Instrumental Variables Again[1]

STA2101 Fall 2019

[1] See last slide for copyright information.

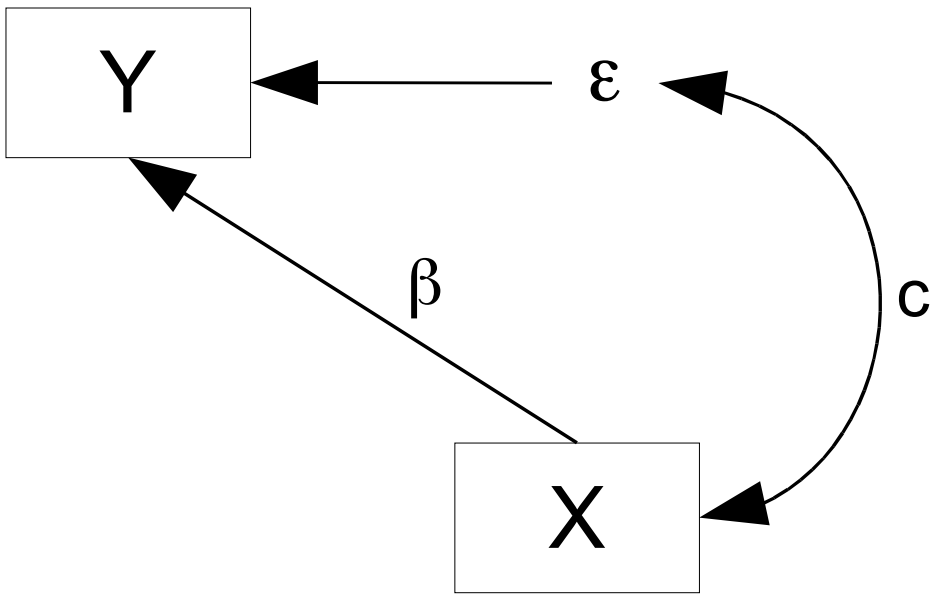
Overview

1. Omitted Variables
2. Including Measurement Error

Remember the problem of omitted variables

Example: X is income, Y is credit card debt.

- Omitted explanatory variables are part of the error term.
- Usually they are correlated with explanatory variables that are in the model.
- This makes the error term correlated with X.



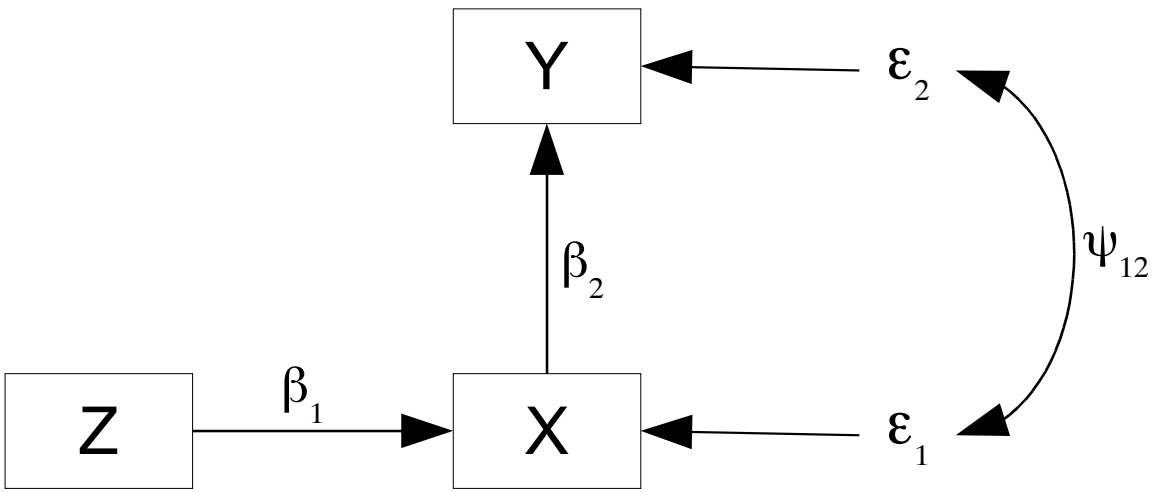
- Parameters are not identifiable.
- Estimation and inference fail.

Instrumental variable method saved the day

Phillip Wright, 1928

An instrumental variable (for an explanatory variable)

- Is related to the explanatory variable in question.
- Is unrelated to any error term in the model.
- Is connected to the response variable only through X.



- Real estate agents: X is income, Y is credit card debt, Z is median home price.
- Interest is in β_2.

Technically everything worked great

$X_i = \alpha_1 + \beta_1 W_i + \epsilon_{i1}$ and $Y_i = \alpha_2 + \beta_2 X_i + \epsilon_{i2}$

$\boldsymbol{\Sigma} =$

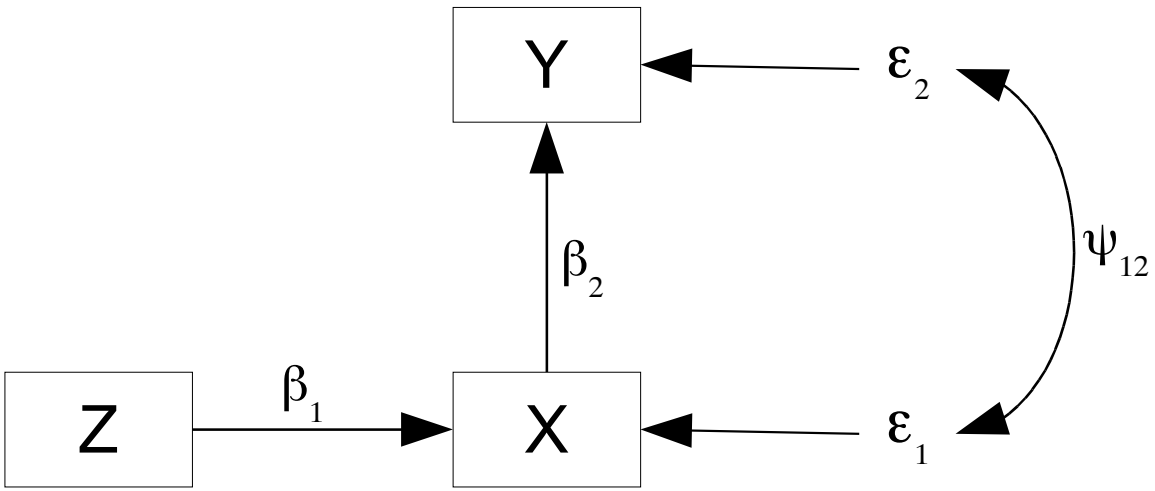
	Z	X	Y
Z	σ_z^2	$\beta_1\sigma_z^2$	$\beta_1\beta_2\sigma_z^2$
X	$\cdot$	$\beta_1^2\sigma_z^2 + \sigma_1^2$	$\beta_2(\beta_1^2\sigma_z^2 + \sigma_1^2) + c$
Y	$\cdot$	$\cdot$	$\beta_1^2\beta_2^2\sigma_z^2 + \beta_2^2\sigma_1^2 + 2\beta_2 c + \sigma_2^2$

- Nine moment structure equations in 9 unknown parameters.
- $\beta_2 = \frac{\sigma_{13}}{\sigma_{12}}$.
- All the other parameters are identifiable too.
- But of course there is measurement error.

The model needs improvement

X is income, Y is credit card debt, Z is median home price.

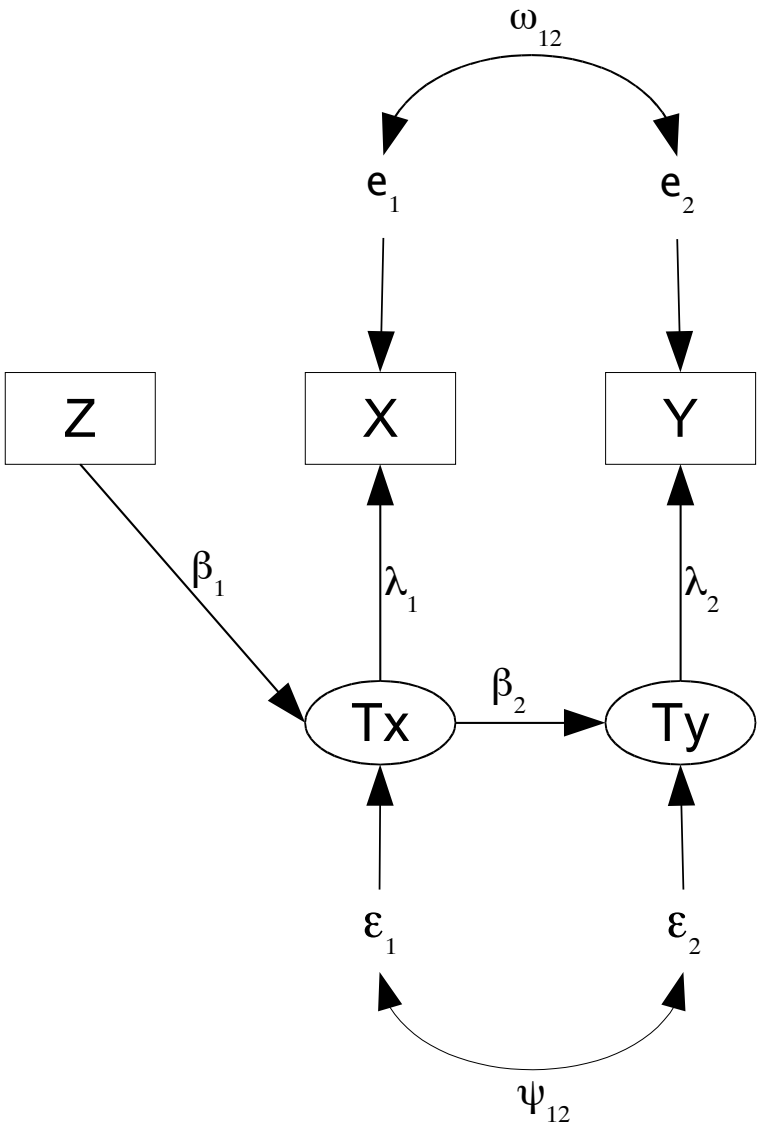
Same picture:



- X = Income is measured with error.
- So is Y = Debt.
- There are still unmeasured variables that impact them both.

An improved Model

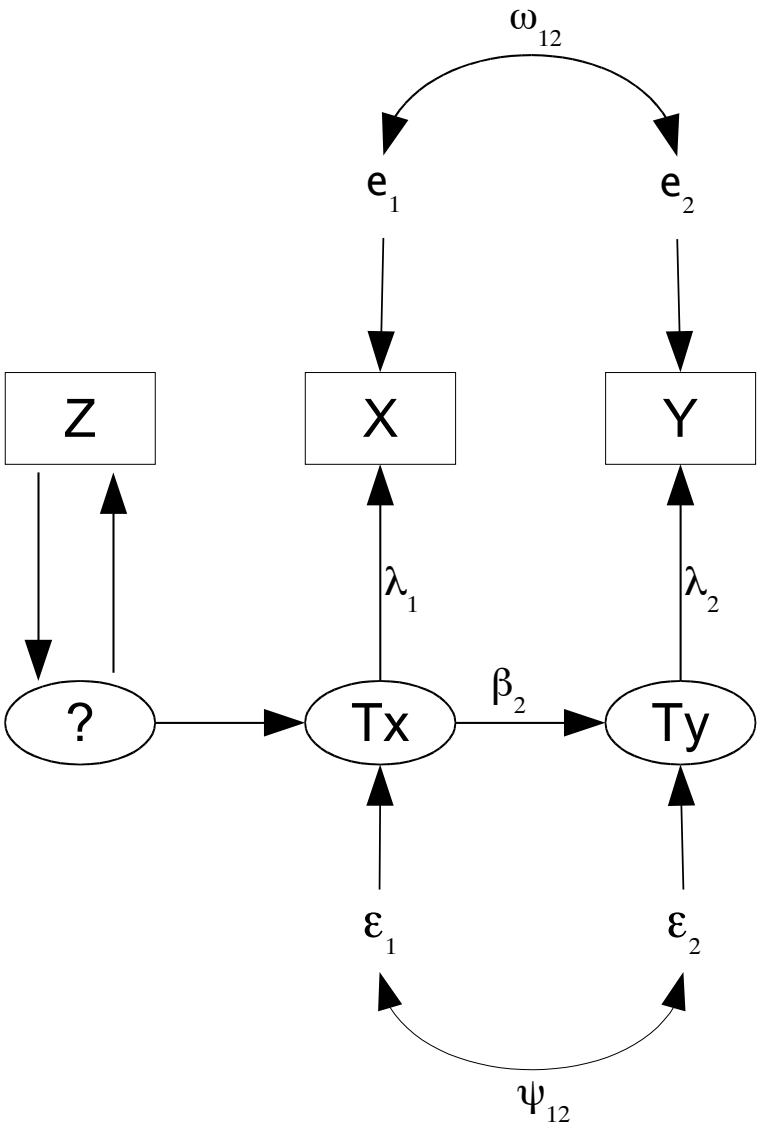
X is income, Y is credit card debt, Z is median home price.



- Common omitted variables are affecting true X and true Y.
- Common omitted variables are affecting measurement of X and measurement of Y.
- Factor loadings are realistic: Positive but not $= 1$.
- Six covariance structure equations in 11 unknowns.
- And it's still not realistic enough.
- Housing prices are only estimated.

Easier to defend, but impossible to estimate

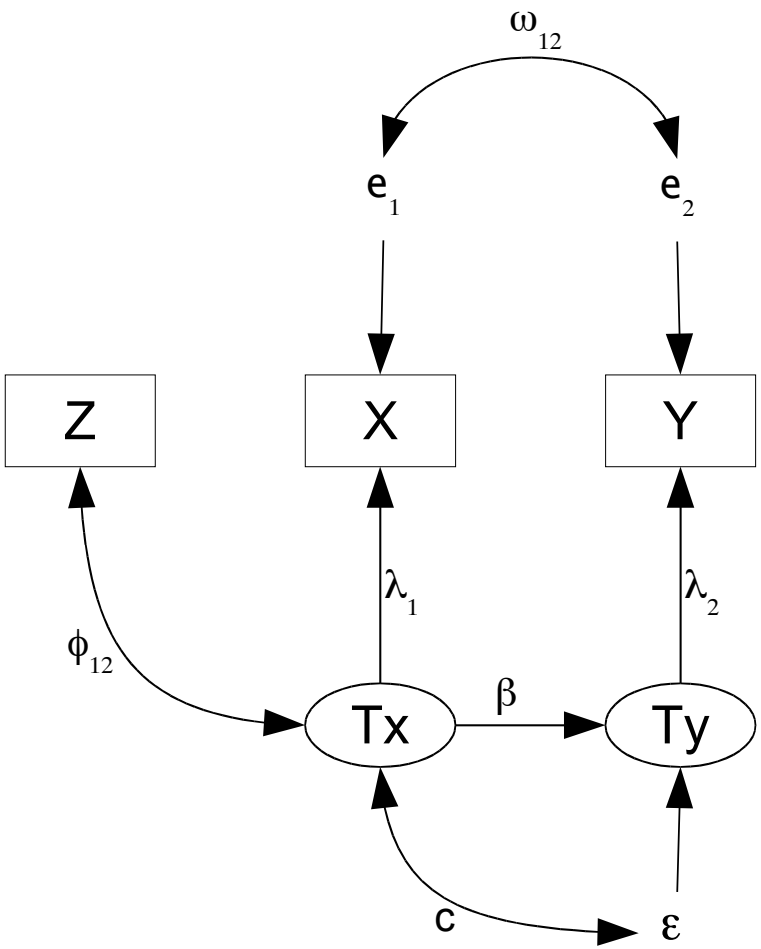
X is income, Y is credit card debt, Z is median home price.



Fortunately the instrumental variable only has to be *correlated* with the explanatory variable.

Here's the Model

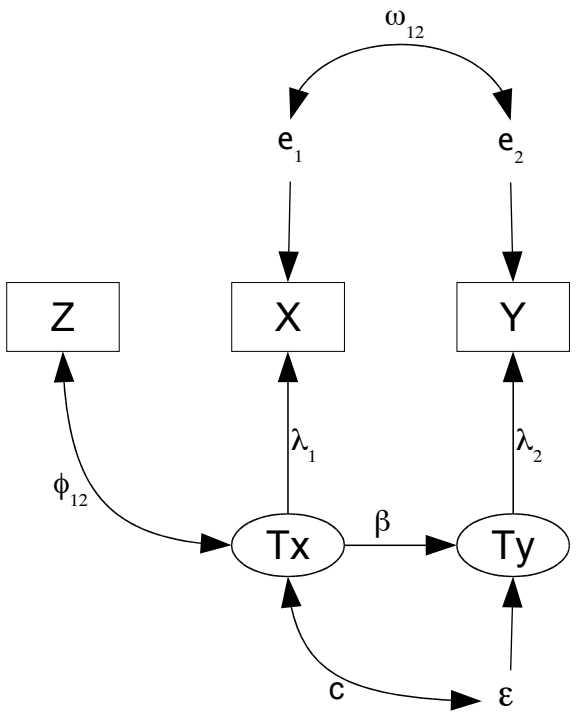
X is reported income, Y is reported credit card debt, Z is estimated median resale home price.



- Fairly realistic.
- Still six covariance structure equations in 11 unknowns (poison).
- Explanatory variable correlated with the error term (poison).
- Correlated measurement errors (poison).
- But we have an instrumental variable.
- Calculate the covariance matrix.

Show part of the calculation

Z is estimated median resale home price, Y is reported credit card debt

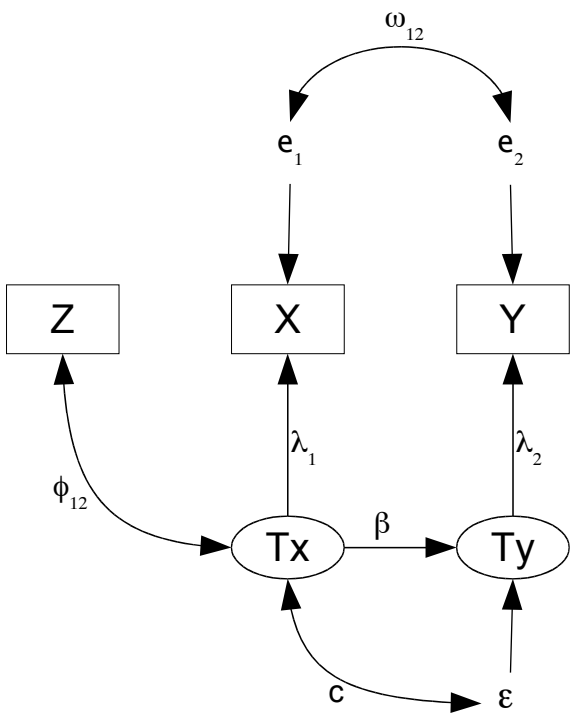


$$
\begin{aligned}
Cov(Z, Y) &= Cov(Z, \lambda_2 T_y + e_2) \\
&= Cov(Z, \lambda_2(\beta T_x + \epsilon) + e_2) \\
&= Cov(Z, \lambda_2 \beta T_x + \lambda_2 \epsilon + e_2) \\
&= \lambda_2 \beta Cov(Z, T_x) + \lambda_2 Cov(Z, \epsilon) + Cov(Z, e_2) \\
&= \lambda_2 \beta \phi_{12} + 0 + 0
\end{aligned}
$$

Covariance matrix of the observable data

Z is estimated median resale home price, X is reported income, Y is reported credit card debt

$$cov\begin{pmatrix} Z \\ X \\ Y \end{pmatrix} = \begin{pmatrix} \phi_{11} & \lambda_1\phi_{12} & \beta\lambda_2\phi_{12} \\ \cdot & \lambda_1^2\phi_{22}+\omega_{11} & \beta\lambda_1\lambda_2\phi_{22}+c\lambda_1\lambda_2+\omega_{12} \\ \cdot & \cdot & \beta^2\lambda_2^2\phi_{22}+2\,\beta c\lambda_2^2+\lambda_2^2\psi+\omega_{22} \end{pmatrix}$$

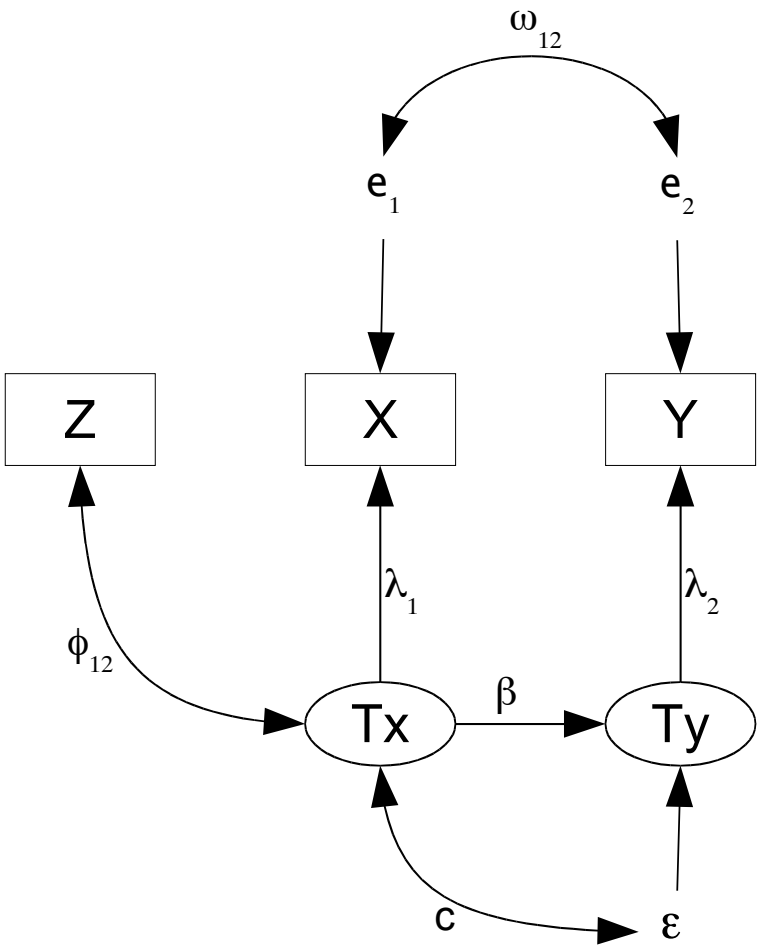


- β is not identifiable.
- But $\phi_{12} > 0$ and $\lambda_2 > 0$.
- So the sign of β is identifiable from σ_{13}.
- $H_0 : \beta = 0$ is testable.
- It's possible to answer the basic question of the study.

It's a miracle

- Instrumental variables can help with measurement error and omitted variables at the same time.
- If there is measurement error, regression coefficients of interest are not identifiable and cannot be estimated consistently, but their signs can.
- Often, that's all you really want to know.
- Matrix version is available.
- The usual rule in Econometrics is (at least) one instrumental variable for each explanatory variable.

Independence of the instrumental variable and error terms is critical.



- Instrumental variables need to come from another world.
- 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 at 40.
  - Z is birth mother's IQ score.

It's a partial solution

- Good instrumental variables are not easy to find.
- They will not be in a data set casually collected for other purposes.
- Advance planning is needed.
- The ultimate instrumental variable is randomly assigned.

Copyright Information