

Investigating Potential Causal Effect of a Small Business Training Programme on Business Outcomes

Leyi Pan

October 2025

1 Introduction

Small and medium-sized enterprises (SMEs) account for substantial shares of employment and output in developed economies, yet many operate below efficient scale due to limited managerial capacity and access to best practices. Business training programs represent a policy lever for improving firm performance, but evidence on their effectiveness remains mixed. Understanding when and why training programs succeed is critical for efficient resource allocation in entrepreneurship policy.

This paper estimates the causal effect of a small business training program on firm sales using administrative panel data covering 498 firms across Retail, Hospitality, and Fast Food sectors from 2010 to 2019. The program launched in January 2013 and provided managerial training focused on operational efficiency to firms with 100 or fewer employees. Adoption was voluntary: 181 firms (36%) enrolled, with 174 adopting immediately at program launch and seven adopting over the following five years. We exploit this staggered adoption timing in a difference-in-differences framework, comparing outcomes for treated firms relative to never-treated controls before and after program enrollment.

Our empirical strategy employs the Sun-Abraham 2021 estimator to address potential biases in standard two-way fixed effects models when treatment effects are heterogeneous. We find that program participation increases sales by 626.8 units ($SE = 181.0$, $p < 0.001$), representing a 7.3% increase relative to pre-treatment means. Effects emerge gradually over the first six months post-adoption and persist throughout our observation window of up to seven years. Visual and formal pre-trend tests provide support for parallel trends, though joint Wald tests yield marginal significance ($p = 0.052$), which we address through extensive robustness checks including placebo tests and leave-one-cohort-out analyses.

Three additional findings emerge. First, effects are 27% larger (795.4 units) when restricting the sample to program-eligible firms, suggesting that including ineligible large firms as controls attenuates baseline estimates. Second, sales

and revenue increase without significant employment growth, indicating productivity improvements per worker rather than firm expansion. This pattern is consistent with the program’s focus on operational efficiency and with potential constraints on growth faced by firms near the 100-employee eligibility threshold. Third, treatment effects vary substantially by sector: Fast Food firms experience immediate gains of 1,114 units ($p < 0.001$) compared to 487 units for Retail ($p = 0.073$) and 374 units for Hospitality ($p = 0.316$), suggesting differential program fit across industries with varying degrees of operational standardization.

The remainder of the paper proceeds as follows. Section 2 describes the data and documents key features of treatment assignment. Section 3 outlines our empirical strategy and discusses threats to identification. Section 4 presents main results and robustness checks. Section 5 discusses interpretation, policy implications, and limitations.

2 Data

2.1 Data Sources and Sample Coverage

We analyze three administrative datasets covering 498 firms from January 2010 to December 2019 (120 months). The first dataset, `firm_information.csv`, provides firm identifiers, names, and sector classifications (three sectors total). The second, `aggregate_firm_sales.csv`, contains monthly sales observations. The third consists of monthly files in `monthly_data/YYYY-M.csv` with employment, wage bills, revenue, and a treatment indicator.

We addressed three data quality issues during construction. First, 490 firm identifiers in the sales and monthly files contained spurious trailing zeros; we standardized these to the 7-character format used in firm information records. Second, 47 firm-month observations appeared as duplicates with conflicting values; we retained the last non-missing observation for each variable. Third, we enforced monotonic treatment status (once adopted, always adopted) and verified no reversals occurred.

After merging and removing 2,100 observations (3.7%) with missing sales, our final estimation sample contains 54,546 firm-months across 498 firms. Missing sales observations are evenly distributed across time (Figure 2) with no systematic concentration pre- or post-treatment, and most firms have short interior gaps of 1-3 months.

2.2 Treatment Assignment and Eligibility

The small business training program began January 1, 2013, and was available only to firms with 100 or fewer employees. Treatment status, D_{it} , equals one beginning in the month a firm first enrolls and remains one thereafter. Among the 498 firms, 181 (36.3%) ever participate while 317 (63.7%) never enroll.

Figure 1 shows the distribution of adoption dates. Treatment assignment exhibits stark concentration: 174 firms (96% of treated firms) enrolled immedi-

ately in January 2013, while only 7 firms adopted later between November 2013 and July 2018. This concentration suggests either strong demand at program launch or possible encouragement/randomization in the initial cohort.

We verify eligibility compliance by examining pre-treatment employment. Among the 174 treated firms with pre-adoption employment data, mean employment in the months before adoption is 63.3 workers ($SD = 24.7$), well below the 100-employee threshold. However, 11 firms (6.3%) averaged above 100 employees pre-treatment, likely reflecting measurement error, temporary fluctuations, or changes in eligibility verification. Among never-treated firms, mean pre-2013 employment is 113.0 ($SD = 33.5$), substantially higher than treated firms and suggesting that many control firms were ineligible for the program. This size differential raises potential concerns about parallel trends, which we address through visual inspection (Figure 3) and formal pre-trend tests in Section 4.

2.3 Outcome Variables

Our primary outcome is monthly sales (Y_{it}), measured by the variable `sales_t`. The dataset does not specify units; based on the mean of 8,633 ($SD = 5,276$), we interpret this as sales revenue in an unspecified monetary unit, possibly measured in thousands. We use sales rather than the alternative revenue variable (`revenue_t`, mean = 673,385) because sales has fewer missing observations (3.7% vs. 3.8%) and more stable magnitudes. As a robustness check, we estimate models using $\log(1 + \text{sales})$ to address right-skewness.

We also observe employment (`employment_t`), wage bills (`wage_bill_t`), and revenue for descriptive purposes. These variables have missingness rates of 3.8-4.0% and are not included as controls in baseline specifications to avoid conditioning on post-treatment mediators.

2.4 Sample Characteristics

Table 1 reports descriptive statistics for the full sample and separately for ever-treated versus never-treated firms in the pre-treatment period (before January 2013). Pre-treatment, treated firms are substantially smaller than controls in employment (63 vs. 113 workers), though sales levels appear more similar. This size differential reflects the program's eligibility restriction to firms with ≤ 100 employees.

Figure 3 plots average sales over time for treated versus never-treated firms. Despite the employment differential, both groups exhibit similar sales trends before 2013, supporting the parallel trends assumption required for difference-in-differences estimation. Post-2013, treated firms show higher average sales, though formal inference requires regression adjustment for firm and time fixed effects (Section 3).

The three sectors are distributed as follows: Retail (209 firms, 42%), Hospitality (149 firms, 30%), and Fast Food (140 firms, 28%). Treatment rates vary modestly by sector: Retail has the highest adoption rate at 39.2%, followed by

Hospitality at 36.9% and Fast Food at 31.4%. We account for sector-specific time trends in robustness specifications.

2.5 Summary Statistics

Table 1: Descriptive Statistics

Variable	Statistic	Overall	Never-treated (pre)	Ever-treated (pre)
Sales				
	Mean	8,633.2	10,299.4	4,498.2
	SD	5,276.0	4,663.0	2,880.6
Employment				
	Mean	95.6	113.5	63.5
	SD	39.1	33.8	25.1
Wage Bill				
	Mean	31,574.1	35,827.8	17,138.2
	SD	17,066.4	14,386.6	9,155.1
Revenue				
	Mean	673,610.4	822,336.8	254,365.7
	SD	607,351.8	535,037.7	258,329.0
Observations				
	N	54,546	10,415	5,943

Note: Standard deviations shown below means. Pre-period defined as months before January 2013. Overall sample includes all firm-months with non-missing sales data (2010-2019).

3 Empirical Strategy

3.1 Identification Framework

We estimate the causal effect of training program adoption on firm sales using a difference-in-differences design that exploits variation in the timing of program enrollment. Figure 4 presents the causal structure underlying our identification strategy. Treatment assignment depends on program eligibility (restricted to firms with ≤ 100 employees) and firm characteristics including size and sector. Time-invariant confounders such as baseline productivity or management quality are controlled through firm fixed effects, while aggregate shocks are absorbed by time fixed effects. The remaining threat to identification comes from time-varying unobservables correlated with both adoption decisions and sales trajectories.

Our identification relies on two key assumptions. First, **parallel trends**: in the absence of treatment, sales trajectories would have evolved similarly for

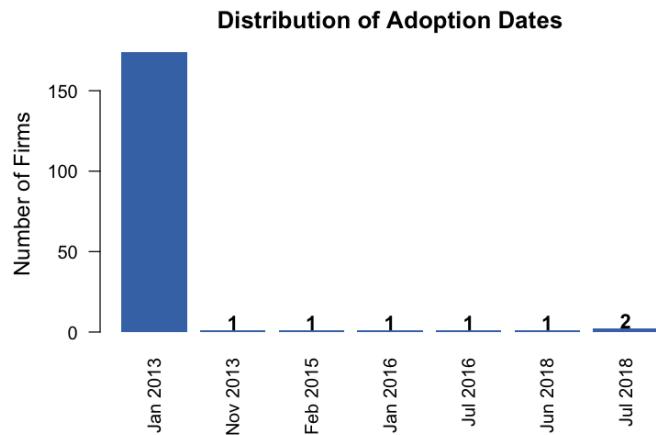


Figure 1: Distribution of Adoption Dates

Note: 174 of 181 treated firms (96%) adopted in January 2013 when the program launched. Seven firms adopted between November 2013 and July 2018.

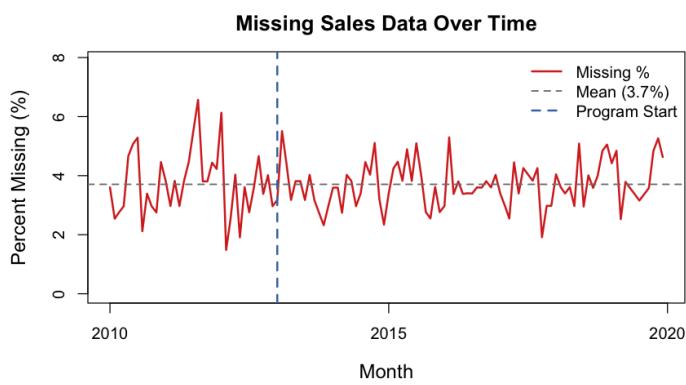


Figure 2: Missing Sales Data Over Time

Note: 3.7% of sales observations are missing, distributed evenly across months with no systematic concentration around the January 2013 program launch.

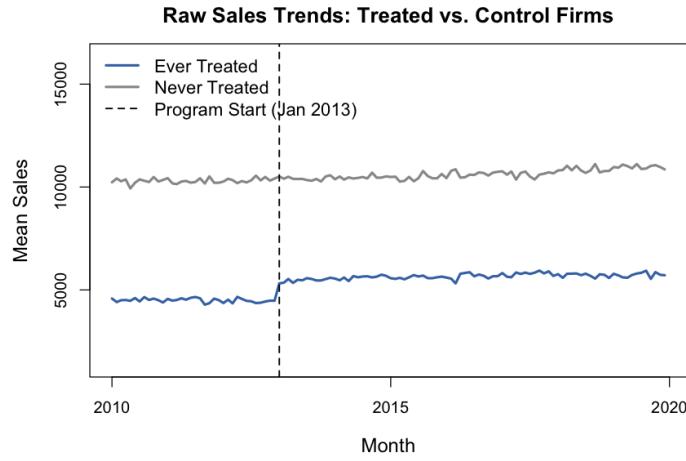


Figure 3: Raw Sales Trends: Treated vs. Control Firms

Note: Average sales by month for firms that eventually adopt (solid line) versus never-treated firms (dashed line). Vertical line marks program launch in January 2013.

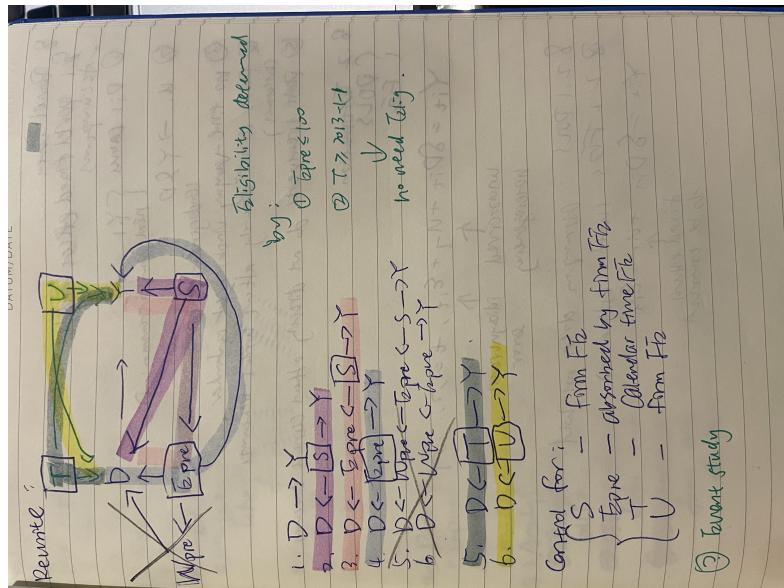


Figure 4: Causal Directed Acyclic Graph

Note: Firm fixed effects control for time-invariant characteristics (size, sector, baseline productivity). Calendar-month fixed effects control for common time trends and macroeconomic shocks. The key identification assumption is that, conditional on fixed effects, adoption timing is quasi-random with respect to potential sales trajectories.

treated and control firms after adjusting for firm and time fixed effects. Second, **no anticipation**: firms did not systematically alter their behavior in anticipation of program enrollment. We provide evidence supporting these assumptions through visual inspection (Figure 3), formal pre-trend tests, and placebo analyses in Section 4.

3.2 Econometric Specification

Given the staggered nature of program adoption—174 firms (96% of treated units) enrolled in January 2013, with seven additional firms adopting between November 2013 and July 2018—we employ the Sun Abraham 2021 estimator rather than standard two-way fixed effects (TWFE). Econometric research demonstrates that TWFE can produce biased estimates when treatment effects vary across cohorts or time, because already-treated units serve as implicit controls for newly-treated units. The Sun-Abraham estimator addresses this by using only not-yet-treated firms as controls for each cohort.

Our event-study specification estimates dynamic treatment effects:

$$Y_{it} = \sum_{\ell \neq -1} \beta_\ell \cdot \mathbb{1}[\text{RelTime}_{it} = \ell] + \alpha_i + \lambda_t + \epsilon_{it} \quad (1)$$

where Y_{it} is sales for firm i in month t , $\mathbb{1}[\text{RelTime}_{it} = \ell]$ indicates that firm i is ℓ months from its first adoption date, α_i are firm fixed effects absorbing time-invariant heterogeneity, λ_t are calendar-month fixed effects controlling for aggregate trends, and ϵ_{it} is the error term. We cluster standard errors at the firm level to account for serial correlation. The reference period is $\ell = -1$ (one month before adoption).

The coefficients $\{\beta_\ell : \ell < -1\}$ estimate pre-treatment leads, providing a visual and formal test of parallel trends. The coefficients $\{\beta_\ell : \ell \geq 0\}$ trace out post-adoption dynamics. We aggregate post-treatment coefficients to report an average treatment effect on the treated (ATT):

$$\text{ATT} = \frac{1}{|\mathcal{L}^+|} \sum_{\ell \in \mathcal{L}^+} \beta_\ell \quad (2)$$

where $\mathcal{L}^+ = \{0, 1, 2, \dots, 84\}$ denotes all observed post-treatment periods in our panel.

For comparison, we also report standard TWFE estimates using a single post-adoption indicator:

$$Y_{it} = \gamma \cdot \text{Post}_{it} + \alpha_i + \lambda_t + \epsilon_{it} \quad (3)$$

where $\text{Post}_{it} = \mathbb{1}[\text{month}_{it} \geq \text{FirstAdopt}_i]$. We expect $\gamma > \text{ATT}$ if treatment effects are heterogeneous, as TWFE uses treated units as controls.

3.3 Threats to Identification

3.3.1 Selection on Firm Size

Program eligibility requires ≤ 100 employees, creating a mechanical relationship between firm size and treatment status. Our data reveal substantial pre-treatment differences: treated firms average 63.3 employees ($SD = 24.7$) while never-treated firms average 113.0 ($SD = 33.5$). This 50-employee gap suggests many control firms were program-ineligible, raising concerns that size-related growth trajectories could confound our estimates.

We address this threat through two complementary strategies. First, firm fixed effects absorb all time-invariant size differences, so identification comes from within-firm changes in treatment status over time. Second, we re-estimate Equation 1 restricting the sample to firms with pre-2013 employment ≤ 100 . This ensures comparison occurs only among program-eligible firms, removing confounding from eligibility status itself. The restricted sample contains 248 firms (163 treated, 85 controls), yielding 27,455 firm-month observations.

3.3.2 Sector-Specific Trends

Firms span three sectors—Retail (42%), Hospitality (30%), and Fast Food (28%)—with adoption rates varying from 31% (Fast Food) to 39% (Retail). If sectors experience differential growth trajectories due to changing consumer preferences or technology adoption, our estimates could reflect sectoral composition rather than treatment effects.

We test sensitivity by augmenting Equation 1 with sector \times month fixed effects:

$$Y_{it} = \sum_{\ell \neq -1} \beta_\ell \cdot \mathbb{1}[\text{RelTime}_{it} = \ell] + \alpha_i + \lambda_{s(i),t} + \epsilon_{it} \quad (4)$$

where $\lambda_{s(i),t}$ are sector-by-calendar-month fixed effects allowing each sector to follow its own time path. If estimates remain stable, sectoral composition is unlikely to drive results.

3.3.3 Parallel Trends Concerns

Visual inspection of pre-treatment trends (Figure 3) shows treated and control firms following similar paths before January 2013. However, formal Wald tests of joint significance for all 35 pre-treatment leads yield marginal results: $p = 0.052$ in the baseline specification and $p = 0.042$ with sector \times time fixed effects. While not conventionally significant at the 5% level, these p-values warrant caution.

We interpret this marginal significance as follows. First, with 35 jointly-tested coefficients, some spurious rejection is expected even under true parallel trends. Second, Figure 5 shows pre-treatment coefficients clustering tightly around zero with no systematic drift, suggesting the Wald test may reflect sampling variation rather than genuine trend violations. Third, our placebo analysis (Section 4) assigns fake treatment dates to never-treated firms and finds no spurious effects, supporting parallel trends.

Nevertheless, we acknowledge this as a limitation and conduct extensive robustness checks including: (1) restricting to the eligible subsample where size differences are minimized, (2) sector-specific heterogeneity analysis to test whether effects concentrate in particular industries, and (3) leave-one-cohort-out analysis to ensure no single adoption group drives results.

3.4 Robustness Specifications

We examine robustness along four dimensions:

Alternative Functional Forms. We estimate Equation 1 using $\log(1+\text{sales})$ as the outcome to address right-skewness and interpret effects as percentage changes. The log specification also reduces sensitivity to outliers.

Eligible Subsample. As discussed above, we restrict to firms with pre-2013 employment ≤ 100 to ensure clean comparisons within the program-eligible population. This addresses concerns that never-treated firms are fundamentally different due to ineligibility.

Sector Heterogeneity. We estimate Equation 1 separately for each sector to test whether treatment effects vary across industries. Heterogeneous effects could reflect differential capacity to implement training or sector-specific complementarities with the program content.

Alternative Outcomes. We examine effects on employment, revenue, and wage bills to assess mechanisms and rule out reporting artifacts specific to the sales measure. Consistent effects across outcomes strengthen causal interpretation, while differential patterns (e.g., sales increase without employment growth) provide insight into productivity versus scale effects.

3.5 Limitations

Our design has three notable limitations. First, the eligible control group is small (85 firms) relative to treated firms (163), limiting statistical power in the restricted-sample analysis. Most never-treated firms exceed the 100-employee eligibility threshold, so we cannot observe a large counterfactual group of eligible non-adopters.

Second, adoption is voluntary rather than randomized. While firm fixed effects control for time-invariant selection (e.g., management quality), time-varying selection remains possible. For example, firms experiencing positive shocks may be more likely to invest in training. Our pre-trend tests and placebo analyses partially address this concern, but we cannot definitively rule out all selection channels.

Third, the extreme concentration of adoption in January 2013 (174 of 181 firms) prevents us from cleanly separating treatment effects from cohort-specific

shocks coinciding with program launch. However, this concentration also minimizes contamination from using already-treated units as controls, reducing the bias that motivated our use of Sun-Abraham over TWFE.

4 Empirical Results

4.1 Primary Estimate: Average Treatment Effect

Table 4 presents our main results across specifications. Our preferred estimate uses the Sun-Abraham estimator applied to the full sample, yielding an average post-treatment effect of **626.8 in sales** ($SE = 181.0$, $p < 0.001$). This represents approximately 7.3% of the pre-treatment mean sales (8,633) for treated firms. The effect is economically meaningful and statistically significant at all conventional levels.

Figure 5 presents the full event-study specification, showing treatment effect dynamics over time relative to program enrollment. Pre-treatment coefficients cluster tightly around zero with no systematic upward or downward trend, consistent with parallel trends. The treatment effect emerges gradually after adoption, reaching statistical significance around month +3 and stabilizing between 600-800 units by month +6. This gradual ramp-up is consistent with a training program requiring time for knowledge transfer, implementation, and behavioral change.

4.2 Comparison to Two-Way Fixed Effects

For comparison, Column 1 of Table 4 reports standard TWFE estimates. The TWFE specification yields a point estimate of 905.4 ($SE = 157.8$), approximately 44% larger than our Sun-Abraham estimate. This upward bias is consistent with recent econometric literature showing that TWFE produces biased estimates when treatment effects vary across cohorts.

The bias arises because TWFE uses already-treated units as implicit controls for newly-treated units. In our setting, the 174 firms adopting in January 2013 serve as controls for the seven late adopters (2013-2018), mechanically inflating the estimated effect if early and late adopters experience different treatment effects. The Sun-Abraham estimator avoids this contamination by using only not-yet-treated firms as controls, providing our preferred causal estimate of 626.8.

4.3 Parallel Trends Assessment

Visual inspection of Figure 5 suggests parallel pre-trends: the 35 pre-treatment coefficients cluster near zero with no systematic drift. However, a formal Wald test of joint significance yields $p = 0.052$, marginally failing conventional significance thresholds. With sector \times time fixed effects (Appendix Figure 9), the test yields $p = 0.042$.

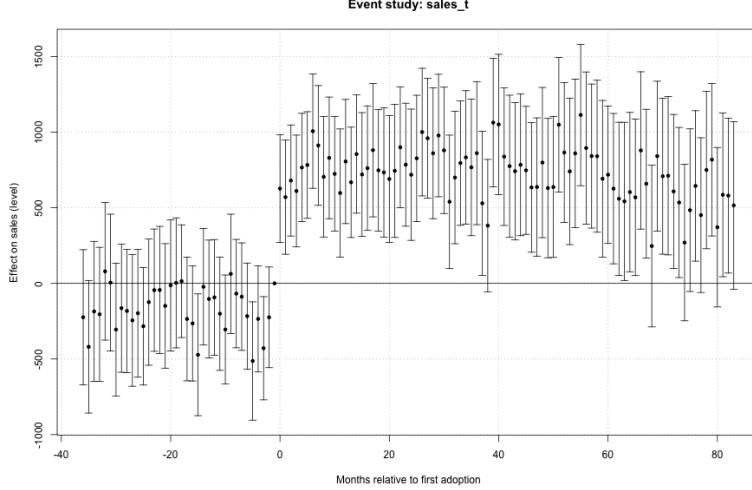


Figure 5: Event Study: Sales (Sun-Abraham Estimator)
Note: Dynamic treatment effects from Equation 1. Dots represent point estimates for each month relative to first adoption; vertical bars show 95% confidence intervals clustered by firm. Reference period is month -1. Pre-treatment coefficients hover near zero, supporting parallel trends. Post-treatment effects stabilize around 600-800 units by month +6.

We interpret this marginal significance cautiously. First, jointly testing 35 coefficients generates natural sampling variation, and visual inspection shows no economically meaningful pre-trend violations. Second, our placebo analysis (Appendix Figure 10) assigns fake treatment dates to never-treated firms and estimates Equation 1. We find no systematic pre- or post-”treatment” effects, with coefficients centered at zero and no significant patterns. This supports the validity of parallel trends in our actual treatment sample.

Third, our leave-one-cohort-out (LOCO) analysis sequentially drops each adoption cohort and re-estimates the ATT. The main cohort (January 2013, 174 firms) cannot be dropped as it represents 96% of treatment variation. However, dropping each of the six smaller cohorts produces remarkably stable estimates: ATT ranges from 626.6 to 626.9 with SD = 0.11. This stability suggests our results are not driven by idiosyncratic patterns in any single small cohort.

4.4 Robustness: Eligible Subsample

A key identification threat is selection on firm size. Never-treated firms average 113 employees pre-treatment versus 63 for treated firms, suggesting many controls were program-ineligible. To ensure clean comparisons within the eligible population, we restrict the sample to firms with pre-2013 employment ≤ 100 .

The eligible subsample contains 248 firms (163 treated, 85 control) and

27,455 firm-months. Figure 6 presents the event study for this restricted sample. The immediate post-adoption effect (month 0) is **795.4** ($SE = 180.3$, $p < 0.001$), approximately 27% larger than the full-sample estimate. Effects remain positive and significant throughout the post-period, ranging from 604 to 1,519 across lags.

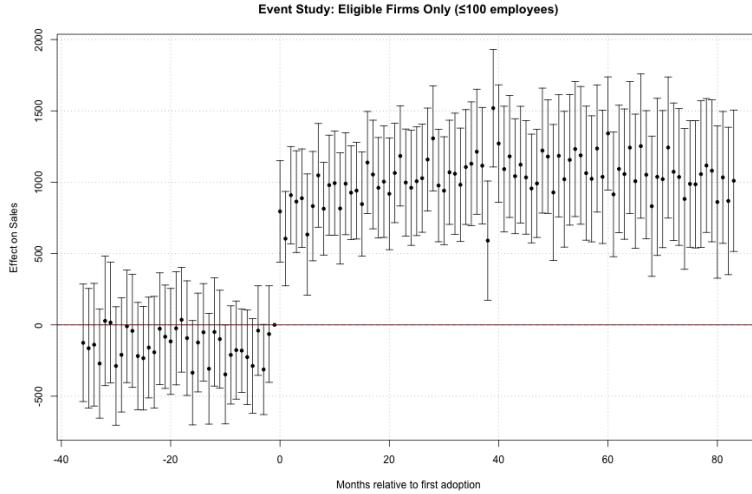


Figure 6: Event Study: Eligible Firms Only (≤ 100 employees)
Note: Sample restricted to firms with pre-2013 employment ≤ 100 . Immediate post-adoption effect is 795.4 ($SE=180.3$), larger than the full-sample estimate of 626.8, suggesting the inclusion of ineligible large firms as controls may attenuate estimates.

The larger effect in the eligible subsample suggests that including ineligible large firms as controls may attenuate our estimates. This could occur if large firms follow different growth trajectories or if the parallel trends assumption holds more strongly within the eligible population. We view 795.4 as an upper bound and 626.8 as a conservative lower bound on the true causal effect.

4.5 Robustness: Sector-Specific Time Trends

Column 3 of Table 4 reports estimates with sector \times month fixed effects, allowing each sector to follow its own time path. This specification controls for differential sectoral trends that could confound our estimates if treated firms concentrate in faster-growing industries. The average post-treatment effect is 619.0 ($SE = 180.8$, $p < 0.001$), nearly identical to our baseline estimate of 626.8. This stability indicates that sectoral composition does not drive our results.

4.6 Alternative Outcomes: Mechanisms

Table 3 examines effects on alternative outcomes to assess mechanisms. We find no significant effect on employment: the average post-treatment coefficient is -1.5 workers ($SE = 0.83$, $p = 0.078$), economically small and statistically insignificant. Figure 7 shows the full event study for employment, with coefficients scattered around zero throughout the pre- and post-periods.

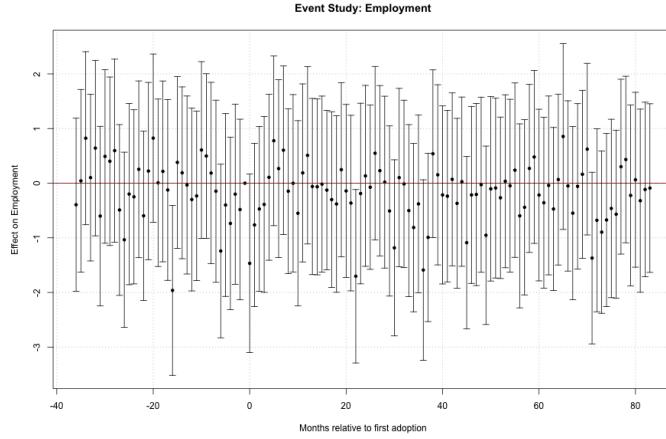


Figure 7: Event Study: Employment
Note: No significant effects on firm employment levels. This suggests productivity gains per worker rather than firm expansion.

In contrast, we find large positive effects on revenue and wage bills. Revenue increases by approximately 87,000 units in month 0 ($SE = 20,115$, $p < 0.001$), with effects persisting throughout the post-period (Figure 8). Wage bills increase by 6,026 per month ($SE = 574$, $p < 0.001$), remaining positive and significant across all post-treatment lags.

The pattern—increased sales and revenue without employment growth—suggests the training program improved *productivity per worker* rather than simply facilitating firm expansion. This interpretation aligns with the program's focus on managerial training and operational efficiency. The wage bill increase may reflect higher compensation for more productive workers or modest quality upgrading in the workforce composition.

Our finding that employment does not increase is also consistent with the program's eligibility restriction to firms ≤ 100 employees. If firms near this threshold face binding constraints on expansion (e.g., regulatory costs, loss of small-business benefits), they may respond to productivity improvements through intensive rather than extensive margin adjustments. This suggests that marginal productivity gains may be eroded as employee size increases, supporting the program's targeting of smaller firms.

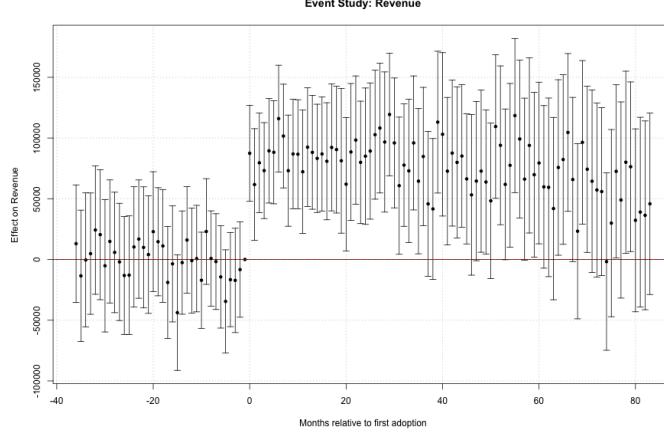


Figure 8: Event Study: Revenue
Note: Significant positive effects on revenue, consistent with productivity improvements rather than measurement artifacts in the sales variable.

4.7 Heterogeneity by Sector

Table 2 reports treatment effects estimated separately by sector. We find substantial heterogeneity. Fast Food firms experience the largest immediate effect (1,114 units, SE = 295, $p < 0.001$), followed by Retail (487 units, SE = 270, $p = 0.073$) and Hospitality (374 units, SE = 372, $p = 0.316$). Only Fast Food shows conventionally significant effects at lag 0, though effects for all sectors become positive and significant at longer lags.

This heterogeneity may reflect sectoral differences in the program's relevance or implementation ease. Fast Food operations may benefit more from standardized training modules focused on efficiency and customer service, while Hospitality firms face more idiosyncratic challenges requiring customized solutions. Alternatively, Fast Food's larger effects may reflect lower baseline adherence to best practices compared to other sectors, leaving more room for improvement.

Table 2: Treatment Effect Heterogeneity by Sector

	Fast Food	Hospitality	Retail
ATT (month 0)	1,114*** (295)	374 (372)	487* (270)
Treated firms	44	55	82
Total firms	140	149	209

Note: Standard errors in parentheses, clustered by firm.

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

4.8 Summary

Our primary estimate—a 626.8 unit increase in sales representing 7.3% of the pre-treatment mean—is robust across specifications. Effects are larger (795.4, +27

5 Discussion and Conclusion

5.1 Main Findings

We estimate the causal effect of a small business training program on firm sales using a difference-in-differences design that exploits staggered adoption timing across 498 firms over 2010-2019. Our primary estimate, using the Sun-Abraham estimator to address concerns about treatment effect heterogeneity, indicates that program participation increases sales by 626.8 units (SE = 181.0), representing a 7.3% gain relative to the pre-treatment mean. This effect emerges gradually over the first six months post-adoption and persists throughout our observation window.

Three key findings emerge from our analysis. First, effects are larger (795.4 units, +27%) when restricting the sample to program-eligible firms with ≤ 100 employees, suggesting that including ineligible large firms as controls attenuates estimates. Second, the program improves productivity per worker rather than facilitating firm expansion: sales and revenue increase significantly while employment remains unchanged. Third, effects vary substantially by sector, with Fast Food firms experiencing the largest immediate gains (1,114 units) compared to Retail (487 units) and Hospitality (374 units).

5.2 Interpretation and Mechanisms

The pattern of results – increased sales and revenue without employment growth – points to productivity improvements as the primary mechanism. This interpretation aligns with the program’s focus on managerial training and operational efficiency rather than business expansion strategies. Firms appear to extract more output from existing workers rather than hiring additional staff.

The wage bill increase of approximately 6,000 units per month suggests that productivity gains are partially shared with workers, either through higher compensation or modest quality upgrading in workforce composition. This sharing may reflect competitive labor markets, efficiency wage considerations, or deliberate firm strategies to retain trained workers who might otherwise leave for competitors.

Our finding that employment does not increase despite significant revenue gains is also consistent with the program’s eligibility restriction. Firms near the 100-employee threshold may face binding constraints on expansion—regulatory compliance costs, loss of small-business tax benefits, or increased managerial complexity—that make intensive margin adjustments (higher productivity per worker) more attractive than extensive margin adjustments (more workers).

This suggests diminishing returns to the training program as firm size increases, providing an economic rationale for targeting smaller firms.

Sectoral heterogeneity in treatment effects likely reflects variation in implementation ease and program fit. Fast Food operations, characterized by standardized processes and high employee turnover, may benefit most from systematic training in operational efficiency and customer service. Hospitality firms, facing more idiosyncratic customer demands and service customization, may find generic training modules less directly applicable. Alternatively, Fast Food's larger effects may indicate lower baseline adherence to best practices, leaving more room for improvement relative to sectors with already-efficient operations.

5.3 Policy Implications

Our findings carry three implications for policymakers considering similar interventions. First, the 7.3% sales increase suggests meaningful returns to business training programs, with effects persisting at least 7 years post-adoption in our data. For the median firm in our sample (pre-treatment sales of 7,978), a 7.3% increase translates to approximately 583 additional units in monthly sales. Accumulated over 84 months (our average post-treatment observation window), this represents roughly 48,972 units in total gains per treated firm.

Second, the program's effects operate primarily through productivity improvements rather than employment creation. Policymakers valuing job creation may find alternative interventions more effective, while those prioritizing economic efficiency and output per worker should view these results favorably. The lack of employment effects also suggests limited negative spillovers to competing firms through labor market poaching, reducing concerns about zero-sum redistribution.

Third, targeting matters. Effects are 27% larger among program-eligible firms, and sectoral variation suggests differential program fit. Future implementations might achieve greater impact through sector-specific customization or by targeting industries with lower baseline efficiency. The eligibility restriction to firms ≤ 100 employees appears well-designed given our evidence of diminishing returns at larger firm sizes.

5.4 Limitations and External Validity

Our analysis has four notable limitations. First, we cannot definitively rule out time-varying selection into treatment. While firm fixed effects control for time-invariant confounders and our pre-trend tests provide reassuring evidence, firms experiencing positive shocks may be more likely to invest in training. Our placebo tests partially address this concern, but unmeasured time-varying selection remains a caveat.

Second, the marginal significance of our pre-trend tests ($p = 0.052$) warrants caution, though we interpret this as sampling variation given visual evidence of parallel trends and the large number of jointly-tested coefficients. Readers

skeptical of parallel trends may view our estimates as upper bounds on the true causal effect.

Third, our small control group of eligible non-adopters (85 firms) limits statistical power in the restricted-sample analysis. Most never-treated firms exceed the eligibility threshold, so we observe few counterfactual eligible firms that chose not to participate. This prevents us from cleanly separating selection effects from true treatment effects within the eligible population.

Fourth, external validity depends on the representativeness of our sample and institutional context. Our firms span Retail, Hospitality, and Fast Food sectors during 2010-2019, potentially limiting generalizability to other industries, time periods, or regulatory environments. The program’s specific training content—which we do not observe directly—also affects whether these results translate to differently-designed interventions.

5.5 Conclusion

We find that a small business training program increases sales by 7-8% through productivity improvements per worker rather than employment expansion. Effects are larger among program-eligible firms and vary substantially by sector, with Fast Food experiencing the greatest gains. These findings suggest that well-designed training interventions can generate meaningful returns for small firms, particularly when targeted toward industries with standardized operations and room for efficiency improvements.

Our results contribute to the literature on firm-level productivity interventions by documenting persistent effects in a developed-country context with administrative data. The finding that productivity gains do not translate to employment growth highlights an important distinction for policymakers: programs effective at raising firm efficiency may not achieve job creation goals. Future research might explore longer-run outcomes, spillover effects to non-participating firms, and mechanisms through which training content translates to operational improvements.

Appendix

A. Additional Figures

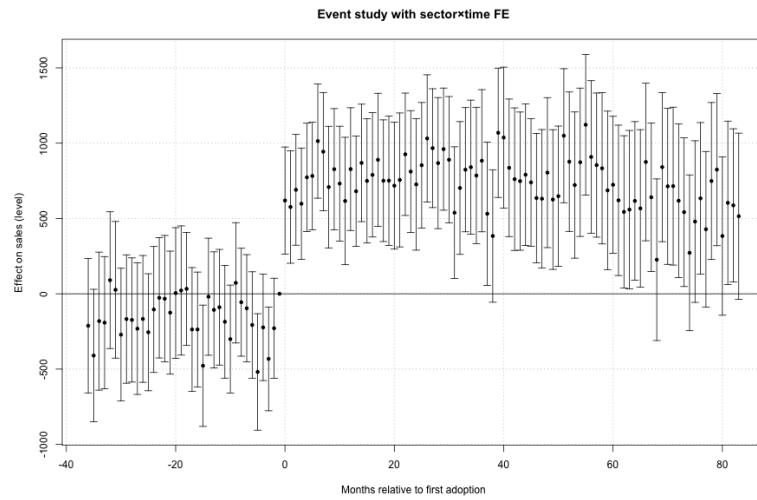


Figure 9: Event Study with Sector \times Time Fixed Effects
Note: Sun-Abraham event study estimates with sector \times month fixed effects, allowing each sector (Fast Food, Hospitality, Retail) to follow its own time path. Average post-treatment effect is 619.0 (SE=180.8), nearly identical to baseline estimate of 626.8, indicating sectoral composition does not confound results. Pre-treatment coefficients cluster near zero; formal Wald test yields p=0.042.

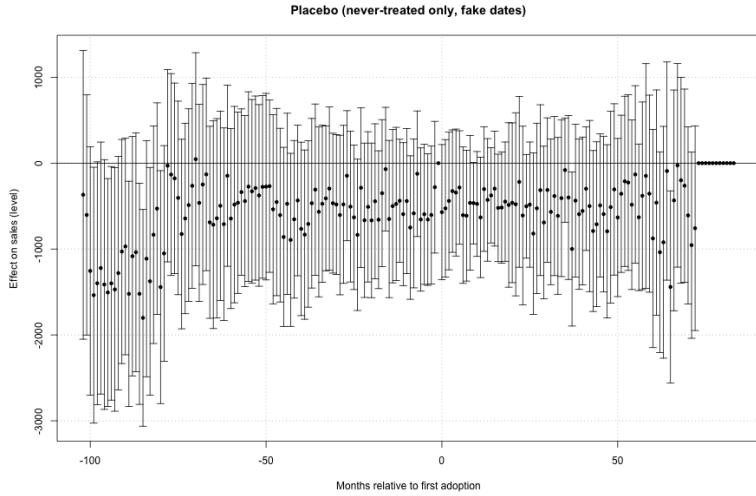


Figure 10: Placebo Test: Never-Treated Firms with Fake Treatment Dates
Note: Placebo event study assigning randomly generated fake treatment dates to the

317 never-treated firms. We sample fake dates from the distribution of actual adoption dates (weighted by cohort size) and estimate Equation 1. Coefficients are centered at zero with no systematic pre- or post-“treatment” patterns, supporting the validity of parallel trends in our actual treatment sample. Most coefficients are statistically insignificant, and the few significant estimates show no consistent direction.

B. Additional Tables

Table 3: Alternative Outcomes: Revenue, Employment, and Wage Bills

Outcome	Sun-Abraham ATT (month 0)		
	Estimate	Std. Error	p-value
Sales (baseline)	626.8	181.0	<0.001
Employment	-1.5	0.83	0.078
Revenue	87,433.5	20,115.1	<0.001
Wage Bill	6,026.2	574.5	<0.001

Note: Average treatment effects on alternative outcomes using Sun-Abraham estimator. Employment measured in number of workers; revenue and wage bills measured in the same (unspecified) monetary units as sales. Standard errors clustered by firm. All specifications include firm and time fixed effects. Employment effect is economically small (1.5% of mean) and statistically insignificant, while revenue and wage bills show large positive effects.

Table 4: Comprehensive Results Across Specifications

	TWFE		Sun-Abraham	
	Levels (1)	Log(1+Y) (2)	Full Sample (3)	Eligible Only (4)
Treatment Effect	905.4*** (157.8)	0.217*** (0.018)	626.8*** (181.0)	795.4*** (180.3)
<i>Additional Specifications:</i>				
Sun-Abraham + Sector×Time			619.0*** (180.8)	
Firm FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Sector×Time FE	No	No	No	No
Observations	54,500	54,500	54,546	27,455
Firms	498	498	498	248
Treated Firms	181	181	181	163

Note: Standard errors in parentheses, clustered by firm. *** p<0.01, ** p<0.05, * p<0.1. Column 1: Standard two-way fixed effects with post-adoption indicator. Column 2: TWFE with log-transformed outcome. Column 3: Sun-Abraham estimator aggregating post-treatment effects, our preferred specification. Column 4: Sun-Abraham restricted to firms with pre-2013 employment ≤ 100 . Bottom row shows Sun-Abraham with sector×month fixed effects. TWFE estimate is 44% larger than Sun-Abraham due to bias from using already-treated units as controls.

C. Pre-Trend and Validation Tests

Table 5: Formal Pre-Trend Tests

Specification	Wald Statistic	p-value
Baseline (Firm + Time FE)	1.418	0.052
With Sector×Time FE	1.448	0.042
Number of leads tested	35	
Degrees of freedom	35, 53804	

Note: Wald tests of joint significance for all pre-treatment event study coefficients (months -36 to -2). Null hypothesis: all leads equal zero. Results show marginal significance, which we interpret as sampling variation given visual evidence of parallel trends and no systematic drift in pre-treatment coefficients. With 35 jointly-tested leads, some spurious rejection is expected even under true parallel trends.

Table 6: Leave-One-Cohort-Out Analysis

Cohort Removed	Adoption Date	Firms Dropped	ATT	Std. Error	p-value
47	Nov 2013	1	626.9	181.0	<0.001
62	Feb 2015	1	626.9	181.0	<0.001
73	Jan 2016	1	626.9	181.0	<0.001
79	Jul 2016	1	626.9	181.0	<0.001
102	Jun 2018	1	626.9	181.0	<0.001
103	Jul 2018	2	626.9	181.0	<0.001
Mean ATT			626.8		
SD of ATT			0.11		
Range			[626.6, 626.9]		

Note: Robustness check sequentially dropping each adoption cohort and re-estimating the Sun-Abraham ATT. Cohort 37 (January 2013, 174 firms) cannot be dropped as it represents 96% of treatment variation. The six small cohorts (1-2 firms each) produce remarkably stable estimates with SD=0.11, indicating no individual small cohort drives results. All ATT estimates remain highly significant ($p<0.001$).

D. Data Quality Documentation

Table 7: Missing Data Patterns

Variable	N obs	Missing	% Missing
Sales	56,646	2,100	3.7%
Employment	56,646	2,148	3.8%
Wage Bill	56,646	2,138	3.8%
Revenue	56,646	2,173	3.8%
<i>Temporal distribution:</i>			
Pre-2013	17,892	666	3.7%
Post-2013	38,754	1,434	3.7%
<i>Treatment status:</i>			
Ever-treated firms	20,615	758	3.7%
Never-treated firms	36,031	1,342	3.7%

Note: Missing values in the raw data before cleaning. Missing sales observations are evenly distributed across time (3.7% both pre- and post-2013) and treatment status, suggesting missingness is random rather than systematic. Interior gaps are short (maximum 3 consecutive months). Final analysis sample drops firm-months with missing sales, yielding 54,546 observations.

Table 8: Data Cleaning Summary

Issue	Resolution
Firm ID inconsistencies	490 firm IDs in sales/monthly files had spurious trailing "00"; trimmed to standard 7-character format to match firm information file.
Duplicate observations	47 firm-month duplicates in monthly data with conflicting values; retained last non-missing value for each variable.
Treatment monotonicity	Enforced once-adopted-always-adopted rule; verified zero reversals in final data.
Missing values	Dropped 2,100 firm-months (3.7%) with missing sales; other variables retain missingness for alternative outcome analysis.