Difference-in-differences

Joacim Tåg

Spring 2024

Overview

Theory

Basic idea

Assumptions and DiD regressions

Multiple periods and variation in treatment timing

Applications

Liquidity Provision in Banking Crises

What Is the Cost of Privatization for Workers?

Technology Transfer in Mergers and Acquisitions and the Careers of Workers

Tolerating Losses for Growth: How US Venture Capitalists Invest Abroad

Summary

Theory

Overview

Theory

Basic idea

Assumptions and DiD regressions

Multiple periods and variation in treatment timing

Applications

Liquidity Provision in Banking Crises

What Is the Cost of Privatization for Workers?

Technology Transfer in Mergers and Acquisitions and the Careers of Workers

Tolerating Losses for Growth: How US Venture Capitalists Invest Abroad

Summary

Theory

Basic idea

Potential outcomes

- Observed variables:
  - Treatment, D_i, is observed as either 0 or 1 for each person i
  - Actual outcomes, Y_i , are observed for each person i
- The Rubin Causal Model introduces the notion of potential outcomes:
  - We can imagine person i having potentially experienced one of the two possible different states of the world.
  - E.g. either she/he is hospitalized or she/he is not.

Potential outcomes

- Each person has two potential outcomes, but only one observed outcome.
- Unobserved counterfactual variables:
  - Each individual i has two potential outcomes that in principle could exist, but only one of them is observed:

$$\text{Potential outcome } = \begin{cases} Y_i^1 & \text{if } D_i = 1 \\ Y_i^0 & \text{if } D_i = 0 \end{cases} \tag{1}$$

- Switching equation: $Y_i = D_i Y_i^1 + (1 - D_i)\, Y_i^0$

Individual-level causal effect

- Individual-level causal effect = Individual treatment effect

$$\delta_i \equiv Y^1{}_i - Y_i^0 \tag{2}$$

- What would be the change in i's health if he/she visits hospital ($D_i = 1$), as compared to him/her not visiting hospital ($D_i = 0$)?
  - The "what would have been" nature defines the causal states of interest
  - There is a key "ceteris paribus" assumption assumption here

Fundamental problem of causal inference

- Holland (1986): It is impossible to observe both Y_i^1 and Y_i^0 for the same individual and so individual causal effects, δ_i, are unknowable

Group	$\mathbf{Y}_i{}^1$	$\mathbf{Y}_i{}^0$
Treatment $(D_i = 1)$	Observable as $\mathbf{Y}_i$	Counterfactual
Control $(D_i = 0)$	Counterfactual	Observable as $\mathbf{Y_i}$

(3)

- Causal effects are defined within rows, but only the diagonal of the table is observable

The missing data problem

- Missing data problem 1:
  - One can never observe the potential outcome under the treatment state for those observed in the control state.
- Missing data problem 2:
  - One can never observe the potential outcome under the controls state for those observed in the treatment state.
- Consequence of 1 and 2:
  - One can never calculate individual-level causal effects.
- But, under certain assumptions, we can calculate various average treatment (AT) effects.

Average Treatment Effect (ATE)

$$
\begin{aligned}
ATE &= E[\delta_i] \\
&= E\left[Y_i^1 - Y_i^0\right] \\
&= E\left[Y_i^1\right] - E\left[Y_i^0\right]
\end{aligned}
\tag{4}
$$

- The Average Treatment Effect (ATE):
  - The expected (“average”, “mean”) effect of the treatment in population
  - The average change in outcomes that would be realized if all units were required to choose $D = 1$ (counterfactually) compared to the case in which all units were required to choose $D = 0$ (counterfactually).

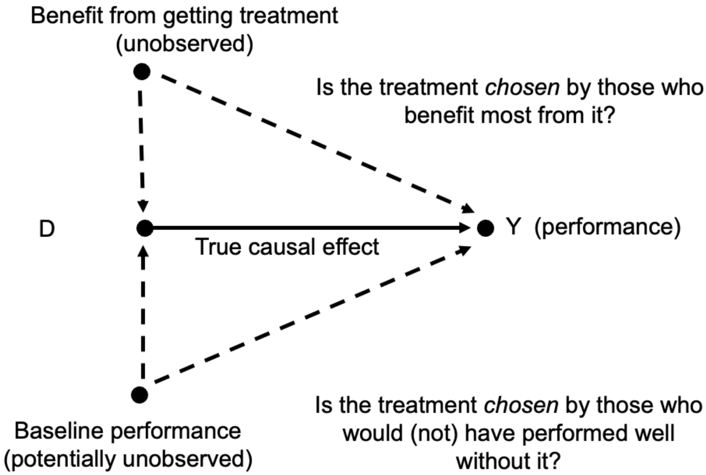
Average Treatment on the Treated (ATT)

$$\begin{aligned} ATT &= E\left[\delta_i \mid D_i = 1\right] \\ &= E\left[Y_i^1 - Y_i^0 \mid D_i = 1\right] \\ &= E\left[Y_i^1 \mid D_i = 1\right] - E\left[Y_i^0 \mid D_i = 1\right] \end{aligned} \tag{5}$$

- ATT is the population mean treatment effect for the group of units that had been sorted to the treatment group.

- ATT might be relevant for evaluating e.g. optional government programs (e.g. labor market programs), since it measures the benefit to those who choose (or are chosen) to receive the treatment.

The selection problem



RCTs vs Observational studies

- Experiments may be impractical due to:
  - Cost (experiments need to be powered correctly)
  - Ethics (some things are not ethical to randomize)
  - Feasibility (some things cannot be randomized)
- Observational studies: How are units assigned to treatment?
  - Selection on observables?
    - Regression and matching
  - Selection on unobservables?
    - Instrumental variables (exogenous variation in treatment)
    - Regression discontinuity design (known selection mechanism)
    - Difference-in-differences (panel data around treatment)

Difference-in-differences

- DiD is a research design for estimating causal effects when there is certain type of selection on unobservables

- DID makes use of panel data and asks:

  ”What can we identify by comparing the effect of a treatment on a treatment group to that of a control group, using longitudinal data from some kind of a natural experiment?”

Difference-in-differences

- ”Natural experiment”: when units are exposed to the treatment vs. control conditions that are determined by nature, policy, or by other factors
  - That are outside the control of the researchers
  - An event that occurs naturally which causes exogenous variation in some treatment variable of interest for some units
  - A circumstance such that a consequential treatment was handed to some people and denied to others haphazardly
- Note: natural experiments are neither “an estimator” or “an experiment”

How difference-in-differences work

- DiD can be used to estimate e.g. the effects of
  - certain policy interventions and policy changes (”natural experiments”)
  - that do not affect everybody
  - at the same time and in the same way
- Using data on the outcomes of four groups, DiD calculates the effect of a treatment on an outcome by comparing
  - the (average) change over time in the outcome variable for the treatment group with
  - the (average) change over time for the control group

How difference-in-differences work

- DID requires data from pre-/post-intervention. This can be:
  - panel data (e.g., individual level data over time)
  - repeated cross-sectional data (e.g at individual or group level, for various years)
  - Note: repeated cross-sectional analyses can lead to compositional bias
- Four groups (2x2):
  1. Group which received the treatment (post-treatment treated)
  2. The treated prior to their treatment (pre-treatment treated)
  3. The non-treated in the periods before the treatment occurs to the treated (pre-treatment non-treated)
  4. The non-treated in the periods after the treatment occurs to the treated (post-treatment non-treated)

Example

- Suppose Finland reforms its taxi regulation in 2018:
  - What is the effect of the regulation on taxi prices (in 2019)?
  - Let δ be the true size of the price effect (”treatment effect”)
- Two possible comparisons:
  - Comparison to similar peer (comparison of Finland vs Sweden in 2019)
  - Before-after comparison (comparison in Finland between 2018 and 2019)

Example

Country	2018	2019
FIN	P_F	$P_F + T + \delta$
SWE	P_S	$P_F + T$

- Data: 2x2 matrix
  - P: price
  - T: time trend in price
  - δ: treatment effect of reform

Comparison to peers

Country	2018	2019
FIN	P_F	$P_F + T + \delta$
SWE	P_S	$P_S + T$

- Comparing prices in 2019 between FIN and SWE
  - $SDO = P_F + \delta - P_S$
  - Comparing peers nets out common time trends
  - However, the SDO is biased because of initial price level differences ($P_F - P_S$)

Before-after comparison

Country	2018	2019
FIN	P_F	$P_F + T + \delta$
SWE	P_S	$P_S + T$

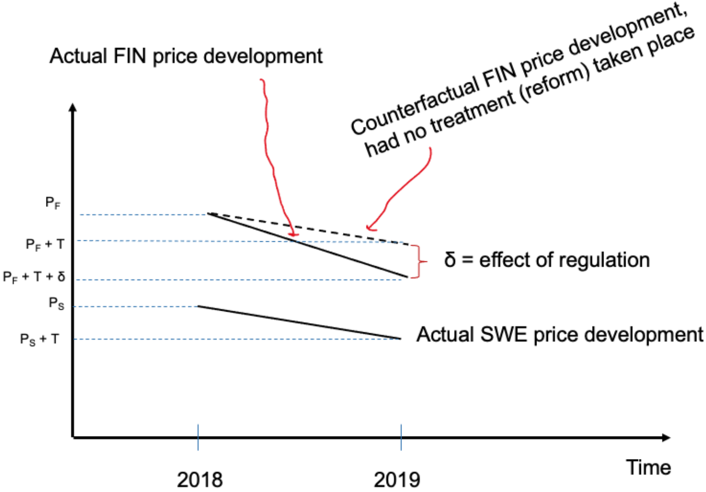
- Comparing prices in FIN between 2018 and 2019
  - $SDO = T + \delta$
  - Comparing before after nets out initial price levels
  - However, the SDO is biased because of underlying trends (T)

Difference-in-differences

Country	2018	2019	Diff
FIN	P_F	$P_F + T + \delta$	T+δ
SWE	P_S	$P_S + T$	T
Diff	$P_F - P_S$	$P_F - P_S + \delta$	δ

- Comparing prices in FIN between 2018 and 2019
  - $Diff - in - Diff = (T + \delta) - T = \delta$
  - Double comparison nets out all initial level differences and shared trends
  - Nets out level and trend biases from both observed and unobserved variables

Difference-in-differences



Theory

Assumptions and DiD regressions

What does DiD estimate?

- Recall the Rubin causal model (potential outcomes)
  - $Y_i^0(t) =$ unit i's outcome in non-treated state (0) at time t
  - $Y_i^1(t) =$ unit i's outcome in treated state (1) at time t
- The ATT is equation

$$\text{ATT=E}\Big[Y_i^1(1) - Y_i^0(1) \mid D_i = 1\Big] = E\Big[Y_i^1(1) \mid D_i = 1\Big] - E\Big[Y_i^0(1) \mid D_i = 1\Big]$$

- Note:
  - Observed (what is outcome of treated after treatment): $E\Big[Y_i^1(1) \mid D_i = 1\Big]$
  - Unobserved (what is outcome of treated if not treated): $\text{E}\Big[Y_i^0(1) \mid D_i = 1\Big]$

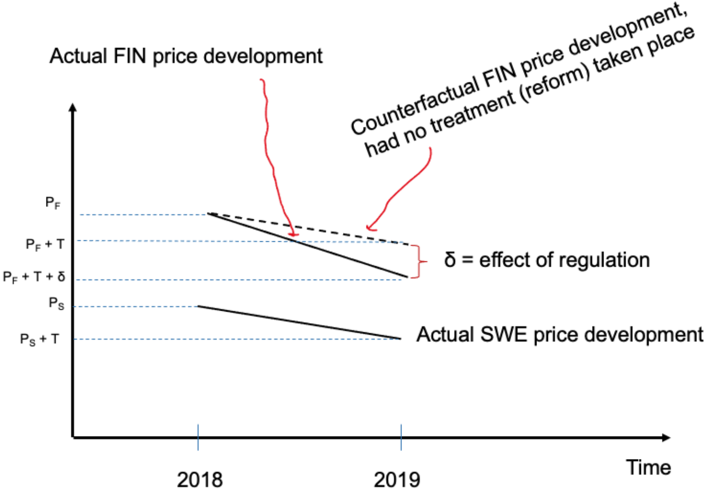
The parallel trends assumption

- The key assumption in DiD is the parallel trends assumption
- Verbally:
  - ”In absence of treatment, the average outcome for the treated and the average outcome of the non-treated would have been the same”
- Formally:

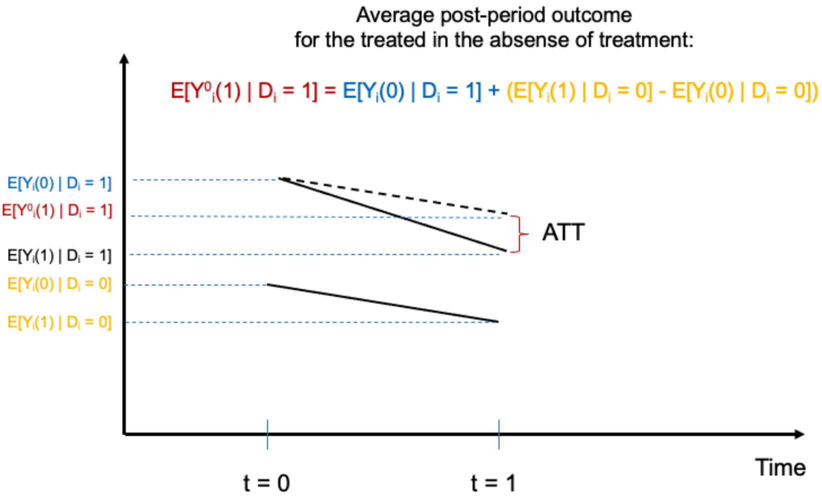
$$E\left[Y_i^0(1) \mid D_i = 1\right] - E\left[Y_i^0(0) \mid D_i = 1\right] = E\left[Y_i^0(1) \mid D_i = 0\right] - E\left[Y_i^0(0) \mid D_i = 0\right]$$

- Graphically:

The parallel trends assumption



The parallel trends assumption



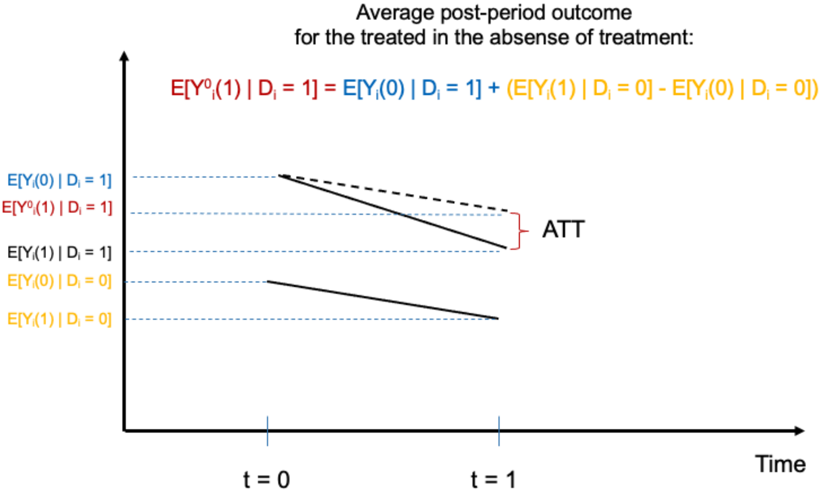
Three central assumptions to keep in mind

- Assumption 1: Parallel trends. Failure leads to non-parallel trends bias.
- Assumption 2: SUTVA:
  - ”Treatment cannot affect the controls”
  - E.g. a policy in one region may affect behavior in control regions if subject learn about it
  - E.g. a large firm bankruptcy in a local labor market may affect workers in the firm but also have local labor market spillovers on workers in other firms (that may be used as controls)
- Assumption 3: No anticipation effects, i.e. treatment has no causal effect prior to its implementation. Important to avoid treatment effects in the pre-period.

The parallel trends assumption

- Observations on the parallel trends assumption:
  - Is about change over time, not about level difference

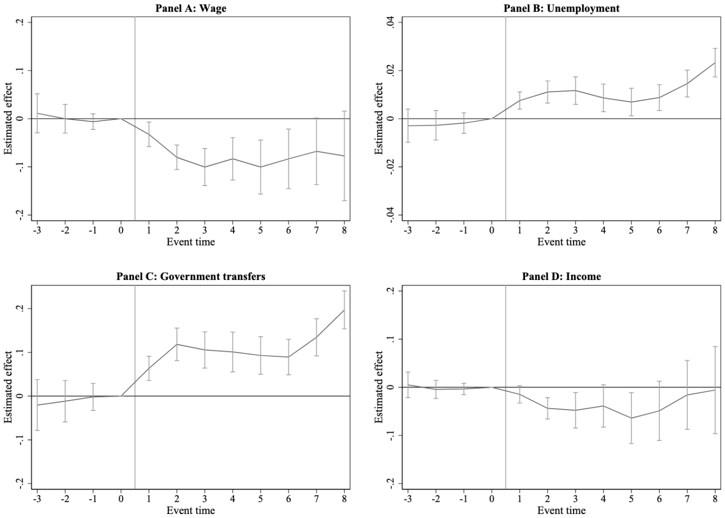
The parallel trends assumption



The parallel trends assumption

- Observations on the parallel trends assumption:
  - Is about change over time, not about level difference
  - Is NOT invariant to the functional form of the econometric model (Y vs $log(y)$)
  - Is impossible to check since the counterfactual is unobserved
  - What people do: assess pre-trends. They should be parallel, suggesting that shocks in the past have affected the groups in a similar way

The parallel trends assumption



The parallel trends assumption

- Observations on the parallel trends assumption:
  - Is about change over time, not about level difference
  - Is NOT invariant to the functional form of the econometric model (Y vs $log(y)$)
  - Is impossible to check since the counterfactual is unobserved
  - What people do: assess pre-trends. They should be parallel, suggesting that shocks in the past have affected the groups in a similar way
  - Note: one must still worry about unobservable chances that take place at the same time as treatment

The parallel trends assumption

- Often policymakers will select the treatment and controls based on pre-existing differences in outcomes – this often leads to non-parallel trends bias

- Example 1: Regional targeting. Target regions that are most promising (or likely to develop badly) leads to a ”selection bias” and violates parallel trends

- Example 2: The “Ashenfelter dip”.
  - Participants in job training program often experience a dip in earnings just prior to entering the program
  - Since wages have a natural tendency to mean reversion, program participants’ wages would have grown rapidly also in absence of program
  - Thus: Comparing wages of participants and non-participants using DiD leads to an upward biased estimate of the program effect

The parallel trends assumption

- Example 3: Venture Capital Funding
  - Suppose the outcome is firm growth before (t=0) and after (t=1) VC funding and that the treatment is the event of obtaining VC funding
  - Problem 1: the firms that received VC funding would likely have grown more relative to control firms not receiving funding (non-parallel trend bias, but pre-trends can be checked)
  - Problem 2: an underlying unobservable may have caused both VC funding and firm turnover, i.e. a granted patent (selection/omitted variable bias)
  - Problem 3: some other control firms might now compete with a better capitalized fast growing firm (SUTVA fails to hold)

DiD in a regression framework

- Typical regression model:

$$\text{Y}_{it} = \alpha + \beta_1 \text{ Treat }_i + \beta_2 \text{ Post }_t + \beta_3(\text{ Treat } \times \text{ Post })_{it} + \varepsilon_{it} \tag{6}$$

- This captures all four groups:
  - Group 1: Group which received the treatment (post-treatment treated)
  - Group 2: The treated prior to their treatment (pre-treatment treated),
  - Group 3: The non-treated in the period before the treatment occurs to the treated (pre-treatment non-treated)
  - Group 4: The non-treated in the current period (post-treatment non-treated).

DiD in a regression framework

- Typical regression model:

$$Y_{it} = \alpha + \beta_1 \text{ Treat }_i + \beta_2 \text{ Post }_t + \beta_3(\text{ Treat } \times \text{ Post })_{it} + \varepsilon_{it} \tag{7}$$

- This captures all four groups:
  - Group 1 (post-treatment treated): $Treat_i = 1$, $Post_t = 1$
  - Group 2 (pre-treatment treated): $Treat_i = 1$, $Post_t = 0$
  - Group 3 (pre-treatment non-treated): $Treat_i = 0$, $Post_t = 0$ (omitted)
  - Group 4 (post-treatment non-treated): $Treat_i = 0$, $Post_t = 1$

DiD model with leads and lags

- Typical regression model:

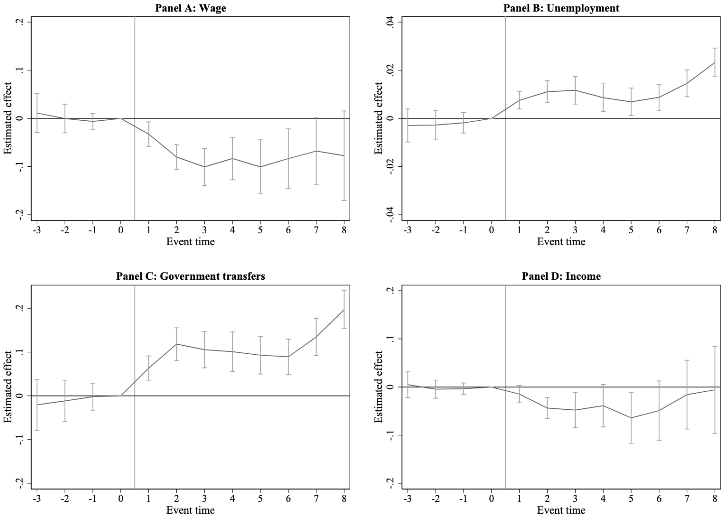
$$Y_{it} = \alpha + \beta_1 \text{ Treat }_i + \beta_2 \text{ Post }_t + \beta_3 (\text{ Treat } \times \text{ Post })_{it} + \varepsilon_{it} \tag{8}$$

- Model often extended to allow event specific estimates to check pre-trends.

$$Y_{it} = \alpha + \beta \text{Treat}_i + \tau_t + \beta_t \sum \tau_t \times \text{Treat}_i + \varepsilon_{it} \tag{9}$$

- Here, τ_t is now year specific dummies that replace the $Post_t$ dummy

DiD model with leads and lags



DiDiD regressions

- Triple-diff regressions (DDD or DiDiD) splits the sample across subgroups:

$$Y_{it} = \alpha_0 + \alpha_1 T_i + \alpha_2 Post_k + \beta_1 DID_{ik} + \mu_1 G_i + \mu_2 G_i * T_i + \mu_3 G_i * Post_k + \beta_2 DIDID_{ik} + \epsilon_{ikt} \quad (10)$$

- β_2 captures comparison before-after, treated-control, and between subgroups G_i

DiDiD regressions

- Relevant parallel trends assumption is in triple differences
  - The triple-differencing nets out treatment group specific shocks, since the subgroups G_i are present within treated and control, before after (observed AND unobserved)
  - Highly useful for heterogeneity analyses, i.e. what is the differential effects of minimum wage increases on women relative to men
  - Note: the triple difference specification gives you tests of statistical significance between subgroup means, however it imposes a linearity assumption on the relationships in the model (unless you include flexible controls).
  - Good practice to show both subsample DiDs + triple-diffs (example will come later)

Synthetic controls

- Sometimes a suitable control group is not available
- Then, it may be possible to create a synthetic control group
  - Idea is to ”optimally” weight together different control groups to produce a better one
  - A benefits is that the weights are observed
  - Ideally, we obtain a synthetic control group that is identical to the treated group
  - Often used when the units of analysis are a few aggregate units (think case study comparisons)
  - Bridges the gap between qualitative and quantitative researchers
- Issue: many ways to choose the weights $\Rightarrow$ researchers can often choose the results they want

Theory

Multiple periods and variation in treatment timing

Multiple periods and variation in treatment timing

- Standard DiD presumes that a specific treatment date exists (e.g reform year)
- Often, however, treatment is staggered over time:
  - Staggered introduction of policies across countries/states
  - Acquisitions of firms in different years
- Other names:
  - Two-way fixed effects with differential timing
  - Dynamic DiD models
  - Event studies
  - Staggered DiD

Standard approach

- Standard approach has (up until about 2019) been to estimate a TWFE regression:

$$y_{it} = \alpha_0 + \delta D_{it} + X_{it} + \alpha_i + \alpha_t + \epsilon_{it} \tag{11}$$

- Here, α_i are unit FE, α_t are time FE and D_{it} a treatment dummy that turns on after treatment for the treated
- Standard intuition has been that it is straight forward to extend the DiD assumptions to this case, but this is not true.
- Key issue: TWFE regression compares both ”treated and not-yet-treated” as well as ”between already-treated”. With heterogeneous treatment effects, this biases δ.

Recent advances

- Two papers to be aware of that cover most of the recent advances in detail (in case you end up using this approach in your master's thesis):
  - Baker, Andrew C., David F. Larcker, and Charles C.Y. Wang. "How Much Should We Trust Staggered Difference-in-Differences Estimates?" Journal of Financial Economics 144, no. 2 (May 2022): 370–95
  - Roth, Jonathan, Pedro H.C. Sant'Anna, Alyssa Bilinski, and John Poe. "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature." Journal of Econometrics 235, no. 2 (August 2023): 2218–44.

Roth et al 2023

Table 1
A checklist for DiD practitioners.

- Is everyone treated at the same time?

If yes, and panel is balanced, estimation with TWFE specifications such as (5) or (7) yield easily interpretable estimates.

If no, consider using a "heterogeneity-robust" estimator for staggered treatment timing as described in Section 3. The appropriate estimator will depend on whether treatment turns on/off and which parallel trends assumption you're willing to impose. Use TWFE only if you're willing to restrict treatment effect heterogeneity.

- Are you sure about the validity of the parallel trends assumption?

If yes, explain why, including a justification for your choice of functional form. If the justification is (quasi-)random treatment timing, consider using a more efficient estimator as discussed in Section 6.

If no, consider the following steps:

1. If parallel trends would be more plausible conditional on covariates, consider a method that conditions on covariates, as described in Section 4.2.
2. Assess the plausibility of the parallel trends assumption by constructing an event-study plot. If there is a common treatment date and you're using an unconditional parallel trends assumption, plot the coefficients from a specification like (16). If not, then see Section 4.3 for recommendations on event-plot construction.
3. Accompany the event-study plot with diagnostics of the power of the pre-test against relevant alternatives and/or non-inferiority tests, as described in Section 4.4.1.
4. Report formal sensitivity analyses that describe the robustness of the conclusions to potential violations of parallel trends, as described in Section 4.5.

- Do you have a large number of treated and untreated clusters sampled from a super-population?

If yes, then use cluster-robust methods at the cluster level. A good rule of thumb is to cluster at the level at which treatment is independently assigned (e.g. at the state level when policy is determined at the state level); see Section 5.2.

If you have a small number of treated clusters, consider using one of the alternative inference methods described in Section 5.1.

If you can't imagine the super-population, consider a design-based justification for inference instead, as discussed in Section 5.2.

Roth et al 2023

Table 2
Statistical packages for recent DiD methods.

Heterogeneity Robust Estimators for Staggered Treatment Timing		
Package	Software	Description
did, csdid	R, Stata	Implements Callaway and Sant'Anna (2021)
did2s	R, Stata	Implements Gardner (2021), Borusyak et al. (2021), Sun and Abraham (2021), Callaway and Sant'Anna (2021), Roth and Sant'Anna (2021)
didimputation, did_imputation	R, Stata	Implements Borusyak et al. (2021)
DIDmultiplegt, did_multiplegt	R, Stata	Implements de Chaisemartin and D'Haultfoeuille (2020)
eventstudyinteract	Stata	Implements Sun and Abraham (2021)
flexpaneldid	Stata	Implements Dettmann (2020), based on Heckman et al. (1998)
fixest	R	Implements Sun and Abraham (2021)
stackedev	Stata	Implements stacking approach in Cengiz et al. (2019)
staggered	R	Implements Roth and Sant'Anna (2021), Callaway and Sant'Anna (2021), and Sun and Abraham (2021)
xtevent	Stata	Implements Freyaldenhoven et al. (2019)
DiD with Covariates		
Package	Software	Description
DRDID, drdid	R, Stata	Implements Sant'Anna and Zhao (2020)
Diagnostics for TWFE with Staggered Timing		
Package	Software	Description
bacondecomp, ddtiming	R, Stata	Diagnostics from Goodman-Bacon (2021)
TwoWayFEWeights	R, Stata	Diagnostics from de Chaisemartin and D'Haultfoeuille (2020)
Diagnostic/ Sensitivity for Violations of Parallel Trends		
Package	Software	Description
honestDiD	R, Stata	Implements Rambachan and Roth (2022b)
pretrends	R	Diagnostics from Roth (2022)

Note: This table lists R and Stata packages for recent DiD methods, and is based on Asjad Naqvi's repository at https://asjadnaqvi.github.io/DiD/. Several of the packages listed under "Heterogeneity Robust Estimators" also accommodate covariates.

One ”simple” option: Stacked DiD regressions

- Suppose a state panel with staggered policy changes across states
  1. Create a separate balanced dataset for each treatment year (cohort). E.g. one treatment state and all others as controls.
  2. Normalize time in each dataset around the treatment year
  3. ”Stack” the datasets (append them)
  4. Run a standard DiD or TWFE model (event time and state differences) on the stacked data with state-by-cohort FE and cohort FE to account for repeated inclusion of the same controls across multiple cohorts (or make sure to not select the SAME controls for each cohort)

- Stacking balances the data in event time, so there is no differential timing → we are back in the standard DiD world

- Can use the Stata command ”stackedev” (github.com/joshbleiberg/stackedev)

Applications

Overview

Theory

- Basic idea
- Assumptions and DiD regressions
- Multiple periods and variation in treatment timing

Applications

- Liquidity Provision in Banking Crises
- What Is the Cost of Privatization for Workers?
- Technology Transfer in Mergers and Acquisitions and the Careers of Workers
- Tolerating Losses for Growth: How US Venture Capitalists Invest Abroad

Summary

Applications

Liquidity Provision in Banking Crises

Motivation

- Banking is a business built on confidence and trust
  - Banks lend to businesses and property owners in the expectation that most loans will be paid off when they come due
  - Depositors trust they'll be able to withdraw their funds on demand
  - Banks hold less cash than needed to pay all depositors, because most deposits are out on loan
  - Maturity mismatch happens
- If confidence falters, the system breaks down and we have a bank run
- Motives liquidity provision in banking crises
- Can such provision help banks survive? (Richardson and Troost, 2009)

Setting

- Analyzed the Great Financial Crisis in the US 1929–1933
  - Collapse of Caldwell and Company, ”We Bank on the South”, due to mismanagement and the stock market crash
- Different branches of the Federal Reserve monitored different banks
  - Atlanta Fed (6th): support the banks
  - St. Luis Fed (8th): let them fail (moral hazard)
  - District lines divided the state of Missisippi in two

Analysis

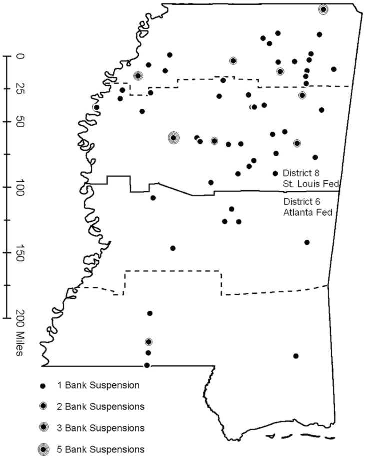


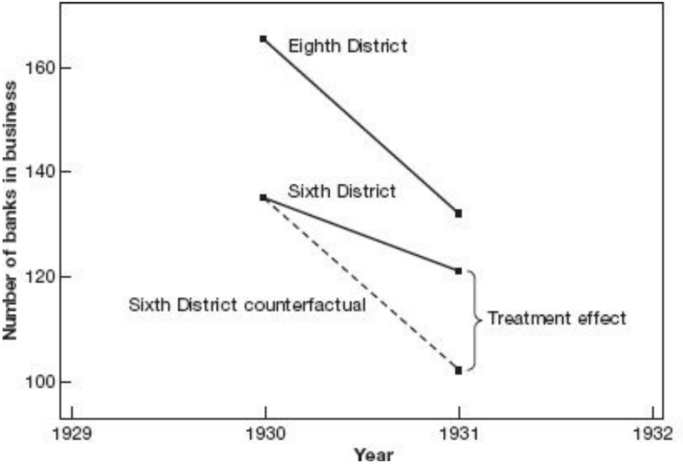
FIG. 1.—Mississippi's division into Federal Reserve districts and bank suspensions between October 1930 and March 1931. Source: See Section II. The solid line represents the Federal Reserve district border. The dotted lines enclose the counties for which at least half the area lies within 1 degree latitude of the district border.

Analysis

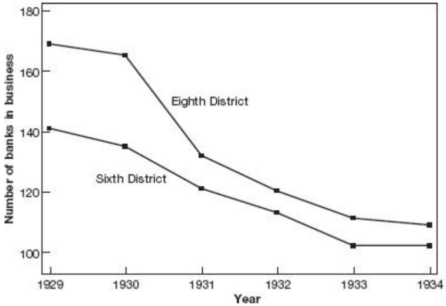
- Did liquidity provision affect number of banks that survived?

$$\begin{aligned}\delta_{DD} &= \left(Y_{6,1931} - Y_{6,1930}\right) - \left(Y_{8,1931} - Y_{8,1930}\right) \\ &= (121 - 135) - (132 - 165) \\ &= -14 - (-33) = 19.\end{aligned} \tag{12}$$

Analysis

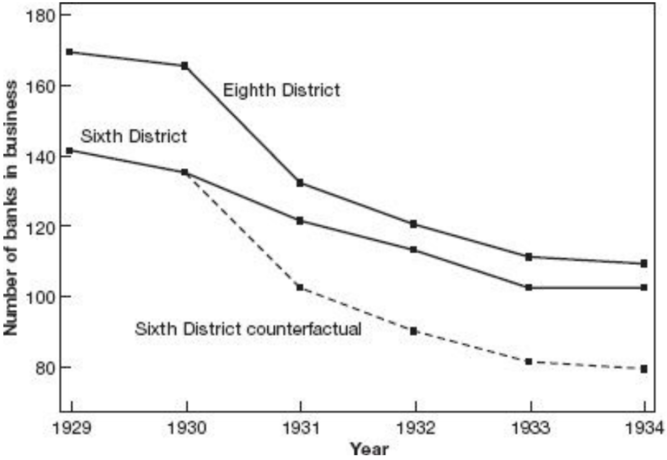

Eighth District
Sixth District
Sixth District counterfactual
Treatment effect
160
140
120
100
Number of banks in business
1929
1930
1931
1932
Year

Trends in bank failures in the Sixth and Eighth Federal Reserve Districts



Note: This figure shows the number of banks in operation in Mississippi in the Sixth and Eighth Federal Reserve Districts between 1929 and 1934.

Analysis


180
160
140
120
100
80
Number of banks in business
Eighth District
Sixth District
Sixth District counterfactual
1929
1930
1931
1932
1933
1934
Year

	1929	1933	Difference (1933–1929)
Panel A. Number of wholesale firms			
Sixth Federal Reserve District (Atlanta)	783	641	−142
Eighth Federal Reserve District (St. Louis)	930	607	−323
Difference (Sixth–Eighth)	−147	34	181
Panel B. Net wholesale sales ($ million)			
Sixth District Federal Reserve (Atlanta)	141	60	−81
Eighth District Federal Reserve (St. Louis)	245	83	−162
Difference (Sixth–Eighth)	−104	−23	81

Notes: This table presents a DD analysis of Federal Reserve liquidity effects on the number of wholesale firms and the dollar value of their sales, paralleling the DD analysis of liquidity effects on bank activity in Figure 5.1.

Summary

- Does liquidity provision help? YES
  - More banks survived
  - More businesses survived (real effects)
- Application of DiD is simple, if the data (”natural experiment”) is there
- Plotting the raw data in various ways can be very convincing

Applications

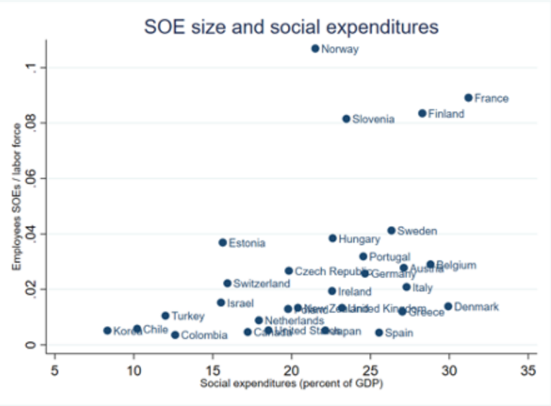
What Is the Cost of Privatization for Workers?

Olsson and Tåg 2023

- Privatization is on the agenda of policymakers across the globe
- Clear benefits for governance, productivity, and performance
- Less evidence on effects on labor despite key policy focus
    - + Improved performance → higher labor demand
    - -/+ Ownership change can trigger labor reallocation
        - Political goals can benefit workers
        - Soft budget constraints of managers lead to less layoffs
        - New owners might breach of implicit contracts
- Olsson and Tåg (2023): What Is the Cost of Privatization for Workers?

Data

- Workers observed over two decades for a multitude of outcomes
  - Long-run career effects
  - Non-wage outcomes and transfers
- Firm financial statements, boards, and management
  - Mechanisms: does labor reallocation drive productivity gains?
  - Cost/benefit analysis
- Substantial number of privatizations (369 firms, 53k workers) distributed over time and across industries
  - Identification
  - Heterogeneity



Magnitudes relevant for:

- Scandinavian countries
- Large part of rest of EU: France, Germany, Italy, Belgium, Austria, Poland, Portugal, Slovenia

Less relevant for:

- Countries without SOEs (US, Canada, Spain, Japan)
- Countries with weaker social safety nets (Korea, Chile, Mexico, Columbia, Turkey)

Empirical design

- Problems
    - staggered adoption (TWFE not OK)
    - selection
    - “employment anticipation” effects
- Solution
    - Stacked difference-in-difference regressions with matching
    - DiDiD regressions for heterogeneity
- Matching
    - Cohort specific cell matching on gender, industry, and year
    - Accounts for macro and industry shocks

Empirical design

- Model outcome Y of worker i at event year k at calendar year t as

  $\mathrm{Y}_{ikt} = \alpha + \pi AFTER_k + \gamma TREATED_i + \beta DID_{ik} + \omega_t + X_i + X_f + \epsilon_{ikt}$

  - DiD_{ik} is the interaction between $AFTER_k$ and $TREATED_i$
  - β captures ITT (intention to treat)
  - X_i includes controls for individual age, sex, immigrant status, labor market experience, tenure, educational fixed effects and municipality fixed effects
  - X_f includes firm size and industry fixed effects

- For dynamic effects, we replace $AFTER_k$ with event time dummies (or period dummies):

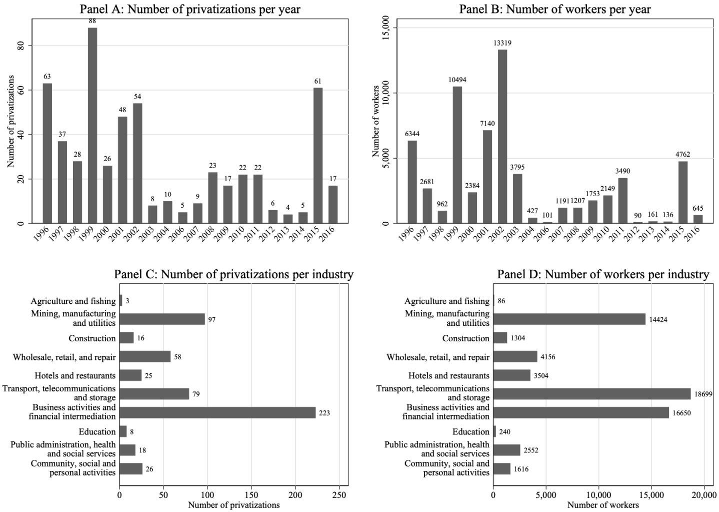
  $\mathrm{Y}_{ikt} = \alpha + \tau_k + \gamma TREATED_i + \beta_k \sum_{k=-3}^{k=8} \tau_k \times TREATED_i + \omega_t + X_i + X_f + \epsilon_{ikt}$

Stacked DiD regressions

- Stacking:
  1. Create a separate dataset for each treatment year (cohort). Done using 1-1 matching within cohort, then picking up the panel.
  2. Normalize time in each dataset around the treatment year
  3. ”Stack” the datasets (append them)
  4. Run a standard DiD model (event time and state differences) on the stacked data. Cluster or include individual-by-cohort FE to account for repeated inclusion of the same controls across multiple cohorts

- Recall: Stacking balances the data in event time, so there is no differential timing $\rightarrow$ we are back in the standard DiD world

Data



Panel A: Number of privatizations per year
Number of privatizations
0
20
40
60
80
63
37
28
88
26
48
54
8
10
5
9
23
17
22
22
6
4
5
61
17
1996
1997
1998
1999
2000
2001
2002
2003
2004
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
Panel B: Number of workers per year
Number of workers
0
5,000
10,000
15,000
6344
2681
962
10494
2384
7140
13319
3795
427
101
1191
1207
1753
2149
3490
90
161
136
4762
645
Panel C: Number of privatizations per industry
Agriculture and fishing
Mining, manufacturing and utilities
Construction
Wholesale, retail, and repair
Hotels and restaurants
Transport, telecommunications and storage
Business activities and financial intermediation
Education
Public administration, health and social services
Community, social and personal activities
3
97
16
58
25
79
223
8
18
26
0
50
100
150
200
250
Number of privatizations
Panel D: Number of workers per industry
86
14424
1304
4156
3504
18699
16650
240
2552
1616
0
5,000
10,000
15,000
20,000
Number of workers

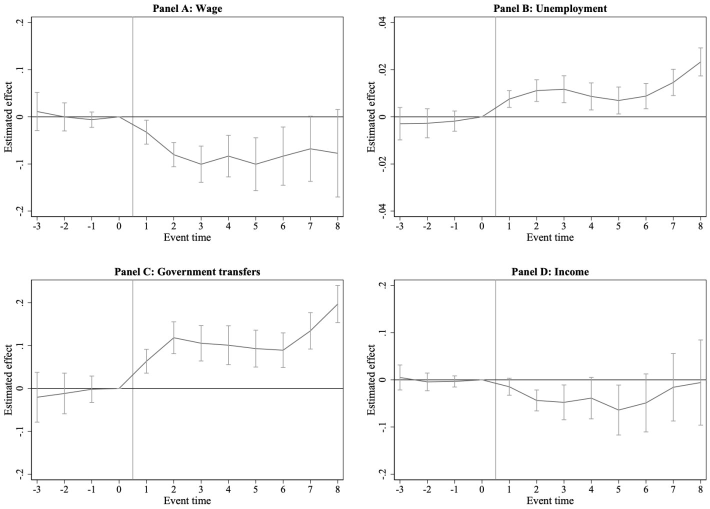
Data

	Treated	Control	Difference	Norm. T-value
	1	2	3	4
Background				
Female	36%	36%	0%	0,00
Immigrant	15%	13%	2%	0,05
Age				
20-33	30%	30%	0%	0,00
34-43	26%	26%	0%	0,00
44-52	24%	24%	0%	0,00
53-60	20%	20%	0%	0,00
Education				
Basic	15%	12%	3%	0,06
High School	53%	53%	0%	0,00
Vocational	16%	16%	0%	-0,01
University	16%	19%	-2%	-0,05
Career				
Wage (thousands SEK)	2710,78	2657,19	53,59	0,02
Labor market experience				
0–5	18%	18%	-1%	-0,01
6–10	11%	11%	0%	0,00
11–20	25%	25%	0%	0,00
21–30	23%	23%	0%	-0,01
30+	23%	22%	1%	0,02
Tenure				
0–2	59%	56%	4%	0,05
3–5	21%	21%	0%	0,00
6–10	15%	19%	-4%	-0,08
11–15	3%	3%	0%	0,00
16+	2%	2%	0%	0,02
Observations	63231	63231		

Individual results

Dependent variable Specification	Wage 1	Unemployment 2	Transfers 3	Income 4
Panel A: Average effect				
Full period	-0.079	0.013	0.119	-0.035
	(-2.96)	(5.16)	(5.82)	(-1.52)
%-change	-7.9%	12.6%	11.9%	-3.5%
Adjusted R^2	0.121	0.072	0.064	0.129
Panel B: Dynamic effect				
Short run (1-2 years)	-0.058	0.011	0.099	-0.029
	(-3.47)	(4.02)	(4.15)	(-2.61)
Medium run (3-4 years)	-0.093	0.012	0.112	-0.043
	(-4.13)	(4.27)	(5.27)	(-2.15)
Long run (5-8 years)	-0.084	0.015	0.135	-0.034
	(-2.23)	(5.27)	(6.09)	(-1.00)
%-change				
Short run	-5.8%	10.7%	9.9%	-2.9%
Medium run	-9.3%	11.5%	11.2%	-4.3%
Long run	-8.4%	14.3%	13.5%	-3.4%
Adjusted R^2	0.123	0.072	0.065	0.131
Mean dependent variable	7.995	0.105	0.723	8.142
Number of observations	1414270	1414270	1414270	1414270

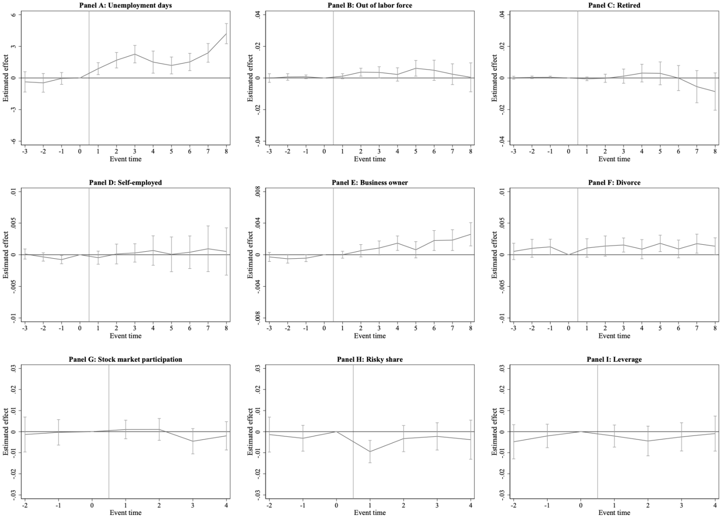
Individual results



Individual results

Dependent variable	Unem. Days	Out of LF	Retirement	Self-employed	Business owner	Divorce	Mortality	Stock market Part.	Risky share	Debt ratio
Specification	1	2	3	4	5	6	7	8	9	10
Panel A: Average effect										
Full period	2.119	0.003	-0.001	0.001	0.001	0.001		-0.001	-0.003	0.000
	(5.99)	(1.33)	(-0.34)	(0.53)	(3.61)	(2.22)		(-0.15)	(-0.88)	(-0.05)
%-change	19.5%	17.6%	-39.5%	19.1%	96.8%	8.3%		-0.1%	-0.7%	0.0%
Adjusted R^2	0.035	0.084	0.221	0.011	0.005	0.002		0.045	0.032	0.108
Panel B: Dynamic effect										
Short run (1-2 years)	1.501	0.002	-0.001	0.000	0.001	0.001		0.002	-0.005	-0.001
	(3.85)	(2.05)	(-0.63)	(0.13)	(1.81)	(1.11)		(0.48)	(-1.64)	(-0.24)
Medium run (3-4 years)	2.100	0.003	0.002	0.001	0.001	0.001		-0.003	-0.002	0.001
	(4.94)	(1.43)	(0.72)	(0.85)	(3.08)	(1.14)		(-0.67)	(-0.32)	(0.12)
Long run (5-8 years)	2.487	0.003	-0.003	0.001	0.002	0.001				
	(6.27)	(1.03)	(-0.60)	(0.49)	(3.52)	(2.34)				
%-change										
Short run	13.8%	13.2%	-23.5%	2.7%	36.9%	6.9%		0.3%	-1.0%	-0.2%
Medium run	19.3%	16.3%	71.7%	25.4%	96.3%	6.8%		-0.4%	-0.3%	0.1%
Long run	22.8%	20.8%	-108.1%	25.2%	131.7%	10.0%				
Adjusted R^2	0.035	0.085	0.221	0.011	0.005	0.002		0.045	0.032	0.108
Mean dependent variable	10.894	0.016	0.003	0.003	0.002	0.008		0.639	0.470	0.560
Number of observations	1414270	1414270	1414270	1414270	1414270	1414270		342554	342554	342554

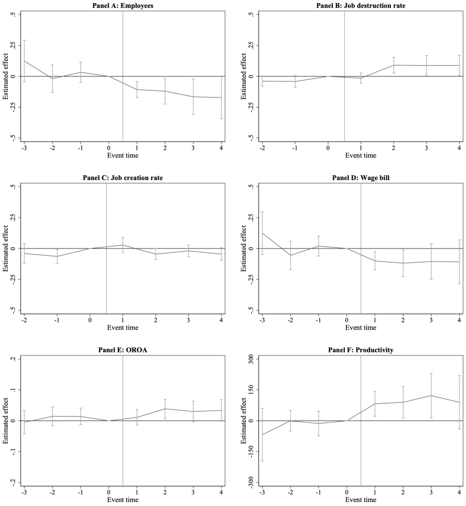
Firm results



Firm results

Dependent variable	Employees	Job destruction	Job creation	Payroll	OROA	Productivity
Specification	1	2	3	4	5	6
Panel A: Average effect						
Full period	-0.163	0.109	-0.025	-0.122	0.021	109.755
	(-2.82)	(4.92)	(-1.40)	(-2.15)	(1.82)	(2.77)
%-change	-16.3%	10.9%	-2.5%	-1.1%	321.3%	35.7%
Adjusted R^2	0.186	0.448	0.721	0.224	0.074	0.072
Panel B: Dynamic effect						
Short run (1-2 years)	-0.141	0.118	0.018	-0.124	0.018	102.247
	(-2.81)	(4.40)	(1.46)	(-2.45)	(1.46)	(2.91)
Medium run (3-4 years)	-0.196	0.098	0.026	-0.122	0.026	122.811
	(-2.32)	(3.23)	(1.69)	(-1.45)	(1.69)	(2.12)
%-change						
Short run	-14.1%	11.8%	-2.7%	-12.4%	273.8%	33.3%
Medium run	-19.6%	9.8%	-2.2%	-12.2%	392.1%	39.9%
Adjusted R^2	0.186	0.448	0.721	0.224	0.074	0.073
Mean dependent variable	0.186	0.072	0.375	11.392	0.007	307.497
Number of observations	4804	4611	4611	4804	4804	4804

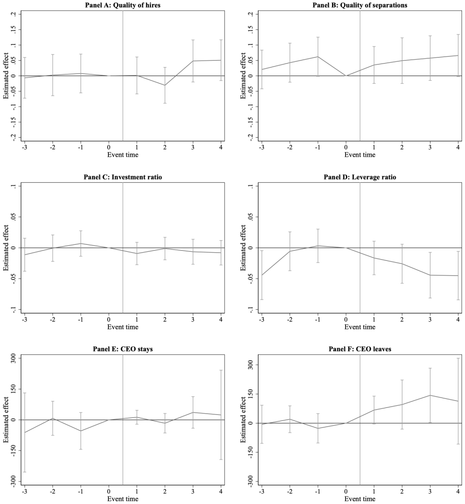
Firm results



Mechanisms

Dependent variable	Quality Hir	Quality Sep	Investment ratio	Leverage	Productivity	Productivity
Sample	Full	Full	Full	Full	CEO remains	CEO replaced
Specification	1	2	3	4	5	6
Panel A: Average effect						
Full period	0.012	0.020	-0.006	-0.022	40.009	98.024
	(0.72)	(1.15)	(-0.93)	(-1.61)	(0.81)	(2.00)
%-change	1.2%	1.9%	-8.4%	-3.5%	12.8%	31.4%
Adjusted R^2	0.004	0.025	0.063	0.108	0.116	0.077
Panel B: Dynamic effect						
Short run	-0.015	0.011	-0.005	-0.012	26.725	81.907
	(-0.78)	(0.52)	(-0.76)	(-0.91)	(0.64)	(1.82)
Medium run	0.048	0.031	-0.007	-0.036	58.086	130.291
	(2.12)	(1.41)	(-0.91)	(-2.02)	(0.77)	(1.65)
%-change						
Short run	-1.5%	1.1%	-7.2%	-1.9%	8.5%	26.2%
Medium run	4.8%	3.0%	-9.9%	-5.7%	18.6%	41.7%
Adjusted R^2	0.006	0.026	0.063	0.108	0.116	0.077
Mean dependent variable	1.011	1.012	0.069	0.630	312.662	312.662
Number of observations	2957	2757	4804	4804	2433	2278

Mechanisms


Panel A: Quality of hires
Panel B: Quality of separations
Panel C: Investment ratio
Panel D: Leverage ratio
Panel E: CEO stays
Panel F: CEO leaves
Estimated effect
Event time

Summary

- Relevant parts:
  1. Describing the data + institutional environment
  2. Identification (”recovering the ATE”) / Policy evaluation
  3. Mechanisms

- Discussion on assumptions:
  - Parallel trends: anticipation effects
  - SUTVA: externalities/GE effects

- Other analyses:
  - Alternative definitions of ”treatment” (partial, share issue, buyers)
  - Accounting for selection effects through attrition
  - Costs vs benefits (policy)

Summary

- What Is the Cost of Privatization for Workers?
  - Wage declines of 5.8-9.3% offset in the long run (5+ years) by government transfers
  - Permanent increases in unemployment (around 12%) and government transfers (up to 8 years) $\rightarrow$ fiscal externality
  - Firm level productivity gains of 35% + employment declines of 16%
  - Concentrated to privatizations involving CEO turnover
- Policy:
  - Indications of positive ”net benefit” to society
  - Mitigating negative effects involve getting workers out of long run unemployment

Applications

Technology Transfer in Mergers and Acquisitions and the Careers of Workers

Gardberg, Heyman and Tåg 2024

- **M&As catalyst for restructuring, labor reallocation and technology adoption**
  - Many countries have ambitions goals on digitization, robots and AI
  - Country and industry differences in technological sophistication
  - Technological transfers through all M&As?
- **Tech change puts pressure on firms and workers to adapt:**
  - can potentially affect many different firm and worker long-run outcomes
  - effects on workers not necessarily homogenous

Motivation

- **M&As vehicle for transferring technologies across firm boundaries**
  $\longrightarrow$ spillovers on workers long-run careers

- **Study workers in Swedish firms acquired by foreign firms w. employer-employee data:**
  - Foreign acquirer heterogeneity: software and database intensity / robot intensity
  - Worker task heterogeneity: exposure to software, AI or robotics

- **Acquisitions from technologically intense regions:**
  - Technology transfers to the acquired firm
  - Disproportionately affects workers performing tasks exposed to acquirer's technological intensity
  - ”Triangulation”: results only in subsamples where the mechanism is in play

Empirical strategy

- Exact individual level cell matching to find control workers
  - Workers in Swedish firms never acquired by a foreign firm
  - Match on occupation, major city residence, firm type (MNE vs local), calendar year
- Time series for each worker stacked into panel with aligned treatment/matching timing
  - Stacked DD: Standard DD regressions with normalized time (as if treatment occurred contemporaneously for all cohorts)
  - Approach avoids problems with staggered TWFE models
  - Stacked DDD: heterogeneity across exposure

Data 1996-2015

467 acquired firms

- Acquirer nationality (Swedish Agency for Economic & Regional Growth)
- General firm data (Statistics Sweden)

$\sim$ 160,000 acquired workers:

- Full-time equivalent real wages, occupation (Salary Structure Statistics)
- Exposure to Software, Robotics and AI (Webb 2022)
    - Overlap of patent and job task descriptions
    - High exposed: workers in 90:th percentile

Data 1996-2015

Acquirer country and industry:

- Software intensity: Software and database capital to total capital (EU Klems)
- Robot intensity: Robot stock to employment (IFR Robot Database)
    - High intensity: If higher than in target industry that year

Stacked regression models

Standard DD of ln wage w of worker i at event year k in calendar year t:

$$w_{ikt} = \alpha_0 + \alpha_1 T_i + \alpha_2 Post_k + \beta DD_{ik} + \omega_t + X_i + X_f + \epsilon_{ikt}$$

To capture dynamic effects we replace $Post_k$ with event time dummies:

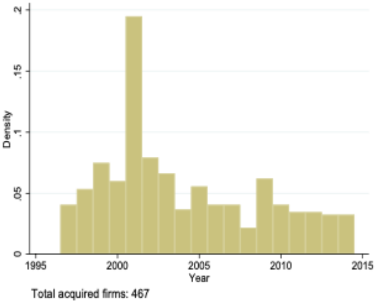
$$w_{ikt} = \alpha_0 + \tau_k + \alpha_1 T_i + \beta_k \sum_{k=-4}^{k=8} \tau_k \times T_i + \omega_t + X_i + X_f + \epsilon_{ikt}$$

DDD estimator for worker heterogeneity:

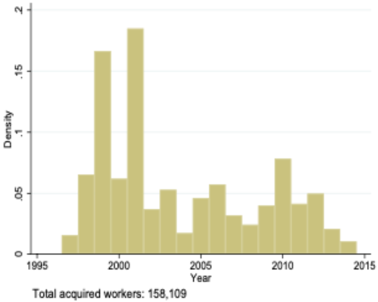
$$\begin{aligned} w_{ikt} = \alpha_0 + \alpha_1 T_i + \alpha_2 Post_k + \beta_1 DD_{ik} + \mu_1 G_i + \mu_2 G_i * T_i + \mu_3 G_i * Post_k + \beta_2 DDD_{ik} \\ + \omega_t + X_i + X_f + \epsilon_{ikt} \end{aligned}$$

Firm level IT expenditures

(A) Acquired Firms per year



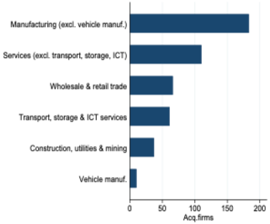
(B) Acquired Workers per year

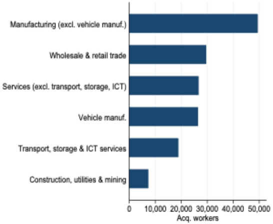


(C) Acquired Firms per Industry

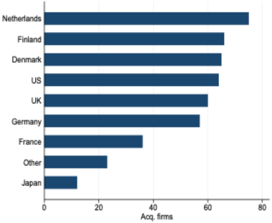
(D) Acquired Workers per Industry

Firm level IT expenditures

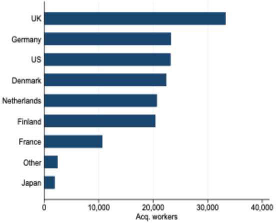




(E) Acquired Firms per Country



(F) Acquired Workers per Country



Firm level IT expenditures

	Treated (1)	Control (2)	Difference (3)	Norm. T-value (4)
Individual variables				
ln wage	9.988	9.980	0.008	0.018
Software exposure	0.541	0.541	0	0.000
Robot exposure	0.512	0.512	0	0.000
AI exposure	0.528	0.528	0	0.000
Age	39.39	40.97	-1.581	-0.128
Education (1-7)	3.712	3.657	0.055	0.028
Experience	20.67	22.33	-1.654	-0.125
Experience^2	513.8	588.4	-74.55	-0.126
Female (%)	0.348	0.341	0.007	0.011
Major city resident (%)	0.693	0.693	0	0.000
Prev. unemp (%)	0.117	0.104	0.013	0.030
$\geq$ 3 year tenure (%)	0.556	0.666	-0.110	-0.161
Firm variables				
ln Firm size	7.158	7.223	-0.065	-0.027
Share high skilled (%)	0.289	0.300	-0.011	-0.034
Swedish MNE (%)	0.524	0.524	0	0.000
VA/L	0.556	0.669	-0.112	-0.135
Observations	158,109	158,109	316,218	

Foreign acquisitions wage effects

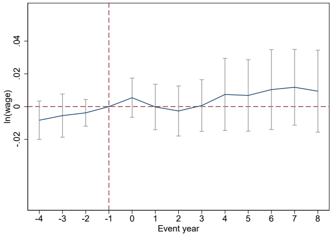


Figure 1: Full sample

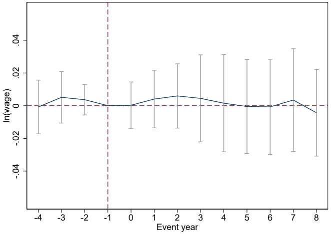


Figure 2: High Software Intensity

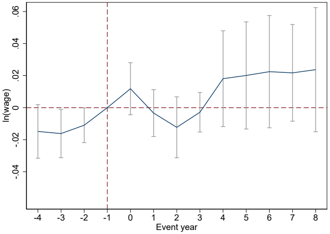


Figure 3: Low Software Intensity

High Software Exposed DDD

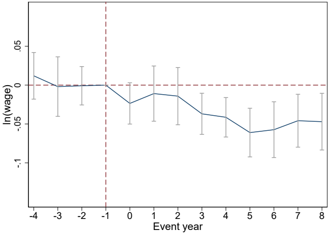


Figure 4: Full sample

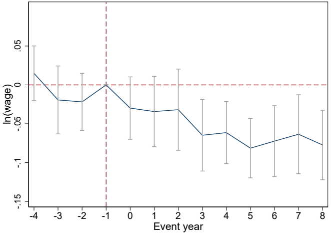


Figure 5: High Software Intensity

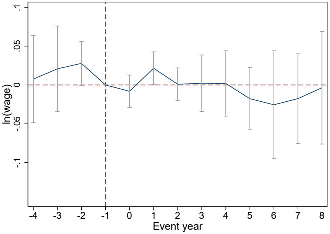


Figure 6: Low Software Intensity

High Software Exposed Subsample

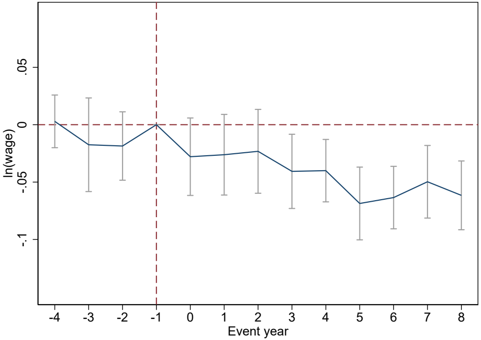


Figure 7: High Software Intensity

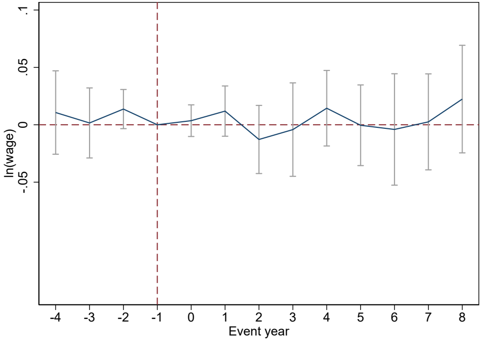


Figure 8: Low Software Intensity

Stayers: High Software Exposed DDD

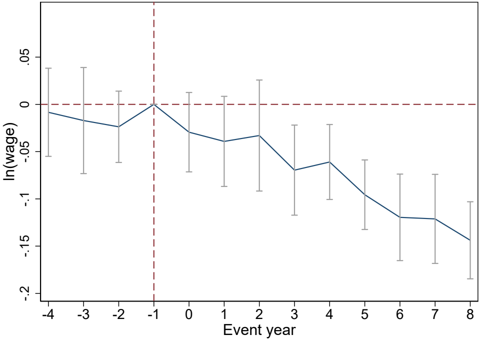


Figure 9: High Software Intensity

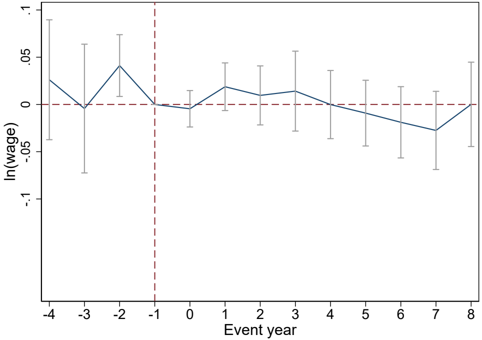


Figure 10: Low Software Intensity

High vs Low Exposed

High Robot Exposure DDD

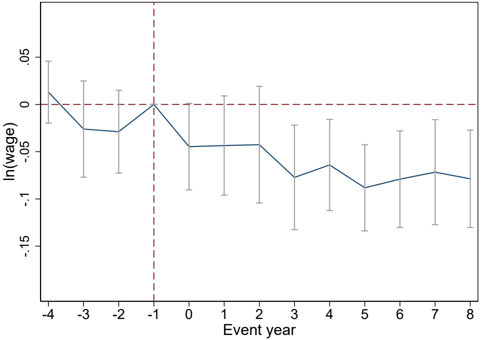


Figure 11: High Robot Intensity

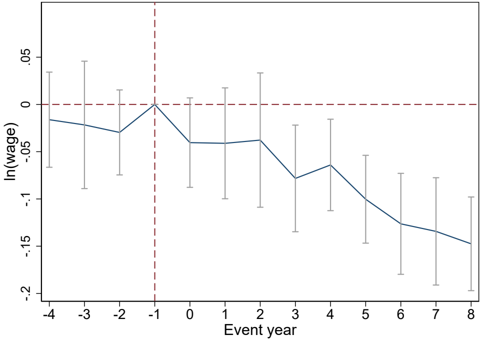


Figure 12: Stayers: High Robot Intensity

High AI Exposure

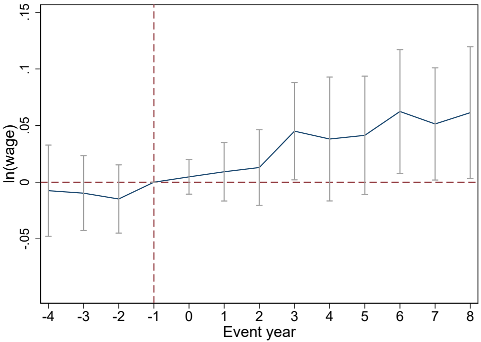


Figure 13: DDD, High Software Intensity

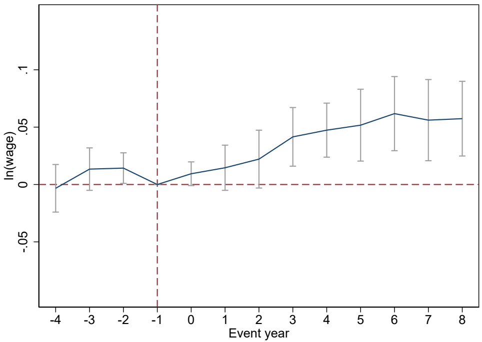
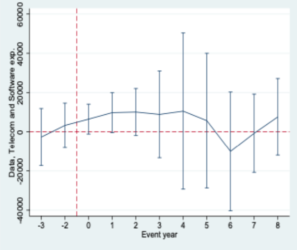


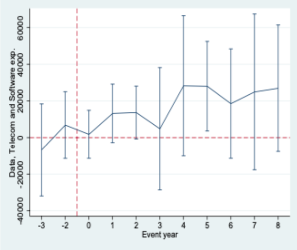
Figure 14: High Intensity & High AI Exposed sample

Firm level IT expenditures

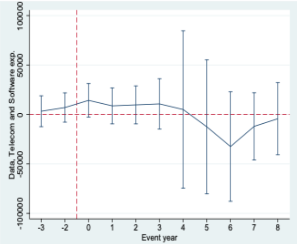
(A) All



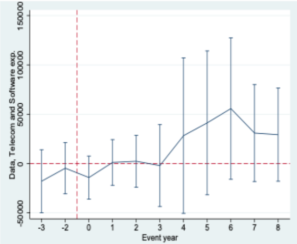
(B) High Intensity



(C) Low Intensity



(D) DDD



Conclusion

- **Technologically intense M&As long-run spillover effects on workers careers**
    - Technological specificity is key
- **Foreign software intense acquisitions lead to:**
    - relative wage losses of 4.2% for software-exposed workers
    - relative wage gains for AI-exposed workers
    - increases investments in software and telecommunications (14.8M SEK)
- **Foreign robotics intense acquisitions lead to:**
    - relative wage losses of 3.5 % for robotics-exposed workers
- **Implications:**
    - Technologically intense acquisitions can spur technological diffusion...
    - ... but heterogeneous effects on workers exposed to the acquirer's technological intensity

Applications

Tolerating Losses for Growth: How US Venture Capitalists Invest Abroad

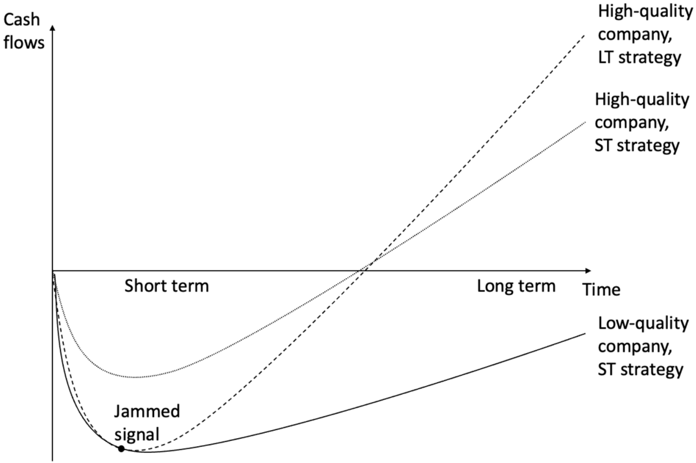
Hellmann, Montag and Tåg 2024

- A fundamental challenge for start-ups is the trade-off between short-term profitability and long-term growth
- Often more ambitious development or growth strategies involve lower short-term profitability (e.g. Spotify, Uber).
- Requires investors that are willing to tolerate prolonged financial losses and imposes financing risk on start-ups (risk of not obtaining follow-on funding)
- Debate in EU about lack of unicorns and VCs that are “playing it too safe”

Motivation

- Question: What determines loss tolerance in VC investing?
  - What are key factors determining loss tolerance?
  - What are the implications for future financing, company growth and exits?
  - Do certain VCs have a more “loss tolerant style” in investing?
- Approach:
  - Develop a theory of loss tolerance to obtain predictions to take to the data
  - Analyze if some VCs have a “loss tolerant” investment style (US vs non-US)

Motivation



Cash flows
High-quality company, LT strategy
High-quality company, ST strategy
Short term
Long term
Time
Low-quality company, ST strategy
Jammed signal

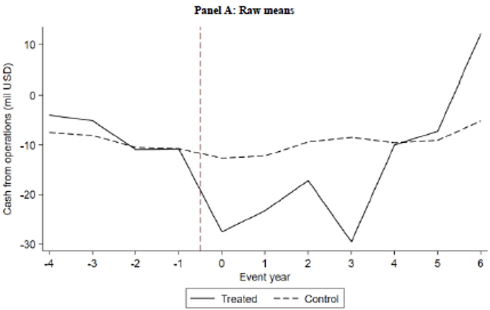
Empirical challenge

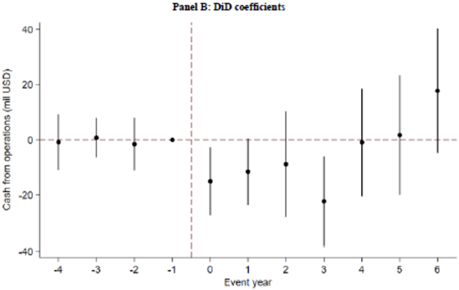
- Challenge: Need a credible measure of financial losses for VC-backed companies
- This paper: Exploit the fact that private Swedish limited liability companies must submit annual reports to Swedish Companies Registration Office by law
  - Construct company-fiscal year panel for companies that ever receive VC funding
  - Compare VC-backed companies that either do or do not also have a US investor
  - Examine relationship between US VC investments and financial losses
  - Show that US VCs have lower financing risk and better exit options

Data

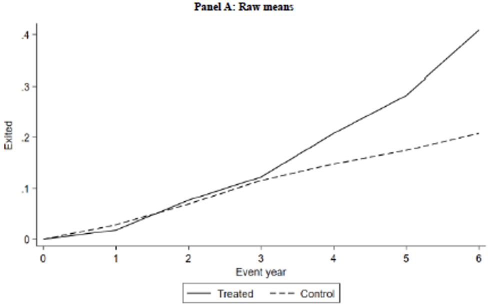
- Principal data source: Swedish Companies Registration Office
  - Annual reports and company events (e.g., bankruptcies)
  - VC investments and exit events from Crunchbase, Pitchbook, ThomsonOne, and Preqin
  - Data on population of Swedish limited liability companies between 1998 and 2020
  - Must submit annual reports to the Companies Registration Office (by law)
- Construct company-fiscal year panel for companies that ever receive VC funding
  - 26,061 company-fiscal years for 2,342 companies with 3,777 VC rounds
  - 14% of companies receive US VC funding, and 12% of rounds include a US VC

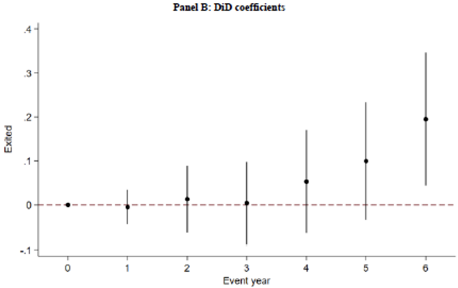
Cash from operation around initial USVC investment





Exit trajectories after initial USVC investment

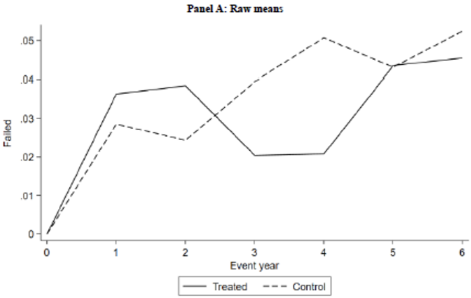


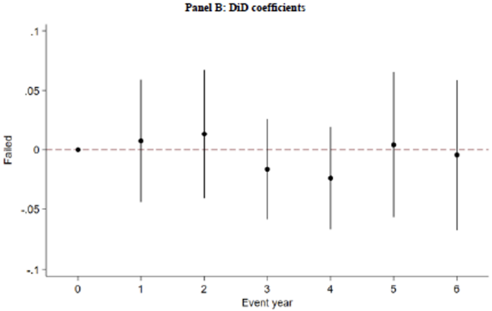


Average value for US VC backed Exit (IPO): $572M ($454M)

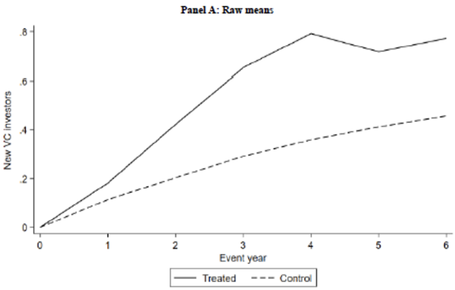
Average value for non-US VC backed Exit (IPO): $220M ($165M)

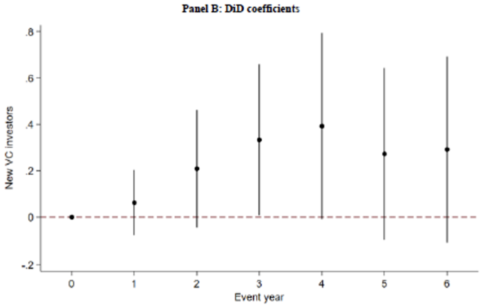
Failure trajectories after initial USVC investment



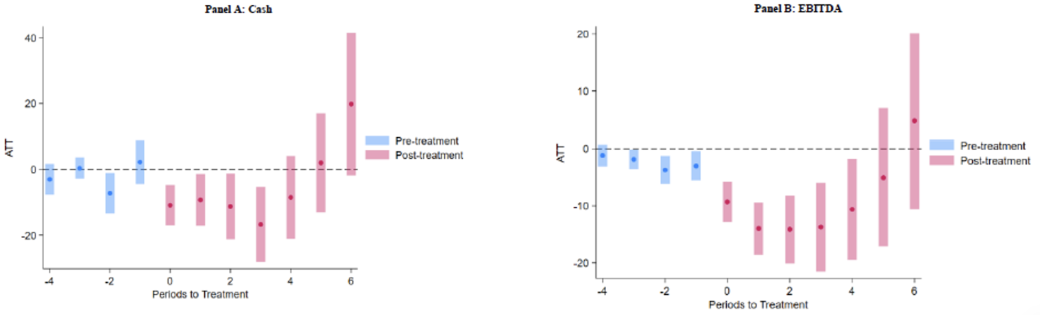


New investors after initial USVC investment





CSDID robustness



No stacking and no concurrent VC funding

Motivation

- Question: What determines loss tolerance in VC investing?
- Approach: Develop a theory of loss tolerance and take the predictions to data using stacked DiD
- Results: Companies with US VC funding:
  - Incur more cumulative losses (higher burn rate), especially in the short run (deeper J-curve)
  - Eventually have better exit outcomes
  - Have same failure rates
- Punchline: Evidence suggests US VCs have a more loss tolerant investment style

Summary

Overview

Theory

Basic idea

Assumptions and DiD regressions

Multiple periods and variation in treatment timing

Applications

Liquidity Provision in Banking Crises

What Is the Cost of Privatization for Workers?

Technology Transfer in Mergers and Acquisitions and the Careers of Workers

Tolerating Losses for Growth: How US Venture Capitalists Invest Abroad

Summary

Summary

Theory

Basic idea

Assumptions and DiD regressions

Multiple periods and variation in treatment timing

Applications

Liquidity Provision in Banking Crises

What Is the Cost of Privatization for Workers?

Technology Transfer in Mergers and Acquisitions and the Careers of Workers

Tolerating Losses for Growth: How US Venture Capitalists Invest Abroad

Summary