Causal Inference: Panel Data and Fixed Effects Models

Ken Stiller

28th May 2024

Last weeks

- Non Compliance: Does information increase turnout?
- What happens if we cannot randomize?
- Selection on observables and the prospects for causal inference
- Matching
- Regression

Let's begin with a recap!

Extrapolations

- The “else equal” principle is often satisfied only through extrapolations beyond the range of the available data.
- Such extrapolations are in turn based on assumptions, which are typically untestable and ‘invisible’ within the regression framework.
- Matching, thus makes the stage of making units similar with regard to covariates more transparent.
- Imagine candidate 7 also went to a public school; thus, comparing Candidate 1 and 7 would provide an estimate of ATE.
- Again this is also the case with regression: extrapolations require attaching greater weight to most similar units. *Regressions?*

Dimensionality

- In the original study, there are more than 400 observations available.
- But: many more covariates are taken into account
- As the number of covariates used to "match" units increases, it becomes exponentially more difficult to find perfect matches.
- Exact matching fails in finite samples if the dimensionality of X is large: not enough information. Far too demanding for the vast majority of research questions and data available.
- With more than one continuous variable, it is also sub-optimal (Abadie & Imbens, 2006).

Matching in Multidimensional Space

“Else being similar”

With many X’s and typically also with continuous X’s, estimation of ATT is based on the detection of the closest possible control unit to match every treated unit:
$\hat{\tau} = \frac{1}{N_1} \sum_{D_i=1} (Y_i - Y_{j(i)})$
where $Y_{j(i)}$ is the outcome of an untreated observation such that $X_{j(i)}$ is the closest value to X_i among the untreated observations.

Problem

- How to decide which control is closest?

Defining Closensess

Think of it as a distance metric. Let $X_i = (X_{i1}, X_{i2}, \dots, X_{ik})$ and $X_j = (X_{j1}, X_{j2}, \dots, X_{jk})$ be covariate vectors for i and j. We want to find ways to link rows according to their similarities in their values in each of these vectors. This is done with distance metrics.

The Propensity Score

Another way to reduce dimensionality: `match` on the Propensity Score

Definition

The probabity to receive treatment (also known as the selection probability) conditional on the set of pre-treatment covariates: $p(X) = P(D = 1|X)$

Identification Assumptions

1. $(Y_1, Y_0) \perp D|X$ (Selection on Observables)
2. $0 < Pr(D = 1|X) < 1$ (common support)

Propensity Score Properties

Balancing: Balancing of pre-treatment variables given the propensity score: $D \perp X|p(X)$

Unconfoundedness: If $Y_1, Y_0 \perp D|X$, then $Y_1, Y_0 \perp D|p(X)$.

How to Estimate the Propensity Score

- Regress D_i on the set of $X's$ using a logit or probit function to estimate the score.
- Take the predicted values of D_i. These predicted values represent the probability of being assigned to treatment, given X(the Propensity Score).
- Choose closest control on $p(X_i)$ (Call this the Nearest Neighbor (NN))
- Test for balance: If not satisfactory: redo by changing matching criteria.
- Repeat until balance is satisfactory:
  Estimate PrScore → Check Balance → Re-Estimate → Check Balance

Checklist

- Always establish balance before you even look at the Y
- Look for balance not only at the characteristics included in matching but higher polynomials and on other covariates. Balance should extend beyond X, if X is correctly specified.
- Do not simply think of matching as an alternative or final resort when design-based identification is not provided. Conversely, use it when there is some design that allows you to make the conditional-on-observables assumption more credible.

Wrap-up–Matching

Prince Charles

Male
Born in 1948
Raised in the UK
Married Twice
Lives in a castle
Wealthy and Famous

Ozzy Osbourne

Male
Born in 1948
Raised in the UK
Married Twice
Lives in a castle
Wealthy and Famous

Wrap-up–Matching

- Useful method to create exchangeable units
- (Too) many alternative algorithms
- (Too) many options
- Sensitivity
- Do we really take care of the unobservability problem (inherent in regression models)?
- **Q for discussion**: Can we draw causal inferences by selecting on observables?

Recap

Strategies for estimating effects of treatments so far:

- Randomize treatment and take the DIGM
- Identify and control for confounding variables such that the CIA (Conditional Independence Assumption) holds

Today: Use observations at more than one point in time

Types of samples

- Cross section: observe many units at one point in time. So far!
- Time series: observe one unit at many points in time
- Repeated cross sections: observe many units at many points in time, and the units are (potentially) different over time
- Panel data: observe many units at many points in time, and the units are the same over time

Repeated Cross Sections

- At one point in time ($t = 1$) we take a random sample of n_1 individuals ($i = 1 ... n_1$)
- At another later point in time ($t = 2$) we take another random sample of potentially different n_2 individuals ($i = 1 ... n_2$)
- Every period we repeat the process of obtaining a cross section
- This data is not panel data because the individuals are not the same over time (or they are and we still do not treat them as such)
- Why would we want to have cross sections at different points in time? Larger N, study whether the slope estimate changes over time as a result of policy change (or shock in time)
- Often the case in surveys

Panel Data

- At one point in time ($t = 1$) we take a random sample of n individuals/units ($i = 1...n$)
- At another later point in time ($t = 2$) we sample again the same n individuals/units ($i = 1...n$)
- Every period we repeat the process of **sampling the same individuals/units**: the same individuals/units are measured at different points in time
- If every individual appearing in $t = 1$ also appears in $t = 2, 3...$: **balanced panel**

Country ID	Year	Var1	Var 2
Afghanistan	2018	x1	y1
Afghanistan	2019	x1	y1
Afghanistan	2020	x1	y1
Croatia	2018	x2	y2
Croatia	2019	x2	y2
Croatia	2020	x2	y2
United Kingdom	2018	x3	y3
United Kingdom	2019	x3	y3
United Kingdom	2020	x3	y3

Empirical example: Number of Women MPs and Domestic Abuse Legislation

Let's imagine we are interested in the question whether the number of Women MPs change the nature of legislation, particularly whether it changes Domestic Abuse legislation passed. Suppose we observe 20 countries over a 20 year period.

$$\text{Domestic Abuse Legislation}_{i,t} = \alpha + \beta \text{Number of Women MPs}_{i,t} + u_{i,t}$$

- What are the potential omitted variables?
- What about time?
- Has the number of women MPs changed through time? Has domestic abuse legislation changed through time?

Empirical example: Number of Women MPs and Domestic Abuse Legislation

Let's imagine we are interested in the question whether the number of Women MPs change the nature of legislation, particularly whether it changes Domestic Abuse legislation passed. Suppose we observe 20 countries over a 20 year period.

$$\text{Domestic Abuse Legislation}_{i,t} = \alpha + \beta \text{Number of Women MPs}_{i,t} + u_{i,t}$$

- What are the potential omitted variables?
- What about time?
- Has the number of women MPs changed through time? Has domestic abuse legislation changed through time?

Empirical example: Number of Women MPs and Domestic Abuse Legislation

Let's imagine we are interested in the question whether the number of Women MPs change the nature of legislation, particularly whether it changes Domestic Abuse legislation passed. Suppose we observe 20 countries over a 20 year period.

$$\text{Domestic Abuse Legislation}_{i,t} = \alpha + \beta \text{Number of Women MPs}_{i,t} + u_{i,t}$$

- What are the potential omitted variables?
- What about time?
- Has the number of women MPs changed through time? Has domestic abuse legislation changed through time?

Empirical example: Number of Women MPs and Domestic Abuse Legislation

Let's imagine we are interested in the question whether the number of Women MPs change the nature of legislation, particularly whether it changes Domestic Abuse legislation passed? We observe 20 countries over a 20 year period (2000-2020).

$$\text{Domestic Abuse Legislation}_{i,t} = \alpha_1 + \alpha_2 \text{decade2010}_t +$$

$$\beta \text{Number of Women MPs}_{i,t} + u_{i,t} (1)$$

- What is the interpretation of $\hat{\alpha_1}$?
- What is the interpretation of $\hat{\alpha_2}$?
- Does the inclusion of decade fixed effects help with selection bias concerns?

Empirical example: estimation

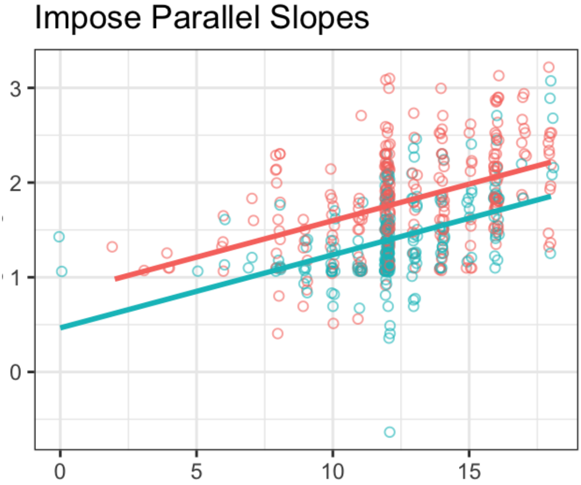


Figure: Different intercepts for two decades, same slope

Empirical example: Number of Women MPs and Domestic Abuse Legislation

We could also add individual year fixed effects for every year in our sample:

$$\text{Domestic Abuse Legislation}_{i,t} = \alpha_1 + \alpha_2 y2002_t + \alpha_3 y2003_t + \alpha_4 y2004_t + \alpha_5 y2005_t$$

$$+ \alpha_6 y2006_t + (...) + \alpha_{20} y2020_t + \beta \text{Number of Women MPs}_{i,t} + u_{i,t} (2)$$

This is equivalent to writing:

$$\text{Domestic Abuse Legislation}_{i,t} = \alpha_1 + \delta_t + \beta \text{Number of Women MPs}_{i,t} + u_{i,t} \qquad (3)$$

- What do the δs estimate?
- Does the inclusion of year fixed effects help with selection bias concerns?
- Control for any time varying unobservable factors. The variation used to estimate the effect is the number of women MPs in a given year but across countries/units.

Panel Data: First Difference Estimator

- Panel data can help control for some omitted variable bias
- We sample the same individuals over two periods and assume that:
  - the intercepts are different for the two periods
  - the slopes (causal effects of interest) are the same over the two periods
- We can write a model for the first period:

$$y_{i,1} = \beta_{0,1} + \beta_1 x_{i11} + \beta_2 x_{i12} + (...) + \beta_k x_{i1k} + \epsilon_{i,1}$$

- We can write a model for the second period:

$$y_{i,2} = \beta_{0,2} + \beta_1 x_{i21} + \beta_2 x_{i22} + (...) + \beta_k x_{i2k} + \epsilon_{i,2}$$

Panel Data: First Difference Estimator

Assume that the error term in period t, consists of two parts:

$$\epsilon_{i,t} = a_i + v_{i,t}$$

where a_i: part that is constant over time (fixed effect)
$v_{i,t}$: part that changes every period (idiosyncratic error)

Examples:
For individuals:

- a_i : genetic material, ability, character,... time invariant
- $v_{i,t}$: luck, experience, performance...

For countries:

- a_i : geography (coastal access), colonial history, ... time invariant
- $v_{i,t}$: economic or political shocks...

Panel Data: First Difference Estimator

The first period:

$$y_{i,1} = \beta_{0,1} + \beta_1 x_{i11} + \beta_2 x_{i12} + (...) + \beta_k x_{i1k} + \underbrace{a_i + v_{i1}}_{\epsilon_{i,1}}$$

The second period:

$$y_{i,2} = \beta_{0,2} + \beta_1 x_{i21} + \beta_2 x_{i22} + (...) + \beta_k x_{i2k} + \underbrace{a_i + v_{i2}}_{\epsilon_{i,2}}$$

Subtracting one period from the other:

$$\Delta y_i = \delta + \beta_1 \Delta x_{i1} + \beta_2 \Delta x_{i2} + (...) + \beta_k \Delta x_{ik} + \Delta v_i$$

The fixed effect a_i has disappeared. Anything observable or not observable that was a confounder and is time invariant at the unit level is not a concern anymore.

Fixed Effects Estimator for Panel Data

Assume that we have a panel dataset over twenty years.
→ Fixed Effects Estimator:

- Regression with Entity/Unit Fixed Effects: Method for controlling for omitted variables in panel data when omitted variables vary across individuals but not over time.
- This is similar to a first-differences estimator (equivalent if $t=2$).
- Regression with Time Fixed Effects: Method for controlling for omitted variables in panel data when omitted variables vary across time but not across individuals.
- Examples?

Regression with Time Fixed Effects

- Time fixed effects can control for variables that are constant across entities but evolve over time.
- For example, in a panel of UK electoral districts/regions this may be variables like UK annual GDP, financial crisis, Brexit, etc...
- If an omitted variable X_t is unobserved, its influence can be exogenised because it varies over time but not across units.
- This means specifying a model that has a different intercept λ_t for each time period (T-1 binary indicators). Then the terms $\lambda_1 ... \lambda_T$ are know as time fixed effects.

$$y_{i,t} = \beta_0 + \beta_1 X_{i,t} + \delta_2 Q2_t + (...) + \delta_t QT_t + u_{i,t}$$

- Where $\delta_2, .., \delta_T$ are unknown coefficients and where $Q2_t = 1$ if $t = 2$, and so forth.

Regression with Entity Fixed Effects

- Entity fixed effects can control for variables that are constant across time but vary across entities.
- That is, *unit-specific, time-invariant* heterogeneity is not a problem anymore.
- This means specifying a model with entity specific intercepts $\alpha_1, \alpha_2, ...\alpha_n$

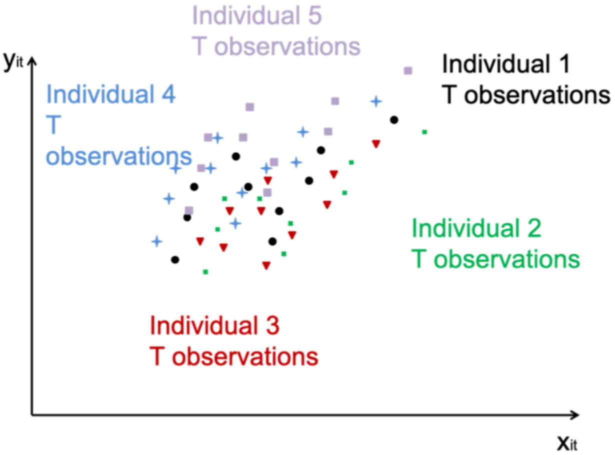
$$y_{i,t} = \beta_0 + \beta_1 X_{i,t} + \gamma_2 D2_i + (...) + \gamma_n DN_i + u_{i,t}$$

- where $D2_i = 1$ if $i = 2$, $D3_i = 1$ if $i = 3$, etc..
- Equivalent to writing the below where $\alpha_1 ... \alpha_n$ are entity specific intercepts.

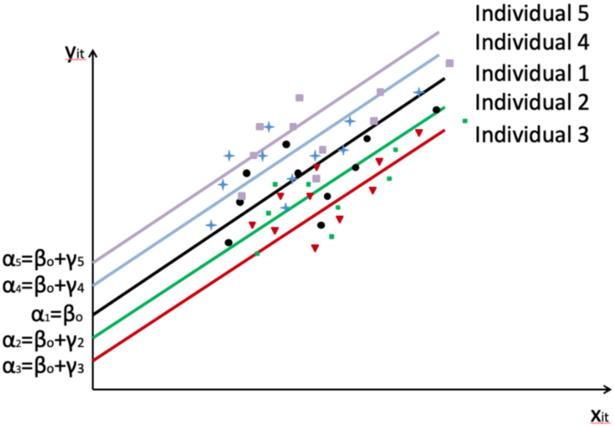
$$y_{i,t} = \beta_1 X_{i,t} + \alpha_i + u_{i,t}$$

Regression with Entity Fixed Effects: Example of Data

We have information for all individuals over time:



Regression with Entity Fixed Effects: Example of Estimation



Regression with Entity Fixed Effects and Time Fixed Effects

Imagine we go back to our original example of examining (a similar) question of the effect of a women policymakers on the representation of women's issues across a panel of US states and 20 years:

$$\text{Domestic Abuse Funding}_{i,t} = \alpha_i + \delta_t + \beta \text{Congresswoman}_{i,t} + u_{i,t} \quad (4)$$

- What are we controlling for with α_i and δ_t?
- Give examples of selection bias that has been resolved.
- What is the leftover variation in $Congresswoman_{i,t}$ used to estimate β?
- What could be the sources of remaning selection bias?

Regression with Entity Fixed Effects and Time Fixed Effects and Potential Outcomes Framework

We need to satisfy: $P(D_i = 1) \perp\!\!\!\perp Y_0, Y_1$

Experiments

- Randomization ensures unconfoundedness without selection on observables: $P(D_i = 1) \perp\!\!\!\perp Y_0, Y_1$ which translates into:
- $E(Y_0|D = 1) = E(Y_0|D = 0)$

Panel data fixed effects models (observational studies)

- Unconfoundedness can be assumed to hold only after conditioning on a set of pre-treatment variables and time and entity fixed effects, which translates into:
- $E(Y_0|D = 1, X_{i,t}, \alpha_i, \delta_t) = E(Y_0|D = 0, X_{i,t}, \alpha_i, \delta_t)$

Table of Contents

Assumptions for standard errors

What does the standard error mean?

$$se(\hat{\beta}) = \sqrt{\frac{v\hat{a}r(e_i)}{n \times v\hat{a}r(X_i)}}$$

Basic assumptions behind OLS standard errors:

- Variance of regression errors independent of X (homoskedastic)
- Regression errors independent of each other (uncorrelated across observations)

Is second assumption likely to be met in the panel data case?

Problem with repeated observations: Addressing correlations among errors

- We cannot make the assumption anymore that regression errors are independent of each other - likely they are correlated within time or within unit.
- We then make the common assumption in panel data that regression errors are independent except within clusters → **cluster-robust standard errors.**
- Clustered standard errors allow for heteroskedasticity and autocorrelation within an entity, but treat the errors as uncorrelated across entities.
- Typically, the clusters are the panel units (e.g. municipalities, states). It doesn't work with 2 clusters → usually need at least thirty or more clusters.

See `estimatr` or `lfe` packages (in `R`).
In Stata, see `cluster()`.

Random Effects Models

We begin with a similar unobserved effects model as before:

$$y_{i,t} = \beta_0 + \beta_1 X_{i,t,1} + \beta_2 X_{i,t,2} + \alpha_i + u_{i,t}$$

- We explicitly include an intercept so we can make the assumption that the unobserved effect α_i has zero mean.
- Suppose we think that α_i is uncorrelated with each explanatory variable in all time periods.
- $cov(x_{i,t,j}, \alpha_i) = 0$ for $t = 1..T; j = 1...k$.
- In this case, β_j is unbiased and can be consistently estimated by using a single cross section ($E[u|X] = 0$ is satisfied). But eliminating a_i from the specification will result in inefficient estimators.

Diagnostic Tests: F-Test

Are the fixed effects predictive of the outcome?

$$y_{i,t} = \beta_1 X_{i,t} + \alpha_i + u_{i,t}$$

- Specify H_0 and H_1.
- $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = ... = \alpha_n = 0$
- $H_1 : \alpha_1 \neq \alpha_2 \neq \alpha_3 \neq ...\alpha_n \neq 0$ (or at least one)
- Run the restricted model and retrieve SSR: $y_{i,t} = \beta_1 X_{i,t} + u_{i,t}$
- Conduct F-test with a set F critical value. Reject the null hypothesis if $F > F_c$

$$F = \frac{(SSR_R - SSR_{UR})/q}{SSR_{UR}/n - k - 1}$$

Diagnostic tests: Hausman Test

- The Hausman test can be used to differentiate between fixed effects model and random effects model in panel analysis. In this case, Random effects (RE) is preferred under the null hypothesis due to higher efficiency, while under the alternative Fixed effects (FE) is unbiased and at least as consistent and thus preferred.
- You can run a Hausman test (which tests whether the unique errors are correlated with the regressors, the null is they are not). If the p-value indicates statistical significance, then you choose fixed effects (since the unique errors are correlated with the regressors).
- $H_0 : cov(u_{i,t}, \alpha_i) = 0$
- $H_1 : cov(u_{i,t}, \alpha_i) \neq 0$