Part G: Models with Multiplicative Effects

Kirill Borusyak

ARE 213 Applied Econometrics

UC Berkeley, Fall 2025

G Outline

1. Modeling multiplicative effects

2. Multiplicative effects and potential outcomes

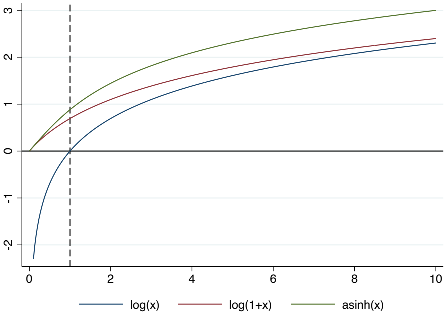
- *Reading*: Chen and Roth (2024)

Multiplicative effects

- With non-negative outcomes, it's often more natural to think that the effects are multiplicative, rather than additive
  - E.g. $X_i =$ regional trade agreement between countries, $Y_i =$ import value
  - E.g. $X_i =$ 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 $\mathbb{E}[\varepsilon_i \mid X_i] = 0$ (i.e., $\mathbb{E}[\log Y_i \mid X_i] = \beta' X_i$), 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 $\operatorname{arcsinh}(Y_i) \equiv \log\left(Y_i + \sqrt{1 + Y_i^2}\right)$ (well-defined for $Y_i < 0$ and symmetric)



- 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! Define $\varepsilon_i = \exp(\beta' X_i)(U_i - 1)$
    - Both models impose $\mathbb{E}[Y_i \mid X_i] = \exp(\beta' X_i)$
      (or equivalently $\log \mathbb{E}[Y_i \mid X_i] = \beta' X_i$ instead of $\mathbb{E}[\log Y_i \mid X_i] = \beta' X_i$ in #1)

Estimating the exponential model

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

- Could use moment conditions $\mathbb{E}[g(X_i) \cdot (Y_i - \exp(\beta' X_i))] = 0$ for any vector $g(\cdot)$
  - Most efficient $g(X_i) \propto \frac{\partial \mu(X_i,\beta)}{\partial \beta'} / \mathrm{Var}[Y_i \mid X_i] = \frac{\exp(\beta' X_i)}{\mathrm{Var}[Y_i|X_i]} X_i$
- Nonlinear least squares (NLLS): $g(X_i) = \exp(\beta' X_i) \cdot X_i$
  - Efficient when $\mathrm{Var}[Y_i \mid X_i] =$ const (usually very wrong in practice)
- Poisson Pseudo Maximum Likelihood (PPML): $g(X_i) = X_i$
  - Efficient when $\mathrm{Var}[Y_i \mid X_i] \propto \mu(X_i, \beta)$
- Gamma-PML: $g(X_i) = \frac{1}{\exp(\beta' X_i)} X_i$
  - Efficient when $\mathrm{Var}[Y_i \mid X_i] \propto \mu(X_i, \beta)^2$

Nonlinear least squares (NLLS)

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

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

- This is MLE under $Y_i \mid X_i \sim \mathcal{N}\left(\exp\left(\beta' X_i\right), \sigma^2\right)$
  - But consistent under $\mathbb{E}\left[Y_i \mid X_i\right] = \exp\left(\beta' X_i\right)$ without normality and homoskedasticity
  - And efficient if additionally $\mathrm{Var}\left[Y_i \mid X_i\right] = \text{const}$, without normality

PPML

- 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: $\mathcal{L} = \sum_i (Y_i \cdot \beta' X_i - \exp(\beta' X_i)) + \text{const}$
  - 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 $\mathbb{E}[Y_i \mid X_i] = \exp(\beta' X_i)$
  - Efficient if $\text{Var}[Y_i \mid X_i] = \sigma^2 \mathbb{E}[Y_i \mid X_i]$ (not limited to "equi-dispersion," $\sigma^2 = 1$, as under actual Poisson model)

PPML with fixed effects

- We mentioned in part D1 that most estimators for nonlinear models with FE suffer from an **incidental parameters problem**
  - PPML is quite special
- Wooldridge (1999) considers a short panel with $\mathbb{E}\left[Y_{it} \mid \boldsymbol{X}_i, \alpha_i\right] = \exp\left(\alpha_i + \beta' X_{it}\right)$ with strictly exogenous 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: $\mathbb{E}\left[Y_{it} \mid \boldsymbol{X}_i, \alpha_i\right] = \exp\left(\alpha_i + \gamma_t + \beta' X_{it}\right)$
  - 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 $O(1/N + 1/T)$
- Correia et al. (2020): fast implementation (in Stata) with multi-dimensional FEs

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:

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

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), Table 5)

Outline

A causal interpretation and heterogeneous effects

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

Case without zeros

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

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

- 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 (2025))*

- 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:

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

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

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

- As in Wooldridge (2025), 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

Thanks!

Please fill in the evaluations NOW!

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

References I

Chen, J. and J. Roth (2024): "Logs with Zeros? Some Problems and Solutions," *Quarterly Journal of Economics*, 139, 891–936.

Correia, S., P. Guimarães, and T. Z. Zylkin (2020): "Fast Poisson estimation with high-dimensional fixed effects," *Stata Journal*, 20, 95–115.

Fernandez-Val, I. and M. Weidner (2016): "Individual and time effects in nonlinear panel models with large N, T," *Journal of Econometrics*, 192, 291–312, arXiv: 1311.7065 Publisher: Elsevier B.V.

Lee, D. S. (2009): "Training, wages, and sample selection: Estimating sharp bounds on treatment effects," *Review of Economic Studies*, 76, 1071–1102.

Santos Silva, J. and S. Tenreyro (2006): "The log of gravity," *Review of Economics and Statistics*, 88, 641–658.

Semenova, V. (2025): "Generalized Lee Bounds," *Journal of Econometrics*, arXiv: 2008.12720.

Wooldridge, J. M. (1999): "Distribution-free estimation of some nonlinear panel data models," *Journal of Econometrics*, 90, 77–97.

References II

——— (2023): "Simple Approaches to Nonlinear Difference-in-Differences with Panel Data," *Econometrics Journal*.

——— (2025): "Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators," *Empirical Economics*.