Empirical Methods for Policy Evaluation

Matteo Bobba

Toulouse School of Economics (TSE)

TSE PhD Program (MRes)
Fall 2025

Difference-in-Differences and Event Studies (5 Classes)

- Two-way fixed effect regressions
  - DID $\neq$ TWFE
  - Event Study regressions
  - Heterogeneity-robust DID estimators

- Application
  - A quick recap on job search models

  $\Rightarrow$ Informality, job search, and schooling investments (Bobba et al, IER 2022)

DID $\neq$ TWFE

Groups and Time Periods

- We consider observations that can be divided into G groups and T periods
- For every $(g, t) \in \{1, ..., G\} \times \{1, ..., T\}$: $=$ nb of obs in group g at period t
  - $\Rightarrow$ For simplicity, we assume hereafter balanced panel of groups:

$$\text{For all } (g, t) \in \{1, ..., G\} \times \{1, ..., T\}, N_{g,t} > 0$$

- Panel/repeated cross-section data set where groups are, e.g.
  - $\Rightarrow$ Individuals, firms, counties, etc.
  - $\Rightarrow$ Cross-section dataset where cohort of birth plays the role of time
  - $\Rightarrow$ One may have $N_{g,t} = 1$, e.g. b/c group=individual or a firm

Treatment and Design

- $D_{g,t} \in R^+$: treatment of group g and at period t

$\Rightarrow$ Staggered adoption design: $D_{g,t}$ increases only once, constant otherwise

- In some cases the treatment may vary across individuals within a group
  - Fuzzy DID, not considered here

$\Rightarrow$ We assume that $D_{g,t}$ is constant within g

Potential Outcomes, SUTVA, and Covariates

- Let $(d_1, ..., d_T)$ denote a treatment trajectory
- Corresponding potential outcomes: $Y_{g,t}(d_1, ..., d_T)$
- Then observed outcome: $Y_{g,t} = Y_{g,t}(D_{g,1}, ..., D_{g,T})$

$\Rightarrow$ We maintain the usual SUTVA:

$$(Y_{g,1}(d_1, ..., d_T), ..., Y_{g,T}(d_1, ..., d_T)) \amalg (D_{g',t'})_{g' \neq g, t'=1,...,T}, \forall (g, t, d_1, ..., d_T)$$

- For any variable $X_{g,t}$, let $\boldsymbol{X}_g = (X_{g,1}, ..., X_{g,T})$ and $\boldsymbol{X} = (\boldsymbol{X}_1, ..., \boldsymbol{X}_G)$.

The Pervasiveness of Two-way Fixed Effect Regressions

$\Rightarrow$ Researchers often consider two-way fixed effects (TWFE) models of the kind:

$$Y_{g,t} = \alpha_g + \gamma_t + \beta_{fe} D_{g,t} + \epsilon_{g,t}.$$

- 26 out of the 100 most cited 2015-2019 AER papers estimate TWFE
- Also commonly used in other social sciences
- Other popular method: event-study regressions=dynamic version of TWFE

TWFE Parameters

- $\widehat{\beta}_{fe}$= OLS coeff. of $D_{g,t}$ in a reg. of $Y_{g,t}$ on group FEs, time FEs and $D_{g,t}$
- $\widehat{\beta}_{fd}$= OLS coeff. of $D_{g,t} - D_{g,t-1}$ in a reg. of $Y_{g,t} - Y_{g,t-1}$ on time FEs and $D_{g,t} - D_{g,t-1}$
- $\widehat{\beta}^{X}_{fe}$, when include covariates $X_{g,t}$ in the regression

$\Rightarrow$ We first focus on β_{fe}, but we will extend the results to β_{fd} and β^{X}_{fe}

The Simplest Setup, 2x2 Case

- Two groups $g \in \{s, n\}$ and two time periods $t \in \{1, 2\}$
- $D_{g,t} \in \{0, 1\}$, such that $D_{s,1} = D_{n,1} = 0$, $D_{s,2} = 1$, and $D_{n,2} = 0$
- $Y_{g,t}$ is the observed outcome in location g at period t
- $Y_{g,t}(0), Y_{g,t}(1)$ are potential outcomes without and with treatment

The Parallel (//) Trend Assumption

- In the absence of treatment, same average outcome evolution across groups

$$\mathbb{E}[Y_{s,2}(0) - Y_{s,1}(0)] = \mathbb{E}[Y_{n,2}(0) - Y_{n,1}(0)]$$

- Weaker than imposing that s and n have same untreated-outcome levels

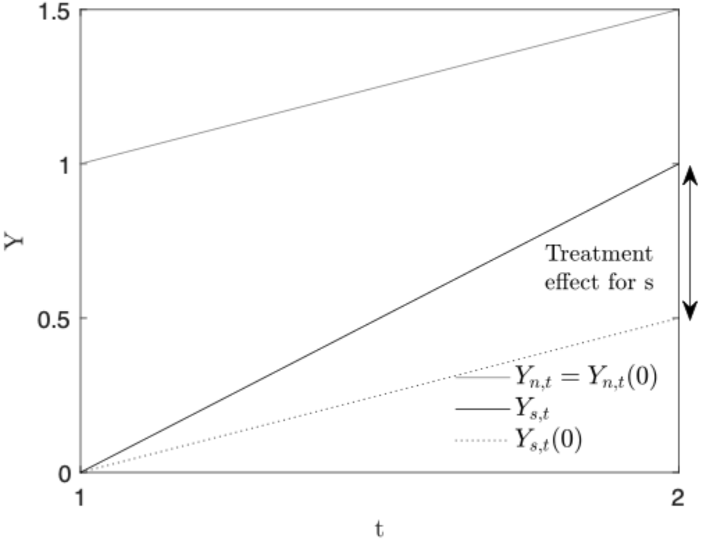
$$\mathbb{E}[Y_{s,t}(0)] = \mathbb{E}[Y_{n,t}(0)] \text{ for all } t$$

- Also weaker than imposing no variation in average untreated outcomes

$$\mathbb{E}[Y_{g,2}(0)] = \mathbb{E}[Y_{g,1}(0)] \text{ for all } g$$

$\Rightarrow$ Appeal of // trends: has testable implications (no pre-trends)

Illustration of Parallel-trends



In the Simplest Set-up, TWFE = DID

- Under // trends, DID is unbiased for the ATE in location s at period 2

$$\begin{aligned}\mathbb{E}(DID) &= \mathbb{E}[Y_{s,2} - Y_{s,1} - (Y_{n,2} - Y_{n,1})] \\ &= \mathbb{E}[Y_{s,2}(1) - Y_{s,1}(0) - (Y_{n,2}(0) - Y_{n,1}(0))] \\ &= \mathbb{E}[Y_{s,2}(1) - Y_{s,2}(0)] + \mathbb{E}[Y_{s,2}(0) - Y_{s,1}(0)] - \mathbb{E}[Y_{n,2}(0) - Y_{n,1}(0)] \\ &= \mathbb{E}[Y_{s,2}(1) - Y_{s,2}(0)]\end{aligned}$$

- This result extends beyond the 2x2 case as long as

$$D_{g,t} = 1\{t > T_0\}D_g, \text{ with } T_0 \geq 1, D_g \in \{0,1\}$$

$\Rightarrow$ The before-after diff is combined with the treated-control diff

Additive Separability of TWFE

- Static case

$$D_{g,t} \in R^{+} \text{and for all } (g, t, d_1, ..., d_T), Y_{g,t}(d_1, ..., d_T) = Y_{g,t}(d_t)$$

- Parallel trends: for all $t \geq 2$, $E[Y_{g,t}(0) - Y_{g,t-1}(0)] = \gamma_t$

- It follows that: $E[Y_{g,t}(0) - Y_{g,1}(0)] = \gamma_t$, and let $\alpha_g = E[Y_{g,1}(0)]$

$\Rightarrow$ Then, $E[Y_{g,t}(0)] = E[Y_{g,1}(0)] + E[Y_{g,t}(0) - Y_{g,1}(0)] = \alpha_g + \gamma_t$

Parameter of Interest

- Average treatment response

$$\Delta^{TR} = \frac{1}{\sum_{g,t} D_{g,t}} \sum_{g,t} (Y_{g,t}(D_{g,t}) - Y_{g,t}(0))$$

- Then, let $\delta^{TR} = E[\Delta^{TR}]$. With a binary D, δ^{TR} =ATT

- Analogously, in (g,t):

$$\Delta_{g,t} = \frac{1}{D_{g,t}} [Y_{g,t}(D_{g,t}) - Y_{g,t}(0)] \text{ if } D_{g,t} \neq 0$$

- Then:

$$\delta^{TR} = E\left[\sum_{(g,t):D_{g,t}>0} W_{g,t}\Delta_{g,t}\right], \quad \text{with } W_{g,t} = \frac{D_{g,t}}{\sum_{(g,t):D_{g,t}>0} D_{g,t}}$$

β_{fe} = weighted sum of ATEs under // trends

- Let $\epsilon_{fe,g,t}$= resid. of the reg. of $D_{g,t}$ on a constant, group FEs, and time FEs
- It can be shown that:

$$\beta_{fe} = E\left[\sum_{(g,t):D_{g,t}>0} W_{fe,g,t}\Delta_{g,t}\right]$$

- $W_{fe,g,t} = \frac{D_{g,t}\epsilon_{fe,g,t}}{\sum_{(g,t):D_{g,t}>0} D_{g,t}\epsilon_{fe,g,t}}$
- In general, $\beta_{fe} \neq \delta^{TR}$ because $W_{fe,g,t} \neq W_{g,t}$

$\Rightarrow$ We may have $W_{fe,g,t} < 0$: if $\epsilon_{fe,g,t} < 0$ while $D_{g,t} > 0$

$\Rightarrow$ $E\left[\widehat{\beta}_{fe}\right]$ may be < 0 even if $Y_{g,t}(d) > Y_{g,t}(0)$ for all (g,t) and $d > 0$

Characterizing (g, t) cells weighted negatively by β_{fe}

- Let $D_{g,.}$ =average treat. rate of g and $D_{.,t}$ =average treat. rate at t
- Under // trends, $W_{fe,g,t}$ is decreasing with $D_{g,.}$ and $D_{.,t}$
- β_{fe} more likely to assign negative weight to:
  - $\Rightarrow$ Periods with many vs few treated groups
  - $\Rightarrow$ Groups treated for many vs few periods
- In staggered adoption designs $(D_{g,t} \geq D_{g,t-1})$, $W_{fe,g,t} < 0$ more likely
  - In the last periods and for groups adopting the treatment earlier
  - $\Rightarrow$ Remove negative weights by removing always treated groups

Forbidden Comparison 1: Switchers Vs. Always Treated

- D binary and design staggered: $\widehat{\beta}_{fe} =$ weighted avg of two types of DIDs:
  - DID_1: comparing s from untreated to treated to n untreated at both dates
  - DID_2: comparing switching group s to group a treated at both dates.

$\Rightarrow$ Negative weights in β_{fe} originate from the second type of DIDs

Forbidden Comparison 1: An Example

- Group e treated at $t=2$, group ℓ treated at $t=3$. Then:

$$\widehat{\beta}_{fe} = \frac{1}{2} \times \underbrace{DID^{1-2}_{e-\ell}}_{DID_1} + \frac{1}{2} \times \underbrace{DID^{2-3}_{\ell-e}}_{DID_2}$$

- At periods 2 and 3, e's outcome $=$ treated potential outcome, so

$$Y_{e,3} - Y_{e,2} = Y_{e,3}(1) - Y_{e,2}(1) = Y_{e,3}(0) + \Delta_{e,3} - (Y_{e,2}(0) + \Delta_{e,2}).$$

- On the other hand, group ℓ only treated at period 3, so

$$Y_{\ell,3} - Y_{\ell,2} = Y_{\ell,3}(0) + \Delta_{\ell,3} - Y_{\ell,2}(0)$$

Forbidden Comparison 1: An Example (continued)

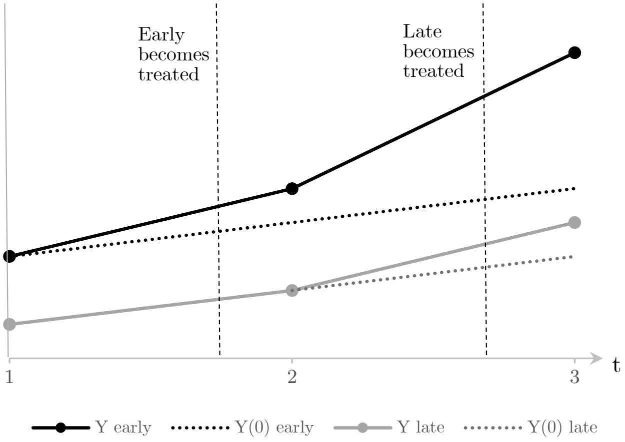
$\Rightarrow$ $E\left[DID_{\ell-e}^{2-3}\right] = E\left[Y_{\ell,3} - Y_{\ell,2} - (Y_{e,3} - Y_{e,2})\right] = E\left[\Delta_{\ell,3} + \Delta_{e,2} - \Delta_{e,3}\right]$

- Note: if $\Delta_{e,2} = \Delta_{e,3}$, $E[DID_{\ell-e}^{2-3}] = E\left[\Delta_{\ell,3}\right]$
- More generally, if $\Delta_{g,t} = \Delta_{g,t'}$, $W_{fe,g,t} \geq 0$. But restrictive!
- Note:

$$Y_{g,t}(0) - Y_{g,t-1}(0) = Y_{g,t}(1) - Y_{g,t-1}(1) \Longleftrightarrow \Delta_{g,t} = \Delta_{g,t-1}$$

- This assumption is actually equivalent to time-invariant treat. effects

Forbidden Comparison 1: Graphical Illustration



Forbidden Comparison 2: Switching more vs Switching less

- Suppose the treatment D is not binary
- $\widehat{\beta}_{fe}$= (group m whose D increases more) - (group ℓ whose D increases less)
- In fact, with two groups m and ℓ and two periods,

$$\widehat{\beta}_{fe} = \frac{Y_{m,2} - Y_{m,1} - (Y_{\ell,2} - Y_{\ell,1})}{D_{m,2} - D_{m,1} - (D_{\ell,2} - D_{\ell,1})}$$

$\Rightarrow$ This "Wald-DID" estimator may not estimate convex combination of effects

Forbidden Comparison 2: An Example

- Assume m goes from 0 to 2 units of treatment while ℓ goes from 0 to 1

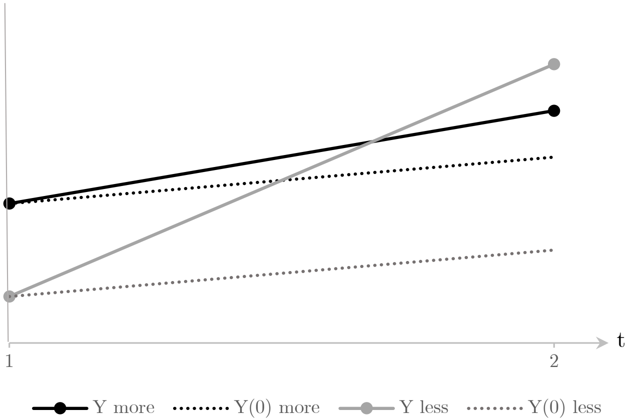
$\Rightarrow$ Denominator of the Wald-DID is $2-0-(1-0)=1$

- Potential outcomes: $Y_{m,t}(d)=Y_{m,t}(0)+\delta_m d$ and $Y_{\ell,t}(d)=Y_{m,t}(0)+\delta_\ell d$

- Then:

$$
\begin{aligned}
E\left[\widehat{\beta}_{fe}\right] =& E\left[Y_{m,2}-Y_{m,1}-(Y_{\ell,2}-Y_{\ell,1})\right] \\
=& E\left[Y_{m,2}(0)+2\delta_m-Y_{m,1}(0)-(Y_{\ell,2}(0)+\delta_\ell-Y_{\ell,1}(0))\right] \\
=& E\left[Y_{m,2}(0)-Y_{m,1}(0)\right]-E\left[Y_{\ell,2}(0)-Y_{\ell,1}(0)\right]+2\delta_m-\delta_\ell \\
=& 2\delta_m-\delta_l
\end{aligned}
$$

Forbidden Comparison 2: Graphical Illustration



Extensions

- This logic extends to β_{fd}, but with different weights $W_{fd,g,t}$

$\Rightarrow$ If $\beta_{fd} \neq \beta_{fe}$, we reject homogeneous TE under // trends

- With covariates, we modify the // trends by assuming that for some λ,

$$\begin{aligned} & E[Y_{g,t}(0) - Y_{g,t-1}(0) - (X_{g,t} - X_{g,t-1})'\lambda | \boldsymbol{D}_g, \boldsymbol{X}_g] \\ = & E[Y_{g,t}(0) - Y_{g,t-1}(0) - (X_{g,t} - X_{g,t-1})'\lambda], \end{aligned}$$

- Let $\epsilon^X_{fe,g,t}$= resid. of reg. of $D_{g,t}$ on a const., group FEs, time FEs and $X_{g,t}$

- Then, same result as above but with $\epsilon^X_{fe,g,t}$ instead of $\epsilon_{fe,g,t}$ in $W_{fe,g,t}$.

Software Implementations

- `bacondecomp` computes DIDs and their corresponding weights entering in $\widehat{\beta}_{fe}$
- `twowayfeweights` computes the weights $W_{fe,g,t}$ and $W_{fd,g,t}$, with/o X
  - $\Rightarrow$ Worst-case scenario of std dev on $\Delta_{g,t}$ (weights are max corr. with TE)
  - $\Rightarrow$ Correlation between weights and proxies of $\Delta_{g,t}$

Example: Effect of Newspapers on Electoral Turnout

- Gentzkow et al. (AER, 2011) use US county data on presidential elections
- Reg change in turnout in g on change in $\#$ newspapers $+$ state-year FE
- One could also estimate the FE regression

Regression	$\widehat{\beta}$ (s.e.)	% of < 0 weights	Sum of < 0 weights
$\widehat{\beta}_{fe}$	-0.0011 (0.0011)	40.1%	-0.53
$\widehat{\beta}_{fd}$	0.0026 (0.0009)	45.7%	-1.43

$\Rightarrow$ Under // trends, we reject the null hypothesis that $\Delta_{g,t} = \Delta \ \forall (g,t)$

Event Study Regressions

Main Assumptions

- Assuming past treatments do not affect current outcomes is restrictive
- We now generalize the previous set-up as follows ($\mathbf{0}_t$ =vector of t zeros)
- Univariate, dynamic framework w/o anticipation, $D_{g,t} \in R^+$:

$$Y_{g,t}(d_1, ..., d_T) = Y_{g,t}(d_1, ..., d_t)$$

- // trends, v2
  - $\Rightarrow$ $\forall g$ and $t \geq 2$, $E[Y_{g,t}(\mathbf{0}_t) - Y_{g,t-1}(\mathbf{0}_{t-1})|\boldsymbol{D}_g] = E[Y_{g,t}(\mathbf{0}_t) - Y_{g,t-1}(\mathbf{0}_{t-1})]$
  - $\Rightarrow$ For all $t \geq 2$, $E[Y_{g,t}(\mathbf{0}_t) - Y_{g,t-1}(\mathbf{0}_{t-1})] = \gamma_t$

Staggered Design: Notation and Parameters

- Staggered treatment: $D_{g,t} = 1\{t \geq F_g\}, F_g =$period when g becomes treated
- Exclude always treated here and let $F_g = T + 1$ if never treated
- Let $\mathbf{1}_t$ be a vector of t ones and define:

$$\begin{aligned}\Delta_g(\ell) =& Y_{g,F_g+\ell}(\mathbf{0}_{F_g-1}, \mathbf{1}_{\ell+1}) - Y_{g,F_g+\ell}(\mathbf{0}_{F_g+\ell}),\\ \Delta(\ell) =& \frac{1}{N_{\leq}(T-\ell)} \sum_{g:F_g+\ell\leq T} \Delta_g(\ell)\end{aligned}$$

where $N_{\leq}(t) = \text{card}\{g : F_g \leq t\}$

$\Rightarrow$ $\Delta_g(\ell)$: effect for g of having received the treatment for $\ell + 1$ periods

$\Rightarrow$ $\Delta(\ell)$: effect of having received the treatment for $\ell + 1$ periods

Staggered design: ES Regressions

- Regression of $Y_{g,t}$ on group and time FEs, and $(1\{F_g = t - \ell\})_{\ell=-K,\dots,L,l\neq-1}$
- When $t - \ell > T$, we set $1\{F_g = t - \ell\} = 0$

$\Rightarrow$ For $\ell \geq 0$, β_ℓ is supposed to estimate $\Delta(\ell)$

$\Rightarrow$ For $\ell \leq -2$, $\beta_\ell =$ placebo used to check // trends

An Aside: "Basic" Problems with ES Regressions

- Let us consider the fully-dynamic specification where

$$K = \overline{K} := \max_{g: F_g \leq T} F_g - 1, \quad L = \overline{L} := T - \min_g F_g$$

- If all groups are eventually treated, then:

- $(\beta_\ell)_{\ell=-\overline{K},\dots,\overline{L},\ell\neq -1}$ not sep. identified from $(\beta_\ell + \kappa(\ell+1))_{\ell=-\overline{K},\dots,\overline{L},\ell\neq -1}$

$\Rightarrow$ Proof: $\sum_{\substack{\ell=-\overline{K} \\ \ell\neq -1}}^{\overline{L}} 1\{F_g = t-\ell\}(\ell+1) = \sum_{\ell=-\overline{K}}^{\overline{L}} 1\{F_g = t-\ell\}(\ell+1)$

$$= (t+1-F_g)\sum_{\ell=-\overline{K}}^{\overline{L}} 1\{F_g = t-\ell\}$$
$$= \underbrace{t+1}_{\text{enter in time FE}} \quad \underbrace{-F_g}_{\text{enter in group FE}}$$

An Aside: "Basic" Problems with ES Regressions

$\Rightarrow$ The fully-dynamic specification requires never treated groups

- Otw some placebo coeffs should be removed, or dyn. effects be restricted
- In the latter case, a common practice is to choose $L < \overline{L}$
- But even if TE are homogenous, this makes sense only if

$$Y_{g,F_g+\ell}(\mathbf{0}_{F_g-1}, \mathbf{1}_{\ell+1}) = Y_{g,F_g+\ell}(\mathbf{0}_{F_g}) \quad \text{for } \ell > L.$$

- More natural to assume that TE stabilize:

$$\Delta_g(\ell) = \Delta_g(L)(= \Delta(L)), \;\; \text{for } \ell > L$$

Decomposition Result for ES Regressions

- $\epsilon_{g,t}$: res. of $1\{F_g = t - \ell\}$ on group/time FEs and $(1\{F_g = t - \ell'\})_{\ell'=-K,..,\overline{L}}$
- Define $W_{g,\ell,\ell'} = \frac{\epsilon_{g,F_g+\ell'}}{\sum_g \epsilon_{g,F_g+\ell}}, \forall(\ell,\ell')$
- Suppose that $D_{g,t}$ is binary and design is staggered. Then, for $\ell = 0, ..., \overline{L}$,

$$E\left[\widehat{\beta}_\ell\right] = E\left[\underbrace{\sum_g W_{g,\ell,\ell}\Delta_g(\ell)}_{\text{sum of effect of } \ell+1 \text{ treat. periods}} + \underbrace{\sum_{\ell'\neq\ell}\sum_g W_{g,\ell,\ell'}\Delta_g(\ell')}_{\text{sum of effects of } \ell'+1 \text{ treat periods}}\right]$$

- $\sum_g W_{g,\ell,\ell} = 1$ and $\sum_g W_{g,\ell,\ell'} = 0 \; \forall \ell, \; \ell' \neq \ell$

ES Even Less Robust to Heterog. TE than Static TWFE

- As in the static case, $W_{g,l,l}$ may be < 0

$\Rightarrow$ $\widehat{\beta}_l$ contaminated by effects of $l' + 1$ treatment periods

$\Rightarrow$ $\Delta_g(l) = \Delta_{g'}(l)\ \forall g, g'$ not sufficient for $\widehat{\beta}_l$ to be unbiased. We also need

$$\Delta_g(l') = \Delta_{g'}(l')\ \forall l' \neq l$$

Origin: Again, Forbidden Comparisons

- As in the static case, $\widehat{\beta}_\ell$ can still be written as a linear combination of DID
- But in some of these DIDs:
  - The control group has been treated for $l' \neq \ell$ period at baseline/endline
  - If $\ell > 0$, the treated group has been treated for $\ell' > 0$ period at the baseline
  - The treated group has been treated for $\ell' < \ell$ period at the endline

$\Rightarrow$ Contamination bias

A Simple Example

- Consider a design with $G = 2$, $T = 3$, $K = 0$
  - Group 1 treated at $t = 2$
  - Group 2 treated at $t = 3$

- Then, some algebra shows that:

$$\widehat{\beta}_1 = \underbrace{Y_{1,3} - Y_{1,2} - (Y_{2,3} - Y_{2,2})}_{DID_1} + 2[\underbrace{Y_{1,2} - Y_{1,1} - (Y_{2,2} - Y_{2,1})}_{DID_2}]$$

A Simple Example (cont'd)

- In DID_1, control group $g = 2$ treated for 1 period at $t = 3$ and $g = 1$ also treated for 1 period at $t = 2$. Then,

$$E[DID_1] = E[\Delta_1(1) - (\Delta_1(0) + \Delta_2(0))].$$

- DID_2 identifies the effect of having been treated for 1 rather than 2 periods:

$$E[DID_2] = E[\Delta_1(0)]$$

$\Rightarrow$ We obtain $E[\widehat{\beta}_1] = E[\Delta_1(1) + (\Delta_1(0) - \Delta_2(0))]$

Software Implementations

- `eventstudyweights` Stata package computes weights in decomposition:

  ```
  eventstudyweights {rel_time_list}, absorb(i.groupid i.timeid)
  cohort(first_treatment) rel_time(ry),
  ```

  with `rel_time_list`=list of the $1\{F_g = t - l\}$, `first_treatment`=F_g (missing if never treated), and `ry`=`timeid - first_treatment`

- Can easily include covariates

Heterogeneity-robust DID estimators

Robust DIDs (Static Case)

- Avoid making the forbidden comparisons leveraged by TWFE:
  - $\Rightarrow$ Never compare switcher to switcher: only compare switcher to stayer
  - $\Rightarrow$ Never compare a switcher to a stayer with a different baseline treatment
- Depends on whether groups' outcome at t only depends on treatment at t
- If yes, we can consider each pair of consecutive time periods independently
  - $\Rightarrow$ $t-1$ to t switchers: groups whose treatment changes from $t-1$ to t
  - $\Rightarrow$ $t-1$ to t stayers: groups whose treatment does not change from $t-1$ to t, with same $t-1$ treatment as switchers

Robust DIDs (Dynamic Case)

- in more general case, we need to control for groups' full treatment history
  - $\Rightarrow$ $t-1$ to t first-time switchers: treat changes for the first time from $t-1$ to t
  - $\Rightarrow$ 1 to $t+\ell$ stayers: treat does not change from 1 to $t+\ell$, with same $t-1$ treatment as switchers
- Allowing for dynamic effects may lead to less precise and interpretable effects

Parameters of interest

- Suppose first that D is binary
- Let us define

$$\mathcal{S} = \{(g,t) : t \geq 2,\ D_{g,t} \neq D_{g,t-1},\ \exists g' :\ D_{g',t} = D_{g',t-1} = D_{g,t-1}\}$$

- N_S =card($\mathcal{S}$)

$\Rightarrow$ Then, ATE for matchable switchers is

$$\delta^S = E\left[\frac{1}{N_S}\sum_{(g,t)\in\mathcal{S}} Y_{g,t}(1) - Y_{g,t}(0)\right]$$

Assumptions for identifying δ^S

- δ^S can be unbiasedly estimated under the following // trends conditions

  $\Rightarrow$ $E[Y_{g,t}(0) - Y_{g,t-1}(0)|\boldsymbol{D}_g] = E[Y_{g,t}(0) - Y_{g,t-1}(0)] = \gamma_{0,t}$

  $\Rightarrow$ $E[Y_{g,t}(1) - Y_{g,t-1}(1)|\boldsymbol{D}_g] = E[Y_{g,t}(1) - Y_{g,t-1}(1)] = \gamma_{1,t}$

- Usual // trends on $Y_{g,t}(0)$ sufficient if we focus on switchers in

$$\mathcal{S}_+ = \{(g,t) : t \geq 2,\ D_{g,t} = 1 > D_{g,t-1} = 0,\ \exists g' : D_{g',t} = D_{g',t-1} = 0\}$$

- Weaker exogeneity assumption sufficient to consistently estimate δ^S:

$$E[Y_{g,t}(0) - Y_{g,t-1}(0)|D_{g,1}, ..., D_{g,t}] = E[Y_{g,t}(0) - Y_{g,t-1}(0)]$$

- Allows for possibility that $Y_{g,t}(0) - Y_{g,t-1}(0)$ affects $D_{g,t+1}$ etc.

Weighted averages of DIDs identify δ^S

- For all $t \in \{1, ..., T\}$ and $d = 0, 1$, let
  - $N_{+,t} = \text{card}\ \{g : D_{g,t} > D_{g,t-1}\}$ and $N_{-,t} = \text{card}\ \{g : D_{g,t} < D_{g,t-1}\}$
  - $N_{=d,t} = \text{card}\ \{g : D_{g,t} = D_{g,t-1} = d\}$
  - $DID_{+,t} = \sum_{g:D_{g,t}>D_{g,t-1}} \frac{1}{N_{+,t}} (Y_{g,t} - Y_{g,t-1}) - \sum_{g:D_{g,t}=D_{g,t-1}=0} \frac{1}{N_{=0,t}} (Y_{g,t} - Y_{g,t-1})$
  - $DID_{-,t} = \sum_{g:D_{g,t}=D_{g,t-1}=1} \frac{1}{N_{=1,t}} (Y_{g,t} - Y_{g,t-1}) - \sum_{g:D_{g,t}<D_{g,t-1}} \frac{1}{N_{-,t}} (Y_{g,t} - Y_{g,t-1})$
- Then

$$E[DIDM] = E\left[\sum_{t=2}^{T} \frac{N_{+,t}}{N_S} DID_{+,t} + \frac{N_{-,t}}{N_S} DID_{-,t}\right] = \delta^S$$

Intuition for $DIDM$

- $DID_{+,t}$: ΔY between those treated between $t-1$ and t, and untreated

$\Rightarrow$ Under // trends on $Y(0)$, it identifies TE in groups switching into treatment

$\Rightarrow$ Under // trends on $Y(1)$, $DID_{-,t}$ identifies TE for switchers out of treat

- Finally, $DIDM$ is a weighted average of those DID estimands

Placebo Estimators

$\Rightarrow$ Focus on groups that are stayers one period before switchers switch

- Compare switchers' and stayers' ΔY, one period before switchers switch
- Also compare switchers and stayers 2, 3 periods etc. before switchers switch

Controlling for Time-varying Covariates

- Rationale: // trends only hold if we account for change in covariates

$$E(Y_{g,t}(d) - Y_{g,t-1}(d)|\boldsymbol{D}_g, \boldsymbol{X}_g) = \gamma_{d,t} + (X_{g,t} - X_{g,t-1})'\lambda_d \quad \forall d \in \mathcal{D}$$

- Special case is group-specific linear trends $X_{g,t} = (1\{g=2\} \times t, .., 1\{g=G\} \times t)$

- Let $\epsilon_{g,t}(d)$ residual of the reg. of $Y_{g,t} - Y_{g,t-1}$ on period FEs and $X_{g,t} - X_{g,t-1}$ for (g,t) s.t. $D_{g,t} = D_{g,t-1} = d \in \mathcal{D}$

- Define $DIDM^X$ as $DIDM$, but using $\epsilon_{g,t}(D_{g,t-1})$ instead of $Y_{g,t} - Y_{g,t-1}$

- Separate reg. for each $d \in \mathcal{D}$, estimated in sample of d-stayers

Software Implementation

- R and Stata command: `did_multiplegt`
  - Options to relax the standard // trends
  - Control for time-varying, time-invariant covariates, or linear time trends
  - Flexibly specifies the number of placebos to be estimated

Example (continued): Gentzkov et al. (AER, 2011)

Table: Estimates of the effect of one additional newspaper on turnout

	Estimate	Standard error	N
$\widehat{\beta}_{fd}$	0.0026	0.0009	15,627
$\widehat{\beta}_{fe}$	-0.0011	0.0011	16,872
$DIDM$	0.0043	0.0014	16,872
$DIDM$ Placebo	-0.0009	0.0016	13,221

$\Rightarrow$ $DIDM$ is 66% larger (t-stat=1.77) than $\widehat{\beta}_{fd}$ and opposite sign to $\widehat{\beta}_{fe}$

Allowing for dynamic effects: potential outcome notation

- We maintain previous assumption of no anticipation effects

$\Rightarrow$ Potential outcomes $Y_{g,t}(d_1, ..., d_t)$ do not depend on future treatments

- $\mathbf{0}_k$: vector of k zeros, $\mathbf{1}_k$: vector of k ones
- $Y_{g,t}(\mathbf{0}_k, \mathbf{1}_{t-k})$: g at t if untreated from 1 to k and treated from $k+1$ to t

Parameters of interest

- Binary staggered design: $D_{g,t} = 1\{t \geq F_g\}$

$\Rightarrow$ ATE of cohort getting treated from c to $c + l$, i.e. for $l + 1$ periods

$$\delta(c,\ell) = E\left[\frac{1}{N_{=}(c)}\sum_{g:F_g=c}\left(Y_{g,c+\ell}(\mathbf{0}_{c-1},\mathbf{1}_{\ell+1}) - Y_{g,c+\ell}(\mathbf{0}_{c+\ell})\right)\middle|\boldsymbol{D}\right]$$

- Where $N_{=}(c) = \text{card}\{g : F_g = c\}$

- ATE across cohorts $\ell + 1$ periods after they started receiving treat

$$\delta(\ell) = E\left[\frac{1}{\sum_{c=2}^{T-\ell} N_{=}(c)}\sum_{c=2}^{T-\ell} N_{=}(c)\delta(c,\ell)\middle|\boldsymbol{D}\right]$$

DID estimators using never-treated groups

- To estimate $\delta(c,\ell)$

$$\widehat{\delta}_1(c,\ell) = \frac{1}{N_=(c)} \sum_{g:F_g=c} (Y_{g,c+\ell} - Y_{g,c-1}) - \frac{1}{N_=(T+1)} \sum_{g:F_g=T+1} (Y_{g,c+\ell} - Y_{g,c-1})$$

- Recall that $\{g : F_g = T+1\}$ = never-treated groups

$\Rightarrow$ $\widehat{\delta}_1(c,\ell)$: ΔY between $c-1$ and $c+\ell$ in cohort c and in never-treated groups

- Then let

$$\widehat{\delta}_1(\ell) = \frac{1}{\sum_{c=2}^{T-\ell} N_=(c)} \sum_{c=2}^{T-\ell} N_=(c)\widehat{\delta}_1(c,\ell)$$

DID estimators using other control groups

- If no never-treated groups, use instead the last treated/not-yet-treated groups

$$\widehat{\delta}_2(c,\ell) = \frac{1}{N_=(c)} \sum_{g:F_g=c} (Y_{g,c+\ell} - Y_{g,c-1}) - \frac{1}{N_>(c+\ell)} \sum_{g:F_g>c+\ell} (Y_{g,c+\ell} - Y_{g,c-1})$$

- $N_>(c+\ell) = \text{card}\{g : F_g > c+\ell\}$. Define $\widehat{\delta}_2(\ell)$ accordingly

- There are more not-yet-treated groups than never-treated groups

$\Rightarrow$ $\widehat{\delta}_2(c,l)$ possibly more precise than $\widehat{\delta}_1(c,\ell)$

- Also, never-treated groups may be very different from other groups

$\Rightarrow$ // trends may not hold when only never treated used as controls

Software Implementation

- The `csdid` (Stata) and `did` (R) commands. Syntax:

  `csdid outcome, time(timeid) gvar(cohort)`

  where `cohort`$= F_g$ (=0 for never treated)

- The `eventstudyinteract` Stata command. Syntax:

  `eventstudyinteract outcome {rel_time_list}, absorb(i.groupid i.timeid) cohort(first_treatment) control_cohort(controlgroup)`

A Quick Recap on Job Search Models

The Optimal Stopping Model

- Risk neutral individual in discrete time with preferences in $t = 0$ given by

$$\sum_{t=0}^{\infty} \beta^t c_t, \quad \beta \in (0,1)$$

- Start as unemployed, with consumption equal to b
- Jobs sampled sequentially. Each job is for life and identical except for wage
- Wages are drawn from an exogenous stationary distribution $F(w)$

$\Rightarrow$ Given draw $w_t \in W$ agent decides whether to take it or continue searching

Dynamic Programming Formulation

- Value function for the agent when he has sampled a job of $w \in W$ is

$$V(w) = \max\left\{\frac{w}{1-\beta}, \beta V + b\right\}$$

where V is the continuation value of not accepting a job:

$$V = \int_{\omega \in \Omega} V(\omega) dF(\omega)$$

$\Rightarrow$ Combine these two equations and get:

$$V(w) = \max\left\{\frac{w}{1-\beta}, b + \beta \int_{\omega \in \Omega} V(\omega) dF(\omega)\right\}$$

Reservation Wage

- $V(w)$ is non-decreasing, decision rule has a reservation value property
- Reservation wage is given by

$$\frac{w^\star}{1-\beta} = b + \beta \int_{\omega \in \Omega} V(\omega) dF(\omega)$$

$\Rightarrow$ Decision rule: $\forall w < w^\star$, $V(w) = \frac{w^\star}{1-\beta}$ and $\forall w \geq w^\star$, $V(w) = \frac{w}{1-\beta}$

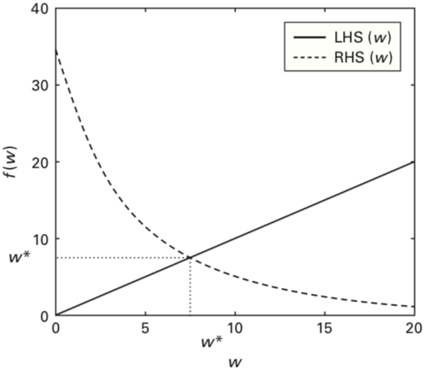
- Therefore, reservation wage can be written as

$$\frac{w^\star}{1-\beta} = b + \beta \left[\frac{w^\star F(w^\star)}{1-\beta} + \int_{w \geq w^\star} \frac{w}{1-\beta} dF(w) \right]$$

Reservation Wage

- Since $\frac{w^\star}{1-\beta} = \int_{w<w^\star} \frac{w^\star}{1-\beta} dF(w) + \int_{w \geq w^\star} \frac{w^\star}{1-\beta} dF(w)$

$$w^\star = b + \frac{\beta}{1-\beta}\left[\int_{w \geq w^\star} (w - w^\star) dF(w)\right]$$



Taking the Model to the Data

- For a random sample of N workers we observe $\{\tilde{t}_u(i), w(i)\}_{i=1}^N$
- Job offers/termination arrive at random times with density between offers

$$\begin{aligned} q_u(t_u) &= \lambda \exp(-\lambda t_u), \lambda > 0 \\ q_e(t_e) &= \eta \exp(-\eta t_e), \eta > 0 \end{aligned}$$

- Reservation wage in continuous time

$$w^\star = b + \frac{\lambda}{\rho + \eta} \int_{w^\star} (w - w^\star) dF(w)$$

$\Rightarrow$ It is easy to show that $\partial w^\star / \partial \eta < 0, \partial w^\star / \partial \rho < 0, \partial w^\star / \partial \lambda > 0$

Simulating Labor Market Data

Hypothetical early labor market career

Event number	State	Time of event	Duration draw	Match value
1	U	0.891	0.891	6.243
2	U	3.168	2.277	4.329
3	U	15.554	12.386	3.871
4	U	15.558	0.004	10.918
5	E	38.921	23.363	—
6	U	44.236	5.315	7.891
7	U	56.793	12.557	12.119
8	E	157.421	100.628	—
9	U	164.772	7.351	10.145
10	E	322.510	157.738	—
⋮	⋮	⋮		⋮

Steady-state Proportions

- The probability that an individual is unemployed is

$$
\begin{aligned}
p(u) &= \frac{\mathbb{E}(t_u)}{\mathbb{E}(t_e) + \mathbb{E}(t_u)} = \frac{[\lambda \tilde{F}(w^\star)]^{-1}}{\eta^{-1} + [\lambda \tilde{F}(w^\star)]^{-1}} \\
&= \frac{\eta}{\eta + \lambda \tilde{F}(w^\star)}
\end{aligned}
$$

- And conversely

$$
p(e) = 1 - p(u) = \frac{\lambda \tilde{F}(w^\star)}{\eta + \lambda \tilde{F}(w^\star)}
$$

Offered and Accepted Wages

- Optimal decision rule $\Rightarrow$ truncation in the accepted wage distribution

$$g(w) = \frac{f(w)}{\tilde{F}(w^\star)}, w \geq w^\star$$

- This density is well defined: integrates to 1 and non-negative for all $w \geq w^\star$

Likelihood Contributions

- The likelihood of an ongoing unemployment spell is

$$L(t_u, u) = f_u(t_u)p(u) = \lambda \tilde{F}(w^\star) \exp[-\lambda \tilde{F}(w^\star) t_u] \times \frac{\eta}{\eta + \lambda \tilde{F}(w^\star)}$$

- The likelihood of employed and earning a wage w is

$$L(w, e) = \frac{f(w)}{\tilde{F}(w^\star)} \times \frac{\lambda \tilde{F}(w^\star)}{\eta + \lambda \tilde{F}(w^\star)} = \frac{\lambda f(w)}{\eta + \lambda \tilde{F}(w^\star)}$$

Likelihood Function

- The likelihood function for a random sample of N individuals is then

$$L(w_1,..,w_{N_e},t_1,..,t_{N_u}) = \prod_{i\in e}\left[\frac{\lambda f(w(i))}{\eta+\lambda\tilde{F}(w^\star)}\right]\times\prod_{i\in u}\left[\frac{\eta\lambda\tilde{F}(w^\star)\exp[-\lambda\tilde{F}(w^\star)t_u(i)]}{\eta+\lambda\tilde{F}(w^\star)}\right]$$

- And the associated log-likelihood is

$$\ln L = -N\ln[\eta+\lambda\tilde{F}(w^\star)] + N\ln\lambda + \sum_{i\in e}\ln[f(w(i))] +$$
$$+ N_u\ln[\tilde{F}(w^\star)] + N_u\ln(\eta) - \lambda\tilde{F}(w^\star)\sum_{i\in u}t_u(i)$$

Identification

- The primitive parameters that explicitly enter in $\ln L$ are λ, η and F
- Parameters b and ρ only enter through $w^{\star}$
- Equilibrium object $w^{\star}$ is part of the support of F

$\Rightarrow$ This feature generates a non-standard likelihood function

Flinn and Heckman (1982)

1. Estimate the reservation wage as the minimum accepted wage

$$\hat{w}^\star = \min(w_1, ..., w_{N_e})$$

$\Rightarrow$ Order statistics are super-consistent (i.e. converge at rate N)

2. Maximize log likelihood with respect to λ, η and μ conditional on $\hat{w}^\star$

$\Rightarrow$ $F(w)$ needs to be recoverable: $F(w|w \geq \hat{w}^\star) = \frac{F(w)-F(w^\star)}{\bar{F}(w^\star)}, \forall w \geq \hat{w}^\star$

3. Plug estimated parameters into equation for $w^\star$ and solve for either b or ρ

$\Rightarrow$ Usually fix ρ and recover the value of b

Informality, job search, and schooling investments

Bobba, Flabbi and Levy (IER, 2022)

- An equilibrium search model where:
  - Search frictions generate mobility between formal and informal jobs
  - Match productivity and bargaining generate overlapping wage distributions

  $\Rightarrow$ Both ingredients generates a mix of formal and informal jobs in equilibrium

- One long-term "cost of informality": Under-investment in education
  - Same features that create informality may also distort returns to schooling

  $\Rightarrow$ Trade-off between welfare in the labor market and pre-market HK

Context: Labor Markets in Latin America

1. More than half of the labor force is in the informal sector
   - Workers not contributing to and not covered by the social security system

   $\Rightarrow$ Informal employees and (most of the) self-employed

2. Neither a segmented or a competitive labor market
   - Individuals transit back and forth between formal and informal jobs
   - Wage/productivity distributions overlap
   - Mix of formality status within the same firm

3. Informal workers gained access to non-contributory social programs

The Model Environment

- Timing
  1. Schooling decision
  2. Searching status decision
  3. Labor market dynamics
- Labor Market States
  1. Unemployed
  2. Self-employed
  3. Informal Employee
  4. Formal Employee

Schooling Decision

- Irrevocable decision about schooling level $h \in \{0, 1\}$
- Individual-specific heterogeneity
  - costs $\kappa \sim T(\kappa)$
  - opportunity cost - PDV of participating in LMK as $h = 0$

$\Rightarrow$ Only agents with $\kappa < \kappa^\star(y)$ will acquire $h = 1$

- All labor market parameters are allowed to be schooling-specific

Searching-status Decision

- Irrevocable decision $s \in \{0, 1\}$:
  - Self-employed $(s = 1)$
  - Unemployed $(s = 0)$
- Search for a job in both states but receive offers at different rates: $\gamma_h < \lambda_h$
- Self-employment income $y \sim R(y|h)$

$\Rightarrow$ Only agents with $y \geq y^{\star}(h)$ search while also working as self-employed

Labor Market Dynamics

State	PDV	Shock	Flow Utility
Workers:			
Unemployed	$U(h)$	λ_h	$\xi_h + \beta_{0,h} B_0$
Self-Employed	$S(y,h)$	γ_h	$y + \beta_{0,h} B_0$
Informal Employee	$E_0[w,y,h]$	η_h, χ_h	$w_0(x;y,h) + \beta_{0,h} B_0$
Formal Employee	$E_1[w,y,h]$	η_h, χ_h	$w_1(x;y,h) + \beta_{1,h} B_1[w_1(x;y,h)]$
Firms:			
Vacancy	$V[h]$	ζ_h	ν_h
Filled Informal Job	$F_0[x,y,h]$	η_h, χ_h	$x - w_0(x;y,h)$
Filled Formal Job	$F_1[x,y,h]$	η_h, χ_h	$x - (1+t)w_1(x;y,h)$

$\Rightarrow$ Match-specific productivity: $x \sim G(x|h)$

$\Rightarrow$ One-shot penalty for firms hiring illegally: $c_h w_0(x;y,h)$

$\Rightarrow$ Matching function determines $\{\lambda_h, \gamma_h, \zeta_h\}$: $m_h = (u_h + \psi_h s_h)^{\iota_h} (v_h)^{1-\iota_h}$

Labor Market Institutions and Wage Determination

- Non-wage workers' flow value:
  - formal employee $= \beta_{1,h} B_1[w_1(x;y,h)] = \beta_{1,h}[\tau t w_1(x;y,h) + b_1]$
  - informal employee $= \beta_{0,h} B_0$

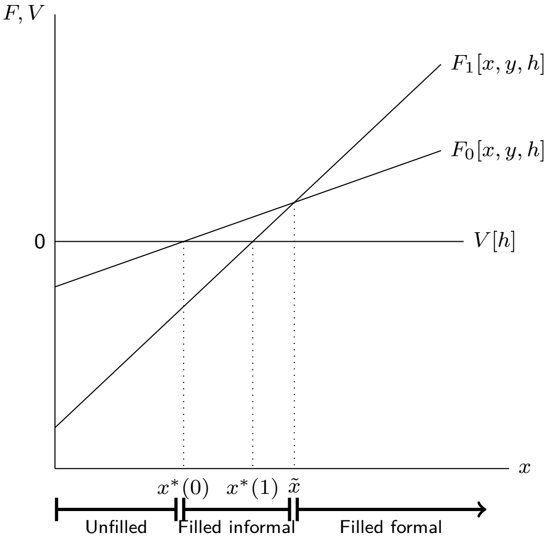
  $\Rightarrow$ b_1 generates spillovers within and between schooling levels

- Nash-bargaining wage schedules (under free-entry of firms) are:

$$\begin{aligned} w_0(x;y,h) &= \frac{\alpha_h}{1+\chi_h c_h} x + (1-\alpha_h)[\rho Q(y,h) - \beta_{0,h} B_0] \\ w_1(x;y,h) &= \frac{\alpha_h}{1+t} x + \frac{(1-\alpha_h)}{1+\beta_{1,h}\tau t}[\rho Q(y,h) - \beta_{1,h} b_1] \end{aligned}$$

$\Rightarrow$ $Q(y,h) \equiv \max\{S(y,h), U(h)\}$

Equilibrium Representation



Empirical Implications

$\Rightarrow$ Main stylized facts of informal labor markets are replicated in equilibrium:

1. A mixture of formal and informal jobs is realized
2. Formal employees have on average higher wages than informal employees
3. But their accepted wage distributions overlap
4. Informal employees and self-employed have different labor market dynamics
5. Firms hire formal or informal workers at different points in time
6. Workers transit over time between different formality status

Data Sources

1. Mexico's Labor Force Survey (ENOE) in 2005
   - Nonagricultural, full-time, male, private-sector workers

   $\Rightarrow$ Secondary-school between the ages of 25 and 55 who reside in urban areas
   - $w \equiv$ Hourly wages as employee, main job after labor contributions
   - $y \equiv$ Hourly labor income as self-employed, without paid employees
   - $f = 1$ if employee is contributing to the social-security fund; $= 0$ otherwise
   - $h = 1$ if Upper secondary completed $= 0$ if Lower secondary completed

2. Aggregate labor shares for Mexico in 2005
   - Total compensations per employee as percentage of GDP

3. Vacancy rates for 2005
   - Good coverage of vacancy posting in urban areas

   $\Rightarrow$ Detailed information on the schooling level required for the job

Identification: Search-Matching-Bargaining Parameters

- $G(x|h)$: Has to be "recoverable"
  - $\Rightarrow$ We assume lognormal with parameters $\{\mu_{x,h}, \sigma_{x,h}\}$
- $\lambda_h, \gamma_h, \eta_h$: stationarity + optimal decision rules identify mobility rates from
  - $\Rightarrow$ Transitions
  - $\Rightarrow$ Steady state distributions over labor market states
- ρ, ξ_h: Use $Q(y, h)$ to obtain their joint identification
- Nash Bargaining coefficient: $\alpha_1 = \alpha_0 = \alpha$
  - $\Rightarrow$ Use labor shares $(w_f(x; y, h)/x)$

Identification: Matching Function + Demand Side

- $\{\psi_h, \iota_h\}$: use vacancy rate and define mkt tightness $\omega_h \equiv \frac{v_h}{u_h+\psi_h s_h}$, so that

$$\begin{aligned}\psi_h &= \frac{\gamma_h}{\lambda_h}\\ \iota_h &= \frac{\ln\omega_h - \ln\lambda_h}{\ln\omega_h}\end{aligned}$$

- Then, back out the demand side parameters
  - $\zeta_h = \omega_h^{-\iota_h}$
  - ν_h: use firm's value function and impose free entry

Identification: Informality Parameters (β_1 and c_h)

- Different transition rates out of formal jobs and informal jobs identify χ_h
- Overlap between formal and informal accepted wage distributions

$$w_0(\tilde{x}(y,h);y,h) - w_1(\tilde{x}(y,h);y,h) > 0$$

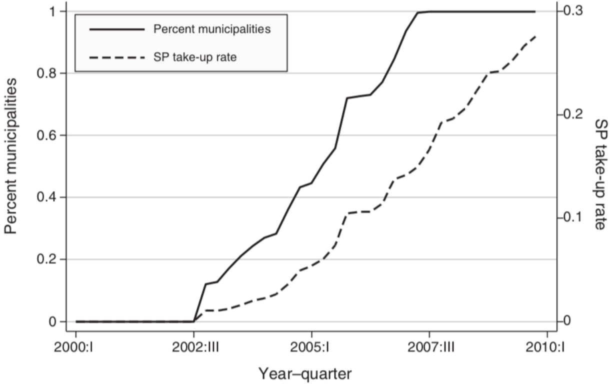
$\Rightarrow$ Given x, formal employees receive higher non-wage benefits

$\Rightarrow$ β_1 and c_h alter location and overlap of accepted wage distributions

- Variation in y is useful variation to separately identify the parameters

Identification: Informality Parameters (β_0)

- The identification of β_0 requires the use of additional information
- We exploit staggered entry of the *Seguro Popular* (SP) program in 2005



$\Rightarrow$ In terms of our model, SP $\approx \uparrow$ in B_0 by 25%

Identification: Informality Parameters (β_0, cont'd)

- Variation in B_0 identify β_0 if uncorrelated with changes in model primitives

$\Rightarrow$ Labor market outcomes pre-policy (2002) seem balanced

	Hourly Wages (log)			Labor Market Proportions			
	Formal	Informal	Self	Formal	Informal	Self	Unempl
SP in 2005 (1=yes)	-0.041	0.048	-0.035	-0.034	0.035	-0.004	0.003
	(0.036)	(0.055)	(0.062)	(0.026)	(0.019)	(0.014)	(0.006)
Complete Sec. (1=yes)	0.218	0.288	0.092	0.061	-0.036	-0.029	0.003
	(0.017)	(0.032)	(0.033)	(0.011)	(0.008)	(0.008)	(0.003)
Number of Obs.	7865	5474	2777	16458	16458	16458	16458

Identification: Self-employment and Schooling Parameters

- $R(y|h)$: Identified by observed self-employment earnings, once we assume a recoverable primitive distribution

  $\Rightarrow$ We assume lognormal with parameters $\{\mu_{y,h}, \sigma_{y,h}\}$

- $T(\kappa)$: The threshold crossing decision rule allows for the identification of one parameter from the proportions of individuals in the two schooling levels

$$\frac{1}{n}\sum_{i=1}^{n} h_i = \int_y T(\kappa^*(y))dR(y|0)$$

  $\Rightarrow$ We assume a negative exponential with parameters δ

Identification: Unobserved Ability Types

- Type is known to the individual but unobserved in the data

- We denote each type with k and its proportion in the population with π_k

$$x|k = a_k^G x$$
$$y|k = a_k^R y$$
$$\kappa|k = a_k^T \kappa$$

$\Rightarrow$ Duration dependence in unemployment identifies these parameters

  - Hazard rates at three and six months for both schooling levels

- Assume $K = 2$
  - type $k = 1$ normalized to $a_1^T = a_1^R = a_1^G = 1$
  - type $k = 2$ exhibiting $a_2^T < 1; a_2^R > 1; a_2^G > 1$

Estimation in Two Steps

1. For $s \in \{0, 1\}$ and SP $\in \{0, 1\}$, we match the following moments
   - Proportions of individuals in each labor market state
   - Accepted wage distributions of formal and informal employees
     - ⇒ Mean and SD: overall and by quintiles
     - ⇒ Overlap: % of formal empl. for each quintile of the informal wage distribution
   - Accepted earnings distributions of self-employed
     - ⇒ Mean and SD
   - Transitions between LMK states (yearly)
   - Hazard rates out of unemployment (at 3 and 6 months)
   - Labor Shares
2. Estimate demand-side parameters using vacancy rates

Parameter Estimates (selected coeffs)

	Low Schooling $h=0$		High Schooling: $h=1$	
	Coeff.	Std. Error	Coeff.	Std. Error
	Search, Matching, and Bargaining			
λ_h	0.4679	0.0035	0.5167	0.0098
γ_h	0.0349	0.0042	0.0306	0.0014
η_h	0.0326	0.0007	0.0190	0.0052
$\mu_{x,h}$	2.7616	0.0367	2.6749	0.0382
$\sigma_{x,h}$	0.6243	0.0132	0.7970	0.0038
$\mu_{y,h}$	1.6718	0.0188	1.9497	0.0763
$\sigma_{y,h}$	0.7754	0.0028	0.8027	0.0258
ξ_h	-103.46	1.6661	-158.05	4.6038
α	0.5630	0.0169	0.5630	0.0169
	Preferences and Informality			
$\beta_{1,h}$	0.7949	0.0044	0.6091	0.0043
$\beta_{0,h}$	0.9862	0.0038	0.9807	0.0015
χ_h	0.0079	0.0004	0.0113	0.0008
c_h	12.882	0.7045	16.574	1.3932
	Matching Function and Demand Side			
ψ_h	0.0745	0.0088	0.0592	0.0034
ι_h	0.7321	0.0253	0.7281	0.0184
ζ_h	7.9718	1.6278	5.8569	0.8742
ν_h	-496.01	288.80	-773.80	111.34

Returns to Schooling

	Ability: Low $k=1$	High $k=2$
PDV of Labor Market Search:		
$\int_y Q(y,h)dR(y\|h)$	0.309	0.278
Average Accepted Wages:		
F: $E_h\left[w_1 \mid \tilde{x}(y,h) \leq x\right]$	0.479	0.435
I: $E_h\left[w_0 \mid x_0^*(y,h) \leq x < \tilde{x}(y,h)\right]$	0.281	0.296
Average Offered Wages:		
F: $E_h\left[w_1 \mid y < y^*(h)\right]$	0.213	0.210
F: $E_h\left[w_1 \mid y \geq y^*(h)\right]$	0.213	0.204
I: $E_h\left[w_0 \mid y < y^*(h)\right]$	0.133	0.134
I: $E_h\left[w_0 \mid y \geq y^*(h)\right]$	0.142	0.136

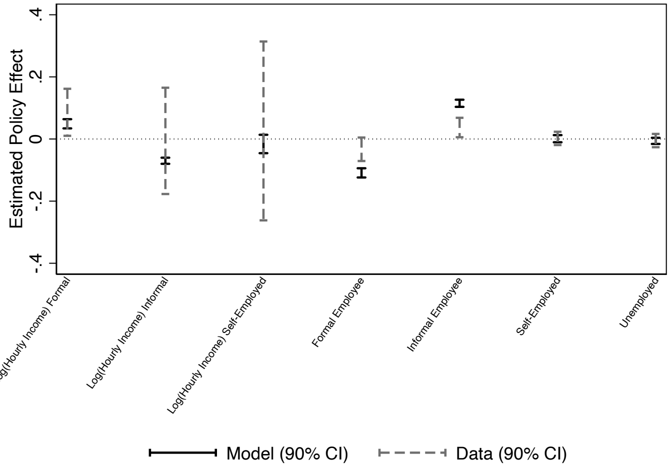
Out-of-Sample Model Validation

- Estimate the effect of $\uparrow B_0$ using SP roll-out one year later (2006)

$$y_{i,q} = \theta d_{m(i),q} + \vartheta h_i + \varrho_{m(i)} + \varphi_q + \epsilon_{i,q}$$

- Predict change in LMK outcomes with B_0^{2006} using estimated model
- Estimate TWFE-DID specifications on both actual and simulated data

Out-of-Sample Model Validation

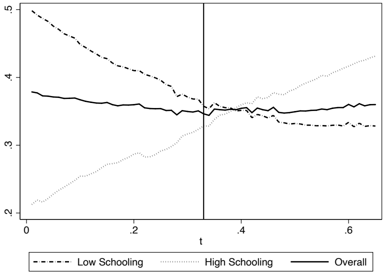


Counterfactual 1: The Equilibrium Effects of Informality

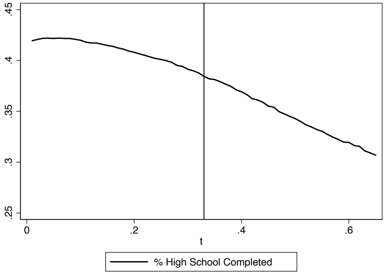
Model:	Firms can only offer a formal contract			
Specifications:	Baseline Model	Exogenous Schooling	Exogenous Contact Rates	Hosios-like Condition ($\alpha = \iota$)
Flow Welfare:				
Total	-0.0596	-0.0750	-0.0020	0.0478
Workers	-0.0460	-0.0599	0.0166	0.0570
Firms	-0.2821	-0.3219	-0.3055	-0.1589
Labor Market Proportions:				
Unemployed	0.0213	0.0636	0.0019	-0.0459
Self-employed	0.3353	0.3526	0.3625	0.2329
Formal Employees	0.0275	-0.0146	-0.0376	0.0076
Schooling Outcomes:				
% HS Completed	0.1029	–	0.0781	0.1501
% High Ability in HS	0.0538	–	0.0569	0.0628

NOTE: Relative changes wrt the benchmark model. Hosios increases α from 0.56 to 0.73.

Counterfactual 2: Changes in Payroll Tax Rate (t)



(a) Informality



(b) Schooling

- Composition effects over schooling/ability explain no impact on informality
- Balanced-budget policy with $\tau = 0 \rightarrow 10\%$ increase in high-school completion

Main Takeways from the Estimated Model

1. Returns to schooling are substantial
2. Informality is welfare improving but:
   - Significantly more so for firms than workers
   - Reduces human capital accumulation (hold-up problem)
3. Payroll tax rate has a non-intuitive impact on equilibrium outcomes
   - Informality rate not a good indicator for policy
   - Redistributive forces within the formal system are key

DID $\Leftrightarrow$ Economic Model

- Relevant institutional features are included in the model in a tractable way
- These parameters are hard to separately identify using labor market data
- The staggered roll-out of the policy provides additional variation to:
  - $\Rightarrow$ Identify the (average) valuation of non-contributory benefits
  - $\Rightarrow$ Validate the model on a different time period by simulating one-step ahead