Lecture VIII: Instrumental Variables

Stanislao Maldonado

Universidad del Rosario
stanislao.maldonado@urosario.edu.co

Impact Evaluation
Universidad del Rosario
March 28th, 2017

1. Motivation for IV

- IV was developed by Philip Wright in 1928, who was interested in studying the effect of a tax on imported agricultural goods
- To do that, you need to estimate the demand and supply curves, but you only observe equilibrium points (simultaneity)! So, running OLS of quantity on prices will provide biased estimators
- They solve this problem by exploiting a third variable that shifted supply but not demand:
    - Correlated with prices (affects supply curve)
    - But uncorrelated with the unobservable variables (demand remains stable)

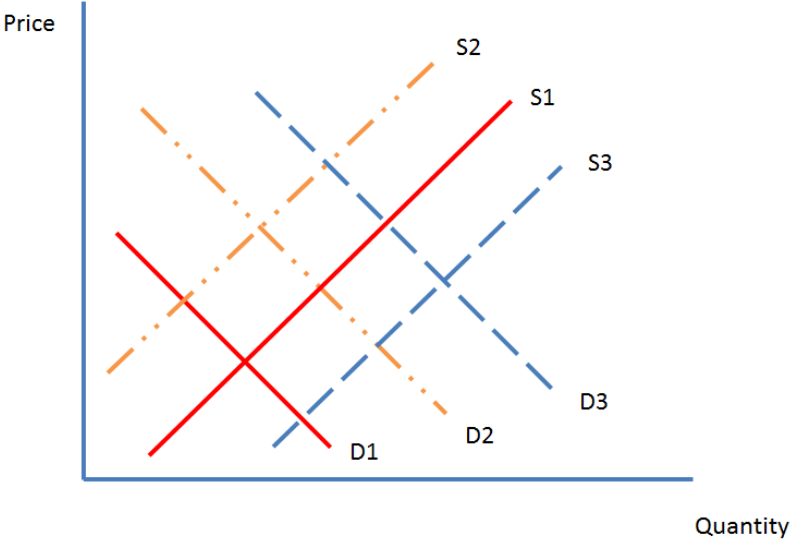
1. Motivation for IV

- IV was developed by Philip Wright in 1928, who was interested in studying the effect of a tax on imported agricultural goods
- To do that, you need to estimate the demand and supply curves, but you only observe equilibrium points (simultaneity)! So, running OLS of quantity on prices will provide biased estimators
- They solve this problem by exploiting a third variable that shifted supply but not demand:
  - Correlated with prices (affects supply curve)
  - But uncorrelated with the unobservable variables (demand remains stable)

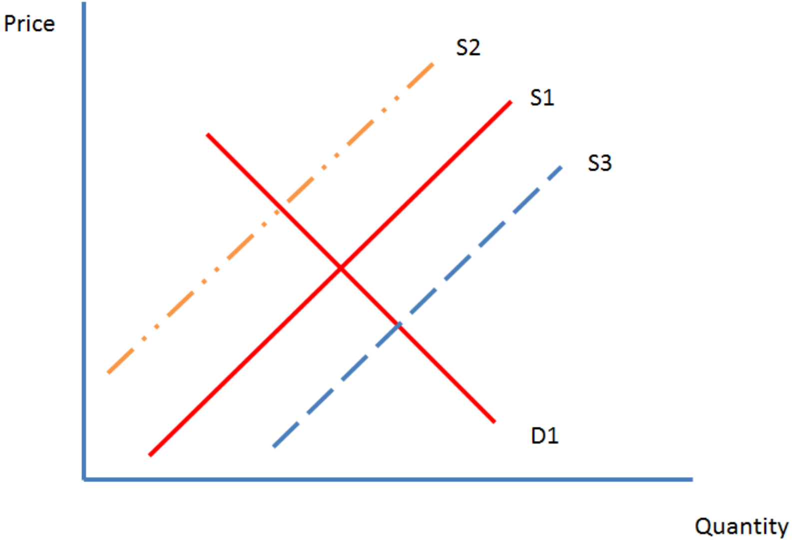
1. Motivation for IV

- IV was developed by Philip Wright in 1928, who was interested in studying the effect of a tax on imported agricultural goods
- To do that, you need to estimate the demand and supply curves, but you only observe equilibrium points (simultaneity)! So, running OLS of quantity on prices will provide biased estimators
- They solve this problem by exploiting a third variable that shifted supply but not demand:
  - Correlated with prices (affects supply curve)
  - But uncorrelated with the unobservable variables (demand remains stable)

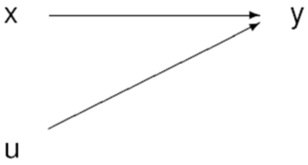
Example: Demand and supply in 3 periods



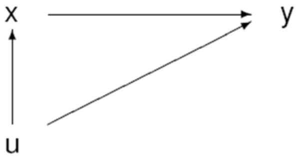
Example: Equilibrium when only the supply curve shifts



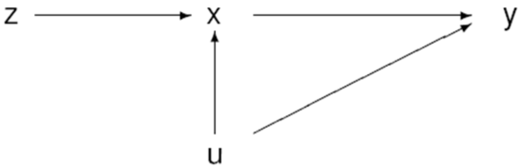
- IV allows a consistent estimation of the parameter of interest
- Recall: OLS consistent



- Recall: OLS inconsistent



- IV logic:



2. Standard IV

- Consider:

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + \mu \tag{1}$$

$$\mathbb{E}(\mu) = 0 \tag{2}$$

$$\mathrm{Cov}(x_j, \mu) = 0, j = 1, 2, ...K - 1 \tag{3}$$

$$\mathrm{Cov}(x_K, \mu) \neq 0 \tag{4}$$

- Suppose there exist a variable q in the error term that creates endogeneity in x_K
- Problem: all betas are inconsistent!
- IV provides a general solution to the endogeneity problem

2. Standard IV

- Consider:

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + \mu \quad (1)$$

$$\mathbb{E}(\mu) = 0 \quad (2)$$

$$\mathrm{Cov}(x_j, \mu) = 0, j = 1, 2, ...K - 1 \quad (3)$$

$$\mathrm{Cov}(x_K, \mu) \neq 0 \quad (4)$$

- Suppose there exist a variable q in the error term that creates endogeneity in x_K
- Problem: all betas are inconsistent!
- IV provides a general solution to the endogeneity problem

2. Standard IV

- Consider:

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + \mu \tag{1}$$

$$\mathbb{E}(\mu) = 0 \tag{2}$$

$$\mathrm{Cov}(x_j, \mu) = 0, j = 1, 2, ...K - 1 \tag{3}$$

$$\mathrm{Cov}(x_K, \mu) \neq 0 \tag{4}$$

- Suppose there exist a variable q in the error term that creates endogeneity in x_K
- Problem: all betas are inconsistent!
- IV provides a general solution to the endogeneity problem

2. Standard IV

- Consider:

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + \mu \tag{1}$$

$$\mathbb{E}(\mu) = 0 \tag{2}$$

$$\mathrm{Cov}(x_j, \mu) = 0, j = 1, 2, ...K - 1 \tag{3}$$

$$\mathrm{Cov}(x_K, \mu) \neq 0 \tag{4}$$

- Suppose there exist a variable q in the error term that creates endogeneity in x_K
- Problem: all betas are inconsistent!
- IV provides a general solution to the endogeneity problem

- We need to observe a variable, z, not in (1) that satisfies:
  1. z must be uncorrelated with μ:

$$\mathrm{Cov}(z, \mu) = 0 \tag{5}$$

  2. The relationship between z and x_K must follow the following criteria:

$$x_K = \delta_0 + \delta_1 x_1 + ... + \delta_{K-1} x_{K-1} + \theta_1 z + r_K \tag{6}$$

- Where $\theta_1 \neq 0$
- When there is only one variable this expression reduces to:

$$\mathrm{Cov}(z, X_K) \neq 0 \tag{7}$$

- We need to observe a variable, z, not in (1) that satisfies:
  1. z must be uncorrelated with μ:

$$\mathrm{Cov}(z, \mu) = 0 \tag{5}$$

  2. The relationship between z and x_K must follow the following criteria:

$$x_K = \delta_0 + \delta_1 x_1 + ... + \delta_{K-1} x_{K-1} + \theta_1 z + r_K \tag{6}$$

- Where $\theta_1 \neq 0$
- When there is only one variable this expression reduces to:

$$\mathrm{Cov}(z, X_K) \neq 0 \tag{7}$$

- Replacing (6) in (1):

$$y = \alpha_0 + \alpha_1 x_1 + ... + \alpha_{K-1} x_{K-1} + \lambda_1 z + v \tag{8}$$

- Where:

$$v = \mu + \beta_K r_K \tag{9}$$

$$\alpha_j = \beta_j + \beta_K \delta_j \tag{10}$$

$$\lambda_1 = \beta_K \theta_1 \tag{11}$$

- v is uncorrelated with all explanatory variables, so OLS can consistently estimate the reduced form parameters in (8)
- Sometimes reduced-form estimates are interesting
- How IV solves the identification problem for β_j in (1)?

- Replacing (6) in (1):

$$y = \alpha_0 + \alpha_1 x_1 + ... + \alpha_{K-1} x_{K-1} + \lambda_1 z + v \tag{8}$$

- Where:

$$v = \mu + \beta_K r_K \tag{9}$$

$$\alpha_j = \beta_j + \beta_K \delta_j \tag{10}$$

$$\lambda_1 = \beta_K \theta_1 \tag{11}$$

- v is uncorrelated with all explanatory variables, so OLS can consistently estimate the reduced form parameters in (8)
- Sometimes reduced-form estimates are interesting
- How IV solves the identification problem for β_j in (1)?

- Replacing (6) in (1):

$$y = \alpha_0 + \alpha_1 x_1 + ... + \alpha_{K-1} x_{K-1} + \lambda_1 z + v \tag{8}$$

- Where:

$$v = \mu + \beta_K r_K \tag{9}$$

$$\alpha_j = \beta_j + \beta_K \delta_j \tag{10}$$

$$\lambda_1 = \beta_K \theta_1 \tag{11}$$

- v is uncorrelated with all explanatory variables, so OLS can consistently estimate the reduced form parameters in (8)
- Sometimes reduced-form estimates are interesting
- How IV solves the identification problem for β_j in (1)?

- Replacing (6) in (1):

$$y = \alpha_0 + \alpha_1 x_1 + ... + \alpha_{K-1} x_{K-1} + \lambda_1 z + v \tag{8}$$

- Where:

$$v = \mu + \beta_K r_K \tag{9}$$

$$\alpha_j = \beta_j + \beta_K \delta_j \tag{10}$$

$$\lambda_1 = \beta_K \theta_1 \tag{11}$$

- v is uncorrelated with all explanatory variables, so OLS can consistently estimate the reduced form parameters in (8)
- Sometimes reduced-form estimates are interesting
- How IV solves the identification problem for β_j in (1)?

- Re-write (1) as follows:

$$y = X\beta + \mu \tag{12}$$

- Where:

$$X \equiv (1, x_2, ..., x_K)$$

- Define:

$$Z \equiv (1, x_2, ..., x_{K-1}, Z)$$

- From (2), (3), and (5) we have the following orthogonality condition:

$$\mathbb{E}(Z'\mu) = 0 \tag{13}$$

- Multiplying (12) by Z', taking expectations and using (13):

$$\mathbb{E}(Z'y) = [\mathbb{E}(Z'X)]\beta + \mathbb{E}(Z'\mu) = [\mathbb{E}(Z'X)]\beta \tag{14}$$

- Re-write (1) as follows:

$$y = X\beta + \mu \tag{12}$$

- Where:

$$X \equiv (1, x_2, ..., x_K)$$

- Define:

$$Z \equiv (1, x_2, ..., x_{K-1}, Z)$$

- From (2), (3), and (5) we have the following orthogonality condition:

$$\mathbb{E}(Z^{'}\mu) = 0 \tag{13}$$

- Multiplying (12) by $Z^{'}$, taking expectations and using (13):

$$\mathbb{E}(Z^{'}y) = [\mathbb{E}(Z^{'}X)]\beta + \mathbb{E}(Z^{'}\mu) = [\mathbb{E}(Z^{'}X)]\beta \tag{14}$$

- Re-write (1) as follows:

$$y = X\beta + \mu \tag{12}$$

- Where:

$$X \equiv (1, x_2, ..., x_K)$$

- Define:

$$Z \equiv (1, x_2, ..., x_{K-1}, Z)$$

- From (2), (3), and (5) we have the following orthogonality condition:

$$\mathbb{E}(Z^{'}\mu) = 0 \tag{13}$$

- Multiplying (12) by $Z^{'}$, taking expectations and using (13):

$$\mathbb{E}(Z^{'}y) = [\mathbb{E}(Z^{'}X)]\beta + \mathbb{E}(Z^{'}\mu) = [\mathbb{E}(Z^{'}X)]\beta \tag{14}$$

- Re-write (1) as follows:

$$y = X\beta + \mu \tag{12}$$

- Where:

$$X \equiv (1, x_2, ..., x_K)$$

- Define:

$$Z \equiv (1, x_2, ..., x_{K-1}, Z)$$

- From (2), (3), and (5) we have the following orthogonality condition:

$$\mathbb{E}(Z^{'}\mu) = 0 \tag{13}$$

- Multiplying (12) by $Z^{'}$, taking expectations and using (13):

$$\mathbb{E}(Z^{'}y) = [\mathbb{E}(Z^{'}X)]\beta + \mathbb{E}(Z^{'}\mu) = [\mathbb{E}(Z^{'}X)]\beta \tag{14}$$

- This system has a unique solution iff $\mathbb{E}(Z^{'}X)$ has full rank. Therefore:

$$\beta = [\mathbb{E}(Z^{'}X)]^{-1}\mathbb{E}(Z^{'}y) \tag{15}$$

- Given a random sample of X,y and Z, the **instrumental variables estimator** of β is:

$$\beta = [Z^{'}X]^{-1}Z^{'}y = \left(N^{-1}\sum_{i=1}^{N} z_i^{'}x_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N} z_i^{'}y_i\right) \tag{16}$$

- This system has a unique solution iff $\mathbb{E}(Z^{'}X)$ has full rank. Therefore:

$$\beta = [\mathbb{E}(Z^{'}X)]^{-1}\mathbb{E}(Z^{'}y) \tag{15}$$

- Given a random sample of X,y and Z, the **instrumental variables estimator** of β is:

$$\beta = [Z^{'}X]^{-1}Z^{'}y = \left(N^{-1}\sum_{i=1}^{N} z_i^{'}x_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N} z_i^{'}y_i\right) \tag{16}$$

2. Two-Stage Least Squares

- Consider again the following equations:

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + \mu \tag{17}$$

$$\mathbb{E}(\mu) = 0 \tag{18}$$

$$\mathrm{Cov}(x_j, \mu) = 0, j = 1, 2, ...K - 1 \tag{19}$$

- Assume now that more than one instrument is available for x_K
- Let $z_1, z_2, ..., z_M$ be variables such that:

$$\mathrm{Cov}(z_h, \mu) = 0, h = 1, 2, ...M \tag{20}$$

- To illustrate the method of 2SLS, define the vector of exogenous variables:

$$Z \equiv (1, x_2, ..., x_{K-1}, z_1, ..., z_M)$$

2. Two-Stage Least Squares

- Consider again the following equations:

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + \mu \tag{17}$$

$$\mathbb{E}(\mu) = 0 \tag{18}$$

$$\mathrm{Cov}(x_j, \mu) = 0, j = 1, 2, ...K - 1 \tag{19}$$

- Assume now that more than one instrument is available for x_K
- Let $z_1, z_2, ..., z_M$ be variables such that:

$$\mathrm{Cov}(z_h, \mu) = 0, h = 1, 2, ...M \tag{20}$$

- To illustrate the method of 2SLS, define the vector of exogenous variables:

$$Z \equiv (1, x_2, ..., x_{K-1}, z_1, ..., z_M)$$

2. Two-Stage Least Squares

- Consider again the following equations:

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + \mu \tag{17}$$

$$\mathbb{E}(\mu) = 0 \tag{18}$$

$$\mathrm{Cov}(x_j, \mu) = 0, j = 1, 2, ...K - 1 \tag{19}$$

- Assume now that more than one instrument is available for x_K
- Let $z_1, z_2, ..., z_M$ be variables such that:

$$\mathrm{Cov}(z_h, \mu) = 0, h = 1, 2, ...M \tag{20}$$

- To illustrate the method of 2SLS, define the vector of exogenous variables:

$$Z \equiv (1, x_2, ..., x_{K-1}, z_1, ..., z_M)$$

- The reduced form:

$$x_K = \delta_0 + \delta_1 x_1 + ... + \delta_{K-1} x_{K-1} + \theta_1 z_1 + ... + \theta_M z_M + r_K \quad (21)$$

- 2SLS chooses the linear combination of z that is most correlated with x_K
  - If x_K were exogenous, then it would be the best instrument for itself
  - If x_K is not exogenous, then the linear combination of z most highly correlated is (21)
- Since z is uncorrelated with μ, then x_K^* in (22) is also uncorrelated with μ:

$$x_K^* = \delta_0 + \delta_1 x_1 + ... + \delta_{K-1} x_{K-1} + \theta_1 z_1 + ... + \theta_M z_M \quad (22)$$

- The reduced form:

$$x_K = \delta_0 + \delta_1 x_1 + ... + \delta_{K-1} x_{K-1} + \theta_1 z_1 + ... + \theta_M z_M + r_K \qquad (21)$$

- 2SLS chooses the linear combination of z that is most correlated with x_K
  - If x_K were exogenous, then it would be the best instrument for itself
  - If x_K is not exogenous, then the linear combination of z most highly correlated is (21)
- Since z is uncorrelated with μ, then x_K^* in (22) is also uncorrelated with μ:

$$x_K^* = \delta_0 + \delta_1 x_1 + ... + \delta_{K-1} x_{K-1} + \theta_1 z_1 + ... + \theta_M z_M \qquad (22)$$

- The reduced form:

$$x_K = \delta_0 + \delta_1 x_1 + ... + \delta_{K-1} x_{K-1} + \theta_1 z_1 + ... + \theta_M z_M + r_K \tag{21}$$

- 2SLS chooses the linear combination of z that is most correlated with x_K
  - If x_K were exogenous, then it would be the best instrument for itself
  - If x_K is not exogenous, then the linear combination of z most highly correlated is (21)
- Since z is uncorrelated with μ, then x_K^* in (22) is also uncorrelated with μ:

$$x_K^* = \delta_0 + \delta_1 x_1 + ... + \delta_{K-1} x_{K-1} + \theta_1 z_1 + ... + \theta_M z_M \tag{22}$$

- Therefore, x_K^* is the part of x_K that is uncorrelated with μ
- Of course x_K^* is not directly observable (coefficients in 22 are population parameters), but it can be consistently estimated using OLS if there are no exact linear dependences among the exogenous variables
- The sample analogues of the x_K^* for each observation i are simply the OLS fitted values:

$$\widehat{x_K} = \widehat{\delta}_0 + \widehat{\delta}_1 x_1 + ... + \widehat{\delta_{K-1}} x_{K-1} + \widehat{\theta}_1 z_1 + ... + \widehat{\theta_M} z_M \tag{23}$$

- We can define for each observation i the vector:

$$\widehat{X}_i \equiv (1, x_{i1}, x_{i2}, ..., x_{iK-1}, \widehat{x_{iK}}) \tag{24}$$

- Therefore, x_K^* is the part of x_K that is uncorrelated with μ
- Of course x_K^* is not directly observable (coefficients in 22 are population parameters), but it can be consistently estimated using OLS if there are no exact linear dependences among the exogenous variables
- The sample analogues of the x_K^* for each observation i are simply the OLS fitted values:

$$\widehat{x_K} = \widehat{\delta}_0 + \widehat{\delta}_1 x_1 + ... + \widehat{\delta_{K-1}} x_{K-1} + \widehat{\theta}_1 z_1 + ... + \widehat{\theta_M} z_M \tag{23}$$

- We can define for each observation i the vector:

$$\widehat{X}_i \equiv (1, x_{i1}, x_{i2}, ..., x_{iK-1}, \widehat{x_{iK}}) \tag{24}$$

- Therefore, x_K^* is the part of x_K that is uncorrelated with μ
- Of course x_K^* is not directly observable (coefficients in 22 are population parameters), but it can be consistently estimated using OLS if there are no exact linear dependences among the exogenous variables
- The sample analogues of the x_K^* for each observation i are simply the OLS fitted values:

$$\widehat{x_K} = \widehat{\delta_0} + \widehat{\delta_1} x_1 + ... + \widehat{\delta_{K-1}} x_{K-1} + \widehat{\theta_1} z_1 + ... + \widehat{\theta_M} z_M \tag{23}$$

- We can define for each observation i the vector:

$$\widehat{X_i} \equiv (1, x_{i1}, x_{i2}, ..., x_{iK-1}, \widehat{x_{iK}}) \tag{24}$$

- Therefore, x_K^* is the part of x_K that is uncorrelated with μ
- Of course x_K^* is not directly observable (coefficients in 22 are population parameters), but it can be consistently estimated using OLS if there are no exact linear dependences among the exogenous variables
- The sample analogues of the x_K^* for each observation i are simply the OLS fitted values:

$$\widehat{x_K} = \widehat{\delta_0} + \widehat{\delta_1} x_1 + ... + \widehat{\delta_{K-1}} x_{K-1} + \widehat{\theta_1} z_1 + ... + \widehat{\theta_M} z_M \tag{23}$$

- We can define for each observation i the vector:

$$\widehat{X}_i \equiv (1, x_{i1}, x_{i2}, ..., x_{iK-1}, \widehat{x_{iK}}) \tag{24}$$

- Therefore, using $\widehat{X}_i$ as the instruments for X_i gives the IV estimator:

$$\beta_{2SLS} = [\widehat{X}'X]^{-1}\widehat{X}'y = \left(N^{-1}\sum_{i=1}^{N}\widehat{x}_i'x_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\widehat{x}_i'y_i\right) \tag{25}$$

$$= [X'Z(Z'Z)^{-1}Z'X]^{-1}[X'Z(Z'Z)^{-1}Z'y] \tag{26}$$

- Notice that:

$$\widehat{X} = Z(Z'Z)^{-1}Z'X = P_Z X \tag{27}$$

- Therefore:

$$\widehat{X}'X = X'P_Z X = (P_Z X)'P_Z X = \widehat{X}'\widehat{X} \tag{28}$$

- Therefore, using $\widehat{X}_i$ as the instruments for X_i gives the IV estimator:

$$\beta_{2SLS} = [\widehat{X}'X]^{-1}\widehat{X}'y = \left(N^{-1}\sum_{i=1}^{N}\widehat{x}_i'x_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\widehat{x}_i'y_i\right) \quad (25)$$

$$= [X'Z(Z'Z)^{-1}Z'X]^{-1}[X'Z(Z'Z)^{-1}Z'y] \quad (26)$$

- Notice that:

$$\widehat{X} = Z(Z'Z)^{-1}Z'X = P_ZX \quad (27)$$

- Therefore:

$$\widehat{X}'X = X'P_ZX = (P_ZX)'P_ZX = \widehat{X}'\widehat{X} \quad (28)$$

- Therefore, using $\widehat{X}_i$ as the instruments for X_i gives the IV estimator:

$$\beta_{2SLS} = [\widehat{X}' X]^{-1} \widehat{X}' y = \left(N^{-1} \sum_{i=1}^{N} \widehat{x}_i' x_i \right)^{-1} \left(N^{-1} \sum_{i=1}^{N} \widehat{x}_i' y_i \right) \tag{25}$$

$$= [X' Z (Z' Z)^{-1} Z' X]^{-1} [X' Z (Z' Z)^{-1} Z' y] \tag{26}$$

- Notice that:

$$\widehat{X} = Z (Z' Z)^{-1} Z' X = P_Z X \tag{27}$$

- Therefore:

$$\widehat{X}' X = X' P_Z X = (P_Z X)' P_Z X = \widehat{X}' \widehat{X} \tag{28}$$

3. Where instruments come from?

- In practice, it is hard to find good instruments. There are two approaches:
  1. Use economic theory to suggest instruments
  2. Look for exogenous source of variation that induces changes in the endogenous regressor

Some examples (Dunning 2012)

Source of instrumental variable	Units in study group	Treatment variable	Outcome variables
Lotteries			
Military drafts	Soldiers	Military service	Earnings, attitudes
Prize lotteries	Lottery players	Overall income	Political attitudes
Judge lotteries	Prisoners	Prison terms	Recidivism
Training invitations	Job-seekers	Job trainings	Wages
School vouchers	Students	Private-school attendance	Educational achievement
Weather shocks			
Rainfall growth	Countries	Economic growth	Civil war
Natural disasters	Countries	Oil prices	Democracy
Age			
Quarter-of-birth	Students	Education	Earnings
Twin studies			
Twin births	Mothers	Number of children	Earnings
Institutional variation			
Electoral cycles	States	Police presence	Crime
Land tenure types	States	Inequality	Public goods
Historical shocks			
Deaths of leaders	Countries	Colonial annexation	Development
Colonial settler mortality	Countries	Current institutions	Economic growth

4. Instrumental variables: Linear GMM with instruments

- IV is leading example of GMM
- Consider a linear regression model $y = \mathbf{x}'\beta + \mu$ in which some of the components of x are correlated with μ, therefore β is inconsistent
- Assume the existence of instruments z that are correlated with x but satisfy that $E(\mu|z) = 0$. Then, $E(y - \mathbf{x}'\beta|z) = 0$
- Using LIE, let's multiply by z to get the K unconditional population moment conditions:

$$E[z(y - \mathbf{x}'\beta)] = 0 \tag{29}$$

- The MM estimator solves the corresponding sample condition:

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i(y - \mathbf{x}_i'\beta) = \mathbf{0} \tag{30}$$

4. Instrumental variables: Linear GMM with instruments

- IV is leading example of GMM
- Consider a linear regression model $y = \mathbf{x}'\beta + \mu$ in which some of the components of x are correlated with μ, therefore β is inconsistent
- Assume the existence of instruments z that are correlated with x but satisfy that $E(\mu|z) = 0$. Then, $E(y - \mathbf{x}'\beta|z) = 0$
- Using LIE, let's multiply by z to get the K unconditional population moment conditions:

$$E[z(y - \mathbf{x}'\beta)] = 0 \tag{29}$$

- The MM estimator solves the corresponding sample condition:

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i(y - \mathbf{x}_i'\beta) = \mathbf{0} \tag{30}$$

4. Instrumental variables: Linear GMM with instruments

- IV is leading example of GMM
- Consider a linear regression model $y = \mathbf{x}^{'}\beta + \mu$ in which some of the components of x are correlated with μ, therefore β is inconsistent
- Assume the existence of instruments z that are correlated with x but satisfy that $E(\mu|z) = 0$. Then, $E(y - \mathbf{x}^{'}\beta|z) = 0$
- Using LIE, let's multiply by z to get the K unconditional population moment conditions:

$$E[z(y - \mathbf{x}^{'}\beta)] = 0 \tag{29}$$

- The MM estimator solves the corresponding sample condition:

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i(y - \mathbf{x}_i^{'}\beta) = \mathbf{0} \tag{30}$$

4. Instrumental variables: Linear GMM with instruments

- IV is leading example of GMM
- Consider a linear regression model $y = \mathbf{x}^{'}\beta + \mu$ in which some of the components of x are correlated with μ, therefore β is inconsistent
- Assume the existence of instruments z that are correlated with x but satisfy that $E(\mu|z) = 0$. Then, $E(y - \mathbf{x}^{'}\beta|z) = 0$
- Using LIE, let's multiply by z to get the K unconditional population moment conditions:

$$E[z(y - \mathbf{x}^{'}\beta)] = 0 \tag{29}$$

- The MM estimator solves the corresponding sample condition:

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i(y - \mathbf{x}_i^{'}\beta) = \mathbf{0} \tag{30}$$

4. Instrumental variables: Linear GMM with instruments

- IV is leading example of GMM
- Consider a linear regression model $y = \mathbf{x}^{'}\beta + \mu$ in which some of the components of x are correlated with μ, therefore β is inconsistent
- Assume the existence of instruments z that are correlated with x but satisfy that $E(\mu|z) = 0$. Then, $E(y - \mathbf{x}^{'}\beta|z) = 0$
- Using LIE, let's multiply by z to get the K unconditional population moment conditions:

$$E[z(y - \mathbf{x}^{'}\beta)] = 0 \tag{29}$$

- The MM estimator solves the corresponding sample condition:

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i(y - \mathbf{x}_i^{'}\beta) = \mathbf{0} \tag{30}$$

- If $dim(z) = K$, then $\beta_{MM} = \beta_{IV}$
- What if $dim(z) > K$?
  - Use just K instruments, but that solution is inefficient
  - Use GMM to choose $\widehat{\theta}_{GMM}$ to make vector $\frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i(y - \mathbf{x}_i^{'}\beta)$ as small as possible using quadratic loss
  - The $\widehat{\beta}_{GMM}$ minimizes:

$$Q_N(\beta) = \left[\frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i(y_i - \mathbf{x}_i^{'}\beta) \right]^{'} \mathbf{W}_N \left[\frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i(y_i - \mathbf{x}_i^{'}\beta) \right]$$

  - Using a matrix form:

$$Q_N(\beta) = \left[\frac{1}{N}(\mathbf{y} - \mathbf{X}\beta)^{'}\mathbf{Z} \right] \mathbf{W}_N \left[\frac{1}{N}\mathbf{Z}^{'}(\mathbf{y} - \mathbf{X}\beta) \right] \quad (31)$$

- If $dim(z) = K$, then $\beta_{MM} = \beta_{IV}$
- What if $dim(z) > K$?
  - Use just K instruments, but that solution is inefficient
  - Use GMM to choose $\widehat{\theta}_{GMM}$ to make vector $\frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i(y - \mathbf{x}_i^{'}\beta)$ as small as possible using quadratic loss
  - The $\widehat{\beta}_{GMM}$ minimizes:

$$Q_N(\beta) = \left[\ \frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i(y_i - \mathbf{x}_i^{'}\beta) \ \right]^{'} \mathbf{W}_N \left[\ \frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i(y_i - \mathbf{x}_i^{'}\beta) \ \right]$$

  - Using a matrix form:

$$Q_N(\beta) = \left[\ \tfrac{1}{N}(\mathbf{y} - \mathbf{X}\beta)^{'}\mathbf{Z} \ \right] \mathbf{W}_N \left[\ \tfrac{1}{N}\mathbf{Z}^{'}(\mathbf{y} - \mathbf{X}\beta) \ \right] \tag{31}$$

- If $dim(z) = K$, then $\beta_{MM} = \beta_{IV}$
- What if $dim(z) > K$?
  - Use just K instruments, but that solution is inefficient
  - Use GMM to choose $\widehat{\theta}_{GMM}$ to make vector $\frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i(y - \mathbf{x}_i^{'}\beta)$ as small as possible using quadratic loss
  - The $\widehat{\beta}_{GMM}$ minimizes:

$$Q_N(\beta) = \left[\ \frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i(y_i - \mathbf{x}_i^{'}\beta) \ \right]^{'} \mathbf{W}_N \left[\ \frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i(y_i - \mathbf{x}_i^{'}\beta) \ \right]$$

  - Using a matrix form:

$$Q_N(\beta) = \left[\ \frac{1}{N}(\mathbf{y} - \mathbf{X}\beta)^{'}\mathbf{Z} \ \right] \mathbf{W}_N \left[\ \frac{1}{N}\mathbf{Z}^{'}(\mathbf{y} - \mathbf{X}\beta) \ \right] \tag{31}$$

- If $dim(z) = K$, then $\beta_{MM} = \beta_{IV}$
- What if $dim(z) > K$?
  - Use just K instruments, but that solution is inefficient
  - Use GMM to choose $\widehat{\theta}_{GMM}$ to make vector $\frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i(y - \mathbf{x}_i^{'}\beta)$ as small as possible using quadratic loss
  - The $\widehat{\beta}_{GMM}$ minimizes:

$$Q_N(\beta) = \left[\ \frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i(y_i - \mathbf{x}_i^{'}\beta) \ \right]^{'} \mathbf{W}_N \left[\ \frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i(y_i - \mathbf{x}_i^{'}\beta) \ \right]$$

  - Using a matrix form:

$$Q_N(\beta) = \left[\ \frac{1}{N}(\mathbf{y} - \mathbf{X}\beta)^{'}\mathbf{Z} \ \right] \mathbf{W}_N \left[\ \frac{1}{N}\mathbf{Z}^{'}(\mathbf{y} - \mathbf{X}\beta) \ \right] \tag{31}$$

- If $dim(z) = K$, then $\beta_{MM} = \beta_{IV}$
- What if $dim(z) > K$?
  - Use just K instruments, but that solution is inefficient
  - Use GMM to choose $\widehat{\theta}_{GMM}$ to make vector $\frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i(y - \mathbf{x}_i^{'}\beta)$ as small as possible using quadratic loss
  - The $\widehat{\beta}_{GMM}$ minimizes:

$$Q_N(\beta) = \left[\ \frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i(y_i - \mathbf{x}_i^{'}\beta) \ \right]^{'} \mathbf{W}_N \left[\ \frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i(y_i - \mathbf{x}_i^{'}\beta) \ \right]$$

  - Using a matrix form:

$$Q_N(\beta) = \left[\ \tfrac{1}{N}(\mathbf{y} - \mathbf{X}\beta)^{'}\mathbf{Z} \ \right] \mathbf{W}_N \left[\ \tfrac{1}{N}\mathbf{Z}^{'}(\mathbf{y} - \mathbf{X}\beta) \ \right] \tag{31}$$

- The F.O.C:

$$\frac{\partial Q_N(\beta)}{\partial \beta} = -2\left[\ \frac{1}{N}\mathbf{X}'\mathbf{Z}\ \right]\mathbf{W}_N\left[\ \frac{1}{N}\mathbf{Z}'(\mathbf{y}-\mathbf{X}\beta)\ \right] = \mathbf{0} \qquad (32)$$

- Solving this give us the **GMM estimator for the linear IV model**:

$$\widehat{\beta}_{GMM} = [\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{y} \qquad (33)$$

- The F.O.C:

$$\frac{\partial Q_N(\beta)}{\partial \beta} = -2\left[\ \frac{1}{N}\mathbf{X}'\mathbf{Z}\ \right]\mathbf{W}_N\left[\ \frac{1}{N}\mathbf{Z}'(\mathbf{y}-\mathbf{X}\beta)\ \right] = \mathbf{0} \tag{32}$$

- Solving this give us the **GMM estimator for the linear IV model**:

$$\widehat{\beta}_{GMM} = [\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{y} \tag{33}$$

Distribution of Linear GMM Estimator

- Recall:

$$\widehat{\beta}_{GMM} = \beta + [(N^{-1}\mathbf{X}'\mathbf{Z})\mathbf{W}_N(N^{-1}\mathbf{Z}'\mathbf{X})]^{-1}(N^{-1}\mathbf{X}'\mathbf{Z})\mathbf{W}_N(N^{-1}\mathbf{Z}'\mu)$$

- Consistency requires that plim $N^{-1}\mathbf{Z}'\mu = \mathbf{0}$
- The limit distribution for is based on $\sqrt{N}(\widehat{\beta}_{GMM} - \beta)$. Then, the asymptotic normal distribution for $\widehat{\beta}_{GMM}$ with mean β and estimated asymptotic variance:

$$\widehat{V}(\widehat{\beta}_{GMM}) = N[\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{X}]^{-1}[\mathbf{X}'\mathbf{Z}\mathbf{W}_N\widehat{\mathbf{S}}\mathbf{W}_N\mathbf{Z}'\mathbf{X}][\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{X}]^{-1}$$

- Where $\widehat{\mathbf{S}}$ is a consistent estimate of:

$$\mathbf{S} = \text{plim}\,\frac{1}{N}\sum_{i=1}^{N} E[\mu_i^2 z_i z_i']$$

Distribution of Linear GMM Estimator

- Recall:

$$\widehat{\beta}_{GMM} = \beta + [(N^{-1}\mathbf{X}'\mathbf{Z})\mathbf{W}_N(N^{-1}\mathbf{Z}'\mathbf{X})]^{-1}(N^{-1}\mathbf{X}'\mathbf{Z})\mathbf{W}_N(N^{-1}\mathbf{Z}'\mu)$$

- Consistency requires that plim $N^{-1}\mathbf{Z}'\mu = \mathbf{0}$
- The limit distribution for is based on $\sqrt{N}(\widehat{\beta}_{GMM} - \beta)$. Then, the asymptotic normal distribution for $\widehat{\beta}_{GMM}$ with mean β and estimated asymptotic variance:

$$\widehat{V}(\widehat{\beta}_{GMM}) = N[\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{X}]^{-1}[\mathbf{X}'\mathbf{Z}\mathbf{W}_N\widehat{\mathbf{S}}\mathbf{W}_N\mathbf{Z}'\mathbf{X}][\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{X}]^{-1}$$

- Where $\widehat{\mathbf{S}}$ is a consistent estimate of:

$$\mathbf{S} = \text{plim}\,\frac{1}{N}\sum_{i=1}^{N} E[\mu_i^2 z_i z_i']$$

Distribution of Linear GMM Estimator

- Recall:

$$\widehat{\beta}_{GMM} = \beta + [(N^{-1}\mathbf{X}'\mathbf{Z})\mathbf{W}_N(N^{-1}\mathbf{Z}'\mathbf{X})]^{-1}(N^{-1}\mathbf{X}'\mathbf{Z})\mathbf{W}_N(N^{-1}\mathbf{Z}'\mu)$$

- Consistency requires that plim $N^{-1}\mathbf{Z}'\mu = \mathbf{0}$
- The limit distribution for is based on $\sqrt{N}(\widehat{\beta}_{GMM} - \beta)$. Then, the asymptotic normal distribution for $\widehat{\beta}_{GMM}$ with mean β and estimated asymptotic variance:

$$\widehat{V}(\widehat{\beta}_{GMM}) = N[\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{X}]^{-1}[\mathbf{X}'\mathbf{Z}\mathbf{W}_N\widehat{\mathbf{S}}\mathbf{W}_N\mathbf{Z}'\mathbf{X}][\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{X}]^{-1}$$

- Where $\widehat{\mathbf{S}}$ is a consistent estimate of:

$$\mathbf{S} = \text{plim}\,\frac{1}{N}\sum_{i=1}^{N} E[\mu_i^2 z_i z_i']$$

Distribution of Linear GMM Estimator

- Recall:

$$\widehat{\beta}_{GMM} = \beta + [(N^{-1}\mathbf{X}'\mathbf{Z})\mathbf{W}_N(N^{-1}\mathbf{Z}'\mathbf{X})]^{-1}(N^{-1}\mathbf{X}'\mathbf{Z})\mathbf{W}_N(N^{-1}\mathbf{Z}'\mu)$$

- Consistency requires that plim $N^{-1}\mathbf{Z}'\mu = \mathbf{0}$
- The limit distribution for is based on $\sqrt{N}(\widehat{\beta}_{GMM} - \beta)$. Then, the asymptotic normal distribution for $\widehat{\beta}_{GMM}$ with mean β and estimated asymptotic variance:

$$\widehat{V}(\widehat{\beta}_{GMM}) = N[\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{X}]^{-1}[\mathbf{X}'\mathbf{Z}\mathbf{W}_N\widehat{\mathbf{S}}\mathbf{W}_N\mathbf{Z}'\mathbf{X}][\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{X}]^{-1}$$

- Where $\widehat{\mathbf{S}}$ is a consistent estimate of:

$$\mathbf{S} = \text{plim}\, \frac{1}{N}\sum_{i=1}^{N} E[\mu_i^2 z_i z_i']$$

Alternative Linear GMM estimators

- Consider the IV estimator:
  - In this case, $r = K$, then $[\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{X}]^{-1}$ simplifies to $(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{W}_N^{-1}(\mathbf{X}'\mathbf{Z})^{-1}$
  - Therefore:

$$\begin{aligned}\widehat{\beta}_{IV} &= (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{W}_N^{-1}(\mathbf{X}'\mathbf{Z})^{-1}\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{y} \\ &= (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}\end{aligned}$$

  - The estimated asymptotic variance:

$$\widehat{V}(\widehat{\beta}_{IV}) = N[\mathbf{Z}'\mathbf{X}]^{-1}\widehat{\mathbf{S}}[\mathbf{X}'\mathbf{Z}]^{-1}$$

- Consider the Optimal-weighted or two-step GMM:
  - For this case, $\mathbf{W}_N = \widehat{\mathbf{S}}^{-1}$. Then:

$$\widehat{\beta}_{OGMM} = [(\mathbf{X}'\mathbf{Z})\widehat{\mathbf{S}}^{-1}(\mathbf{Z}'\mathbf{X})]^{-1}(\mathbf{X}'\mathbf{Z})\widehat{\mathbf{S}}^{-1}(\mathbf{Z}'\mathbf{y})$$

Alternative Linear GMM estimators

- Consider the IV estimator:
  - In this case, $r = K$, then $[\mathbf{X}^{'}\mathbf{Z}\mathbf{W}_N\mathbf{Z}^{'}\mathbf{X}]^{-1}$ simplifies to $(\mathbf{Z}^{'}\mathbf{X})^{-1}\mathbf{W}_N^{-1}(\mathbf{X}^{'}\mathbf{Z})^{-1}$
  - Therefore:

$$\begin{aligned}\widehat{\beta}_{IV} &= (\mathbf{Z}^{'}\mathbf{X})^{-1}\mathbf{W}_N^{-1}(\mathbf{X}^{'}\mathbf{Z})^{-1}\mathbf{X}^{'}\mathbf{Z}\mathbf{W}_N\mathbf{Z}^{'}\mathbf{y} \\ &= (\mathbf{Z}^{'}\mathbf{X})^{-1}\mathbf{Z}^{'}\mathbf{y}\end{aligned}$$

  - The estimated asymptotic variance:

$$\widehat{V}(\widehat{\beta}_{IV}) = N[\mathbf{Z}^{'}\mathbf{X}]^{-1}\widehat{\mathbf{S}}[\mathbf{X}^{'}\mathbf{Z}]^{-1}$$

- Consider the Optimal-weighted or two-step GMM:
  - For this case, $\mathbf{W}_N = \widehat{\mathbf{S}}^{-\mathbf{1}}$. Then:

$$\widehat{\beta}_{OGMM} = [(\mathbf{X}^{'}\mathbf{Z})\widehat{\mathbf{S}}^{-\mathbf{1}}(\mathbf{Z}^{'}\mathbf{X})]^{-1}(\mathbf{X}^{'}\mathbf{Z})\widehat{\mathbf{S}}^{-\mathbf{1}}(\mathbf{Z}^{'}\mathbf{y})$$

- Consider the Optimal-weighted or two-step GMM (Cont..):
  - The estimated asymptotic variance:

$$\widehat{V}(\widehat{\beta}_{OGMM}) = N[\mathbf{X}'\mathbf{Z}\widehat{\mathbf{S}}^{-1}\mathbf{Z}'\mathbf{X}]^{-1}$$

- Consider the two-stage least squares (2SLS):
  - For this case, if errors are homoskedastic, $\widehat{S}^{-1} = [s^2 N^{-1}\mathbf{Z}'\mathbf{Z}]^{-1}$. Then $\mathbf{W}_N = [N^{-1}\mathbf{Z}'\mathbf{Z}]^{-1}$
  - Where $s^2 = (N-K)^{-1}\sum_{i=1}^{N}\widehat{\mu}_i^2$
  - Then:

$$\widehat{\beta}_{2SLS} = [(\mathbf{X}'\mathbf{P_Z}\mathbf{X})]^{-1}(\mathbf{X}'\mathbf{P_Z}\mathbf{y})$$

  - Where $\mathbf{P_Z} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$
  - The estimated asymptotic variance (homoskedastic case):

$$\widehat{V}(\widehat{\beta}_{2SLS}) = s^2[\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}$$

- Consider the Optimal-weighted or two-step GMM (Cont..):
  - The estimated asymptotic variance:

$$\widehat{V}(\widehat{\beta}_{OGMM}) = N[\mathbf{X}^{'}\mathbf{Z}\widehat{\mathbf{S}}^{-\mathbf{1}}\mathbf{Z}^{'}\mathbf{X}]^{-1}$$

- Consider the two-stage least squares (2SLS):
  - For this case, if errors are homoskedastic, $\widehat{S}^{-1} = [s^2 N^{-1}\mathbf{Z}^{'}\mathbf{Z}]^{-1}$. Then $\mathbf{W}_N = [N^{-1}\mathbf{Z}^{'}\mathbf{Z}]^{-1}$
  - Where $s^2 = (N-K)^{-1}\sum_{i=1}^{N}\widehat{\mu}_i^2$
  - Then:

$$\widehat{\beta}_{2SLS} = [(\mathbf{X}^{'}\mathbf{P_Z}\mathbf{X})]^{-1}(\mathbf{X}^{'}\mathbf{P_Z}\mathbf{y})$$

  - Where $\mathbf{P_Z} = \mathbf{Z}(\mathbf{Z}^{'}\mathbf{Z})^{-1}\mathbf{Z}^{'}$
  - The estimated asymptotic variance (homoskedastic case):

$$\widehat{V}(\widehat{\beta}_{2SLS}) = s^2[\mathbf{X}^{'}\mathbf{Z}(\mathbf{Z}^{'}\mathbf{Z})^{-1}\mathbf{Z}^{'}\mathbf{X}]^{-1}$$

5. Practical issues with IV

- IV designs require care to avoid problems that compromise its causal interpretation:
  - Instrument can be correlated with the non-observables
  - Instrument is weakly correlated to the endogenous variable
  - Economic meaning can be difficult to understand
- Current good practice in IV for non-experimental settings includes the following steps (Murray 2010):
  1. Check the significance and estimated signs of the instrument(s)
  2. Avoid bad instruments
  3. Test for a weak instrument
  4. Conduct hypothesis tests using the conditional likelihood ratio (CLR) test
  5. Build confidence intervals based on the CLR test or its variants
  6. Obtain point estimates using Fuller's (1977) estimators rather than 2SLS
  7. Interpret the IV results with care

5. Practical issues with IV

- IV designs require care to avoid problems that compromise its causal interpretation:
  - Instrument can be correlated with the non-observables
  - Instrument is weakly correlated to the endogenous variable
  - Economic meaning can be difficult to understand
- Current good practice in IV for non-experimental settings includes the following steps (Murray 2010):
  1. Check the significance and estimated signs of the instrument(s)
  2. Avoid bad instruments
  3. Test for a weak instrument
  4. Conduct hypothesis tests using the conditional likelihood ratio (CLR) test
  5. Build confidence intervals based on the CLR test or its variants
  6. Obtain point estimates using Fuller's (1977) estimators rather than 2SLS
  7. Interpret the IV results with care

5.1 Choosing promising IVs

- An ideal IV design has a strong to instrument that is credibly unrelated to unobservable dimensions in the principal equation
- Recall:
  - Instrument must be relevant (testable)
  - Instrument must be excludable (untestable)
- Discussing the validity of the exclusion restriction requires a lot of care and also creativity
- Example: Settler mortality as instrument for institutions (Acemoglu et al 2001)
  - Institutions are endogenous
  - Studying the relationship between institutions and economic performance is plagued of endogeneity issues
  - Settler mortality influenced the type of institutions colonizers decided to implement in the new world
  - Institutions persisted over time and explain differences in living standards across countries

5.1 Choosing promising IVs

- An ideal IV design has a strong to instrument that is credibly unrelated to unobservable dimensions in the principal equation
- Recall:
  - Instrument must be relevant (testable)
  - Instrument must be excludable (untestable)
- Discussing the validity of the exclusion restriction requires a lot of care and also creativity
- Example: Settler mortality as instrument for institutions (Acemoglu et al 2001)
  - Institutions are endogenous
  - Studying the relationship between institutions and economic performance is plagued of endogeneity issues
  - Settler mortality influenced the type of institutions colonizers decided to implement in the new world
  - Institutions persisted over time and explain differences in living standards across countries

5.1 Choosing promising IVs

- An ideal IV design has a strong to instrument that is credibly unrelated to unobservable dimensions in the principal equation
- Recall:
  - Instrument must be relevant (testable)
  - Instrument must be excludable (untestable)
- Discussing the validity of the exclusion restriction requires a lot of care and also creativity
- Example: Settler mortality as instrument for institutions (Acemoglu et al 2001)
  - Institutions are endogenous
  - Studying the relationship between institutions and economic performance is plagued of endogeneity issues
  - Settler mortality influenced the type of institutions colonizers decided to implement in the new world
  - Institutions persisted over time and explain differences in living standards across countries

5.1 Choosing promising IVs

- An ideal IV design has a strong to instrument that is credibly unrelated to unobservable dimensions in the principal equation
- Recall:
  - Instrument must be relevant (testable)
  - Instrument must be excludable (untestable)
- Discussing the validity of the exclusion restriction requires a lot of care and also creativity
- Example: Settler mortality as instrument for institutions (Acemoglu et al 2001)
  - Institutions are endogenous
  - Studying the relationship between institutions and economic performance is plagued of endogeneity issues
  - Settler mortality influenced the type of institutions colonizers decided to implement in the new world
  - Institutions persisted over time and explain differences in living standards across countries

Example: Settler mortality and institutions

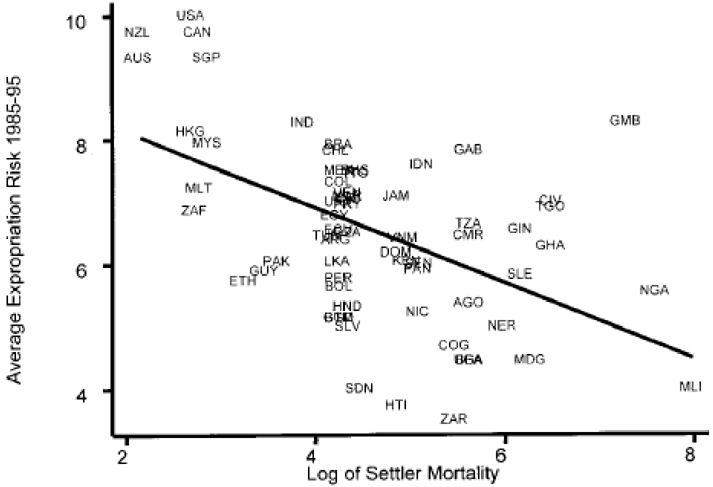


FIGURE 3. FIRST-STAGE RELATIONSHIP BETWEEN SETTLER MORTALITY AND EXPROPRIATION RISK

5.2 Avoiding bad IV

- Some strategies to consider:
  1. Tests of over-identifying restrictions
  2. Check for serially correlated disturbances
  3. Use alternative instruments
  4. Preclude links between the instruments and the disturbances
  5. Use information from other populations
  6. Be diligent about omitted explanators
  7. When feasible, randomize
  8. Use economic theory
  9. Use intuition and reduced forms
  10. Relax the exclusion restriction and compute bounds on the treatment effects

a. Test of over-identifying restrictions

- When more of one instrument is available, it is possible to evaluate the validity of the over-identifying instruments in the over-identified model when the model is estimated using optimal GMM
- Starting point is the fitted value of criterion function in (3):

$$Q_N(\widehat{\beta}) = \left[\ \frac{1}{N}(\mathbf{y} - \mathbf{X}\widehat{\beta})'\mathbf{Z} \ \right] \widehat{\mathbf{S}}^{-1} \left[\ \frac{1}{N}\mathbf{Z}'(\mathbf{y} - \mathbf{X}\widehat{\beta}) \ \right]$$

- Notice that if the population moment conditions $E[z(y - \mathbf{x}'\beta)] = 0$ are correct, then $z(y - \mathbf{x}'\beta) \simeq 0$. Therefore, $Q_N(\widehat{\beta})$ should be close to zero
- The null hypothesis is $H_0 : E[z(y - \mathbf{x}'\beta)] = 0$, then the OIR test is asymptotically chi-squared distributed with df=number of restrictions

a. Test of over-identifying restrictions

- When more of one instrument is available, it is possible to evaluate the validity of the over-identifying instruments in the over-identified model when the model is estimated using optimal GMM
- Starting point is the fitted value of criterion function in (3):

$$Q_N(\widehat{\beta}) = \left[\ \frac{1}{N}(\mathbf{y} - \mathbf{X}\widehat{\beta})'\mathbf{Z}\ \right] \widehat{\mathbf{S}}^{-1} \left[\ \frac{1}{N}\mathbf{Z}'(\mathbf{y} - \mathbf{X}\widehat{\beta})\ \right]$$

- Notice that if the population moment conditions $E[z(y - \mathbf{x}'\beta)] = 0$ are correct, then $z(y - \mathbf{x}'\beta) \simeq 0$. Therefore, $Q_N(\widehat{\beta})$ should be close to zero
- The null hypothesis is $H_0 : E[z(y - \mathbf{x}'\beta)] = 0$, then the OIR test is asymptotically chi-squared distributed with df=number of restrictions

a. Test of over-identifying restrictions

- When more of one instrument is available, it is possible to evaluate the validity of the over-identifying instruments in the over-identified model when the model is estimated using optimal GMM
- Starting point is the fitted value of criterion function in (3):

$$Q_N(\widehat{\beta}) = \left[\ \frac{1}{N}(\mathbf{y} - \mathbf{X}\widehat{\beta})'\mathbf{Z} \ \right] \widehat{\mathbf{S}}^{-1} \left[\ \frac{1}{N}\mathbf{Z}'(\mathbf{y} - \mathbf{X}\widehat{\beta}) \ \right]$$

- Notice that if the population moment conditions $E[z(y - \mathbf{x}'\beta)] = 0$ are correct, then $z(y - \mathbf{x}'\beta) \simeq 0$. Therefore, $Q_N(\widehat{\beta})$ should be close to zero
- The null hypothesis is $H_0 : E[z(y - \mathbf{x}'\beta)] = 0$, then the OIR test is asymptotically chi-squared distributed with df=number of restrictions

a. Test of over-identifying restrictions

- When more of one instrument is available, it is possible to evaluate the validity of the over-identifying instruments in the over-identified model when the model is estimated using optimal GMM
- Starting point is the fitted value of criterion function in (3):

$$Q_N(\widehat{\beta}) = \left[\ \tfrac{1}{N}(\mathbf{y} - \mathbf{X}\widehat{\beta})'\mathbf{Z} \ \right] \widehat{\mathbf{S}}^{-1} \left[\ \tfrac{1}{N}\mathbf{Z}'(\mathbf{y} - \mathbf{X}\widehat{\beta}) \ \right]$$

- Notice that if the population moment conditions $E[z(y - \mathbf{x}'\beta)] = 0$ are correct, then $z(y - \mathbf{x}'\beta) \simeq 0$. Therefore, $Q_N(\widehat{\beta})$ should be close to zero
- The null hypothesis is $H_0 : E[z(y - \mathbf{x}'\beta)] = 0$, then the OIR test is asymptotically chi-squared distributed with df=number of restrictions

- Results of this approach have to be interpreted with care:
  - Validity of the overidentifying is neither sufficient nor necessary for the validity of the moment conditions
  - Test does not check whether $E[z(y-\mathbf{x}^{'}\beta)]=0$ is true but rather whether there is some vector $\widehat{\beta}$ such that $E[z(y-\mathbf{x}^{'}\widehat{\beta})]=0$. Notice that:

$$E[z(y-\mathbf{x}^{'}\widehat{\beta})]=0$$
$$E[z(\mathbf{x}^{'}\beta+\mu-\mathbf{x}^{'}\widehat{\beta})]=0$$
$$E[z\mathbf{x}^{'}](\widehat{\beta}-\beta)]=E[z\mu]$$

  - When instruments are valid, then $E[z\mu]=0$ but it is possible to find a vector $(\widehat{\beta}-\beta)$ such that the last condition holds even if $E[z\mu]$ is different from zero
  - Rejection of the overidentifying restrictions can be the result of parameter heterogeneity

b. Check for serially correlated disturbances

- In panel settings, lagged dependent variable explanators are likely to be correlated with the disturbances if the disturbances are autoregressive
- Testing for serial correlation in model with lagged dependent variables is a useful check for whether or not IV estimation is needed
- Arellano and Bond (1991) offer a test for serial correlation that can be useful in this context

b. Check for serially correlated disturbances

- In panel settings, lagged dependent variable explanators are likely to be correlated with the disturbances if the disturbances are autoregressive
- Testing for serial correlation in model with lagged dependent variables is a useful check for whether or not IV estimation is needed
- Arellano and Bond (1991) offer a test for serial correlation that can be useful in this context

b. Check for serially correlated disturbances

- In panel settings, lagged dependent variable explanators are likely to be correlated with the disturbances if the disturbances are autoregressive
- Testing for serial correlation in model with lagged dependent variables is a useful check for whether or not IV estimation is needed
- Arellano and Bond (1991) offer a test for serial correlation that can be useful in this context

c. Use alternative instruments

- Sometimes is not feasible to include all the instruments in a single specification and use over-identification tests
- It is advisable to still explore how the results are affected when different instruments are used in order to identify a causal relationship
  - If parameters using different instruments differ appreciably and seemingly significant from one another, then the validity of the instrument would be hard to sustain
- Example: Hoxby (2002) on the effect of class size on children's test scores
  - Two IV strategies: a) unpredictable surprises in class sizes due to year to year random fluctuations in local births; and b) rules for class sizes
  - These two instruments exploit variation at different level (school and cohort) that cannot be accommodated in a single regression framework

c. Use alternative instruments

- Sometimes is not feasible to include all the instruments in a single specification and use over-identification tests
- It is advisable to still explore how the results are affected when different instruments are used in order to identify a causal relationship
    - If parameters using different instruments differ appreciably and seemingly significant from one another, then the validity of the instrument would be hard to sustain
- Example: Hoxby (2002) on the effect of class size on children's test scores
    - Two IV strategies: a) unpredictable surprises in class sizes due to year to year random fluctuations in local births; and b) rules for class sizes
    - These two instruments exploit variation at different level (school and cohort) that cannot be accommodated in a single regression framework

c. Use alternative instruments

- Sometimes is not feasible to include all the instruments in a single specification and use over-identification tests
- It is advisable to still explore how the results are affected when different instruments are used in order to identify a causal relationship
  - If parameters using different instruments differ appreciably and seemingly significant from one another, then the validity of the instrument would be hard to sustain
- Example: Hoxby (2002) on the effect of class size on children's test scores
  - Two IV strategies: a) unpredictable surprises in class sizes due to year to year random fluctuations in local births; and b) rules for class sizes
  - These two instruments exploit variation at different level (school and cohort) that cannot be accommodated in a single regression framework

d. Use information from other populations

- An instrument is valid if it is not an omitted explanator in the model
- With a single instrument, researchers need to look for creative ways to show that the instrument must be excluded in the model
- Knowledge of the institutional setting is critical
- Example: the impact of resource booms on public officials' demand for bribes (Maldonado 2012)
  - The author exploits the fact that the police and judicial workers are paid by the central government whereas the wages of municipality workers depends on the local government
  - Impact the boom should be expected only for municipality workers
  - Author finds no evidence of resource boom on the demand of bribes for the case of policemen and judicial workers

d. Use information from other populations

- An instrument is valid if it is not an omitted explanator in the model
- With a single instrument, researchers need to look for creative ways to show that the instrument must be excluded in the model
- Knowledge of the institutional setting is critical
- Example: the impact of resource booms on public officials' demand for bribes (Maldonado 2012)
  - The author exploits the fact that the police and judicial workers are paid by the central government whereas the wages of municipality workers depends on the local government
  - Impact the boom should be expected only for municipality workers
  - Author finds no evidence of resource boom on the demand of bribes for the case of policemen and judicial workers

d. Use information from other populations

- An instrument is valid if it is not an omitted explanator in the model
- With a single instrument, researchers need to look for creative ways to show that the instrument must be excluded in the model
- Knowledge of the institutional setting is critical
- Example: the impact of resource booms on public officials' demand for bribes (Maldonado 2012)
  - The author exploits the fact that the police and judicial workers are paid by the central government whereas the wages of municipality workers depends on the local government
  - Impact the boom should be expected only for municipality workers
  - Author finds no evidence of resource boom on the demand of bribes for the case of policemen and judicial workers

d. Use information from other populations

- An instrument is valid if it is not an omitted explanator in the model
- With a single instrument, researchers need to look for creative ways to show that the instrument must be excluded in the model
- Knowledge of the institutional setting is critical
- Example: the impact of resource booms on public officials' demand for bribes (Maldonado 2012)
  - The author exploits the fact that the police and judicial workers are paid by the central government whereas the wages of municipality workers depends on the local government
  - Impact the boom should be expected only for municipality workers
  - Author finds no evidence of resource boom on the demand of bribes for the case of policemen and judicial workers

5.3 Dealing with weak IV

- This problem arises when there is only a weak correlation between the instrument and the endogenous regressor
- Assume the following setting:

$$Y_i = \beta X_i + e_i \tag{34}$$

$$X_i = Z_i \pi + v_i \tag{35}$$

- The standard 2SLS estimator is $\beta_{2SLS} = \frac{X' P_Z Y}{X' P_Z X}$
- We make the extra assumption that error terms e_i and v_i are independently drawn from a normal distribution with variances σ_e^2 and σ_v^2, and a correlation ρ
- Let's introduce **the concentration parameter**:

$$\mu^2 = \pi' Z' Z \pi / \sigma_v^2 \tag{36}$$

5.3 Dealing with weak IV

- This problem arises when there is only a weak correlation between the instrument and the endogenous regressor
- Assume the following setting:

$$Y_i = \beta X_i + e_i \tag{34}$$
$$X_i = Z_i \pi + v_i \tag{35}$$

- The standard 2SLS estimator is $\beta_{2SLS} = \frac{X' P_Z Y}{X' P_Z X}$
- We make the extra assumption that error terms e_i and v_i are independently drawn from a normal distribution with variances σ_e^2 and σ_v^2, and a correlation ρ
- Let's introduce **the concentration parameter**:

$$\mu^2 = \pi' Z' Z \pi / \sigma_v^2 \tag{36}$$

5.3 Dealing with weak IV

- This problem arises when there is only a weak correlation between the instrument and the endogenous regressor
- Assume the following setting:

$$Y_i = \beta X_i + e_i \tag{34}$$
$$X_i = Z_i \pi + v_i \tag{35}$$

- The standard 2SLS estimator is $\beta_{2SLS} = \frac{X' P_Z Y}{X' P_Z X}$
- We make the extra assumption that error terms e_i and v_i are independently drawn from a normal distribution with variances σ_e^2 and σ_v^2, and a correlation ρ
- Let's introduce **the concentration parameter**:

$$\mu^2 = \pi' Z' Z \pi / \sigma_v^2 \tag{36}$$

5.3 Dealing with weak IV

- This problem arises when there is only a weak correlation between the instrument and the endogenous regressor
- Assume the following setting:

$$Y_i = \beta X_i + e_i \tag{34}$$

$$X_i = Z_i \pi + v_i \tag{35}$$

- The standard 2SLS estimator is $\beta_{2SLS} = \frac{X' P_Z Y}{X' P_Z X}$
- We make the extra assumption that error terms e_i and v_i are independently drawn from a normal distribution with variances σ_e^2 and σ_v^2, and a correlation ρ
- Let's introduce **the concentration parameter**:

$$\mu^2 = \pi' Z' Z \pi / \sigma_v^2 \tag{36}$$

5.3 Dealing with weak IV

- This problem arises when there is only a weak correlation between the instrument and the endogenous regressor
- Assume the following setting:

$$Y_i = \beta X_i + e_i \tag{34}$$
$$X_i = Z_i\pi + v_i \tag{35}$$

- The standard 2SLS estimator is $\beta_{2SLS} = \frac{X' P_Z Y}{X' P_Z X}$
- We make the extra assumption that error terms e_i and v_i are independently drawn from a normal distribution with variances σ_e^2 and σ_v^2, and a correlation ρ
- Let's introduce **the concentration parameter**:

$$\mu^2 = \pi' Z' Z\pi/\sigma_v^2 \tag{36}$$

- Consider also the following standard normal Gaussian random variables:

$$\xi_e = \frac{\pi' Ze}{\sqrt{\pi' Z' Z\pi}\sigma_e} \tag{37}$$

$$\xi_v = \frac{\pi' Zv}{\sqrt{\pi' Z' Z\pi}\sigma_v} \tag{38}$$

- Consider also the following quadratic forms of normal random variables with respect to P_Z:

$$S_{vv} = \frac{v' P_Z v}{\sigma_v^2} \tag{39}$$

$$S_{ev} = \frac{e' P_Z v}{\sigma_e \sigma_v} \tag{40}$$

- Consider also the following standard normal Gaussian random variables:

$$\xi_e = \frac{\pi' Ze}{\sqrt{\pi' Z' Z\pi}\sigma_e} \tag{37}$$

$$\xi_v = \frac{\pi' Zv}{\sqrt{\pi' Z' Z\pi}\sigma_v} \tag{38}$$

- Consider also the following quadratic forms of normal random variables with respect to P_Z:

$$S_{vv} = \frac{v' P_Z v}{\sigma_v^2} \tag{39}$$

$$S_{ev} = \frac{e' P_Z v}{\sigma_e \sigma_v} \tag{40}$$

- It can be shown that the joint distribution of normal variables and the quadratic forms is known, depends on ρ and does not depend on the sample size or π
- Under the assumptions above, Rothenberg (1984) derived the following exact finite-sample distribution of the 2SLS estimator:

$$\mu(\widehat{\beta}_{2SLS} - \beta_0) = \frac{\sigma_e}{\sigma_v} \cdot \frac{\xi_e + S_{ev}/\mu}{1 + 2\xi_v/\mu + S_{vv}/\mu^2} \tag{41}$$

- μ plays the role of sample size:
  - When μ tends to infinity, then $\mu(\widehat{\beta}_{2SLS} - \beta_0)$ asymptotically converges to a normal distribution
  - Otherwise, the finite sample distribution of $\widehat{\beta}_{2SLS}$ is non-standard and far from normal

- It can be shown that the joint distribution of normal variables and the quadratic forms is known, depends on ρ and does not depend on the sample size or π
- Under the assumptions above, Rothenberg (1984) derived the following exact finite-sample distribution of the 2SLS estimator:

$$\mu(\widehat{\beta}_{2SLS} - \beta_0) = \frac{\sigma_e}{\sigma_v} \cdot \frac{\xi_e + S_{ev}/\mu}{1 + 2\xi_v/\mu + S_{vv}/\mu^2} \tag{41}$$

- μ plays the role of sample size:
  - When μ tends to infinity, then $\mu(\widehat{\beta}_{2SLS} - \beta_0)$ asymptotically converges to a normal distribution
  - Otherwise, the finite sample distribution of $\widehat{\beta}_{2SLS}$ is non-standard and far from normal

- It can be shown that the joint distribution of normal variables and the quadratic forms is known, depends on ρ and does not depend on the sample size or π
- Under the assumptions above, Rothenberg (1984) derived the following exact finite-sample distribution of the 2SLS estimator:

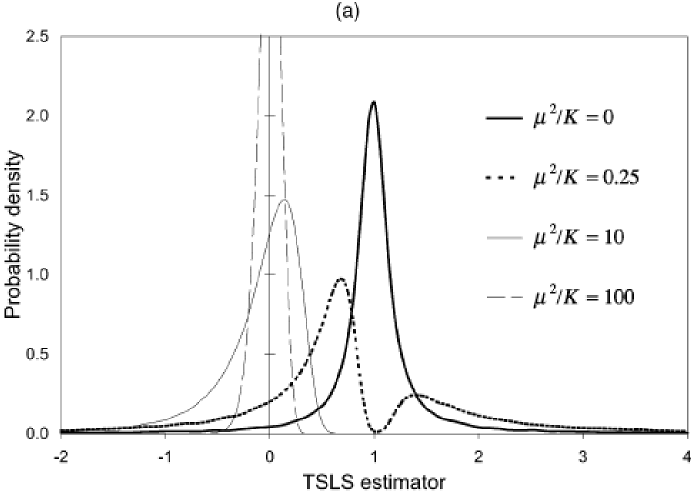
$$\mu(\widehat{\beta}_{2SLS} - \beta_0) = \frac{\sigma_e}{\sigma_v} \cdot \frac{\xi_e + S_{ev}/\mu}{1 + 2\xi_v/\mu + S_{vv}/\mu^2} \tag{41}$$

- μ plays the role of sample size:
  - When μ tends to infinity, then $\mu(\widehat{\beta}_{2SLS} - \beta_0)$ asymptotically converges to a normal distribution
  - Otherwise, the finite sample distribution of $\widehat{\beta}_{2SLS}$ is non-standard and far from normal

- μ depends on the size of π. When π is small (weak instrument), then the finite sample distribution of $\widehat{\beta}_{2SLS}$ is non-normal
- The idea of weak correlation depends on a significant way on the sample size n, however this does not mean that the problem of weak instruments is a "small sample problem"
  - Staiger and Stock(1997) show that, even for very large sample sizes, there are some values of the correlation between the instrument and the endogenous regressor for which the quality of the normal approximation is poor

- μ depends on the size of π. When π is small (weak instrument), then the finite sample distribution of $\widehat{\beta}_{2SLS}$ is non-normal
- The idea of weak correlation depends on a significant way on the sample size n, however this does not mean that the problem of weak instruments is a "small sample problem"
    - Staiger and Stock(1997) show that, even for very large sample sizes, there are some values of the correlation between the instrument and the endogenous regressor for which the quality of the normal approximation is poor

Probability density function for 2SLS estimator and the concentration parameter



Testing for Weak Instruments

- Testing for weak instruments requires to formally define what we mean by "weak"
- Stock and Yogo (2005) suggest two questions to think about this:
  - How small must μ^2 be for instruments to be weak?
  - Since π is unknown, how is an applied researcher to know whether μ^2 is in fact sufficient small and that his/her instruments are weak?
- Stock and Yogo (2005) define a weak instrument using two key ideas:
  - A set of instruments is strong if μ^2/r is sufficiently large so that the 2SLS relative bias (with respect to OLS) is at most (say) 10% (bias approach)
  - A set of instruments is strong if μ^2/r is large enough that a 5% hypothesis test rejects no more than (say) 15% of the time (size approach)

Testing for Weak Instruments

- Testing for weak instruments requires to formally define what we mean by "weak"
- Stock and Yogo (2005) suggest two questions to think about this:
  - How small must μ^2 be for instruments to be weak?
  - Since π is unknown, how is an applied researcher to know whether μ^2 is in fact sufficient small and that his/her instruments are weak?
- Stock and Yogo (2005) define a weak instrument using two key ideas:
  - A set of instruments is strong if μ^2/r is sufficiently large so that the 2SLS relative bias (with respect to OLS) is at most (say) 10% (bias approach)
  - A set of instruments is strong if μ^2/r is large enough that a 5% hypothesis test rejects no more than (say) 15% of the time (size approach)

Testing for Weak Instruments

- Testing for weak instruments requires to formally define what we mean by "weak"
- Stock and Yogo (2005) suggest two questions to think about this:
  - How small must μ^2 be for instruments to be weak?
  - Since π is unknown, how is an applied researcher to know whether μ^2 is in fact sufficient small and that his/her instruments are weak?
- Stock and Yogo (2005) define a weak instrument using two key ideas:
  - A set of instruments is strong if μ^2/r is sufficiently large so that the 2SLS relative bias (with respect to OLS) is at most (say) 10% (bias approach)
  - A set of instruments is strong if μ^2/r is large enough that a 5% hypothesis test rejects no more than (say) 15% of the time (size approach)

- They propose a test for weak instruments based on the F for the first stage. Recall that the first stage regression with one instrument is:

$$X_i = \pi' Z_i + \delta' W_i + v_i \tag{42}$$

- The test is a F test for the null hypothesis that μ^2/r is less than or equal to the weak instrument threshold against the alternative that it exceeds the threshold
  - Stock and Yogo (2005) computed the critical values for this F-statistic for both bias and size
  - The null hypothesis is that the coefficients on the elements of Z are all zero
  - When there is more than one endogenous variable X, a Cragg and Donald's (1993) statistic should be used instead of the F test

- They propose a test for weak instruments based on the F for the first stage. Recall that the first stage regression with one instrument is:

$$X_i = \pi' Z_i + \delta' W_i + v_i \tag{42}$$

- The test is a F test for the null hypothesis that μ^2/r is less than or equal to the weak instrument threshold against the alternative that it exceeds the threshold
  - Stock and Yogo (2005) computed the critical values for this F-statistic for both bias and size
  - The null hypothesis is that the coefficients on the elements of Z are all zero
  - When there is more than one endogenous variable X, a Cragg and Donald's (1993) statistic should be used instead of the F test

Critical values for the Stock and Yogo's (2005) test

Table 1. Selected Critical Values for Weak Instrument Tests for TSLS Based on the First-stage F statistic

Number of instruments (K)	Relative bias > 10%		Actual size of 5% test > 15%	
	Threshold μ^2/K	F statistic 5% critical value	Threshold μ^2/K	F statistic 5% critical value
1			1.82	8.96
2			4.62	11.59
3	3.71	9.08	6.36	12.83
5	5.82	10.83	9.20	15.09
10	7.41	11.49	15.55	20.88
15	7.94	11.51	21.69	26.80

NOTE: The second column contains the smallest values of μ^2/K that ensure that the bias of TSLS is no more than 10% of the inconsistency of OLS. The third column contains the 5% critical values applicable when the first-stage F statistic is used to test the null that μ^2/K is less than or equal to the value in the second column against the alternative that μ^2/K exceeds that value. The final two columns present the analogous weak-instrument thresholds and critical values when weak instruments are defined so that the usual nominal 5% TSLS t test of the hypothesis $\beta = \beta_0$ has size potentially exceeding 15%. (Source: Stock and Yogo 2001.)

- One limitation with the Stock and Yogo's (2005) test is that assumes that errors are conditionally homoskedastic and serially uncorrelated
- Montiel and Pflueger (2013) propose a extension of the Stock and Yogo (2005) that is robust to heteroskedasticity, autocorrelation and clustering:
  - They show that when instruments are weak, these problems can increase bias and distort test sizes
  - Their test evaluate the null hypothesis that the estimator approximate asymptotic bias exceed a threshold of a worst case benchmark related to the OLS bias when errors are conditionally homoskedastic and serially uncorrelated
  - They define an **effective F-statistic**, which is a scaled version of the non-robust first stage F:

$$\widehat{F}_{ef} = \frac{X' P_Z X}{Tr(\widehat{W})} \tag{43}$$

- One limitation with the Stock and Yogo's (2005) test is that assumes that errors are conditionally homoskedastic and serially uncorrelated
- Montiel and Pflueger (2013) propose a extension of the Stock and Yogo (2005) that is robust to heteroskedasticity, autocorrelation and clustering:
  - They show that when instruments are weak, these problems can increase bias and distort test sizes
  - Their test evaluate the null hypothesis that the estimator approximate asymptotic bias exceed a threshold of a worst case benchmark related to the OLS bias when errors are conditionally homoskedastic and serially uncorrelated
  - They define an **effective F-statistic**, which is a scaled version of the non-robust first stage F:

$$\widehat{F}_{ef} = \frac{X' P_Z X}{Tr(\widehat{W})} \tag{43}$$

5.4 Robust inference under weak instruments

- Once a weak instrument problem has been detected, the next step is to use inference techniques robust to weak instruments
- We are interested in testing the null hypothesis $H0 : \beta = 0$ in a context where confidence intervals are potentially of infinity lenght for standard procedures when the instrument is weak
- One approach to accurately perform inferences robust to weak instruments is to find statistics whose distributions do not depend on the value of the concentration parameter μ^2
- Two tests have been proposed:
  - Anderson and Rubin's (1949) statistic (AR)
  - Lagrange Multiplier (LM)

5.4 Robust inference under weak instruments

- Once a weak instrument problem has been detected, the next step is to use inference techniques robust to weak instruments
- We are interested in testing the null hypothesis $H0 : \beta = 0$ in a context where confidence intervals are potentially of infinity lenght for standard procedures when the instrument is weak
- One approach to accurately perform inferences robust to weak instruments is to find statistics whose distributions do not depend on the value of the concentration parameter μ^2
- Two tests have been proposed:
  - Anderson and Rubin's (1949) statistic (AR)
  - Lagrange Multiplier (LM)

5.4 Robust inference under weak instruments

- Once a weak instrument problem has been detected, the next step is to use inference techniques robust to weak instruments
- We are interested in testing the null hypothesis $H0 : \beta = 0$ in a context where confidence intervals are potentially of infinity lenght for standard procedures when the instrument is weak
- One approach to accurately perform inferences robust to weak instruments is to find statistics whose distributions do not depend on the value of the concentration parameter μ^2
- Two tests have been proposed:
  - Anderson and Rubin's (1949) statistic (AR)
  - Lagrange Multiplier (LM)

5.4 Robust inference under weak instruments

- Once a weak instrument problem has been detected, the next step is to use inference techniques robust to weak instruments
- We are interested in testing the null hypothesis $H0 : \beta = 0$ in a context where confidence intervals are potentially of infinity lenght for standard procedures when the instrument is weak
- One approach to accurately perform inferences robust to weak instruments is to find statistics whose distributions do not depend on the value of the concentration parameter μ^2
- Two tests have been proposed:
  - Anderson and Rubin's (1949) statistic (AR)
  - Lagrange Multiplier (LM)

a. Anderson-Rubin statistic

- The **AR statistic** is defined in the following way:

$$AR(\beta) = \frac{(y - \beta X)' P_Z (y - \beta X)/r}{(y - \beta X)' M_Z (y - \beta X)/(n - r)} \tag{44}$$

- Where $M_Z = 1 - P_Z$. Under quite general assumptions, the asymptotic distribution does not depend on μ^2 and converges in large samples to χ^2_r/r
- Large values of the AR statistic indicate violations of the null hypothesis

b. Lagrange multiplier statistic

- Consider the reduced-form from (6) and (7):

$$X_i = \beta\pi^{'} Z_i + w_i \tag{45}$$

- Where $w_i = e_i + \beta v_i$
- Let Ω be the covariance matrix of error terms $(w_i, v_i)'$. A natural estimator of Ω is $\widehat{\Omega} = \tilde{Y}' M_Z \tilde{Y}/(n-r)$, where $\tilde{Y} = [y, X]$
- Moreira (2002) introduced the following statistics:

$$\widehat{S} = \frac{(Z'Z)^{-1/2} Z' \tilde{Y} b}{\sqrt{b'\widehat{\Omega} b}} \tag{46}$$

$$\widehat{T} = \frac{(Z'Z)^{-1/2} Z' \tilde{Y} \widehat{\Omega}^{-1} a}{\sqrt{a'\widehat{\Omega}^{-1} a}} \tag{47}$$

- Where $b = [1, -\beta]$ and $a = [\beta, 1]'$
- The **LM statistic** is of the following form:

$$LM(\beta) = \frac{(\widehat{S}'\widehat{T})^2}{\widehat{T}'\widehat{T}} \tag{48}$$

- The LM statistic has an asymptotic distribution χ^2_1. A high value of the LM statistics indicates violation of the null hypothesis

c. Moreira's conditional approach

- Moreira(2003) has proposed a conditional approach that allows the use of critical values that are functions of the data instead of being fixed
- The **Conditional likelihood-ratio test** uses the statistic:

$$LR(\beta) = \frac{1}{2}\left(\widehat{S}'\widehat{S} - \widehat{T}'\widehat{T} + \sqrt{(\widehat{S}'\widehat{S} + \widehat{T}'\widehat{T})^2 - 4((\widehat{S}'\widehat{S})(\widehat{T}'\widehat{T}) - (\widehat{S}'\widehat{T})^2}\right)$$

- Under the conditional argument, Andrews et al (2007) computes conditional critical values that are computationally less intensive than the original Moreira's (2003) approach based on Monte Carlo simulations with stronger assumptions about Z
- The resulting test is asymptotically valid uniformly over all values of the concentration parameter (Mikusheva 2010)

5.5 Robust confidence intervals

- The previous tests can be inverted to derive confidence sets since the problem of constructing a confidence set is dual to the problem of testing
- The confidence sets for the AR and LM statistics can be derived analytically (see Mikusheva and Poi 2006)
- Mikusheva (2010) suggested an algorithm to invert the CLR test

2.6 Estimation under weak instruments

- 2SLS has poor properties under weak instruments. Therefore, the literature have proposed several alternative estimators that are asymptotically equivalent to 2SLS but with better finite-sample properties
- Some estimators:
  - Limited-information maximum likelihood (LIML)
  - Split-sample IV (Angrist and Krueger 1995)
  - Jackknife IV
  - Independently weighted IV

a. Limited-information maximum likelihood

- LIML is based on the assumption of joint-normality of errors in structural and first-stage equations
- LIML and 2SLS differ in finite-samples due to differences in terms of the weights placed on instruments
- Recent scholarship have shown that LIML has smaller bias than either 2SLS or GMM
- LIML and 2SLS are both special cases of k-class estimators. This class of estimator is defined as:

$$\widehat{\beta}_{k-class} = [X'(I - kM_Z)^{-1}X]^{-1}X'(I - kM_Z)^{-1}y \qquad (49)$$

- The LIML estimator sets k equal to the minimum eigenvalue of $(Y'M_Z Y)^{-1/2} Y' M_X Y (Y' M_Z Y)^{-1/2}$

- Fuller (1977) suggested a variant of LIML that obtain finite moments, especially when values of parameter a are equal to 1 (the estimator is approximately unbiased) or 4 (the estimator is biased but it has a lower mean squared error)
- Some debate about what is the best approach:
  - Some econometricians prefer Fuller's (1977) estimator because delivers more precise estimates
  - Others prefer LIML because its median is approximately equal to the coefficient of interest

b. Split-sample IV

- Recall the IV interpretation:

$$\widehat{\beta}_{IV} = (\widehat{X}'X)^{-1}\widehat{X}'y \tag{50}$$

- Where $\widehat{X}' = X'P_Z$. Substituting for y yields:

$$\widehat{\beta}_{IV} = \beta + (\widehat{X}'X)^{-1}\widehat{X}'\mu \tag{51}$$

- We know that β is consistent but is biased. Since X and μ are correlated, then $\widehat{X} = P_Z X$ is correlated with μ. Therefore, $E[\widehat{X}'\mu] \neq 0$
- An alternative is to use the instrument prediction $\tilde{X}$, which have the property of that $E[\tilde{X}'\mu] = 0$. Therefore,

$$\tilde{\beta}_{IV} = (\tilde{X}'X)^{-1}\tilde{X}'y \tag{52}$$

- Notice that this estimator is still biased but it is less biased compared to the original estimator because $E[\tilde{X}'\mu] = 0$ does not imply that $E[(\tilde{X}'X)^{-1}\tilde{X}'\mu] = 0$
- Angrist and Krueger (1995) proposed to obtain such instruments by splitting the sample into two sub-samples (y_1, X_1, Z_1) and (y_2, X_2, Z_2):
  - First sample (y_1, X_1, Z_1) is used to obtain estimates of π_1 from a regression of X_1 on Z_1
  - Second sample is used to obtain the IV estimator with the instrument $\tilde{X}_2 = Z_2\widehat{\pi}_1$
- Angrist and Krueger (1995) define the **unbiased split sample IV estimator** as follows:

$$\tilde{\beta}_{USSIV} = (\tilde{X}_2'X_2)^{-1}\tilde{X}_2'y_2 \tag{53}$$

- A variant is the split-sample IV $\tilde{\beta}_{SSIV} = (\tilde{X}_2'\tilde{X}_2)^{-1}\tilde{X}_2'y_2$ that is based on the Theil interpretation.
- These estimators have finite-samples bias towards zero, unlike the traditional IV estimators that are biased towards the OLS estimate
- Limitation: loss of efficiency