Part F: Miscellaneous

Kirill Borusyak

ARE 213 Applied Econometrics

UC Berkeley, Fall 2024

F Outline

1. More on statistical inference
2. Multiplicative effects
3. Final thoughts

More on statistical inference

- We have previously covered:
  - Heteroskedasticity-robust (Eicker-Huber-White) SE
  - Cluster-robust (clustered) SE
  - Spatially-clustered (Conley) SE
  - Exposure-robust (shift-share) SE
  - CI from test inversion: randomization inference, Anderson-Rubin for weak IV
- Still not covered:
  - At what level to cluster? *[See also MacKinnon, Nielsen, Webb (2023)]*
  - Multi-way clustering
  - The case of few clusters [See Cameron and Miller (2014), Imbens and Kolesar (2016), and MacKinnon, Nielsen, Webb (2023)]
  - Bootstrap, wild bootstrap, Bayesian bootstrap, ... [Too big of a topic]

At what level to cluster?

- General principle: observations i, j should be in the same cluster if $\mathbb{E}\left[Z_i^\perp \varepsilon_i^\perp \varepsilon_j^\perp Z_j^{\perp\prime}\right] \neq 0$
  - Here $\perp$ = "residualized on included covariates"; assume few covariates first
- "Design-based" inference: cluster at the level of random assignment of the treatment

$$\mathbb{E}\left[Z_i^\perp Z_j^\perp \mid \boldsymbol{\varepsilon}, \text{covariates}\right] = 0 \qquad \text{for } c(i) \neq c(j)$$

  - Mutual correlations of the error terms do not matter
- "Model-based" inference: cluster at the level at which error terms are mutually uncorrelated:

$$\mathbb{E}\left[\varepsilon_i^\perp \varepsilon_j^\perp \mid \boldsymbol{Z}, \text{covariates}\right] = 0 \qquad \text{for } c(i) \neq c(j)$$

  - Assignment process of the instrument/treatment does not matter

At what level to cluster? (2)

- Be careful with many covariates, e.g. FEs with a small $\#$ of observations per FE
  - Demeaning introduces correlations among $Z_i^{\perp}\varepsilon_i^{\perp}$!
- Unit FEs in a short panel $\Longrightarrow$ must cluster by unit (in addition to anything else)
- Minimum wage triple-diff with many states, two periods, and two groups (restaurant workers & lawyers)
  - ✓ to cluster by state, ✗ to cluster by cross-sectional unit (state-group pair)
  - We have state-by-year FEs and only two observations per state-year
  - Triple-diff is equivalent to DiD across states and periods where the outcome is difference between the two groups

At what level to cluster? (3)

- More "conservative" clustering does not necessarily increase standard errors
  - If clustering by county is correct, clustering by state will give very similar SE
- But beware of too conservative clustering
  - Too few clusters $\Longrightarrow$ downward bias in SE

Multi-way clustering

- Consider OLS estimation of $Y_i = \beta' X_i + \varepsilon_i$ (extends naturally to IV)
  - Assume each unit belongs to group $g(i) \in \{1, \ldots, G\}$
  - And each unit belongs to (non-nested) group $h(i) \in \{1, \ldots, H\}$
- Examples:
  - Workers belong to state $g(i)$ and industry $h(i)$
  - Bilateral trade flow corresponds to exporter $g(i)$ and importer $h(i)$
- Two-way (a.k.a. double) clustering assumption *(cf. Cameron, Gelbach, Miller 2011)*:

$$\mathbb{E}\left[X_i \varepsilon_i \varepsilon_j X_j'\right] = 0 \quad \text{unless } g(i) = g(j) \textbf{ or } h(i) = h(j), \text{ or both}$$

  - Allows correlation of $X_i \varepsilon_i$ between unit pairs that share at least one cluster

Multi-way clustering (2)

- Variance estimator: for $G, H \to \infty$,

$$\widehat{Var}(\hat{\beta}) = (X'X)^{-1}\,\Omega\,(X'X)^{-1}, \quad \Omega = \sum_{i,j=1}^{N} X_i \hat{\varepsilon}_i \hat{\varepsilon}_j X_j' \cdot \mathbf{1}\left[g(i) = g(j) \text{ or } h(i) = h(j)\right]$$

- *Warning:* do not confuse it with one-way clustering by $(g(i), h(i))$ pair:

$$\Omega = \sum_{i,j=1}^{N} X_i \hat{\varepsilon}_i \hat{\varepsilon}_j X_j' \cdot \mathbf{1}\left[g(i) = g(j) \textbf{ and } h(i) = h(j)\right]$$

  - E.g. two-way clustering by state and industry $\neq$ clustering by state-industry
  - The former is more conservative than clustering by state; the latter is less conservative
  - "I cluster by state and industry" is ambiguous

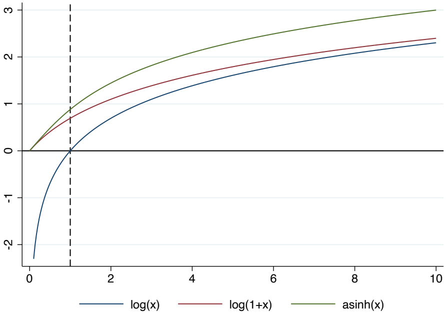
Outline

Multiplicative effects

- *Reading*: Chen and Roth (2024)
- With non-negative outcomes, it's often more natural to think that the effects are multiplicative, rather than additive
    - E.g. $D_i =$ regional trade agreement between countries, $Y_i =$ import value
    - E.g. $D_i =$ log price, $Y_i =$ demand
- How should we deal with this case? Models? Estimands? Estimators?
    - Start with homogeneous effects; then add heterogeneity
- Common practice: use $\log Y_i$ as outcome, $\log Y_i = \beta' X_i + \varepsilon_i$
    - Assuming $\mathbb{E}[\varepsilon_i \mid X_i] = 0$, OLS in logs is consistent for (constant effect) β

Issue of zeros

- If there are zeros ($Pr(Y_i = 0) > 0$), $\log Y_i$ is not well-defined
- Common to use log-like transformations: $\log(1 + Y_i)$ or inverse hyperbolic sine $\operatorname{arcsinh}(Y_i) \equiv \log\left(Y_i + \sqrt{1 + Y_i^2}\right)$



- With and without zeros, are these good ideas? Are there other options?

Modeling multiplicative effects

- Compare three models:

  1. $\log Y_i = \beta' X_i + \varepsilon_i, \quad \mathbb{E}[\varepsilon_i \mid X_i] = 0$
  2. $Y_i = \exp(\beta' X_i) U_i, \quad \mathbb{E}[U_i \mid X_i] = 1$
  3. $Y_i = \exp(\beta' X_i) + \varepsilon_i, \quad \mathbb{E}[\varepsilon_i \mid X_i] = 0$

- Are #1 and #2 equivalent? No! *(Santos Silva and Tenreyro 2006)*
- Are #2 and #3 equivalent? Yes! $\mathbb{E}[Y_i \mid X_i] = \exp(\beta' X_i)$in both; $\varepsilon_i = \mathbb{E}[\beta' X_i](U_i - 1)$

Estimating the exponential model

How would we estimate the model $\mathbb{E}\left[Y_i \mid X_i\right] = \exp\left(\beta' X_i\right) \equiv \mu(X_i, \beta)$?

- Could use moment conditions $\mathbb{E}\left[g(X_i) \cdot (Y_i - \exp(\beta' X_i))\right] = 0$ for any vector $g(\cdot)$
  - Most efficient $g(X_i) \propto \frac{\partial \mu(X_i,\beta)}{\partial \beta'} / \mathrm{Var}\left[Y_i \mid X_i\right] = \frac{\exp(\beta' X_i)}{\mathrm{Var}[Y_i|X_i]} X_i$
- Nonlinear least squares (NLLS):

$$\hat{\beta}_{NLLS} = \arg\min_b \sum_i \left(Y_i - \exp\left(b' X_i\right)\right)^2$$

$$\text{FOC:} \quad 0 = \sum_i \left(Y_i - \exp\left(\hat{\beta}'_{NLLS} X_i\right)\right) \cdot \exp\left(\hat{\beta}'_{NLLS} X_i\right) X_i$$

  - Consistent when $\mathbb{E}\left[Y_i \mid X_i\right] = \exp\left(\beta' X_i\right)$; efficient when $\mathrm{Var}\left[Y_i \mid X_i\right] = \text{const}$
  - E.g. because this is MLE assuming $Y_i \mid X_i \sim \mathcal{N}(\exp\left(\beta' X_i\right), \sigma^2)$
  - In practice, $\mathrm{Var}\left[Y_i \mid X_i\right]$ increases with $\mathbb{E}\left[Y_i \mid X_i\right] \Longrightarrow$ NLLS is very inefficient

Estimating the exponential model (2)

- Poisson regression: originates as MLE for count data, $Y_i \in \{0, 1, 2, \dots\}$:

$$Y_i \mid X_i \sim \text{Poisson}(\mu(X_i, \beta)), \qquad \text{i.e. } Pr(Y_i = k \mid X_i) = \frac{\mu(X_i, \beta)^k \exp\left(-\mu(X_i, \beta)\right)}{k!}$$

  - Log-likelihood: $\mathcal{L} = \sum_i \left(Y_i \cdot \beta' X_i - \exp\left(\beta' X_i\right)\right) + \text{const}$
  - FOC: $\sum_i \left(Y_i - \exp\left(\hat{\beta} X_i\right)\right) X_i = 0$
- But this $\hat{\beta}_{PPML}$ is well-defined for any data $Y_i \geq 0$ — **Poisson pseudo-maximum likelihood (PPML)** estimator
  - Consistency only requires $\mathbb{E}\left[Y_i \mid X_i\right] = \exp\left(\beta' X_i\right)$
  - Efficient if $\mathrm{Var}\left[Y_i \mid X_i\right] = \sigma^2 \mathbb{E}\left[Y_i \mid X_i\right]$ (not limited to "equi-dispersion," $\sigma^2 = 1$, as under actual Poisson model)
- There is also Gamma-PML: solves $\sum_i \left(\frac{Y_i - \exp\left(\hat{\beta} X_i\right)}{\exp\left(\hat{\beta} X_i\right)}\right) X_i = 0$

PPML with fixed effects

- We mentioned in part C1 that most nonlinear models with fixed effects suffer from an **incidental parameters problem**
  - PPML is quite special
- Wooldridge (1999) considers a short panel with $\mathbb{E}[Y_{it}] = \exp(\alpha_i + \beta' X_{it})$
  - PPML is consistent for β
- Fernandez-Val and Weidner (2016) consider a long panel ($N, T \to \infty$) with two-way fixed effects: $\mathbb{E}[Y_{it}] = \exp(\alpha_i + \gamma_t + \beta' X_{it})$
  - An equivalent setting: gravity model for Y_{ij}
  - Various estimators are consistent (because many observations per FE) but PPML doesn't suffer from bias of order $O_p(1/N + 1/T)$
- Correia, Guimaraes, Zylkin (2020): fast implementation (in Stata) with multi-dimensional fixed effects

Jeffrey Wooldridge
@jmwooldridge

Poisson regression can get one so far with so little trouble, why do so many still resist? Especially with panel data. It’s too bad we can’t give it another name to reflect the fact that its a fully robust estimator of conditional mean parameters.

Jeffrey Wooldridge
@jmwooldridge

Saying “I can’t use Poisson regression because of overdispersion” is tantamount to saying “I can’t use OLS because of heteroskedasticity.” In other words, nonsense.

Jeffrey Wooldridge
@jmwooldridge

“Here Lies Jeffrey Wooldridge. He Defended Poisson Regression Until the End.”

Application

Santos Silva and Tenreyro (2006) estimate the gravity equation of international trade across country pairs, assuming:

$$\mathbb{E}\left[Exports_{ij}\right] = \exp\left(\beta' X_{ij} + \alpha_i + \gamma_j\right)$$

Estimator: Dependent variable:	OLS ln (T_{ij})	OLS ln (1 + T_{ij})	Tobit ln (a + T_{ij})	NLS T_{ij}	PPML $T_{ij} > 0$	PPML T_{ij}
Log distance	−1.347**	−1.332**	−1.272**	−0.582**	−0.770**	−0.750**
	(0.031)	(0.036)	(0.029)	(0.088)	(0.042)	(0.041)
Contiguity dummy	0.174	−0.399*	−0.253	0.458**	0.352**	0.370**
	(0.130)	(0.189)	(0.135)	(0.121)	(0.090)	(0.091)
Common-language dummy	0.406**	0.550**	0.485**	0.926**	0.418**	0.383**
	(0.068)	(0.066)	(0.057)	(0.116)	(0.094)	(0.093)
Colonial-tie dummy	0.666**	0.693**	0.650**	−0.736**	0.038	0.079
	(0.070)	(0.067)	(0.059)	(0.178)	(0.134)	(0.134)
Free-trade agreement dummy	0.310**	0.174	0.137**	1.017**	0.374**	0.376**
	(0.098)	(0.138)	(0.098)	(0.170)	(0.076)	(0.077)
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9613	18360	18360	18360	9613	18360
RESET test p-values	0.000	0.000	0.000	0.000	0.564	0.112

(Santos Silva and Tenreyro 2006, Fig. 5)

A causal interpretation and heterogeneous effects

- We now have several *models* + *estimators*:
  - $\mathbb{E}[\log Y_i \mid X_i] = \beta' X_i$ or $\mathbb{E}[\log(1 + Y_i) \mid X_i] = \beta' X_i$ + OLS
  - $\mathbb{E}[Y_i \mid X_i] = \exp(\beta' X_i)$ + NLLS, PPML, Gamma-PML
- But for causal questions all of this is a wrong starting point
  - Potential outcomes is our model!
  - What are the estimands of different estimators? And what do we want?
  - Assume for now random assignment of treatment D_i

Case without zeros

- OLS of $\log Y_i$ on D_i: $\tau = \mathbb{E}\left[\log Y_i(1) - \log Y_i(0)\right]$
- Poisson of Y_i on D_i: $\mathbb{E}\left[Y_i(0)\right] = \exp\left(\beta_0\right)$, $\mathbb{E}\left[Y_i(1)\right] = \exp\left(\beta_0 + \tau\right) \implies$

$$\tau = \log \mathbb{E}\left[Y_i(1)\right] - \log \mathbb{E}\left[Y_i(0)\right]; \qquad \exp\left(\tau\right) - 1 = \frac{\mathbb{E}\left[Y_i(1) - Y_i(0)\right]}{\mathbb{E}\left[Y_i(0)\right]}$$

- What are the differences? (Cf. *Chen and Roth 2024*)
- Poisson identifies the ATE in levels, rescaled by the control mean
  - The effect may be dominated by the right tail of $Y_i(0)$
  - That may be what the policymaker cares about

Case with zeros

- Additional issue with zero: there are two types of responses
  - **Extensive margin**: $Pr(Y_i(1) = 0) - Pr(Y_i(0) = 0)$
  - **Intensive margin:** $\mathbb{E}\left[Y_i(1) - Y_i(0) \mid Y_i(0) > 0, Y_i(1) > 0\right]$
- Log-like transformations are very dependent on measurement units of Y_i *(Chen and Roth 2024)*:
  - If extensive margin $\neq 0$, by rescaling Y_i any real number can become the estimand!
- Pros and cons of other methods depend on the goal
- Case 1: you don't care about separating extensive and intensive margins
  - E.g. $Y_i = \#$ of publications in a year: 0 has no special meaning
  - PPML still yields $\mathbb{E}\left[Y_i(1) - Y_i(0)\right] / \mathbb{E}\left[Y_i(0)\right]$, a mix of the two margins

Case with zeros (2)

- Case 2: you want to isolate the two margins
  - E.g. $Y_i = \#$ of hours worked per week; extensive margin $=$ non-employment
  - For extensive margin, can regress $\mathbf{1}[Y_i > 0]$ on D_i
  - Intensive margin is not point identified because of selection: can't just drop zeros
    - But "Lee bounds" are available *(Lee 2009, Semenova 2023)*
- Case 3: you can take a stand on how to combine the two margins: e.g.

$$\mathcal{U}(Y_i) = \begin{cases} \log Y_i, & Y_i > 0 \\ -c, & Y_i = 0 \end{cases}$$

  - Then give up on scale invariance and regress $\mathcal{U}(Y_i)$ on D_i

PPML diff-in-diff

- Wooldridge (2023) extends DiD imputation and its regression implementation to the multiplicative model with staggered adoption
- Assume multiplicative parallel trends at the cohort level:

$$\mathbb{E}\left[Y_{it}(0) \mid E_i = e\right] = \exp\left(\alpha_t + \beta_e\right)$$

- Use untreated data to estimate α_t and β_e by TWFE PPML, then estimate CATT (in levels)

$$CATT_{et} = \bar{Y}_{t|E_i=e} - \exp\left(\hat{\alpha}_t + \hat{\beta}_e\right)$$

- As in Wooldridge (2021), can implement in a single step: PPML regression on TWFE and dummies for each treated cohort-period
  - Coefficients are interpreted as $\log \mathbb{E}\left[Y_{it}(1) \mid E_i = e\right] - \log \mathbb{E}\left[Y_{it}(0) \mid E_i = e\right]$
  - Can convert ATT as % of untreated mean or ATT in levels

Outline

Good practices (in my view)

1. Motivate the treatment variable

- Use a formal or informal model (causal or structural), not identification concerns
- Goal: to avoid unmodeled spillovers, have relatively homogeneous effects (for external validity), more useful estimand
- Unit of analysis?
  - Autor, Dorn, Hanson (2013): import competition by region rather than industry, to capture spillovers
- Binary or continuous treatment? In levels or logs? Rescaled by something?
  - Broda and Parker (2014): receiving stimulus payment as a dummy or $ amount?
  - Autor et al. (2013): import competition growth by industry measured as...

$$\frac{Imports_t - Imports_{t-1}}{Empl_{t-1}} \quad \text{or } \Delta \frac{Imports_t}{Empl_t} \quad \text{or } \Delta \log Imports_t$$

Good practices (in my view)

2. Think and talk about the error term/potential outcomes

- Exogeneity/endogeneity don't mean anything without specifying them
- Contextual knowledge helps: what else affects the outcome
- Theory helps: e.g. demand and supply

Good practices (in my view)

3. Don't confuse the model (setting + assumptions), estimand, and estimator

- For causal questions, potential outcomes or a DAG is the model — not the regression you run
- Be clear about the estimand, especially when spillovers are relevant
- Questionable practice: discussing threats to identification without stating the identification assumptions (and the estimand)

Good practices (in my view)

4. Distinguish natural and quasi experiments

- **Natural experiments** (or design-based identification strategies) are "serendipitous randomized trials" *(DiNardo 2008)*
  - You can describe an experiment that your treatment or instrument approximates, with many randomly determined shocks
  - Don't fake it: e.g. the RD cutoff is not randomized
- **Quasi-experiments** (or model-based identification strategies):
  - You can describe a treatment and control group.
  - They are imbalanced (or small) but you'll still cautiously compare them in some way: e.g. on trends but not on levels in a DiD
- **Neither:** One big shock, with no clear treatment and control group

Good practices (in my view)

5. Be careful with notation

- Find several mistakes:

$$Y_{it} = \alpha_i + \alpha_s \times \alpha_t + \sum_{k \neq -1} \beta_k D_k + X_{it} + \varepsilon_{it}$$

where $\alpha_s \times \alpha_t$ are state-by-year fixed effects and D_k are indicators of k periods since treatment

Thanks!

If you suspect a typo or mistake, send me an email