Correlation or Causality?

Wanchuan Lin

Guang Hua, PKU

April 1, 2008

Question of the day

- Before you came to my class, what did you hear about me?

Skirt Lengths and the Economy?

- Ralph Rotnem, a stock broker, once observed that the long-term trends of stock prices and of the hemlines on women's skirts appear to be in concert.

Everything has a unique history, here below is the story of skirts.

- Women in Paris had to wear pants before the debut of the skirts.

Everything has a unique history, here below is the story of skirts.

- Women in Paris had to wear pants before the debut of the skirts.
- In 1929,long skirts became popular. While, the Dow Jones Industrial Index (DJII) plunged from 395 to 195 and to 40 two years later.

Everything has a unique history, here below is the story of skirts.

- Women in Paris had to wear pants before the debut of the skirts.
- In 1929,long skirts became popular. While, the Dow Jones Industrial Index (DJII) plunged from 395 to 195 and to 40 two years later.
- In 1960's,DJII rushed to 1000. At the same time, short skirts showed up.

Everything has a unique history, here below is the story of skirts.

- Women in Paris had to wear pants before the debut of the skirts.
- In 1929,long skirts became popular. While, the Dow Jones Industrial Index (DJII) plunged from 395 to 195 and to 40 two years later.
- In 1960's,DJII rushed to 1000. At the same time, short skirts showed up.
- In 1970's,DJII fell to 590.Summer seemed cold these years and Paris women began to wear long skirts again.

Everything has a unique history, here below is the story of skirts.

- Women in Paris had to wear pants before the debut of the skirts.
- In 1929,long skirts became popular. While, the Dow Jones Industrial Index (DJII) plunged from 395 to 195 and to 40 two years later.
- In 1960's,DJII rushed to 1000. At the same time, short skirts showed up.
- In 1970's,DJII fell to 590.Summer seemed cold these years and Paris women began to wear long skirts again.
- In 1990's,mini skirt debuted. DJII rushed to 10000.

Everything has a unique history, here below is the story of skirts.

- Women in Paris had to wear pants before the debut of the skirts.
- In 1929,long skirts became popular. While, the Dow Jones Industrial Index (DJII) plunged from 395 to 195 and to 40 two years later.
- In 1960's,DJII rushed to 1000. At the same time, short skirts showed up.
- In 1970's,DJII fell to 590.Summer seemed cold these years and Paris women began to wear long skirts again.
- In 1990's,mini skirt debuted. DJII rushed to 10000.
- In 2000's, bikini became a nice choice for Paris girls. Guess what, DJII was high up to 13,000.

Look!! the history of the skirt is the history of DJII waving

- In 2008, the American stock market plunged because of the sub-prime crisis.

Look!! the history of the skirt is the history of DJII waving

- In 2008, the American stock market plunged because of the sub-prime crisis.
- Should we then conclude that the girls should only wear pants in the time of the subprime mortgage crisis?

Stock Price and lengths of skirts: What are the possible explanations?

- Can it be a simultaneity problem (i.e. Reverse Causality problem)?

Stock Price and lengths of skirts: What are the possible explanations?

- Can it be a simultaneity problem (i.e. Reverse Causality problem)?
  - Yes. Simultaneity arises when at least one of the explanatory variables is determined simultaneously along with y. If skirt length is determined partly as a function of stock price, then skirt length and the error term are generally correlated.

Stock Price and lengths of skirts: What are the possible explanations?

- Can it be an Omitted Variable Problem?

Stock Price and lengths of skirts: What are the possible explanations?

- Can it be an Omitted Variable Problem?
  - Yes. For example, *people's mood of the time* can be the omitted variable that affects both the stock price and the lengths of skirts. When people are in good mood, they invest more in the stock and at the same time they wear short skirts.

Stock Price and lengths of skirts: What are the possible explanations?

- Can it be an Omitted Variable Problem?
  - Yes. For example, *people's mood of the time* can be the omitted variable that affects both the stock price and the lengths of skirts. When people are in good mood, they invest more in the stock and at the same time they wear short skirts.
  - Yes. For example, *economic condition* can be an omitted variable. When economic conditions are good, stock price increases. In addition, when economic conditions are good, people's basic needs are fulfilled, which might decrease the crime rates and thus people feel safe to wear skirts.

Stock Price and lengths of skirts: What are the possible explanations?

- Can it be an Omitted Variable Problem?
  - Yes. For example, *people's mood of the time* can be the omitted variable that affects both the stock price and the lengths of skirts. When people are in good mood, they invest more in the stock and at the same time they wear short skirts.
  - Yes. For example, *economic condition* can be an omitted variable. When economic conditions are good, stock price increases. In addition, when economic conditions are good, people's basic needs are fulfilled, which might decrease the crime rates and thus people feel safe to wear skirts.
  - Of course, you can tell a similar story such as when the economic conditions are good, men become rich and thus girls try to seduce men by wearing short skirts. If one buys this story, then economic conditions is one omitted variable that affects both the stock price and the length of skirts.

Ice Cream and Crime ?

In a small Midwestern town, a phenomenon was discovered that defied any logic. The local police chief observes that as ice cream consumption increases, crime rates tend to increase as well. Quite simply, if you measured both, you would find the relationship was direct, meaning that as people eat more ice cream, the crime rate increases. And as you might expect, as they ate less ice cream, the crime rate went down.

- The two variables, crime rates and ice cream consumptions, must share something or have something in common with one another. It must be something that relates to both level of ice cream consumption and level of crime rate.

- The two variables, crime rates and ice cream consumptions, must share something or have something in common with one another. It must be something that relates to both level of ice cream consumption and level of crime rate.
- Can you guess what that is?

Omitted Variable Problem

- The *outside temperature* is what they both have in common. When it gets warm outside, such as in the summertime, more crimes are committed (it stays light longer, people leave the windows open, etc.). And because it is warmer, people enjoy the ancient treat and art of eating ice cream. And conversely, during the long and dark winter months, less ice cream is consumed and fewer crimes are committed as well.

Omitted Variable Problem

- The *outside temperature* is what they both have in common. When it gets warm outside, such as in the summertime, more crimes are committed (it stays light longer, people leave the windows open, etc.). And because it is warmer, people enjoy the ancient treat and art of eating ice cream. And conversely, during the long and dark winter months, less ice cream is consumed and fewer crimes are committed as well.
- The *economic conditions* might be another omitted variable. For example, when the economy is good, people are richer and thus buy more food including ice cream while at the same time when the economic conditions are good, the income inequality also increases exactly as a result of the high economic growth. This high income inequality might then lead to high crime rates because poor people feel less content.

Omitted Variable Problem

- The *outside temperature* is what they both have in common. When it gets warm outside, such as in the summertime, more crimes are committed (it stays light longer, people leave the windows open, etc.). And because it is warmer, people enjoy the ancient treat and art of eating ice cream. And conversely, during the long and dark winter months, less ice cream is consumed and fewer crimes are committed as well.
- The *economic conditions* might be another omitted variable. For example, when the economy is good, people are richer and thus buy more food including ice cream while at the same time when the economic conditions are good, the income inequality also increases exactly as a result of the high economic growth. This high income inequality might then lead to high crime rates because poor people feel less content.
- Any other possible explanations?

A Very Cleaver Paper which Makes Use of the Relationship between Crime Rates and Weather

- The Dynamics of Criminal Behavior: Evidence from Weather Shocks
  by Rian Jacob, Lars Lefgre and Enrico Moretti
  Journal of Human Resources, 42(3), 2007

Question of Interests

- Determine the true persistence of criminal activity by estimating the causal relationship between crime rates in different time periods within the same locality.

Question of Interests

- Determine the true persistence of criminal activity by estimating the causal relationship between crime rates in different time periods within the same locality.
- If there is a transitory increase in the cost (benefit) of crime at time t, so that criminal activity at time t decreases (increases), what happens to crime at time t+1?

Importance of the Question

- This issue is particularly salient today given the growing consensus among criminologists and law enforcement officials on the benefits of highly targeted crime prevention strategies that focus police resources on "hot spots" and "hot times" rather than relying on more broad-based interventions.

Importance of the Question

- This issue is particularly salient today given the growing consensus among criminologists and law enforcement officials on the benefits of highly targeted crime prevention strategies that focus police resources on "hot spots" and "hot times" rather than relying on more broad-based interventions.
- Law enforcement officials have long been worried that the benefits of *targeted crime prevention strategies may be mitigated by the temporal and spatial displacement of crime*.

Possible Reasons for the *Persistence* of Criminal Activity.

Higher crime today in any particular area is associated with higher crime tomorrow.

Possible Explanations?

1. Potential offenders are influenced by the criminal behavior of others.

Possible Reasons for the *Persistence* of Criminal Activity.

Higher crime today in any particular area is associated with higher crime tomorrow.

Possible Explanations?

1. Potential offenders are influenced by the criminal behavior of others.
2. The persistence in crime rates over time could also be explained by unobserved heterogeneities across localities. The persistence of unobserved factors that determine the costs ad benefits of criminal activity such as police presence and poverty levels will lead to a positive correlation in crime rates over time even in the absence of a true causal relationship.

Possible Reasons for a *Negative* Relationship Between Crime Rates

Possible Explanations?

- Displacement - the shifting of criminal activity from one time or location to another.

Possible Reasons for a *Negative* Relationship Between Crime Rates

Possible Explanations?

- Displacement - the shifting of criminal activity from one time or location to another.
    - In the case of violent crime, the evidence is consistent with a model where the marginal utility of violence is decreasing in the amount of violence committed during the prior week.

Possible Reasons for a *Negative* Relationship Between Crime Rates

Possible Explanations?

- Displacement - the shifting of criminal activity from one time or location to another.
  - In the case of violent crime, the evidence is consistent with a model where the marginal utility of violence is decreasing in the amount of violence committed during the prior week.
  - In the case of property crime, the existence of displacement is consistent with a model in which transitory fluctuations in the cost of crime create an income effect that is manifested for multiple periods.

Possible Reasons for a *Negative* Relationship Between Crime Rates

Possible Explanations?

- Displacement - the shifting of criminal activity from one time or location to another.
  - In the case of violent crime, the evidence is consistent with a model where the marginal utility of violence is decreasing in the amount of violence committed during the prior week.
  - In the case of property crime, the existence of displacement is consistent with a model in which transitory fluctuations in the cost of crime create an income effect that is manifested for multiple periods.
- E.g. A study of juvenile curfew in Detroit, for example, found that afternoon crime nearly doubled after the introduction of the curfew.

Identification Strategy - 2SLS

- Use $weather_{i,t-1}$ as IV for $crime_{i,t-1}$

Identification Strategy - 2SLS

- Use $weather_{i,t-1}$ as IV for $crime_{i,t-1}$
- First Stage

$$crime_{i,t-1} = \Gamma X_{it} + \gamma_1 weather_{i,t-1} + \gamma_2 weather_{it} + \eta_{it}$$

Identification Strategy - 2SLS

- Use $weather_{i,t-1}$ as IV for $crime_{i,t-1}$
- First Stage

$$crime_{i,t-1} = \Gamma X_{it} + \gamma_1 weather_{i,t-1} + \gamma_2 weather_{it} + \eta_{it}$$

- Second Stage

$$crime_{i,t} = BX_{it} + \beta_1 crime_{i,t-1} + \beta_2 weather_{it} + \varepsilon_{it}$$

Identification Strategy - 2SLS

- Use $weather_{i,t-1}$ as IV for $crime_{i,t-1}$
- First Stage

$$crime_{i,t-1} = \Gamma X_{it} + \gamma_1 weather_{i,t-1} + \gamma_2 weather_{it} + \eta_{it}$$

- Second Stage

$$crime_{i,t} = BX_{it} + \beta_1 crime_{i,t-1} + \beta_2 weather_{it} + \varepsilon_{it}$$

- where i is indexes for jurisdictions and t indexes for time period (weeks), X_{it} is a vector of control variables and weather it is a vector of weather variables.

Q1:Can the first stage error and second stage error term be uncorrelated with each other?

- The answer is NO. Why not?

Q1:Can the first stage error and second stage error term be uncorrelated with each other?

- The answer is NO. Why not?
- The reason we need IV is that $corr(\varepsilon_{it}, crime_{i,t-1}) \neq 0$.

$$
\begin{aligned}
corr(\varepsilon_{it}, crime_{i,t-1}) &\neq 0 \\
iff \quad corr(\varepsilon_{it}, \Gamma X_{it} + \gamma_1 weather_{i,t-1} + \gamma_2 weather_{it} + \eta_{it}) &\neq 0 \\
iff \;\; corr(\varepsilon_{it}, \eta_{it}) &\neq 0
\end{aligned}
$$

so by construction, $\quad corr(\varepsilon_{it}, \eta_{it}) \neq 0$
That is, the error term from the first stage and second stage are correlated by construction.

Q2: Should we include the second stage exogenous variables in our first stage estimation?

- The answer is yes!! If not, we will, in general, get in consistent estimators.

Q2: Should we include the second stage exogenous variables in our first stage estimation?

- The answer is yes!! If not, we will, in general, get in consistent estimators.
- but how do we prove that?

Q2: Should we include the second stage exogenous variables in our first stage estimation?

- The answer is yes!! If not, we will, in general, get in consistent estimators.
- but how do we prove that?
- Model

$$\begin{aligned} y_1 &= y_2\beta_1 + x_2\beta_2 + \varepsilon_1 \\ E(y_2\varepsilon_1) &\neq 0 \end{aligned} \quad (1)$$

y_2 is endogenous and we use z_1 as IV for y_2
assume that z_1 is a valid IV.

Below are our first and second stage regressions.
When we do not include x_2 in our first stage regression, our first stage regression takes the following form.

$$\begin{aligned} y_2 &= \alpha_0 + \alpha_1 z_1 + v_1 \\ y_2 &= \widehat{y}_2 + \widehat{v}_1 \\ & \quad \widehat{y}_2 \text{ is orthogonal to } \widehat{v}_1 \end{aligned}$$

From (1), we can get

$$\begin{aligned} y_1 &= (\widehat{y}_2 + \widehat{v}_1)\beta_1 + x_2\beta_2 + \varepsilon_1 \\ &= \widehat{y}_2\beta_1 + x_2\beta_2 + (\varepsilon_1 + \widehat{v}_1\beta_1) \\ &= \widehat{y}_2\beta_1 + x_2\beta_2 + u_1 \end{aligned}$$

Our second stage: Regress y_1 on $\widehat{y}_2$ and x_2. New error term is $u_1 \equiv \varepsilon_1 + \widehat{v}_1\beta_1$

Should we include the second stage exogenous variables in our first stage estimation?

- The answer is: It depends!!

Should we include the second stage exogenous variables in our first stage estimation?

- The answer is: It depends!!
- Depends on what?

Should we include the second stage exogenous variables in our first stage estimation?

- The answer is: It depends!!
- Depends on what?
- Case 1: *when y_2 and x_2 are correlated*,
  Then, $\widehat{v}_1$ and x_2 are correlated exactly because x_2 was not included in the linear projection for y_2.
  Therefore, $\widehat{\beta}_2$ is inconsistent. By (*Theorem*1), if x_2 is correlated with $\widehat{y}_2$, then $\widehat{\beta}_1$ is also inconsistent.

Should we include the second stage exogenous variables in our first stage estimation?

- The answer is: It depends!!
- Depends on what?
- Case 1: *when* y_2 *and* x_2 *are correlated*,
  Then, $\widehat{v}_1$ and x_2 are correlated exactly because x_2 was not included in the linear projection for y_2.
  Therefore, $\widehat{\beta}_2$ is inconsistent. By $(Theorem1)$, if x_2 is correlated with $\widehat{y}_2$, then $\widehat{\beta}_1$ is also inconsistent.
- Case2: *when* y_2 *and* x_2 *are uncorrelated*,
  then you are fine not to include the uncorrelated exogenous variable in your first stage regression.

See what Wooldridge says on this issue:

A Quote from Wooldridge Page 91
" the following seemingly sensible, two-step procedure is generally *inconsistent*:

1. regress x_k on $1, z_1, z_2, ...z_M$ and obtain the fitted value say $\widetilde{x}_k$;
2. run the regression of y on $1, x_{1,} x_2 .... x_{k-1}, \widetilde{x}_k$

- Problem 5.11 in Wooldridge's book asks you to show that the omitting $x_{1,} x_2 .... x_{k-1}$ in the first-stage regression and then explicitly doing the second stage regression produces *inconsistent* estimators of the $\beta_{j.}$"

- Note that this is not just an efficiency issue. By not including the second stage exogenous variable in the second stage, you get an inconsistent estimators.

In conclusion: Should we include the second stage exogenous variables in our first stage estimation?

Answer:

- Not including exogenous variables in the first stage is okay when your endogenous variable is not correlated with your exogenous variables.

In conclusion: Should we include the second stage exogenous variables in our first stage estimation?

Answer:

- Not including exogenous variables in the first stage is okay when your endogenous variable is not correlated with your exogenous variables.
- However, it doesn't hurt if you put exogenous variables that are uncorrelated with your endogenous variables in your first stage.

In conclusion: Should we include the second stage exogenous variables in our first stage estimation?

Answer:

- Not including exogenous variables in the first stage is okay when your endogenous variable is not correlated with your exogenous variables.
- However, it doesn't hurt if you put exogenous variables that are uncorrelated with your endogenous variables in your first stage.
  1. It is just that the coefficients for those exogenous variables are not going to be significantly different from zero.

In conclusion: Should we include the second stage exogenous variables in our first stage estimation?

Answer:

- Not including exogenous variables in the first stage is okay when your endogenous variable is not correlated with your exogenous variables.
- However, it doesn't hurt if you put exogenous variables that are uncorrelated with your endogenous variables in your first stage.
  1. It is just that the coefficients for those exogenous variables are not going to be significantly different from zero.
  2. That is not to say that more often than not, your endogenous variables are correlated with your exogenous variables.

In conclusion: Should we include the second stage exogenous variables in our first stage estimation?

Answer:

- Not including exogenous variables in the first stage is okay when your endogenous variable is not correlated with your exogenous variables.
- However, it doesn't hurt if you put exogenous variables that are uncorrelated with your endogenous variables in your first stage.
  1. It is just that the coefficients for those exogenous variables are not going to be significantly different from zero.
  2. That is not to say that more often than not, your endogenous variables are correlated with your exogenous variables.
- If the endogenous variable is correlated with exogenous variables, by not including the exogenous variables in the first stage, your IV estimator is inconsistent. Therefore, the general practice is that, people always put all of the exogenous variables in their first stage.

Theorem 1

Model:

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

Suppose
x1 and u is uncorrelated
x2 and u is correlated
x1 and x2 are correlated
Since x2 is endogenous, $\widehat{\beta}_2$ is inconsistent

- The question is whether $\widehat{\beta}_1$ is a consistent estimator or not

Theorem 1

Model:

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

Suppose
x1 and u is uncorrelated
x2 and u is correlated
x1 and x2 are correlated
Since x2 is endogenous, $\widehat{\beta}_2$ is inconsistent

- The question is whether $\widehat{\beta}_1$ is a consistent estimator or not

- The answer is NO!! (Theorem1)

Proof for Theorem 1

Model

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

Suppose
x1 and u is uncorrelated
x2 and u is correlated
x1 and x2 are correlated
Since x2 is endogenous, $\widehat{\beta}_2$ is inconsistent
The question is that whether $\widehat{\beta}_1$ is consistent or not?

Proof for Theorem 1 (Continued)

Model:

$$y_i = x'_{1i}\beta_1 + x'_{2i}\beta_2 + ui$$

$$\begin{aligned} where,\ E(x_{2i}, u_{i)} &\neq 0 \\ E(x_{1i}, u_i) &= 0 \end{aligned}$$

1. Is Beta1 consistent?

Proof for Theorem 1 (Continued)

Model:

$$y_i = x'_{1i}\beta_1 + x'_{2i}\beta_2 + ui$$

$$\begin{aligned} where,\ E(x_{2i}, u_{i)} &\neq 0 \\ E(x_{1i}, u_i) &= 0 \end{aligned}$$

1. Is Beta1 consistent?
2. Proof:

$$\begin{aligned}
\widehat{\beta}_1 &= (x'_1 M_2 x_1)^{-1} x'_1 M_2 y \\
&= (x'_1 M_2 x_1)^{-1} x'_1 M_2 (x'_1\beta_1 + x'_2\beta_2 + u) \\
&= \beta_1 + (x'_1 M_2 x_1)^{-1} x'_1 M_2 x'_2 \beta_2 + (x'_1 M_2 x_1)^{-1} x'_1 M_2 u \\
&= \beta_1 + (x'_1 M_2 x_1)^{-1} x'_1 M_2 u \quad \text{since } M_2 x'_2 = (I - x_2(x'_2 x_2)^{-1} x'_2) x'_2 = \\
&= \beta_1 + (x'_1 M_2 x_1)^{-1} (x'_1 (I - x_2 (x'_2 x_2)^{-1} x'_2) u) \\
&= \beta_1 + (x'_1 M_2 x_1)^{-1} (x'_1 u + (x'_1 x_2)(x'_2 x_2)^{-1} (x'_2 u)) \\
&= \beta_1 + (\frac{1}{n} x'_1 M_2 x_1)^{-1} (\frac{1}{n} x'_1 u + (\frac{1}{n} x'_1 x_2)(\frac{1}{n} x'_2 x_2)^{-1} (\frac{1}{n} x'_2 u))
\end{aligned}$$

What is wrong with the proof below?

- $$\begin{aligned}\hat{\beta}_1 &= \beta_1 + (x_1' M_2 x_1)^{-1}(x_1' u + (x_1' x_2)(x_2' x_2)^{-1}(x_2' u)) \\ &= \beta_1 + (x_1' M_2 x_1)^{-1}(x_1' u + (x_1' x_2)(x_2^{-1} x_2'^{-1})(x_2' u)) \\ &= \beta_1 + (x_1' M_2 x_1)^{-1}(x_1' u + x_1'(x_2 x_2^{-1})(x_2'^{-1} x_2')u) \\ &= \beta_1 + (x_1' M_2 x_1)^{-1}(x_1' u + x_1' u)\end{aligned}$$

What is wrong with the proof below?

- $$\begin{aligned}\widehat{\beta}_1 &= \beta_1 + (x_1' M_2 x_1)^{-1}(x_1' u + (x_1' x_2)(x_2' x_2)^{-1}(x_2' u)) \\ &= \beta_1 + (x_1' M_2 x_1)^{-1}(x_1' u + (x_1' x_2)(x_2^{-1} x_2'^{-1})(x_2' u)) \\ &= \beta_1 + (x_1' M_2 x_1)^{-1}(x_1' u + x_1'(x_2 x_2^{-1})(x_2'^{-1} x_2') u) \\ &= \beta_1 + (x_1' M_2 x_1)^{-1}(x_1' u + x_1' u)\end{aligned}$$

- $$\begin{aligned}p\lim \widehat{\beta}_1 &= \beta_1 + (p\lim(\frac{1}{n} x_1' M_2 x_1)^{-1} p\lim(\frac{1}{n} x_1' u + \frac{1}{n} x_1' u)) \\ \because &\quad E(x_1, u) = 0 \\ \therefore &\quad p\lim(\frac{1}{n} x_1' u) = 0 \\ \therefore &\quad p\lim \widehat{\beta}_1 = \beta_1\end{aligned}$$

What is wrong with the proof below?

- $$\begin{aligned}
\widehat{\beta}_1 &= \beta_1 + (x_1' M_2 x_1)^{-1}(x_1' u + (x_1' x_2)(x_2' x_2)^{-1}(x_2' u)) \\
&= \beta_1 + (x_1' M_2 x_1)^{-1}(x_1' u + (x_1' x_2)(x_2^{-1} x_2'^{-1})(x_2' u)) \\
&= \beta_1 + (x_1' M_2 x_1)^{-1}(x_1' u + x_1'(x_2 x_2^{-1})(x_2'^{-1} x_2') u) \\
&= \beta_1 + (x_1' M_2 x_1)^{-1}(x_1' u + x_1' u)
\end{aligned}$$

- $$\begin{aligned}
p\lim \widehat{\beta}_1 &= \beta_1 + (p\lim(\frac{1}{n} x_1' M_2 x_1)^{-1} p\lim(\frac{1}{n} x_1' u + \frac{1}{n} x_1' u)) \\
\because &\quad E(x_1, u) = 0 \\
\therefore &\quad p\lim(\frac{1}{n} x_1' u) = 0 \\
\therefore &\quad p\lim \widehat{\beta}_1 = \beta_1
\end{aligned}$$

- What is wrong with the proof above?

- What is the condition for

$$(AB)^{-1} = B^{-1}A^{-1}$$

to hold?

- What is the condition for

$$(AB)^{-1} = B^{-1}A^{-1}$$

  to hold?
- A and B are square matrixes and the inverse of both matrixes exist.

Proof for Theorem 1 (Continued)

$$
\begin{aligned}
p\lim \widehat{\beta}_1 &= \beta_1 + (p\lim(\frac{1}{n}x_1' M_2 x_1)^{-1})(p\lim(\frac{1}{n}x_1' u) \\
&\quad + p\lim((\frac{1}{n}x_1' x_2)(\frac{1}{n}x_2' x_2)^{-1}(\frac{1}{n}x_2' u)) \\
&= \beta_1 + p\lim((\frac{1}{n}x_1' x_2)(\frac{1}{n}x_2' x_2)^{-1}(\frac{1}{n}x_2' u)) \\
&\because \quad p\lim(\frac{1}{n}x_1' u) = p\lim \frac{1}{n}\Sigma x_{1i} u_i = E(x_{1i} ui) = 0
\end{aligned}
$$

$$
\begin{aligned}
p\lim \widehat{\beta}_1 &= \beta_1 \text{ iff} \\
p\lim(\frac{1}{n}x_1' x_2(\frac{1}{n}x_2' x_2)^{-1}\frac{1}{n}x_2' u)) &= 0 \quad \text{iff} \ (\text{ note : } E(x_{2i}, u_i) \neq 0) \\
p\lim \frac{1}{n}x_1' x_2 &= 0 \quad \text{iff} \\
E(x_{1i} x_{2i}') &= 0
\end{aligned}
$$

QED

Identification Strategy - 2SLS

Use $weather_{i,t-1}$ as IV for $crime_{i,t-1}$
First Stage

$$crime_{i,t-1} = \Gamma X_{it} + \gamma_1 weather_{i,t-1} + \gamma_2 weather_{it} + \eta_{it}$$

Second Stage

$$crime_{i,t} = BX_{it} + \beta_1 crime_{i,t-1} + \beta_2 weather_{it} + \varepsilon_{it}$$

where i is indexes for jurisdictions and t indexes for time period (weeks), X_{it} is a vector of control variables and weather it is a vector of weather variables.

- What is the underlying identification assumption for the instrument to be valid?

Two Identification Assumptions

1. Rank condition

$$\gamma_1 \neq 0$$

Two Identification Assumptions

1. Rank condition

$$\gamma_1 \neq 0$$

2. The instrument must be exogenous

$$cov(\varepsilon_{it}, weather_{i,t-1} | weather_{i,t}, X_{it}) = 0$$

Two Identification Assumptions

1. Rank condition

$$\gamma_1 \neq 0$$

2. The instrument must be exogenous

$$cov(\varepsilon_{it}, weather_{i,t-1} | weather_{i,t}, X_{it}) = 0$$

- What does this mean in Chinese?

Two Identification Assumptions

1. Rank condition

$$\gamma_1 \neq 0$$

2. The instrument must be exogenous

$$cov(\varepsilon_{it}, weather_{i,t-1} | weather_{i,t}, X_{it}) = 0$$

- What does this mean in Chinese?
- or in English?

Two Identification Assumptions

1. Rank condition

$$\gamma_1 \neq 0$$

2. The instrument must be exogenous

$$cov(\varepsilon_{it}, weather_{i,t-1} | weather_{i,t}, X_{it}) = 0$$

- What does this mean in Chinese?
- or in English?
- Conditional on weather at time t and other covariates, weather at time $t-1$ cannot directly influence crime at time t.

Threats to Identification 1

1. In the absence of good controls for current weather, lagged weather may be directly correlated with current crime because it will contain information regarding current weather.
2. Some unobserved aspects of current weather might be in the ε_{it} and thus ε_{it} is correlated with $weather_{i,t-1}$ because weather, when measured at high frequency, is serially correlated.

Problem

(Homework): Show that the combination of serial correlation in weather, along with imperfect measures of weather conditions in any one period, will violate the assumptions necessary for satisfactory identification and result in a positive bias in the coefficient on lagged crime.

Treats to Identification 2

1. Weather may affect the intensity of non-criminal activities, and therefore, may directly affect the benefits of crime. For example, if weather affects the number of shoppers in a shopping mall, the opportunities for car theft at the mall may also be affected.
2. The identification might yield biased estimates if (1) weather displaces non-criminal activities and (2) the amount of these non-criminal activities affects the benefits/costs of criminal activity.

Any other Identification Treats you can come up with?

1. Whether to commit crime in period t depends on individual wealth at time t and this wealth is not controlled from in our second stage regression. And suppose that bad weather in period t-1 destroys people's wealth at time t-1 and the wealth at time t-1 is correlated with wealth at time t. Therefore, we prove that if this story is true, then the IV is not valid.

Problem

(homework) Find one paper that use weather as instrumental variable for something. State the identification assumptions and possible identification threats.

Problem

(homework) Find out how Jacob et al deal with the above two identification threats.

Problem

(homework) Find two econ papers that use IV. You should answer the following questions. 1. List the name, authors, journal published for the paper. 2. What are the IV(s) they use? 3. What are their first and second stage regressions look like? 4. What is their identification assumptions? 5. What are the possible identification threats? (List at least two) 6. What are the solutions the authors use to deal with the identification threats?

IV Regression with Interaction Terms

- David Card, (1995). "Using Geographic Variation in College Proximity to Estimate the Return to Schooling". In L.N. Christofides, E.K. Grant, and R. Swidinsky, editors, Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp. Toronto: University of Toronto Press, 1995.

IV Regression with Interaction Terms

- David Card, (1995). "Using Geographic Variation in College Proximity to Estimate the Return to Schooling". In L.N. Christofides, E.K. Grant, and R. Swidinsky, editors, Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp. Toronto: University of Toronto Press, 1995.
- (College Proximity as an IV for Education): Using wage data for 1976, Card (1995) uses a dummy variable that indicates whether a man grew up in the vicinity of a four-year college as an instrumental variable for years of schooling. He also includes several other controls. In the equation with experience and its square, a black indicator, southern and urban indicators, and regional and urban indicators for 1966, the instrumental variables estimate of the return to schooling is .132, or 13.2 percent, while the OLS estimate is 7.5 percent. The IV estimate is almost twice as large as the OLS estimate. This result would be counterintuitive if we thought that an OLS analysis suffered from an upward omitted variable bias.

$$\log(wage) = \alpha_0 + \alpha_1 \cdot edu + \alpha_2 \cdot black + \alpha_3 \cdot X + \varepsilon \quad (2)$$

Card (1995) use "nearcollege" as IV for education in equation(2).
Suppose we want to investigate whether the return for education is higher for African Americans or not. What should then be the econometric specification?
We want to estimate equation (3) instead.

$$\log(wage) = \beta_0 + \beta_1 \cdot edu + \beta_2 \cdot black \cdot edu + \beta_3 \cdot black + \beta_4 \cdot X + \varepsilon \quad (3)$$

Problem

*(Homework): Explain why black*nearcollege is a potential IV for black*edu*

Problem

*(Homework): Suppose instead of using black*nearcollege as IV for black*edu, now we use 2SLS with black* $\widehat{edu}$ *as IV for black*edu, where* $\widehat{edu}$ *are the fitted values from the first stage regression of edu on all exogenous variables (including nearcollege). What do you find?*

Problem

*(Homework): Explain the differences between using black*edu as IV and using black* $\widehat{edu}$ *as IV. Are both IV estimators consistent? If no, why not? If yes, which one is more efficient than the other?*

IV for Just Identified Case

The term Just Identified means that the number of instruments equals to the number of endogenous variables.

Given the regression equation $y = X\beta + \varepsilon$, Let z_i is for one individual and we have n individuals. Suppose that there exists some $Z = (z_1, z_2, ..., z_n)^T$ such that:

- $p\lim_n \frac{1}{n} Z^T X = p\lim_n \frac{1}{n} \sum_{i=1}^{n} z_i x_i^T = E\left(z_i x_i^T\right) \equiv \sum_{ZX}$
  where $\sum_{ZX}$ is a $K \times K$ matrix with $\det\left(\sum_{ZX}\right) \neq 0$
- $p\lim_n \frac{1}{n} Z^T \varepsilon = p\lim_n \frac{1}{n} \sum_{i=1}^{n} z_i \varepsilon_i = E\left(z_i \varepsilon_i\right) \equiv \sum_{Z\varepsilon} = 0$
  where $\sum_{Z\varepsilon}$ is a $K \times 1$ vector and 0 is a $K \times 1$ vector of zero $\left(E\left(z_i \varepsilon_i\right) = 0, \forall i\right)$.

Then the variables z_i are called instrumental variables.

a. IV Estimator

We can easy get the IV estimator $\widehat{\beta}_{IV}$ of β:

$$\begin{aligned}\widehat{\beta}_{IV} &= \left(Z^T X\right)^{-1} Z^T y \\ &= \left(Z^T X\right)^{-1} Z^T \left(X\beta + \varepsilon\right) \\ &= \beta + \left(\frac{1}{n} Z^T X\right)^{-1} \left(\frac{1}{n} Z^T \varepsilon\right)\end{aligned}$$

So

$$p\lim_n \widehat{\beta}_{IV} = \beta + \left(p\lim_n \frac{1}{n} Z^T X\right)^{-1} \left(p\lim_n \frac{1}{n} Z^T \varepsilon\right)$$

By **WLLN**,

$$p\lim_n \frac{1}{n} Z^T X = p\lim_n \frac{1}{n} \sum_i z_i x_i^T = E\left(z_i x_i^T\right) = \sum_{ZX}$$

$$p\lim_n \frac{1}{n} Z^T \varepsilon = p\lim_n \frac{1}{n} \sum_i z_i \varepsilon_i = E\left(z_i \varepsilon_i\right) = \sum_{Z\varepsilon} = 0$$

Therefore,

$$p\lim_n \widehat{\beta}_{IV} = \beta$$

So IV estimator is a consistent estimator of β.

b. Method of Moments Interpretation of IV estimator

From the proof of the consistent of IV estimator, the key property is

$$E\left(z_i\varepsilon_i\right) = \sum_{Z\varepsilon} = 0.$$

The property of IV solves the problem exactly and $\widehat{\beta}_{IV}$ makes $\frac{1}{n}\sum_i z_i\widehat{\varepsilon}_i = \frac{1}{n}\sum_i z_i\left(y_i - x_i^T\widehat{\beta}_{IV}\right) = 0$ hold, where $\widehat{\varepsilon}_i = y_i - x_i^T\widehat{\beta}_{IV}$.

c. Asymptotic Properties of IV Estimators

Suppose that $E\left(z_i\varepsilon_i\right)=0$, then by **CLT**

$$\frac{1}{\sqrt{n}}Z^T\varepsilon=\frac{1}{\sqrt{n}}\sum_i z_i\varepsilon_i \xrightarrow{d} N\left(0,\Phi\right)$$

where $\Phi=E\left(\varepsilon_i^2 z_i z_i^T\right)$

By Slutsky Theorem, the product $\left(\frac{1}{\sqrt{n}}\sum_i z_i x_i^T\right)^{-1}\frac{1}{\sqrt{n}}\sum_i z_i\varepsilon_i$ converges to the product of two limits, i.e.,

$$\begin{aligned}\sqrt{n}\left(\widehat{\beta}_{IV}-\beta\right) &= \left(\frac{1}{n}Z^T X\right)^{-1}\left(\frac{1}{n}Z^T\varepsilon\right)\\ &= \left(\frac{1}{n}\sum_i z_i x_i^T\right)^{-1}\frac{1}{\sqrt{n}}\sum_i z_i\varepsilon_i\end{aligned}$$

$$\xrightarrow{d} N\left(0,\sum\nolimits_{ZX}^{-1}\Phi\sum\nolimits_{ZX}^{-1}\right)$$

- Consistent estimator for $\sum_{ZX}$

$$\widehat{\sum}_{ZX} = \frac{1}{n} \sum_i z_i x_i^T$$

- Consistent estimator for $\Phi = E\left(\varepsilon_i^2 z_i z_i^T\right)$.

$$\widehat{\Phi} = \frac{1}{n} \sum_i \widehat{\varepsilon}_i^2 z_i z_i^T \xrightarrow{p} \Phi$$

This is known as Heteroskedasticity-Robust (Robust, White or Eicker-White) Estimator.
Under conditional homoskedasticity, i.e., $E\left(\varepsilon_i^2 | z_i\right) = \sigma^2$, then $\widehat{\Phi}$ simplified to $\widehat{\Phi} = \widehat{\sigma}^2 \left(\frac{1}{n} Z^T Z\right)$

where $\widehat{\sigma}^2$ is any consistent estimator of σ^2, e.g., the unbiased estimator $\frac{1}{n-K} \sum_i \widehat{\varepsilon}_i^2$ or the biased one $\frac{1}{n} \sum_i \widehat{\varepsilon}_i^2$.

2SLS Estimation in Stata

assuming conditional homoskedasticity...

Second Stage : price = a0+a1*price+a2*displacement+a3*trunk+e
trunk is endogenous
use headroom to iv for trunk
First Stage : trunk=b0+b1*headroom+b2*price+b3*displacement

2SLS Estimation in Stata

assuming conditional homoskedasticity...

```
sysuse auto, clear
ivreg price displacement (trunk=headroom)
```

Just one line command in stata!!

Can you do the IV estimation manually?

assuming conditional homoskedasticity...

- Being able to do estimation by yourself is important when you need to do more advanced stuff.

Can you do the IV estimation manually?

assuming conditional homoskedasticity...

- Being able to do estimation by yourself is important when you need to do more advanced stuff.
- Below is the code that you can use to produce the above IV regression results.

Can you do the IV estimation manually?

assuming conditional homoskedasticity...

- Being able to do estimation by yourself is important when you need to do more advanced stuff.
- Below is the code that you can use to produce the above IV regression results.
-
```
* BEGIN OF DO FILE
sysuse auto, clear
ivreg price displacement (trunk=headroom)
regress trunk headroom disp
predict double trunk_hat
regress price trunk_hat displacement
local dof=e(df_r) /* Where dof corresponds to the degrees
of freedom of the residuals from the
previous regression */
```

- ```
  replace trunk_hat=trunk
  predict double e_hat,residuals
  generate double ee_hat=e_hat^2
  summarize ee_hat, meanonly
  scalar sigsq_hat_pr= r(sum)/`dof' /* Signa hat square */
  matrix V=e(V)
  matrix xx_1=e(V)/(e(rmse)^2)
  matrix V=sigsq_hat_pr*xx_1
  matrix b=e(b) /* Where e(b) contains the point estimates
  from the second stage regression */
  ereturn post b V, dof(`dof')
  ereturn display
  ```
  ```

- replace trunk_hat=trunk
  predict double e_hat,residuals
  generate double ee_hat=e_hat^2
  summarize ee_hat, meanonly
  scalar sigsq_hat_pr= r(sum)/‘dof’ /* Signa hat square */
  matrix V=e(V)
  matrix xx_1=e(V)/(e(rmse)^2)
  matrix V=sigsq_hat_pr*xx_1
  matrix b=e(b) /* Where e(b) contains the point estimates
  from the second stage regression */
  ereturn post b V, dof(‘dof’)
  ereturn display
- * END OF DO FILE

2SLS Estimation

assume conditional homoskedasticity...

Problem

(Homework) Please derive the asy. variance of the IV estimator by yourself and then link the theoretical framework with the stata commands. For each of the above stata commands, please write a brief explanation on what is each command trying to achieve and its connection with the theoretical framework that you derived.

2SLS Estimation-

under heteroskedasticity!!!

```
sysuse auto, clear
ivreg price displacement (trunk=headroom),robust
```

Problem

(Homework) The above stata code is written assuming conditional homoskedasticity, however, under heteroskedasticity, you have to add robust after your ivreg command. Your task now is to write stata commands that can produce results that are identical to the results reported by using "ivreg price displacement (trunk=headroom),robust" command.

Hint: The standard errors that calculated from the codes above are the correct standard errors if you do not specify the -robust- option in -ivreg-. For a discussion on how to adjust this variance to be robust, please refer to pages 97-99 of the Stata 10 [R] Q-Z manual. Specifically on page 99, you will see formulas that apply to the case where you have instrumental variables. Note that you can check out STATA 10 Manual from the Guang Hua Library.

Please type your answer and email it to appliedeconometrics2008@gmail.com by 4/7 midnight. You can work independently or form a group of two people. Please include also the questions in your answer. Please use "homework 1" followed by your name(s) as the title of your email. Thanks and Good Luck!!

Thanks You!!

Q & A