Instrumental Variables[1]

STA431 Winter/Spring 2015

[1]See last slide for copyright information.

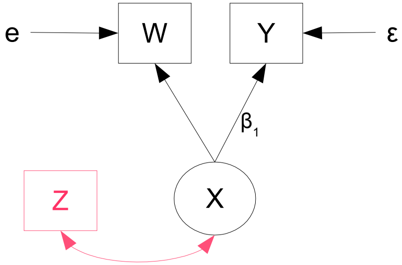
The problem

- Double measurement solves the measurement error identifiability problem.
- But sometimes two independent measurements are not available.
- Often, the data are already collected.
- Maybe collected for some other purpose.
- For example, companies might be required to provide information about water pollution, but they're not going to measure it twice.

Definition

Definition: An *instrumental variable* for an explanatory variable is an observable response variable that has

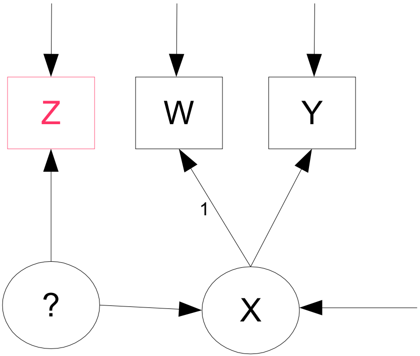
- Non-zero covariance with the explanatory variable.
- Zero covariance with the error term of the regression.



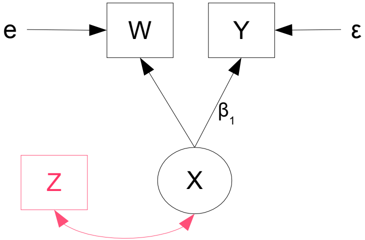
- Could have $X \to Z$, $Z \to X$, or something more complicated.

Example of an instrumental variable

- Participants in the survey are real estate agents, nationwide.
- X is income.
- Y is credit card debt.
- Z is median selling price of a home in the sales area.
- The instrumental variable is Z.



Identifiability



- Consider the covariance matrix only.
- All variables are invisibly centered.

$$
\begin{aligned}
Y_i &= \beta_1 X_i + \epsilon_i \\
W_i &= X_i + e_i
\end{aligned}
$$

$$
V\left(\begin{array}{c} X_i \\ Z_i \end{array}\right) = \left(\begin{array}{cc} \phi_{11} & \phi_{12} \\ & \phi_{22} \end{array}\right), \; Var(\epsilon) = \psi, \; Var(e) = \omega
$$

Covariance structure equations

$$
\begin{array}{rcl}
Y_i &=& \beta_1 X_i + \epsilon_i \\
W_i &=& X_i + e_i
\end{array}
$$

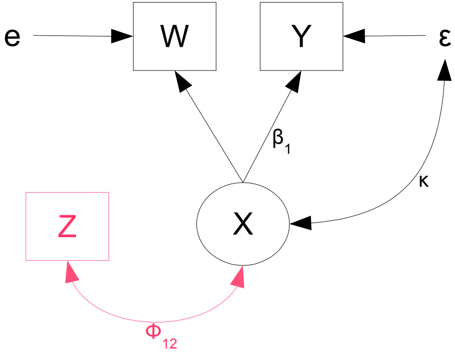
$$
V\left(\begin{array}{c} X_i \\ Z_i \end{array} \right) = \left(\begin{array}{cc} \phi_{11} & \phi_{12} \\ & \phi_{22} \end{array} \right), \; Var(\epsilon) = \psi, \; Var(e) = \omega
$$

Checking the parameter count rule …

$$
\boldsymbol{\Sigma} = \left(\begin{array}{ccc} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ & \sigma_{22} & \sigma_{23} \\ & & \sigma_{33} \end{array} \right) =
\begin{array}{c|ccc}
 & W & Z & Y \\ \hline
W & \phi_{11} + \omega & \phi_{12} & \beta_1 \phi_{11} \\
Z & & \phi_{22} & \beta_1 \phi_{12} \\
Y & & & \beta_1^2 \phi_{11} + \psi
\end{array}
$$

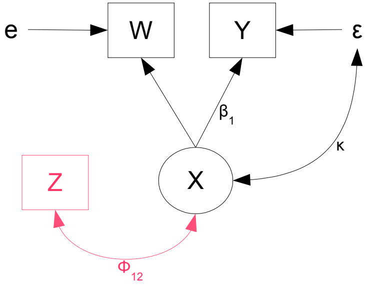
A miracle (Phillip Wright, 1928)

- Other things influence credit card debt besides income.
- And they are no doubt correlated with income.
- The deadly problem of omitted variables.
- Remember, the ϵ in a regression means "everything else."



Covariance structure equations are *mostly* the same

Six equations in seven unknowns



$$\begin{array}{rcl} Y_i & = & \beta_1 X_i + \epsilon_i \\ W_i & = & X_i + e_i \\ V\left(\begin{array}{c} X_i \\ Z_i \\ \epsilon_i \end{array}\right) & = & \left(\begin{array}{ccc} \phi_{11} & \phi_{12} & \kappa \\ & \phi_{22} & 0 \\ & & \psi \end{array}\right) \end{array}$$

$$\boldsymbol{\Sigma} = \left(\begin{array}{ccc} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ & \sigma_{22} & \sigma_{23} \\ & & \sigma_{33} \end{array}\right) = \begin{array}{c|ccc} & W & Z & Y \\ \hline W & \phi_{11}+\omega & \phi_{12} & \beta_1\phi_{11} + \kappa \\ Z & & \phi_{22} & \beta_1\phi_{12} \\ Y & & & \beta_1^2\phi_{11} + 2\beta_1\kappa + \psi \end{array}$$

We only care about β_1 anyway.

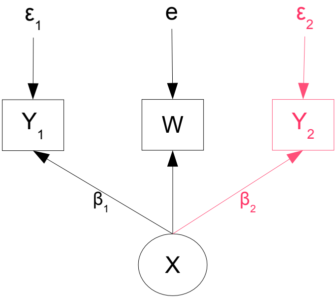
Generalizing ...

$$\boldsymbol{\Sigma} = \left(\begin{array}{ccc} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ & \sigma_{22} & \sigma_{23} \\ & & \sigma_{33} \end{array} \right) = \begin{array}{c|ccc} & W & Z & Y \\ \hline W & \phi_{11}+\omega & \phi_{12} & \beta_1\phi_{11}+\kappa \\ Z & & \phi_{22} & \beta_1\phi_{12} \\ Y & & & \beta_1^2\phi_{11}+2\beta_1\kappa+\psi \end{array}$$

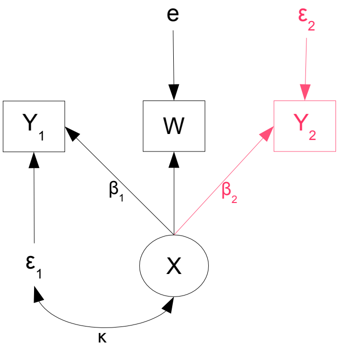
- Matrix version is straightforward.
- The usual rule in Econometrics is (at least) one instrumental variable for each explanatory variable.
- The $p \times p$ matrix of covariances between $\mathbf{X}$ and $\mathbf{Z}$ must have an inverse.
- Instrumental variables are related to $\mathbf{X}$ for reasons that are *separate* from why $\mathbf{X}$ is related to $\mathbf{Y}$.
- For example, does academic ability contribute to higher salary?
    - Study adults who were adopted as children.
    - X is academic ability.
    - Y is salary at age 40.
    - W is measured IQ.
    - Z is birth mother's IQ (there are studies like this).

Watch Out! Independence of the instrumental variable and ϵ is critical.

Y_2 is an instrumental variable.



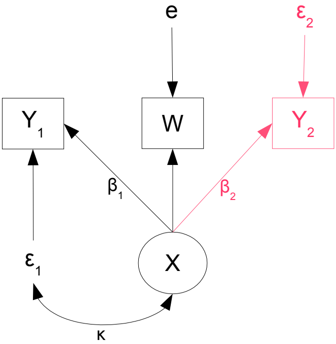
Y_2 is *not* an instrumental variable.



$$E\{Y_2\epsilon_1\} = E\{(\beta_2 X + \epsilon_2)\epsilon_1\} = \beta_2 E\{X\epsilon_1\} + E\{\epsilon_1\}E\{\epsilon_2\} = \beta_2\kappa.$$

The second model

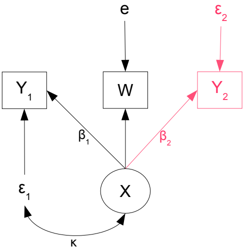
All variables are centered



$$
\begin{aligned}
W_i &= X_i + e_i \\
Y_{i,1} &= \beta_1 X_i + \epsilon_{i,1} \\
Y_{i,2} &= \beta_2 X_i + \epsilon_{i,2}
\end{aligned}
$$

$$
\boldsymbol{\Sigma} = \left(\begin{array}{ccc} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ & \sigma_{22} & \sigma_{23} \\ & & \sigma_{33} \end{array} \right) =
\begin{array}{c|ccc}
 & W & Y_1 & Y_2 \\ \hline
W & \phi + \omega & \beta_1 \phi + \kappa & \beta_2 \phi \\
Y_1 & & \beta_1^2 \phi + 2\beta_1 \kappa + \psi_1 & \beta_2(\beta_1 \phi + \kappa) \\
Y_2 & & & \beta_2^2 \phi + \psi_2
\end{array}
$$

β_1 is not identifiable



$$\boldsymbol{\Sigma} = \begin{array}{c|ccc} & W & Y_1 & Y_2 \\ \hline W & \phi+\omega & \beta_1\phi+\kappa & \beta_2\phi \\ Y_1 & & \beta_1^2\phi+2\beta_1\kappa+\psi_1 & \beta_2(\beta_1\phi+\kappa) \\ Y_2 & & & \beta_2^2\phi+\psi_2 \end{array}$$

- Infinitely many (β_1, κ) pairs yield the same covariance matrix.
- β_1 could be positive, negative or zero and you can't tell.
- β_1 and κ conceal one another.

Conclusions

- Instrumental variables can potentially solve the problems of omitted variables and measurement error at the same time.
- The recipe is one instrumental variable for each latent explanatory variable.
- If you can find them.
- If there is correlation between an explanatory variable and the error term of the regression (because of omitted variables), an additional response variable that is influenced by that explanatory variable is *not* an instrumental variable.
- The ultimate instrumental variable is an experimental manipulation.
- Sometimes it's a "natural experiment," like tax rates. (Tobacco taxes, smoking and health.)

Copyright Information