Part D: Panel Data Methods

D1: Linear Panel Data Methods

Kirill Borusyak

ARE 213 Applied Econometrics

UC Berkeley, Fall 2025

Panel data methods: Outline

1. Linear panel data methods with constant effects
2. Canonical DiD and event studies
3. DiD with staggered adoption
4. DiD extensions
5. Synthetic control methods and factor models

D1 outline

1. Estimation, efficiency, inference
2. Extensions
3. Application: Mover designs and the AKM model

Readings: IW Lecture 3 (also CT Ch.21-22 and JW Ch. 10-11)

Motivation

- Selection on observables is rarely convincing in cross-sectional data
  - Self-selection is complex, too many unknown confounders
- And we don't always have a convincing instrument
- To allow for selection on (some) unobservables, leverage repeated observations for the same unit over time — **panel data**
  - How do outcomes change when treatment changes?
  - This doesn't resolve the fundamental problem of causal inference but helps control for unobserved confounders that are time-invariant
- Panel data are also helpful to evaluate effect dynamics

Linear panel model

- For $i = 1, \ldots, I$ and $t = 1, \ldots, T$, consider a constant-effects model

  $$Y_{it} = \beta' X_{it} + \alpha_i + \varepsilon_{it}$$

  where β is of interest; α_i is additive "unobserved heterogeneity"

- Denote $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{iT})$, $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{iT})$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})$

- Every (i, t) is observed $\Longrightarrow$ **balanced panel**; with missing data: unbalanced panel

Asymptotic sequences

- Conventional asymptotic (our focus): **short panel**
  - A growing sample of I units for a fixed number of periods, T
  - Appropriate when $I \gg T$
- Alternative asymptotic: **long panel**
  - The sample grows by increasing both I and T (at the same or different rates)
  - More appropriate when $I \approx T$

Notions of exogeneity

- **Contemporaneous exogeneity**: $\mathbb{E}\left[\varepsilon_{it} \mid X_{it}, \alpha_i\right] = 0$
  - Interpretation: no time-varying confounders
  - A natural assumption... but insufficient to identify β
- (Main focus) **Strict exogeneity**: $\mathbb{E}\left[\varepsilon_{it} \mid X_{i1}, \ldots, X_{iT}, \alpha_i\right] = 0$
  - For $h > 0$, $\mathrm{Cov}\left[\varepsilon_{it}, X_{i,t-h}\right] = 0$: precludes unmodeled dynamic effects (but solved by including lags of RHS variables)
  - $\mathrm{Cov}\left[\varepsilon_{it}, X_{i,t+h}\right] = 0$: precludes "**feedback effects**" from Y_{it} to $X_{i,t+h}$
    - ★ Also precludes $X_{it} \equiv Y_{i,t-1}$ as one of the RHS variables — "**dynamic panel**" models
- (For later) **Sequential exogeneity**: $\mathbb{E}\left[\varepsilon_{it} \mid X_{i1}, \ldots, X_{it}, \alpha_i\right] = 0$
  - Allows feedback effects and dynamic panels models

Random and fixed effects

- If selection on α_i is allowed, $\mathbb{E}\left[\alpha_i \mid \boldsymbol{X}_i\right] \neq$ const, α_i is called **fixed effect**
  - If $\mathbb{E}\left[\alpha_i \mid \boldsymbol{X}_i\right] =$ const (no selection on unobservables) $\Longrightarrow \alpha_i$ is a **random effect**
  - These labels are not about whether α_i is stochastic (always think random sample of $(\alpha_i, \boldsymbol{X}_i, \boldsymbol{\varepsilon}_i, \boldsymbol{Y}_i)_{i=1}^N$)
- To estimate β in the FE model, remove α_i in different ways:
  - "FE regression": Dummy variable regression $=$ within transformation
  - First differences
  - Long differences
- *Note:* random effects model allows for more estimation methods
  - "Pooled OLS" regression of Y_{it} on X_{it} across i, t identifies β
  - "Random effects" Generalized Least Squares estimator

Estimation by dummy variables (FE) regression

- View $\{\alpha_i\}$ as a set of nuisance parameters multiplying dummies for each unit:

$$Y_{it} = \beta' X_{it} + \sum_{j=1}^{I} \alpha_j W_{ji} + \varepsilon_{it}, \qquad W_{ji} \equiv \mathbf{1}\left[i = j\right] \tag{$*$}$$

  - *Note*: don't write e.g. $Y_{it} = \beta' X_{it} + \sum_j \alpha_j + \varepsilon_{it}$

Estimation by dummy variables (FE) regression

- By FWL, OLS estimation of $(*)$ = simple OLS after "**within transformation**":

$$\dot{Y}_{it} = \beta' \dot{X}_{it} + \dot{\varepsilon}_{it}, \qquad \dot{V}_{it} \equiv V_{it} - \bar{V}_i = V_{it} - \frac{1}{T}\sum_{s=1}^{T} V_{is}$$

  - *Exercise*: prove this
  - This is much faster than using I dummies! Use *reghdfe* in Stata, *fixest* in R
- For consistency of $\hat{\beta}$, we need

$$\mathbb{E}\left[\dot{X}_{it}\dot{\varepsilon}_{it}\right] = \mathbb{E}\left[\left(X_{it} - \frac{1}{T}\sum_{s=1}^{T} X_{is}\right)\left(\varepsilon_{it} - \frac{1}{T}\sum_{s=1}^{T}\varepsilon_{is}\right)\right] = 0$$

  - Holds under strict exogeneity but not contemporaneous/sequential exogeneity
  - *Question:* why is strict exogeneity necessary for $(*)$?

Estimation by first & long differencing

- The FE model also implies the first-difference (FD) specification:

$$\Delta Y_{it} = \beta' \Delta X_{it} + \Delta \varepsilon_{it}, \qquad \Delta V_{it} \equiv V_{it} - V_{i,t-1}, \quad t = 2, \dots, T$$

  - $\mathbb{E}\left[\Delta X_{it} \cdot \Delta \varepsilon_{it}\right] = 0$ under strict exogeneity (but not contemporaneous/sequential)
  - *Exercise:* with $T = 2$, OLS estimators of FE and FD equations are identical (even in unbalanced panels) — but not otherwise

- And a long-difference specification: for $h > 1$,

$$Y_{it} - Y_{i,t-h} = \beta' \left(X_{it} - X_{i,t-h}\right) + \left(\varepsilon_{it} - \varepsilon_{i,t-h}\right), \qquad t = h+1, \dots, T$$

with $\mathbb{E}\left[\left(X_{it} - X_{i,t-h}\right) \cdot \left(\varepsilon_{it} - \varepsilon_{i,t-h}\right)\right] = 0$ under strict exogeneity

Choosing between FE and FD estimators

- FE and FD estimate β from the same model under the same assumption (strict exogeneity). So how do we choose?

1. Efficiency:
    - FE estimator is efficient when ε_{it} are serially uncorrelated
    - FD estimator is efficient when ε_{it} follow a random walk (i.e., $\Delta\varepsilon_{it}$ are serially uncorrelated)
    - It's not about persistence in the outcome (which could come from α_i) but about differential persistence of observations close in time
    - FD can lose a lot of data in unbalanced panels

Choosing between FE and FD estimators (2)

2. Robustness to violations of assumptions:
    - When contemporaneous or sequential exogeneity holds but strict exogeneity doesn't, bias($\hat{\beta}_{FE}$) $= O(1/T)$ while bias($\hat{\beta}_{FD}$) $= O(1)$ (for weakly dependent data)
    - Differencing can exacerbate measurement error in X_{it}
3. Fads: FE estimation is more popular; it *appears* that you've controlled for more
    - False, plus FD & long-diffs are more transparent (especially in more complex situations)
    - If you use a FE specification, always rewrite it in FD

https://x.com/KhoaVuUmn/status/1630576551325495296

Inference

- In short panels, always cluster SE by i
  - See Bertrand, Duflo, and Mullainathan (2004) for an illustration
- In FD regressions, non-clustered (heteroskedasticity-robust) SEs require

$$\mathrm{Cov}\left[\Delta X_{it}\Delta\varepsilon_{it}, \Delta X_{is}\Delta\varepsilon_{is}\right] = 0$$

  - Unless X_{it} or ε_{it} is a random walk, ΔX_{it} and $\Delta\varepsilon_{it}$ are serially correlated
- In FE regressions, need $\mathrm{Cov}\left[\dot{X}_{it}\dot{\varepsilon}_{it}, \dot{X}_{is}\dot{\varepsilon}_{is}\right] = 0$
  - But X_{it} and ε_{it} are often serially correlated
  - And even if ε_{it} are uncorrelated, $\dot{\varepsilon}_{it}$ are *negatively* serially correlated in short panels (although DoF correction by $\frac{IT}{IT-I-\dim(X)}$ can correct for that)

Remarks

- FE model cannot estimate the effects of time-invariant variables
  - They get killed by both within transformation and differencing
- There is no intercept in the FE model
  - But can estimate $\hat{\alpha}_i = \frac{1}{T}\sum_t \left(Y_{it} - \hat{\beta}' X_{it}\right)$: unbiased for α_i but not consistent in short panels
  - And report $\frac{1}{I}\sum_i \hat{\alpha}_i$ that is consistent for $\mathbb{E}\left[\alpha_i\right]$
- With FEs, R^2 is often very high and not very informative
  - Instead, can use "**within** R^2"
  - I.e., R^2 after within-transformation: % of $\mathrm{Var}\left[\dot{Y}_i\right]$ explained by $\mathrm{Var}\left[\dot{X}_i\right]$

Outline

Extensions

1. Random effects estimation
2. Two-way fixed effects
3. Models with random slopes
4. Modeling dynamic effects
5. Nonlinear panel models
6. Fixed effects beyond panel data

Random effects estimation

- In addition to $\mathbb{E}\left[\varepsilon_{it} \mid \boldsymbol{X}_i, \alpha_i\right] = 0$, assume $\mathbb{E}\left[\alpha_i \mid \boldsymbol{X}_i\right] = 0$
- Then **Pooled OLS** estimator is consistent:

$$Y_{it} = \beta' X_{it} + v_{it}, \quad v_{it} = \alpha_i + \varepsilon_{it}, \quad \mathrm{Cov}\left[v_{it}, X_{it}\right] = 0$$

  - But inefficient: v_{it} are correlated across all t

- RE estimator: feasible GLS for $\mathrm{Var}\left[\boldsymbol{\varepsilon}_i \mid \boldsymbol{X}_i, \alpha_i\right] = \sigma_\varepsilon^2 \mathbb{I}_T$ and $\mathrm{Var}\left[\alpha_i \mid \boldsymbol{X}_i\right] = \sigma_\alpha^2$:

$$\hat{\beta}_{RE} = \left(\sum_i \boldsymbol{X}_i' \hat{\Omega}^{-1} \boldsymbol{X}_i\right)^{-1} \left(\sum_i \boldsymbol{X}_i' \hat{\Omega}^{-1} \boldsymbol{Y}_i\right), \quad \hat{\Omega} = \begin{pmatrix} \hat{\sigma}_\alpha^2 + \hat{\sigma}_\varepsilon^2 & \hat{\sigma}_\alpha^2 & \dots & \hat{\sigma}_\alpha^2 \\ & \dots & & \\ \hat{\sigma}_\alpha^2 & \hat{\sigma}_\alpha^2 & \dots & \hat{\sigma}_\alpha^2 + \hat{\sigma}_\varepsilon^2 \end{pmatrix}$$

Random effects estimation (2)

- Consistent even without homoskedastic and serially uncorrelated ε_{it}
- Allows for time-invariant X_{it}
- Can be obtained via **quasi-differencing**: OLS for

$$\left(Y_{it} - \hat{\lambda}\bar{Y}_i\right) = \beta' \left(\boldsymbol{X}_{it} - \hat{\lambda}\bar{\boldsymbol{X}}_i\right) + \left(v_{it} - \hat{\lambda}\bar{v}_i\right)$$

where $\hat{\lambda} = 1 - \left(1 + T\hat{\sigma}_\alpha^2/\hat{\sigma}_\varepsilon^2\right)^{-1/2}$

  - $\hat{\lambda} \approx 1$, so $\hat{\beta}_{RE} \approx \hat{\beta}_{FE}$, when T is large or $\mathrm{Var}\left[\alpha_i\right] \gg \mathrm{Var}\left[\varepsilon_{it}\right]$

Two-way fixed effects (TWFE)

- Besides α_i, we may want to include (additive) period effects γ_t to capture shocks that affect all units:
$$Y_{it} = \beta' X_{it} + \alpha_i + \gamma_t + \varepsilon_{it} \tag{\#}$$
- Unlike α_i, period FEs are non-stochastic parameters (on period dummies)
- (#) requires an innocuous normalization on $\{\alpha_i\}$ or $\{\gamma_t\}$
- FE estimator: in balanced panels, OLS from **double-differenced** specification:
$$\ddot{Y}_{it} = \beta' \ddot{X}_{it} + \ddot{\varepsilon}_{it}, \qquad \ddot{V}_{it} \equiv \left(V_{it} - \bar{V}_i\right) - \left(\bar{V}_t - \bar{V}\right)$$
  - Follows from FWL, because unit and period dummies are in-sample orthogonal
- FD estimator: OLS from $\Delta Y_{it} = \beta' \Delta X_{it} + \Delta\gamma_t + \Delta\varepsilon_{it}$ where $\Delta\gamma_t$ are period FEs

Models with random slopes

- Consider $Y_{it} = \alpha_i + \beta' X_{it} + \gamma_i' Z_{it} + \varepsilon_{it}$
  - E.g. unit-specific linear trends: $Y_{it} = \alpha_i + \beta' X_{it} + \gamma_i' t + \varepsilon_{it}$
- Dummy variable representation and estimation:

$$Y_{it} = \sum_{j=1}^{I} \alpha_j W_{ji} + \beta' X_{it} + \sum_{j=1}^{I} \gamma_j' Z_{jt} W_{ji} + \varepsilon_{it}$$

- FWL: for each *i* separately, residualize Y_{it}, X_{it} on $(1, Z_{it})$ in the time series

Modeling dynamic effects

- **Distributed lags** model can be accommodated with no change:

$$Y_{it} = \beta_0' X_{it} + \beta_1' X_{i,t-1} + \dots \beta_L' X_{i,t-L} + \alpha_i + \varepsilon_{it}$$

- **Lagged dependent variable** on the RHS: dynamic panel model

$$Y_{it} = \rho Y_{i,t-1} + \beta' X_{it} + \alpha_i + \varepsilon_{it}$$

  - Violates strict exogeneity: ε_{it} can't be mean-independent of $Y_{i,s-1}$ for $s = t+1$
  - At best, we can hope for sequential exogeneity of $Y_{i,t-1}$:

$$\mathbb{E}\left[\varepsilon_{it} \mid Y_{i1}, \dots, Y_{i,t-1}, \boldsymbol{X}_i, \alpha_i\right] = 0$$

  - FE estimator is not consistent in short panels: "Nickell bias" $\propto 1/T$
  - FD estimator has bias $O(1)$. But FD specification can be estimated by IV...

Arellano-Bond estimator

- FD specification: $\Delta Y_{it} = \rho \Delta Y_{i,t-1} + \beta' \Delta X_{it} + \Delta \varepsilon_{i,t}$
- OLS doesn't work: $\mathrm{Cov}\left[\Delta Y_{i,t-1}, \Delta \varepsilon_{it}\right] = \mathrm{Cov}\left[Y_{i,t-1} - Y_{i,t-2}, \varepsilon_{it} - \varepsilon_{i,t-1}\right] \neq 0$
- But can use $Y_{i1}, \dots, Y_{i,t-2}$ as IVs for $\Delta Y_{i,t-1}$ $\Rightarrow$ Arellano and Bond (1991) estimator: GMM for moment conditions collected across t

$$\mathbb{E}\left[(\Delta Y_{it} - \rho \Delta Y_{i,t-1} - \beta' \Delta X_{it}) \cdot (Y_{i1}, \dots Y_{i,t-2}, \Delta X'_{it})'\right] = 0, \qquad t = 3, \dots, T$$

- These IVs can be weak, especially when ρ is close to 1
  - Improvements are available, e.g. Blundell and Bond (1998)
- Similar ideas can be applied with sequentially exogenous X_{it} other than $Y_{i,t-1}$ (with feedback effects of $\varepsilon_{i,t-1}$ on X_{it})

Nonlinear panel data models

Nonlinear models with fixed effects are much more complicated. Consider binary choice models:

$$Pr(Y_{it} = 1 \mid X_{it}, \alpha_i) = F(\beta' X_{it} + \alpha_i), \qquad F = \text{probit or logistic}$$

- Likelihood estimation of β along with $\{\alpha_i\}$ results in the **incidental parameter problem**: $\hat{\beta}$ is inconsistent in short panels
  - The problem doesn't arise with linear F because the within transformation kills α_i
- For logistic regression (but not probit), "conditional logit" estimator eliminates α_i yielding consistent estimates for β:
  - But not for average partial effects which depend on α_i
- More progress with long panels; see Fernández-Val and Weidner (2018)

FEs beyond panel data

There are other types of data with repeated observations:

1. Twin studies = repeated observations in the same family
   - E.g. Ashenfelter and Rouse (1998) estimate returns to schooling for twins
     - X_i = years of schooling
     - Family FEs control for genetic differences
   - *Warning*: why does education vary between twins?
     - Need to explain why confounders are the same between the twins while X_{it} is not
     - Bound and Solon (1999): first-borns have higher weight, IQ, schooling

Khoa Vu
@KhoaVuUmn

Your honor, those are my emotional support fixed effects.

FEs beyond panel data (2)

2. County-level cross-section = repeated observations for the same state
   - State FEs control for additive state-level unobservables
   - E.g. state-level policies with constant effects
3. In a county-level panel, can include state-by-year FEs:

$$Y_{it} = \beta' X_{it} + \alpha_i + \gamma_{s(i),t} + \varepsilon_{it}$$

   - Including $\gamma_{s(i),t}$ removes state-year-specific means from Y_{it} and X_{it}
   - *Note*: $\sum_{s',t'} \gamma_{s't'} \mathbf{1}\left[s(i) = s'\right] \times \mathbf{1}\left[t = t'\right]$ is correct notation; $\gamma_{s(i)} \times \delta_t$ is wrong

FEs beyond panel data (3)

4. Repeated cross-sections: in each year a new random sample from each state
   - Can't control for individual heterogeneity
   - But state FEs control for additive state-level unobservables
   - Cluster at the state-level if X_{it} only varies by state
5. Dyadic data: e.g. how does distance X_{ij} between exporting country i and importing country j affect log trade flow Y_{ij}? (Gravity equation)
   - *Note:* by construction, it's like a long panel, $I \times I$
   - *Exercise:* Which FE would you include? How would you cluster SE?

Outline

Application: Are there good firms?

- Abowd, Kramarz, and Margolis (1999): Are there good firms that pay higher wages to the same workers?
    - How much variation in wages is explained by worker characteristics? By firm characteristics?
    - Do "better" workers tend to work for "better" firms?
- Use 10 years of employer-employee matched data for France
    - A panel of workers i: observe employer ID $\boldsymbol{j}(it)$, experience, wages

AKM model

Model of log-wages Y_{it}:

$$Y_{it} = \beta' X_{it} + \alpha_i + \gamma_{\boldsymbol{j}(it)} + \varepsilon_{it}$$

- X_{it} are time-varying observables, e.g. experience — **not** of interest
- $\{\alpha_i\}$ are worker FEs; $\{\gamma_j\}$ are firm FEs
- ε_{it} captures match-specific wage premium

Identification

- Relative firm FEs are identified by **movers**:

$$\Delta Y_{it} = \beta' \Delta X_{it} + \left(\gamma_{\boldsymbol{j}(it)} - \gamma_{\boldsymbol{j}(i,t-1)}\right) + \Delta \varepsilon_{it}$$

  - (actual estimation via dummy variable regression, not in first-differences)
  - Identification only within a connected component of the worker-firm graph

- Requires strict exogeneity: $\mathbb{E}\left[\varepsilon_{it} \mid X_{i1}, \dots, X_{iT}, \alpha_i, \boldsymbol{j}(i1), \dots, \boldsymbol{j}(iT)\right] = 0$
  - Matching of firms and workers can depend on FEs but not on match quality ε_{it} (**"exogenous mobility"**)

Testing exogenous mobility (Card, Heining, and Kline (2013))

Do movers from high-γ to low-γ firms lose *less* than movers from low-γ to high-γ gain?

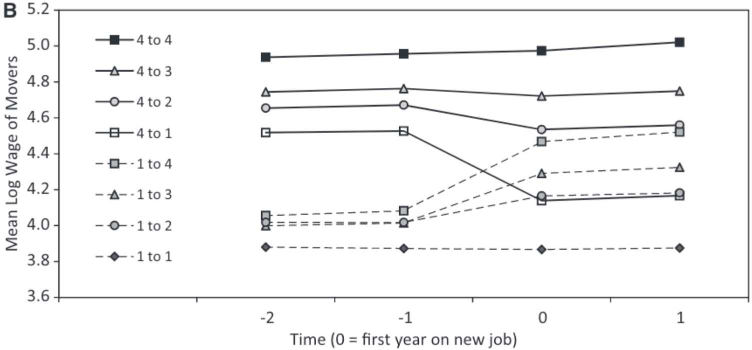

FIGURE V

(use quartiles of average wages paid to other workers; German employer-employee data)

Estimation issues

- Card et al. (2013) compute:
  - Variances of $\hat{\alpha}_i$ and $\hat{\gamma}_{\boldsymbol{j}(it)}$ as worker and firm contributions to wage inequality
  - Covariance of $\hat{\alpha}_i$ and $\hat{\gamma}_{\boldsymbol{j}(it)}$ as a measure of sorting
- But FEs are not estimated consistently!
  - For all workers, at most 8 wage observations $\Rightarrow \mathrm{Var}\left[\hat{\alpha}_i\right]$ biased $\uparrow$
  - For many firms, there are only a few movers $\Rightarrow \mathrm{Var}\left[\hat{\gamma}_{\boldsymbol{j}(it)}\right]$ biased $\uparrow$ and $\mathrm{Cov}\left[\hat{\alpha}_i, \hat{\gamma}_{\boldsymbol{j}(it)}\right]$ biased $\downarrow$
- Kline, Saggio, and Solvsten (2020) provide a bias correction
  - Consistent estimates of $\mathrm{Var}\left[\alpha_i\right], \mathrm{Var}\left[\gamma_{\boldsymbol{j}(it)}\right], \mathrm{Cov}\left[\alpha_i, \gamma_{\boldsymbol{j}(it)}\right]$ without consistent estimates of the FEs

Kline et al. (2020) findings (for Veneto region in Italy)

TABLE II

VARIANCE DECOMPOSITION[a]

	Pooled	Younger Workers	Older Workers
Variance of Firm Effects			
Plug in (PI)	$\boxed{0.0358}$	0.0368	0.0415
Homoscedasticity Only (HO)	0.0295	0.0270	0.0350
Leave Out (KSS)	$\boxed{0.0240}$	0.0218	0.0204
Variance of Person Effects			
Plug in (PI)	$\boxed{0.1321}$	0.0843	0.2180
Homoscedasticity Only (HO)	0.1173	0.0647	0.2046
Leave Out (KSS)	$\boxed{0.1119}$	0.0596	0.1910
Covariance of Firm, Person Effects			
Plug in (PI)	0.0039	−0.0058	−0.0032
Homoscedasticity Only (HO)	0.0097	0.0030	0.0040
Leave Out (KSS)	0.0147	0.0075	0.0171
Correlation of Firm, Person Effects			
Plug in (PI)	$\boxed{0.0565}$	−0.1040	−0.0334
Homoscedasticity Only (HO)	0.1649	0.0726	0.0475
Leave Out (KSS)	$\boxed{0.2830}$	0.2092	0.2744
Coefficient of Determination (R^2)			
Plug in (PI)	$\boxed{0.9546}$	0.9183	0.9774
Homoscedasticity Only (HO)	0.9029	0.8184	0.9524
Leave Out (KSS)	$\boxed{0.8976}$	0.8091	0.9489

Extensions

- See Kline (2024) for a modern treatment
- Hull (2018) interprets mover designs under heterogeneous effects
- Versions of the AKM model have been applied in other settings:
  - Wages depend on worker FE and city FE (Glaeser and Mare (2001))
  - Log healthcare utilization depends on person FE ("demand") and location FE ("supply") (Finkelstein, Gentzkow, and Williams (2016))
  - Changes in credit depends on client firm FE (demand) and bank FE (supply) (Amiti and Weinstein (2018), in a cross-section)

References I

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High wage workers and high wage firms," *Econometrica*, 67, 251–333.

Amiti, M. and D. E. Weinstein (2018): "How Much Do Idiosyncratic Bank Shocks Affect Investment? Evidence from Matched Bank-Firm Loan Data," *Journal of Political Economy*, 126, 525–587.

Arellano, M. and S. Bond (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *The Review of Economic Studies*, 58, 277.

Ashenfelter, O. and C. Rouse (1998): "Income, Schoolig, and Ability: Evidence from a New Sample of Identical Twins," *Quarterly Journal of Economics*, 113, 253–284.

Bertrand, M., E. Duflo, and S. Mullainathan (2004): "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics*, 119, 249–275.

Blundell, R. and S. Bond (1998): "Initial conditions and moment restrictions in dynamic panel data models," *Journal of Econometrics*, 87, 115–143.

Bound, J. and G. Solon (1999): "Double trouble: on the value of twins-based estimation of the return to schooling," *Economics of Education Review*, 18, 169–182.

References II

CARD, D., J. HEINING, AND P. KLINE (2013): "Workplace Heterogeneity and the Rise of West German Wage Inequality," *Quarterly Journal of Economics*, 128, 967–1015.

FERNÁNDEZ-VAL, I. AND M. WEIDNER (2018): "Fixed Effects Estimation of Large- T Panel Data Models," *Annual Review of Economics*, 10, 109–138.

FINKELSTEIN, A., M. GENTZKOW, AND H. WILLIAMS (2016): "Sources of Geographic Variation in Health Care: Evidence From Patient Migration," *The Quarterly Journal of Economics*, 131, 1681–1726.

GLAESER, E. L. AND D. C. MARE (2001): "Cities and Skills," *Journal of Labor Economics*, 19, 316–342.

HULL, P. (2018): "Estimating Treatment Effects in Mover Designs," *Working Paper*.

KLINE, P. (2024): "Firm wage effects," in *Handbook of Labor Economics*, Elsevier, vol. 5, 115–181.

KLINE, P., R. SAGGIO, AND M. SOLVSTEN (2020): "Leave-Out Estimation of Variance Components," *Econometrica*, 88, 1859–1898.