposes H_0: $\boldsymbol{\Theta} = \mathbf{0}$.

• If the exclusion is incorrect –i.e., $\theta = \theta_0 \neq 0$–, $\boldsymbol{\beta}$ will show an omitted variables bias problem. In the simple one exogenous variable & one IV case, it is easy calculate the bias:

$$\mathbf{b}_{IV} = \beta + \theta_0/\pi$$

The smaller π, the bigger the bias (the bias is worse with weak IVs).

Instrument Exogeneity: Detection & Remedies

• Detection of instrument exogeneity:
  – Endogenous IV's: Inconsistency of $\mathbf{b}_{IV}$ that makes it no better (and probably worse) than $\mathbf{b}_{OLS}$.
  – Durbin-Wu-Hausman test: Endogeneity of the problem regressor(s). But, DWH tests do not have good properties in the presence of weak instruments.

• Remedy:
  – Some modifications of the DWH have been suggested under weak instruments, see Hahn and Hausman (2002, 2005).
  – Avoid endogenous weak instruments.
  – General problem: It is not easy to find good instruments in theory and in practice. Find *natural experiments*.

IV: Remarks (Greene)

• Finding good instruments –i.e., meet both conditions- is not easy.

• When only the relevant condition is emphasized, OLS can be better than IV ("the cure can be worse..."). Even "clever" IVs can have low correlations with **X** and create severe finite-sample bias. The bias tends to be worse when there are many overidentifying restrictions (l is large relative to k).

• For the simple case of one endogenous variable, the F-stat in the 1st stage can help to identify weak IVs. With many IVs, Stock and Yogo (2005) provide rules of thumb regarding the weakness of the IVs based on a statistic due to Cragg and Donald (1993).

• Large T will not help. A&K and Consumption CAPM tests have very large samples!

IV: Remarks (Greene)

• Just identified IV is approximately unbiased (or less biased) even with weak instruments (although it is not possible to see this from the bias formula.)

What should you do in practice? (Pischke)

• Report 1st-stage and think about whether it makes sense. Are the magnitude and sign as you would expect?

• Report the *F*-stat on the excluded IVs. The bigger this is, the better. *F*s above 10 to 20 are considered relatively safe, lower *F*s put you in the danger zone.

• Pick your best single IV and report just-identified estimates using this one only. Just-identified IV is approximately median-unbiased.

• Check over-identified 2SLS estimates with LIML. If the LIML estimates are very different, or SEs are much bigger, worry.

• Check coefficients, t-stats, and *F*-stats for excluded IVs in the reduced-form regression of dependent variables on IVs. The reduced-form estimates are just OLS, so they are unbiased. If the relationship you expect is not in the reduced form, it is probably not there.

IV: Final General Remarks

• Finding good instruments is not easy. A good natural experiment, which defines the IV, is worthy of a paper.

• Angrist: "Tell a story about why a particular IV is a good instrument." In the omitted variable case: "Does the IV, for all intents and purposes, randomize the endogenous regressor?"

• IV models can be very informative, but it is your job (as author of the paper) to convince the audience.

• Good IV models are generally interesting in their own right, and should not be treated as "robustness" checks.

• The emphasis on IV with natural experiments is part of the *quasi-experimental revolution*, which shifted the emphasis in applied economics (and finance) from theory to empirical experiments.

IV: Final General Remarks

• This shift in mainly microeconomics (and, lately in other areas of economics and finance) from theory to empirical experiments: *"From Mas-Colell to Angrist and Pischke."*

• Criticism to IV estimation using natural experiments:

- No theory. For example, there is no optimizing labor (structural) model behind A&K (1991). Q: Where does the reduced form equation come from?

- Difficult to interpret the results. For example, LATE is an average, which may or may not contain information about the parameter of interest. It may not be useful (see Heckman and Urzua (2009), for an example, where LATE is not informative).

IV: Final General Remarks

• Answer from natural experimentalists to the criticism: Big skepticism about structural models. Thus, no modeling is good! Angrist and Pischke (2010): "*The explosion of IV methods, including LATE estimation, has led to greater "credibility" in applied econometrics."*

• References:

- "Mostly Harmless Econometrics: An Empiricist's Companion," textbook by Angrist and Pischke (2009).

- "Instruments, Randomization, and Learning about Development," by Deaton (2010, JEL).

- "The Empirical Economist's Toolkit: From Models to Methods," by Panhans and Singleton (2015, Duke Working Paper).