Econ 2120: Section 5

Panel Data: Omitted Variables Bias

Ashesh Rambachan

Fall 2018

Outline

Outline

Twins Study

Data: Random sample from population of families with twins.

$D_i = (Y_{i1}, Y_{i2}, Z_{i1}, Z_{i2})$ for $i = 1, \ldots, n$.

Y_{it} is earnings of twin t and Z_{it} is education of twin t.

Concern: There is an unobserved variable A_i (family background) that is correlated with both education and earnings.

The coefficient on education in a regression of earnings on education will be biased! What can we do?

Twins Study

Assume that (D_i, A_i) are i.i.d. and we are interested in

$$E[Y_{it}|Z_{it}, A_i]$$

A_i is unobserved. So what can we do?

Idea: Exploit within-family variation in Y_{it}, Z_{it} to estimate the partial effect of interest.

Outline

Latent Variable Model

Consider the regression function for Y_{it} given Z_{i1}, Z_{i2}, A_i. We **assume** that

$$E[Y_{it}|Z_{i1}, Z_{i2}, A_i] = g_t(Z_{it}, A_i).$$

That is, the conditional expectation of Y_{it} only depends on Z_{it}, A_i.

This assumption is known as **strict exogeneity**.

- It is an "exclusion restriction."
- Allows for correlation between the latent variable A_i and the covariates, Z_{i1}, Z_{i2}.

We additionally place a **functional form assumption** on the conditional expectation function.

$$E[Y_{it}|Z_{i1}, Z_{i2}, A_i] = \gamma_{1t} + \gamma_{2t} Z_{it} + \gamma_{3t} A_i.$$

- Note: Typically to assume that $\gamma_{3t} = 1$ because A_i is unobserved. But, we will not directly assume that to make things general.

Latent Variable Model

We will additionally assume (for now) that the coefficients do not depend on t:

$$E[Y_{it}|Z_{i1}, Z_{i2}, A_i] = \gamma_1 + \gamma_2 Z_{it} + \gamma_3 A_i.$$

This is the **latent variable model**.

- It is a "structural model" – the coefficients are of direct interest to us.
- But the conditional expectation function has no observable, sample counterpart as A_i is unobserved.
- **Question**: Can we learn about γ_2 using reduced-form relations that are observable (e.g. the best linear predictor of Y_it given Z_{i1}, Z_{i2})?

Latent Variable Model: The Linear Predictors

The linear predictors for Y_{it} given $1, Z_{i1}, Z_{i2}$ are

$$E^*[Y_{i1}|1, Z_{i1}, Z_{i2}] = \gamma_1 + \gamma_2 Z_{i1} + \gamma_3 E^*[A_i|1, Z_{i1}, Z_{i2}]$$
$$E^*[Y_{i2}|1, Z_{i1}, Z_{i2}] = \gamma_1 + \gamma_2 Z_{i2} + \gamma_3 E^*[A_i|1, Z_{i1}, Z_{i2}]$$

The best linear predictor for A_i is

$$E^*[A_i|1, Z_{i1}, Z_{i2}] = \lambda_0 + \lambda_1 Z_{i1} + \lambda_2 Z_{i2}.$$

What can we learn about γ_2 from these reduced form relations?

Outline

Strict Exogeneity

Our set-up:

$$Y_{it} = \gamma_1 + \gamma_2 Z_{it} + \gamma_3 A_i + u_{it}$$

and strict exogeneity implies that

$$E[u_{it}|Z_{i1}, Z_{i2}, A_i] = 0$$

for $t = 1, 2$. This implies that

$$E[z_{it} u_{is}] = 0 \quad s, t = 1, 2.$$

General set-up:

$$Y_{it}. = X_{it}'\beta + A_i + u_{it},$$

where $i = 1, \ldots, n$, $t = 1, \ldots, T$, x_{it} is a $K \times 1$ vector of covariates, and

$$E[u_{it}|X_{i1}, \ldots, X_{iT}, A_i] = 0$$

for all $t = 1, \ldots, T$. This implies that

$$E[u_{it} X_{is}] = 0 \quad s, t = 1, \ldots, T.$$

Strict Exogeneity: Examples

When might be a reasonable assumption?

> You have to evaluate this on a case-by-case basis.

The panel data machinery we will build all rely on the assumption of strict exogeneity. Need to think carefully about when it is reasonable.

Let's look at some examples.

> Based on Wooldridge (2002), Ch. 11.

Strict Exogeneity: Example – Job Training

Typical model for estimating the effects of job training on subsequent wages:

$$log(\text{wage}_{it}) = \theta_t + Z_{it}'\gamma + \beta \text{prog}_{it} + c_i + u_{it},$$

i indexes individuals, t indexes time, Z_{it} is some vector of observable characteristics.

Common approach: $T = 2$.

$T = 0$: No one has participated in the program so $\text{prog}_{i0} = 0$ for all i.

$T = 1$: Some individuals are chosen to participate or choose to participate.

c_i: Worried that individuals enter the program based on unobservables – "self-selection problem."

Strict Exogeneity: Example – Job Training

Is strict exogeneity plausible?

Could u_{it} be correlated with $\text{prog}_{i,t+1}$?

> Concern: Perhaps individuals select into the program at $t + 1$ due to random shocks to past wages? If true, strict exogeneity may not hold.

Strict Exogeneity Example: Lagged Dependent Var.

Consider a simple model for wage dynamics

$$log(\text{wage}_{it}) = \beta_1 log(\text{wage}_{i,t-1}) + c_i + u_{it} \quad t = 1, \ldots, T.$$

We are interested in β_1 – how persistent are wages?

Let $y_{it} = log(\text{wage}_{it})$. A typical assumption is

$$E[u_{it}|y_{i,t-1}, \ldots, y_{i,0}, c_i] = 0.$$

Combined with first equation $\implies$ all dynamics are captured by the first lag.

Strict exogeneity fails under these assumptions.

At $t+1$, the covariate is y_{it}, which is by definition correlated with u_{it}. So, u_{it} is not uncorrelated with future covariates.

In math,

$$E[y_{it} u_{it}] = E[u_{it}^2] > 0.$$

Outline

Outline

Identifying γ_2 due to symmetry

Suppose Z_{i1}, Z_{i2} are symmetric with respect to A_i. That is,

$$\lambda_1 = \lambda_2.$$

Then,

$$E^*[A_i|1, Z_{i1}, Z_{i2}] = \lambda_0 + \lambda_1(Z_{i1} + Z_{i2}).$$

With this assumption, γ_2 is identified.

The linear predictors for Y_{it} are

$$E^*[Y_{i1}|1, Z_{i1}, Z_{i2}] = (\gamma_1 + \gamma_3\lambda_0) + \gamma_2 Z_{i1} + \gamma_3\lambda_1(Z_{i1} + Z_{i2})$$
$$E^*[Y_{i2}|1, Z_{i1}, Z_{i2}] = (\gamma_1 + \gamma_3\lambda_0) + \gamma_2 Z_{i2} + \gamma_3\lambda_1(Z_{i1} + Z_{i2})$$

Identifying γ_2 due to symmetry

In particular, γ_2 is identified from either linear predictor. BUT, the latent variable model contains additional restrictions – the coefficients across the linear predictors are the same.

- We can exploit this in our inference on γ_2 – we perform system estimation using the generalized linear predictor. See Lecture Note 6 for details.

Is symmetry a plausible assumption?

- Seems reasonable in the twins example from Lecture Note 6.
- Probably not in other applications (wage regressions?)

Outline

Set-Up

The latent variable model is:

$$Y_{it} = \gamma Z_{it} + A_i + U_{it}$$

where we'll now assume that

$$E[U_{it}|Z_{i1}, \ldots, Z_{iT}, A_i] = E[U_{it}|A_i] \quad t = 1, \ldots, T$$

Referred to as **strict exogeneity conditional on** A_i in lecture note 6.

- Interpretation: Conditional on A_i, U_{it} is uncorrelated with $Z_{i1}, \ldots, Z_{iT}$.
- Weaker assumption than the strict exogeneity assumption given earlier – no longer saying that A_i is exogenous with respect to U_{it}.
    - Individual's with different fixed unobservables may have a different types of "shocks" over time.

Set-Up

So, we have that

$$E[Y_{it}|Z_i, A_i] = \gamma Z_{it}. + A_i + E[U_{it}|A_i].$$

We assume that

$$E[U_{it}|Z_i, A_i] = \phi_{1t} + (1 + \phi_{2t})A_i$$

and let

$$E^*[A_i|1, Z_i] = \lambda_0 + \lambda_1 Z_{i1} + \ldots + \lambda_T Z_{iT}.$$

Then, we get that

$$\begin{aligned} E[Y_{it}|Z_i, A_i] &= \gamma Z_{it} + \phi_{1t} + (1 + \phi_{2t})A_i \\ E^*[Y_{it}|1, Z_i] &= \gamma Z_{it} + \delta_{1t} + \delta_{2t}(\lambda_1 Z_{i1} + \ldots + \lambda_T Z_{iT}), \end{aligned}$$

where

$$\delta_{1t} = \phi_{1t} + (1 + \phi_{2t})\lambda_0, \quad \delta_{2t} = (1 + \phi_{2t})$$

for $t = 1, \ldots, T$.

Set-Up

$$E[U_{it}|Z_{i1},\ldots,Z_{iT},A_i] = E[U_{it}|A_i] \quad t=1,\ldots,T$$

Go back to farm example: Y_{it} = output, Z_{it} = labor input, A_i = unobserved soil quality, U_{it} = weather conditions.

> **Assumption**: Something like – "Conditional on soil quality, there is no selection of labor based on weather in any period" but "soil quality may be correlated with climate."

If you go to Wooldridge (2002), this will be set-up as

$$E[u_{it}|Z_i,A_i] = 0,$$

which is a stronger assumption.

First Differences

Assume that $\phi_{1t} = \phi_1$, $\phi_{2t} = \phi_2$ are constant over time.

> A_i does not have a time-varying effect on Y_{it} and the constant is not time-varying.

Then,

$$E[Y_{it}|Z_i, A_i] = \gamma Z_{it} + \phi_1 + \phi_2 A_i.$$

And so, we can first-difference to eliminate A_i. We have

$$E[Y_{it} - Y_{i,t-1}|Z_i, A_i] = \gamma(Z_{it} - Z_{i,t-1}).$$

By iterated expectations,

$$E[Y_{it} - Y_{i,t-1}|Z_i] = \gamma(Z_{it} - Z_{i,t-1})$$

We can do inference using the generalized linear predictor.

Outline

Within-Group Estimator/Fixed Effects Estimator

Assume that $\phi_{1t} = \phi_1$, $\phi_{2t} = \phi_2$ are constant over time.

Define

$$\bar{Z}_i = \frac{1}{T}\sum_{t=1}^{T} Z_{it}, \quad \bar{Y}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}.$$

We can show that

$$E[Y_{it} - \bar{Y}_i | Z_i, A_i] = \gamma(Z_{it} - \bar{Z}_i)$$

and by iterated expectations,

$$E[Y_{it} - \bar{Y}_i | Z_i] = \gamma(Z_{it} - \bar{Z}_i).$$

We can do inference with the generalized linear predictor.

Outline

Time-Varying Coefficients

Leave the details for you – great exercise in showing that parameters are identified.

Idea: Allow ϕ_{1t}, ϕ_{2t} to vary across periods.

> Form best linear predictors of Y_{it} given a constant and Z_i for each period t and write the best linear predictor coefficients in terms of the structural parameters.
>
> Is the number of reduced form parameters $\geq$ number of structural parameters? Can you solve for the structural parameter of interest from the reduced form parameters?

Quite a general strategy.

Outline

Teacher Effects

Suppose we data on n teachers over T years. The data are $\{D_i\}_{i=1}^n$ with $D_i = (Y_i, Z_i)$. Let

$$Y_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{iT} \end{pmatrix}, \quad Z_i = \begin{pmatrix} Z'_{i1} \\ \vdots \\ Z'_{iT} \end{pmatrix},$$

where Y_{it} is a scalar (average test score of class) and Z_{it} is a $J \times 1$ vector of class characteristics (previous test scores, family income, etc.)

Problem: Predict $Y_{i,T+1}$ for a given teacher i if you are told $Z_{i,T+1}$.

Teacher Effects

Start with generalized linear predictor $E^*_\Phi[Y_i|R_i] = R_i\beta$, where R_i is $T \times K$ matrix constructed from Z_i.

Residualize and form $U_i = Y_i - R_i\beta$ so,

$$Y_{it} = R'_{it}\beta + U_{it}$$

If we observe $R_{i,t+1}$, predicting $Y_{i,t+1}$ effectively boils down to predicting $U_{i,t+1}$ – this is the "effect that is not predictable from class characteristics."

> This prediction of $U_{i,T+1}$ is the **teacher effect** (i.e. predictable variation in the component of test scores that is unpredictable from classroom characteristics).

Teacher Effects

Assume $U_{it} = 0$ (R'_{it} includes a constant) for $t = 1, \ldots, T+1$. Consider

$$E^*[U_{i,T+1} | U_{i1}, \ldots, U_{iT}] = U_i' \delta, \quad \delta = E[U_i U_i']^{-1} E[U_i U_{i,T+1}].$$

What can we do:

We can construct $\hat{\beta}$ and so, we can form $\hat{U}_i = Y_i - R_i \hat{\beta}$.

We can estimate $E[U_i U_i']$ with $n^{-1} \sum_{i=1}^{n} \hat{U}_i \hat{U}_i'$.

What about $E[U_i U_{i,T+1}]$? We need to model the covariance structure of U_{it}.

Teacher Effects

Suppose that

$$U_{it} = V_i + \epsilon_{it} \quad t = 1, \ldots, T+1$$

where $E[V_i] = E[\epsilon_{it}] = 0$ and

$$E[V_i^2] = \sigma_v^2, \quad E[\epsilon_{it}^2] = \sigma_\epsilon^2, \quad E[V_i \epsilon_{it}] = 0, \quad E[\epsilon_{it}\epsilon_{is}] = 0$$

where the last is for $s \neq t$.

V_i is the **teacher effect** and that's what we are after.

Teacher Effects

With this, we have that

$$E[U_i U_i'] = \sigma_v^2 ll' + \sigma_\epsilon^2 I_T, \quad E[U_i U_{i,T+1}] = \sigma_v^2 l$$

where l is a $T \times 1$ vector of ones.

You can show that (just algebra)

$$(\sigma_v^2 ll' + \sigma_\epsilon^2 I_T)^{-1} = \frac{1}{\sigma_\epsilon^2}(I_T - \frac{\sigma_v^2}{\sigma_\epsilon^2 + T\sigma_v^2} ll')$$

Immediately get that

$$\delta = \frac{\sigma_v^2}{T\sigma_v^2 + \sigma_\epsilon^2} l$$

Teacher Effects

So, we immediately get that

$$E^*[U_{i,T+1}|U_i] = E^*[V_i|U_i] = U_i'\delta = \frac{T\sigma_v^2}{T\sigma_v^2 + \sigma_\epsilon^2}\bar{U}_i,$$

where $\bar{U}_i = T^{-1}\sum_{t=1}^{T} U_{it}$. Re-written, this becomes

$$E^*[U_{i,T+1}|U_i] = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_\epsilon^2/T}\bar{U}_i.$$

Again, we have a shrinkage-type estimator.

The lecture note shows how to extend this model to allow for serial correlation in the ϵ_{it} term.

Estimating Teacher Effects

We can use the generalized linear predictor to impose restrictions on the covariance structure of the reduced form errors, U_i.

Let Q_i be a column vector that is formed by stacking the unique elements of $U_i U_i'$ – this is a symmetric matrix so we can just focus on the lower triangular portion.

$$Q_i = \begin{pmatrix} U_{i1}^2 \\ U_{i2}U_{i1} \\ \vdots \\ U_{iT}U_{i1} \\ U_{i2}^2 \\ U_{i3}U_{i2} \\ \vdots \\ U_{iT}U_{i2} \\ \vdots \\ U_{iT}^2 \end{pmatrix}$$

Estimating Teacher Effects

The restrictions we placed

$$U_{it} = V_i + \epsilon_{it} \quad t = 1, \ldots, T+1$$

$$E[V_i^2] = \sigma_v^2, \quad E[\epsilon_{it}^2] = \sigma_\epsilon^2, \quad E[V_i \epsilon_{it}] = 0, \quad E[\epsilon_{it}\epsilon_{is}] = 0$$

implies a particular structure for $E[Q_i]$.

Estimating Teacher Effects

In particular, we have that

$$E[Q_i] = \begin{pmatrix} \sigma_v^2 + \sigma_\epsilon^2 \\ \sigma_v^2 \\ \vdots \\ \sigma_v^2 \\ \sigma_v^2 + \sigma_\epsilon^2 \\ \sigma_v^2 \\ \vdots \\ \sigma_v^2 \\ \vdots \\ \sigma_v^2 + \sigma_\epsilon^2 . \end{pmatrix}$$

Estimating Teacher Effects

We can write this as

$$E[Q_i] = A \begin{pmatrix} \sigma_v^2 \\ \sigma_\epsilon^2, \end{pmatrix}$$

where

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ \vdots & \\ 0 & 1 \\ 1 & 1 \\ 0 & 1 \\ \vdots & \\ 0 & 1 \\ \vdots & \\ 1 & 1 \end{pmatrix}$$

Estimating Teacher Effects

We can then plug this into our generalized linear predictor framework. If we are not willing to assume that the population satisfies these assumptions about the covariate structure (reasonable since we are very over-identified), we use a weight matrix:

$$E^*_\Omega[Q_i|A] = A\sigma^2, \quad \sigma^2 = \begin{pmatrix} \sigma_v^2 \\ \sigma_\epsilon^2 \end{pmatrix}.$$

Then, $\left(\sigma^2 = (A'\Omega A)^{-1} A'\Omega E[Q_i]\right)$

Estimating Teacher Effects

We can implement this by constructing the reduced-form errors

$$\hat{U}_i = Y_i - R_i\hat{\beta}$$

and forming

$$\hat{\sigma}^2 = (A'\Omega A)^{-1}A'\Omega\Big(n^{-1}\sum_{i=1}^{n}\hat{Q}_i\Big).$$

Then, our estimate of the teacher effect is

$$\hat{E}^*[V_i|U_i] = \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2 + \hat{\sigma}_\epsilon^2/T}\bar{U}_i.$$