I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 18233

The (In)effectiveness of Targeted Payroll Tax Reductions

Alessandra Fenizia
Nicholas Li
Luca Citino

OCTOBER 2025

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 18233

The (In)effectiveness of Targeted Payroll Tax Reductions

Alessandra Fenizia
George Washington University

Nicholas Li
George Washington University and IZA

Luca Citino
Bank of Italy

OCTOBER 2025

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

The (In)effectiveness of Targeted Payroll Tax Reductions*

This paper studies the cost-effectiveness of targeted payroll taxes for stimulating labor demand. It uses rich administrative data to study the effects of an Italian reform that raised social security contributions for apprenticeship contracts but granted a substantial discount for firms with 9 employees or less. The discount does not increase demand for apprenticeship contracts. Instead, it subsidizes inframarginal hiring. This reform is not cost-effective. Point estimates imply that each million euros of foregone social security contributions supports the employment of 29 apprentices for one year and no permanent contracts (these estimates are not statistically different from zero).

JEL Classification: J01, J08, H20

Keywords: targeted payroll taxes, cost effectiveness, labor demand

Corresponding author:
Alessandra Fenizia
George Washington University
2121 I St NW, Washington, DC 20052
USA

E-mail: afenizia@gwu.edu

* We thank David Card and Jörn-Steffen Pischke for their guidance and support during the early stages of this project. We also thank Fabrizio Balassone, Francesco D'Amuri, Daniel G. Garrett, Andrew Goodman-Bacon, Steven Hamilton, Hilary Hoynes, Elira Kuka, Francesca Lotti, Matteo Paradisi, Enrico Rubolino, Vincenzo Scrutinio, Bryan Stuart, Roberto Torrini, Dario Tortarolo, Eliana Viviano, and Valeria Zurla for the useful discussion and comments. The realization of this project was possible thanks to the "VisitInps Scholars" program. We are very grateful to the staff of Direzione Centrale Studi e Ricerche for their invaluable support with the data and the institutional setting. The opinions expressed in this paper are those of the authors alone and do not necessarily reflect the views of the Bank of Italy or INPS.

1 Introduction

Policymakers often turn to targeted payroll tax reductions to combat high unemployment rates among the young, the low-skilled, and the long-term unemployed (OECD, 2003, 2011). However, targeting workers at the margins of the labor market presents dilemmas for both policy and policy evaluation. From a policy perspective, targeted payroll tax reductions may not be cost-effective because they may subsidize inframarginal employment that would exist absent incentives (Saez et al., 2019). How strongly firms respond to incentives to hire workers for whom they have little interest is an empirical question.

From a program evaluation perspective, estimating targeting's cost-effectiveness faces two identification requirements. First, targeted workers should be different from untargeted workers, *by design*. Second, targeted payroll tax cuts benefit workers by increasing labor demand, a firm decision, so incentives should be (quasi-)random *across firms*. A subsidy that does not satisfy the first requirement is not targeted, and thus, its estimated effects may not be portable to marginalized workers with intrinsically lower labor demand. A subsidy that does not satisfy the second requirement is not conducive to credibly estimating causal firm responses, making it challenging to determine how much workers ultimately benefit. Thus, the empirical requirements of estimating the cost-effectiveness of targeting are surprisingly steep: an ideal subsidy would target non-random workers at randomly selected firms. To the best of our knowledge, there is little evidence that satisfies both requirements.

This paper analyzes changes in firm behavior in response to a reduction in payroll taxes targeted to apprentices. In 2007, employers were required to pay higher social security contributions (SSCs) for apprentices; however, Italian firms with at most 9 full-time equivalent employees (excluding apprentices) were given temporary relief from the increases. The relief was equivalent to roughly two months of earnings per apprentice, 8% of the earnings for a typical 19-month apprenticeship, and phased out over time. The SSC discount for apprentices satisfies the demanding requirements for evaluating targeted subsidies. (1) The discontinuity across firm size generates quasi-random variation across firms; and (2) the subsidies apply only to apprentices. Our analysis of confidential matched employer-employee data furnished by the Italian Social Security Institute (INPS) compares firm outcomes above and below the 9-employee discontinuity in SSCs in a difference-in-discontinuities design. Concretely, the reduced-form estimates measure "intention-to-treat" effects using narrow variation in a neighborhood of the policy threshold. We use the policy variation as an instrument for firms' tax payments to measure jobs supported per unit of lost tax revenue.

The design provides a strong first stage—we find discontinuous effects on subsidy take-up and SSCs. At the same time, we provide ample evidence supporting the validity of the reduced form effects and, correspondingly, the instrument's exclusion restriction. First, we find no evidence of manipulation in firm size (the running variable) at the threshold, mitigating concerns that the reform generates costly firm-size distortions observed in other settings (Garicano et al., 2016; Caicedo et al., 2020). Second, there are no pre-trends in the estimated discontinuity in the outcomes. Third, there are no estimated effects on observed firm characteristics, industry composition, or geographical composition; ergo, our results are not confounded by comparing observably different firms over time. Fourth, the design is robust to changes in the relationship between firm outcomes and firm size that may arise from factors unrelated to the policy—such as macroeconomic trends or mean reversion.

We find that the targeted SSC discount does not increase the demand for apprenticeship contracts. Our reduced-form employment estimates are precisely zero. Instead, the policy primarily subsidized inframarginal firms (i.e., those who did not change their hiring behavior in response to the reform). We also find that the policy did not increase the rate at which existing apprentices were given permanent contracts (i.e., transformations). One reason why payroll tax cuts may have little effect on employment is if they result in higher wages. We find that firms do not adjust apprentices' earnings in response to the reform. Our findings are consistent with the precise null employment effects observed in response to a large wage subsidy in Jain et al. (2025).

Because our policy generates variation across firms, we can examine whether treatment firms responded in undesirable or unintended ways relative to control firms. Treatment firms did not substitute toward or away from apprentices to other contract types, did not opportunistically re-label existing contracts, did not churn through more apprentices, and did not hire lower-quality workers. We show that the null effects are unlikely to be driven by the size of the subsidy, low salience or awareness of the policy, the firm's incentives to limit growth to maintain eligibility, the fact that the subsidy applies to training contracts, the subsidy's temporary nature, or the Great Recession. Instead, our null results imply that the demand for apprenticeship contracts is simply inelastic (in line with the findings of Egebark and Kaunitz, 2013 and Huttunen et al., 2013). We show that the demand for apprentices is consistently inelastic across industries and regions; for firms that did or did not employ apprentices at baseline; for firms that pay their apprentices more or less; and for firms that do or do not face liquidity constraints. Despite the robustness of these results, our design provides internally valid estimates for firms in a neighborhood of 9 employees, and we cannot rule out different effects for larger firms. Nevertheless, the results suggest that one cannot induce firms to hire more apprentices simply by lowering their labor costs. This interpretation is corroborated by the RIL, a survey of Italian firms. When asked why they do or do not hire apprentices, firms rarely respond that cost was a primary consideration.

Finally, we formally measure the cost-effectiveness of targeted subsidies and benchmark the estimates to those in the literature. In contrast to the back-of-the-envelope estimates in the literature, we use instrumental variables to estimate the number of jobs created per unit of foregone revenue. This approach allows us (1) to compare effect sizes across studies using a unified metric and (2) to assess the statistical precision of these estimates. Point estimates imply that each million euros of foregone social security contributions supports the employment of 29 apprentices for one year and no permanent contracts (these estimates are not statistically different from zero). Unsurprisingly, these estimates are consistent with the reduced-form results that show no employment effects. However, they also imply that (1) the larger apprentice subsidies in the pre-reform status quo carried large fiscal costs without meaningfully increasing apprentice employment, and (2) the payroll tax increase on the control group substantially increased revenues with no meaningful reductions in apprentice employment.

Are these estimates outliers? We argue they are not. We benchmark our results against other findings in the literature. While other evaluations highlight positive employment effects, these effects often come at enormous costs. After normalizing employment estimates against their costs, there is actually little evidence that targeted subsidies are a cost-effective way of increasing the employment of workers at the margins of the labor market.

This paper makes two contributions to the literature on the effects of targeted wage subsidies and payroll taxes. First, the research design satisfies the dual empirical requirements of providing quasi-random variation in incentives across firms and targeting non-random, marginalized workers. This allows us to credibly estimate labor demand for marginalized workers. Second, it provides a new perspective on the cost-effectiveness of using targeted subsidies to support the employment of workers at the margins of the labor market.

Satisfying the dual empirical requirements yields two key benefits. The first benefit comes from analyzing exogenous variation in incentives across firms. In contrast, national studies (e.g., Bozio et al., 2020; Egebark and Kaunitz, 2013; Huttunen et al., 2013; Jain et al., 2025; Saez et al., 2019, 2012, 2021; Rubolino, 2021) estimate labor demand responses by comparing the aggregate employment of targeted and untargeted workers. The estimated effects may be biased if firms substitute untargeted workers for targeted ones. Such a SUTVA violation could lead one to overstate the effectiveness of a policy since the increased demand for targeted workers comes at the cost of decreased demand for untargeted workers, a confounding that would not show up in parallel pre-trend tests. Because our specifications are not based on comparisons of workers, we can distinguish between increased demand and substitution. The second benefit comes from the policy targeting specifically marginalized workers. Cross-regional studies often study firms, but targeted workers are not necessarily marginalized (e.g., Bennmarker et al., 2009; Benzarti and Harju, 2021a; Bohm and Lind, 1993; Guo, 2024; Korkeamaki and Uusitalo, 2006).[1] Since labor demand for all workers is inherently higher than labor demand for marginalized workers, estimates derived from broad-based policies may overstate the cost-effectiveness of subsidies targeted to marginalized workers.

Our precise, null reduced-form employment estimates stand in stark contrast to recent estimates of positive employment effects (Benzarti and Harju, 2021a; Cahuc et al., 2019; Saez et al., 2019, 2021). However, our second contribution is to point out that the large employment effects are costly. We provide a new perspective on the mixed employment findings across the literature. Taking existing estimates at face value, we show that there are essentially no reliable instances in the literature where targeted subsidies were a cost-effective way of supporting the employment of marginalized workers. While our precise null reduced-form estimates stand in contrast with the literature, our cost-effectiveness estimates stand in accord with the literature in this new light. Our IV estimates are quantitatively small and not statistically different from zero across the board. Our IV approach offers two key advantages: it provides a unified metric for comparing policy effectiveness across different settings and allows for a transparent assessment of the estimates' statistical precision.

Finally, our paper also contributes to the literature that critically examines difference-in-differences designs and what researchers can learn from parallel pre-trends (Borusyak et al., 2021; De Chaisemartin and D'haultfœuille, 2023; Goodman-Bacon, 2021; Rambachan and Roth, 2023; Roth and Sant'Anna, 2023; Roth et al., 2023; Sun and Abraham, 2021). Researchers often evaluate the impact of policies that apply to units above (or below) a given threshold by comparing the outcome of units above the threshold with those below using a standard difference-in-differences design (Benzarti et al., 2020; Bozio et al., 2017; Cahuc et al., 2019; Goos and Konings, 2007; Saez et al., 2019). Our paper illustrates the perils of this approach. While this seems like a transparent and reasonable design, we show that it can lead to misleading conclusions even in the presence of parallel pre-trends. Such a design does not distinguish level shifts at the threshold with rotations of the conditional expectation function, leading to potentially spurious estimated effects. Our design is robust to such rotations.

The paper is structured as follows. Section 2 describes the institutional setting. Section 3 presents the data. Section 4 develops the empirical strategy and presents the main results. Section 5 evaluates the cost-effectiveness of the policy. Finally, Section 6 concludes.

[1] A notable exception is Van Reenen (2003), which studies the New Deal for Young People in the UK (See also Bell et al., 1999). Comparing young workers in pilot regions to other regions, he finds large exits from unemployment.

2 Institutional Background

This section describes the legal framework for apprenticeship contracts in Italy and the policy variation we exploit in our empirical analysis.

2.1 Apprenticeship Contracts in Italy

Apprenticeships are labor contracts that allow workers to earn a professional qualification and a salary in exchange for labor services (Snell, 1996; Ryan, 2012). By law, apprentices receive at least 120 hours of training per year dictated by collective bargaining agreements (CBAs): 80 hours to occupation-specific training and 40 hours to general training. In practice, training requirements are poorly enforced.

During the period of study, only private-sector workers aged 18–29 are eligible to work as apprentices. New apprentices go through a short probationary period of less than two months. Most apprenticeship contracts last fewer than three years, though in rare circumstances they can last up to six years. At the end of the contract, firms can either hire apprentices or let them go at no cost. Many choose to retain them as full-time employees.

Firms have two incentives to hire apprentices: they can pay apprentices up to 2 levels below the target pay grade in the CBA, and they pay lower payroll taxes. Firms also pay lower SSCs for one year if they hire apprentices permanently (Law 56/1987). Firms cannot employ more apprentices than regular workers, but this constraint rarely binds (column 2 in Table 1).

Table A.1 reports the summary statistics for the apprentices at baseline (January 2006). The typical apprentice is male (65.7%), 22.5 years old, earns 1050 euros per month, and has 3.7 years of experience. The vast majority of apprentices are native (88%) and have had at least one previous job in the private sector (98.5%). By the end of our sample period, 44.6% transition to an open-ended position at their training firm.

2.2 The 2007 Budget Bill

The 2007 Budget Bill (Law n.296/2006) increased employers' SSCs on apprenticeship contracts to finance paid sick leave for apprentices. However, SSCs were discounted for firms with at most 9 employees, generating a clean discontinuity in incentives.

Figure 1 illustrates how SSCs change for an average apprenticeship earning 12,000 euros per year. Before 2007, SSCs were a fixed 2.85 euros per contract per week or about 148 euros per year (green triangles). Beginning on January 1, 2007, firms with more than nine employees paid 10% of earnings (1,200 euros per year, hollow blue circles). Firms with nine or fewer employees received a discount. They paid 1.5% of earnings in the first year (180 euros), 3% in the second year (360 euros), and full contributions in subsequent years (orange circles). The savings amount to roughly two months of earnings per apprentice, 8% of a typical 19-month apprenticeship.

The eligibility for discounted SSCs was based on *policy-relevant* firm size, total full-time equivalent employment minus apprentices, temporary agency workers, workers on leave, and workers with an on-the-job training contract. Our rich administrative data allow us to follow this definition closely (see Appendix B for more details). The increase in SSCs applied to both existing apprenticeship contracts and those signed after January 1, 2007. For pre-existing contracts, the eligibility was determined based on the average firm size in 2006. For contracts signed after January 1, 2007, eligibility was determined by the firm size at the time of hiring.

The discount was not applied automatically. Firms claimed the discount by flagging a box when filing their monthly report to the Italian Social Security Agency. No other pre-existing or concurrent policy was discontinuous at nine employees.[2]

The top panel of Figure 3 illustrates the relationship between the share of firms receiving the subsidy in January 2007 (take-up rate) and the policy-relevant firm size in the same month. The monthly take-up rate is approximately 2% for firms below the 9-employee threshold, sharply decreases around nine employees, and converges to 0.4% for firms above the threshold. These relatively low monthly take-up rates primarily reflect two facts: (1) relatively few firms hire anyone, much less apprentices, in any given *month* (Table 1), and (2) 75% of firms do not hire apprentices at all. Generally, firms are aware of the policy: 80% of eligible firms receive the subsidy (Appendix Figure A.1).

Figure 3 highlights two important facts. First, there is no appreciable discontinuity at the threshold. This is partly due to mismeasurement in policy-relevant firm size *at the time of hiring*: we measure policy-relevant firm size over the course of the month, but eligibility is determined instantaneously. Second, the take-up rate does not drop to zero past the threshold. Some firms receive the payroll tax reduction despite being ineligible, reflecting firms self-reporting eligibility and imperfect compliance. As we discuss in Section 4.1, two-way non-compliance will lead intention-to-treat (ITT) estimates to be smaller than the treatment effect on the treated (TOT).

[2] Consistent with the absence of other policies, Figure 2 shows the cumulative density function of policy-relevant firm size before (orange lines) and after the reform (green lines). There is no discontinuity at the 9-employee threshold, and the distribution of policy-relevant firm size remains stable over time.

3 Data

In this section, we describe the data that form the basis of our empirical analysis and how we construct our sample.

3.1 Data and Sample Selection

Social Security Records. Our main source of data is the confidential matched employer-employee dataset collected by the Italian Social Security Institute (*Istituto Nazionale di Previdenza Sociale*—INPS hereafter), derived from monthly tax filings. These data cover the universe of all private non-agricultural firms with at least one employee from 1983 to today. Firms are identified by a unique tax number and workers are identified by their social security number. The data include firms' location; detailed industry codes; juridical status; and opening and closing dates. For each job spell, we observe the beginning and end dates; earnings net of SSCs; part- versus full-time status; coarse occupation categories (apprentice, blue-collar, white-collar, or manager); and worker demographics. Crucially, the data contain detailed information on applicable tax policies covering the contract, including employer SSCs and whether the firm received the SSC discount.

We utilize the full data to construct workers' employment histories (including their previous earnings), and our main analysis focuses on firm outcomes between January 2003 and December 2009. We restrict our main sample to firms with policy-relevant firms size between 3 and 15 employees. This yields a sample of 1,015,619 firms. Our sample is skewed toward small firms by construction. However, 90% of Italian firms have 15 or fewer employees, and these firms employ 65% of all apprentices.

RIL data. We complement the confidential social security records with a representative survey of firms that collected data on the demand for different contracts in 2005, the RIL (i.e., *Rilevazione Longitudinale su Imprese e Lavoro*). For consistency, we restrict this sample to firms between 3 and 15 employees (N=10,191).

3.2 Descriptive Statistics

Table 1 displays the summary statistics for firm characteristics in our main sample at baseline (i.e., in January 2006). Column 1 reports the characteristics for the full sample; columns 2 and 3 display the statistics for firms that hire apprentices and firms that ever take up the subsidy, respectively.

The average firm in our sample is a Limited Liability Company (LLC) established in the early 90s and employs 7 workers. Full-time equivalent employment is roughly the same as average number of employees because most workers are employed full-time. 94% (=6.63/7.088) of the employees have a permanent contract. Apprenticeship contracts are nominally permanent contracts and make up approximately 6% (=0.427/7.088) of the contracts in our sample.

An average apprenticeship lasts for 19 months. Apprentices experience a substantial amount of turnover: in any given month, firms hire on average 0.030 apprentices and separate from 0.015 apprentices.

While some firms employ many apprentices, 75% (=99,311/398,412) of firms in our sample do not hire any. Among firms in our sample, those that employed at least one apprentice in January 2006 (column 2) are marginally larger and have more apprentices than the average firm in our sample (column 1). However, firms that employ apprentices are similar in their hiring and separation behavior. By construction, firms that take up the SSCs relief (column 3) are smaller and have more apprentices than the average firm in the sample (column 1), but do not appear to be different on other dimensions.

Mirroring Table 1, Appendix Table A.2 compares the industry shares of firms in our sample. Firms that hire apprentices (column 2) are more likely to be in manufacturing than the average firm in our sample (column 1), and less likely to be in agriculture or public administration, education and health. Public sector workers are not eligible to be apprentices.

4 The Effect of the Policy Threshold

This section is organized as follows. First, we formally lay out the difference-in-discontinuities approach. Second, we illustrate the approach using two key outcomes as examples: take-up and apprentice hiring. Third, we provide evidence of model validity, showing that the difference in covariates of firms just above and below the discontinuity does not change over time. Finally, we document the policy's null effects across other outcomes.

4.1 Difference-in-Discontinuities Design

The incentives generated by the law suggest comparing firms with policy-relevant firm size above and below the eligibility threshold of 9. Define Z_{it} as the distance of policy-relevant firm size from the discontinuity, and $T_{it} = \mathbf{1}[Z_{it} \leq 0]$. Firm size is not randomly assigned, so firms of different sizes differ in dimensions other than program eligibility.

We address this challenge using a difference-in-discontinuities approach. Our estimated discontinuities come from normalizing period-specific discontinuity estimates to the baseline period, January 2006:

$$\begin{aligned} Y_{it} &= a_{1t} + a_{2t}T_{it} + g_{1t}Z_{it} + g_{2t}Z_{it} \times T_{it} + u_{it} \quad \forall t \\ b_t &\equiv a_{2t} - a_{2,\text{Jan 2006}}. \end{aligned} \tag{1}$$

The first line is a standard regression discontinuity (RD) specification, estimated in each period. The second line subtracts the estimated discontinuity at baseline in January 2006. The bias in cross-sectional RD comes from unmodeled non-linearities in the relationship between the outcome and the running variable; intuitively, our approach debiases using estimates from the baseline pre-period, January 2006. Thus, the reduced-form effects are given by b_t, the *changes* in the estimated discontinuity at the threshold relative to January 2006. Mirroring validity tests of difference-in-differences designs, we can assess pre-trends in the estimated discontinuity from periods prior to January 2007. We cluster the errors at the firm level. We contrast our approach with a standard RD design in Appendix C.

To ensure that more weight comes from observations closest to the discontinuity, we follow the standard approach in the RD literature and weight observations according to a triangular kernel function (Calonico et al., 2014). To avoid estimated null results coming from measurement error, we exclude firms within firm-size 1 of the discontinuity.[3]

Our empirical strategy identifies an ITT effect. Because these reduced-form estimates do not adjust for imperfect compliance and include firms regardless of whether or not they hire apprentices, our estimated effect will be smaller than the TOT (those who took the subsidy). Section 4.2 illustrates how our approach estimates the effects of the policy on take-up and apprentice hiring, respectively.

4.2 Illustrating the Design with Take-up and Apprentice Hiring

Figure 3 and Figure 4 deconstruct the regression specification. The top panel of each figure is a binned scatter plot approximating the conditional expectation function in January 2007 of tax-break take-up and apprentice hiring, respectively. Overlaid in grey are best-fit lines excluding different windows of data, and overlaid in black excluding a window of firm-size

[3]Our approach differs from standard applications of "donut-hole" RD for two reasons. First, the usual impetus for excluding data near the discontinuity in other settings is the manipulation of the running variable, but we find no evidence of manipulation, and our specification passes all tests of validity (Section 4.3). Second, our longitudinal data allows us to estimate the bias associated with extrapolation in the baseline period (January 2006) and subtract it from all other estimates. Regardless, our (null) results do not appear to be driven by the inclusion (or exclusion) of data closest to the discontinuity (see Appendix Figure A.2).

1, our preferred estimates. The research design in Equation 1 repeats this estimation in each period, shown in the second panel. The third panel plots the measured discontinuity over time, and our reduced-form difference-in-discontinuities estimates are obtained by subtracting the value at the base period, January 2006.

In Figure 3, the likelihood of take-up increases by 2 p.p. per month. The plot shows the change is abrupt, and our design exploits variation over time. Naturally, because the policy did not exist prior to January 2007, the estimates for take-up are zero in the pre-period, and the estimates normalized to January 2006 are mechanically identical. The policy's effect on take-up declines through the end of our analysis period.

Figure 4 is constructed analogously, and it examines the policy's effect on apprentice hiring. None of the binned scatterplots show any visual sign of discontinuity. The time-series of the discontinuity estimates shows no appreciable change in January 2007 or subsequently—the estimates normalized to January 2006 are virtually identical. The noisy appearance of the time-series belies the precision of the estimates owed to the large administrative sample.

In the middle subplots, Figure 4 also shows that the conditional expectation function is rotating clockwise, coincident with a general slowdown in overall hiring and apprentice hiring through the end of 2009. A clockwise rotation would drive down the mean of the unsubsidized firms above the threshold. In a standard difference-in-differences specification, this would lead to conclusions that the subsidy supported hiring. See Appendix C.

4.3 Tests of Validity

Here, we show that our design consistently compares observationally similar firms, so our results are unlikely to be driven by changing patterns of selection. First, we show covariate balance by estimating our main specification with firm characteristics measured at baseline. The covariate differences between firms just above and just below the threshold are constant and do not depend on when policy-relevant firm size is measured. Second, we show that the marginal distribution of policy-relevant firm size is constant over time, exhibiting no bunching or manipulation.

4.3.1 Covariate Balance and Observable Differences

Covariate differences between firms just above and just below the threshold do not change over time. Table 2 and Table 3 show covariate stability over firm age and type; firm industry; and firm location, respectively. These tables report the effects of being below the threshold (b_t) from the main difference-in-discontinuities specification in Equation 1, where the outcome variables are general firm characteristics. For parsimony, we report a subset of the estimates.[4] The first two columns report the pre-reform estimates for $t_0 - 48$ (January 2003) and $t_0 - 24$ (January 2005). Columns 3–5 report the post-reform estimates for t_0 (January 2007), $t_0 + 12$ (January 2008), and $t_0 + 35$ (December 2009), respectively. The last three columns report Wald F-statistics testing the null that all the coefficients, the pre-reform coefficients, and the post-reform coefficients are zero, respectively. There is no imbalance along age or firm type in Panel A of Table 2.

While the vast majority of covariates show no signs of imbalance, in Panel B of Table 2, the balance tests detect statistically significant coefficients for Manufacturing and Transportation and Construction dummies. Similarly, most region dummies are strongly balanced (Table 3), but some coefficients for Lombardy, Liguria, Umbria, and Molise dummies are significant at the 10% level. These are not the consequence of systematic changes but rather random variation plus precision from our large administrative data.

To summarize the covariate balance validity checks, we assess the policy's effects on a covariate index, the predicted values from a regression of apprentice hiring on time-invariant firm characteristics. Figure 5 shows that being above versus below the cutoff does not correspond to changes in covariates that systematically predict apprentice hiring. The estimates are extremely precise. The statistically insignificant point estimates fluctuate between -0.0005 and +0.0005, almost two orders of magnitude smaller than the statistically insignificant effects on apprentice hiring that fluctuate between -0.01 and +0.01 (Appendix Figure A.3). Altogether, there is no evidence that measuring policy-relevant firm size contemporaneously results in compositional shifts or comparisons between observationally different groups.

4.3.2 Stability of Marginal Distributions and Unobservable Differences

Section 4.3.1 shows that the differences in observable characteristics are stable over time, evidence that our empirical specification compares observationally similar firms over time. To provide evidence that firms are not sorting across the policy threshold on unobservable dimensions—i.e., manipulating firm size to become eligible for the subsidy—we plot the CDFs of the running variable for each of the 84 periods in Figure 2. CDFs prior to January 2007 are plotted in orange; those starting from January 2007 are plotted in green. The marginal distributions are highly stable. The 84 CDFs are virtually identical and exhibit almost no change in the periods before and after the reform.

[4]The results for the full subset of estimates are available upon request.

4.4 Reduced-Form Effects of Subsidizing Apprentice Hiring

Figure 6 shows the reduced form effects of being *below* the policy threshold on social security contributions. Despite requiring firms to opt in, the policy has bite. In a given month, *smaller* firms pay 25 euros *less* per month in social security contributions than *larger* firms. The fiscal impact of 25 euros per month per firm may seem small at first glance; however, we emphasize that these reduced-form estimates do not adjust for imperfect compliance and include firms regardless of whether or not they hire apprentices. We discuss the cost-effectiveness (or lack thereof) in further detail in Section 5.

The pre-trends are flat, and the differences between eligible and ineligible firms do not emerge until January 2007. The differences between smaller and larger firms are largest at the onset of the policy in January 2007 and decline through 2009.

Despite paying lower SSCs, Figure 7 shows that firms just below the policy cutoff do not have relatively more or fewer apprentices than they did before the enactment of the policy. We also see no effects on the net apprentice wage bill, the total pecuniary compensation for the firm's apprentices net of taxes and SSCs (Figure 8). From a welfare perspective, the policy is efficient. The negative fiscal impact accompanies a null behavioral response, so the subsidy is essentially a pure transfer, and the marginal value of public funds is essentially 1 (Hendren and Sprung-Keyser, 2020).[5]

No Heterogeneity. We evaluate whether the null results on the number of apprentices mask heterogeneity across groups using a simplified version of Equation 1 that pools all periods after January 2007. Appendix Figure A.4 reports the estimates by industry and plots them against the share of apprentices employed in each industry. We find no heterogeneity across industries. Similarly, we do not find any heterogeneity across regions (Appendix Figure A.5), baseline apprentice earnings (Appendix Figure A.6), contemporaneous apprentice earnings (Appendix Figure A.7), three different measures of liquidity constraints (Appendix Figure A.8), and whether firms employed apprentices at baseline (Appendix Figure A.9). We find no evidence of heterogeneous treatment effects regardless of how we group firms.

4.5 Reduced-Form Effects On Other Outcomes

In this section, we examine firms' strategic responses to the reform. We see no effect.

First, firms do not churn through more apprentices. Panel A of Appendix Table A.3 shows that the reform does not impact the contract length, the number of new apprenticeship

[5] While the null wage bill effects are precise in absolute terms, they are not precise enough to reliably apportion the tax break's incidence, owing to variation in apprentice tax bill across firms.

contracts, or apprentice separations. Notably, firms do not decrease their rate of transforming apprentices to permanent workers.

Second, firms do not "re-label" existing contracts as apprenticeships to take advantage of the lower social contributions in the first two years. In Panel A of Appendix Table A.3, we see no increases in the number of new apprenticeship contracts or decreases in the number of apprentices hired from outside the firm.

Third, firms do not "lower the bar" to hire more apprentices. In Panel B of Appendix Table A.3, we see no changes in characteristics of hired apprentices, including those that correlate with ability such as previous salary, previous experience, or starting salary.

Fourth, firms do not substitute away from temporary workers to apprentices. Appendix Table A.3 shows no effects on temporary worker hires and separations (Panel A) and, consequently, no effects on the stock of temporary workers (Panel C).

Fifth, firms do not substitute toward or away from permanent workers to apprentices. Whether they are substitutes or complements, directly estimating the threshold's effects on permanent workers is complicated by the fact that they are used to compute policy-relevant firm size, the running variable pivotal to our design. However, the stability of the marginal distributions of policy-relevant firm size and absence of bunching over time (Figure 2) point away from the policy's incentives affecting the firm's permanent labor demand.

Altogether, our evidence suggests that the policy subsidized inframarginal decisions with no corresponding increases in labor demand or substitution effects.

4.6 Why No Reduced-Form Effects?

Here, we consider several explanations for our null results: (1) measurement error; (2) the size of the subsidy; (3) a lack of saliency or awareness; (4) firm incentives to maintain eligibility; (5) the temporary nature of the subsidy; (6) training costs; and (7) the Great Recession. None of these can explain our findings. We conclude that the demand for apprentices is simply inelastic, which we corroborate with data from the RIL survey.

Measurement error? No. Our monthly data is high-quality and high-frequency. However, we do not measure the running variable, policy-relevant firm size, at the precise moment that firms hire apprentices. If firm size fluctuates within a given month, measurement error may attenuate the reduced-form results toward zero.

To avoid our null results being a consequence of measurement error, our preferred specifications exclude firms within a window of 1 of the threshold, relying on the pre-period discontinuity to remove the bias associated with extrapolation. Our null results on apprentice hiring are robust to the amount of excluded data (Appendix Figure A.2). Moreover, measurement error does not prevent us from finding significant effects on fiscal outcomes like SSC. We find it unlikely that measurement error in the running variable affects only the treatment effects of employment outcomes.

Is the subsidy too small? No. The size of the subsidy is substantial, worth roughly two months of earnings for the average apprenticeship contract or 8% of the earnings for a typical 19-month apprenticeship. This amounts to a subsidy of 960 euros per apprentice per year for firm paying average earnings and reaches 1,460 euros per apprentice per year for businesses paying the 95th percentile of the apprentices' earnings distribution. The SSC subsidy is similar in size to the subsidy studied by Cahuc et al. (2019) and Guo (2024), who find large employment effects on targeted workers. Specifically, the size of our 8% subsidy is in the same ballpark as the one analyzed by Cahuc et al. (2019), which amounts 4% of labor costs for workers paid 30% more than the minimum wage and can range from a minimum of 0% to a maximum of 12%. Our subsidy (960 euros per worker per year) is larger than the one examined by Guo (2024) (200-600 dollars per worker per year).

Are firms unaware of the policy? No. One possibility is that firms do not respond to the SSC discount because they were unaware of it. It is worth noting that the SSC discount is not applied automatically: firms must claim it. Figure A.1 plots the share of firms that take up the policy among those that hire apprentices against policy-relevant firm size. Ultimately, 80% of eligible firms that hired apprentices received the discount and must be aware of the policy.

Do firms restrict apprentice hiring to maintain eligibility? No. Importantly, hiring apprentices does not affect eligibility because apprentices are not included in policy-relevant firm size. Moreover, firms do not appear to restrict permanent employee hiring dynamically. First, the policy never induces firms to hire an additional apprentice. Second, the marginal distribution of firm size is remarkably stable. If firms are systematically keeping their firm size below the threshold to maintain eligibility, we should observe increased mass below the policy threshold and decreased mass above it. None of these patterns emerges.

Does the temporary nature of the subsidy hinder its effectiveness? No. The subsidy covers the first two years of each apprenticeship contract at eligible firms. Because the typical apprenticeship contract lasts 19 months, most contracts are effectively subsidied for their entire duration. Moreover, previous studies suggest that temporary subsidies should be, if anything, more effective than permanent ones (Cahuc et al., 2019).

Are the null results driven by training requirements? No. One may be concerned that because of the training requirement, firms do not respond to the policy and hire apprentices. Three pieces of evidence push against this concern. First, training requirements are poorly enforced (Tiraboschi, 2014). Second, we find no effects among firms that hired apprentices prior to the policy (Appendix Figure A.9), firms that should face lower (fixed) training costs. Third, only a small fraction of firms report that training costs deter them from hiring apprentices (Panel A of Figure 9).

Are the effects suppressed by the Great Recession? Unlikely. There are several reasons why we do not believe that our results are driven by the Great Recession. First, the policy was not created in response to the Recession. The first full year of the policy took place prior to the recession. We see no effects during that time period. Second, recessions may make firms liquidity constrained; however, the effects are not different based on whether firms are liquidity constrained or not (Appendix Figure A.8). Finally, recent research suggests that effects may in fact be larger in recessions (Cahuc et al., 2019; Benzarti and Harju, 2021a).

Inelastic demand. We conclude that measurement error, the size of the subsidy, the lack of saliency, firm incentives to maintain eligibility, the temporary nature of the subsidy, the training requirements, and the Great Recession are unlikely to explain our results. Firms simply exhibit inelastic demand for apprentices. Survey evidence corroborates this argument. When asked why they do not hire apprentices, firms' most common reason is that they do not need more people (Figure 9, Panel A). When asked why they do hire apprentices, firms' most common reason is to provide training prior to hiring a new permanent employee (Panel B). In neither case is cost a primary consideration (Aepli et al., 2024). These results are in line with Egebark and Kaunitz (2013) and Huttunen et al. (2013), who find very modest to null effects of comparable policies.

5 Cost Effectiveness

The objective of this paper is to measure the cost-effectiveness of payroll tax reductions as jobs supported per unit of foregone revenue. This section is organized as follows. First, we explain the advantages of formally measuring cost-effectiveness using an instrumental variable strategy. Second, we report IV estimates of apprenticeships supported per unit of revenue. Lastly, we compare the IV estimates derived from the Italian reform to back-of-the-envelope measures of jobs per unit revenue reported in previous studies.

5.1 Measuring Cost-Effectiveness using Instrumental Variables

Denote firm i's payroll tax payments as R_i, and their employment as L_i^*. The number of jobs the policy supports per unit of revenue the policy forgoes is given by $\gamma = -\frac{\mathbf{E}\left[L_i^*(1)-L_i^*(0)\right]}{\mathbf{E}[R_i(1)-R_i(0)]}$, where $L_i^*(z)$ and $R_i^*(z)$ index firm i's potential outcomes, with and without the policy. γ can be estimated in a back-of-the-envelope calculation. Instead, we incorporate our difference-in-discontinuity design in an IV regression, allowing us to compute standard errors for γ. Specifically, we estimate the following system using 2SLS:

$$\begin{aligned} L_{ijt}^* &= -\gamma R_{it} + g_L(Z_{it}, t) + \varepsilon_{it} \\ R_{it} &= bT_{it} \times Post_t + g_R(Z_{it}, t) + \eta_{it}, \end{aligned} \tag{2}$$

where L_{ijt}^* measures employment of type j (the outcome), and R_{it} measures social security contributions (the endogenous regressor). The excluded instrument is $T_{it} \times Post_t$ and $g_Y(Z_{it}, t)$ are controls for time dummies and the running variable in each period.

Equation 2 differs from Equation 1 only because it averages the dynamic effects into a single parameter so that the system is just-identified. For example, the first-stage equation for R_{it} is identical to its reduced-form specification, except there is a single parameter b corresponding to a single $T_{it} \times Post_t$ indicator rather than the set b_t parameters corresponding to each of the time dummies $T_{it} \times \Delta_t$. Appendix Table A.4 reports the first stage coefficient estimate, which is highly statistically significant with an F-statistic of 230.

Before presenting the IV results, it is important to note that estimating the reciprocal cost per job, $\frac{1}{\gamma}$, using instrumental variables is unlikely to yield meaningful insights. Since the "first stage" in this context refers to the effect of the reform on employment, a reform that has no impact on job creation—such as the one analyzed in this paper—lacks a valid first stage. As a result, the standard errors would be extremely large, rendering the estimates effectively uninformative for assessing cost-effectiveness.

Our instrumental variable approach to estimating policy cost-effectiveness offers two key advantages: it provides a consistent metric—jobs supported per unit of foregone revenue—for comparing effectiveness across different policy contexts, and it allows for the assessment of statistical precision even when the reform has no measurable impact on employment.

5.2 IV Estimates of Cost-Effectiveness

Table 4 reports the IV estimates. In each month, the point estimates imply that €1M of lost social security contribution revenue supports the employment of 29 apprentices for one year. The effects are not statistically different from zero. By comparison, for €1M one can hire 79 apprentices at their prevailing wage (1M/1050), making direct hiring of apprentices 2.7 times (79/29) as cost effective as subsidizing firms.

Increased apprenticeships are only an intermediate goal; the ultimate goal of subsidizing apprenticeships is increasing permanent employment. Only a subset of subsidized apprentices become permanent employees. Thus, one can alternatively evaluate the subsidy against the ultimate goal, using as the endogenous variable the number of apprentices that transformed into permanent contracts. In line with our point estimates, €1M of lost social security contribution revenue does not support any transformed contracts (the point estimate is negative). Altogether, these estimates suggest that targeted payroll tax cuts are not a cost-effective method of supporting both the temporary and permanent employment of marginalized workers in the short term. Importantly, our cost-effectiveness estimates reflect only the direct fiscal cost of the policy and do not incorporate fiscal externalities or administrative and political costs.

While we find that offering small firms a discount on social security contributions for apprentices was not cost-effective, it is important to note that this result reflects inelastic labor demand. Consequently, taking our estimates at face value suggests both that (1) the pre-reform apprentice subsidies were little more than a transfer to firms and (2) the increase in social security contributions for larger employers effectively raised revenue without causing significant employment losses.

5.3 Measures of Cost Effectiveness Across Studies

Considering reduced-form employment effects and ignoring costs, our study adds a precise zero to the collection of mixed results on payroll taxes (Benzarti and Harju, 2021b,a; Bohm and Lind, 1993; Bennmarker et al., 2009; Korkeamaki and Uusitalo, 2006; Saez et al., 2019, 2021). However, the wage subsidy programs are difficult to compare because they have different features and vary in fiscal costs. Only a small subset of studies have evaluated the cost-effectiveness of these reforms (Cahuc et al., 2019; Egebark and Kaunitz, 2013; Neumark, 2013; Saez et al., 2021). Examining differing policies across different countries is inherently difficult, but normalizing employment effects against fiscal costs offers a unified way of comparing results across studies. Here, we compute the implied number of jobs supported by € 1 million of foregone revenue implied by structural or back-of-the-envelope estimates and compare the literature to our IV estimates, emphasizing that the policies examined by the included studies differ in targeted populations.[6]

[6]The specific studies are Bartik (2001); Bartik and Erickcek (2010); Dupor and Mehkari (2016); Dupor and McCrory (2018); Egebark and Kaunitz (2013); Feyrer and Sacerdote (2011); Neumark (2013); Saez et al. (2021); Wilson (2012). The estimates of cost-effectiveness for Bartik (2001) and Bartik and Erickcek (2010)

Figure 10 reports the results. We find that payroll tax cuts (orange triangles) and most wage subsidies, more broadly, are not cost-effective. With two notable exceptions (Bartik, 2001; Cahuc et al., 2019), Figure 10 suggests that the cost of generating employment effects is extremely high, even for programs that generate positive employment effects (Saez et al., 2019, 2021). This figure suggests that hiring credits (hollow circles) may be more cost-effective than payroll tax cuts. Firms must hire new employees to receive hiring credits, making it less likely that the policy subsidizes inframarginal employment.

6 Conclusion

This paper studies the cost-effectiveness of a targeted payroll tax cut in stimulating labor demand. Using a difference-in-discontinuities framework, we find that the reduction in SSCs did not have employment effects for either apprentices or their substitutes. The program also did not have discernible effects on apprentice earnings. Its only effects were on tax revenue.

To evaluate the cost-effectiveness of the policy, we use the policy variation in an instrumental variables strategy to estimate the number of jobs sustained by each euro of foregone revenues. Over the first three years, each €1 million euro of lost social security contribution supports the employment of 29 apprentices for one month and no open-ended positions (and the estimates are not statistically significant). This both implies that giving tax relief to small firms did not have the desired effect of supporting employment and that raising taxes on large firms did not come with large disemployment effects.

Our precise null employment effects contrast with the literature, which lacks consensus on the responsiveness of labor demand to policy. However, when benchmarking other studies against their fiscal cost, our IV approach yields estimates that generally accord with other studies: increasing employment using wage subsidies comes at enormous cost.

However, our study is not without limitations. The estimates are internally valid for firms in the neighborhood of 9 employees, and our well-identified partial equilibrium analysis cannot measure spillovers or other general equilibrium effects. Nevertheless, these results suggest caution in the use of payroll tax credits to stimulate employment.

are taken from Neumark (2013). When available, we used estimates of the policies' effects on job-years. When not, we used estimates on the number of jobs. We do not include confidence intervals because the studies generally did not include standard errors on their estimates.

References

Aepli, M., Muehlemann, S., Pfeifer, H., Schweri, J., Wenzelmann, F., and Wolter, S. C. (2024). The impact of hiring costs for skilled workers on apprenticeship training: A comparative study.

Bartik, T. and Erickcek, G. A. (2010). The employment and fiscal effects of michigan's mega tax credit program.

Bartik, T. J. (2001). Jobs for the poor: Can labor demand policies help?

Bell, B., Blundell, R., and Reenen, J. V. (1999). Getting the unemployed back to work: the role of targeted wage subsidies. *International tax and public finance*, 6(3):339–360.

Bennmarker, H., Mellander, E., and Öckert, B. (2009). Do regional payroll tax reductions boost employment? *Labour Economics*, 16(5):480–489.

Benzarti, Y. and Harju, J. (2021a). Can payroll tax cuts help firms during recessions? *Journal of Public Economics*, 200:104472.

Benzarti, Y. and Harju, J. (2021b). Using Payroll Tax Variation to Unpack the Black Box of Firm-Level Production. *Journal of the European Economic Association*, 19(5):2737–2764.

Benzarti, Y., Harju, J., and Matikka, T. (2020). Does mandating social insurance affect entrepreneurial activity? *American Economic Review: Insights*, 2(2):255–268.

Bohm, P. and Lind, H. (1993). Policy evaluation quality: A quasi-experimental study of regional employment subsidies in sweden. *Regional Science and Urban Economics*, 23(1):51–65.

Borusyak, K., Jaravel, X., and Spiess, J. (2021). Revisiting event study designs: Robust and efficient estimation. *arXiv preprint arXiv:2108.12419*.

Bozio, A., Breda, T., and Grenet, J. (2017). Incidence of social security contributions: evidence from france. *Paris School of Economics Working Paper*.

Bozio, A., Breda, T., and Grenet, J. (2020). Does Tax-Benefit Linkage Matter for the Incidence of Social Security Contributions? Working paper.

Cahuc, P., Carcillo, S., and Le Barbanchon, T. (2019). The effectiveness of hiring credits. *The Review of Economic Studies*, 86(2):593–626.

Caicedo, S., Espinosa, M., and Seibold, A. (2020). Unwilling to train? firm responses to the colombian apprenticeship regulation. Technical report, CESifo Working Paper.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326.

De Chaisemartin, C. and D'haultfœuille, X. (2023). Two-way fixed effects and differences-in-differences estimators with several treatments. *Journal of Econometrics*, 236(2):105480.

Dupor, B. and McCrory, P. B. (2018). A cup runneth over: Fiscal policy spillovers from the 2009 recovery act. *The Economic Journal*, 128(611):1476–1508.

Dupor, B. and Mehkari, M. S. (2016). The 2009 recovery act: Stimulus at the extensive and intensive labor margins. *European Economic Review*, 85:208–228.

Egebark, J. and Kaunitz, N. (2013). Do payroll tax cuts raise youth employment? Working Paper Series 2013:27, IFAU - Institute for Evaluation of Labour Market and Education Policy.

Feyrer, J. and Sacerdote, B. (2011). Did the stimulus stimulate? real time estimates of the effects of the american recovery and reinvestment act. Technical report, National Bureau of Economic Research.

Garicano, L., Lelarge, C., and Van Reenen, J. (2016). Firm size distortions and the productivity distribution: Evidence from france. *American Economic Review*, 106(11):3439–79.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.

Goos, M. and Konings, J. (2007). The impact of payroll tax reductions on employment and wages: A natural experiment using firm level data. Technical report, LICOS Discussion Paper.

Grembi, V., Nannicini, T., and Troiano, U. (2016). Do fiscal rules matter? *American Economic Journal: Applied Economics*, pages 1–30.

Guo, A. (2024). Payroll tax incidence: Evidence from unemployment insurance. *Journal of Public Economics*, 239:105209.

Hendren, N. and Sprung-Keyser, B. (2020). A unified welfare analysis of government policies*. *The Quarterly Journal of Economics*, 135(3):1209–1318.

Huttunen, K., Pirttilä, J., and Uusitalo, R. (2013). The employment effects of low-wage subsidies. *Journal of Public Economics*, 97:49–60.

Jain, M., Mommaerts, C., and Weaver, J. (2025). Generating economic opportunity through hiring subsidies: Evidence from the work opportunity tax credit. Working Paper.

Korkeamaki, O. and Uusitalo, R. (2006). Employment Effects of a Payroll-Tax Cut: Evidence from a Regional Tax Exemption Experiment. Discussion Papers 407, VATT Institute for Economic Research.

Neumark, D. (2013). Spurring job creation in response to severe recessions: Reconsidering hiring credits. *Journal of Policy Analysis and Management*, 32(1):142–171.

OECD (2003). *Employment Outlook, Towards More and Better Jobs.*

OECD (2011). *Taxation and Employment.*

Rambachan, A. and Roth, J. (2023). A More Credible Approach to Parallel Trends. *The Review of Economic Studies*, 90(5):2555–2591.

Roth, J. and Sant'Anna, P. H. (2023). When is parallel trends sensitive to functional form? *Econometrica*, 91(2):737–747.

Roth, J., Sant'Anna, P. H., Bilinski, A., and Poe, J. (2023). What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics.*

Rubolino, E. (2021). Taxing the Gender Gap: Labor Market Effects of a Payroll Tax Cut for Women in Italy. Working paper.

Ryan, P. (2012). Apprenticeship: between theory and practice, school and workplace. In *The future of vocational education and training in a changing world*, pages 402–432. Springer.

Saez, E., Matsaganis, M., and Tsakloglou, P. (2012). Earnings determination and taxes: Evidence from a cohort-based payroll tax reform in greece. *The Quarterly Journal of Economics*, 127(1):493–533.

Saez, E., Schoefer, B., and Seim, D. (2019). Payroll taxes, firm behavior, and rent sharing: Evidence from a young workers' tax cut in sweden. *American Economic Review*, 109(5):1717–63.

Saez, E., Schoefer, B., and Seim, D. (2021). Hysteresis from employer subsidies. *Journal of Public Economics*, 200:104459.

Snell, K. D. (1996). The apprenticeship system in british history: the fragmentation of a cultural institution. *History of Education*, 25(4):303–321.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.

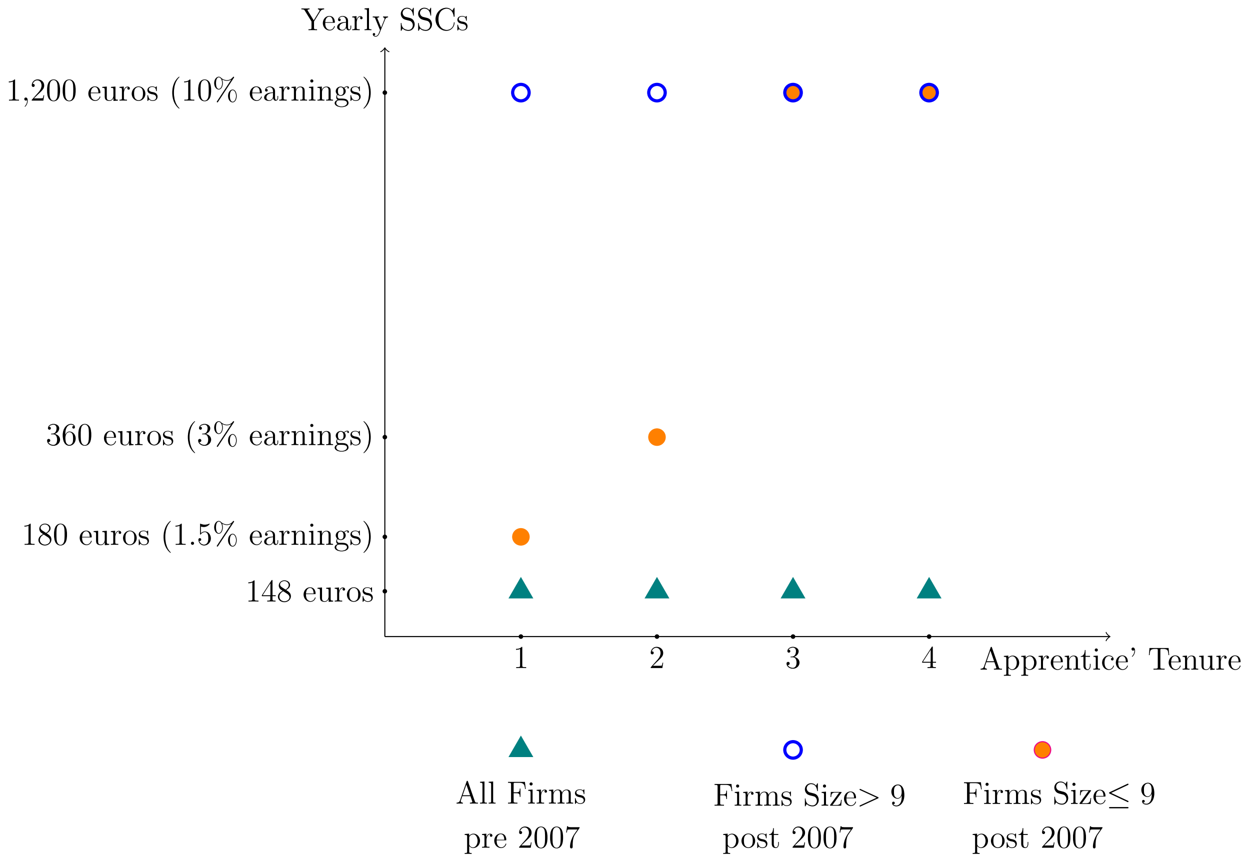
Tiraboschi, M. (2014). Young workers in recessionary times: A caveat (to continental europe) to reconstruct its labour law? *GundersonM. FazioF.(Eds.), Tackling youth unemployment*, pages 3–26.

Van Reenen, J. (2003). Active labour market policies and the british new deal for the young unemployed in context.

Wilson, D. J. (2012). Fiscal spending jobs multipliers: Evidence from the 2009 american recovery and reinvestment act. *American Economic Journal: Economic Policy*, 4(3):251–282.

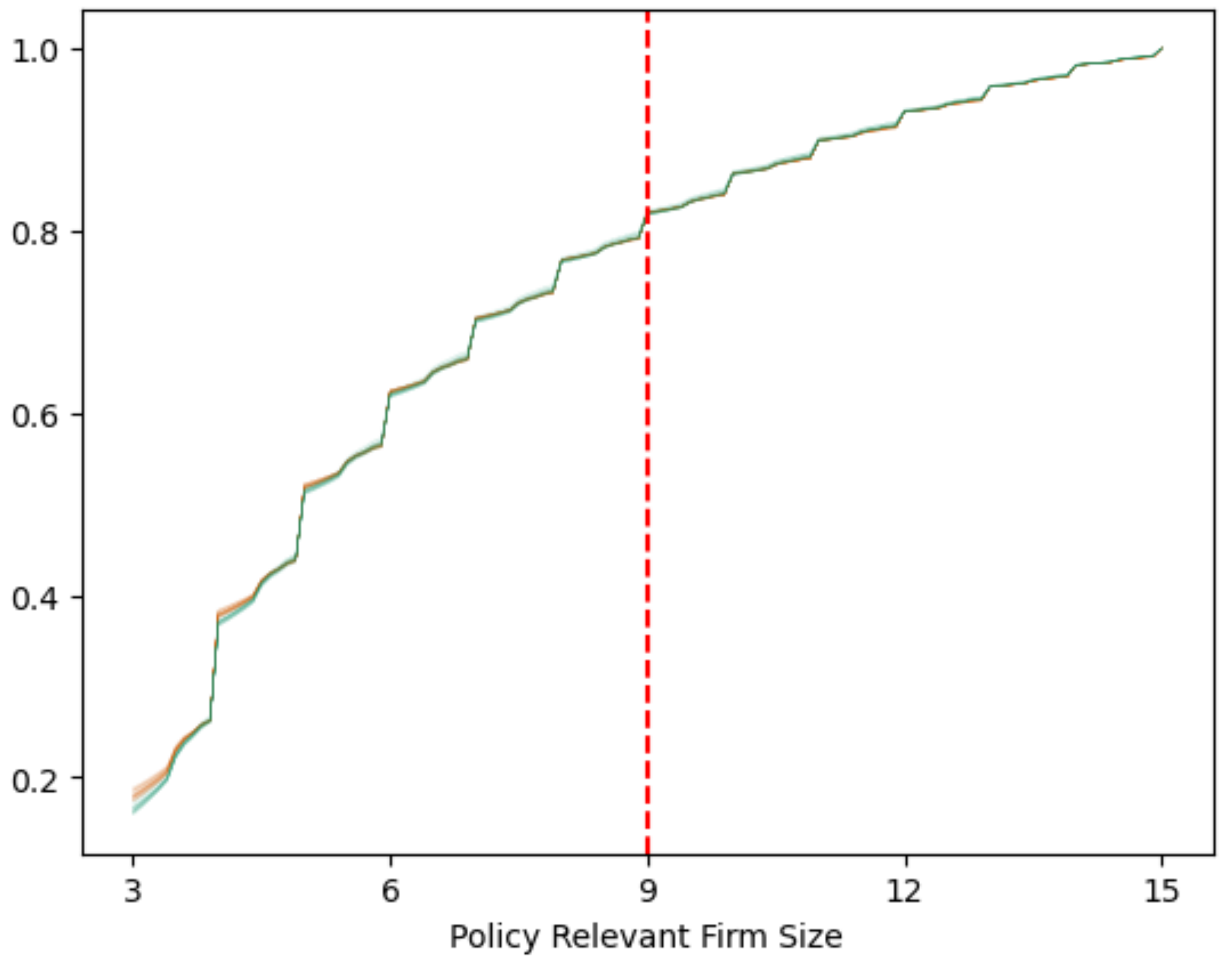
7 Figures

Figure 1: Social Security Contributions for Apprenticeship Contracts



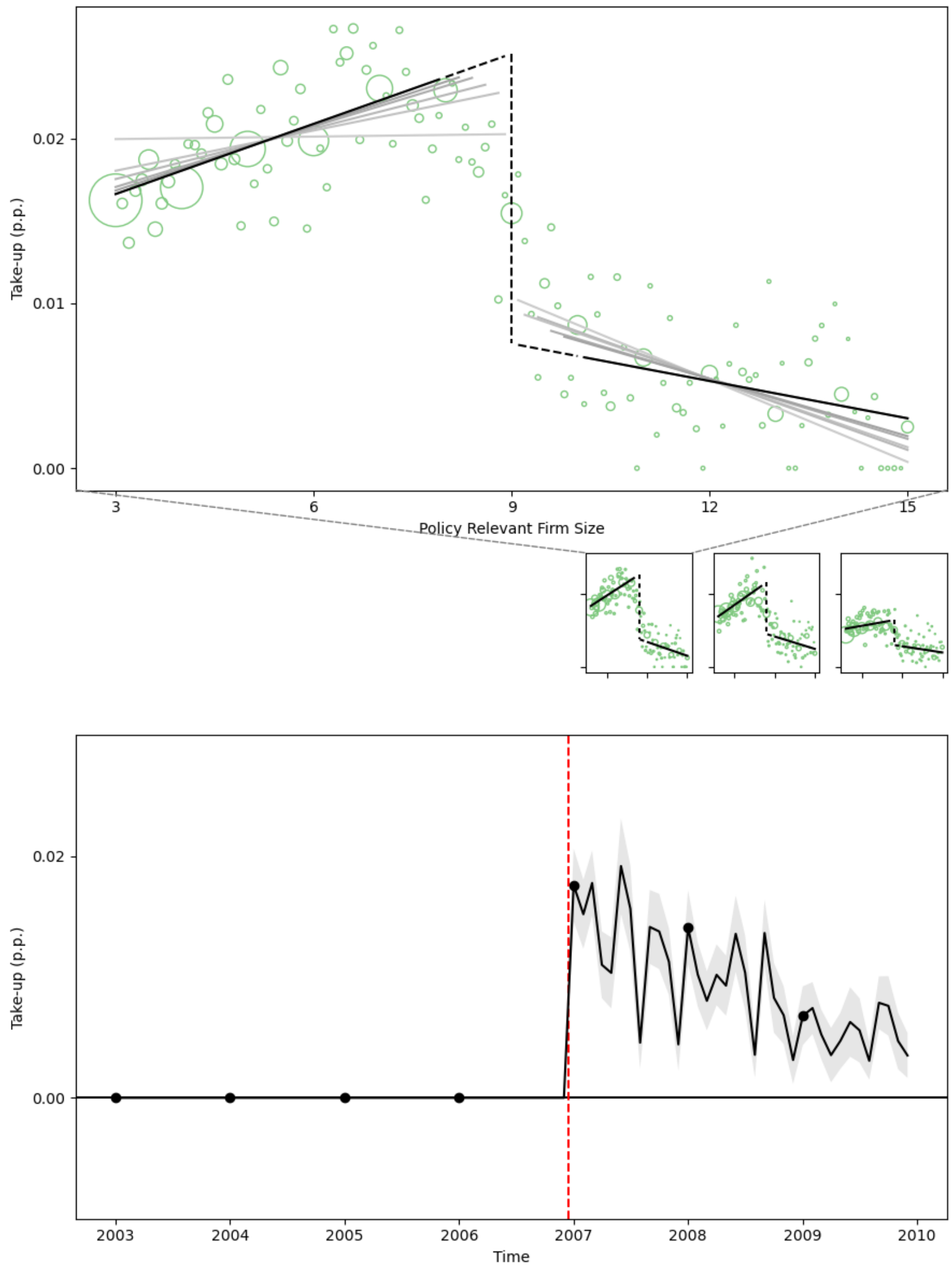
Notes: This figure illustrates how yearly social security contributions (SSCs) for apprenticeship contracts changed in response to the 2007 Budget Bill. Before 2007, employers paid a fixed weekly fee of 2.85 euros per apprenticeship contract. The yearly social contributions are computed as $2.85 \times 52 = 148.2$ euros (green triangles). After January 1, 2007, yearly social contributions are computed as a percentage of the apprentice's yearly earnings; their schedule differs between firms below or above the 9-employee threshold. Social contributions amount to 10% of the apprentice's earnings for firms with more than 9 employees (blue hollow circles). Firms with 9 employees or less pay 1.5% of the apprentice's earnings in the first year of the contract, 3% in the second year, and 10% in the third year and all the following ones (orange circles). To compute the change in social contributions implied by this policy, we use the average 2006 yearly earnings, which are equal to 12,000 euros.

Figure 2: Empirical CDFs of Policy Relevant Firm Size, Jan 2003–Dec 2009



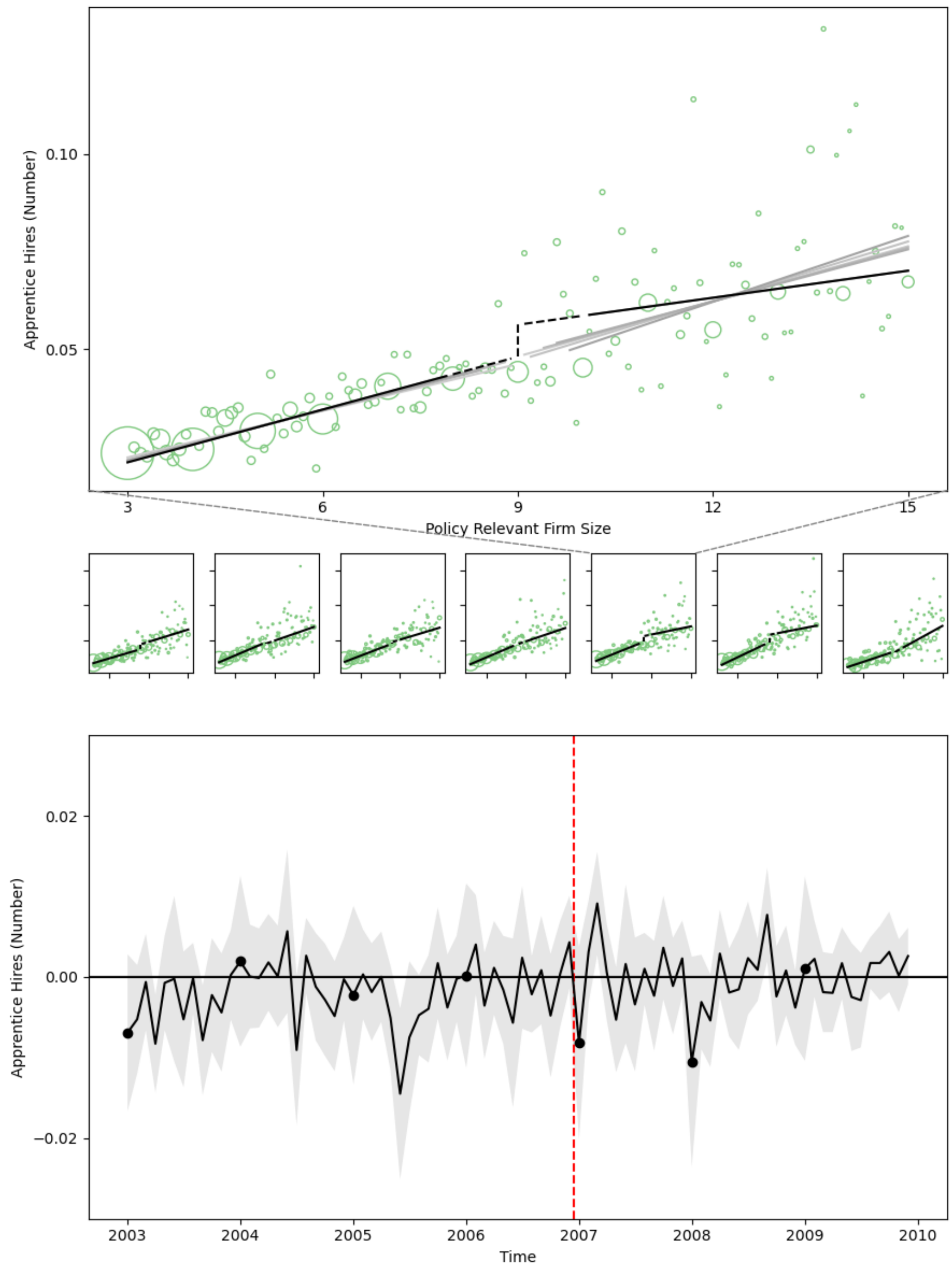
Notes: Social Security Administration data (January 2003–December 2009). This figure overlays all 84 monthly CDFs of policy-relevant firm size from Jan 2003 to Dec 2009 for firms with a policy-relevant firm size between 3 and 15. CDFs prior to Jan 2007 are plotted in orange. Those subsequent to Jan 2007 are plotted in green. Because they overlap, most CDFs are not visible.

Figure 3: Reduced-form Effects on Take-up



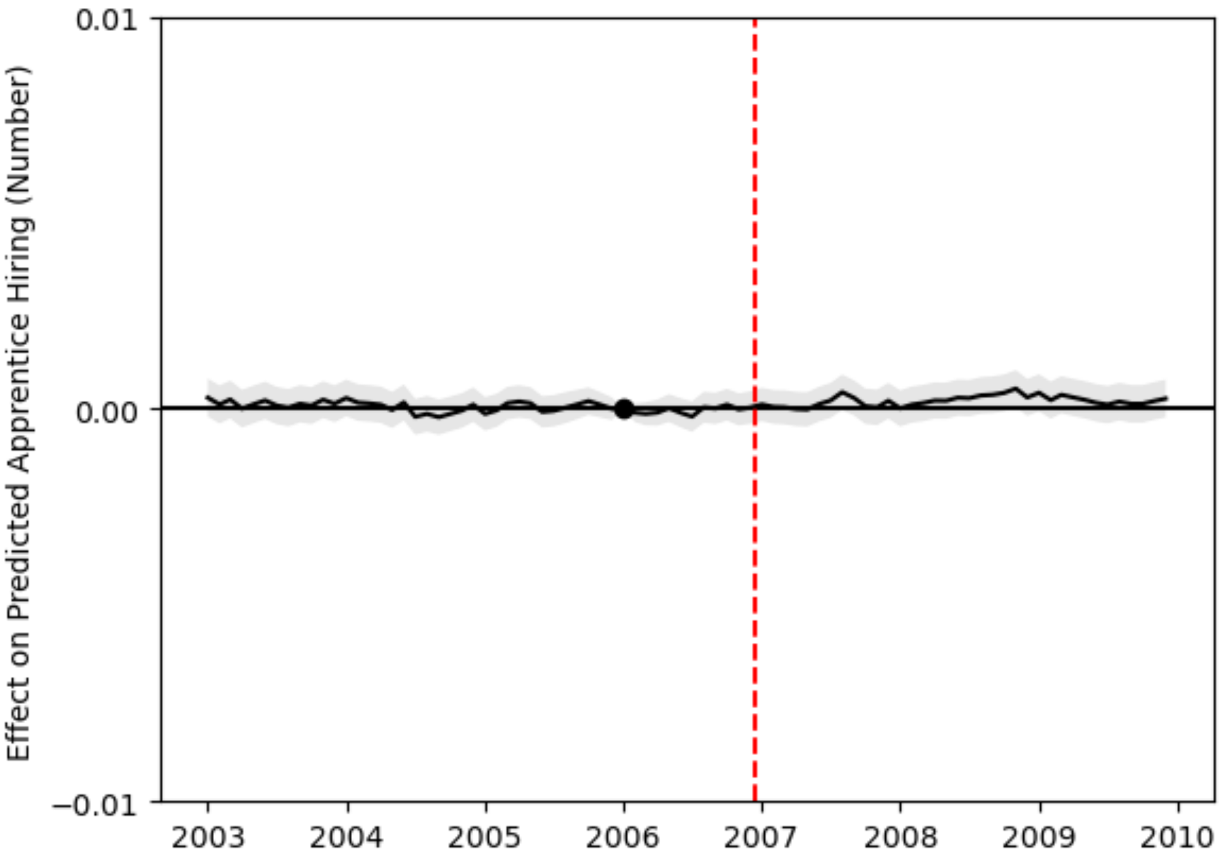
Notes: Social Security Administration data (January 2003–December 2009). This figure shows the effect of the policy on take-up. The top panel shows a binned scatterplot of take-up against policy-relevant firm size in January 2007, the first month of the policy. The size of the green dots indicates the number of firms within the bin. Fitted values from piece-wise linear regressions are overlayed. The black line indicates regressions estimated, excluding a window of 1 around the discontinuity. (Grey lines are fit using windows of 0, 0.2, 0.4, and 0.8.) The first panel is a zoomed example of the conditional expectation function in each period, shown in the second panel. The third panel plots a time series of the discontinuity estimates. 95% confidence intervals are shaded in grey. Note that take-up is mechanically zero before January 2007.

Figure 4: Reduced-form Effects on Apprentice Hiring



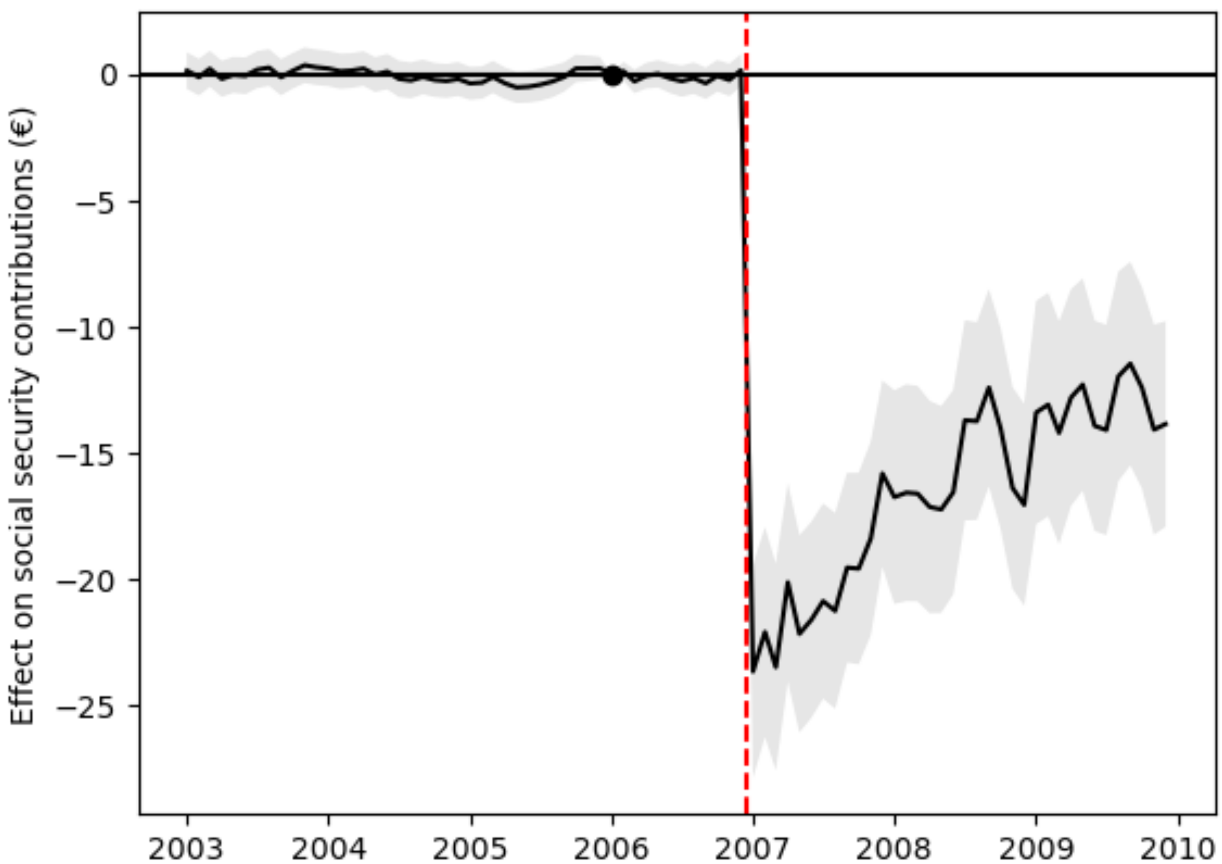
Notes: Social Security Administration data (January 2003–December 2009). This figure shows the effect of the policy on apprentice hiring, mirroring Figure 3. See notes for Figure 3 for details.

Figure 5: Reduced Form Estimates of Threshold on Covariate Index



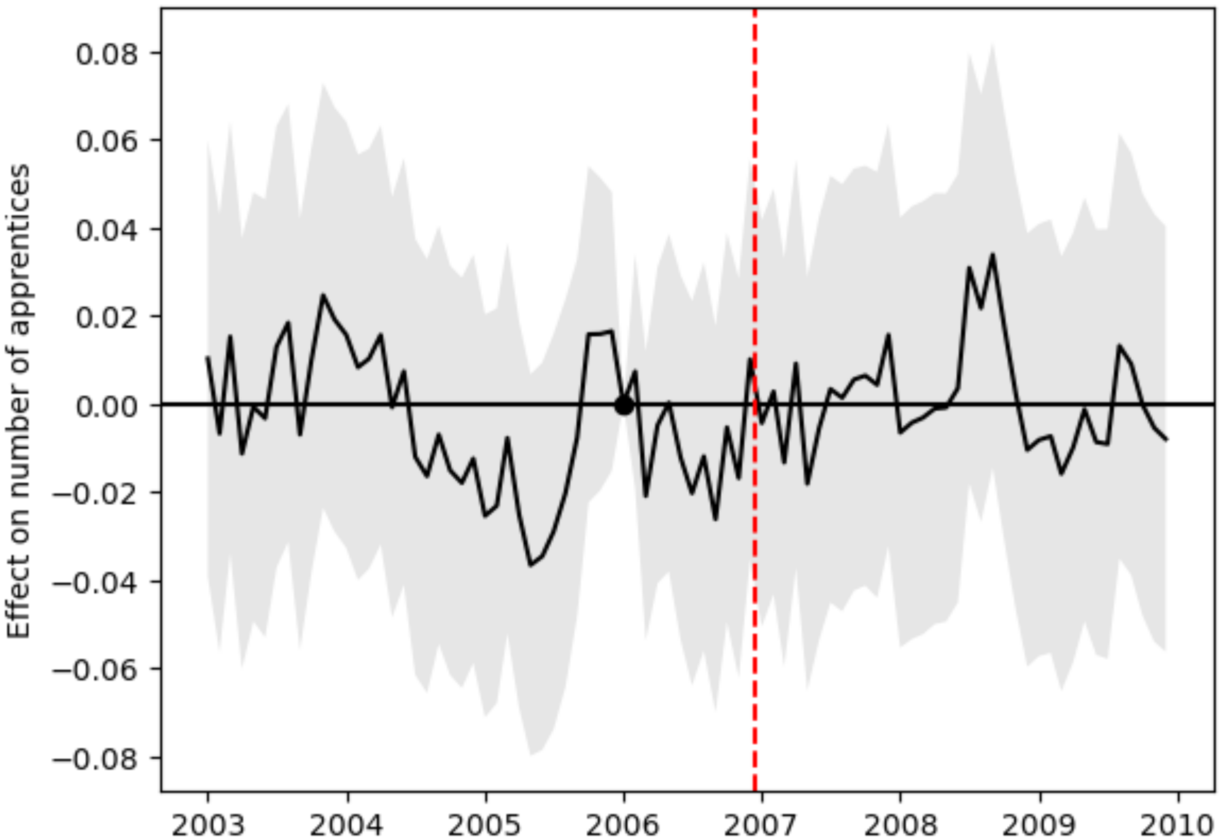
Notes: Social Security Administration data (January 2003–December 2009). This figure reports the effects of being below the subsidy threshold (b_t) from the main DD specification in equation Equation 1 where the outcome variable is a covariate index, the predicted values from a regression of apprentice hiring on time-invariant firm characteristics. Estimates are relative to January 2006, the omitted category. 95% confidence intervals are shaded in grey.

Figure 6: Reduced Form Estimates of Threshold on Social Security Contributions



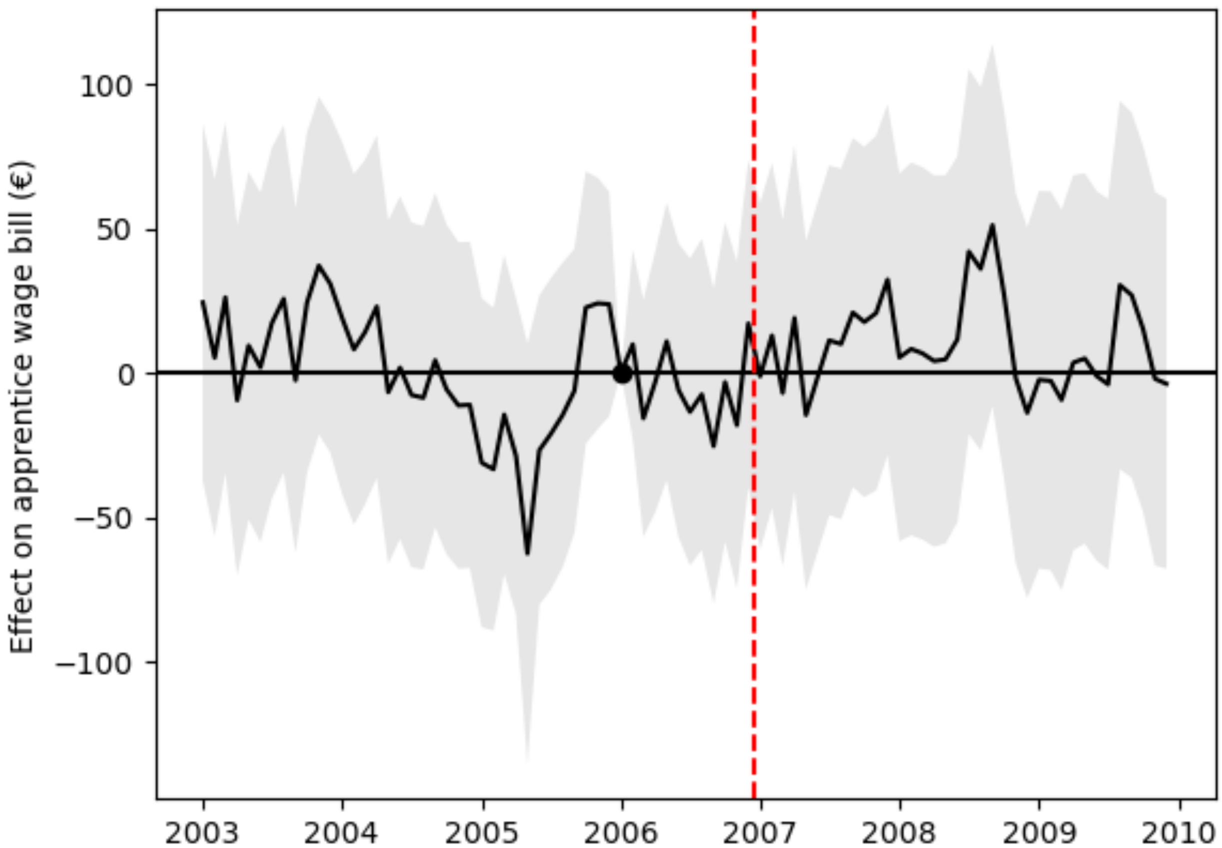
Notes: Social Security Administration data (January 2003–December 2009). This figure reports the effects of being below the subsidy threshold (b_t) from the main DD specification in Equation 1 where the outcome variable is firm's total social security contribution. See Figure 5 notes for details.

Figure 7: Reduced Form Estimates of Threshold on Number of Apprentices



Notes: Social Security Administration data (January 2003–December 2009). This figure reports the effects of being below the subsidy threshold (b_t) from the main DD specification in Equation 1 where the outcome variable is the firm's number of apprentices. See Figure 5 notes for details.

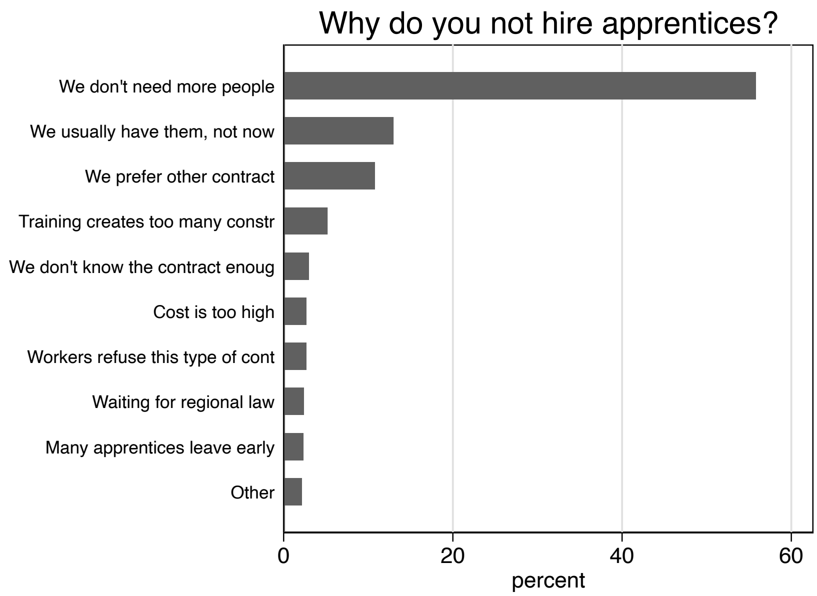
Figure 8: Reduced Form Estimates of Threshold on Apprentice Wage Bill



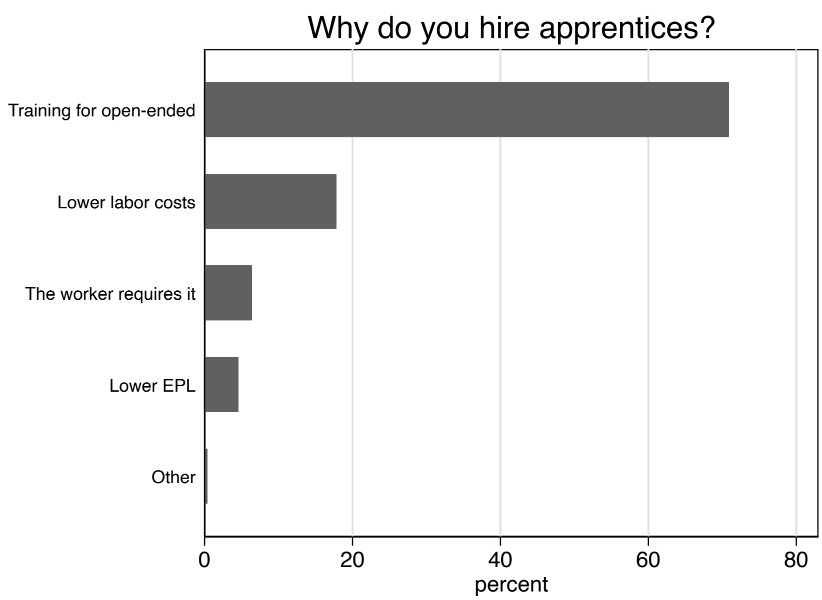
Notes: Social Security Administration data (January 2003–December 2009). This figure reports the effects of being below the subsidy threshold (b_t) from the main DD specification in Equation 1, where the outcome variable is the firm's wage bill for their apprentices. See Figure 5 notes for details.

Figure 9: Labor Demand For Apprentices

(a)

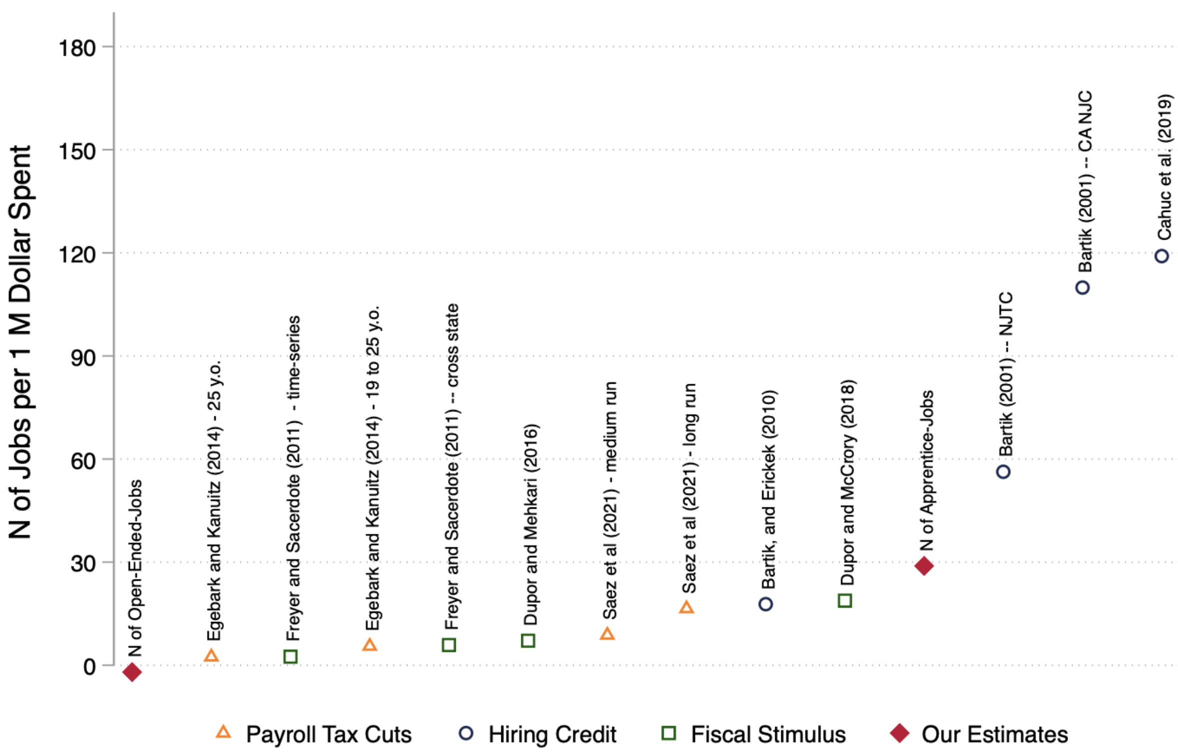


(b)



Notes: RIL data (2005). Panels (a) and (b) illustrate firms' answers to the questions "Why don't you hire apprentices?" and "Why do you hire apprentices?", respectively.

Figure 10: The Cost Effectiveness of Wage Subsidies



Notes: This figure reports the number of jobs/job-years supported by 1 million dollars spent. We compare our estimates (red diamonds) with those from other studies on payroll tax cuts (orange triangles), hiring credits (blue circles), and fiscal stimulus (green squares).

8 Tables

Table 1: Characteristics of Firms in January 2006

	(1) All firms	(2) Firms with apprentices	(3) Firms that ever take-up
Employees	7.088	8.796	7.056
	[3.634]	[3.822]	[3.015]
Full-time equivalents	6.875	8.606	6.893
	[3.475]	[3.719]	[2.922]
Permanent workers	6.629	8.322	6.611
	[3.552]	[3.727]	[2.979]
Temp workers	0.429	0.458	0.414
	[1.129]	[1.067]	[1.006]
Seasonal workers	0.033	0.028	0.039
	[0.478]	[0.449]	[0.515]
Apprentices	0.427	1.712	0.963
	[0.954]	[1.205]	[1.325]
Apprentice contract length	19.062	19.062	19.895
	[15.073]	[15.073]	[15.402]
Apprentice wage bill	518.750	2081.100	1149.000
	[1188.200]	[1553.200]	[1612.000]
Apprentice wage bill and SSC	524.770	2105.300	1162.700
	[1201.100]	[1568.600]	[1629.800]
Apprentice SSC	6.027	24.180	13.602
	[13.476]	[17.018]	[18.716]
All hires	0.332	0.430	0.312
	[1.354]	[1.640]	[1.122]
Young hires	0.109	0.206	0.137
	[0.538]	[0.817]	[0.579]
Apprentice hires	0.030	0.122	0.066
	[0.261]	[0.512]	[0.352]
Temp hires	0.072	0.078	0.068
	[0.441]	[0.431]	[0.369]
All separations	0.182	0.221	0.153
	[0.690]	[0.741]	[0.464]
Young separations	0.061	0.107	0.070
	[0.309]	[0.414]	[0.293]
Apprentice separations	0.015	0.060	0.032
	[0.140]	[0.276]	[0.191]
Temp separations	0.034	0.036	0.032
	[0.242]	[0.228]	[0.213]
Year established	1992.400	1993.100	1993.200
	[10.630]	[9.796]	[9.681]
Share sole proprietorship	0.217	0.209	0.213
	[0.412]	[0.406]	[0.409]
Share LLC	0.783	0.791	0.787
	[0.412]	[0.407]	[0.409]
N	398,412	99,311	59,670

Notes: Social Security Administration data (January 2006). This table reports the summary statistics for the firms in our sample at baseline (January 2006). The standard deviation is reported in brackets. All statistics are calculated across firm observations. The apprentice contract length is measured in months and is computed among firms that employ apprentices.

Table 2: Covariate Balance: Firm Characteristics and Industry Shares

	(1) t_0-48	(2) t_0-24	(3) t_0	(4) t_0+12	(5) t_0+35	(6) Full	(7) Post	(8) Pre
Panel A: Firm Characteristics								
Age	0.431	-0.064	0.155	0.154	0.297	0.855	0.958	0.729
	(0.223)	(0.198)	(0.199)	(0.216)	(0.231)	<0.824>	<0.554>	<0.884>
General	0.002	0.009	0.003	-0.012	0.003	0.924	0.640	1.271
Partnership	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	<0.673>	<0.974>	<0.128>
LLC	0.002	-0.009	-0.001	0.008	0.002	0.811	0.784	0.799
	(0.011)	(0.009)	(0.009)	(0.010)	(0.010)	<0.894>	<0.857>	<0.798>
Panel B: Industry Shares								
Agriculture	0.001	-0.001	-0.001	-0.001	-0.001	0.911	0.697	1.217
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	<0.704>	<0.943>	<0.174>
Manufacturing	-0.006	0.005	0.003	-0.006	-0.000	1.460	1.260	1.711
	(0.010)	(0.009)	(0.009)	(0.009)	(0.010)	<0.004>	<0.109>	<0.005>
Utilities	-0.002	0.000	-0.004	-0.001	-0.001	1.056	1.153	1.077
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	<0.343>	<0.219>	<0.346>
Transportation	0.013	0.003	-0.002	0.004	0.003	1.310	1.070	1.659
and Construction	(0.008)	(0.007)	(0.007)	(0.008)	(0.008)	<0.031>	<0.345>	<0.008>
Trading	-0.016	0.001	-0.014	-0.003	-0.012	0.834	0.852	0.863
	(0.008)	(0.007)	(0.007)	(0.008)	(0.008)	<0.861>	<0.754>	<0.702>
Services	0.008	-0.010	0.010	0.003	0.009	1.110	1.231	1.007
	(0.008)	(0.007)	(0.007)	(0.008)	(0.008)	<0.230>	<0.133>	<0.457>
Public Admin, Health,	0.005	0.005	0.006	0.004	0.005	0.842	0.986	0.673
and Education	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	<0.847>	<0.499>	<0.932>

Notes: Social Security Administration data (January 2003–December 2009). N=24,532,943. This table reports the effects of being below the threshold (b_t) from the main DD specification in Equation 1 where the outcome variables are general firm characteristics. Each row reports the estimates for a different outcome variable. Estimates are relative to $t_0 - 12$ (January 2006). The first two columns report the pre-reform DD estimates for $t_0 - 48$ (January 2003) and $t_0 - 24$ (January 2004). Columns 3-5 report the post-reform estimates for t_0 (January 2007), t_0+12 (January 2008), and t_0+35 (December 2009), respectively. The last three columns report Wald F-statistics testing the null that all the DD coefficients, the pre-reform coefficients, and the post-reform coefficients are zero, respectively. The dependent variables are firm characteristics and industry dummies in Panels A and B, respectively. Robust standard errors clustering by firms reported in parenthesis. p-values from Wald tests are reported in triangular brackets.

Table 3: Covariate Balance: Regional Shares

	(1) t_0-48	(2) t_0-24	(3) t_0	(4) t_0+12	(5) t_0+35	(6) Full	(7) Post	(8) Pre
Valle d'Aosta	0.000	0.000	-0.000	-0.001	0.000	0.798	0.835	0.717
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	<0.911>	<0.781>	<0.895>
Lombardy	-0.011	0.002	0.004	-0.002	0.006	1.370	1.377	1.326
	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	<0.014>	<0.044>	<0.091>
Piedmont	-0.006	-0.001	0.006	0.001	-0.003	0.830	0.867	0.805
	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	<0.868>	<0.728>	<0.790>
Liguria	0.010	0.006	0.002	0.005	0.001	1.228	1.332	1.104
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	<0.077>	<0.063>	<0.307>
Veneto	0.001	0.004	0.006	0.006	0.008	1.018	1.317	0.692
	(0.007)	(0.006)	(0.006)	(0.006)	(0.007)	<0.434>	<0.071>	<0.918>
Trentino-Alto Adige	-0.005	-0.001	-0.004	-0.000	-0.001	1.012	1.248	0.708
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	<0.449>	<0.119>	<0.904>
Friuli-Venezia Giulia	0.001	-0.006	-0.001	-0.003	-0.003	0.918	1.124	0.693
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	<0.688>	<0.259>	<0.917>
Emilia-Romagna	0.002	-0.004	-0.002	-0.008	0.003	0.888	0.977	0.772
	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)	<0.758>	<0.517>	<0.834>
Tuscany	0.005	0.002	0.002	0.006	0.011	0.995	0.547	1.522
	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	<0.493>	<0.995>	<0.023>
Abruzzo	0.005	0.002	0.004	0.005	0.005	1.062	0.841	1.474
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	<0.330>	<0.771>	<0.033>
Marche	0.000	0.002	-0.006	-0.003	-0.002	1.016	0.972	1.182
	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)	<0.438>	<0.527>	<0.210>
Umbria	0.000	-0.001	-0.001	0.006	0.001	1.390	1.423	1.291
	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	<0.011>	<0.030>	<0.113>
Molise	-0.000	0.001	0.001	-0.001	0.001	1.353	1.362	1.409
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	<0.018>	<0.050>	<0.053>
Basilicata	0.000	0.001	-0.003	-0.000	-0.000	1.134	1.127	1.195
	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	<0.190>	<0.254>	<0.196>
Lazio	-0.005	-0.003	-0.003	-0.005	-0.003	0.701	0.619	0.838
	(0.006)	(0.005)	(0.005)	(0.005)	(0.006)	<0.983>	<0.981>	<0.742>
Campania	0.002	0.001	-0.003	0.000	-0.005	1.087	0.998	1.135
	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	<0.275>	<0.477>	<0.266>
Calabria	-0.002	-0.003	-0.001	-0.003	-0.007	1.019	0.969	1.065
	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	<0.431>	<0.533>	<0.363>
Sicily	0.004	0.001	0.001	-0.000	-0.001	0.904	1.029	0.681
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	<0.721>	<0.418>	<0.927>
Sardinia	0.003	0.001	0.004	0.004	0.001	1.116	1.156	1.073
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	<0.220>	<0.216>	<0.351>
Apulia	-0.005	-0.004	-0.007	-0.007	-0.013	0.954	0.995	1.027
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	<0.599>	<0.482>	<0.424>

Notes: Social Security Administration data (January 2003-–December 2009). N=24,532,943. This table reports the effects of being below the threshold (b_t) from the main DD specification in Equation 1 where the outcome variables are region dummies. See notes to Table 2 for details.

Table 4: IV Estimates of Cost-Effectiveness.

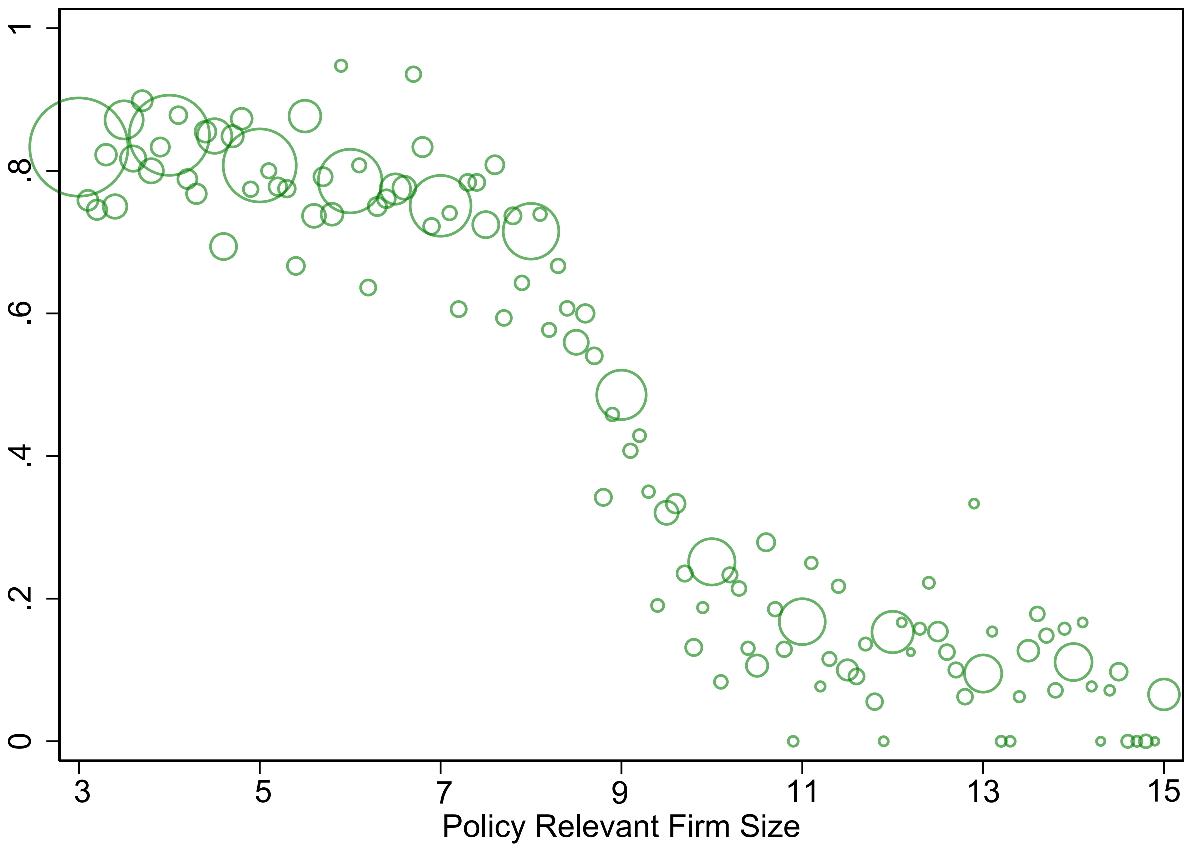
Apprentice Years per €1M	Total Apprentice Compensation per €1M	Transformations per €1M
29	647,237	-2
(58)	(921,320)	(21)

Notes: Social Security Administration data (January 2003–December 2009). N=24,523,943. This table reports IV coefficient estimates of apprentice jobs supported and apprentice compensation supported (β) per €1M of lost social security contributions from Equation 2. The excluded instrument is a dummy variable for being below the policy cut-off in a month after January 2007. Each IV regression controls for policy-relevant firm size and policy-relevant firm size interacted with being below the threshold in each month, mirroring the reduced-form estimates. The first-stage F-statistic is 230 (see Appendix Table A.4). Robust standard errors clustering by firms reported in parenthesis.

Online Appendix

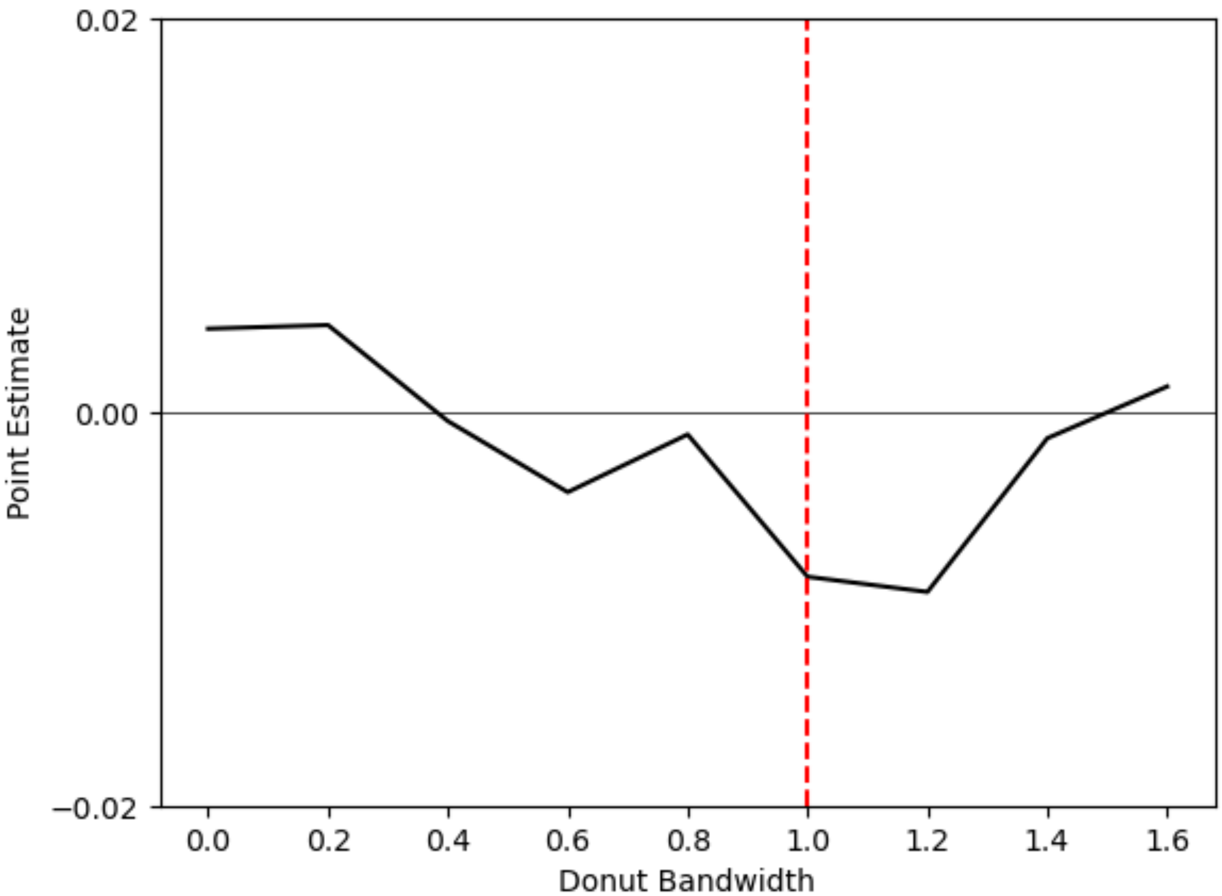
A Appendix Figures and Tables

Figure A.1: Monthly Take Up for Firms that Hire Apprentices



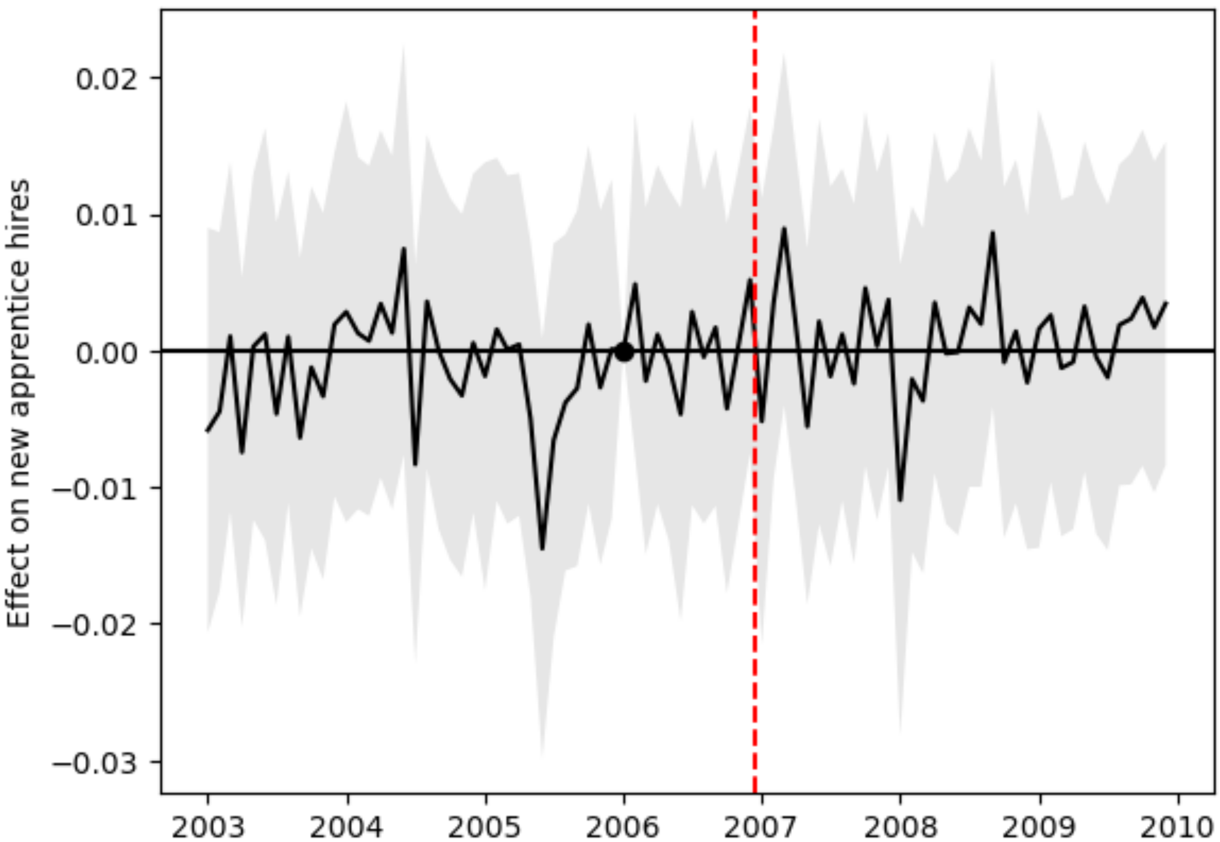
Notes: Social Security Administration data (January 2003–December 2009). This figure shows a binned scatterplot of take-up among firms that hired apprentices in January 2007, plotted against policy-relevant firm size. The size of the green dots reflects the number of firms in each bin. Take-up is defined as a binary variable equal to one if the firm received the subsidy, and zero otherwise.

Figure A.2: Sensitivity of Apprentice Hiring Effects to Amount of Excluded Data



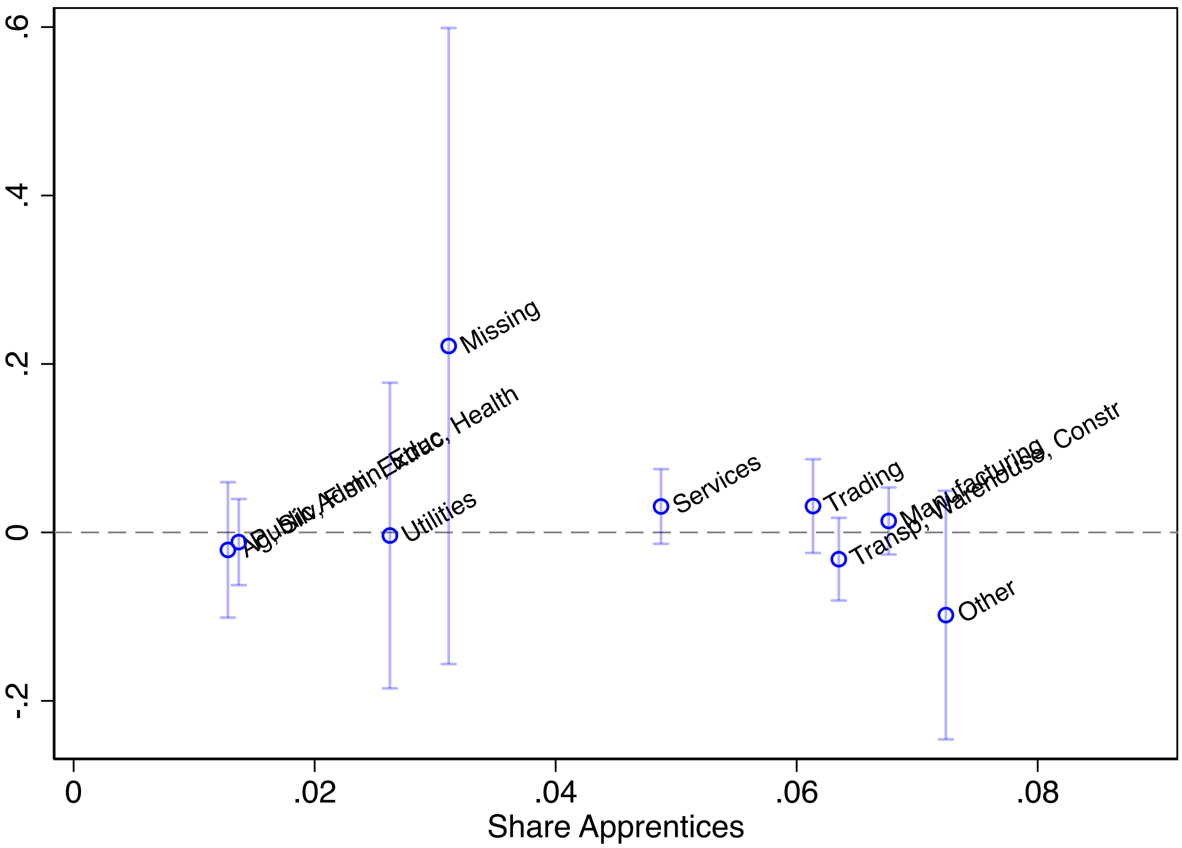
Notes: Social Security Administration data (January 2003–December 2009). This figure shows how the amount of excluded data—the donut bandwidth—affects the coefficient for being below the subsidy threshold in January 2007, ($b_{\text{Jan. 2007}}$) in Equation 1. The outcome variable in this figure is new apprentice hires.

Figure A.3: Reduced Form Estimates of Threshold on Apprentice Hiring



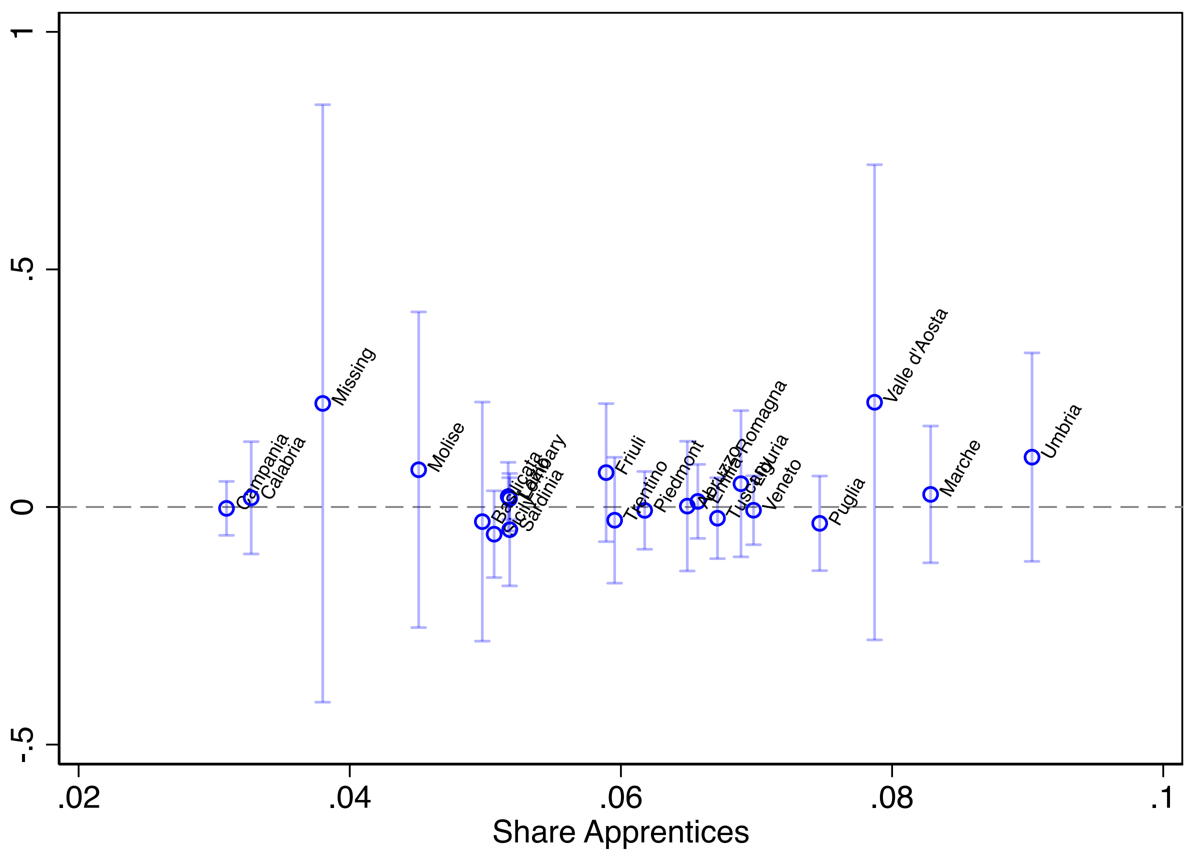
Notes: Social Security Administration data (January 2003–December 2009). This figure reports the effects of being below the subsidy threshold (b_t) from the main DD specification in Equation 1 where the outcome variable is new apprentice hires. See Figure 5 notes for details.

Figure A.4: Heterogeneity by Industry: Reduced Form Estimates of Threshold on Number of Apprentices



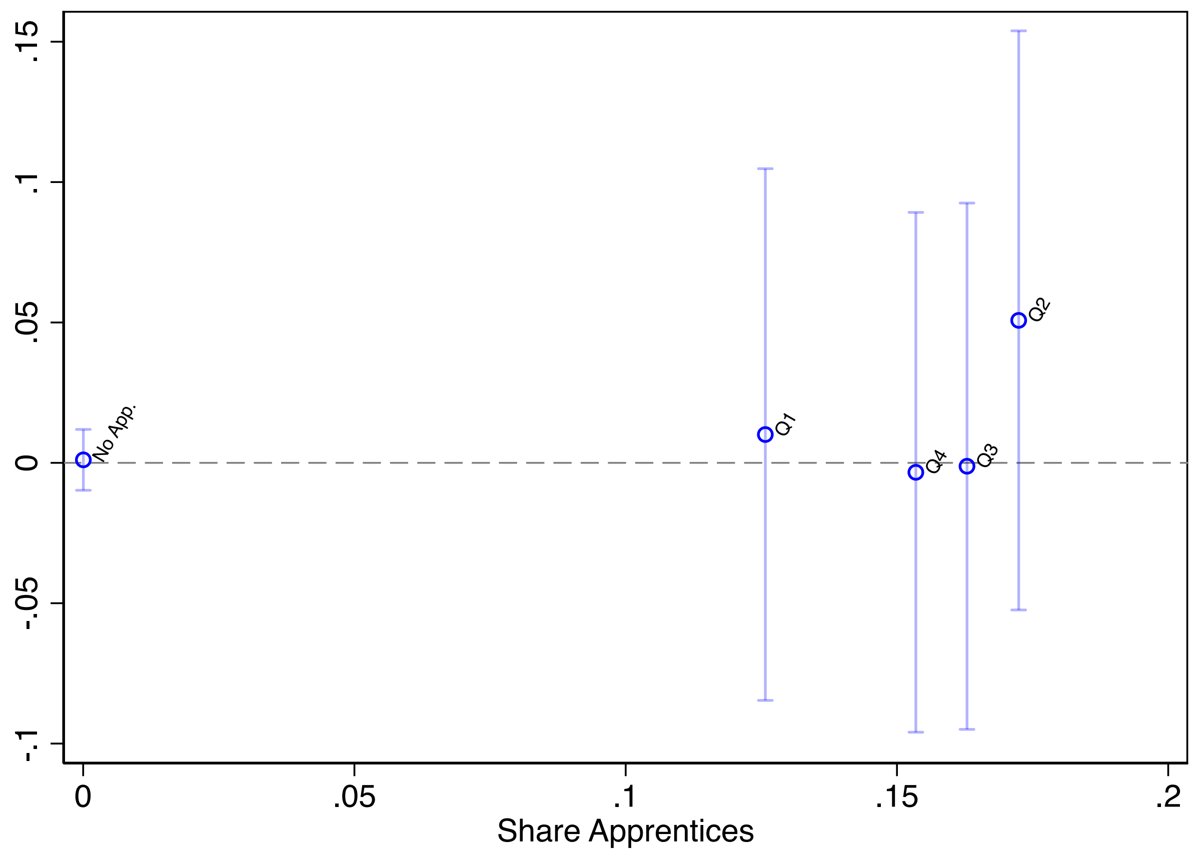
Notes: Social Security Administration data (January 2003–December 2009). We estimate separately by industry the effects of being below the subsidy threshold (b_t) from a model that pools all "post" periods in Equation 1. This figure plots the pooled estimates and their corresponding 95% confidence intervals (vertical axis) against the share of apprentices employed in each industry (horizontal axis). The outcome variable is the number of apprentices. See Figure 5 notes for details.

Figure A.5: Heterogeneity by Region: Reduced Form Estimates of Threshold on Number of Apprentices



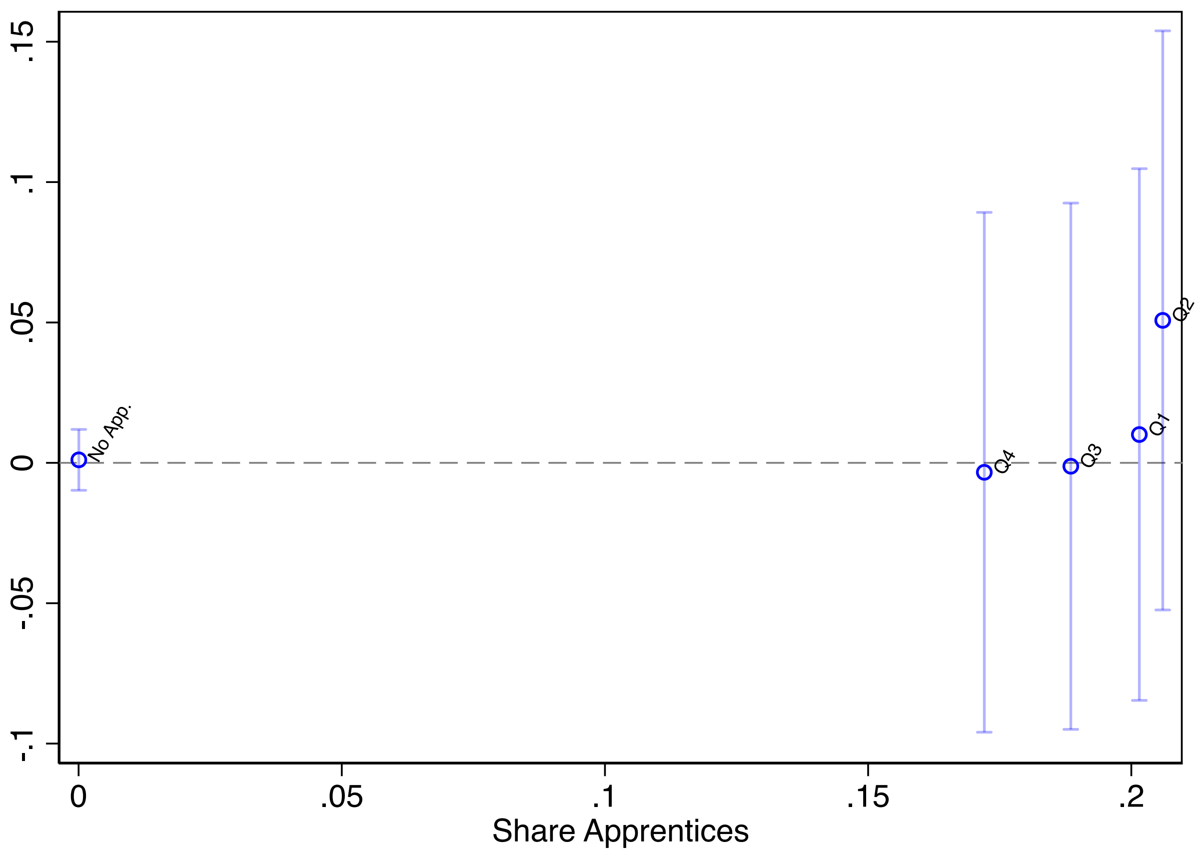
Notes: Social Security Administration data (January 2003–December 2009). We estimate separately by region the effects of being below the subsidy threshold (b_t) from a model that pools all "post" periods in Equation 1. This figure plots the pooled estimates and their corresponding 95% confidence intervals (vertical axis) against the share of apprentices employed in each region (horizontal axis). The outcome variable is the number of apprentices. See Figure 5 notes for details.

Figure A.6: Heterogeneity by Baseline Apprentice Earnings: Reduced Form Estimates of Threshold on Number of Apprentices



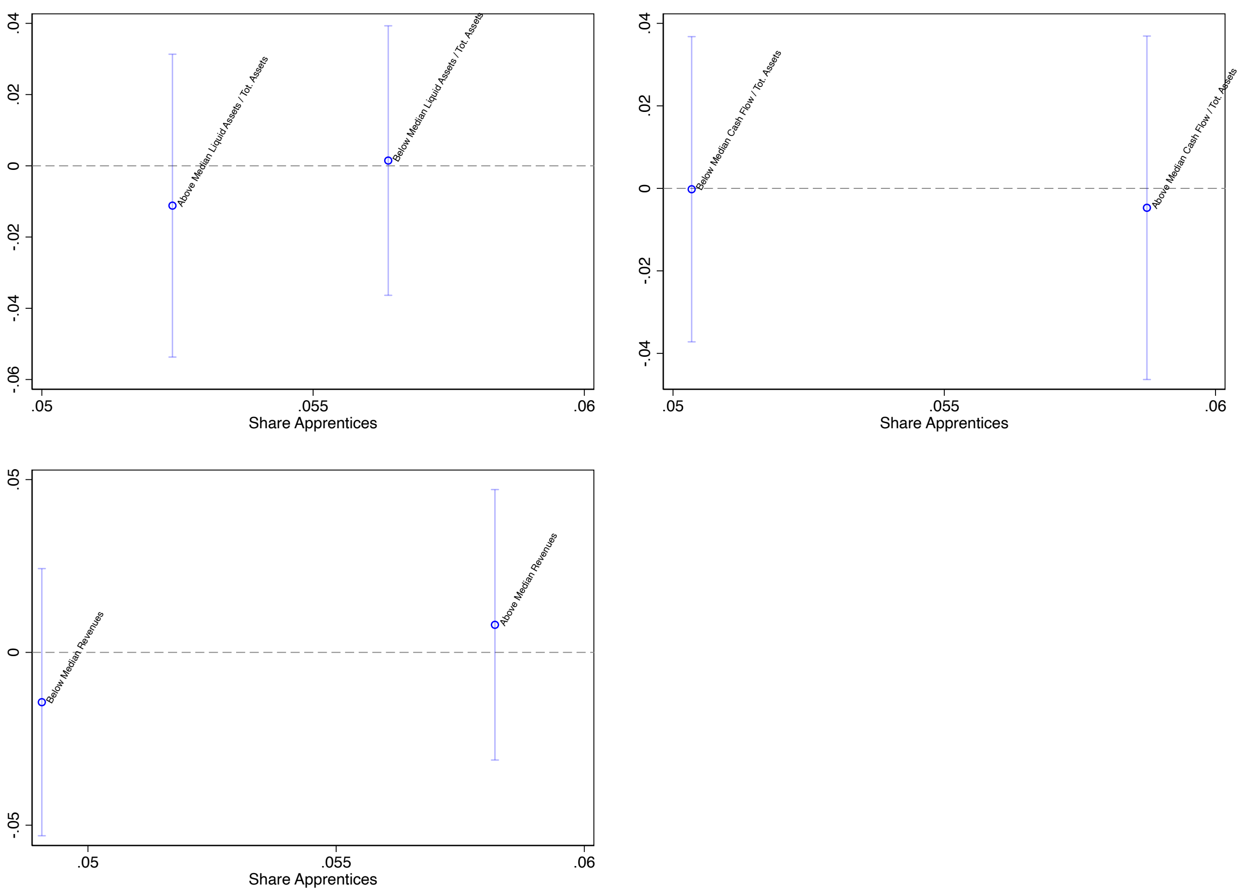
Notes: Social Security Administration data (January 2003–December 2009). We estimate separately by quantile of 2006 apprentice earnings the effects of being below the subsidy threshold (b_t) from a model that pools all "post" periods in Equation 1. We grouped all firms that did not employ any apprentice in 2006 in a category called "No App.". "Q1" represents the first quartile of 2006 apprentice earnings distribution. "Q2" through "Q4" are defined analogously. This figure plots the pooled estimates and their corresponding 95% confidence intervals (vertical axis) against the share of apprentices employed in each group of firms (horizontal axis). The outcome variable is the number of apprentices. See Figure 5 notes for details.

Figure A.7: Heterogeneity by Contemporaneous Apprentice Earnings: Reduced Form Estimates of Threshold on Number of Apprentices



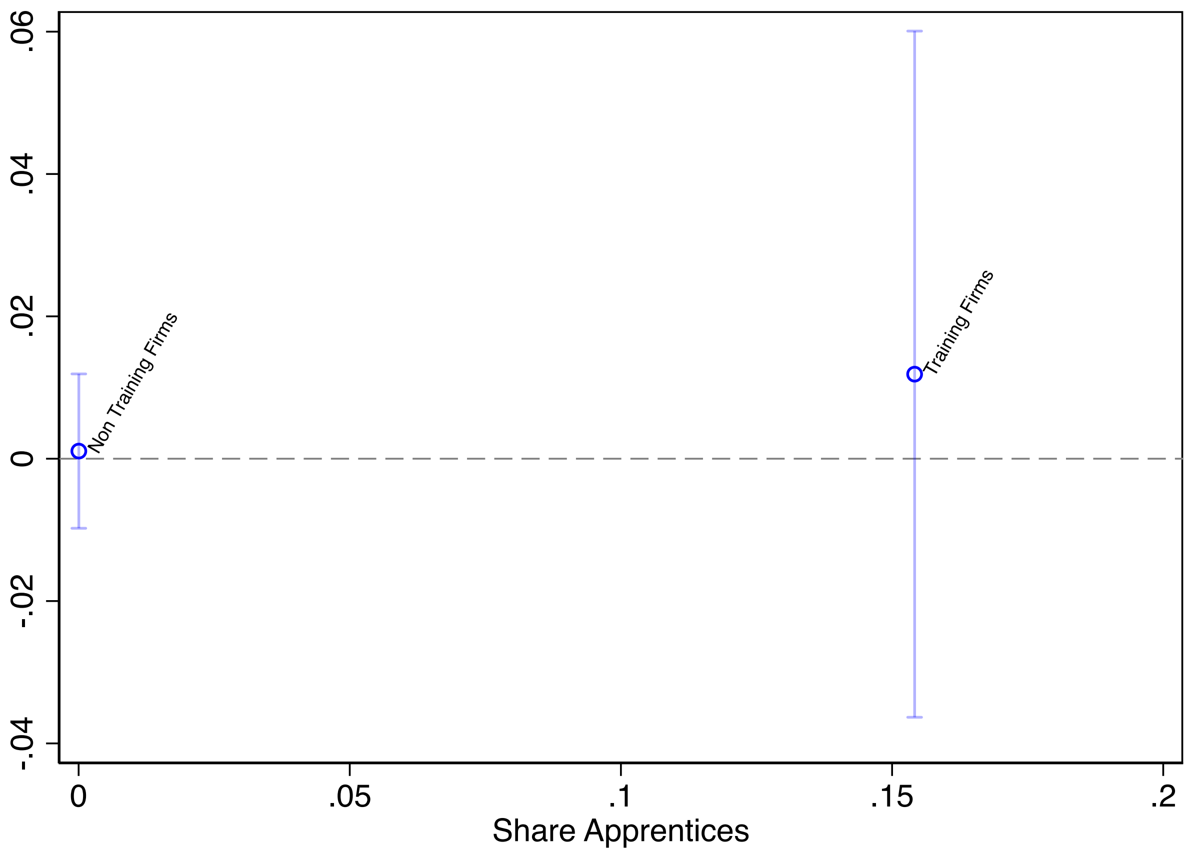
Notes: Social Security Administration data (January 2003–December 2009). We estimate separately by quantile of contemporaneous apprentice earnings the effects of being below the subsidy threshold (b_t) from a model that pools all "post" periods in Equation 1. We grouped all firms that do not employ any apprentice in a category called "No App.". "Q1" represents the first quartile of 2006 apprentice earnings distribution. "Q2" through "Q4" are defined analogously. This figure plots the pooled estimates and their corresponding 95% confidence intervals (vertical axis) against the share of apprentices employed in each group of firms (horizontal axis). The outcome variable is the number of apprentices. See Figure 5 notes for details.

Figure A.8: Heterogeneity by Liquidity Constraints: Reduced Form Estimates of Threshold on Number of Apprentices



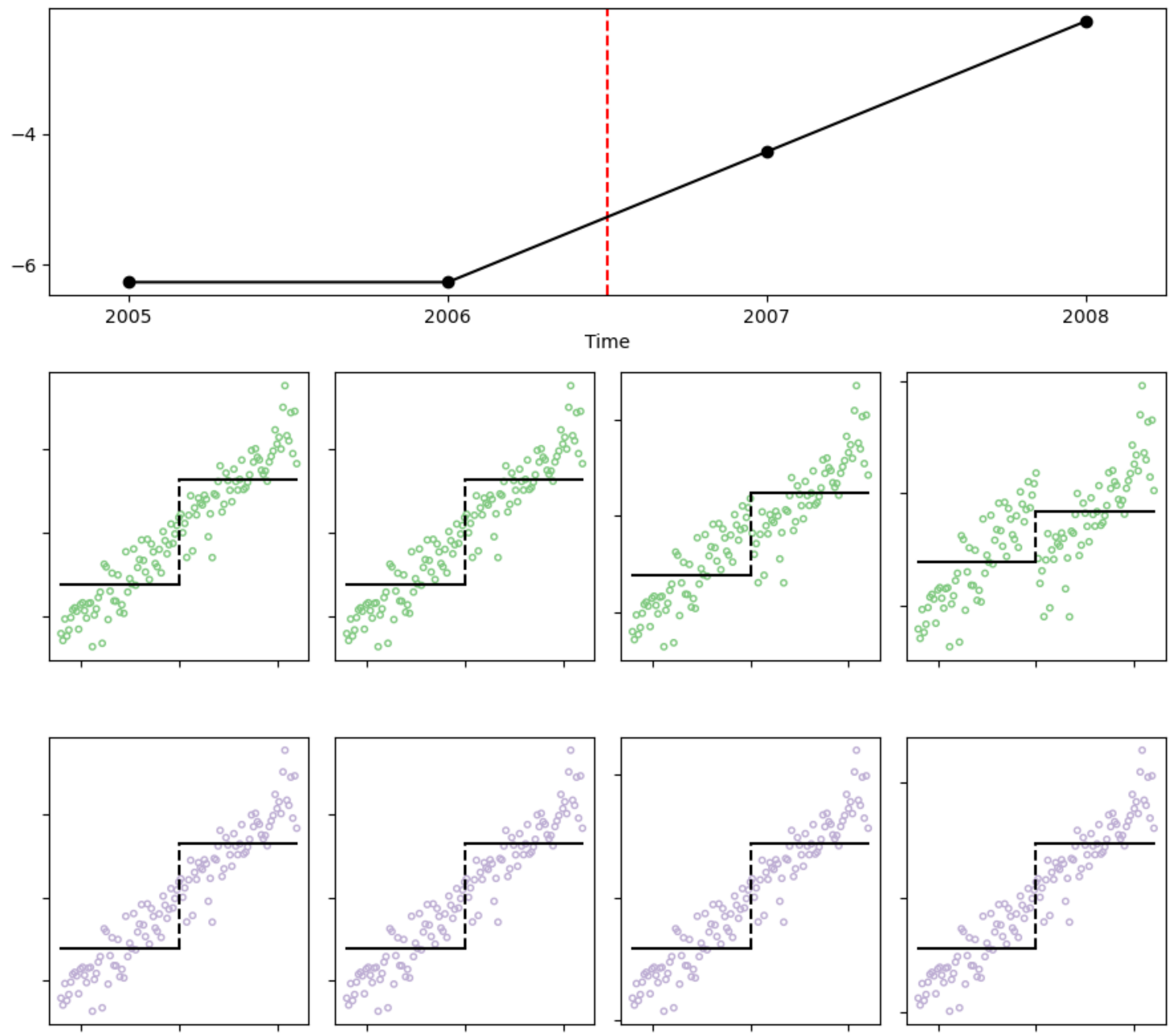
Notes: Social Security Administration data (January 2003–December 2009). We estimate separately by liquidity constraint status the effects of being below the subsidy threshold (b_t) from a model that pools all "post" periods in Equation 1. We follow Saez et al. (2019) and use three measures of liquidity constraints: i) liquid assets over total assets, ii) cash flow over total assets, and iii) revenues. For each measure of liquidity, we divide firms into two groups based on whether they fall above vs. below the median of each proxy for liquidity constraints. Each panel plots the pooled estimates and their corresponding 95% confidence intervals (vertical axis) against the share of apprentices employed in each group of firms (horizontal axis). The outcome variable is the number of apprentices. See Figure 5 notes for details.

Figure A.9: Heterogeneity by Training Status: Reduced Form Estimates of Threshold on Number of Apprentices



Notes: Social Security Administration data (January 2003–December 2009). We estimate separately by training status the effects of being below the subsidy threshold (b_t) from a model that pools all "post" periods in Equation 1. We define as "training firms" those that employed at least one apprentice in 2006 and "non-training firms" those who did not. This figure plots the pooled estimates and their corresponding 95% confidence intervals (vertical axis) against the share of apprentices employed in each group of firms (horizontal axis). The outcome variable is the number of apprentices. See Figure 5 notes for details.

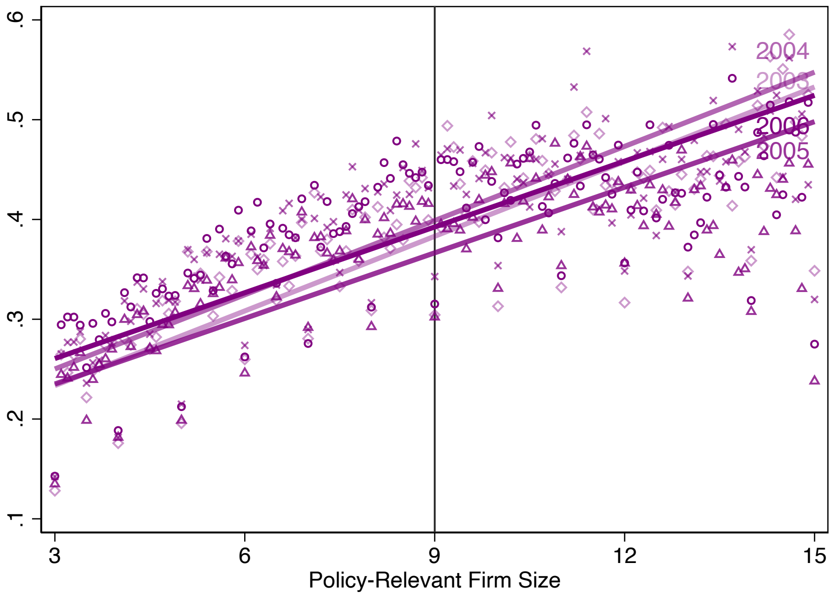
Figure A.10: Two examples of joint distributions that generate observationally equivalent difference-in-differences estimates.



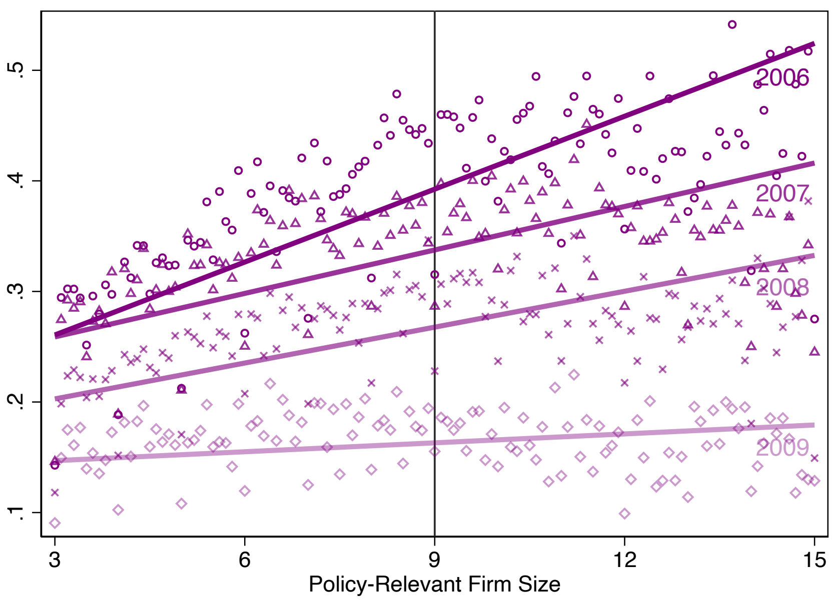
Notes: This figure uses simulated data to show that the same DD estimates can come from two different relationships of the outcome and the targeted characteristic. The top row plots a time series of first difference estimates. A standard DD specification would subtract the difference at a baseline period (e.g. 2006). The second row of figures plots the underlying relationship between the outcome Y and the targeted characteristic Z in green. A discontinuity emerges in 2007 and grows in 2008. The third row plots an alternative relationship between Y and Z in purple that generates the same estimates. The conditional expectation function is stable in the pre-period and only rotates in the post-period. There is little evidence that the outcome changes discontinuously at the targeted threshold.

Figure A.11: The Rotation of the Conditional Expectation Function

(a)

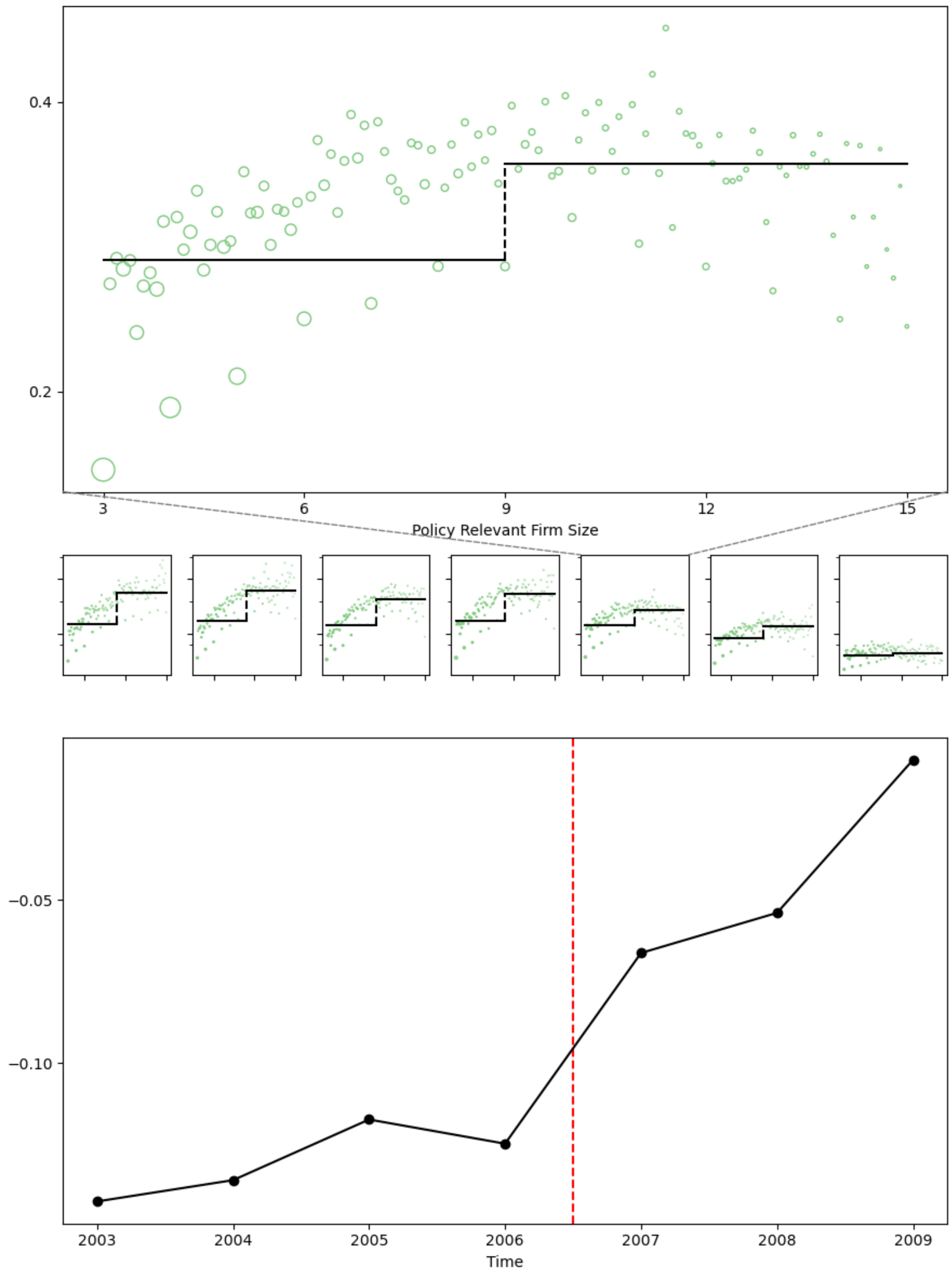


(b)



Notes: Social Security Administration data (January 2003–December 2009). This Figure shows a binned scatterplot of apprentice hiring against 2006 policy-relevant firm size. Panels (a) and (b) illustrate the relationship between 2003 and 2006, and between 2006 and 2009, respectively. 2006 appears in both graphs to enhance comparability.

Figure A.12: Spurious Effects on Apprentice Hiring under Standard Diff-in-Diff Specification



Notes: Social Security Administration data (January 2003–December 2009). This figure decomposes the comparisons made by the standard difference-in-differences specification. Treated firms are those whose average policy-relevant firm size over 2006 is less than 9. The top panel shows a binned scatterplot of annual apprentice hiring against *average baseline* policy-relevant firm size in 2007, the first year of the policy. The size of the green dots indicates the number of firms within the bin. Means conditional on being in treatment on control—a piecewise zeroth order polynomial fit—are overlayed as black lines. The first panel is a zoomed example of the fitted means in each period, shown in the second panel. The third panel plots a time series of the mean difference between treated and control firms.

Table A.1: Characteristics of Apprentices in January 2006

	(1)
Male	0.657
	[0.475]
Native	0.881
	[0.324]
Age	22.458
	[2.819]
Previously employed	0.985
	[0.123]
Experience	3.759
	[2.572]
Monthly (net) earnings	1050.300
	[334.690]
N	169,581

Notes: Social Security Administration data (January 2006). This table reports the summary statistics for the apprentices in our sample at baseline (January 2006). The standard deviation is reported in brackets. All statistics are calculated across apprentice observations.

Table A.2: Industry Composition of Firms in January 2006

	(1) All firms	(2) Firms with apprentices	(3) Firms that ever take-up
Ag., silviculture, fishing, and extraction	0.015	0.004	0.003
	[0.120]	[0.060]	[0.056]
Manufacturing	0.297	0.358	0.327
	[0.457]	[0.479]	[0.469]
Utilities	0.005	0.003	0.003
	[0.072]	[0.053]	[0.051]
Transportation, warehouse, and construction	0.226	0.225	0.219
	[0.418]	[0.418]	[0.414]
Trading	0.205	0.205	0.208
	[0.404]	[0.404]	[0.406]
Services	0.187	0.163	0.195
	[0.390]	[0.369]	[0.396]
Public admin, education, and health	0.032	0.010	0.011
	[0.175]	[0.097]	[0.106]
Other	0.032	0.032	0.033
N	398,412	99,311	59,670

Notes: Social Security Administration data (January 2006). This table reports the summary statistics for the firms in our sample at baseline (January 2006). The standard deviation is reported in brackets. All statistics are calculated across firm observations.

Table A.3: Reduced Form Estimates of Threshold on Other Outcomes

	(1) t_0-48	(2) t_0-24	(3) t_0	(4) t_0+12	(5) t_0+35
Panel A: Flows					
New apprentice contracts	-0.007	-0.002	-0.008	-0.011	0.002
	(0.008)	(0.008)	(0.008)	(0.009)	(0.006)
New apprentice hires	-0.006	-0.002	-0.005	-0.011	0.003
	(0.008)	(0.008)	(0.008)	(0.009)	(0.006)
New temporary hires	0.002	0.008	-0.003	-0.007	0.022
	(0.015)	(0.016)	(0.016)	(0.016)	(0.014)
All hires	0.015	-0.025	0.032	-0.025	0.016
	(0.046)	(0.047)	(0.045)	(0.047)	(0.035)
New hires (under age 30)	-0.011	-0.003	-0.004	-0.003	0.004
	(0.019)	(0.018)	(0.018)	(0.018)	(0.014)
Apprentice separations	-0.001	-0.004	0.001	-0.002	-0.012
	(0.003)	(0.003)	(0.003)	(0.003)	(0.006)
Temporary separations	-0.002	-0.002	-0.002	-0.011	-0.018
	(0.006)	(0.007)	(0.007)	(0.008)	(0.016)
All separations	0.008	0.001	0.037	0.016	-0.050
	(0.019)	(0.019)	(0.019)	(0.020)	(0.039)
Separations (under age 30)	-0.001	-0.004	0.004	-0.002	-0.034
	(0.009)	(0.008)	(0.008)	(0.008)	(0.016)
Apprentice transformations	0.002	0.001	0.003	-0.000	0.003
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
Panel B: Apprentice Characteristics					
Apprentice avg. age	0.524	0.279	-0.080	0.448	-0.395
	(0.331)	(0.338)	(0.350)	(0.353)	(0.546)
Apprentice avg. experience	0.365	0.656	0.301	0.433	0.253
	(0.311)	(0.316)	(0.333)	(0.331)	(0.473)
Apprentice male share	0.002	-0.016	0.021	-0.051	-0.048
	(0.058)	(0.057)	(0.055)	(0.056)	(0.083)
Apprentice native share	0.064	0.011	0.005	0.004	-0.004
	(0.038)	(0.039)	(0.039)	(0.041)	(0.061)
Apprentice prev. employed share	0.011	0.015	0.047	0.021	-0.034
	(0.052)	(0.049)	(0.049)	(0.047)	(0.076)
Wage bill (new hires)	11.663	-30.117	-6.689	21.862	119.946
	(46.411)	(46.150)	(48.261)	(49.788)	(84.921)
Contract length	-0.943	-2.779	-1.490	-0.941	0.977
	(1.739)	(1.753)	(1.794)	(1.733)	(1.894)
Panel C: Stocks					
Number of Temporary Workers	0.034	0.011	0.028	-0.027	-0.007
	(0.032)	(0.032)	(0.034)	(0.036)	(0.042)

Notes: Social Security Administration data (January 2003–December 2009). N=24,532,943. This table reports the effects of being below the threshold (b_t) from the main DD specification in Equation 1, where the outcome variables are general firm characteristics. Each row reports the estimates for a different outcome variable. Estimates are relative to $t_0 - 12$ (January 2006). The first two columns report the pre-reform DD estimates for $t_0 - 48$ (January 2003) and $t_0 - 24$ (January 2004). Columns 3-5 report the post-reform estimates for t_0 (January 2007), $t_0 + 12$ (January 2008), and $t_0 + 35$ (December 2009), respectively. Robust standard errors clustered by firms are reported in parenthesis.

Table A.4: First Stage

	(1) Social Security Contributions
Below × Post	-16.411
	(1.082)
N firms	857,587
N obs	24,532,943
F-stat	230

Notes: Social Security Administration data (January 2003–December 2009). This table reports the first stage estimates from the main IV specification in Equation 2. Robust standard errors clustering by firms reported in parenthesis.

B The Policy-Relevant Firm Size

The 2007 Budget Bill does not define how to compute the policy-relevant firm size and delegates this task to the Italian Social Security Agency (INPS). INPS details how to compute the policy-relevant firm size in a provision issued in January 2007 (*circolare n. 22, 2007*). We follow this definition closely.

The firm size that determines the eligibility for the SSC discount is full-time equivalent employment excluding apprentices, temporary agency workers, workers who are on leave (unless the firm hires a substitute), and workers who have been hired with an on-the-job training contract. The types of job training contracts that are excluded from the computation of firm size are those created under the following provisions: exD.lgs.251/2004, D.lgs.n.276/2003, law n.223/1991.

Our rich administrative data contains detailed information on workers' contracts and allows us to construct an accurate measure for the policy-relevant firm size. In this context, there are two sources of potential measurement error. First, INPS data does not contain a flag for the on-the-job training contracts created under the exD.lgs.251/2004. Anecdotally, this contractual arrangement is very rare and it is unlikely to generate substantial measurement error. Second, our proxy does not account for workers who are on temporary leave (e.g., sick leave or maternity leave).

C Pitfalls of Standard Difference-in-Differences when Program Eligibility is Defined Using a Continuous Variable

This section formalizes the argument that discretizing a continuous treatment in a standard difference-in-differences (DD) approach can inadvertently use variation unrelated to policy changes, leading to erroneous conclusions about the effect of the policy. First, we show that rotations of the conditional expectation function are a form of omitted variable bias in standard DD models. Second, we illustrate that a difference-in-discontinuities approach is robust to rotations of the conditional expectation function over time because it controls flexibly for the running variable in each period. Finally, we illustrate our findings using a concrete example.

C.1 RD *or* Diff-in-diff

We begin by stating the standard fuzzy RD assumptions.

Assumption 1 (Potential Outcomes and Exclusion). *In each period t, each firm draws a pair of potential outcomes, potential choices under treatment, and the running variable $(Y_{it}(0), Y_{it}(1), D_{it}(0), D_{it}(1), Z_{it})$, and the observed outcome is $Y_{it}(D_{it}) = Y_{it}(0)\cdot(1-D_{it}) + Y_{it}(1)\cdot D_{it}$.*

Assumption 2 (Regression Discontinuity). *Assume:*

1. *Continuity in potential outcomes: $\mathbf{E}[Y_{it}(D_{it})|Z_{it}=z]$ is continuous in z for each D_{it}*

2. *Continuity in take-up rate: $\mathbf{E}[D_{it}(T_{it})|Z_{it}=z]$ is continuous in z for each T_{it}*

Local linear regression estimators of regression discontinuity also typically requires that the density of the running variable is continuous. In our setting, firm size bunches at round numbers (Figure 2), rendering infeasible standard RD estimators that compare observed outcomes in a small neighborhood around the discontinuity.

An alternative especially common in the literature on wage subsidies is to apply a difference-in-differences approach, comparing mean differences between large and small firms and subtracting selection bias by measuring pre-existing differences prior to the intervention (see e.g., Cahuc et al., 2019). This approach unwittingly imposes additional assumptions on firms' potential outcomes away from the threshold. To see this formally, consider the standard parallel trends assumption:

Assumption (Strong Parallel Trends). *Assume that potential outcomes can represented by*

$$Y_{it}(0) = a_i + c_t + u_{it}$$
$$Y_{it}(1) = a_i + c_t + b_{it} + u_{it}$$

with u_{it} independent.

The difference-in-differences regression specification masks heterogeneity away from the threshold because it recodes a continuous variable, effectively approximating the conditional expectation function with horizontal lines (Figure A.10, Panel B). Difference-in-differences specifications are often operationalized by estimators derived from saturating indicator variables for time and their interactions with treatment,

$$Y_{it} = a_1 + a_2 T_{it} + \sum_{\tau \neq -12} a_3^{\tau}\Delta_t^{\tau} + b_4^{\tau}(T_{it} \times \Delta_t^{\tau}) + u_{it}, \tag{3}$$

where Δ_t^τ are dummies for each time period. Parallel pre-trends that check that $b_\tau = 0 \quad \forall \tau < 0$ are testing that u_{it} is mean independent of T_i, $\mathbf{E}\left[u_{it}|T_i\right] = 0$. [7] However, the strong parallel trends assumption also requires that u_{it} is fully independent of Z_{it}. Testing the significance of b_τ does not exhaust the available validity tests of the assumption.

Concretely, let $\mathbf{E}[u_{it}|Z_{it} = z] = g_t(z)$. By assuming that $u_{it} \perp Z_{it}$, a strict parallel trends assumption not only implies parallel trends in intercepts ($\mathbf{E}[g_t(Z)|T] = 0$) but also parallel trends in the slopes of the conditional expectation function of Y given Z ($\mathbf{E}[g_t'(Z)|T] = 0$).[8] If the conditional expectation function rotates over time, then Z_{it} is correlated with u_{it} and is an omitted variable. This can lead one to find no effect with regression discontinuity but find a spurious effect with difference-in-differences.

Appendix Figure A.10 simulates two scenarios that produce identical DD estimates. The DD specification cannot distinguish between a treatment effect generated by the discontinuity (green scatter plots) and rotations of the conditional expectation function (purple scatterplots), i.e., the conditional expectation function becomeing more/less flat over time. Failing to isolate variation close to the discontinuity means that RD estimates and DD estimates can diverge, even assuming constant treatment effects.

Notably, many empirical analyses often measure Z_{it} in some base year because it is not subject to manipulation and therefore less "endogenous." However, the conditional expectation function will often regress to the mean, generating a rotation.[9]

C.2 RD *and* Diff-in-diff: Difference in Discontinuities

Even without treatment effect heterogeneity ($b_{it} = b$ in the strong parallel trends assumption), the previous discussion shows how RD and difference-in-differences can yield different estimates. Differences-in-discontinuities rectifies this problem. If changing slopes are an omitted variable, a simple fix is to allow flexibility in the slope of the conditional expectation, isolating variation adjacent to the discontinuity to infer the causal effects of the policy. (One way to view differences-in-discontinuities is as an alternative to local linear regression methods to debiasing RD estimates.)

Formally, we make a weaker parallel trends assumption:

[7] When T_{it} is time-invariant, one can include unit fixed effects to obtain equivalent estimates with greater statistical power.

[8] For the identifying assumption to hold, $Cov[g(Z_{it}) \times \Delta_{it}^\tau, \varepsilon_{it}]$ for any function $g(\cdot)$.

[9] In a simple error-in-variables (white noise) model, $|Cov[Y_{it}, Z_{it}]| < |Cov[Y_{it}, Z_{i0}]|$ for $t \neq 0$.

Assumption 3 (Weak Parallel Trends)**.** *Assume that potential outcomes can represented by*

$$Y_{it}(0) = a_i + c_t + u_{it}$$
$$Y_{it}(1) = a_i + c_t + b_{it} + u_{it}$$

with $\mathbf{E}[u_{it}|Z_{it}] - \mathbf{E}^*[u_{it}|Z_{it}] = d\ \forall t$*, where* $\mathbf{E}^*[\cdot]$ *is a linear projection and* d *is a constant.*

Under this assumption, the curvature in the conditional expectation function of untreated potential outcomes is time-invariant.[10] Whereas the literature on RD has focused on minimizing d by estimating local quadratic regressions and restricting estimation to a narrow bandwidth, we subtract the bias generated by non-linearities using the pre-period.[11]

Combining difference-in-differences with regression discontinuity to exploit variation around the threshold yields model (1) in Section 4.1. In order to cluster standard errors at the firm level, we operationalize the difference-in-discontinuities approach with a saturated, stacked regression model,

$$\begin{aligned}
Y_{it} = & \underbrace{a_{1,\text{Jan 2006}}}_{\text{Baseline intercept}} && + \underbrace{\sum_{t \neq \text{Jan 2006}} a_{1t}\Delta_t}_{\text{Time-varying intercepts}} \\
& + \underbrace{a_{2,\text{Jan 2006}} T_{it}}_{\text{Baseline discontinuity}} && + \underbrace{\sum_{t \neq \text{Jan 2006}} \boldsymbol{b}_t \left(\boldsymbol{T}_{it} \times \boldsymbol{\Delta}_t\right)}_{\textbf{Difference-in-discontinuities}} \\
& + \underbrace{g_{1,\text{Jan 2006}} Z_{it}}_{\text{Baseline slope above discontinuity}} && + \underbrace{\sum_{t \neq \text{Jan 2006}} g_{1t}\left(Z_{it} \times \Delta_t\right)}_{\text{Time-varying slope above discontinuity}} \\
& + \underbrace{g_{2,\text{Jan 2006}} Z_{it} \times T_{it}}_{\text{Baseline slope below discontinuity}} && + \underbrace{\sum_{t \neq \text{Jan 2006}} g_{2t}\left(Z_{it} \times T_{it} \times \Delta_t\right)}_{\text{Time-varying slope below discontinuity}} + u_{it},
\end{aligned} \tag{4}$$

where Δ_t are time dummies. The point estimates are identical to estimating separate regression models in each period and subtracting the baseline discontinuity from the measured discontinuity.

Through the lens of this model, the main and interacted terms of Z_{it} can be viewed

[10] Unlike other applications of diff-in-discontinuity designs (see e.g. Grembi et al., 2016), we are not trying to subtract the effect of other policies that share the same discontinuity.

[11] A technical literature has emerged to select a bandwidth that balances bias and precision while debiasing the estimates using controls for higher-order polynomials (Calonico et al., 2014). Calonico et al. (2014) Remark 7 notes that conventional point estimates from a quadratic regression specification coincide with their procedure that allows the point estimate and bias correction specifications to be fit on samples with differing bandwidths.

as omitted variables. The standard DD short regression specification constrains $g_{1t} = 0$ and $g^{\tau}_{2t} = 0$. Applying standard difference-in-differences specifications by discretizing a continuous treatment inadvertently generates an omitted variable bias by failing to control for changing slopes. Changing slopes manifests visually as a rotation, which could lead to spurious inferences.

Our approach combines the strengths of difference-in-differences and regression discontinuity. First, the approach yields pre-trend validity tests that mirror validity tests of difference-in-differences designs. Specifically, we perform a series of placebo tests by examining the difference-in-discontinuities coefficients b_t for $t <$ Jan 2007. Second, our approach uses the longitudinal dimension of the panel data to bias-correct our estimates. Whereas the literature on RD has focused on removing bias by deleting data (i.e., estimating local quadratic regressions and restricting estimation to a narrow bandwidth, see Calonico et al., 2014), we subtract the bias generated by non-linearities and extrapolation using data from the pre-period.[12]

C.3 A Cautionary Tale

As noted previously, the strategy of defining treatment at baseline to avoid simultaneity bias arising from the "endogenous" choice of firm size can itself induce a rotation from the regression coefficient exhibiting mean reversion. Whereas our difference-in-discontinuities specification is robust to rotations because it isolates variation near the discontinuity, the difference-in-differences estimates reflect the variation derived from rotations of the conditional expectation function.

In Figure A.11, we document that defining Z_{it} in the year prior to the policy, the conditional expectation function is very stable between 2003 and 2006 (Panel a) and rotates between 2007 and 2009 (Panel b). In Figure A.12, we decompose the comparisons made by the standard difference-in-differences specification and show that a naive analysis of the subsidy policy generates spurious estimates driven by such a rotation.

Can a rotating conditional expectation function be causal? A discontinuity at the threshold is generally considered to be "good variation" and strong evidence of policy

[12] A technical literature has emerged to select a bandwidth that balances bias and precision while debiasing the estimates using controls for higher-order polynomials (Calonico et al., 2014). Calonico et al. (2014) Remark 7 notes that conventional point estimates from a quadratic regression specification coincide with their procedure that allows the point estimate and bias correction specifications to be fit on samples with differing bandwidths. As we mentioned above, the procedure proposed by Calonico et al. (2014) is not feasible in our setting. Bunching in the running variable at round numbers and the fact that most firms have only a few employees leaves a very narrow bandwidth to estimate the local linear regression.

effects. Nevertheless, it is worth asking whether variation away from the threshold is actually "bad variation." Specifically, if our design focuses on DD estimates just above versus just below the policy threshold, could a design that measures time variation in the slope of the conditional expectation function be consistent with causal effects?

We argue no. Estimating Equation 3 on a rotation would spuriously detect treatment effects in regions without policy variation. Consider the bottom panel of Appendix Figure A.10 and conditioning the analysis sample on firms entirely above or entirely below the policy discontinuity. In such a sample, there is no cross-sectional policy variation. However, the differences between large and small firms *within the subsample* are changing over time.

Robustness to over-identifying placebo tests (i.e., estimating the placebo effects by moving the policy threshold to the left or to the right of the actual policy threshold) may ameliorate concerns, especially in the case of Appendix Figure A.10 when the conditional expectation function is linear. But, if the conditional expectation function exhibits concavity or convexity, a relatively flat portion of the conditional expectation function may rotate less, and the placebo test would fail to find spurious effects.