Instrumental Variable

Professor Ji-Woong Chung
Korea University

Outline

Outline

Motivation [Part 1]

- Consider the following estimation:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

- $cov(x_i, u) = 0$ for $i = 1, ..., k-1$, but $cov(x_k, u) \neq 0$.
  - Will this give us a consistent estimate of β_k?
  - When will we get consistent estimates of the other β's? Is it likely?

Motivation [Part 2]

- Answer #1: No, we won't get a consistent estimate of β_k.
- Answer #2: We will only get a consistent estimate of other β's if x_k is uncorrelated with all other x's.
- Instrumental Variables provide a *potential* solution to this problem...

Instrumental Variables: Intuition

- Think of x_k as having "good" and "bad" variation.
  - Good variation is uncorrelated with u.
  - Bad variation is correlated with u.
- An IV, z, explains the variation in x_k, but does not explain y directly.
  - I.e., it only explains the "good" variation in x_k.
  - If it explains y directly, z is a omitted variable which is part of u.
- IV allows us to isolate the "good" variation in x_k and replace x_k with only that component.

Outline

Instrumental Variables: Formal Conditions

- IVs must satisfy two conditions:
  - **Relevance Condition:** IV must explain variation in x_k.
  - **Exclusion Condition:** IV must be uncorrelated with u.
- Which condition is harder to satisfy? Can we test them?

To illustrate these conditions, let's start with the simplest case, where we have one instrument, z, for the problematic regressor, x_k.

Instrumental Variables: Identification

- For simplicity, assume $y = \beta_0 + \beta_1 x + u$ and $cov(x, u) \neq 0$. Suppose the previous two conditions are satisfied.

$$\begin{aligned} Cov(z, y) &= Cov(z, \beta_0 + \beta_1 x + u) \\ &= \beta_1 Cov(z, x) + Cov(z, u) \\ &= \beta_1 Cov(z, x) \\ \Rightarrow \beta_1 &= \frac{Cov(z, y)}{Cov(z, x)} \end{aligned}$$

- The IV estimand is:

$$\beta^{IV} = \frac{Cov(z, y)}{Cov(z, x)} \Longleftrightarrow \beta^{OLS} = \frac{Cov(x, y)}{Cov(x, x)}$$

Relevance Condition [Part 1]

- In the following model,

$$x_k = \alpha_0 + \alpha_1 x_1 + \cdots + \alpha_{k-1} x_{k-1} + \gamma z + v$$

  the relevance condition is satisfied if $\gamma \neq 0$
- What does this mean in words?
- Answer: z is relevant to explaining x_k after partialling out the effect of **all** other x's.

Relevance Condition [Part 2]

- Easy to test relevance condition:
  - Just regress x_k on all the other x's and the instrument z and see if z explains x_k
  - This is the "first stage" of IV estimation.
- It is important to take note of the sign (and even magnitude) of γ and not just its statistical significance.
  - Arguments for why a variable z makes a good IV candidate for an endogenous explanatory variable x_k should include a discussion about the nature of the relationship between x_k and z.

Exclusion Condition [Part 1]

- In the original model:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

- Exclusion condition: z is uncorrelated with u.
- What does this mean in words?
- Answer: z has no explanatory power with respect to y after conditioning on x's, i.e., z is uncorrelated with omitted variables.
  - Otherwise, both x_k and z should be controlled for. But we can't since $cov(x_k, u) \neq 0$.

Exclusion Condition [Part 2]

- You cannot test the exclusion condition directly.
- Why not?
- Answer: Because u is unobservable. You need a strong economic argument to support it.

What's wrong with this?

- Estimate this regression:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \gamma z + u$$

- If $\gamma = 0$, then exclusion restriction likely holds... i.e., z doesn't explain y after conditioning on the other x's

Answer: If the original regression doesn't give consistent estimates, then neither will this one! $cov(x_k, u) \neq 0$, so the estimates are still biased.

What Makes a Good Instrument?

- A good instrument must be justified with economic arguments.
- The relevance condition can be tested formally. But you should have an economic argument for why
- The exclusion condition cannot be tested formally. A convincing economic argument is needed as to why it explains y, but only through its effect on x_k

Outline

Implementing IV Estimation

- Once you've found a good IV:
  1. **First Stage:** Regress x_k on all other x's and z.
  2. **Second Stage:** Use the predicted x_k in the original model instead of x_k.
- This is called **Two-Stage Least Squares** (2SLS).

First Stage of 2SLS

- Estimate:

$$x_k = \alpha_0 + \alpha_1 x_1 + \cdots + \alpha_{k-1} x_{k-1} + \gamma z + v$$

- Get the predicted value $\hat{x}_k$, where:

$$\hat{x}_k = \hat{\alpha}_0 + \hat{\alpha}_1 x_1 + \cdots + \hat{\gamma} z$$

Second Stage of 2SLS

- Use the predicted values $\hat{x}_k$ to estimate the following in the second stage:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_{k-1} x_{k-1} + \beta_k \hat{x}_k + u$$

- This ensures a consistent estimation when the relevance and exclusion conditions are satisfied.[1]

[1] Equivalence of IV and 2SLS (when just identified): $\hat{\beta}^{2SLS} = (\hat{X}'\hat{X})^{-1}\hat{X}'y$, where $\hat{X} = Z(Z'Z)^{-1}Z'X$. Then $\hat{\beta}^{2SLS} = (X'Z(Z'Z)^{-1}Z'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'y$, which simplifies into $(X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'y = (Z'X)^{-1}(Z'Z)(X'Z)^{-1}X'Z(Z'Z)^{-1}Z'y$, which is $(Z'X)^{-1}Z'y = \hat{\beta}^{IV}$.

Intuition Behind 2SLS

- The predicted values $\hat{x}_k$ represent variation in x_k that is driven by factors uncorrelated with u. Good variation.
- Why not use just the other x's to generate the predicted value? Why need z?
- Answer: The predicted values would be collinear with the other x's in the second stage.

Reduced Form Estimates [Part 1]

- The reduced form estimation is when you regress y directly on the instrument z and the non-problematic x's:[2]

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_{k-1} x_{k-1} + \delta z + u$$

- This gives an unbiased estimate of the effect of z on y, likely through x_k.

[2] Plug the 1st stage regression into the structural model.

Reduced Form Estimates [Part 2]

- The IV estimate of x_k is given by:[3]

$$\hat{\beta}_{IV} = \frac{Cov(z,y)}{Cov(z,x)} = \frac{Cov(z,y)/Var(z)}{Cov(z,x)/Var(z)} = \frac{\hat{\delta}}{\hat{\gamma}}$$

  $\hat{\delta}$ Reduced form coefficient estimate for z (y on z)
  $\hat{\gamma}$ First stage coefficient estimate for z (x on z)

- If there is no effect of z on y in the reduced form, IV is unlikely to work.
  - IV estimates are just scaled versions of the reduced form.

[3] Alternatively, consider the structural model, $y = b_0 + b_1 x + u$, and the 1st stage, $x = a_0 + a_1 z + v$. The reduced form is $y = (b_0 + b_1 a_1) + b_1 a_1 z + (b_1 v + u)$. Hence, b_1 is the reduced form coefficient $(b_1 a_1)$ divided by the 1st stage coefficient (a_1).

Practical Advice [Part 1]

- Don't state in your paper's intro that you use IV unless:
  1. You state what the IV is.
  2. **And** you provide a strong economic argument for why it satisfies the necessary conditions.
- Don't bury the explanation of your IV! If you really have a good IV, you should defend it in the intro.

Practical Advice [Part 2]

- Justify why we should believe the exclusion restriction holds.
- Too many researchers only talk about the relevance condition.
- The exclusion restriction is equally important.

Practical Advice [Part 3]

- Don't do 2SLS manually. Let software (e.g., Stata) do it.
- Common mistakes when doing 2SLS on your own
  (See 4.6.1 in Angrist and Pische)
  1. Standard errors will be wrong.
  2. Nonlinear models in the first stage.
  3. Incorrect use of fitted values.

Practical Advice [Part 3-1]

- Why will standard errors be wrong if you try to do 2SLS on your own?
- Answer: Because the second stage uses 'estimated' values that have their own estimation error. This error needs to be considered when calculating standard errors![4]

---

[4] To see this:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$
$$Cov(x_1, u) = 0, Cov(x_2, u) \neq 0$$

First stage: $x_2^* = \alpha_0 + \alpha_1 x_1 + \gamma z$ in population. To estimate x_2^*, regress x_2 on x_1 and z and get $\hat{x}_2 = \hat{\alpha}_0 + \hat{\alpha}_1 x_1 + \hat{\gamma} z$. Can express $\hat{x}_2 = x_2^* + \epsilon$, where ϵ is an estimation error. Hence, we estimate

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2^* + u + \beta_1 \epsilon,$$

where the error term is $u + \beta_1 \epsilon$, not u.

Practical Advice [Part 3-2]

- Don't use predicted values from nonlinear models (e.g., Probit or Logit) in the second stage of an IV regression.
- Only linear OLS in the first stage guarantees the covariates and fitted values will be uncorrelated with the error term.
- This is called the "forbidden regression." (See p. 190 in Angrist and Pischke)

Practical Advice [Part 3-3]

- In models with quadratic terms, e.g.

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + u$$

  people often try to calculate one fitted value using one instrument, z, and then plug in and into second stage, with $\hat{x}$ and $(\hat{x})^2$
- Seems intuitive, but it is NOT consistent!
- Instead, you should just use z and z^2 as IVs!

Practical Advice [Part 4]

- **All** non-problematic x's need to be included in the first stage!
- You are not doing 2SLS correctly and won't get consistent estimates if this is not done.
- This includes fixed effects, such as firm and year FE.

Practical Advice [Part 5]

- Always report your first stage results and R^2.
- Reasons for this:
  1. It is a direct test of the relevance condition.
  2. It helps determine whether there might be a weak IV problem.

Outline

Weak Instruments Problem

- A weak instrument is an IV that doesn't explain much of the variation in the problematic regressor.
- Why is this an issue?
  - The small sample bias is greater when the instrument is weak; i.e., our estimates, which use a finite sample, might be misleading...
  - t-stats in finite samples can also be misleading.

Weak IV Bias

- Hahn and Hausman (2005) show that the finite sample bias of 2SLS is approximately:

$$\frac{j\rho(1-r^2)}{Nr^2}$$

- Where:
  - j is the number of IVs.
  - ρ is the correlation between x_k and u
  - r^2 is the R^2 from the first-stage regression.
  - N is the sample size.

Weak IV Bias

$$\frac{j\rho(1-r^2)}{Nr^2}$$

- j: More instruments, which we'll talk about later, need not help; they help increase r^2, but if they are weak (i.e., don't increase r^2 much), they can still increase finite sample bias
- r^2: A low explanatory power in the first stage can result in large bias, even if N is large.

Detecting Weak Instruments

- Look for warning flags:
  1. Large standard errors in IV estimates.
     - You'll get large SEs when covariance between instrument and problematic regressor is low
  2. Low F-statistic from the first stage.
     - The higher F-statistic for **excluded** IVs, the better
     - Stock, Wright, and Yogo (2002) suggest that an F-statistic above 10 indicates a strong IV. Or t stat of $\sqrt{10} \approx 3.2$[5]

[5] When there is heteroskedasticity, $F > 20$ (Olea and Pflueger, 2013)

Excluded IVs

- Just some terminology...
  - In some ways, can think of all nonproblematic x's as **included** IVs; they all appear in first stage and are used to get predicted values
  - But, when people refer to **excluded** IVs, they refer to the IVs (i.e., z's) that are excluded from the second stage

Outline

More Than One Problematic Regressor

- Again, consider the following model:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

- Now, there are two problematic regressors, x_{k-1} and x_k.
- IVs can still solve this problem, but each problematic regressor needs its own IV.

Multiple IVs [Part 1]

- If there are multiple problematic regressors, we need one IV for each problematic regressor.
- The 2SLS steps are similar:
  1. Regress x_k on all other x's (except x_{k-1}) and the IVs.
  2. Regress x_{k-1} on all other x's (except x_k) and the IVs.
  3. Use the predicted values in the second stage.

Multiple IVs [Part 2]

- We need at least as many IVs as problematic regressors to ensure predicted values are not collinear with non-problematic regressors.
- If the number of IVs equals the number of problematic regressors, the model is **"just identified"**

Overidentified Models

- When there are more IVs than problematic regressors, the model is said to be "overidentified."
  - m instruments for h problematic regressors, where $m > h$
- Can implement 2SLS just as before...

Overidentified Model conditions

- Necessary conditions very similar
  - Exclusion restriction: none of the instruments are correlated with u
  - Relevance condition
    - Each first stage (there will be h of them) must have at least one IV with non-zero coefficient
    - Of the m instruments, there must be at least h of them that are partially correlated with problematic regressors [otherwise, model isn't identified]
    - E.g., you can't just have one IV that is correlated with all the problematic regressors, and all the other IVs are not

Benefit of Overidentified Model

- Assuming you satisfy the relevance and exclusion conditions, you will get more asymptotic efficiency with more IVs
- Intuition: you can extract more ‘good’ variation from the first stage of the estimation

However, Overidentification Dilemma

- Suppose you find not just h instruments for h problematic regressors, you find $m > h$
- But why might you not want to use the $m - h$ extra instruments?

Answer - Weak instruments

- Again, as we saw earlier, a weak instrument will increase likelihood of finite sample bias and misleading inferences!
  - If have one good IV, not clear you want to add some extra (less good) IVs...

Practical Advice - Overidentified IV

- Helpful to always show results using "just identified" model with your best IVs
  - It is least likely to suffer from small sample bias
  - In fact, the just identified model is approaximately median-unbiased,[6] making weak instruments critique less of a concern
    (See p. 209 in Angrist and Pischke)

[6] I.e., the estimate underestimates just as often as it overestimates. So, it is invariant under one-to-one transformation.

Overidentification Tests [Part 1]

- In an overidentified model, you can "test" the quality of your IVs.
- The logic behind overidentification tests:
  - If all IVs are valid, we can get consistent estimates using any subset of the IVs.
  - So, compare IV estimates from different subsets to check if they are similar. If they are similar, it suggests the IVs are valid.

Overidentification Tests [Part 2]

- What is wrong with this logic?
  - Researcher has overidentified IV model
  - All the IVs are highly questionable in that they lack convincing economic arguments
  - But authors argue that because their model passes some "overidentification test" that the IVs must be okay

Overidentification Tests [Part 3]

- Answer: All the IVs could be junk!
  - The "test" implicitly assumes that some subset of instruments is valid
  - This may not be the case!
- To reiterate
  - There is no test to prove an IV is valid! Can only motivate that the IV satisfies exclusion restriction using economic theory

Outline

Miscellaneous IV Issues

1. IVs with interactions
2. Constructing additional IVs
3. Using lagged y or lagged x as IVs
4. Using group average of x as IV for x
5. Using IV with FE
6. Using IV with measurement error

IVs with Interactions [Part 1]

- Suppose we want to estimate the following:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + u$$

$Cov(x_1, u) = 0, Cov(x_2, u) \neq 0$

- So, both x_2 and $x_1 x_2$ are problematic.
- If we can only find one IV, z, can we get consistent estimates?

IVs with Interactions [Part 2]

- Answer = Yes! We can construct additional instruments from the IV.
  - Use z as IV for x_2.
  - Use $x_1 z$ as IV for $x_1 x_2$.
- The same economic argument supporting z as IV for x_2 will apply to $x_1 z$ as IV for $x_1 x_2$.

Constructing Additional IVs [Part 1]

- Suppose you want to estimate:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

$Cov(x_1, u) = 0, Cov(x_2, u) \neq 0, Cov(x_3, u) \neq 0$

- You only have one IV, z, which you believe is correlated with both x_2 and x_3.
- Can you use z and z^2 as IVs for x_2 and x_3, respectively?

Constructing Additional IVs [Part 2]

- Answer = Technically, yes, but it's probably not advisable.
- Without an economic reason for why z^2 is correlated with x_3 after partialling out z, it's likely a weak instrument.[7]

[7] Because z^2 often provides very little additional correlation with the endogenous variables beyond that provided by z.

Lagged Instruments

- It's common to use lagged variables as IVs.
- Two common forms:
  1. Instrumenting for a lagged y in a dynamic panel model with FE using lagged-lagged y.
  2. Instrumenting for a problematic x or lagged y using the lagged version of the same x.

Using Lagged y as IV in Panel Models

- In dynamic panel models with FE, the lagged dependent variable will be correlated with the error term.
- One solution is to use lagged values of y, e.g., $y_{i,t-2}$ as IV for the problematic $y_{i,t-1}$.
- However, lagged y values are often correlated with changes in errors if errors are serially correlated, which is common in corporate finance.[8]

[8] Arrellano and Bond (1991) suggest 1) estimate FD and 2) use further lagged y's as instruments for the differenced y, under the assumption that the error term is not serially correlated. See Holtz-Eakin, Newey, and Rosen (1988), Arellano and Bond (1991), Blundell and Bond (1998) for more details on these type of IV strategies

Using Lagged x as IV [Part 1]

- Another approach is to make assumptions about how $x_{i,t}$ is correlated with $u_{i,t}$.
- Idea behind the relevance condition is $x_{i,t}$ is persistent and predictive of future x or future y.
- And the exclusion restriction holds if we assume $x_{i,t}$ is uncorrelated with future shocks, u.

Using Lagged x as IV [Part 2]

- It's unclear how plausible this is...
- Serial correlation in u, which is common in corporate finance, often guarantees this IV is invalid.
- A strong economic argument is typically lacking.[9]

[9] See Arellano and Bond (1991), Arellano and Bover (1995) for more details on these type of IV strategies

Using Group Averages as IVs [Part 1]

- Often we see the following:

$$y_{i,j} = \alpha + \beta x_{i,j} + u_{i,j}$$

- $y_{i,j}$ is outcome for observation i (e.g., firm) in group j (e.g., industry)
- Worried that $Cov(x, u) \neq 0$.
- So, they use group average, $\bar{x}_{-i,j}$, as IV

$$\bar{x}_{-i,j} = \frac{1}{J-1} \sum_{i \in j, k \neq i} x_{k,j}$$

Using Group Averages as IVs [Part 2]

- And the papers say...
  - "group average of x is likely correlated with own x", i.e., relevance condition holds
  - "but group average doesn't directly affect y", i.e., exclusion restriction holds

Using Group Averages as IVs [Part 3]

- Is this a good IV?
  - Relevance condition implicitly assumes some common group-level heterogeneity, f_j, that is correlated with x_{ij}
  - But if model has f_j (i.e., group fixed effect), then $\bar{x}_{-i,j}$ must violate exclusion restriction![10]

[10] Consider $y_{i,j} = \beta x_{i,j} + u_{i,j}$ with $cov(x_{i,j}, u_{i,j}) \neq 0$. If f_j is omitted and is correlated with $x_{i,j}$, then f_j is also correlated with $\bar{x}_{-i,j}$.

Other Miscellaneous IVs

- As noted earlier, IVs can be useful in panel estimations:
  - #1 They help identify the effect of variables that don't vary within groups, which cannot be estimated directly in FE models.
  - #2 They can help with measurement error.

#1 IV with FE Models [Part 1]

- Use the following steps to identify variables that don't vary within groups?:
  1. Estimate the FE model (within transformation).
  2. Take the group-averaged residuals and regress them onto variables, x, that don't vary within groups.
     (i.e., the variables you couldn't estimate in FE model)
     - This second step (on its own) problematic ...
     - ... because the group-averaged error will still be correlated with the unobserved f_i (since f_i is time-invariant and influences all residuals within the group).

#1 IV with FE Models [Part 2]

- Solution in the second step is to use IV!
  2* Use the covariates that do vary within groups (from the first step) as instruments in the second step.
    - Which x's from first step are valid IVs?
    - Answer = those that don't co-vary with unobserved heterogeneity but do co-vary with variables that don't vary within groups [again, economic argument needed here]
- This strategy is discussed in Hausman and Taylor (1981).

Example

- Estimate the effect of education on wages.

$$ln(wage_{it}) = \beta_1 Experience_{it} + \beta_2 Education_i + f_i + u$$

  - Standard FE estimates the effect of β_1 consistently, but it drops Education because it's time-invariant and collinear with f_i.
  - The HT approach allows you to estimate the returns to Education β_2 by using the time-varying variable Experience as an instrument for Education in the second stage.
  - Required IV Assumption: Experience is uncorrelated with the individual fixed effect (Ability). If Ability affects both Education and Experience, the IV assumption fails, and the estimate of the return to Education will be biased. The need for this assumption is the core "economic argument" required in the third step.

#2 IV and Measurement Error [Part 1]

- Measurement error can be a problem in FE models.
- IVs provide a potential solution:
  - Find an IV that is correlated with the mismeasured variable but not with the error term.

#2 IV and Measurement Error [Part 2]

- However, identifying a valid instrument requires understanding the exact source of the measurement error.
  - This is because the disturbance, u, will include the measurement error; hence, how can you make an economic argument that z is uncorrelated with it if you don't understand the measurement error?[11, 12]

---

[11] See Almeida, Campello, and Galvao (RFS 2010) for examples of this strategy. Bond and Cummins (2002) use analyst's forecasts as an IV for investment demand. Griliches and Hausman (1986) proposes lagged x as an IV, assuming ME is i.i.d. Biorn (2000) relaxes the assumption: ME autocorrelation is constant.

[12] Consider $y = \beta_0 + \beta_1 x_1^* + \beta_2 x_2 + u$, where x_1^* is unobservable but $x_1 = x_1^* + e_1$ is observable. Hence, we estimate $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (u - \beta_1 e_1)$. Biased if $cov(x_1, e_1) \neq 0$. We need an IV for x_1. Suppose we have another measurement for x_1^*, i.e., $z_1 = x_1^* + \epsilon_1$. The instrument z_1 must be uncorrelated with both the original structural error u and the measurement error e_1. Since e_1 is a component of the disturbance $u - \beta_1 e_1$, one must have a strong theoretical argument about the nature of the measurement error to justify the exclusion restriction.

Outline

Limitations of IV

- Two main limitations:
  1. Finding a good instrument is difficult.
  2. External validity can be a concern.

Subtle Violations of Exclusion Restriction

- Even seemingly good IVs can violate the exclusion restriction.
  - Example: Bennedsen et al. (2007) use gender of the first-born child as an IV for family CEO succession.
  - However, the gender of the first-born may still be correlated with unobservable factors (e.g., talent of the daughter). Roberts and Whited (p.31, 2011)

Bennedsen et al. (2007) [Part 1]

- Paper studies effect of family CEO succession on firm performance
  - IVs for family CEO succession using gender of first-born child
    - Families where the first child was a boy are more likely to have a family CEO succession
    - Obviously, gender of first-born is totally random; seems like a great IV...
  - Any problem?

Bennedsen et al. (2007) [Part 2]

- Problem is that first-born gender may be correlated with u
  - Girl-first families may only turnover firm to a daughter when she is very talented
  - Therefore, effect of family CEO turnover might depend on gender of first born
  - I.e., gender of first born is correlated with u because u includes interaction between problematic x and the instrument, z!

External vs. Internal validity

- External validity is another concern of IV [and other identification strategies]
  - **Internal validity** is when the estimation strategy successfully uncovers a causal effect
  - **External validity** is when those estimates are predictive of outcomes in other scenarios (settings or populations)
- IV (done correctly) gives us internal validity. But it doesn't necessarily give us external validity

External Validity [Part 1]

- Issue is that IV estimates only tell us about subsample where the instrument is predictive
  - Remember, we're only making use of variation in x driven by z
  - So, we are not learning effect of x for observations where z does not explain x.
- It's a version of LATE (local average treatment effect) and affects interpretation.

External Validity [Part 2]

- Again, consider Bennedsen et al. (2007)
  - Gender of first born may only predict likelihood of family turnover in certain firms
  - I.e., family firms where CEO thinks females (including daughters) are less suitable for leadership positions
- Thus, we only learn about effect of family succession for these firms
- Why might this matter?

External Validity [Part 3]

- Answer: These firms might be different in other dimensions, which limits the external validity of our findings
- E.g., Could be that these are poorly run firms
  - If so, then we only identify effect for such poorly run firms using the IV
  - And effect of family succession in well-run firms might be quite different

External Validity [Part 4]

- Possible test for external validity problems
  - Size of residual from first stage tells us something about importance of IV for certain observations.
  - Large residual means IV didn't explain much. Small residual means it did.
  - Compare characteristics (i.e., other x's) of observations of groups with small and large residuals to make sure they don't differ much

Summary [Part 1]

- IV estimation helps overcome identification challenges.
- A good IV satisfies two conditions:
  1. Relevance condition.
  2. Exclusion condition.
- The exclusion condition cannot be tested directly; must use economic argument to support it

Summary [Part 2]

- IV estimation has its limitations:
  1. Finding a good IV is difficult.
  2. Weak instruments can be a problem; particularly when you have more IVs than problematic regressors
  3. External validity can be a concern.