Panel data

Intro

Prep

```r
library(tidyverse)
```

```
## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr
```

```
## Conflicts with tidy packages -------------------------
```

```
## filter(): dplyr, stats
## lag():    dplyr, stats
```

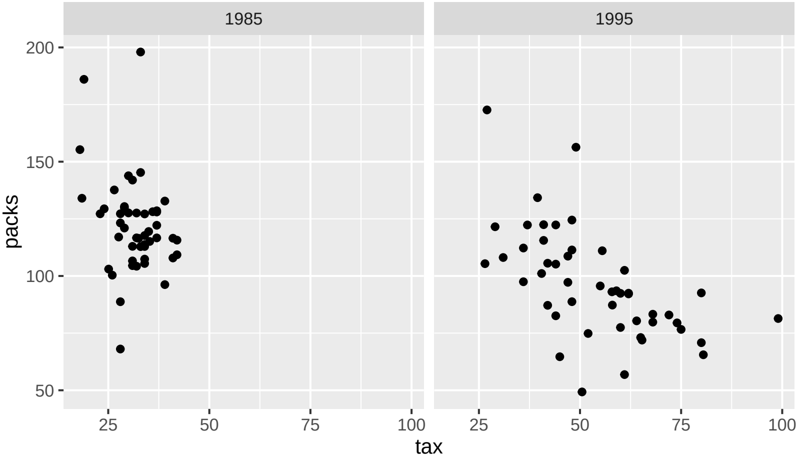
Panel data

A **panel data** set is a data set that has two dimensions:

- cross-sectional
- time

PD: Example

```
data("CigarettesSW",package="AER")
qplot(tax,packs,data=CigarettesSW) + facet_wrap(~year)
```



PD: Example (2)

```
data("PSID7682",package="AER")
head(PSID7682 %>% select(id,year,wage,education),20)
```

```
##    id year wage education
## 1   1 1976  260         9
## 2   1 1977  305         9
## 3   1 1978  402         9
## 4   1 1979  402         9
## 5   1 1980  429         9
## 6   1 1981  480         9
## 7   1 1982  515         9
## 8   2 1976  475        11
## 9   2 1977  500        11
## 10  2 1978  525        11
## 11  2 1979  695        11
## 12  2 1980  810        11
## 13  2 1981  890        11
## 14  2 1982  912        11
```

Repeated measurements

Panel data gives us **repeated measurements** on the same unit.

- multiple **time periods** for a country
- identical twins
- multiple babies to a mother
- multiple outcome variables
- opponents in a tournament

Why use panel data?

1. Panel data sets contain **more information**
   - smaller standard errors than comparable cross-sections
2. Estimate how (conditional) distributions change over time
3. Real innovation:
   - estimate models with heterogeneous individuals
   - Under E.1-E.4, all individuals are the same, conditional on X_i
     - $E[u_i|X_i] = 0$: people are identical, in expectation
   - Repeated measurements allows you to control for the fact that people are different:
     - **unobserved heterogeneity**

Panel data model: sampling

We have a large random sample of size n on

$$(Y_i, X_i) = ((Y_{i1}, \cdots, Y_{iT}), (X_{i1}, \cdots, X_{iT})),$$

for a total number of nT observations.

Panel data model: equation

The **regression equation** for a panel data model looks like:

$$Y_{it} = h(X_{it}, \beta, \alpha_i, \gamma_t, u_{it}),$$

where, in addition to the additional indices Y, X, u we have unit-specific parameters α_i and time-specific parameters γ_t.

Unobserved heterogeneity

The unobserved heterogeneity, or “fixed effect”, α_i allows for individuals to

- structural models: have different production functions / utility functions
- reduced form: have systematically different APEs

Economists are generally unwilling to assume $\alpha_i \perp X_i$. The following examples provide some motivation.

Examples

Intro

(from ECON435)

- To understand why panel data can be useful: examples.
- In each example:

$$y_{it} = \alpha_i + X_{it}\beta + u_{it}$$

  - Repeated measurements for each i across t
  - u_{it} will be an error terms as in the first part of this course
  - α_i represents a unit-specific error term:
    - unobserved (not in X) characteristics of the individual
    - time-invariant
- Questions:
  - What is in α_i?
  - Is it correlated with X_{it}?

Hint:

- Yes, it is correlated with X_{it}.
  - (But: Why?)
- Preview of panel data models:
  - $E(u_{it} + \alpha_i | X_{it}) \neq 0$

Example 1: Birthweight and smoking

- Data on
  - birthweight of newborns (y_{it})
  - smoking behavior of their mother $X_{1,it}$
  - (i, t) refers to the $t-$th newborn of mother i

Linear panel data model:

$$y_{it} = \alpha_i + X_{it}\beta + u_{it}$$

where X_{it} includes

- X_{1it},
- age
- income,
- education

Example 1: Questions

1. What are other factors that could influence a baby's birthweight?
2. Do you believe that those factors do not change over time?
3. Do you believe that those factors are correlated with smoking?

Example 1: Answers

1: genetics, diet, exercise, healthy behavior, diligence in precautions for baby's outcomes

2: genetics: yes, diligence: probably not, healthy behavior: probably pretty persistent

3: genetics: tricky, should be different genes that make you tall and that make you smoke

The *unobserved heterogeneity* can be *specific*, unincluded variables, or *vaguer* terms such as "health awareness", "healthy behavior".

Example 2: Skytrain

- Data on housing prices in two periods, $t = 1, 2$
- In between, a skytrain is built
- D_i is a binary indicator: is skytrain $<$5min walk from house i?

Relationship:

$$y_{it} = \alpha_i + D_i\beta + u_{it} + X_{it}\gamma$$

where X_{it} is a vector of the characteristics of the house:

- number of bedrooms,
- bathrooms, and
- square footage, and the
- year that it was built
- ...

Example 2: Skytrain: Questions

1. What does α_i capture?
2. Why is it correlated with D_i ?

Example 2: Skytrain: Answers

1. Amenities (parks, schools, shopping); location (density, ...)
2. Think about the decision to build the skytrain station. there are several ways in which this decision making process can induce correlation between α_i and D_{it}. If the social planner is trying to develop a new neighhbourhood, they may be looking for a spot with cheap (c.p.) plots. Alternatively, they may be targeting areas with high density, or with a lot of amenities, because it likely increases the usage rate. These are two opposing mechanisms. It is unlikely that they cancel out exactly.

Compare the **incinerator** example in Wooldridge.

Example 3: Texting bans

- Existing literature:

$$P(death|driving + phone) = 4 \times P(death|driving + nophone)$$

- People continue to text. Why?

$$Y_{i,m} = \alpha_i + \delta_m + X_{im}\beta + \omega B_{im} + u_{im}$$

where:

- i is state, m is month
- Y is (log of) traffic fatalities
- X includes
  - population
  - proportion male
  - unemployment
  - gas tax
- B: is a texting ban in place?

Example 3: Texting bans (2)

- What's in α_i?
- Correlated with X?
- Correlated with B?

Example 3: Texting bans

Finding: $\hat{\omega} = -0.0374$.

- Interpret this finding.

Details:

- No effect for "weak bans"
- No effect except for single-occupancy vehicles
- Effect starts when findings are announnced, disappears four months after ban in effect

Example 4: Mafia and public spending

From PS8, AER(2014)

$$Y_{it} = \alpha_i + G_{it}\beta + \gamma_t + u_{it} + X_{it}\beta$$

- i is an Italian province, t is a year (1990-1999)
- Y_{it} is the rate of growth
- G_{it} is government spending on infrastructure in state i
- X_{it}: controls

Example 4: Mafia

In terms of growth rates:

- What is captured by γ_t?
- What is captured by α_i?

Why is α_i correlated with G_{it}?

Note: Paper uses panel data and IV.

Example 5: A community-college teacher like me

Fairlie et al, AER(2014)

$$Y_{ic} = \alpha_i + \lambda_c + \beta 1 Z_{ic} + u_{ic} + X_{ic}\gamma$$

where

- Y_{ic}:
  - dropped course?
  - passed course | finishing
  - grade | finishing
  - good grade? | finishing
  - enrolled in a similar course subsequently?
- Z_{ic} is an indicator for whether student i and j are part of the same minority
- X_{ic} is a vector of controls
- What does α_i capture?
- What does λ_c capture?

Example 5: A

λ_c and α_i

> *control[s] for instructor fixed effects and minority-specific course fixed effects. The former controls for the possibility that minority students take courses from instructors who have systematically different grading policies from other instructors, while the latter controls for selection by comparative advantage where minority students are drawn to courses that are a particularly good match or in which minority students are drawn to courses that are a particularly good match.*

p. 2574, Fairlie, Hoffmann, Oreopolous (2014)

Example 5: Findings

No findings without fixed effects. With fixed effects:

- dropped course?: $-0.02***$
- passed course $|$ finishing: 0.012
- grade $|$ finishing: $0.054***$
- good grade? $|$ finishing $0.024***$
- enrolled in a similar course subsequently? $0.013*$

Example 6: Income and democracy

PS6

$$democracy_{it} = \alpha_i + GDP_{it}\beta + u_{it}$$

1. Reverse causality?
2. What is in α_i?

Does it matter?

Introduction: beertax example

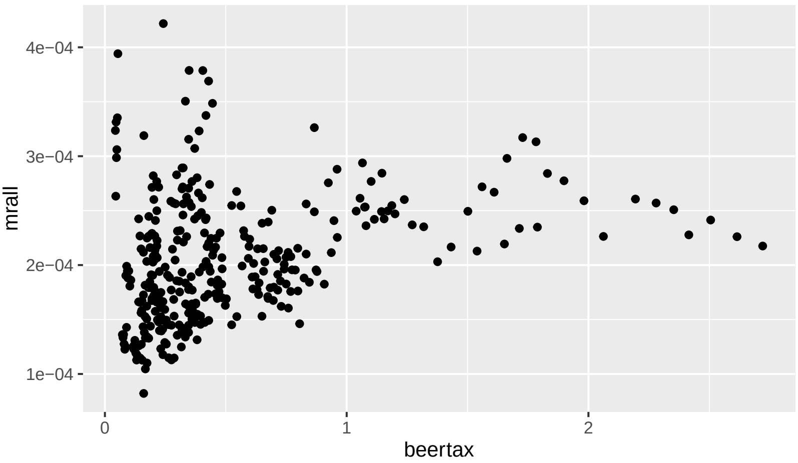
(from Stock and Watson)

```r
## Load the fatality data
library(haven)
beer_fatality <- read_dta("fatality.dta")
summary(beer_fatality)
```

```
##      state            year         spircons          unra
##  Min.   : 1.00   Min.   :1982   Min.   :0.790   Min.    :
##  1st Qu.:18.75   1st Qu.:1983   1st Qu.:1.300   1st Qu.:
##  Median :30.50   Median :1985   Median :1.670   Median :
##  Mean   :30.19   Mean   :1985   Mean   :1.754   Mean    :
##  3rd Qu.:42.50   3rd Qu.:1987   3rd Qu.:2.013   3rd Qu.:
##  Max.   :56.00   Max.   :1988   Max.   :4.900   Max.    :
##
##      perinc           emppop         beertax             s
##  Min.   : 9514   Min.   :42.99   Min.   :0.04331   Min.
##  1st Qu.:12086   1st Qu.:57.69   1st Qu.:0.20885   1st (
```

beertax:plot

```r
library("ggplot2")
qplot(beertax,mrall,data=beer_fatality)
```



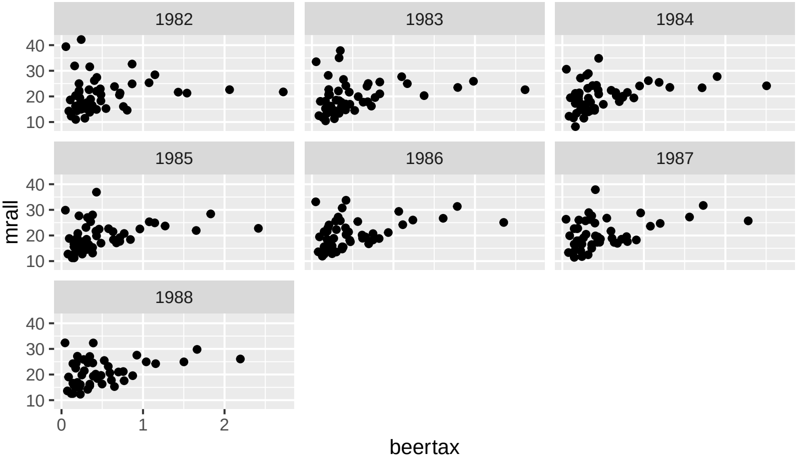
beertax: OLS

```
## Scale up fataiilities for readability
beer_fatality$mrall <- beer_fatality$mrall*100000
ols_reg <- lm(mrall~beertax,data=beer_fatality)
summary(ols_reg)
```

```
##
## Call:
## lm(formula = mrall ~ beertax, data = beer_fatality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.9060  -3.7768  -0.9436   2.8548  22.7643
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  18.5331     0.4357  42.539  < 2e-16 ***
## beertax       3.6461     0.6217   5.865 1.08e-08 ***
## ---
```

beertax: Plot for each year:

```
p <- qplot(beertax,mrall,data=beer_fatality)
p <- p + facet_wrap(~year)
p
```



beertax: Fixed effects results

```
# Fixed effects estimator
fe_reg <- lm(mrall~beertax+as.factor(state),data=beer_fatal
library(stargazer)
```

```
##
## Please cite as:

##  Hlavac, Marek (2015). stargazer: Well-Formatted Regress

##  R package version 5.2. http://CRAN.R-project.org/packag
```

```
stargazer(ols_reg,fe_reg,
          type = "text",
          keep = "beertax",
          keep.stat = c("n"))
```

```
##
## =========================================
```

Takeaways

1. What is in α_i?
   - "unobserved heterogeneity", or:
   - "fixed effect", or:
   - an intercept specific to the cross-section unit, or:
   - omitted variable that does not change over time
2. Key feature of panel data in economics: **Un**observed heterogeneity is generally correlated with the **observ**ables X_{it}

Unobserved heterogeneity: problem

- 1: Geometrically: *sketch81* and *sketch82*
- 2: Information from within- and between dimension
- 3: Between information not reliable when $\alpha_i \not\perp X_i$

Incidental parameters problem

Unobserved heterogeneity

To solve the issue that α_i, when treated as a RV, is correlated with X_i in most economic applications, we can treat it as a parameter to be estimated along with β.

This leads to the **incidental parameters problem**:

- the size of the parameter space grows with n
- not covered by standard extremum estimation proof
- inconsistency for conventional estimators of the common parameters
  - unless $T \to \infty$ or
  - model is linear + static

Example

(from Manuel Arellano)

$$X_{it} \sim \mathcal{N}(\alpha_i, \sigma^2), \ \ i = 1, \cdots, n, \ \ t = 1, \cdots, T$$

where $T \ geq2$, $n \to \infty$ and sampling is iid across (i, t).

Example: likelihood

- This is a fully specified model for X_{it}: use MLE.
- Objective: consistent estimator for the common parameter σ^2

The log-likelihood is given by

$$\mathcal{L}(\sigma, (\alpha_i)_i) \propto -nT \log \sigma - \sum_i \sum_t \frac{1}{2} \left(\frac{X_{it} - \alpha_i}{\sigma} \right)^2$$

Example: individual parameter estimates

Solve FOC for $\hat{\alpha}_i$:

$$0 = \frac{1}{\hat{\sigma}^2} \sum_t (X_{it} - \hat{\alpha}_i)$$

so that

$$\hat{\alpha}_i = \frac{1}{T} \sum_t X_{it} \equiv \bar{X}_i$$

Example: common parameters

Solve FOC for $\hat{\sigma}_i$:

$$0 = -\frac{nT}{\sigma} + \frac{1}{\hat{\sigma}^3} \sum_i \sum_t (X_{it} - \hat{\alpha}_i)^2$$

so that

$$\hat{\sigma}^2 = \frac{1}{nT} \sum_i \sum_t \left(X_{it} - \bar{X}_i \right)^2$$

Example: inconsistency

$$\hat{\sigma}^2 \xrightarrow{p} E\frac{1}{T}\sum_t \left(X_{it} - \bar{X}_i\right)^2$$
$$= \frac{T-1}{T}\sigma^2$$

where the first equality follows from the LLN. For the last equality, remember the degree-of-freedom correction for the sample variance in undergrad.

Conclusion:

- inconsistent by a factor 2 if $T = 2$
- inconsistency disappears as $T \to \infty$

IP: solutions

Solutions to the incidental parameters problem:

0. Bias corrections: $\times T/(T-1)$
1. Transforming the model to remove α_i
2. Finding a sufficient statistic for α_i
3. Large-T solutions + bias-reduction

Taxonomy

Solutions to the IP problem tend to be model- or class-specific. The remainder of this lecture gives you a menu of models and explains how the IP problem has been solved in those settings.

1. Relationship (α_i, X_i): "fixed effects" v "random effects"
2. Number of time periods T: "fixed-T" v "large-T"
3. Linear v nonlinear
4. Dynamic v static
5. One-dimensional v multi-dimensional heterogeneity; additivity

see `sketch72.png`

The static linear model

Model equation

In the static linear model, the regression equation states that for all $i = 1, \cdots, n$ and for all $t = 1, \cdots, T$:

$$Y_{it} = X_{it}\beta + \alpha_i + u_{it}$$

Stack these equations for each i across t into

$$Y_i = X_i\beta + \alpha_i \iota_T + u_i = X_i\beta + v_i$$

where:

- Y_i is a $T \times 1$ vector consisting of the dependent variables for unit i
- X_i is a $T \times K$ matrix
- ι_T is a $T \times 1$ column vector of ones
- v_i is the **composite error**

Model

To turn this equation into a model, we need to impose distributional assumptions on

$$(\alpha_i, X_i, u_i)$$

We will do this - unconventionally - at the end of this lecture. We will first define a class of estimators.

Estimators

The object of interest is β. All estimators for β will be of the form

$$\hat{\beta}_A = \left(\sum_i X_i' A' A X_i\right)^{-1} \left(\sum_i X_i' A' A y_i\right)$$

this corresponds to the OLS estimator in the **transformed** linear model

$$AY_i = AX_i\beta + Av_i$$

Estimator 1: Pooled OLS

- "Run a regression of Y on X".
- Corresponds to $A = I_T$
- Ignores panel aspect

Preview: consistency requires . . . ?

Estimator 2: LSDV

- "Add a dummy variable for each country"
- Corresponds to estimation in

$$\begin{aligned} Y_i &= X_i\beta + \sum_j 1\{i=j\}\alpha_i\iota_T + u_i \\ &= X_i\beta + D_i\alpha + u_i \end{aligned}$$

where D_i is a $T \times n$ matrix in which each column corresponds to a dummy variable for a country.

Estimator 2: LSDV: FWL

By FWL, we know that y on X_1 and X_2 is equivalent to

1. regressing y on X_1, call residuals y^*
2. regressing X_2 on X_1, call residuals X_2^*
3. regressing y^* on X_2^*

Equivalence is for $\hat{\beta}_2$:

Estimator 2: LSDV: FWL

Running OLS of y on $\{1\{i == j\}, j\}$ and X is equivalent to:

1. y on $\{1\{i == j\}, j\}$, call residuals My
2. X on $\{1\{i == j\}, j\}$, call residuals MX
3. My on MX

Estimator 2: LSDV: FWL (2)

- coefficient estimates are $(Z'Z)^{-1}Z'y$, so the
- predictions are $Z(Z'Z)^{-1}Z'y$, so the
- residuals are

$$y - Z(Z'Z)^{-1}Z'y = [I - Z(Z'Z)^{-1}Z']y$$

Estimator 2: LSDV: FWL (3)

$$Z(Z'Z)^{-1}Z' = \frac{1}{T}ZZ'$$

Estimator 2: LSDV: FWL (4)

Conclusion:

$$I - Z(Z'Z)^{-1}Z' = \begin{pmatrix} 1-1/T & -1/T & -1/T & 0 & 0 \\ -1/T & 1-1/T & -1/T & 0 & 0 \\ -1/T & -1/T & 1-1/T & 0 & 0 \\ 0 & 0 & 0 & 1-1/T & -1/T \\ 0 & 0 & 0 & -1/T & 1-1/T \\ 0 & 0 & 0 & -1/T & -1/T \end{pmatrix}$$

so that the LSDV estimator for $\hat{\beta}$ is equivalent to $\hat{\beta}_{A_{\mathrm{FE}}}$ with

$$A_{\mathrm{FE}} = I_T - \frac{1}{T}\iota_T \iota_T'$$

Estimator 2: LSDV + FE

An alternative way to think about the LSDV estimator.

In

$$Y_{it} = X_{it}\beta + \alpha_i + u_{it}$$

take averages across time on both sides of the equality to obtain

$$\bar{Y}_i = \bar{X}_i\beta + \alpha_i + \bar{u}_i$$

Now subtract the latter from the former to obtain

$$\tilde{Y}_i = \tilde{X}_i\beta + \tilde{u}_i$$

Estimator 2: FE

- We use the time-invariance of α_i to eliminate it from the (transformed) model.
- Relevant property of the transformation matrix A is that $A\iota_T = 0$.
- If $A\iota_T = 0$, then $Av_i = Au_i$

Preview Assumption for consistency of FE: $E(Au_i|AX_i) = 0$. No restriction on (α_i, X_i).

Estimator 3: FD

An alternative way to exploit the time-invariance of the unobserved heterogeneity is

$$
\begin{aligned}
Y_{it} &= X_{it}\beta + \alpha_i + u_{it} \\
Y_{it-1} &= X_{it-1}\beta + \alpha_i + u_{it-1} \\
\Delta Y_{it} &= \Delta X_{it}\beta + \Delta u_{it}
\end{aligned}
$$

The resulting, differenced, equation also does not feature α_i.

Estimator 3: FD: Transformation

The corresponding transformation is

$$A_{\mathsf{FD}} = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$

which also satisfies $A\iota_T = 0$.

Estimator 3

Some other possibilities:

$$A_{\text{LD}} = \begin{pmatrix} -1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$

$$A_{\text{FO}} = \begin{pmatrix} 1 & -1/3 & -1/3 & -1/3 \\ 0 & 1 & -1/2 & -1/2 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

Estimator 4: Between

Between estimator sets

$$A_{\text{BE}} = \frac{1}{T}\iota_T \iota_T'$$

- Note that $A_{\text{BE}} = \iota_T$
- Corresponds to running a regression on country averages

Estimator 5: Random effects

The random effects estimator is the optimal linear combination of the between and within (FE) estimators.

Consistency

`sketch73.png`

1. Consistency of OLS
2. Discuss exogeneity and multicollinearity assumptions in the model

$$AY_i = AX_i\beta + u_i$$

What's next

- We have seen the importance of assumptions such as

$$E(u_{is}|X_i t) = 0 \, \forall t \leq s$$

- Verify for a given example: texting bans
- We can build estimators from the ground up using those assumptions as a starting point.
- Do this using GMM

So, first: review of GMM

GMM for panels

GMM: Review

1. structure
2. examples
3. consistency and asymptotic normality
4. optimal weight matrix and efficiency

Panel data GMM

- An observation is a T-vector
- Every (s, t) exogeneity condition gives a moment
  - sequential: $1/2T(T+1)K$
  - strict exogeneity: T^2K
- Combine a transformation with exogeneity conditions

Panel data GMM: example

[FD + strict exogeneity]

Dynamic panels

A useful application of this framework is in dynamic panels.

- Model
- Why does pooled OLS fail?
- Why does FD fail?
- Absence of serial correlation generates instruments!

Dynamic panels

[Arellano-Bond]

Dynamic panels

- What if there is first-degree serial correlation?
- What if $\rho = 0$?
-

Conclusion

0. Linear static model + random effects: OLS or RE
1. Linear static models + fixed effects: FD or FE
2. Linear dynamic model: Arellano-Bond
3. Nonlinear static model?

Nonlinear panel data

Setting

We will study a specific class of models:

- nonlinear:
$$E(Y_{it}|X_i, \alpha_i) = h(\alpha_i + X_{it}\beta)$$
- fixed T
- fixed effects: no restriction on (α_i, X_i)
- strict exogeneity: $u_{is} \perp X_{it}$ for all s, t

Main question

- Q: "Can we formulate a consistent estimator for β?"
- Concern: incidental parameters problem

Binary choice

- We will focus on the binary choice model
- Solution to IP problem for binary choice applies to a handful of related models
- Generally, different nonlinear models require different solutions

Binary choice: model

For each $t \in \{1, \cdots, T\}$,

$$\begin{aligned} Y_{it}^* &= \alpha_i + X_{it}\beta + u_{it} \\ Y_{it} &= 1\{Y_{it}^* \geq 0\} \\ u_{it}|X_i, \alpha_i &\sim F(u|\cdot) \end{aligned}$$

Time homogeneity: α_i, β, F do not depend on t.

Binary choice: model (2)

- available: cross-section for consistent estimation of $L(Y_i, X_i)$
- fixed-T
- fixed effects: no assumptions on (α_i, X_i)
- strict exogeneity
- error terms are identically distributed

Binary choice model: identification

Informal statement of results in Chamberlain (2010):

- Bounded regressors: identification fails if $F \neq \Lambda$
- Unbounded regressors (a la Manski (1988, Assumption 2(c)):
  - identification holds more generally
  - $\sqrt{n}-$consistency only if $F = \Lambda$

Binary choice: logit: model

For each $t \in \{1, \cdots, T\}$,

$$Y_{it}^* = \alpha_i + X_{it}\beta + u_{it}$$
$$Y_{it} = 1\{Y_{it}^* \geq 0\}$$
$$u_i | X_i, \alpha_i \sim IIDLOG(0, 1)$$

where the "IID"-ness refers to the $t-$dimension: errors terms are independent across t

Binary choice: logit: ML (1)

To show: maximum likelihood estimator is inconsistent for β due to the incidental parameters (IP) problem.

Parameters to be estimated:

- β
- $(\alpha_i, i = 1, \cdots, n)$ - IPs

Derivation follows Arellano's notes.

Binary choice: logit: ML (2)

α_i enters only through likelihood contribution for individual i,

$$\sum_t \{ Y_{it} \ln \Lambda(X_{it}\beta + \alpha_i) + (1 - Y_{it}) \ln(1 - \Lambda(X_{it}\beta + \alpha_i)) \}$$

so ML estimator $\hat{\alpha}_i$ solves the FOC

$$\sum_t \left\{ Y_{it}(1 - \Lambda(X_{it}\hat{\beta} + \hat{\alpha}_i)) - (1 - Y_{it})\Lambda(X_{it}\hat{\beta} + \hat{\alpha}_i) \right\} = 0$$

- Q: Verify the above, using your logit skills from cross-sectional binary choice, remembering that $\Lambda' = \Lambda(1 - \Lambda)$.

Binary choice: logit: ML (3)

- Proceed with
  - $T = 2$
  - $X_{i1} = 0, X_{i2} = 1.$
- Leads to analytical solution, leveraging $X_{it} = X_t$

FOC simplifies to:

$$Y_{i1}(1-\Lambda(\hat{\alpha}_i))-(1-Y_{i1})\Lambda(\hat{\alpha}_i)+Y_{i2}(1-\Lambda(\hat{\beta}+\hat{\alpha}_i))-(1-Y_{i2})\Lambda(\hat{\beta}+\hat{\alpha}_i) = 0$$

implying

$$Y_{i1} + Y_{i2} = \Lambda(\hat{\alpha}_i) + \Lambda(\hat{\beta} + \hat{\alpha}_i)$$

Binary choice: logit: ML (4)

- Split by cases.
- For **switchers**, with $Y_{i1} + Y_{i2} = 1$, use logit symmetry
- Leads to

$$\hat{\alpha}_i = \begin{cases} -\infty & \text{if } Y_{i1} + Y_{i2} = 0 \\ -\hat{\beta}/2 & \text{if } Y_{i1} + Y_{i2} = 1 \\ +\infty & \text{if } Y_{i1} + Y_{i2} = 2 \end{cases}$$

- Only switchers are informative about β
- Q: Intuition?

Binary choice: logit: ML (5)

- Next: maximize the log likelihood with respect to β
- Plug in $\hat{\alpha}_i = -\hat{\beta}/2$ to obtain

$$\mathcal{L}_n\left(\hat{\beta}\right) = \sum_i 1\left\{Y_{i1} + Y_{i2} = 1\right\} \times$$
$$\left[Y_{i1} \ln \Lambda\left(-\hat{\beta}/2\right) + (1 - Y_{i1}) \ln\left[1 - \Lambda\left(-\hat{\beta}/2\right)\right]\right.$$
$$\left. + \; Y_{i2} \ln \Lambda\left(\hat{\beta}/2\right) + (1 - Y_{i2}) \ln\left[1 - \Lambda\left(\hat{\beta}/2\right)\right]\right].$$

- For the switchers (effective sample) with $y_{i1} + y_{i2} = 1$,

$$y_{i2} = 1 - y_{i1}$$

- By the symmetry of the logit CDF,

$$\Lambda\left(-\hat{\beta}/2\right) = 1 - \Lambda\left(\hat{\beta}/2\right),$$

Binary choice: logit: ML (6)

Then

$$\begin{aligned}
\mathcal{L}_n\left(\hat{\beta}\right) &= \sum_i 1\left\{Y_{i1}+Y_{i2}=1\right\}\times \\
&\left[(1-Y_{i2})\ln(1-\Lambda\left(\hat{\beta}/2\right)) + (Y_{i2})\ln\left[\Lambda\left(\hat{\beta}/2\right)\right]\right. \\
&\left.+\; Y_{i2}\ln\Lambda\left(\hat{\beta}/2\right) + (1-Y_{i2})\ln\left[1-\Lambda\left(\hat{\beta}/2\right)\right]\right] \\
&= 2\sum_i 1\left\{Y_{i1}+Y_{i2}=1\right\}\left(Y_{i2}\ln\Lambda\left(\hat{\beta}/2\right) + (1-Y_{i2})\ln(1-\Lambda\left(\hat{\beta}\right.\right.
\end{aligned}$$

Binary choice: logit: ML (7)

- FOC is (**verify**)

$$\sum_i 1\{Y_{i1} + Y_{i2} = 1\}\left(\Lambda\left(\hat{\beta}/2\right) - Y_{i2}\right) = 0$$

- The ML estimator for β sets

$$\Lambda\left(\hat{\beta}/2\right) = \frac{\sum_i 1\{Y_{i1} + Y_{i2} = 1\} Y_{i2}}{\sum_i 1\{Y_{i1} + Y_{i2} = 1\}} \equiv \hat{p}$$

- Note that the sample proportion

$$\hat{p} \to p \equiv P(Y_{i1} = 0, Y_{i2} = 1 | Y_{i1} + Y_{i2} = 1)$$

Binary choice: logit: ML (8)

Detour:

- Investigate $p = P(Y_{i1} = 0, Y_{i2} = 1 | Y_{i1} + Y_{i2} = 1, Xi, \alpha_i)$
- Will suppress dependence on (α_i, X_i)

Prep:

1. Remember that
$$P(Y_{i1} = 0) = 1 - \Lambda(\alpha_i)$$
2. Also,
$$P(Y_{i2} = 1) = \Lambda(\alpha_i + \beta)$$
3. Because of serial independence,
$$\begin{aligned} P(Y_{i1} = 0, Y_{i2} = 1) &= P(Y_{i1} = 0)P(Y_{i2} = 1) \\ &= (1 - \Lambda(\alpha_i))\Lambda(\alpha_i + \beta) \end{aligned}$$
4. Similarly,
$$P(Y_{i1} = 1, Y_{i2} = 0) = \Lambda(\alpha_i)(1 - \Lambda(\alpha_i + \beta))$$

Binary choice: logit: ML (9)

Finally,

$$
\begin{aligned}
p &= \frac{P(Y_{i1}=0, Y_{i2}=1, Y_{i1}+Y_{i2}=1)}{P(Y_{i1}+Y_{i2}=1)} \\
&= \frac{P(Y_{i1}=0, Y_{i2}=1)}{P(Y_{i1}=0, Y_{i2}=1)+P(Y_{i1}=1, Y_{i2}=0)} \\
&= \frac{(1-\Lambda(\alpha_i))\Lambda(\alpha_i+\beta)}{(1-\Lambda(\alpha_i))\Lambda(\alpha_i+\beta)+\Lambda(\alpha_i)(1-\Lambda(\alpha_i+\beta))}
\end{aligned}
$$

- Q: Show that α_i drops out!

Binary choice: logit: ML (10)

Using $\Lambda(u) =] exp(u)/(1+\exp(u))$ and $1-\Lambda(u) = 1/(1+\exp(u))$,

$$p = \frac{\exp(\alpha_i + \beta)}{\exp(\alpha_i + \beta) + \exp(\alpha_i)}$$
$$= \Lambda(\beta).$$

1. Does not depend on α_i
2. Does not equal $\Lambda(\beta/2)$

Binary choice: logit: ML (11)

Conclusion for the maximum likelihood estimator:

$$\hat{\beta} \to 2\beta$$

Binary choice: logit: ML (12)

1. Abrevaya (1997, Economics Letters) for the proof that $beta - hat- > 2beta$ in the more general context for an arbitrary number of regressors that can be anything, not just 0, 1
2. Open question: what is the bias when T>2? We know that $plimbeta - hat \in [T/(T-1)beta, 2beta]$ but that's all. If you can show it, you can send it to Abrevaya's journal. I suspect he will publish it.
3. The consistency is severe, likely decreases with T
4. Easy fix for $T = 2 : theta - tilde = theta - hat/2$. However, since the inconsistency is only known for this special case,
5. Differencing? Does not work in nonlinear models.

Binary choice: logit: CMLE (1)

- The solution involves finding a **sufficient statistic** $h(Y_i)$ for the incidental parameter α_i, i.e.

$$P(Y_i = y | h(Y_i), X_i, \alpha_i) = P(Y_i = y | h(Y_i), X_i)$$

- We saw one on the road to showing the inconsistency of MLE: $h(Y_i) = \sum_t Y_{it}$
- Conditional MLE
- Details: Andersen (1973, JASA), Chamberlain (1980, REStud)
- Efficiency: maintained if P is in exponential family (Hahn, 1998, ET)
    - includes binary choice logit
    - includes linear regression model with normal errors

Binary choice: logit: CMLE (2)

- Treat case with $T = 2$, general X_i:
  - see `Panel data - Binary choice CMLE.lyx`
- Algebra for general T is more involved, but the conclusion is similar.
  - see `sketch74,sketch75`

Binary choice: logit: CMLE (3)

- This derivation suggests a new criterion function to maximize which yields an extremum estimator.
- That criterion function will be concave.
- To establish consistency and asymptotic normality, we would then have to check identification, etc.
- Key insight is that the incidental parameter problem has been avoided by conditioning on the sufficient statistics.

Binary choice: Manski

The above trick is specific to logit. Can it be extended?

- Not to other, known distribution functions
- However, at the cost of identifiying β only up to scale, we can deal with the unknown-F case
- Approach turns binary choice model into Manski's semiparametric binary choice model

Binary choice: Manski: model

$$Y_{it}^* = \alpha_i + X_{it}\beta - U_{it}$$
$$Y_{it} = 1(Y_{it}^* \geq 0)$$
$$U_{it}|\alpha_i, X_i \sim F(u|X)$$

Binary choice: Manski: ID

Coming up:

$$med\left(d_{i2}-d_{i1}|\alpha_i, X_i, d_{i1}+d_{i2}=1\right)=sgn\left(\Delta X_i\beta\right) \qquad (1)$$

- Link to Manski's cross-sectional model
- Identical distributions (not necessarily independent) result in median-0

Binary choice: partial effects

Despite having a consistent estimator for β, we do not have an estimator for the APE or marginal effects. Why?

Binary choice: partial effects (2)

1. It depends on α_i, for which no consistent estimator is available!
2. **Partial identification** of a partial effect:
   - Chernozhukov, Fernandez-Val, Hahn, Newey (2013, ECTA)
   - Chernozhukov et al. (2015, JoE)

Binary choice: extending the results

Talk about ordered logit / transformation paper.

Nonlinear panels: other models

- count regression
- binomial choice
- ordered logit
- censored choice
- ...