

Research Project Report

Student ID Name: XXX

Examining the Impact of Paid Family Leave Policy on Firm's Risk-Taking

1. Introduction

Paid Family Leave (PFL) policies have gained increasing attention for their potential effects on workers and firms. While existing literature often focuses on employee outcomes, the impact on firm behavior—particularly risk-taking—remains less explored. This study examines whether the adoption of state-level PFL policies in the United States influences firms' investment-related risk-taking activities. By combining traditional econometric methods with modern machine learning approaches, I aim to uncover both average and heterogeneous treatment effects across different types of firms.

2. Research Design

This study investigates the causal effect of state-level Paid Family Leave (PFL) policies on firm risk-taking behavior in the United States. To estimate this effect, I use multiple causal inference methods, including Difference-in-Differences (DiD), Callaway and Sant'Anna's staggered DiD, and Double Machine Learning (DML).

2.1 Identification Strategy

My main identification strategy is based on a Difference-in-Differences (DiD) framework. Firms in states that implemented Paid Family Leave (PFL) laws serve as the treated group, while firms in other states serve as the control group. I define a binary treatment indicator $treated$ based on whether a state has implemented PFL by a given year (2022), and a post indicator for years after the policy takes effect. The interaction term $treated \times post$ captures the average treatment effect on the treated (ATT). To address potential bias from staggered PFL adoption across states and years, I also apply the method proposed by Callaway and Sant'Anna (2021), which estimates group-time ATT using not-yet-treated or never-treated units as controls. This approach flexibly accommodates heterogeneous treatment timing and avoids potential bias from standard two-way fixed effects (TWFE).

While the identification strategy relies on the parallel trends assumption, the estimation of treatment effects is enhanced using Double Machine Learning (DML). Following Chernozhukov et al. (2018), I use DML as a flexible estimation tool that controls for high-dimensional covariates and reduces bias.

Several machine learning learners are compared (Lasso, Ridge, Random Forest, XGBoost), and model diagnostics such as cross-fitted R^2 and CVC will be used to guide learner selection. Importantly, DML is applied on top of the identification assumptions provided by the DiD framework, not as a replacement for them.

2.2 Treatment and Outcome Variables

The treatment variable is defined at the state level using the policy effective year (*Effective_Year*). I construct the following variables:

- *treated*: 1 if the state adopts PFL by any year, 0 otherwise.
- *post*: 1 if the firm-year is on or after the effective year, 0 otherwise.
- *interaction_dummy*: product of treated and post.

The outcome variable is firm-level risk-taking, measured as the sum of log-transformed investments:

$$Risk_taking = \log(1 + R\&D) + \log(1 + CapEx) + \log(1 + Acquisition)$$

2.3 Covariates and Moderators

I control for both firm-level and state-level characteristics:

- Firm-level covariates include market value (*ln_mv*), physical capital (*PPE*), Tobin's Q (*tobin_q*), financial leverage (*debt_to_assets*), market competition (*hhi*), industry growth (*cagr_munificence*), and volatility (*growth_volatility*).
- State-level covariates include unemployment rate (*unemp_rate*), GDP (*log_GDP*), and population.

To assess heterogeneous effects, I include two moderators:

- *has_nonus_sub*: 1 if the firm has at least one foreign subsidiary.
- *sme_dummy*: 1 if the firm has fewer than 500 employees.

3. Data

This study combines firm-level panel data from the Compustat database with state-level policy and macroeconomic information to estimate the effect of Paid Family Leave (PFL) adoption on firm behavior in the United States.

3.1 Data Sources

- Firm-level data: Extracted from the Compustat database, including financial and organizational characteristics of publicly listed firms.

- PFL policy data: Manually compiled at the U.S. state level based on the effective year of PFL implementation.
- State-level covariates: Unemployment rate, population, and GDP (log-transformed), sourced from public macroeconomic datasets (e.g., BLS, BEA).

The panel covers years 2002 to 2021, depending on data availability, and includes both treated and untreated firms across multiple industries and states.

3.2 Sample Construction

I merge firm-year observations with state-level variables by the firm's state and calendar year. We exclude firms with missing values for key outcome or treatment variables, and winsorize extreme values for financial covariates at the 5th and 95th percentiles to reduce the impact of outliers.

Treatment assignment is defined at the state level:

- *treated* = 1 if the firm is in a state that has adopted PFL by 2022.
- *post* = 1 if the year is greater than or equal to the state's *Effective_Year*.

Matched samples are created using nearest neighbor matching on pre-treatment covariates to reduce imbalance between treated and control firms.

3.3 Key Variables

- Outcome: *Risk_taking* = $\log(1 + \text{R\&D}) + \log(1 + \text{CapEx}) + \log(1 + \text{Acquisitions})$
- Firm-level covariates:
 - *ln_mv*: Log market value of equity, reflects firm size and market valuation.
 - *PPE*: Property, plant, and equipment; physical capital investment.
 - *tobin_q*, *debt_to_assets*: Financial and structural characteristics — growth opportunities and leverage.
 - *hhi*: Herfindahl index for industry competition.
 - *cagr_munificence*, *growth_volatility*: Industry-level environment — captures market growth trends and uncertainty.
- State-level covariates:
 - *unemp_rate*: Labor market condition.
 - *population*, *log_GDP*: Economic scale and market size proxies.
- Moderators:
 - *sme_dummy*: Firm has <500 employees.
 - *has_nonus_sub*: Firm owns foreign subsidiaries

Table 1: Summary Statistics

Variable	N	Mean	SD	P25	P50	P75
ln_mv	58382	6.37	2.30	4.88	6.48	7.91
PPE	58382	718.43	1,539.84	8.91	68.11	479.85
debt_to_assets	58382	0.25	0.22	0.05	0.21	0.40
tobin_q	58382	1.86	2.01	0.72	1.14	2.04
hhi	58382	2,186.88	1,706.82	859.76	1,733.06	2,954.31
cagr_munificence	58382	0.05	0.05	0.01	0.04	0.08
growth_volatility	58382	0.10	0.06	0.05	0.08	0.13
unemp_rate	58382	6.13	1.93	4.60	5.60	7.30
log_GDP	58382	13.39	0.89	12.79	13.31	14.17
population	58382	15,475,628.18	11,250,237.42	6,403,290.00	11,573,400.00	21,690,325.00

4. Methodology

This study adopts a multi-method framework to estimate the causal impact of Paid Family Leave (PFL) policies on firm-level risk-taking. The core identification strategies are:

1. Difference-in-Differences (DiD): to compare changes in risk-taking between treated and control firms before and after PFL adoption.
2. Callaway & Sant’Anna’s staggered DiD: to address staggered treatment timing and estimate group-time-specific average treatment effects (ATT).

To enhance robustness and reduce bias in estimation:

- I use propensity score matching as a preprocessing step to improve covariate balance prior to estimation.
- I implement Double Machine Learning (DML) to flexibly control for high-dimensional covariates during ATT estimation. DML builds on the identification assumptions above and applies modern machine learning algorithms (Lasso, Ridge, Random Forest, XGBoost) for nuisance parameter estimation.

4.1 Baseline Difference-in-Differences (DiD) and TWFE

I begin with a simple Difference-in-Differences (DiD) model that uses a binary treatment indicator, a post-treatment indicator, and their interaction as the key regressor to estimate the average treatment effect on the treated (ATT):

$$Risk_taking_{it} = \alpha + \beta_1 * interaction_dummy_{it} + \varepsilon_{it}$$

This baseline specification does not include fixed effects or covariates and captures the raw difference in outcomes between treated and control groups before and after the policy change.

I then estimate extended versions of the model with covariates and two-way fixed effects (TWFE) by including firm and year fixed effects, to control for time-invariant heterogeneity and common time trends.

4.2 Difference-in-Differences with Moderators

To test for treatment heterogeneity, I include interactions with moderators:

- SME status (firms with <500 employees)
- International exposure (whether the firm has non-U.S. subsidiaries)

This leads to a triple-differences (DDD) model:

$$Risk_taking_{it} = \alpha + \beta_1 * interaction_dummy_{it} + \beta_2 * moderator + \beta_3 * (interaction_dummy \times moderator) + controls + FE + \varepsilon_{it}$$

The coefficient on the triple interaction term reflects the differential treatment effect across subgroups.

4.3 Matching

To improve covariate balance between treated and control units, I apply nearest neighbor matching using firm-level and state-level covariates. The matched dataset is used in both Callaway & Sant'Anna and the TWFE models.

Balance is assessed using standardized mean differences and visualized with `love.plot()` from the `cobalt` package.

4.4 Callaway & Sant'Anna's DiD

To address issues from staggered treatment timing, I estimate group-time average treatment effects (ATT) using the `att_gt()` function. This method allows us to:

- Account for dynamic treatment timing
- Avoid bias from improper TWFE assumptions
- Use not-yet-treated or never-treated units as valid controls

I estimate ATT for each group and year, and aggregate them into an overall ATT using `aggte()`.

4.5 Double Machine Learning (DML)

To improve causal estimation with high-dimensional covariates, I apply Double Machine Learning (DML) following Chernozhukov et al. (2018). The DML framework uses sample-splitting and cross-fitting to reduce bias from overfitting.

In this study, both the outcome model and treatment model are estimated using machine learning algorithms. I include a set of base learners:

- OLS (unregularized benchmark)
- Lasso and Ridge (regularized linear models)
- Random Forest and XGBoost (non-linear models)

While cross-fitted R^2 and Cross-Validation with Confidence (CVC) are for model selection, they could not be implemented successfully in this setting. Therefore, learner choice is discussed in the results section.

Table 2: Summary of Methodology

Method	Purpose	Key Variables Involved
Baseline DiD	Estimate average treatment effect (ATT)	<i>interaction_dummy</i> , <i>Risk_taking</i>
TWFE DiD	Control for firm/year fixed effects	Same as above + fixed effects + covariates
DDD (moderator)	Test heterogeneity by subgroup	Moderator: <i>sme_dummy</i> , <i>has_nonus_sub</i>
Mathcing (NN)	Improve covariate balance pre-estimation	Matching on: <i>emp</i> , <i>ln_mv</i> , <i>PPE</i> , <i>tobin_q</i> , <i>hhi</i> , etc.
Callaway & Sant'Anna DiD	Address staggered treatment timing	<i>Effective_Year</i> ; covariates
Double Machine Learning	Reduce bias, allow flexible covariate control	<i>Risk_taking</i> , <i>interaction_dummy</i> , high-dim covariates; learners: Lasso, RF, XGBoost, etc.

5. Results

This section presents the estimated effects of Paid Family Leave (PFL) policies on firm-level risk-taking. I report results from several causal identification strategies, including Difference-in-Differences (DiD) and Callaway and Sant'Anna's staggered DiD estimator. To enhance estimation accuracy and robustness, I also implement Double Machine Learning (DML) as a post-modeling technique to flexibly control for covariates using various machine learning algorithms.

Subgroup analyses are conducted to examine treatment heterogeneity based on firm size (SMEs vs. large firms) and international exposure (presence of non-U.S. subsidiaries).

5.1 Baseline DiD and TWFE Results

I begin by estimating the average treatment effect using a Difference-in-Differences (DiD) approach. Model (1), which excludes covariates and fixed effects, shows a positive and statistically significant effect of PFL on firm risk-taking: the estimated coefficient on *interaction_dummy* is 0.652 with a standard error of 0.161 ($p < 0.001$).

Model (2) includes firm- and state-level covariates and confirms the significance of the effect, though the estimate is slightly smaller (0.396) and the model fit improves substantially ($R^2 = 0.502$).

In Models (3) and (4), I introduce firm and year fixed effects (TWFE). The estimated effects become statistically insignificant:

- Model (3) without covariates yields an estimate of -0.071 ($p = 0.3135$), and
- Model (4) with covariates gives an even smaller coefficient of -0.020 ($p = 0.737$).

These results suggest that the observed effect in simple DiD models may be driven by time-invariant firm characteristics or time trends. After accounting for these with fixed effects, there is no robust evidence that PFL adoption affects firm risk-taking on average across all firms.

Table 3: Baseline DiD and TWFE regression results

	(1)	(2)	(3)	(4)
	Baseline DiD	Baseline DiD	TWFE	TWFE
	w/o controls	with controls	w/o controls	with controls
interaction_dummy	0.652*** (0.161)	0.396*** (0.111)	-0.071 (0.070)	-0.020 (0.061)
R-squared	0.003	0.502	0.845	0.865
Controls	No	Yes	No	Yes
Fixed Effects	No	No	Firm × Year	Firm × Year
Observations	58382	58382	58382	58382

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.2 Triple-Differences (DDD) with Moderators

To explore treatment effect heterogeneity, I estimate TWFE-based triple-differences (DDD) regressions by interacting the treatment variable (*interaction_dummy*) with two moderators: international firm status (*has_nonus_sub*) and SME status (*sme_dummy*).

International Firms

Model (1) and (2) examine the differential effect of PFL on international vs. domestic firms. In the model without controls (1), the treatment effect is significantly negative for domestic firms (-0.267 ,

$p < 0.01$), while the interaction term is significantly positive (0.271, $p < 0.05$), suggesting that international firms experience a smaller negative effect—or potentially a positive one.

When controls are added in Model (2), both the main and interaction effects remain directionally consistent, though the interaction effect is no longer statistically significant. This suggests that the differential effect of PFL on international firms becomes weaker and less precise after controlling for covariates.

SMEs

Models (3) and (4) test for effect differences by firm size. In both specifications, the interaction term between *interaction_dummy* and *sme_dummy* is negative and statistically significant (-0.442 , $p < 0.001$ without controls; -0.268 , $p < 0.01$ with controls), indicating that SMEs experience a larger reduction in risk-taking behavior after PFL adoption.

Additionally, the main effect of *sme_dummy* is large and negative, confirming that SME status is independently associated with lower baseline risk-taking. The high R^2 values (~ 0.85) across all models suggest good model fit.

Overall, these results reveal substantial heterogeneity:

- Domestic firms reduce risk-taking more than international firms.
- SMEs show stronger negative treatment effects than larger firms.

Table 4: TWFE DDD Regression Results by International and SME Status

	(1)	(2)	(3)	(4)
Moderator	International	International	SME	SME
interaction_dummy	-0.267** (0.094)	-0.127+ (0.077)	0.082 (0.089)	0.074 (0.079)
has_nonus_sub/sme_dummy	0.410*** (0.061)	0.166** (0.051)	-1.463*** (0.077)	-0.755*** (0.067)
interaction_dummy×moderator	0.271* (0.109)	0.147 (0.090)	-0.442*** (0.099)	-0.268** (0.084)
R-squared	0.845	0.865	0.848	0.865
Controls	No	Yes	No	Yes
Fixed Effects	Firm × Year	Firm × Year	Firm × Year	Firm × Year
Observations	58382	58382	58382	58382

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

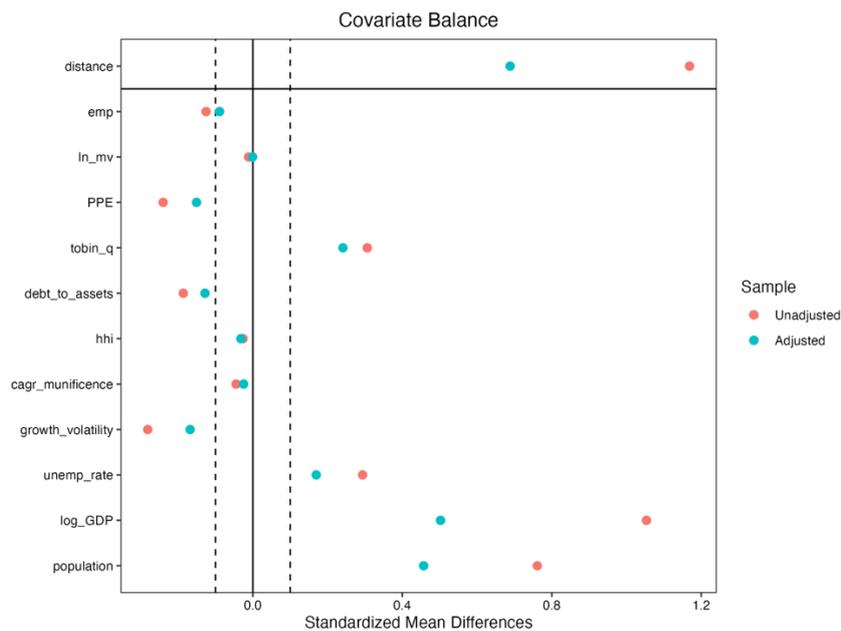
5.3 Matching and Post-Matching DiD

I apply nearest neighbor matching to improve balance between treated and control firms. Post-matching DiD and TWFE estimates confirm the earlier pattern: firms in states with PFL adoption reduce their risk-taking, and this effect is especially strong among SMEs

Table 5: Covariate Balance Before and After Matching

Variables	All Data			Matched Data		
	Means	Means	Std. Mean	Means	Means	Std. Mean
	Treated	Control	Diff	Treated	Control	Diff
distance	0.55	0.26	1.17	0.55	0.38	0.69
emp	5,749.49	7,194.85	-0.13	5,749.49	6,779.56	-0.09
ln_mv	6.35	6.378	-0.01	6.35	6.35	-0.00
PPE	515.70	836.35	-0.24	515.70	716.94	-0.15
tobin_q	2.32	1.59	0.31	2.32	1.75	0.24
debt_to_assets	0.23	0.27	-0.19	0.23	0.26	-0.13
hhi	2,159.26	2,202.95	-0.03	2,159.26	2,212.18	-0.03
cagr_munificence	0.05	0.05	-0.05	0.05	0.05	-0.02
growth_volatility	0.09	0.11	-0.28	0.09	0.10	-0.17
unemp_rate	6.50	5.91	0.29	6.50	6.16	0.17
log_GDP	13.92	13.08	1.05	13.92	13.52	0.50
population	21,890,977.30	11,743,841.97	0.76	21,890,977.30	15,798,953.32	0.46

Figure 1: Covariate Balance (Love Plot)



Post-Matching TWFE Estimation

To further strengthen the causal interpretation, I re-estimate the TWFE model using the matched dataset created via nearest neighbor matching. To accommodate the staggered timing of PFL adoption, I construct a new treatment indicator (*interaction_dummy*) by calculating a subclass-specific hypothetical effective year and coding post-treatment accordingly. Here, *interaction_dummy* serves as a post-matching treatment indicator, constructed as the interaction of group assignment and staggered post-policy timing.

The model includes firm and year fixed effects and controls for a full set of covariates. Table 6 presents the estimation results. The coefficient on *interaction_dummy* is -0.016 with a standard error of 0.062 and is not statistically significant ($p > 0.1$).

Compared to the TWFE model without matching (Model 4 in Table 3), which also yields a small and insignificant negative estimate (-0.020), the results remain consistent. This reinforces the conclusion that there is no robust evidence of a significant average treatment effect of PFL on firm risk-taking, even after improving covariate balance through matching.

Table 6: TWFE with Covariates Using Matched Data

	Risk_taking
interaction_dummy	-0.016
R-squared	0.876
Controls	Yes
Fixed Effects	Firm \times Year
Observations	42942

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.4 Callaway & Sant’Anna DiD Estimates

To better account for staggered treatment timing, I apply the Callaway and Sant’Anna (2021) Difference-in-Differences estimator using the matched dataset. This method accommodates variation in treatment adoption timing and allows estimation of dynamic treatment effects across event time.

In this design, the “not-yet-treated” units serve as the control group, and I use the doubly robust estimation approach to estimate the average treatment effect on the treated (ATT). Due to convergence issues when including covariates after matching, I report estimates based on matched data without additional covariates. Matching already ensures covariate balance, reducing the need for further adjustment in the doubly robust framework. Attempts to include covariates led to instability due to limited within-group variation after matching.

The estimated aggregated ATT is 0.34 with a standard error of 0.44, yielding a 95% confidence interval of $[-0.52, 1.12]$ (Table 7). Although the point estimate is positive, it is not statistically significant. This suggests that, in the matched sample, PFL policy has no clear average effect on firm risk-taking.

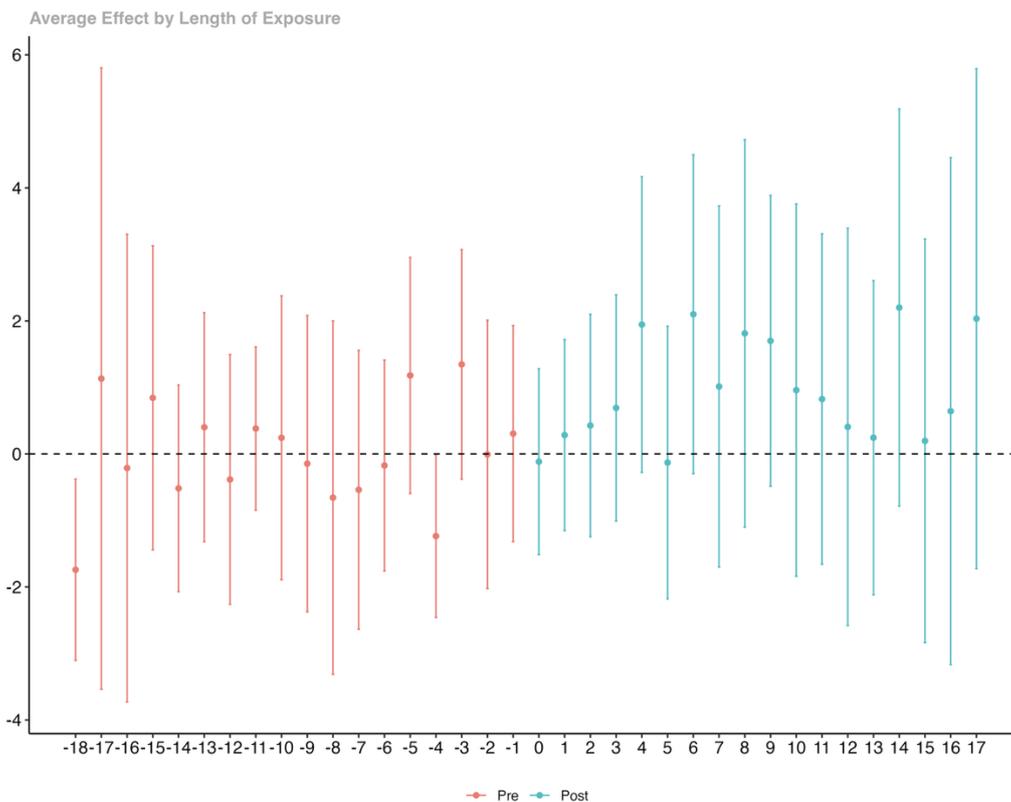
Table 7: Callaway & Sant’Anna Aggregated ATT Estimates Using Matched Data

ATT	Std. Error	95% CI Lower	95% CI Upper
0.3358	0.4353	-0.5174	1.189

To further explore dynamics, I estimate ATT by event time. Figure 2 plots these estimates. During the pre-treatment period (event time -18 to -3), ATT estimates are consistently close to zero with wide confidence intervals, supporting the parallel trends assumption.

In the post-treatment period (event time > 0), the ATT estimates become slightly more positive, but the confidence intervals remain wide, indicating limited statistical power and no clear evidence of a sustained treatment effect.

Figure 2: Dynamic ATT Plot by Event Time



5.5 Double Machine Learning (DML) Results

To flexibly estimate the treatment effect while controlling for high-dimensional covariates, I apply Double Machine Learning (DML) using five base learners: OLS, Lasso, Ridge, Random Forest, and XGBoost. While DML does not constitute an identification strategy itself, it strengthens inference under a valid identification assumption by reducing bias from overfitting and controlling for confounders flexibly.

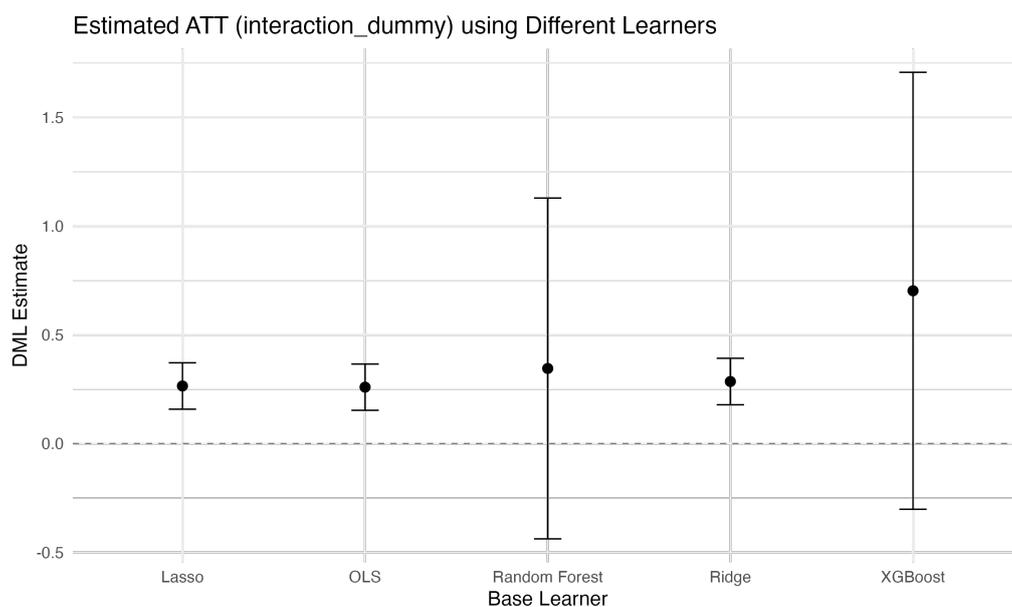
The results suggest that Lasso and Ridge produce relatively stable and statistically significant estimates of the treatment effect (*interaction_dummy*), with OLS results being comparable. In contrast, estimates from Random Forest and XGBoost are more variable and statistically insignificant, reflecting greater instability under the current specification.

Table 8 summarizes the ATT estimates across learners, and Figure 3 visualizes the estimates with 95% confidence intervals.

Table 8: DML results for each learner (Estimate, SE, CI, p-value)

	Estimate	Std. Error	CI_Lower	CI_Upper	p-value
OLS	0.2588206	0.05423520	0.1525196	0.3651216	0.0000
Lasso	0.2641986	0.05439962	0.1575753	0.3708219	0.0000
Ridge	0.2847563	0.05453546	0.1778668	0.3916458	0.0000
Random Forest	0.3447752	0.39964880	-0.4385364	1.1280869	0.3883
XGBoost	0.7018673	0.51243205	- 0.3024995	1.7062341	0.1708

Figure 3: Estimated ATT with 95% CI by base learner



Due to technical limitations, I was unable to compute cross-fitted R^2 and Cross-Validation with Confidence (CVC) to guide model selection. These diagnostic metrics are important for evaluating the predictive performance and reliability of learners, especially in high-dimensional settings. Future work may incorporate these metrics to improve learner choice and model robustness.

Table 9: DML Estimate with Fixed Effects (Lasso)

Learner	Estimate	Std. Error	CI_Lower	CI_Upper	p-value
Lasso (w/ FE)	-0.02170977	0.0294843	-0.07949901	0.03607946	0.4615382

6. Discussion & Conclusion

This study examines the impact of Paid Family Leave (PFL) policies on firms' risk-taking behavior using causal inference methods, including Difference-in-Differences (DiD), Two-Way Fixed Effects (TWFE), and the staggered DiD estimator by Callaway & Sant'Anna. While DiD and TWFE yield mixed and sometimes statistically insignificant results, methods incorporating matching and moderator analysis reveal stronger heterogeneous effects, particularly among domestic firms and small and medium enterprises (SMEs). Although matching improves covariate balance, the Callaway & Sant'Anna estimator remains statistically inconclusive. DML, used as a flexible estimation tool rather than an identification strategy, consistently detects stable negative effects across learners like Lasso and Ridge. However, results remain insignificant when firm fixed effects are incorporated.

Overall, the findings suggest that PFL policies may influence firm-level risk-taking, but the effects are not uniform and depend on firm characteristics such as size and global exposure. Future work may benefit from richer firm-level data and more advanced diagnostics (e.g., cross-fitted R^2) to enhance model evaluation and treatment effect estimation.

References

- Ahrens, A., Chernozhukov, V., Hansen, C., Kozbur, D., Schaffer, M., & Wiemann, T. (2025). An Introduction to Double/Debiased Machine Learning. *arXiv preprint arXiv:2504.08324*.
- Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of econometrics*, 225(2), 200-230.
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters.
- Dube, A., Jacobs, J., Naidu, S., & Suri, S. (2020). Monopsony in online labor markets. *American Economic Review: Insights*, 2(1), 33-46.
- Hsiao, A. Y.-C. (2023). Examining the impact of paid family and medical leave policy on firm innovation: Evidence from quasi-experiments [Master's thesis, National Tsing Hua University].