



**Universitat
Pompeu Fabra**
Barcelona

Department
of Economics and Business

Economics Working Paper Series

Working Paper No. 1743

**Unwilling to train? Firm responses to the
Colombian apprenticeship regulation**

**Santiago Caicedo, Miguel Espinosa,
and Arthur Seibold**

August 2020

UNWILLING TO TRAIN? FIRM RESPONSES TO THE COLOMBIAN APPRENTICESHIP REGULATION*

Santiago Caicedo Miguel Espinosa Arthur Seibold[†]

August 2020

Abstract

We study firm responses to a large-scale change in apprenticeship regulation in Colombia. The reform requires firms to train, setting apprentice quotas that vary discontinuously in firm size. We document strong heterogeneity in responses across sectors, where firms in sectors with high skill requirements tend to avoid training apprentices, while firms in low-skill sectors seek apprentices. Guided by these reduced-form findings, we structurally estimate firms' training costs. Especially in high-skill sectors, many firms face large training costs, limiting their willingness to train apprentices. Yet, we find substantial overall benefits of expanding apprenticeship training, in particular when the supply of trained workers increases in general equilibrium. Finally, we show that counterfactual policies that take into account heterogeneity across sectors can deliver similar benefits from training while inducing less distortions in the firm size distribution and in the allocation of resources across sectors.

JEL Codes: E24, J21, J24, M5

*We thank Ufuk Akcigit, Benito Arruñada, Leonardo Bursztyn, Jonathan Dingel, Christian Dustmann, Jan Eeckhout, Luis Garicano, Alessandro Gavazza, Maria Guadalupe, Ali Hortacsu, Greg Kaplan, Camille Landais, Bob Lucas, Rocco Macchiavello, Sandra McNally, Magne Mogstad, Joan Monras, Massimo Motta, Carlos Ospino, Luis Rayo, Mar Reguant, Esteban Rossi-Hansberg, Uta Schönberg, Sandro Shelegia, Alexandra Spitz-Oener, Claudia Steinwender, Nancy Stokey, John Sutton, Chad Syverson, John van Reenen and numerous conference and seminar participants for helpful comments and discussions. Gabriela Díaz, Verónica Pérez and Lukas Rodrian provided excellent research assistance for this project.

[†]Caicedo: University of Chicago and Universidad de Los Andes, scaiceso@gmail.com, scaicedosoler@uchicago.edu; Espinosa: Universitat Pompeu Fabra, miguel.espinosa@upf.edu; Seibold: University of Mannheim, seibold@uni-mannheim.de

1 Introduction

Labor market opportunities for young people are a major concern across the developing world. Young workers make up a growing share of the labor force, while formal employment opportunities are limited, especially for those without college education. In tackling those issues, apprenticeships combining formal vocational education and training within firms are gaining in popularity.¹ Apprenticeships are often viewed as effective in upgrading the skills and raising the productivity of young workers, and help them acquire early labor market experience. Hence, expanding apprenticeship programs is a common policy recommendation in order to tackle youth unemployment (e.g. [Fazio et al., 2016](#); [ILO, 2017](#); [Kuczera, 2017](#)).

Firms' willingness to take on young individuals and train them has been identified as a key limiting factor in successfully rolling out apprenticeship programs (e.g. [Groh et al., 2016](#); [Alfonsi et al., 2020](#)). They play a crucial role in facilitating knowledge transfer between workers and apprentices, but may lack incentives to train apprentices. If the productivity of untrained individuals is initially too low, their contribution to firm production might not be enough to offset the cost of training them. While there is growing evidence of the effects of training on apprentices or workers both in developed and developing countries (see e.g. the reviews by [McKenzie, 2017](#) and [Card et al., 2018](#)), surprisingly little is known about firm responses to apprenticeship programs and how firms can be effectively incentivized to train apprentices.

In this paper, we aim at filling this gap and make three main contributions. First, we provide reduced-form evidence of firm responses to a large-scale change in apprenticeship regulation in Colombia that requires firms to train apprentices. The reform is successful in expanding overall apprenticeship training, but firm responses are strongly heterogeneous across sectors. In particular, firms in sectors with high skill requirements tend to avoid training apprentices. Second, we uncover firms' underlying training costs based on structural estimation. Many firms face large training costs and these vary substantially across sectors. This can explain both the limited number of apprentices before the reform and heterogeneous firm behavior after the reform. Third, we use the structural model to quantify the effects of the regulation in partial and general equilibrium and to simulate counterfactual policies. The results suggest positive overall welfare effects of the apprenticeship program that could be further enhanced by considering heterogeneity across sectors.

The Colombian setting provides several advantages for our analysis. First, with youth unemployment rates of around 30% and youth informality rates above 70% in the early 2000s, young people in Colombia experienced labor market issues in common with many other low- and middle-income countries. Second, a reform in 2003 introducing a new apprenticeship regulation provides unique, large-scale variation in firms' incentives to train. The reform introduced apprentice quotas featuring multiple discontinuities in firm size, requiring firms to train apprentices between a minimum and maximum quota depending on their size. The reform also changed apprentices' minimum wages and

¹Throughout this paper, we follow [Wolter and Ryan \(2011\)](#) and consider apprenticeships as “programs that comprise both work-based training and formal education, in most countries at upper-secondary level, and lead to a qualification in an intermediate skill, not just to semi-skilled labor” (p.523).

introduced the possibility for firms to pay a costly fee as a “buy-out” from training apprentices. A third advantage is that high-quality administrative records at the firm level are available for the analysis, and they can be linked to a firm census containing additional rich survey information on firm characteristics.

We begin the analysis by showing reduced-form evidence of the effects of the apprenticeship regulation. On aggregate, the reform is highly successful in increasing the number of trained apprentices to more than fifteen times the pre-reform level, but it also induces sizeable labor input responses and firm size distortions. At the firm level, we organize the results into three empirical facts documenting the strong heterogeneity of responses across sectors. First, we use bunching methods to gauge firm size responses to the discontinuities in apprenticeship quotas. We find that firms in sectors with a large fraction of highly skilled workers (henceforth high-skill sectors) reduce their size to locate just *below* the regulation thresholds in order to train fewer apprentices, leaving “missing mass” just above the thresholds. Meanwhile, firms in sectors with a low fraction of highly skilled workers (henceforth low-skill sectors) bunch *at* the regulation thresholds in order to increase the number of apprentices they can hire. Firm size distortions are large: The marginal bunching firm in high-skill sectors reduces their size by two full-time workers to avoid the higher apprentice quota above the threshold, and the marginal bunching firm in low-skill sectors increases its size by around 1.5 workers to be able to train more apprentices.

Second, we show that conditional on their post-reform size, firms in high-skill sectors tend to train the minimum number of apprentices required, while most firms in low-skill sectors train the maximum number of apprentices possible. In fact, training as many apprentices as possible is the most common response in low-skill sectors, where 65% of firms choose the maximum number. Third, many firms in high-skill sectors pay fees to the government as a buy-out from the apprentice quotas. 58% of high-skill sector firms pay fees such that they are allowed not to train any apprentices. In low-skill sectors, on the other hand, this behavior is virtually non-existent. Taken together, the reduced-form results imply that high-skill sector firms tend to avoid training apprentices, while low-skill sector firms seek apprentices. Hence, while the overall number of apprentices increases sharply, most training occurs in low-skill sector firms.

Guided by these empirical results, we develop a parsimonious model of firm production featuring heterogeneous costs of training apprentices, allowing us to recover the unobserved training cost distribution and to further quantify the effects of the regulation. We consider an economy with multiple sectors and heterogeneous firms characterized by their training costs and managerial ability. Firms produce using labor from workers and apprentices. The key difference between these two types of labor is that apprentices require training in order to become productive, which is costly as it requires workers’ time. In addition, apprentices’ productivity per unit of time can differ from that of workers. This theoretical framework is able to capture the differential firm responses to the reform observed in the data via heterogeneity in training costs. High-skill sector firms with high training costs avoid apprentices, while low-skill sector firms with low training costs seek them. Moreover, the model highlights the initially high minimum wage for apprentices as an important source of inefficiency contributing to the low levels of training before the reform.

The structural parameters of the model can be transparently identified, using the observed firm responses to the reform in combination with pre-reform data. First, we estimate the output elasticity of labor and the parameters of the productivity distribution by sector using pre-reform data. We then estimate a non-parametric training cost distribution for each sector via indirect inference, targeting moments that correspond to the key firm responses from the reduced-form analysis: excess mass and missing mass in the post-reform firm size distribution, the fraction of firms choosing the maximum apprentice quota, and the fraction paying the fee. These observed responses are informative of the costs and benefits of training apprentices to firms. They allow us to uncover the distribution of firms' *net* training costs, a combination of the cost of training in terms of workers' time and the lower productivity of apprentices.

The estimated net training costs are high for many firms. Scaled in terms of equivalent time spent by workers, average training costs are 0.75 in low-skill sectors and 1.05 in high-skill sectors. This implies that for the average low-skill sector firm, apprentices are only 25% as productive as workers, and for the average high-skill sector firm apprentices even have slightly negative productivity. These results are consistent with existing studies such as [Alfonsi et al. \(2020\)](#) who find that many firms are reluctant to train apprentices even when they are incentivized by wage subsidies. As expected from the reduced-form results, training costs are higher in high-skill sectors, where apprentices have negative marginal productivity for 60% of firms. This is directly related to the empirical fact that many high-skill sector firms prefer to pay fees over training apprentices, even though the fee nominally exceeds the apprentices' minimum wage. The strong heterogeneity in training costs has another important implication: at any given size of training incentives, only the firms with the lowest training costs take on apprentices. If training expenses by firms are positively correlated with the amount of skills acquired by apprentices, this may also help explain why the literature tends to find small benefits of some types of firm-based training.

Next, we use the estimated model to quantify the effects of the regulation on aggregate outcomes and welfare under three scenarios: i) in partial equilibrium, where wages are fixed, ii) in general equilibrium, where wages adjust and any excess labor supply is absorbed, and iii) in a dynamic scenario, where trained apprentices increase the supply of future workers. In partial equilibrium, the apprenticeship regulation has relatively small aggregate effects. Firms re-optimize and substitute a significant number of workers for apprentices, such that total labor input (apprentices plus workers) does not change much and total output is relatively stable. Around 1.2% of regular workers are displaced. There is some reallocation across sectors: Output increases by 0.4% in low-skill sectors but decreases by 0.4% in high-skill sectors. In addition, we provide further reduced-form evidence of the effect of the regulation on firm outcomes. Results from a difference-in-difference strategy around the regulation thresholds suggest that there are indeed sizeable labor substitution effects, as those firms who hire more apprentices lay off some workers, but the effects on output and profits are small and insignificant. Hence, the reduced-form results are consistent with the model-based partial equilibrium effects.

The general equilibrium and dynamic scenarios then yield further insights beyond the scope of the reduced-form analysis. In these scenarios, the impact on output can become substantial. In

general equilibrium, displaced workers are absorbed by the labor market, as wages adjust downwards. Production increases by 1.3% in low-skill sectors, but also in high-skill sectors the negative partial-equilibrium effects are reversed. Finally, these positive effects are magnified in the dynamic scenario, where aggregate output increases by 3.6% in total. These dynamic effects illustrate an important potential benefit of apprenticeship policies implemented on a large scale, where the overall size of the trained workforce increases, enabling growth in the affected sectors.

In addition, we unpack the aggregate effects and consider the effects on different groups of agents. The apprenticeship regulation creates winners and losers. Most directly related to the goal of improving young individuals' opportunities, apprentices benefit from the regulation in all sectors and under all scenarios, as their wages in the formal sector exceed their outside option. These positive effects may not be surprising, as the large number of individuals entering the program reveals that they derive benefits from apprenticeships. On the firm side, the effects vary by sector: most low-skill sector firms see an increase in profits as they are able to hire productive apprentices at relatively low wages. In high-skill sectors, many firms become less profitable in partial equilibrium. However, these negative effects vanish when considering general equilibrium and dynamic effects, where firms both in low-skill and high-skill sectors benefit from the increased supply of trained workers. The main losers from the expansion of apprenticeships are incumbent trained workers, especially in low-skill sectors. In partial equilibrium, some incumbent workers are displaced, while in general equilibrium they see a decrease in wages.

Finally, we conduct counterfactual simulations based on the estimated model. We begin by decomposing the effects of the different components of the apprenticeship regulation: minimum and maximum quotas, the decrease in the apprentice minimum wage, and the possibility of paying fees. The results indicate that each component plays an important role. Reducing the apprentice minimum wage significantly increases the number of trained apprentices. Mandating a minimum quota ensures that many firms, including those in high-skill sectors, train at least some apprentices. On the other hand, the maximum quota restricts the apprentice intake so that firms in low-skill sectors do not use too many apprentices as "cheap labor", merely substituting regular workers. The possibility of paying the fee instead of training serves as a buffer, reducing the negative effect on firms with very high training costs, but at the same time limits training in high-skill sectors.

We also simulate the effects of counterfactual apprenticeship policies. In particular, we study policies that incentivize firms to train apprentices by lowering the cost of training, while disposing of the size-based quotas that induce firm size distortions. The first alternative policy is a pure subsidy on training costs financed by payroll taxes. This scenario resembles real-world subsidy schemes both in developing and developed countries, and is closely related to experimental studies that provide firms with wage subsidies for apprentices (e.g. [Crépon and Premand, 2019](#), [Alfonsi et al., 2020](#)). We find that a uniform training subsidy performs similarly to the Colombian benchmark regulation on aggregate, but leads to even more concentration of training in low-skill sectors. In fact, training is concentrated only in few firms with the very lowest training costs. As a second counterfactual, we consider sector-specific minimum wages for apprentices. Specifically, the minimum wage is allowed to be lower in high-skill sectors, where firms face higher training costs, but the future benefits of

training are potentially large. We find that this type of policy can yield the same number of trained apprentices as the benchmark regulation, while limiting reallocation of production towards low-skill sectors. Furthermore, differentiating apprentice minimum wages could encourage more training in high-skill sectors. Overall, sector-specific minimum wages generate the largest welfare gain among the policies considered. We conclude that policies that take account of heterogeneity in training costs by allowing for some sector-specific aspects of apprenticeship regulation can further enhance welfare gains.

This paper contributes to the literature on training programs in developing countries. Training programs involve and affect both workers and firms. A number of recent studies provide evidence of the effects of training on workers or apprentices, including [Card et al. \(2011\)](#), [Attanasio et al. \(2011, 2017\)](#), [Hirshleifer et al. \(2016\)](#), [Kugler et al. \(2015\)](#) and [Alfonsi et al. \(2020\)](#). [McKenzie \(2017\)](#) and [Card et al. \(2018\)](#) provide reviews of this literature. Although some studies recognize the role of the demand side in making these programs successful (e.g. [Groh et al., 2016](#); [Alfonsi et al., 2020](#)), there is surprisingly little evidence on firm responses to training programs and how to effectively incentivize firms to train.² Notable exceptions include [Crépon and Premand \(2019\)](#) and [Hardy and McCasland \(2020\)](#) who study the effect of additional apprentices on firm outcomes, and related work by [de Mel et al. \(2019\)](#) who consider wage subsidies for workers. We make several novel contributions to this literature. First, the unique nature of the reform allows us to study firm responses to apprenticeship regulation along various margins, which are informative of their demand for apprentices, and we uncover important heterogeneity in responses across firms and sectors.³ Second, further exploiting this variation, we structurally estimate the underlying training costs of firms, which are key in explaining firm responses. Third, while most of the literature uses randomized experiments, we follow a complementary approach and study a large-scale policy faced by all Colombian firms. This then allows us to quantify general equilibrium and dynamic effects of the apprenticeship regulation. The relatively small partial-equilibrium effects we find are consistent with some of the mixed evidence on the benefits of training programs in the literature. However, our general-equilibrium results support the conjecture that a substantial benefit of expanding apprenticeships is an increase in the overall number of trained workers.

Moreover, our paper is related to the literature on firm-size distortions (e.g. [Besley and Burgess, 2004](#), [Guner et al., 2008](#), [Dabla-Norris et al., 2018](#)). In particular, [Garicano et al. \(2016\)](#) study how firm size-based policies affect the firm size distribution, productivity and the allocation of labor. In our counterfactual simulations, we highlight how similar benefits from apprenticeship training could be achieved by price-based policies without inducing firm size distortions.

We organize the rest of the paper as follows. Section 2 describes the data and institutional context, Section 3 presents the reduced-form evidence, Section 4 outlines the theoretical framework,

²A closely related literature analyzes the effects of apprenticeships in developed countries (see [Wolter and Ryan, 2011](#) for a review). This literature also mostly focuses on the benefits to apprentices (e.g. [Krueger and Pischke, 1995](#), [Fersterer et al., 2008](#) and [Göggel and Zwick, 2012](#)) and provides little evidence on firm responses. Exceptions include [Wolter et al. \(2006\)](#) and [Mohrenweiser and Zwick \(2009\)](#).

³A contemporaneous paper by [Ospino \(2018\)](#) studies the Colombian apprenticeship regulation using survey data and finds small positive effects of the policy on firms' productivity.

Section 5 presents the quantitative results, Section 6 discusses the counterfactual simulations, and finally Section 7 concludes.

2 Institutional Context and Data

2.1 Institutional Context and Apprenticeship Regulation

In the early 2000s, Colombia experienced high levels of informality and youth unemployment. The informality rate in 2002 was 62%, and even 70% for young workers aged 18 to 24 years. The youth unemployment rate was above 30%, twice that of other workers.⁴ These issues motivated a labor market reform in 2003, including a radical overhaul of the apprenticeship regulation.⁵ The goal of the reform was to improve skills of low-productivity individuals, provide links to formal employers and reduce youth unemployment (Gaviria and Nuñez, 2003).

A government agency, the Colombian National Training Service (SENA), is responsible for vocational training and apprenticeships. Before the 2003 reform, firms could train apprentices, but there was no minimum apprentice quota and the regulation was hardly enforced in practice. Only a maximum number of apprentices of no more than 5% of the firm’s total labor force was stated. The most prevalent way of complying was assigning regular workers to evening courses, without actually training new apprentices (Ospino, 2018). As we show in Section 3, prior to 2003 barely any firms trained apprentices.

The 2003 reform establishes a dual vocational training system with two phases, the teaching phase where apprentices are in a formal education institution, and the productive phase where they receive training in the firm.⁶ Any individual with at least basic secondary education (8th grade) can apply for apprenticeships. SENA sets the apprenticeship curriculum and provides the classroom portion of training directly for more than 80% of apprentices, teaching around 400,000 students per year (SENA, 2018). Upon completion of the apprenticeship, trained apprentices receive a certificate from SENA or another accredited institution where they received their classroom instruction.

The post-reform apprenticeship regulation has three main components:

1. **Apprentice Quotas:** First, *apprentice quotas* depending on the number of full-time workers at a firm are established. There is a minimum number of required apprentices: firms with 15 to 29 full-time workers must have at least one apprentice, increasing by one more apprentice in intervals of 20 full-time workers. Thus, if the firm has 30 to 49 full-time workers it must train at least two apprentices, between 50 and 69 it must train at least three, etc. In addition, the quota sets a maximum number of apprentices. The maximum is one apprentice for firms with less than 15 workers, and twice the minimum for firms with more than 15 workers.

⁴Own calculations based on data from the Household Survey (ECH) by the National Department of Statistics (DANE).

⁵Other elements of the reform broadly focused on increasing labor market flexibility. None of the other parts were based on the number of workers or related to the thresholds we exploit in this paper.

⁶The teaching phase lasts between one year (1760 hours) and 1.5 years (2640 hours). The productive phase lasts between 6 months and 2 years.

2. **Apprentices' Minimum Wage:** The reform also lowers the minimum wage of apprentices as an additional incentive for firms to train. Apprentices must be paid at least 50% of the regular minimum wage during the teaching phase and 75% during the productive phase.⁷
3. **Fee:** Finally, firms can pay a fee instead of hiring the minimum required apprentices. This fee is proportional to the difference between the minimum apprentice quota and the number actually hired by the firm. The amount of the fee is 100% of the (regular) minimum wage per missing apprentice.

The reform also provided SENA with tools for strict enforcement of the regulation. Whenever firms are found non-compliant because they fail to train the required number of apprentices or to pay the fee, a fine equivalent to two times the minimum wage per missing apprentice is imposed. All firms with more than 10 workers have to report the total number of hours of regular workers every 6 months. Using this information, the number of equivalent full-time workers is calculated.⁸ SENA checks the reported information by comparing it to independent data coming from payroll taxes. Next, SENA determines the apprentice quota for each firm, and firms have two months to comply with the regulation after receiving notice of the quota. Firms can hire apprentices via a centralized matching system run by SENA, or independently. Besides directly managing the teaching phase, SENA also monitors apprentices' training in firms. Apprentices report their satisfaction with the training firm twice a month, and they can be reallocated if they are dissatisfied with the quality of training.

2.2 Data and Summary Statistics

We use a novel administrative data set provided by SENA, covering the universe of manufacturing firms with at least 10 workers between 1995 and 2009.⁹ For each firm-year observation, the data includes the number of workers, the number of apprentices, and indicators for fees and fines paid by the firms in relation to the apprenticeship regulation. We link the administrative data to the Colombian manufacturing census (EAM), a rich firm-level survey data set collected by the National Department of Statistics (DANE).¹⁰ The survey data includes additional information on workers divided into three production/skill layers (professional, production and administrative workers), as well as on output, sales, wages, inputs and costs. Table 1 shows summary statistics of the data. In the full sample described in Column (1), there are 108,385 firm-year observations, and 14,586 unique firms.

Our main data is at the firm level and does not contain any further individual-level characteristics of apprentices. In order to obtain some additional information on apprentices, Appendix B shows

⁷Apprentices who receive education from a university, as opposed to vocational education institutions, are paid a full minimum wage in both phases. The regulation also specifies that all apprentices have to be paid a full minimum wage if the unemployment rate falls below 10%. This did not occur throughout the sample period.

⁸Only regular employees of the firm are counted. This excludes indirectly-hired workers, such as temporary or outsourced workers.

⁹SENA collects data for all firms in the economy. We focus on manufacturing firms with at least 10 workers, as only this sector is available in the census data in the pre-reform period.

¹⁰The data is collected by DANE at the establishment level and then aggregated at the firm level.

Table 1: Summary Statistics

	(1)	(2)	(3)	(4)
	All Sectors	<i>Full Sample</i> Low-Skill	High-Skill	<i>Threshold Sample</i>
Apprentices	1.08 (4.12)	1.72 (5.58)	0.50 (1.93)	0.74 (1.80)
Workers	49.08 (94.72)	49.54 (101.26)	48.68 (88.49)	29.82 (30.56)
Workers (Survey)	54.58 (127.33)	56.22 (139.16)	53.12 (115.77)	31.23 (34.52)
Fraction Professionals	0.08 (0.14)	0.06 (0.12)	0.10 (0.15)	0.07 (0.12)
Fraction Production Workers	0.55 (0.27)	0.56 (0.28)	0.54 (0.27)	0.60 (0.24)
Fraction Administrative Workers	0.37 (0.26)	0.38 (0.27)	0.36 (0.25)	0.33 (0.21)
Output	11,031,021 (28,153,949)	10,679,502 (27,074,391)	11,343,706 (29,077,247)	5,579,873 (12,491,212)
Value Added	4,755,892 (12,739,041)	4,412,605 (12,081,763)	5,057,618 (13,282,718)	2,243,634 (5,575,936)
Profits	3,113,409 (9,359,584)	2,893,555 (8,868,686)	3,306,645 (9,766,703)	1,456,771 (4,365,586)
Wage bill (permanent workers)	1,240,527 (3,048,647)	1,115,178 (2,839,862)	1,352,028 (3,218,945)	590,291 (1,078,797)
Total wage bill	1,609,439 (3,702,964)	1,463,753 (3,430,131)	1,739,030 (3,925,257)	773,080 (1,354,139)
Wage per worker (permanent workers)	19,733 (17,763)	18,384 (17,649)	20,885 (17,780)	17,104 (13,059)
Capital/Output	0.66 (0.88)	0.64 (0.91)	0.67 (0.85)	0.69 (0.91)
Intermediates/Output	0.54 (0.18)	0.56 (0.19)	0.53 (0.18)	0.55 (0.18)
Observations	108,385	51,024	57,361	14,848
Firms	14,586	7,403	7,986	2,018

Notes: Summary statistics for years 1995 to 2009. All monetary variables in 2009 thousands of pesos. Standard deviations in parentheses.

descriptive statistics based on two other data sources. The first dataset is a small-scale survey on school-to-work transitions of young individuals. The second dataset is administrative data from social insurance records, which has a larger sample but fewer observed characteristics. Unfortunately this data does not coincide with our period of analysis but is only available in 2013 to 2015 and 2015 to 2016, respectively. First, Appendix Table B.1 summarizes apprentices’ characteristics. Apprentices tend to be in their early 20s and 57% are female according to the administrative data. The majority have high-school or some secondary education. In line with the regulation, the average apprentice wage is below the regular minimum wage, which is significantly lower than all other workers as well as other young workers of the same age group. Most apprentices self-report their socioeconomic status as low or medium. It is also worth noting that more than 90% are satisfied with their current apprenticeship/job, and less than 30% want to move jobs, which reflects higher satisfaction than other young workers.

Appendix Figure B.1 shows wages of apprentices around the month they complete the apprenticeship. Immediately upon graduation, there is a jump by around 40%, illustrating the returns to apprenticeship training. The figure also shows a second series for apprentices who do not graduate in the sample period, whose wage stays around the minimum wage. Finally, Appendix Table B.2 shows transition probabilities of graduating apprentices. Around three quarters stay in the same firm after completing the apprenticeship, and the majority of those that leave the firm stay in the same sector.

3 Reduced-Form Results

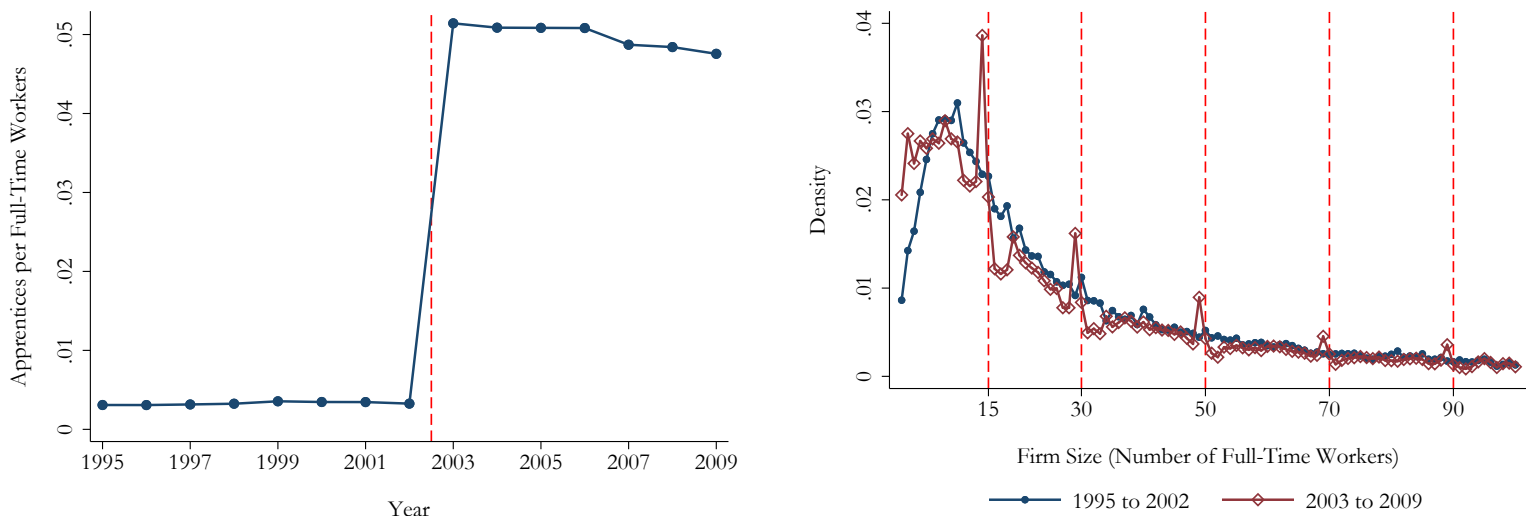
3.1 Aggregate Effects: Number of Apprentices and Firm Size Distribution

The primary objective of the apprentice regulation is to expand the scope of apprenticeship training. Figure 1a shows that the policy is successful in this regard, dramatically increasing the total number of apprentices. Before the reform, there are around 0.3 apprentices per 100 full-time workers, and this increases by an order of magnitude to around five apprentices per 100 full-time worker after the reform. The total number of apprentices in the manufacturing sector increases from below 1000 just before the reform to more than 15,000 just after, while the average number of full-time workers is relatively stable.

However, the regulation also induces changes in the firm size distribution by setting apprentice quotas as a function of the number of full-time workers. Figure 1b shows that such distortions in the post-reform number of full-time workers are sizeable. In the pre-reform years 1995 to 2002, the firm size distribution is relatively smooth. In contrast, the post-reform distribution in the years 2003 to 2009 exhibits pronounced spikes around the regulation thresholds marked by the dashed vertical lines, and holes or “missing mass” on both sides of the thresholds.¹¹ The figure provides first visual evidence that some firms change their labor inputs in response to the policy. Moreover, the fact

¹¹Appendix Figures A.1 and A.2 show distributions year by year, exhibiting the same patterns as the pooled distributions.

Figure 1: Number of Apprentices and Firm Size Distribution



(a) Number of Apprentices over Time

(b) Firm Size Distribution Before and After the Reform

Notes: Panel (a) shows the number of apprentices per full-time worker in the full sample. Panel (b) shows the distribution of full-time workers for all firms before (1995-2002) and after the reform (2003-2009), using a bin size of one.

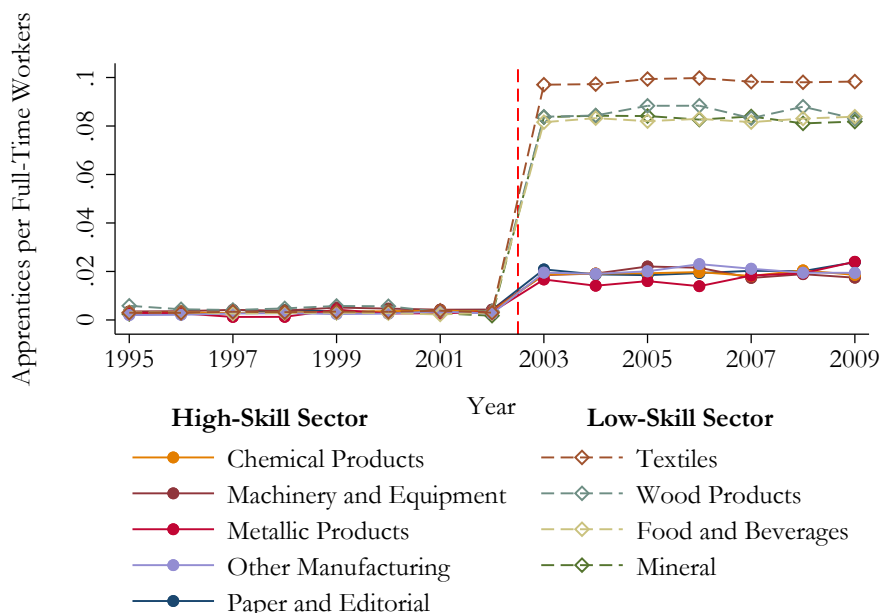
that there is missing mass on both sides of the thresholds gives a first indication of heterogeneous responses, as some firms seem to avoid being just below while others avoid being just above the thresholds.

To study the heterogeneity in firm responses, we divide firms into nine two-digit sectors using the Colombian industrial classification.¹² Figure 2 plots the number of apprentices per full-time worker in each sector. The figure suggests a clear division in apprentice intake across sectors. In the wood products, textiles, food and beverage, and mineral non-metallic products sectors there are between eight and ten apprentices per 100 full-time workers. In contrast, in the paper and editorial, machinery and equipment, metallic products, chemical products and other manufacturing sectors, there are only around two apprentices per 100 full-time workers.

To make sense of these differences in apprentice intake, we rank sectors by the fraction of professional workers with tertiary education out of all workers. This fraction can be interpreted as a measure of skill requirements in each industry, reflecting the difficulty or costs of training apprentices. We denote those below the median as *low-skill sectors* and those above the median as *high-skill sectors*. Using this definition has remarkable power in explaining differences in the number of apprentices in the post-reform period. The four sectors classified as low-skill (wood products, textiles, food and beverage, and mineral non-metallic products) are precisely those with the most apprentices in Figure 2, whereas the five sectors classified as high-skill (paper and editorial, machinery and equipment, metallic products, chemical products, and other manufacturing) take significantly fewer apprentices.

¹²We use the industrial classification by DANE, CIU 3 A.C, which is adapted from the International Standard Industrial Classification (ISIC).

Figure 2: Number of Apprentices By Sector



Notes: The graph shows the number of apprentices per full-time worker by two-digit sector. “High-skill sector” refers to sectors above the median share of skilled workers (professionals), “low-skill sector” refers to sectors below the median.

Columns (2) and (3) of Table 1 show summary statistics for firms in low-skill vs. high-skill sectors. On average, firms in high-skill sectors have almost double the number of highly-skilled (professional) workers (10% vs. 6% in low-skill sectors), and they have fewer unskilled production workers and administrative workers. Although the average high-skill sector firm has fewer workers than the average low-skill sector firm, output, value added and profits are higher. High-skill sector firms pay higher wages per worker, and they use more capital and less intermediate inputs than low-skill sector firms. In addition, Appendix Table A.1 shows some descriptive statistics for each 2-digit sector. The fraction of professionals varies between 3.5% and 7% in low-skill sectors and between 7.3% and 8.6% in high-skill sectors. Moreover, the patterns in some other characteristics that may be naturally related to skill requirements would imply a similar ranking of sectors to our baseline measure. In particular, the table shows average wages of all workers, average wages of production workers, the fraction of permanent full-time workers, and the capital-output ratio. The ranking of 2-digit sectors using these alternative measures remains relatively stable, with the exception of the mineral non-metallic sector where there are few observations.

3.2 Heterogeneous Firm Responses Across Sectors

Next, we provide evidence of heterogeneous firm responses to the apprentice regulation across low-skill and high-skill sectors. We show that firms in high-skill sectors tend to avoid training apprentices, while firms in low-skill sectors tend to train as many apprentices as possible. We organize the results into three empirical facts, studying firm-size distributions, apprentice intake by firm size, and the fraction of firms paying fees in each sector.

Fact 1: Firms in high-skill sectors bunch below the thresholds; firms in low-skill sectors bunch at the thresholds.

Figure 3 presents the firm size distribution around the first five thresholds for firms in high-skill sectors (Panel 3a) and in low-skill sectors (Panel 3b), pooling the post-reform years 2003 to 2009. There are pronounced spikes in both panels, but the figure reveals a crucial difference in the location of bunching across sectors. Firms in high-skill sectors bunch *below* the thresholds, while firms in low-skill sectors bunch exactly *at* the thresholds. Moreover, for high-skill sector firms there is missing mass above the thresholds, while for low-skill sector firms the missing mass is below the thresholds. Taken together, this indicates that some firms in high-skill sectors reduce the number of full-time workers to avoid the higher minimum apprentice quota that would apply above the thresholds, such that they have to train fewer apprentices. Low-skill sector firms, on the other hand, increase the number of full-time workers in order to increase their maximum apprentice quota, such that they can train more apprentices.

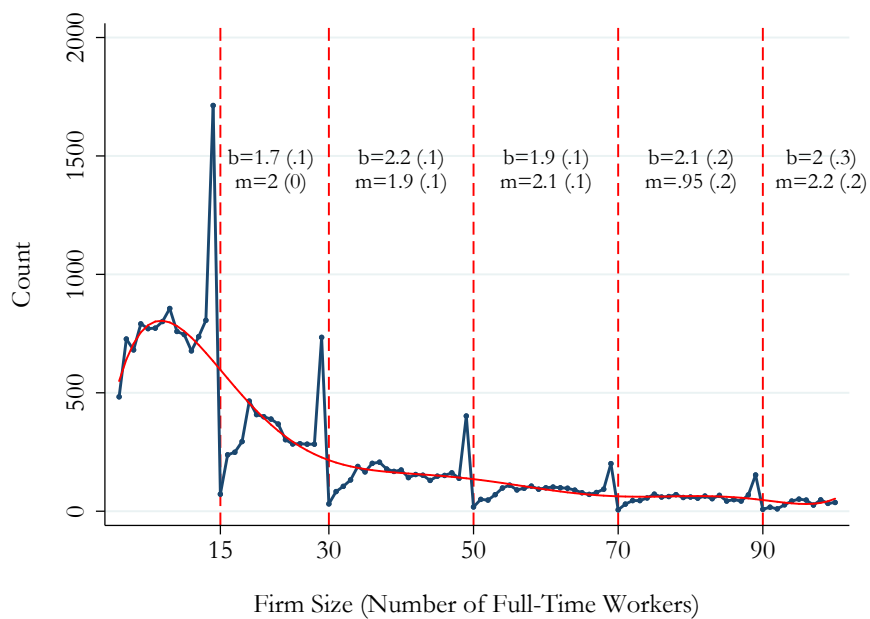
In order to quantify firm size responses, we use the bunching method (Saez, 2010, Chetty et al., 2011, Kleven, 2016). In each panel, we fit a flexible 7th-order polynomial to the distribution of firm size n to construct the smooth counterfactual distribution shown in the solid red line. The bunching and missing mass regions just around the threshold are excluded from this counterfactual estimation.¹³ We then compute the *excess mass* $b = B/h_o(\hat{n})$ as the count of firms B at the threshold \hat{n} relative to the estimated counterfactual $h_o(\hat{n})$. The key identification assumption is that the density would have been smooth in the absence of the policy, which is directly supported by the fact that the pre-reform distribution is smooth in Figure 1b. Similarly, we compute the *missing mass* as the “hole” in the observed distribution relative to the counterfactual, $m = M/h_o(\hat{n})$, where M is the firm count in the missing mass region. For both excess mass and missing mass estimates, bootstrapped standard errors are shown in parentheses.

The estimates in Figure 3 show sizeable and significant excess mass and missing mass in both panels. The fact that excess mass and missing mass are similar at each threshold confirms that bunching responses indeed originate from the neighborhood of the threshold. Comparing across sectors, bunching responses are somewhat larger in high-skill sectors, where the excess mass is between 1.7 and 2.2 across the five thresholds. This can be interpreted in terms of a firm size response: The marginal bunching firm is estimated to reduce the number of workers by around two. In low-skill sectors, on the other hand, the implied firm size response goes in the opposite direction, where the marginal bunching firm increases their firm size by between 0.6 and 1.8 workers. Finally, note that the patterns in the figure indicate that there seem to be little frictions that would prevent firms from adjusting their size. In both Panels (a) and (b), bunching responses are sharp with little or no diffuse excess mass around, and the density drops to close to zero at the thresholds in Panel (a) in

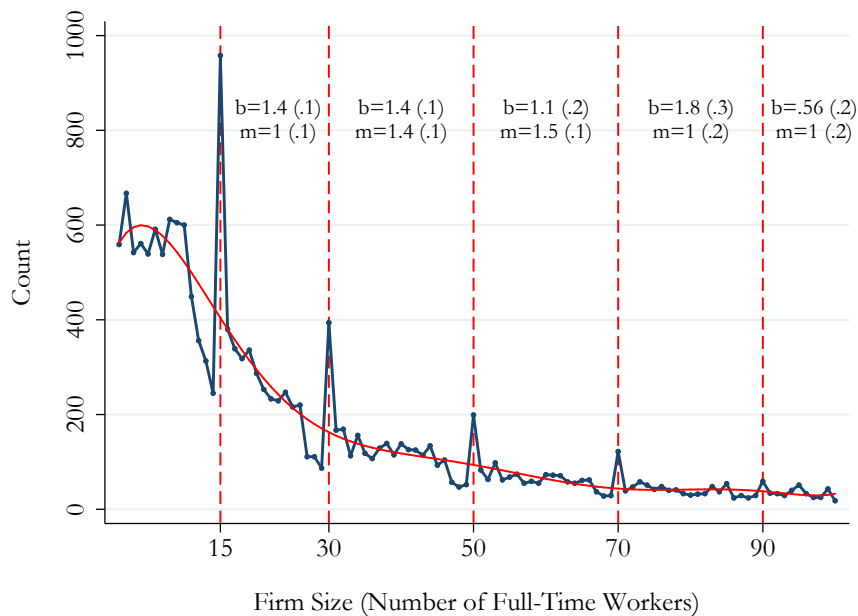
¹³In line with the sharp bunching responses, the bunching region is chosen as one bin below the threshold in Panel (a) and the bin at the threshold in Panel (b). The missing mass region is the threshold plus three bins to the right in Panel (a), and the three bins below the threshold in Panel (b), which fits the observed density hole well. We do not explicitly restrict the missing mass to equal the bunching mass, but the two tend to line up well for our choice of estimation regions.

particular.

Figure 3: Fact 1 - Bunching Responses in High-skill and Low-skill Sectors



(a) High-skill Sectors



(b) Low-skill Sectors

Notes: Distribution of the number of full-time workers in high-skill and low-skill sectors post-reform (2003-2009), using a bin size of one. The dashed vertical lines denote the regulation thresholds. The solid red line shows the fitted counterfactual density. Excess mass b and missing mass m is reported at each threshold, with bootstrapped standard errors in parentheses.

Fact 2: Firms in high-skill sectors tend to choose the minimum number of apprentices; firms in low-skill sectors tend to choose the maximum.

Figure 4a shows the number of apprentices by firm size in high-skill and low-skill sectors, as well as the minimum and maximum apprentice quotas from the regulation. The figure suggests that firms in high-skill sectors generally try to avoid training apprentices while firms in low-skill sectors tend to train as many as possible. Firms in high-skill sectors have an average number of apprentices below the minimum quota of the regulation. Thus, some of these firms must be paying the fee or not comply with the regulation. In contrast, firms in low-skill sectors have an average number of apprentices close to the maximum quota, indicating that some of them may be bounded by this upper limit.¹⁴

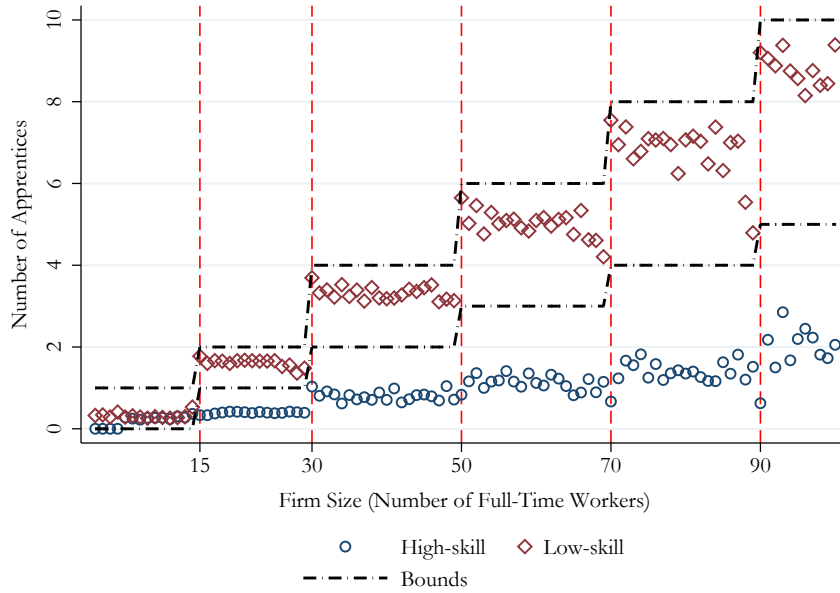
In line with the discontinuously increasing apprentice quotas, there are jumps in the average number of apprentices at the regulation thresholds. These jumps are particularly sharp in the low-skill sectors. Moreover, within each bracket of the regulation, the average number of apprentices follows a decreasing pattern, which is again more marked in low-skill sectors. The number of apprentices is high just above the thresholds, relatively constant for a few bins to the right, and decreases in the highest bins of each bracket. Hence, the overall picture reflects a mixture of the causal effect of the policy and selection. Firms in brackets with higher apprentice quotas have to take more apprentices as a result of the regulation, but locally firms can select into brackets via bunching responses. The firms bunching at or just above the thresholds are those who wish to hire many apprentices, while firms locating just below the thresholds are those who wish to hire few apprentices.

Table 2 summarizes these responses in more detail. The table shows the proportion of firms choosing the maximum number of apprentices, the minimum number, a number between the maximum and the minimum, below or above these bounds, and those that do not train any apprentices. The most common responses to the regulation are choosing exactly the minimum number, exactly the maximum number, or no apprentices at all. Together these responses account for close to 95% of observations across all sectors. However, responses differ strongly across high-skill and low-skill sectors. While 63% of firms in high-skill sectors train no apprentices at all, almost all firms in low-skill sectors (99.7%) train apprentices. Whenever high-skill sector firms have apprentices, they tend to choose exactly the minimum number required. On the other hand, almost two thirds of firms in low-skill sectors choose the maximum number of apprentices, and the remainder either chooses a number of apprentices between the regulation bounds or the minimum number required.

These results have two additional important implications. First, most apprenticeship training happens in low-skill sectors. In the post-reform years, 77% of apprentices are trained in low-skill sector firms, although there are a similar number of firms in low-skill and high-skill sectors. The fact that low-skill sector firms increase their size but high-skill sector firms decrease their size further exacerbates the increasing employment share in low-skill sectors. Second, the results imply an important role of the apprentice minimum wage. In particular, the observation that many firms hire more than the minimum number of apprentices required suggests that it is now worthwhile training

¹⁴Appendix Figure A.4 shows that the pre-reform number of apprentices by firm size is low and similar across all sectors.

Figure 4: Number of Apprentices and Share of Firms Paying Fees



(a) Fact 2 - Number of Apprentices by Firm Size



(b) Fact 3 - Share of Firms Paying Fees by Firm Size

Notes: Panel (a) shows the average number of apprentices by firm size bin for high-skill and low-skill sector firms, post-reform (2003-2009). The black dashed lines show the minimum and maximum apprentice quotas, and the red dashed vertical lines denote the regulation thresholds. Panel (b) the fraction of firms paying fees by firm size bin in high-skill and low-skill sectors, post-reform (2003-2009).

Table 2: Summary of Training Responses

Type of Training Response	(1) Low-Skill Sectors	(2) High-Skill Sectors	(3) All
Maximum Number of Apprentices	65.4%	1.5%	30.2%
Minimum Number of Apprentices	26.9%	31.9%	29.7%
Between Minimum and Maximum	7.0%	3.1%	4.9%
Above Maximum	0.0%	0.0%	0.0%
Below Minimum but >0	0.3%	0.0%	0.2%
Zero Apprentices	0.3%	63.6%	35.1%
Observations	11,824	14,470	26,294

Notes: The table shows the fraction of firms by type of training response to the regulation in the post-reform years 2003 to 2009. Only firms subject to a minimum quota, i.e. with at least 15 full-time workers, are included.

extra apprentices. This cannot be explained by the apprentice quotas alone, but it is consistent with an additional effect of the lower minimum wage for apprentices.

Fact 3: Only high-skill sector firms tend to pay the fee.

Figure 4b shows the fraction of firms paying the fee by firm size. In high-skill sectors, around 60% of firms choose to pay the fee instead of training the minimum number of apprentices required. Note that the nominal cost of hiring an apprentice is smaller (50% to 75% of the minimum wage) than the fee (100% of the minimum wage). Thus, the responses of high-skill sector firms indicate that their training costs must be high. In fact, apprentices must create an overall negative marginal benefit for many firms as they are unwilling to hire apprentices even though it is nominally cheaper than not hiring apprentices. Later on, this fact disciplines the admissible production functions in the theoretical analysis.

Moreover, the fraction of high-skill sector firms paying the fee is relatively stable across different firm sizes. This suggests that training apprentices entails a proportional cost to firms rather than a fixed cost, as in the case of fixed training costs one would expect that large firms subject to higher apprentice quotas should be less likely to pay the proportional fee. This informs the specification of training costs in the model. In contrast, low-skill sector firms rarely pay the fee. In addition, we show that very few firms are fined for non-compliance in Appendix Figure A.3. The vast majority of firms comply with the regulation by either taking the required number of apprentices or paying fees, which is consistent with the strict enforcement described in Section 2.

Correlates of Firm Responses

Appendix Table A.2 shows regressions of indicators for the different types of responses to the regulation on firm characteristics. In order to avoid endogenous changes in these characteristics, the post-reform responses are regressed on the pre-reform firm-level average of each characteristic. The table confirms the strong sectoral divide in responses. Even after controlling for a broad set of firm-level characteristics including firm size, wages, output, profits and inputs, choosing the maximum

apprentice quota is strongly correlated with being in a low-skill sector, whereas choosing the minimum quota or paying the fee is strongly correlated with being in high-skill sectors. The correlation of responses with pre-reform firm size is very small in magnitude, suggesting that responses are similar across the firm size distribution. Choosing the maximum number of apprentices is more common among firms with a lower fraction of production workers and higher wages, but with lower output, labor productivity, profit rate and capital-to-output ratio, whereas the opposite signs apply to choosing the minimum number. Paying the fee to avoid apprentices seems to be common among all types of firms within high-skill sectors, as there is no significant relationship with most firm characteristics except negative correlations with the fraction of professional and production workers.

Finally, Appendix Table A.3 shows the correlation of bunching behavior with the different types of responses to the regulation. In the post-reform period, there are over 4000 observations of firms bunching below the regulation thresholds, of which 88% are in high-skill sectors. As expected, bunching below the thresholds tends to coincide with avoiding apprentices. 49% of bunchers below choose the minimum number of apprentices and 27% pay the fee. Bunching above the thresholds, on the other hand, is concentrated in low-skill sectors and tends to coincide with choosing the maximum number of apprentices, which 72% of bunchers above do.

4 Model: Heterogeneous Training Costs

In this section, we develop a model with heterogeneous firms that rationalizes our empirical findings and allows us to estimate training costs and to further quantify the effects of the policy. Firms differ in two dimensions in the model: their costs of training apprentices and their managerial ability. The training costs come from a sector-specific distribution and explain the heterogeneous firm responses uncovered by the reduced-form analysis, while managerial ability gives rise to the firm size distribution.

4.1 Model Setup and Equilibrium without Regulation

Consider an infinite-period economy composed of $K \geq 1$ sectors with a fixed number of heterogeneous firms in each sector. Firms in sector k are characterized by training costs t_a^k and managerial ability z^k , which follow sector-specific distributions \mathcal{T}^k and \mathcal{Z}^k . These firm characteristics are invariant across time. Managerial ability z^k is an idiosyncratic firm characteristic that can also be interpreted as technological differences or other factors that affect a firm's productivity. Firms produce y_t^k units of good k in period t using labor, which is supplied either by workers l_t^k or by apprentices $l_{a,t}^k$.¹⁵ We suppose that workers are sector-specific, but apprentices can be trained in any sector. All individuals are endowed with a unit of time which they supply inelastically.

Apprentices are trained using workers' time. Let $t_a^k \geq 0$ denote training costs in units of time workers have to spend training the apprentice. Across firms and sectors, apprentices potentially

¹⁵The model could be readily extended to include capital or other production inputs. Here we suppose that labor is the only input to simplify the analysis and emphasize the role of apprentices. In Appendix C.2 we extend the model to allow for other endogenous inputs, showing that the results remain qualitatively the same.

require different amounts of time of training. A sector whose technology requires simple menial tasks probably requires less time to train apprentices compared to a sector with highly specialized tasks. Production combines a firm's managerial ability z^k with the total labor supplied by both types of workers given training costs t_a^k . If a firm hires n_t^k workers and trains $n_{a,t}^k$ apprentices, the total labor supplied by workers is $l_t^k := n_t^k - t_a n_{a,t}^k$ and labor supplied by apprentices is $l_{a,t}^k := \zeta_a^k n_{a,t}^k$, where ζ_a^k is the productivity of apprentices relative to workers. This reflects that apprentices' productivity per unit of time they contribute to production may also differ across sectors. Finally, firms in different sectors also differ in their production technology $f^k(l_t^k, l_{a,t}^k; z^k)$. We assume that the production function is increasing in labor inputs $(l_t^k, l_{a,t}^k)$ and in managerial ability.

Firms maximize profits by choosing the number of workers and apprentices in each period. Firms in sector k take as given the price of the good they produce p_t^k , and the wages of workers and apprentices w_t^k and $w_{a,t}^k$. A firm in sector k with managerial ability z^k and training costs t_a^k solves

$$\max_{(n_t^k), (n_{a,t}^k) \geq 0} \sum_{t=0}^{\infty} \beta^t [p_t^k f^k(n_t^k - t_a^k n_{a,t}^k, \zeta_a^k n_{a,t}^k; z^k) - w_t^k n_t^k - w_{a,t}^k n_{a,t}^k] \quad s.t. \quad t_a^k n_{a,t}^k \leq n_t^k \quad \forall t. \quad (1)$$

The constraints $t_a^k n_{a,t}^k \leq n_t^k$ ensure that the firm must hire enough workers to train the chosen number of apprentices in every period. Note that the firm's choice is static and symmetric across all sectors. We therefore drop sector and time superscripts and subscripts to simplify notation whenever there are no ambiguities.

We can further characterize the solution taking the FOCs of (1),

$$[n]: p \frac{\partial f}{\partial l} \leq w \quad , \quad [n_a]: p \frac{\partial f}{\partial l_a} \leq \frac{w_a + t_a w}{\zeta_a}.$$

The FOC for the number of apprentices intuitively reflects how firms compare the marginal product of an additional apprentice to their marginal cost. Their marginal product is scaled by ζ_a^k relative to the productivity of workers. The marginal cost of an apprentice is not only their wage w_a , but also t_a units of time of a worker that earns wage w , which sums up to $w_a + t_a w$.

Our empirical findings discipline the family of production functions consistent with the observed firm responses. In particular, given that the policy imposes a fee that is nominally larger than the apprentice minimum wage, we show that the production function has to allow for apprentices to have negative marginal productivity for the firms that choose to pay this fee. The production function in (1), $f(n - t_a n_a, \zeta_a n_a; z)$ implies that the marginal product of apprentices is $\frac{\partial f}{\partial n_a} = -f_l t_a + f_{l_a}$, which clearly can be negative when training costs are high and/or the marginal productivity of labor f_l is high.¹⁶ In Appendix C.1 we formalize this argument, and also show the existence and uniqueness of the solution with standard regularity conditions on the production function (see Assumption 1). In addition, the appendix shows that demand for apprentices decreases in their relative wages and in training costs.

¹⁶This simple production function relates to a broader literature highlighting the importance of worker complementarities, knowledge and time constraints in production (e.g. [Kremer, 1993](#), [Garicano and Rossi-Hansberg, 2004](#), [2006](#)).

Labor Markets, Preferences and Equilibrium

We consider a simple supply side, where the number of apprentices trained in a sector increases the number of workers in that sector in the next period. Let L_t^k and $L_{a,t}$ denote the supply of workers in sector k and the total number of untrained apprentices in period t . Workers can do the tasks of apprentices but not the other way around. This implies that in equilibrium, apprentices' wages are smaller or equal to the ones of workers in each sector. In addition, the minimum wage could be binding. Both constraints together imply $w_t^k \geq w_{a,t}^k \geq w_{min} \geq 0, \forall t, k$.

We suppose that the labor market clears separately by sector. The wage constraints imply that some workers could remain unemployed and potential apprentices untrained. Let N_t^k and $N_{a,t}^k$ denote aggregate demand for workers and apprentices, respectively. The market clearing conditions are $N_t^k + U_t^k = L_t^k$, $\sum_k N_{a,t}^k + U_{a,t} = L_{a,t}$, where $U_t, U_{a,t} \geq 0$ denote unemployed workers and untrained apprentices, respectively.

Trained apprentices increase the supply of workers in their sector of training. This dynamic component of the model captures potentially important benefits of the policy. The supply of workers in each sector k one period ahead ($t + 1$) satisfies

$$L_{t+1}^k = L_t^k + \chi_a^k N_{a,t}^k, \quad (2)$$

where χ_a^k denotes the effective units of labor a trained apprentice contributes to next-period supply of labor in the sector.

Having separate labor markets for workers by sector allows us to account for wage differences across sectors as observed in the data. Effectively this amounts to an assumption that the skills of trained apprentices becoming workers are sector-specific but equally useful for firms within the same sector. This assumption is guided by two considerations. First, vocational training is usually tailored to a particular sector, and trained apprentices obtain an industry-specific certificate (see [Fazio et al., 2016](#)). Second, the assumption is further supported by the evidence on apprentices' firm and sector transitions in [Appendix Table B.2](#). The table suggests that the human capital of trained apprentices is to a large extent sector-specific, as the majority stay in the same sector after graduation. In addition, we consider the case with multiple types of workers, where markets clear by occupation instead of sector in [Appendix C.3](#). Theoretical results remain similar to the baseline model, but for a quantitative implementation of such an alternative model we would need information on the number of apprentices trained in each occupation.

In the model, inefficiency in training provision arises due to a combination two factors. First, skills are transferrable across firms within a sector and there is free labor mobility of apprentices after training. Second, apprentices are unable to pay for their training, i.e. to fully compensate firms for their training cost due to the minimum wage floor. As discussed in [Section 3.2](#), the reduced-form results indicate that the initially high minimum for apprentices is an important contributing factor to the low levels of training pre-reform. Setting the apprentice minimum wage below the minimum wage for regular workers expands the scope of training of individuals with low initial productivity, as it enables apprentices to effectively compensate firms for their training costs. Note that we do

not include any additional sources of inefficient training provision in the absence of regulation in the model. Doing so would further exacerbate the positive effects of policies aimed at increasing training.¹⁷

To close the model we adopt a simple preference structure. All individuals have a common utility function $u(\cdot)$ over the goods produced by each sector. They choose a consumption bundle $c_t = (c_t^1, \dots, c_t^K)$, taking prices p_t^k as given. Hence, changes in individuals' welfare come solely from their incomes that depend on the sector and on whether they are trained or untrained apprentices, workers or firm owners/managers. Trained apprentices earn wage $w_{a,t}$, workers in sector k earn wage w_t^k , firm owners earn profits $\pi(z^k, t_a^k)$, and untrained apprentices earn a subsistence income $\underline{w} \leq w_{min}$.¹⁸ Individuals are infinitely lived and maximize the usual discounted lifetime utility. A competitive equilibrium is defined as the set of wages and prices in each sector and period such that firms choose the number of apprentices and workers optimally, all individuals choose their optimal consumption bundles from goods in different sectors and labor and goods markets clear.¹⁹

4.2 Equilibrium with Apprenticeship Regulation

Next, we incorporate the firm-size based apprenticeship regulation into the theoretical framework. First, the regulation imposes apprentice quotas based on the number of workers. Let $(N_j)_{j=0}^\infty$ be a sequence of thresholds, where $N_0 = 0$. If the number of workers is $n \in [N_{j-1}, N_j)$, then the number of required apprentices is $n_a \in [\underline{n}_a^j, \bar{n}_a^j]$, $\forall j \geq 1$. Second, the regulation reduces the minimum wage of apprentices to w_{min}^a , below the minimum wage for workers w_{min} . Third, firms can pay a fee $\mathcal{F}_a(n, n_a)$ instead of training the required apprentices. This fee is a function of the number of workers and apprentices in the firm. It is proportional to the difference between the minimum number of required apprentices \underline{n}_a^j and the apprentices hired n_a .

A firm (z, t_a) facing this regulation solves

$$\begin{aligned} \max_{n, n_a \geq 0, d_f \in \{0, 1\}} \quad & f(n - t_a n_a, n_a; z) - wn - w_a n_a - d_f \mathcal{F}_a(n, n_a) \quad s.t \quad t_a n_a \leq n \\ & (n, n_a, d_f) \in \bigcup_j [N_{j-1}, N_j) \times [\underline{n}_a^j, \bar{n}_a^j] \times \{0\}, \quad \text{or} \\ & (n, n_a, d_f) \in \bigcup_j [N_{j-1}, N_j) \times [0, \underline{n}_a^j] \times \{1\} \quad \text{and} \quad \mathcal{F}_a(n, n_a) = \phi_a (\underline{n}_a^j - n_a)^+. \end{aligned} \quad (3)$$

where $d_f \in \{0, 1\}$ is the decision whether to pay the fee or not, and $\phi_a > 0$ is a positive constant. When a firm decides to pay the fee, it can train fewer apprentices than the minimum required \underline{n}_a^j and pay an amount proportional to the difference between the minimum quota and the apprentices hired, $\mathcal{F}_a(n, n_a) = \phi_a (\underline{n}_a^j - n_a)^+$. As in the actual regulation firms are not allowed to exceed the

¹⁷There is a rich theoretical literature on sources of inefficiency in training provision, including the seminal work of [Becker \(1964\)](#) who emphasizes the role of general vs. firm-specific training and [Acemoglu and Pischke \(1998, 1999\)](#) focusing on imperfect information and imperfect competition. [Garicano and Rayo \(2017\)](#) and [Fudenberg and Rayo \(2019\)](#) argue that knowledge transfer to apprentices can be inefficiently slow.

¹⁸This income represents the outside option of apprentices. It can be interpreted as the wage they would receive in the informal sector or an unemployment benefit.

¹⁹See [Appendix C.1](#) for the details and the formal definition of a competitive equilibrium.

maximum quota. Thus, the fee function only takes into account the positive difference between the minimum quota and the number of apprentices trained by the firm.

Now we characterize the firm's solution to (3) in relation to the optimum without regulation. Let $n_a^*(z, t_a)$ be the optimal number of apprentices for a firm (z, t_a) with no-regulation size $n^*(z, t_a)$. We study three cases: when the optimal number of apprentices is above the maximum quota $n_a^*(z, t_a) > \bar{n}_a^j$, when it is between the bounds $n_a^*(z, t_a) \in [\underline{n}_a^j, \bar{n}_a^j]$ and when it is below the minimum quota $n_a^*(z, t_a) < \underline{n}_a^j$, for some threshold N_j . We show that if the optimal number of apprentices is above the maximum quota, firms either choose the number of apprentices at that maximum or bunch at a threshold to get more apprentices. In the second case, firms do not change their decision as the optimal number of apprentices is within the regulation bounds. In the third case, since the optimal number of apprentices is below the minimum quota, some firms want to avoid training apprentices. They can do so two ways. Some reduce their size just below the threshold to avoid the higher minimum quota. Others choose to pay the fee if it is sufficiently low. Lemma 3 in Appendix C.1 compiles these results.

We add more structure to the production function and focus on the case where the optimal number of apprentices without regulation is a fixed proportion of the labor force, $n_a^* = Bn^*$ with $B \in [0, \infty)$. Concretely, we suppose that the production function satisfies regularity and homogeneity conditions in Assumption 1 that guarantee the existence of this fixed proportion solution to the firms problem without regulation. Below we show that production functions with these properties fit the data well.²⁰

Assumption 1. (Production Function) Suppose $f : \mathbb{R}_+^3 \rightarrow \mathbb{R}_+$ is \mathbb{C}^2 and

- (i) is homogenous of degree $\gamma \in (0, 1)$ in (l, l_a) and has CRS in (l, l_a, z) ,
- (ii) the Inada conditions hold for workers' labor, $\lim_{l \rightarrow 0} \frac{\partial f}{\partial l} = \infty$,
- (iii) has non-negative cross derivatives with respect to z , i.e. $\frac{\partial^2 f}{\partial l \partial z} \geq 0$ and $\frac{\partial^2 f}{\partial l_a \partial z} \geq 0$.

Using these properties, we show that if relative wages of apprentices and the training cost are low enough, firms want to train as many apprentices as possible. In this case, the optimal number of apprentices is above the maximum quota, some firms to the left of the thresholds N_j increase their size and bunch at the thresholds, and firms never pay the fee. In contrast, when relative wages or training costs are sufficiently high, the optimal number of apprentices converges to zero, below the minimum quota. This implies that some firms to the right of the thresholds N_j reduce their size and bunch just below in order to avoid training extra apprentices. Additionally, if the fee is low enough (ϕ_a small), some firms prefer to pay the fee instead of having additional apprentices. Proposition 1 formalizes and compiles these results.

Proposition 1. Suppose Assumption 1 holds and firms solve the maximization problem with regulation (3). Then,

²⁰For completeness, Lemma 4 in Appendix C.1 characterizes all other solutions.

Case 1: there exist $(\frac{w_a}{w}, \bar{t}_a)$ such that for $\frac{w_a}{w} \leq \frac{w_a}{w}$ and $t_a \leq \bar{t}_a$,

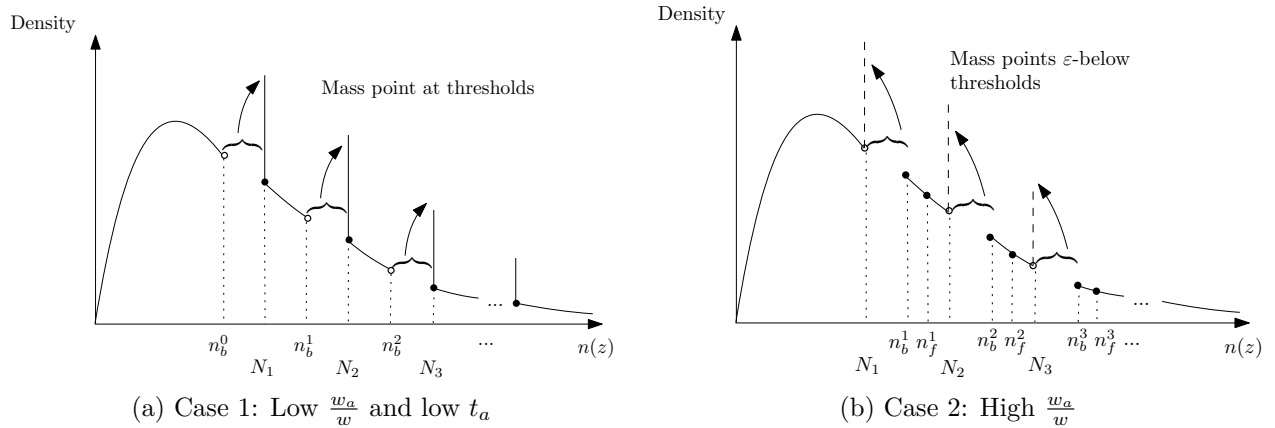
- i. the number of apprentices without regulation is $n_a^* = B_u n^*$ and is above the maximum quota, $n_a^*(z, t_a) > \bar{n}_a^j$.
- ii. there exist cutoffs $\{z_b^j, z_r^j\}_j$ such that firms $z \in [z_b^j, z_r^j]$, increase their size to the threshold N_k with $k > j$, so there is missing mass to the left of the thresholds.
- iii. firms choose the upper-bound of the regulation $n_a^r = \bar{n}_a^j$.
- iv. firms never pay the fee.

Case 2: there exist $\frac{w_a}{w}$ such that for $\frac{w_a}{w} \geq \frac{w_a}{w}$,

- i. the number of apprentices without regulation is $n_a^* = B_s n^*$ and is below the minimum quota, $n_a^*(n) < \underline{n}_a^j$.
- ii. there exist cutoffs $\{z_b^j, z_r^j\}_j$ such that firms $z \in [z_b^j, z_r^j]$, reduce their size ϵ below a lower threshold N_k for $k \leq j$.
- iii. firms that choose to increase apprentices, choose the minimum number \underline{n}_a^j .
- iv. there exist $\bar{\phi}_a > 0$ such that for $\phi_a \leq \bar{\phi}_a$, there is an additional cutoffs z_f^j such that firms $z \in (z_r^j, z_f^j]$ choose to pay the fee.

Figure 5 shows the implications of the policy for the firm size distribution in the two cases from Proposition 1. Panel 5a depicts case 1 where firms increase their size to train more apprentices. There is bunching at each of the thresholds $\{N_j\}_j$ and missing mass of firms between $[n_b^j, N_{j+1})$. Panel 5b illustrates case 2, where firms either reduce their size or pay the fee to avoid training more apprentices. Firms bunch below each threshold, leaving missing mass on $[N_j, n_b^j)$. If the fee is low enough, firms of size $[n_b^j, n_f^j]$ prefer to pay the fee instead of having the required apprentices.

Figure 5: Firm Size Distribution



Notes: The figure shows the predicted firm size distribution as described in proposition 1.

Proposition 1 provides a framework to understand our three empirical facts. Reducing the minimum wage for apprentices makes it profitable for firms in low-skill sectors, where training costs

are low, to hire as many apprentices as possible. These firms bunch at the thresholds to be able to train more apprentices, choose the maximum quota of the regulation and never pay the fee. On the other hand, the decrease in apprentices' wages is not sufficient to persuade firms in high-skill sectors, where training costs are high, to train more than the minimum required. Moreover, many of these firms avoid training additional apprentices by decreasing their size and bunching just below the thresholds, or by paying the fee.

5 Quantitative Exercises

In this section, we parameterize and estimate the model. Based on the reduced-form results, we consider low-skill sectors u and high-skill sectors s . In order to identify the structural parameters of the model, we exploit the same moments of firm responses as in the reduced-form estimation, plus additional pre-reform data on firm size and production. Our main objectives are to uncover the training cost distribution by sector, and to quantify the effect of the regulation on aggregate outcomes and the welfare of apprentices, workers and firms.

5.1 Parameter Estimation and Moments

Training Costs, Production Function and Managerial Ability Distributions

For the quantitative exercises, we have to choose functional forms of the training cost distribution, the production function and the managerial ability distribution. The reduced-form results indicate that training costs differ across sectors, but they also vary to some extent within sector, as for instance some high-skill sector firms pay fees while others train apprentices. To reflect this, we use sector-specific training cost distributions. We estimate a non-parametric distribution $\mathcal{T}^k(\cdot)$ for each sector, identifying points of these distributions using indirect inference and matching the endogenous firm responses to the policy. Concretely, we choose $n_{\mathcal{T}}$ points of the cumulative distribution function $(t_{a,i}^k, \varrho_i^k)$ where $\varrho_i^k = Pr\{\tilde{t}_a \leq t_{a,i}^k\}$, and define the full distribution $\mathcal{T}^k(\cdot)$ as the linear interpolation anchored on these points.²¹

Next, we parameterize firms' production function $f^k(l, l_a; z, t_a)$ as a Cobb-Douglas function featuring managerial ability and labor input, allowing its parameters to vary across sectors $k \in \{u, s\}$. We assume that labor input combines labor from workers and apprentices in a linear function.

$$f^k(l, l_a; z, t_a^k) = z^{1-\gamma^k} (l + l_a)^{\gamma^k} = z^{1-\gamma^k} ((n - t_a^k n_a) + \zeta_a^k n_a)^{\gamma^k}.$$

Linearity implies that apprentices are substitutes for workers in their sector of training. ζ_a^k denotes marginal productivity of a trained apprentice relative to workers in sector k . It is immediate to see that t_a^k captures the training costs in terms of workers' time to make apprentices ζ_a^k times as productive as workers.

²¹Appendix D.4 shows alternative specifications using a calibrated truncated normal or uniform distribution. These yield a similar but somewhat worse fit.

Analyzing this linear case not only simplifies the solution, but also reflects two key properties of the data. First, it allows for the possibility of some firms choosing the nominally more expensive fee instead of training the required apprentices. As discussed in Section 4.1, a necessary condition for this to hold is that apprentices must have negative marginal productivity for some firms. The production function with linear labor input features apprentices with negative marginal productivity whenever training costs are larger than the productivity of trained apprentices, i.e. $(\zeta_a^k - t_a^k) < 0$. Second, as we show in Table 2, most firms choose corner solutions: only very few choose apprentices between the minimum and the maximum quota, and when they pay the fee they choose zero apprentices. This supports the view that the linear labor input function is a good approximation of firm behavior.

Finally, we assume that managerial ability z follows a three-parameter Generalized Extreme Value (GEV) distribution, $z \sim GEV(\lambda^k, \theta^k, \xi^k)$,

$$\mathcal{Z}^k(z) = e^{-\left(1 + \xi^k \left(\frac{z - \lambda^k}{\theta^k}\right)^{-1/\xi^k}\right)},$$

where $\lambda^k \geq 0$ denotes the location parameter of the distribution, $\theta^k > 0$ the scale parameter and $\xi^k > 0$ the shape parameter.²²

Estimation and Identification

Using these functional forms we simulate the model for $n_{sim} = 100,000$ firms. We follow a three-step procedure to match key moments in the data and identify the structural parameters of the model. First, we estimate the production function using pre-reform data. Second, we use maximum likelihood to estimate the parameters of the productivity distribution using pre-reform data. Third, we use indirect inference and the firm responses to estimate the training cost distribution. For each sector $k \in \{u, s\}$, we estimate four parameters, namely the output elasticity of labor γ^k and the managerial ability distribution parameters $\{\lambda^k, \theta^k, \xi^k\}$, as well as $n_{\mathcal{T}} = 13$ points $(t_{a,i}^k, \varrho_i^k)$ of the training cost distribution.

As a first step, we estimate sector-specific production functions using the pre-reform data from 1995 to 2002. We use those firms that do not have apprentices before the reform to estimate the output elasticity of labor γ^k . Since labor is the only production input in the model, we estimate this elasticity by regressing log output of each firm on log number of full-time workers, controlling for time and firm fixed effects. The estimated elasticity would be upward-biased if other production inputs are gross complements of labor. In Appendix D.2, we allow for other inputs, following the procedure of Levinsohn and Petrin (2003) with the Akerberg et al. (2015) correction to estimate output elasticities.²³ Results using these estimated elasticities are quantitatively similar to the baseline specification.

²²Appendix D.1 shows that this distribution provides the best fit of the pre-reform firm size distribution among two and three-parameter distributions typically used to model productivity.

²³We follow the code *prodest* developed by Mollisi and Rovigatti (2017) to estimate the production function using control function methods. The appendix compares results from different production function estimation methods including the one developed by Olley and Pakes (1996), Levinsohn and Petrin (2003) and Wooldridge (2009).

Next, we match the pre-reform firm size distribution using a maximum likelihood estimation procedure. The production function implies that the optimal number of workers without regulation is linear in managerial ability, $n^*(z, t_a) = a^k z$, such that the firm size distribution is also $GEV(a^k \lambda^k, a^k \theta^k, \xi^k)$. To obtain the productivity parameters we target the size distribution of firms that do not train apprentices before the reform. In this case, $a^k = \left(\frac{\gamma^k}{w^k}\right)^{1/(1-\gamma^k)}$, and we can fit this distribution using maximum likelihood. Using the estimated parameters for the firm-size distribution λ_n^k , θ_n^k and ξ^k , we compute the productivity distribution parameters $\lambda^k = \frac{1}{a^k} \lambda_n^k$ and $\theta^k = \frac{1}{a^k} \theta_n^k$, using the estimated γ^k and the observed average pre-reform wages of workers w^k in each sector.

Finally, we use indirect inference to estimate the training cost distribution \mathcal{T}^k . We target three key groups of firm responses in each sector that correspond to the empirical facts from the reduced-form analysis. The estimation is overidentified as we can use 43 moments in each sector: 10 bunching mass points (one for each of the first 10 thresholds), 30 missing mass points (3 bins above/below each of the first 10 thresholds), the percentage of firms that choose the maximum number of apprentices before and after the policy, and the percentage of firms that pay the fee. We choose the values for $(t_{a,i}^k, \varrho_i^k)_{i=1}^{n_{\mathcal{T}}}$ minimizing the weighted sum $\mathcal{L}((t_{a,i}^k, \varrho_i^k)_{i=1}^{n_{\mathcal{T}}})$ of the absolute difference between the model-implied moments and the data moments,

$$\min_{(t_{a,i}^k, \varrho_i^k)_{i=1}^{n_{\mathcal{T}}}} \mathcal{L}((t_{a,i}^k, \varrho_i^k)_{i=1}^{n_{\mathcal{T}}}) := \min_{(t_{a,i}^k, \varrho_i^k)_{i=1}^{n_{\mathcal{T}}}} \sum_{j=1}^{43} \omega_j^k \frac{|model^k(j) - data^k(j)|}{\frac{1}{2}|model^k(j)| + \frac{1}{2}|data^k(j)|},$$

where ω_i^k is the weight assigned to moment i . We choose these weights to give equal importance to each of the four groups of moments.²⁴

We use the regulation specifications for the remaining parameters. $\{N_j, \underline{n}_a^j, \bar{n}_a^j\}_j$ denote the apprentice thresholds $\{N_j\}_j = \{15, 30, 50, \dots, 20(j-1) + 10, \dots\}$ and quotas establishing a minimum \underline{n}_a^j and a maximum number of apprentices \bar{n}_a^j . w_a^{min} is the minimum wage for apprentices which we set to 75% of the minimum wage, $w_a = 0.75w_{min}$, and $\phi_a = w_{min}$ is the fee parameter that is proportional to the difference between required and actual apprentices.²⁵

Intuitively, the observed firm responses identify points of the training cost distribution. For firms with low enough training costs $\tilde{t}_a^k < \zeta_a^k - \frac{w_a}{w^k}$, it is optimal to train as many apprentices as possible. Hence, the fraction of firms choosing the maximum number corresponds to the point in the training cost distribution, $Pr\{t_a \leq \zeta_a^k - \frac{w_a}{w^k}\} = \mathcal{T}\left(\zeta_a^k - \frac{w_a}{w^k}\right)$. We match this before and after the reform. Conversely, if training costs are large, the percentage of firms paying the fee identifies $Pr\{t_a \geq \zeta_a^k + \frac{\phi_a - w_a}{w^k}\} = \mathcal{T}\left(\zeta_a^k + \frac{\phi_a - w_a}{w^k}\right)$. The minimization algorithm first chooses these three points of the CDF, matching the three probabilities. We determine the remaining $n_{\mathcal{T}} - 3$ points choosing percentiles and estimating the $t_{a,i}^k$ that minimizes the objective function. We choose ten additional points, $n_{\mathcal{T}} - 3 = 10$, and linearly interpolate the distribution to compute the moments in each iteration.²⁶ These points of the CDF identify the training cost distribution.

²⁴See Appendix D.3 for the precise definition of weights.

²⁵75% of the minimum wage is the apprentices' wage during the productive phase, which corresponds more closely to the model.

²⁶We use a *pattern search* algorithm which is useful for solving non-smooth optimization problems with simulations.

It is important to note that we cannot identify ζ_a separately from t_a based on our data. To anchor the training cost distribution, we normalize $\zeta_a = 1$ in both sectors. Therefore, our estimates can be interpreted as *net training cost* that encompasses both the costs of workers having to train the apprentice and the fact that apprentices can be less productive than workers during the time they spend in production. Such net training costs are sufficient to characterize firms’ production and training decision for the quantitative results and policy counterfactuals presented later. The scale of net training cost estimates is in terms of workers’ time equivalent.

Table 3 summarizes the estimated parameters, with bootstrapped confidence 95% intervals. With 0.61 and 0.58, high-skill and low-skill sectors have similar labor shares. The managerial ability distribution has a larger location λ and scale parameter θ in high-skill sectors, but a smaller shape parameter ξ . This implies that high-skill sector firms have higher average managerial ability.²⁷

Table 3: Estimated Parameters

Parameter	Description	High-Skill	Low-Skill
γ^k	Labor share in output.	0.61 (0.58, 0.63)	0.58 (0.55, 0.61)
λ^k	Location parameter of productivity distribution $\mathcal{Z}(z)$.	1916 (1427, 2559)	939 (720, 1200)
θ^k	Scale parameter of productivity distribution $\mathcal{Z}(z)$.	1978 (1475, 2655)	1014 (776, 1300)
ξ^k	Shape parameter of productivity distribution $\mathcal{Z}(z)$.	0.85 (0.83, 0.88)	0.88 (0.85, 0.9)

Notes: Confidence intervals at 95% level in parenthesis, computed using 1000 bootstrap samples.

Supply Side

To close the model and compute the effect of the policy on aggregate variables, we calibrate the supply side of the economy. We use data on the number of firms in each sector before the reform and the estimated firm size distribution to get the initial supply of workers in each sector L^k . There are a similar average number of firms in each sector in the pre-reform period, 3,496 firms in high-skill sectors and 3,362 in low-skill sectors. We normalize the minimum wage before the reform to $w_{min} = 1$, such that all nominal variables are in units of the pre-reform minimum wage. Using data from the firm census, we compute the average wages of workers in each sector in 2002, the last year before the reform. Workers in high-skill sectors earn on average 3.95 times the minimum wage, whilst workers in low-skill sectors earn 3.25 times the minimum wage. This difference in wages is important for identifying the training cost distribution parameters as well as for the quantitative exercises. Finally, we assume that the number of potential apprentices L_a is larger than the aggregate demand for apprentices. Such excess supply of apprentices is supported by enrollment figures indicating that

The additional $n_{\mathcal{T}} - 3$ percentiles ϱ_i are chosen as **Chevyshev nodes** around the anchored points. Spreading them uniformly yield similar results.

²⁷The mean of a $GEV(\lambda, \theta, \xi)$ distribution is $\lambda + (1 - \Gamma(1 - \xi))^{\frac{\theta}{\xi}}$.

only half of applicants enrolled in vocational institutions are able to find an apprenticeship (SENA, 2018).

5.2 Goodness of Fit

Table 4 reports the targeted moments in the data and the corresponding values we obtain from the estimated model. In cases where we match a full function a reference is provided to the relevant figure showing the fit. Overall, the estimated model closely resembles the data. We match, almost exactly, the fraction of firms that choose the maximum quota and that pay the fee by sector. Moreover, Figure 6 shows the targeted distribution of full-time workers by sector in the pre-reform period. The distributions are smooth around the thresholds and look similar to the data in both sectors. The GEV distribution approximates the firm-size distributions well, only slightly underestimating the mass of firms between 5 and 10 full-time workers and slightly overestimating the mass of firms with workers between 15 and 40.

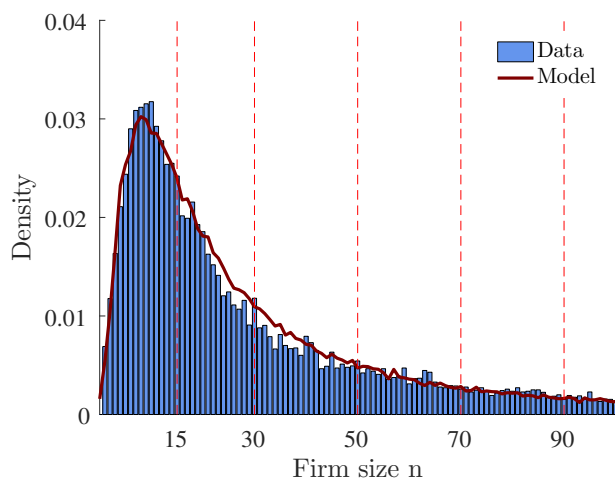
Table 4: Targeted Moments

Moment Description	Model	Data
<i>High-Skill Sectors</i>		
Pre-reform firm-size distribution	see Fig. 6a	see Fig. 6a
Fraction choosing maximum apprentices pre-reform (≥ 15 workers)	1.5%	1.5%
Fraction choosing maximum apprentices post-reform (≥ 15 workers)	1.5%	1.5%
Fraction paying fee (≥ 15 workers)	58.2%	58.2%
Bunching and missing mass	see Fig. 6c	see Fig. 6c
<i>Low-Skill Sectors</i>		
Pre-reform firm-size distribution	see Fig. 6b	see Fig. 6b
Fraction choosing maximum apprentices pre-reform (≥ 15 workers)	1.5%	1.5%
Fraction choosing maximum apprentices post-reform (≥ 15 workers)	65.4%	65.4%
Fraction paying fee (≥ 15 workers)	0.5%	0.5%
Bunching and missing mass	see Fig. 6d	see Fig. 6d

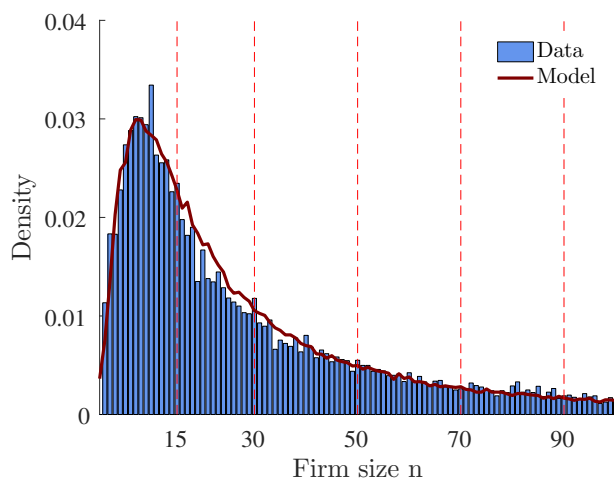
Notes: Value of targeted moments in the data and in simulated model.

Figure 6 further illustrates how the model fits our three empirical facts very well. Panels 6c and 6d show the observed and simulated firm size distribution after the reform, corresponding to empirical fact 1. The model captures the incentives high-skill firms have to bunch just below the threshold, and that of low-skill firms to bunch at the thresholds. Note that the model somewhat overestimates the fraction of firms bunching in high-skill sectors. This happens mainly for two reasons. First, as mentioned above, the number of firms between 15 and 40 full-time workers is slightly overestimated. Second, our simple model abstracts from other margins of substitution that could “smooth” firm

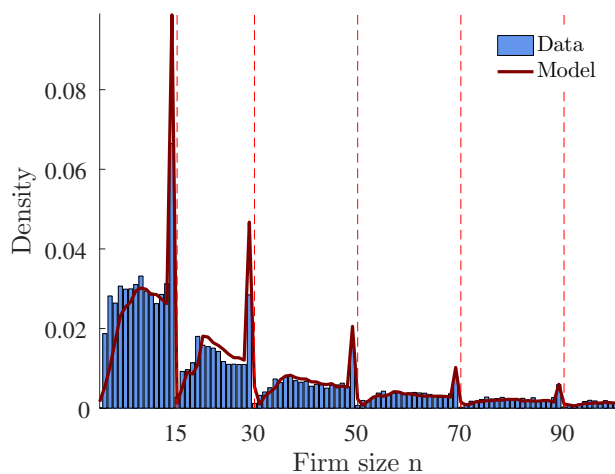
Figure 6: Goodness of Fit



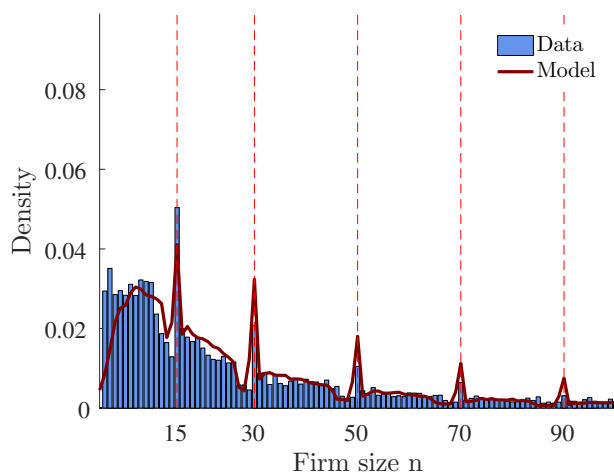
(a) Pre-Reform High-skill Size Distribution



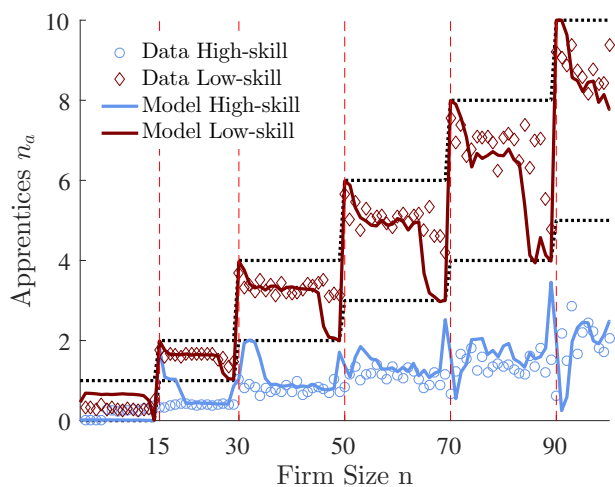
(b) Pre-Reform Low-skill Size Distribution



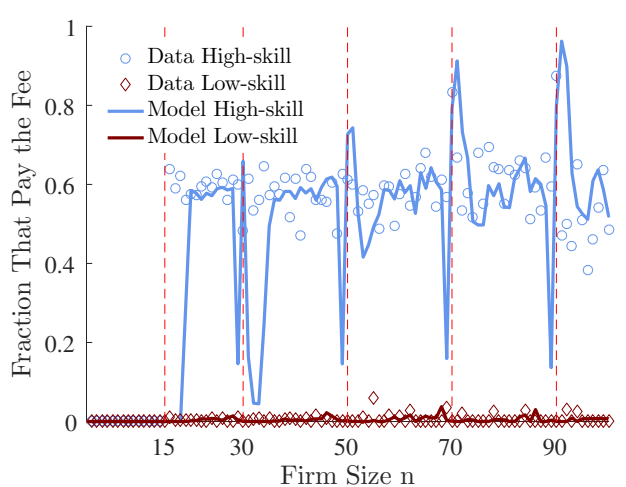
(c) Post-Reform High-skill Size Distribution



(d) Post-Reform Low-skill Size Distribution



(e) Number of Apprentices



(f) Fraction that Pay the Fee

Notes: Panels (a) and (b) show the distribution of full-time workers and model fit by sector for the pre-reform period (1995-2002) and Panels (c) and (d) for the post-reform period (2003-2009). Panel (e) the number of apprentices in each sector by the firm size (full-time workers). Panel (f) the fraction of firms that pay the fee by firm size.

responses. For instance, if other inputs such as capital, material or other types of labor substitute for full-time workers, this may reduce the fraction of firms bunching.

Panel 6e shows the fit of the model to empirical fact 2, the number of apprentices by firm size. High-skill sector firms choose an average number below the minimum quota, while the number is close to the maximum quota for firms in low-skill sectors. The model highlights that within each regulation bracket there is a decreasing pattern for firms in the low-skill sectors. Almost all firms at the regulation thresholds choose the maximum quota of the regulation, while the fraction of firms choosing this maximum decreases in firm size between thresholds. Finally, Panel 6f shows that the simulated model also replicates empirical fact 3, with around 60% of firms in high-skill sectors with more than 15 workers choosing to pay the fee. Meanwhile, as in the data, almost none of the low-skill sector firms pay the fee and the fraction is relatively flat in firm size.

5.3 Training Cost Distribution

As our first quantitative result, Figure 7 shows the estimated net training cost distribution. Training costs are substantial in general. The lower average training costs in low-skill sectors ($\mathbb{E}(t_a^u) = 0.75$) compared to high-skill sectors ($\mathbb{E}(t_a^s) = 1.05$) is consistent with most training happening in low-skilled sectors. The training costs in high-skill sectors are also more dispersed ($Var(t_a^s) = 0.053$) than in low-skill sectors ($Var(t_a^u) = 0.0025$).

Figure 7: Training Cost Distribution and Apprentices' Marginal Productivity

(a) Training Cost Distribution

(b) Apprentices' Marginal Productivity

Notes: Panel (a) shows the estimated non-parametric training cost distribution \mathcal{T}^k . Panel (b) shows the marginal productivity of apprentices for the estimated linear input case. Shaded areas denote the 95% confidence intervals around the estimated distributions computed using bootstrapping.

An interesting additional interpretation of training costs stems from transforming them into apprentices' marginal productivity. As mentioned before, we are not able to separately identify t_a and ζ_a . However, our model identifies the marginal productivity of apprentices $\zeta_a - t_a$ shown in Panel 7b. The units tell us how productive an untrained apprentice is relative to a regular worker. For instance, $\zeta_a - t_a = 0.25$ suggests that the marginal productivity of an apprentice during training is

equal to 25% of the marginal productivity of the average worker in low-skilled sectors. Moreover, the figure shows that apprentices have negative marginal productivity in around 60% of high-skill sector firms, while apprentices’ productivity is positive in most low-skill sector firms. As discussed in Section 4.1, this ties in well with the empirical observation that many firms in high-skill sectors choose to pay fees.

These training cost distributions are crucial for firms’ willingness to train apprentices, a key limiting factor for the success of apprenticeship programs (Groh et al., 2016, Alfonsi et al., 2020). Policies aimed at increasing apprenticeship training must offer sufficient incentives to overcome the high training costs revealed by many firms. Moreover, our results highlight that an important source of training cost heterogeneity stems from sectoral differences. As we show in our counterfactual exercises below, policies that consider this heterogeneity explicitly can yield further benefits compared to uniform policies.

5.4 Quantitative Effects of the Apprenticeship Regulation

Partial Equilibrium, General Equilibrium and Dynamic Effects

We quantify the effects of the apprenticeship regulation under three scenarios: (i) the short-term or *partial equilibrium* effect where wages are constant, (ii) the *general equilibrium* effect where displaced workers are absorbed back into the labor market, and (iii) the *dynamic* effect where trained apprentices increase the future supply of workers.

Wages are key to quantifying the effects of the policy on aggregate outcomes and welfare. Table 5 shows wages (relative to the minimum wage) by sector in each scenario. In partial equilibrium, wages are fixed to the pre-reform levels and calibrated directly from the data. The policy induces firms to substitute from hiring workers to training apprentices. In general equilibrium, labor markets absorb the excess supply of workers, thereby lowering their wages. The response is stronger among firms in low-skill sectors, where wages fall by 1.2%. When trained apprentices become workers in the next period, there is a further decrease in wages for both sectors. Since low-skill sectors train more apprentices, wages fall by more (6.2%) than in high-skill sectors (2%).

Table 5: Wages

	Partial Equilibrium	General Equilibrium	Dynamic Effects
High-skill Sectors	3.95	3.93	3.90
Low-skill Sectors	3.25	3.22	3.11

Notes: All wages are in units relative to the minimum wage. Partial equilibrium wages based on EAM firm census data in the last year before the reform (2002). General equilibrium and dynamic effects wages come from the estimated model.

Aggregate Effects

We begin by quantifying the effects of the policy on aggregate outcomes. In particular, the change in aggregate output reflects the change in overall efficiency, and changes in the total number of workers and apprentices are informative of the mechanisms that affect output.

Table 6 shows the effects under each scenario by sector. First, we document that despite the large labor input responses, the partial equilibrium effects on aggregate output are relatively small. Panel A shows that firms respond to the policy by increasing the number of apprentices, which displaces some workers. In total, around one worker is displaced for each five apprentices. Moreover, the first two columns show a percentage of displaced workers of approximately 1.2%, with more workers laid off in the low-skilled sector. Despite the sizeable increase in training, there are relatively small effects on output. The overall effect is negative but close to zero. However, there is some reallocation of production across sectors: High-skill sector firms decrease production by 0.39% and low-skill sector firms increase production by 0.36%. In Section 5.5, we show additional reduced-form evidence consistent with these partial equilibrium effects of the regulation, and discuss the relationship to existing literature.

In contrast, Panels B and C show that general equilibrium and dynamic effects can lead to a substantial increase in aggregate output. In general equilibrium, wages fall and displaced workers are absorbed back into the labor market, increasing production. Aggregate output increases by 0.7%, there is higher growth in low-skill sectors and the initial output losses are reverted in high-skill sectors. The positive effects on output can be even stronger when apprentices become workers in the dynamic scenario, resulting in an overall increase in output of around 3.6%. While output in high-skill sectors also increases in Panels B and C, it is worth noting that some reallocation of production towards low-skill sectors persists as they see larger output growth under all scenarios.

Winners and Losers

The policy also has a distributional impact, as the effects on the welfare of apprentices, workers and firm owners/managers vary. There are winners and losers from the regulation, both across these groups of agents and across sectors. Given the data and model, the effects on firms can be readily seen. To quantify changes in the welfare of apprentices and workers, we assume that all agents have the same homothetic preferences, such that their utility is linear in income. For apprentices, we normalize their outside-option income to zero, and to compute dynamic effects of the policy we assume that apprentices acquire the skills necessary to replace an average worker in their sector of training.²⁸ For workers, we track the welfare of those who are working before the reform. In the partial equilibrium scenario, we suppose that the displaced workers earn no wage when they become unemployed. Taken together, these assumptions probably imply that we obtain an upper bound on

²⁸The outside option is important to determine the gains for apprentices. The alternatives could be unemployment or informality, both of which have low value. In Colombia, there is no unemployment insurance over the sample period. The average informal wage is below the minimum wage and more than 40% of informal workers earn less than half the minimum wage (Bernal, 2009). Moreover, the excess supply of potential apprentices is suggestive of the benefits they receive from apprenticeships (SENA, 2018).

Table 6: Quantitative Effects of the Apprentice Regulation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	<i>Aggregate Outcomes</i>				<i>Changes in Agents' Welfare $\Delta U_j/U^*$</i>			
	Workers	% Workers	Apprentices	% Output	Apprentices	Workers	Firms	Total
A. Partial Equilibrium								
High-skill Sectors	-1964	-0.92	4527	-0.39	0.25	-0.56	-0.45	-0.77
Low-skill Sectors	-3304	-1.57	17622	0.36	1.12	-0.91	0.14	0.35
Total	-5268	-1.24	22149	-0.06	0.64	-0.72	-0.19	-0.27
B. General Equilibrium								
High-skill Sectors	0	0.00	4557	0.18	0.25	-0.22	-0.24	-0.21
Low-skill Sectors	0	0.00	17898	1.28	1.14	-0.40	0.54	1.28
Total	0	0.00	22454	0.67	0.64	-0.30	0.11	0.45
C. Dynamic Effects ($t + 1$)								
High-skill Sectors	4527	2.16	4642	1.50	1.29	-0.70	0.24	0.83
Low-skill Sectors	17622	8.49	19207	6.17	4.65	-2.36	2.58	4.87
Total	22149	5.31	23849	3.57	2.78	-1.44	1.28	2.62

Notes: The table shows the effects of the policy on aggregate outcomes in columns (1) to (4), namely on the number of workers, the number of apprentices and output. Columns (5) to (7) show the effect on the welfare of apprentices, workers and firm owners. Column (8) shows the sum of welfare effects across the groups of agents from columns (5) to (7).

the welfare consequences for apprentices and workers. For firms we assume that there is no entry or exit, so the change in utility comes solely from the change in profits.

Table 6 shows the change in welfare for each group of agents in each scenario. We compute the change in aggregate utility for each group of agents by sector, comparing the respective scenario to a situation without regulation. The change is divided by the sum of utilities across all these agents before the reform, such that the sum of Columns (5) to (7) is equal to the percentage change in total welfare in Column (8). Note that the sum of agents' welfare changes in Column (8) tends to be slightly smaller than the output effect in Column (4) because output also includes government revenue from fee payments.

In Column (5), apprentices gain from the regulation under all scenarios. In the short-run, their welfare improves as the wages they earn as apprentices are higher than their outside option. In the dynamic scenario, these effects are magnified as their wages increase to those of regular workers after training. Since low-skill sectors train more apprentices, the positive effects are larger there, although wages in high-skill sectors are higher. In contrast, Column (6) shows that incumbent workers are harmed by the regulation. In partial equilibrium, this occurs because some workers are displaced by apprentices. In general equilibrium and dynamically, they are absorbed by the labor market again, but wages are lower. Workers in low-skill sectors are more affected, as more of the incumbent workers lose their jobs and their wages fall more sharply. In the general equilibrium scenario, the negative effects on workers are somewhat attenuated relative to partial equilibrium, but the dynamic drop in wages implies more significant negative effects on incumbent workers, particularly in low-skill sectors.

Firms are unequally affected by the policy. In partial equilibrium, the welfare of firm owners

falls in high-skill sectors, while it increases in low-skill sectors. In general equilibrium, high-skill sector firms experience smaller negative effects and the gains of low-skill sectors are reinforced, as firms benefit from higher overall labor supply at lower wages. In all sectors, these gains are further magnified when apprentices increase the supply of productive workers in period $t + 1$. In addition, there is some heterogeneity in the impact of firms within sectors due to heterogeneous training costs. Appendix Figure A.5 shows the distribution of the change in firm profits by sector. As in Table 6, low-skill sector firms gain on average, but high-skill sector firms lose in partial equilibrium. In Panel (a), the effects on high-skill sectors firms vary between profit reductions of 2% and zero. In Panel (b), the general equilibrium boost in profits is not enough to revert their losses and even in the dynamic scenario some firms still lose. In low-skill sectors, some firms see profit reductions in partial equilibrium, but virtually all firms benefit in general equilibrium and in the dynamic scenario.

5.5 Reduced-Form Evidence on Firm Outcomes

To check the empirical plausibility of these model-based quantitative findings, we next present some additional reduced-form evidence of the effect of the regulation on firm outcomes. In particular, the quantitative exercises suggest that despite substantial labor input adjustments, there are small effects on output in partial equilibrium. To estimate the reduced-form causal effect of the regulation on these outcomes, we use a difference-in-difference approach around the regulation thresholds. Overall, the reduced-form results are well in line with the quantitative findings: there are sizeable effects on the number of apprentices and workers, but small, insignificant effects on output.

We estimate the following difference-in-difference specification:

$$Y_{it} = \alpha_i + \delta_t + \beta Above_i \cdot Post_t + \epsilon_{it} \quad (4)$$

where $Above_i$ is an indicator for firms whose size is above the threshold in the last year pre-reform year 2002, α_i is a firm fixed effect, δ_t is a year fixed effect, $Post_t$ is an indicator for the post-reform years from 2003 onwards and ϵ_{it} is an error term. The specification compares firms on the two sides of the thresholds, where they are subject to different apprentice quotas. As discussed in Section 3.2, firms' actual post-reform firm size is subject to endogenous changes, such that observed differences in outcomes across brackets reflect a mixture of the causal effect of the regulation and selection. Hence, we assign the treatment $Above_i$ based on *pre-reform* firm size, when the distribution is still smooth. We include firms within five size bins above and below the regulation thresholds, pooling across the first ten thresholds. Moreover, the sample is restricted to firms that stay within two adjacent regulation brackets across years.²⁹ Column (4) of Table 1 shows summary statistics of this "threshold sample". Since only firms around the first ten thresholds are included, the main difference to the full sample is the smaller average firm size of around 30 workers (49 in the full sample), but other characteristics not related to size tend to be similar.

Table 7 shows results from the regression for four outcomes, the number of apprentices, the

²⁹More precisely, we exclude firms with a yearly change in full-time workers of more than 50. We also exclude firms with more than 300 workers in 2002.

number of full-time workers, log output and the profit rate (profits over revenue). The first two columns show a strong “first stage”, where firms above the thresholds significantly increase the number of apprentices. As expected, the causal effect of the policy on the number of apprentices is stronger in low-skill sectors, with an increase of 0.8 compared to 0.16 in high-skill sectors. Next, the results indicate that firms training more apprentices as a result of the regulation reduce the number of regular workers. Again, the effects are larger in low-skill sectors, where the point estimate is -1.8. In high-skill sectors, there is a smaller and insignificant effect of -0.86. Finally, the effect on log output and profit rates are small and insignificant in all sectors. Appendix Table A.4 shows the estimated effects on additional outcomes. In Panel A, the effect on the number of workers is largely driven by reductions in administrative and production workers. Panel B shows the effect on the number of workers reported in the survey data, where the estimates for both sectors are remarkably similar to those from the main administrative data. Moreover, there are no significant effects on temporary workers in the firm, nor on any other production inputs such as capital and intermediate inputs.

Table 7: Reduced-Form Effect of the Regulation on Firm Outcomes

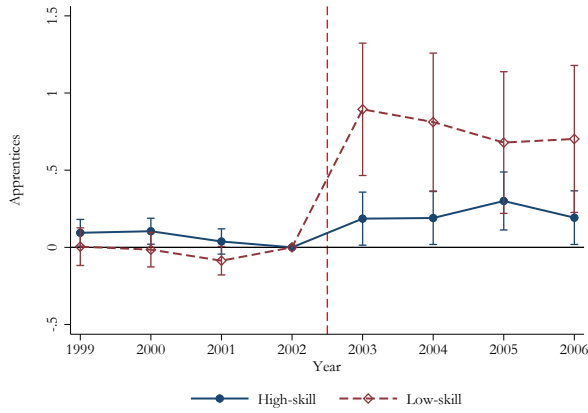
Sector	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Apprentices		Workers		Log Output		Profit Rate Π/Y	
	High	Low	High	Low	High	Low	High	Low
Above*Post	0.159*** (0.0576)	0.802*** (0.224)	-0.855 (0.863)	-1.781* (1.036)	-0.0302 (0.0355)	-0.00697 (0.0472)	-0.0135 (0.0113)	-0.00174 (0.0107)
Mean DV (Pre-reform)	0.161	0.135	30.46	30.30	14.31	14.46	0.212	0.193
Observations	8491	6357	8491	6357	8491	6357	8491	6357
R-squared	0.279	0.600	0.904	0.894	0.866	0.862	0.527	0.488

Notes: All regressions include year and firm FE. Standard errors clustered at the firm level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

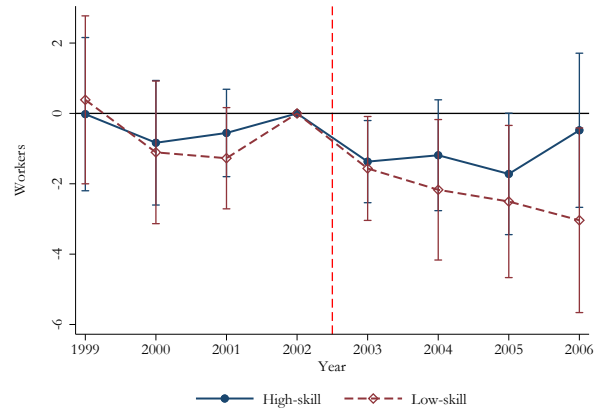
In addition, Figure 8 plots yearly difference-in-difference coefficients around the reform, comparing treated firms (above the thresholds) to control firms (below the thresholds) separately in high-skill and low-skill sectors. In both high-skill and low-skill sectors, there are no significant pre-trends for the number of apprentices, the number of workers, output and profits. Coefficients on the number of apprentices become positive exactly at the time of the reform and stay relatively constant throughout the post-reform period in both sectors. Similarly, there are clearly negative effects on the number of full-time workers from the time of the reform. The effects on output and profits remain insignificant, although the point estimates indicate potential small negative effects, in particular on profits of high-skill sector firms.

Finally, Appendix Figure A.6 shows a comparison of the reduced-form results from Table 7 to analogous specifications from the model. The difference-in-difference results can arguably be interpreted as partial equilibrium effects, as the specification relies on a short-run, within-sector comparison of firms with different apprentice quotas. As predicted by the model, the effect on the

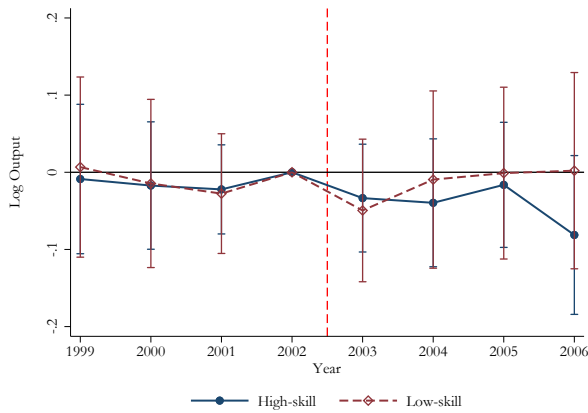
Figure 8: Reduced-Form Effects of the Regulation on Firm Outcomes



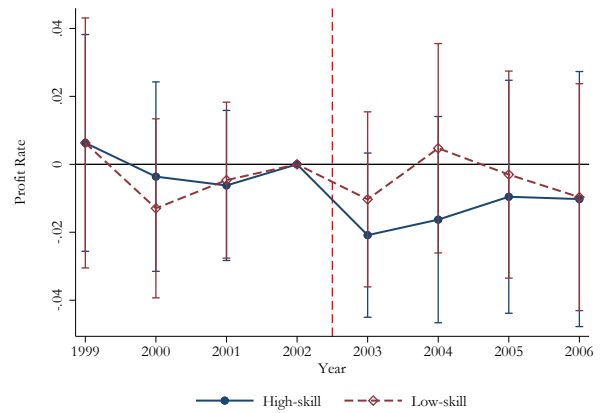
(a) Number of Apprentices



(b) Full-Time Workers



(c) Log Output



(d) Profit Rate (Profits/Revenue)

Notes: Yearly difference-in-difference coefficients for the period 1999 to 2006 using 2002 as base year. The vertical bars show 95% confidence intervals based on standard errors clustered at the firm level.

number of apprentices is stronger in low-skill sectors, although the model somewhat over-predicts the apprentice intake in high-skill sectors. The effect on the number of workers is remarkably similar in model and data, with almost perfectly overlapping confidence intervals. Similarly, the partial-equilibrium effect on output is small and negative both in the model and the data. Profit rates are negatively affected both in model and data, but the model seems to overestimate these negative effects. Overall, these results suggest that the quantitative model not only closely replicates the empirical facts, but also produces predictions consistent with the local effects on firm outcomes around the thresholds.

These additional results are closely related to recent studies that find mixed evidence on the effect changes in labor input composition on firm outcomes such as output and profits (e.g. [de Mel et al., 2019](#); [Hardy and McCasland, 2020](#)). An important difference between our setting and these experimental studies is the Colombian apprenticeship regulation “forces” all firms to take apprentices, or pay costly fees otherwise, whereas they tend to offer wage subsidies or help with finding apprentices/workers. This could explain why we find less positive effects on firm outcomes in partial equilibrium. It is also worth noting that they study very small firms, where initial labor constraints may be more likely to play a role, whereas the average firm in our threshold sample has around 30 workers. However, our quantitative results indicate that there can be large overall benefits of the apprenticeship regulation in general equilibrium and when taking into account dynamics where trained apprentices become productive workers later on.

Moreover, existing studies do not tend to find clear displacement effects of training programs on other workers (e.g. [Attanasio et al., 2017](#), [Alfonsi et al., 2020](#)). On the contrary, we find such displacement effects in the short-run, especially in low-skilled sectors. The model in [Section 4](#) naturally predicts such displacement effects, as workers and apprentices are to some extent substitutable in production and firms are incentivized by the reform to take in more apprentices. The difference-in-difference results are in line with this prediction. Note that the reduced-form results capture the total effect of the regulation on the number of workers in the firm, which includes both direct substitution between workers and apprentices and additional changes in the number of workers due to endogenous firm size responses to the size-based apprentice quotas. Hence, the difference-in-difference results may not directly correspond to a pure labor substitution effect. Finally, our quantitative results highlight that displacement effects are likely to dissipate in the longer-run, when wages adjust and workers re-enter the labor market such that the total size of the skilled workforce grows.

6 Counterfactual Simulations

In this section, we present counterfactual simulations based on the estimated model. First, we decompose the overall effect of the regulation into its different components, and second, we simulate alternative apprenticeship policies.

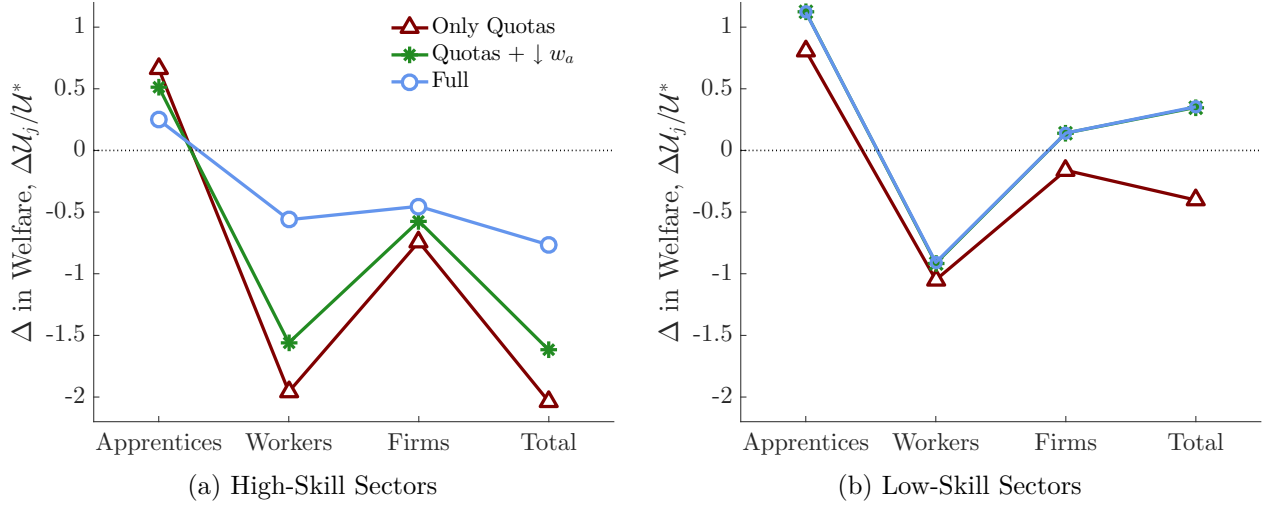
6.1 Decomposing the Effects of the Regulation

The Colombian apprenticeship regulation is a bundle of different components, namely the apprentice quotas, the decrease in the apprentice minimum wage, and the possibility of paying the fee. Our first counterfactual analysis studies the separate effects of each of the components of the policy. Here we present this decomposition in partial equilibrium and study the effects on the welfare of each group of agents. Appendix D.5 shows additional details of the decomposition, in particular regarding the effects on aggregate outcomes. We consider three scenarios: (i) only apprentice quotas, (ii) quotas and the decrease in the apprentices' minimum wage, and (iii) the full regulation.

Figure 9 plots the change in total and group-level welfare under each scenario. As in Table 6 we compute the change in aggregate utility for each group of agents, divided by total utility before the regulation. Panel 9a shows results for high-skill sectors, where the full regulation generates welfare gains for apprentices, but losses for workers and firms. In the scenario with only apprentice quotas, these gains and losses are magnified. When the apprentice minimum wage decreases in addition to the quotas, the overall number of apprentices trained increases significantly, and welfare improves for workers and firms, as firms can take advantage of lower apprentice wages and displace fewer workers. The welfare gain of apprentices is slightly smaller as they are paid less. Finally, when the possibility of paying the fee is added in the full regulation, firms and their workers see smaller losses, since those with the highest training cost now pay the fee instead of training apprentices. However, fewer apprentices are trained, such that the welfare gain of apprentices decreases. Panel 9b shows corresponding results for low-skill sectors. Here, apprentices and firms gain from the full regulation, but workers lose. Compared to quotas alone, the decrease in the apprentice minimum wage is crucial for firms to gain in partial equilibrium. Adding the possibility of paying fees has close to no effect (the respective points in the graph overlap almost perfectly), as low-skill sector firms rarely pay fees. Interestingly, welfare of apprentices increases with the lower minimum wage due to the strongly increased number of apprentices hired by firms. Incumbent workers lose from the regulation in all scenarios.

This decomposition illustrates the role of each of the three main components of the policy. The quotas set tight limits on the number of apprentices a firm has to train. The minimum quota guarantees that many firms train. The maximum quota ensures that firms do not excessively substitute their incumbent workers for apprentices, using apprentices as “cheap labor”. To incentivize firms to train apprentices, the regulation also lowers the apprentice minimum wage. In practice, this benefits mostly firms in low-skill sectors, although it does induce more training in some high-skill sectors firms. Finally, fees partially undo the harm to firms with very high training cost. This reduces the negative impact the regulation can have on firms, but also decreases the positive general equilibrium and dynamic effects as less apprentices are trained in those sectors.

Figure 9: Policy Decomposition



Notes: The figure shows the change in welfare by sector and groups of agents, namely apprentices, incumbent workers and firm owners/managers, relative to the no-regulation total aggregate utility, U^* . *Total* denotes the sum of welfare changes across the three groups.

6.2 Alternative Apprenticeship Policies

Finally, we study two alternative policies, first a subsidy for apprenticeship training, and second, sector-specific minimum wages for apprentices. To make the counterfactual scenarios comparable to the Colombian benchmark regulation, we restrict the number of trained apprentices to be the same across scenarios. Note that both scenarios are price-based policies that avoid firm size distortions as they do not feature size-based apprentice quotas.

Subsidizing Apprenticeship Training

First, we consider a policy that subsidizes apprenticeship training. We focus on budget-balanced policies, where firms are taxed to finance the subsidy. Firms pay payroll taxes at rate τ , and these funds are used to pay a subsidy of $\varsigma\%$ of apprentices' wages. This policy is similar to training subsidies both in developing and developed countries,³⁰ and relates to subsidy interventions analyzed in recent studies (e.g. Crépon and Premand, 2019, Alfonsi et al., 2020).

In this case, a firm (z, t_a) solves

$$\max_{n_a, n \geq 0} p z^{1-\gamma} ((n - t_a n_a) + n_a)^\gamma - w(1 + \tau)n - (1 - \varsigma)w_a n_a \quad \text{s.t.} \quad t_a n_a \leq n. \quad (5)$$

The linear labor inputs again imply corner solutions. Whenever the tax τ or the subsidy ς is large enough, firms have incentives to train more apprentices. These incentives are stronger for firms with low training costs or in sectors where the difference between workers' wages and apprentices'

³⁰For instance, Latin American including Chile and Brazil have implemented training subsidies (see Fazio et al., 2016). Moreover, the UK "Apprenticeship Levy", a subsidy on training costs financed by payroll taxes, is similar to the policy we consider here.

wages is large.³¹ To compare this policy to the benchmark regulation, we compute the tax rate and the subsidy (τ, ς) such that the government’s budget is balanced and the total number of apprentices trained is the same as under the original policy (see Table 6). The solution yields a tax rate of $\tau = 0.11\%$ and a subsidy of $\varsigma = 6.6\%$.

Table 8, Panel A show total tax revenue and subsidy payments by sector in Columns (1) to (3). Firms in high-skill sectors pay more taxes than those in low-skill sectors, but 88% of subsidy payments go to low-skill sector firms, where training costs are lower. Accordingly, Column (6) shows that 95% of apprentices are trained in low-skill sectors. Hence, high-skill sector firms effectively subsidize training in low-skill sectors. Intuitively, if firms in both sectors face the same tax and subsidy rates, it is disproportionately more attractive to firms with low training costs to train apprentices, which benefits low-skill sectors. In fact, the subsidy leads to an extreme concentration of training in firms with the lowest training costs even within low-skill sectors. The percentage of firms with at least one apprentice decreases from around 60% of firms to only 5%. Firms that train often have as many apprentices as possible, hiring only the workers necessary to train these apprentices. Moreover, the effect of the subsidy on total output is slightly more negative than that of the benchmark regulation from Table 6, and the subsidy leads to more displacement of regular workers.

Table 8: Effect of Alternative Apprenticeship Policies

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	<i>Budget Balance</i>			<i>Aggregate Outcomes</i>			
	Tax Revenue	Subsidy Payments	Balance	Workers	% Workers	Apprentices	% Output
A. Training Subsidy							
High-skill Sectors	908	159	749	-730	-0.34	911	-0.15
Low-skill Sectors	718	1467	-749	-6430	-3.05	21239	-0.05
Total	1626	1626	0	-7160	-1.69	22150	-0.11
B. Sector-Specific Apprentice Minimum Wage							
High-skill Sectors	-	-	-	-386	-0.18	4523	0.16
Low-skill Sectors	-	-	-	-4885	-2.31	17621	0.08
Total	-	-	-	-5271	-1.24	22144	0.12

Notes: Columns (1) to (3) present budget balance variables: tax revenues, subsidy payments and net balance. Columns (4) to (7) show (partial equilibrium) effects on aggregate outcomes: the change in the number of workers, the percentage change of workers, the change in the number of trained apprentices and the percentage change in aggregate output. The total number of apprentices in Column (7) exhibits a small difference from Table 8 due to approximation error in the simulation.

This counterfactual highlights the importance of minimum and maximum apprentice quotas when regulation is uniform across all firms, as quotas induce a more even distribution of apprenticeship training across firms both within and across sectors. With a uniform subsidy and no other regulation, training is concentrated only in those firms with the lowest training cost. Note that this does not necessarily imply that subsidizing apprenticeships generally has a negative impact. First, we focus

³¹See Lemma 6 in Appendix D.6 for a full characterization of the solution.

only on budget-balanced subsidies, and second, while the subsidy performs worse than the benchmark regulation, it may still improve general-equilibrium or dynamic welfare compared to no regulation at all.

Sector-Specific Apprenticeship Minimum Wage

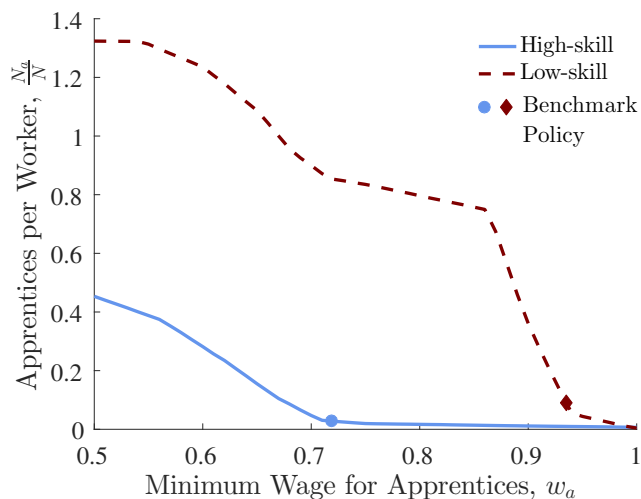
As a second counterfactual, we consider an alternative policy that explicitly takes into account heterogeneity in training costs. Specifically, we allow apprentices' minimum wages to differ across sectors. This policy resembles the situation in some European countries including Germany, where apprentices' wages are governed by sector-specific agreements rather than general restrictions (Steedman, 2012). Again, there are no quotas or fees in the counterfactual. We find sector-specific minimum wages that clear the labor market such that each sector trains as many apprentices as in the benchmark regulation.³² Relative to the minimum wage for regular workers, the apprentice minimum wage is $w_a^{*s} = 0.72$ in high-skill sectors and $w_a^{*u} = 0.94$ in low-skill sectors. Table 8 shows the effects of sector-specific minimum wages set at these levels. Total output increases already in partial equilibrium, which is driven by growth both in high-skill sectors *and* in low-skill sectors. Thus, the policy diminishes the reallocation of resources towards low-skill sectors observed under the benchmark regulation.

This counterfactual is relatively conservative in the sense that it restricts the number of apprentices in each sector to be the same as in the benchmark regulation, where most training occurs in low-skill sectors. In fact, it would be possible to shift this imbalance and increase training in high-skill sectors by further adjusting minimum wages. This is illustrated in Figure 10, which shows the number of apprentices (per worker) by the level of the apprentice minimum wage in each sector. The dots mark the number of apprentices from the benchmark regulation targeted by the main counterfactual. Not surprisingly, low-skill sectors train more apprentices than high-skill sectors at any given level of the minimum wage. Furthermore, the figure implies that a higher number of apprentices in high-skill sectors could be achieved by further reducing the apprentice minimum wage in those sectors, enabling apprentices to compensate firms for the costlier training they receive. For instance, the same apprentice-to-worker ratio as in low-skill sectors under the benchmark regulation could be obtained by setting the high-skill sector apprentice minimum wage to around 0.66.

A potential caveat with setting a lower minimum wage is that apprentices may lack incentives to participate. The simulation does not take into account such supply side considerations. This simplification is guided by three considerations. First, in the Colombian context there is excess supply of apprentices across all industries in the years after the reform (SENA, 2018). Second, the apprentice minimum wages in the main counterfactual are close to or above the actual apprentice minimum wage in Colombia, which is between 0.5 and 0.75 of the minimum wage for regular workers (see Section 2). Third, the minimum wage would have to be lower in high-skill sectors, where apprentices are likely to benefit more from training as future wages are higher. If apprentices value this, they should be willing to accept initially lower wages (Becker, 1964). Nevertheless, when

³²See Appendix D.6 for the analytical details.

Figure 10: Apprentices per Worker



Notes: The figure shows the number of apprentices per worker when varying the apprentice minimum wage in each sector. The dots mark the level of sector-specific minimum wages from the main counterfactual, which target the actual number of apprentices in each sector from the benchmark regulation.

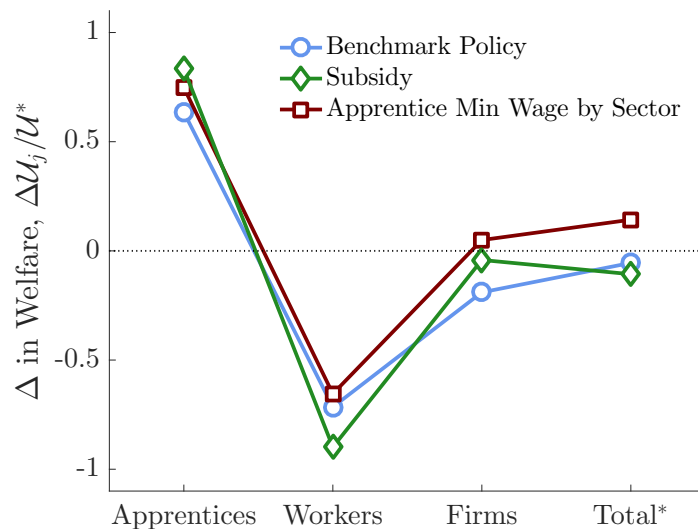
decreasing the minimum wage even more, it may be important to also take into account apprentices' participation responses.

Comparing Apprenticeship Policies

Finally, we compare the effects of the two counterfactual scenarios to the benchmark regulation on the welfare of the different groups of agents in Figure 11. The overall welfare effect of the training subsidy is slightly negative, which is driven by a negative impact on workers and firms, while apprentices gain more strongly from the subsidy than from the benchmark regulation. In terms of total welfare, the subsidy performs slightly worse than the benchmark regulation. Moreover, the figure shows that all groups of agents are better off under the sector-specific minimum wages than under the benchmark regulation. In fact, it is the only policy where firms gain and there are overall welfare gains already in partial equilibrium. Apprentices are also better off and the negative effect on workers is attenuated compared to the benchmark regulation. Overall, we conclude that a policy that takes into account heterogeneity across sectors, such as sector-specific apprentice minimum wages, can magnify the welfare gains of apprenticeship programs.³³

³³Recall that the main counterfactual considered here targets the same number of apprentices as the benchmark regulation *in each sector*. Appendix Figure D.5 shows how aggregate outcomes and welfare would be affected by varying the fraction of apprentices trained in high-skill sectors via further adjustments to apprentice minimum wages, holding the total number of apprentices fixed. The figure suggests that further, albeit modest welfare improvements could be achieved when setting minimum wage levels that increase the share of training of high-skill sectors.

Figure 11: Comparing Apprenticeship Policies: Change in Welfare



Notes: The figure shows the change in welfare by sector and groups of agents, namely apprentices, incumbent workers and firm owners/managers, relative to the no-regulation total aggregate utility, U^* . $Total^*$ denotes the sum of welfare changes across the three groups plus government revenue.

7 Conclusion

There is widespread belief among policymakers that well-designed apprenticeship policies have the potential to improve labor market opportunities for young people. Especially in developing countries, they can help tackle youth unemployment and provide a pathway into higher-productivity, formal employment. Since these training programs tend to yield benefits that are not fully internalized by firms, there is scope for government intervention. Combining reduced-form evidence based on a unique policy change and a structural model, we document important heterogeneity in firm responses to apprenticeship regulation and uncover underlying training costs. We argue that widening the scope of apprenticeship training can have potentially large positive effects on the welfare of apprentices, on firms, and on aggregate production.

These results have important policy implications. First, given the large training costs many firms face, purely price-based interventions such as training subsidies would have to give sizeable incentives to firms to achieve a broad increase in training. The Colombian regulation, featuring a mixture of apprentice quotas and lower minimum wages, does achieve a large increase in the number of apprentices, but it induces firm-size distortions and some reallocation towards low-skill sectors. Second, policies that take into account heterogeneity in training costs, such as sector-specific apprentice minimum wages, can deliver similar benefits from training, while inducing less distortions. These lessons are likely to be relevant for other low- and middle-income countries that suffer from similarly high youth unemployment and informality rates as Colombia did in the early 2000s.

Some limitations of our analysis are worth highlighting. First, the firm-level data provides information on the quantity but not on the quality of apprenticeships. We show some evidence of apprentices' benefits from training and we argue that features of the Colombian regulation commit

firms to standards of training. However, it may be fruitful for future research to consider the quality of training in firms and how it may be related to training costs. Second, for the general equilibrium results, we assume that trained apprentices as well as any displaced workers are eventually absorbed back into the formal labor market at adjusted wages. This assumption is motivated by the notion of a shortage of skilled labor in developing countries, such that firms are willing to hire additional trained workers. If labor market frictions attenuate these general equilibrium effects and some workers remain unemployed or in informal work, this would dampen the positive effects of the apprenticeship regulation. Third, we estimate firms' revealed training costs and emphasize the heterogeneity across sectors, but it remains an important issue how policymakers could identify firms or sectors with high training costs in practice. Exploring these questions further may help strengthen the policy implications of this paper.

References

- Acemoglu, D. and Pischke, J.-S. (1998). Why do Firms Train? Theory and Evidence. *The Quarterly Journal of Economics*, 113(1):79–119.
- Acemoglu, D. and Pischke, J.-S. (1999). The Structure of Wages and Investment in General Training. *Journal of Political Economy*, 107(3):539–572.
- Ackerberg, D. A., Caves, K., and Frazer, G. (2015). Identification Properties of Recent Production Function Estimators. *Econometrica*, 83(6):2411–2451.
- Alfonsi, L., Bandiera, O., Bassi, V., Burgess, R., Rasul, I., Sulaiman, M., and Vitali, A. (2020). Tackling Youth Unemployment: Evidence from a Labor Market Experiment in Uganda. *Econometrica*, Forthcoming.
- Attanasio, O., Guarín, A., Medina, C., and Meghir, C. (2017). Vocational Training for Disadvantaged Youth in Colombia: A Long-Term Follow-Up. *American Economic Journal: Applied Economics*, 9(2):131–143.
- Attanasio, O., Kugler, A., and Meghir, C. (2011). Subsidizing Vocational Training for Disadvantaged Youth in Colombia: Evidence from a Randomized Trial. *American Economic Journal: Applied Economics*, 3(3):188–220.
- Becker, G. S. (1964). Human Capital Theory. *Columbia, New York*, 1964.
- Bernal, R. (2009). The Informal Labor Market in Colombia: Identification and Characterization. *Revista Desarrollo y Sociedad*, (63):145–208.
- Besley, T. and Burgess, R. (2004). Can Labor Regulation Hinder Economic Performance? Evidence from India. *The Quarterly Journal of Economics*, 119(1):91–134.
- Card, D., Ibarra, P., Regalia, F., Rosas-Shady, D., and Soares, Y. (2011). The Labor Market Impacts of Youth Training in the Dominican Republic. *Journal of Labor Economics*, 29(2):267–300.
- Card, D., Kluve, J., and Weber, A. (2018). What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations. *Journal of the European Economic Association*, 16(3):894–931.
- Chetty, R., Friedman, J. N., Olsen, T., and Pistaferri, L. (2011). Adjustment Costs, Firm Responses,

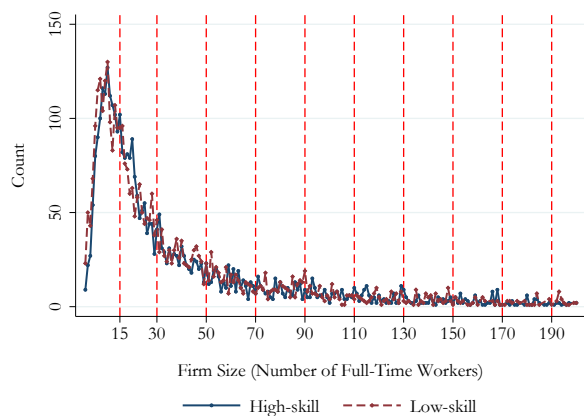
- and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records. *The Quarterly Journal of Economics*, 126(2):749–804.
- Crépon, B. and Premand, P. (2019). Direct and Indirect Effects of Subsidized Dual Apprenticeships. Dabla-Norris, M. E., Jaramillo, L., Lima, F., and Sollaci, A. (2018). *Size Dependent Policies, Informality and Misallocation*. International Monetary Fund.
- de Mel, S., McKenzie, D., and Woodruff, C. (2019). Labor Drops: Experimental Evidence on the Return to Additional Labor in Microenterprises. *American Economic Journal: Applied Economics*, 11(1):202–35.
- Fazio, M. V., Fernández-Coto, R., and Ripani, L. (2016). Apprenticeships for the XXI Century: A Model for Latin America and the Caribbean? *Working Paper*.
- Fersterer, J., Pischke, J.-S., and Winter-Ebmer, R. (2008). Returns to Apprenticeship Training in Austria: Evidence from Failed Firms. *Scandinavian Journal of Economics*, 110(4):733–753.
- Fudenberg, D. and Rayo, L. (2019). Training and Effort Dynamics in Apprenticeship. *American Economic Review*, 109(11):3780–3812.
- 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.
- Garicano, L. and Rayo, L. (2017). Relational Knowledge Transfers. *American Economic Review*, 107(9):2695–2730.
- Garicano, L. and Rossi-Hansberg, E. (2004). Inequality and the Organization of Knowledge. *American Economic Review*, 94(2):197–202.
- Garicano, L. and Rossi-Hansberg, E. (2006). Organization and Inequality in a Knowledge Economy. *The Quarterly Journal of Economics*, 121(4):1383–1435.
- Gaviria, A. and Nuñez, J. A. (2003). Evaluating the Impact of SENA on Earnings and Employment. Technical report, Departamento Nacional de Planeación.
- Göggel, K. and Zwick, T. (2012). Heterogeneous Wage Effects of Apprenticeship Training. *The Scandinavian Journal of Economics*, 114(3):756–779.
- Groh, M., Krishnan, N., McKenzie, D., and Vishwanath, T. (2016). Do Wage Subsidies Provide a Stepping-Stone to Employment for Recent College Graduates? Evidence from a Randomized Experiment in Jordan. *The Review of Economics and Statistics*, 98(3):488–502.
- Guner, N., Ventura, G., and Xu, Y. (2008). Macroeconomic Implications of Size-Dependent Policies. *Review of Economic Dynamics*, 11(4):721–744.
- Hardy, M. and McCasland, J. (2020). Are Small Firms Labor Constrained? Experimental Evidence from Ghana. *Working Paper*.
- Hirshleifer, S., McKenzie, D., Almeida, R., and Ridao-Cano, C. (2016). The Impact of Vocational Training for the Unemployed: Experimental Evidence from Turkey. *The Economic Journal*, 126(597):2115–2146.
- ILO, I. L. O. (2017). ILO Toolkit for Quality Apprenticeships, Volume 1: Guide for Policy Makers.
- Kleven, H. J. (2016). Bunching. *Annual Review of Economics*, 8:435–464.
- Kremer, M. (1993). The O-ring Theory of Economic Development. *The Quarterly Journal of Economics*, 108(3):551–575.

- Krueger, A. B. and Pischke, J.-S. (1995). A Comparative Analysis of East and West German Labor Markets: Before and After Unification. In *Differences and Changes in Wage Structures*, pages 405–446. University of Chicago Press.
- Kuczera, M. (2017). Striking the Right Balance: Costs and Benefits of Apprenticeship. *Working Paper*.
- Kugler, A., Kugler, M., Saavedra, J., and Prada, L. O. H. (2015). Long-Term Educational Consequences of Vocational Training in Colombia: Impacts on Young Trainees and Their Relatives. Technical report, National Bureau of Economic Research.
- Levinsohn, J. and Petrin, A. (2003). Estimating Production Functions using Inputs to Control for Unobservables. *The Review of Economic Studies*, 70(2):317–341.
- McKenzie, D. (2017). How Effective Are Active Labor Market Policies in Developing Countries? A Critical Review of Recent Evidence. Policy Research Working Paper 8011, World Bank Group, Washington, D.C.
- Mohrenweiser, J. and Zwick, T. (2009). Why Do Firms Train Apprentices? The Net Cost Puzzle Reconsidered. *Labour Economics*, 16(6):631–637.
- Mollisi, V. and Rovigatti, G. (2017). Theory and Practice of TFP Estimation: The Control Function Approach using Stata. *Working Paper*.
- Olley, G. S. and Pakes, A. (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, pages 1263–1297.
- Ospino, C. (2018). The Effects of Being Subject to the Colombian Apprenticeship Contract on Manufacturing Firm Performance. *Documentos de Trabajo del CEDLAS*.
- Saez, E. (2010). Do Taxpayers Bunch at Kink Points? *American Economic Journal: Economic Policy*, 2(3):180–212.
- SENA (2018). Colombia Data Report on Vocational Education and Training 2016.
- Steedman, H. (2012). Overview of Apprenticeship Systems and Issues. *ILO*.
- Wolter, S. C., Mühlemann, S., and Schweri, J. (2006). Why Some Firms Train Apprentices and Many Others Do Not. *German Economic Review*, 7(3):249–264.
- Wolter, S. C. and Ryan, P. (2011). Apprenticeship. In *Handbook of the Economics of Education*, volume 3, pages 521–576. Elsevier.
- Wooldridge, J. M. (2009). On Estimating Firm-Level Production Functions using Proxy Variables to Control for Unobservables. *Economics Letters*, 104(3):112–114.

ONLINE APPENDIX TO “UNWILLING TO TRAIN? FIRM RESPONSES TO THE COLOMBIAN APPRENTICESHIP REGULATION”

A Additional Figures and Tables

Figure A.1: Firm Size Distribution Pre-Reform



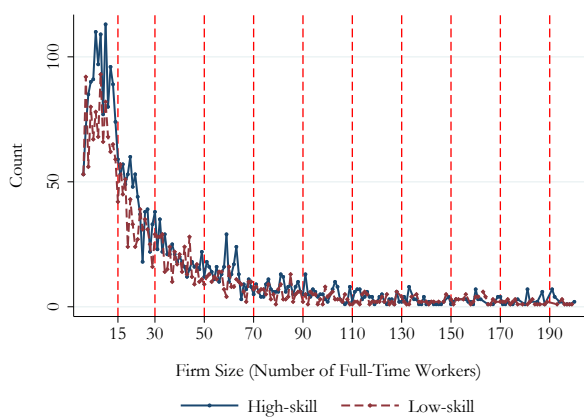
(a) 1996



(b) 1998



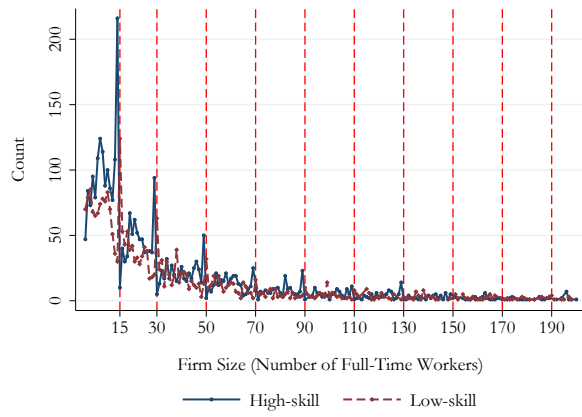
(c) 2000



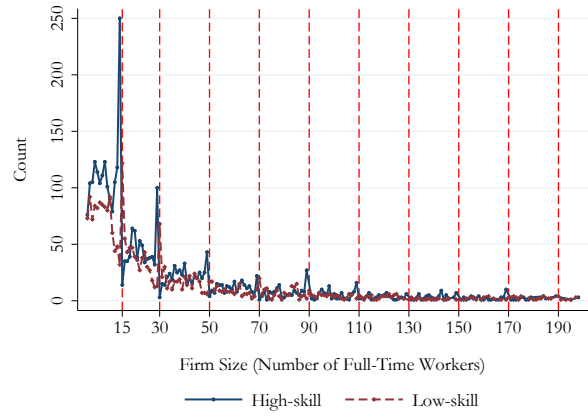
(d) 2002

Notes: Distribution of full-time workers for all firms before in selected years before the reform, using a bin size of one.

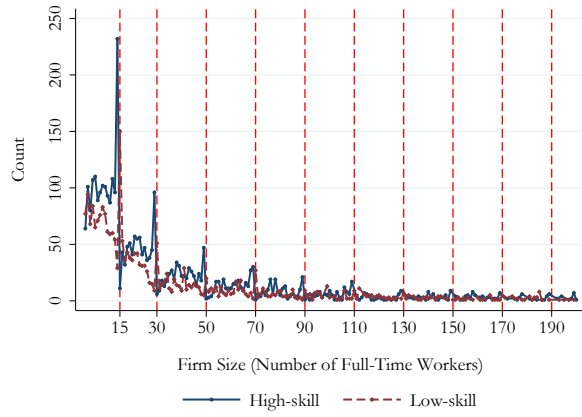
Figure A.2: Firm Size Distribution Post-Reform



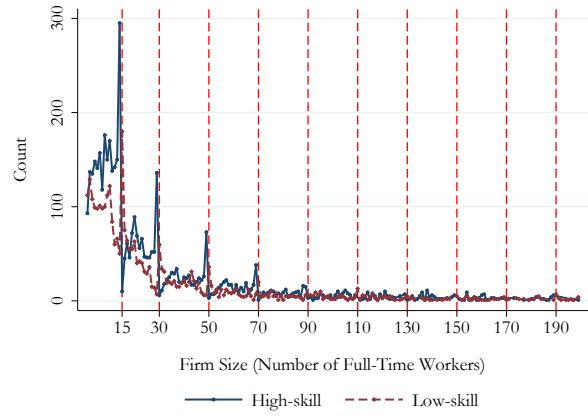
(a) 2003



(b) 2005



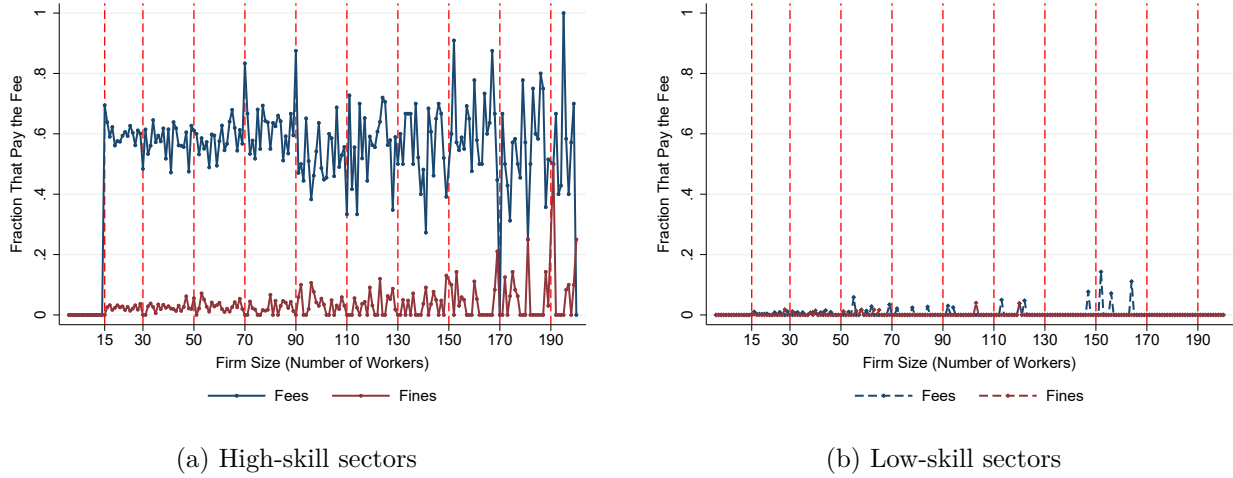
(c) 2007



(d) 2009

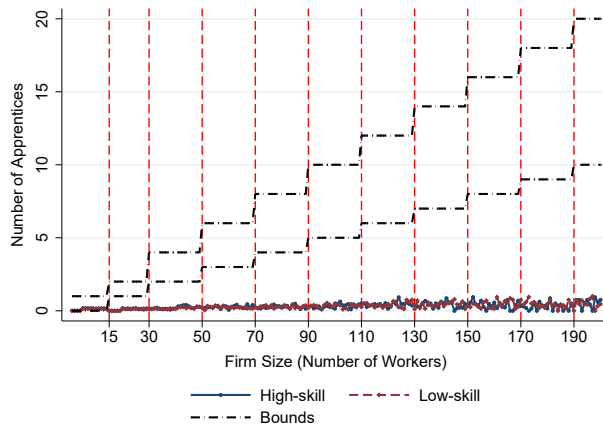
Notes: Distribution of full-time workers for all firms before in selected years after the reform, using a bin size of one.

Figure A.3: Fees and Fines



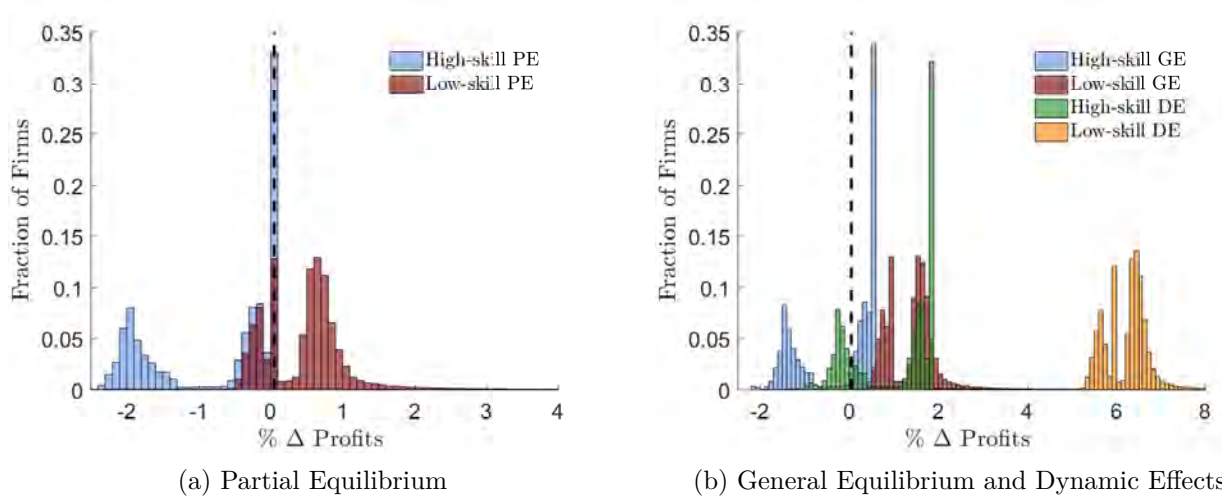
Notes: The figure shows the fraction of firms paying fees and fines in high-skill sectors (Panel a) and low-skill sectors (Panel b) in the post-reform years 2003 to 2009. The red dashed vertical lines denote the regulation thresholds.

Figure A.4: Number of Apprentices by Firm Size, Pre-Reform



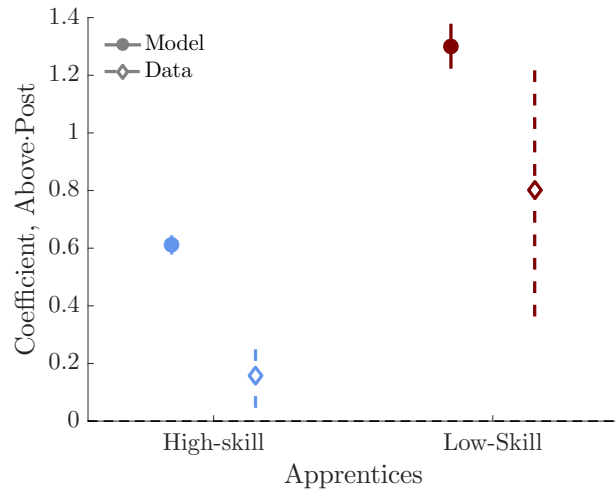
Notes: Average number of apprentices by firm size bin for high-skill and low-skill sector firms, pre-reform (1995-2002). The black dashed lines show the minimum and maximum apprentice quotas, and the red dashed vertical lines denote the regulation thresholds.

Figure A.5: Distribution of Profit Changes

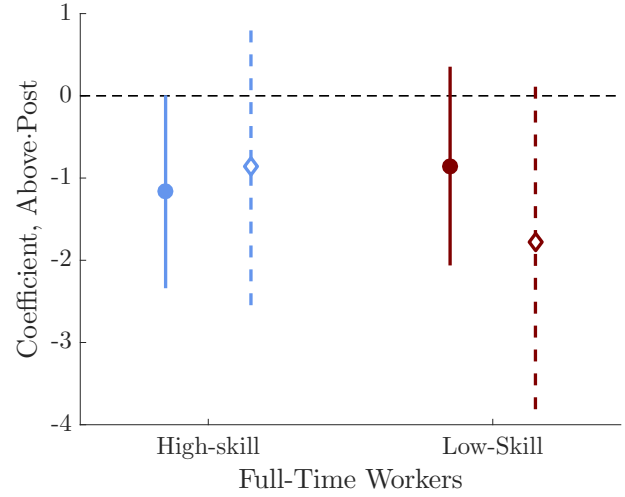


Notes: The figure shows the distribution of the percentage change in firm profits by sector in each scenario. We consider the partial equilibrium (PE), general equilibrium (GE) and dynamic effects (DE) of the policy.

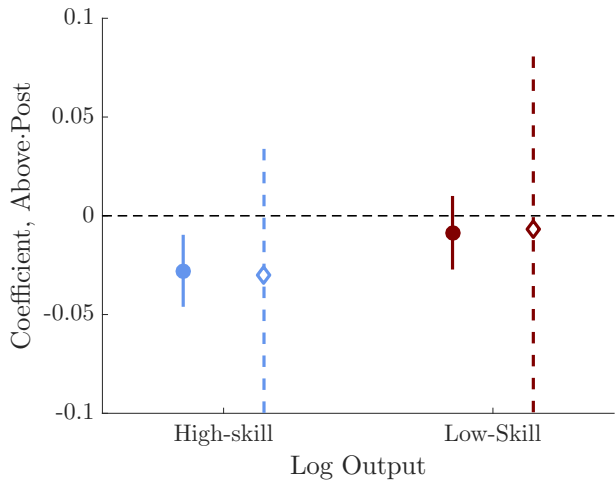
Figure A.6: Reduced-Form Coefficients vs. Model Prediction



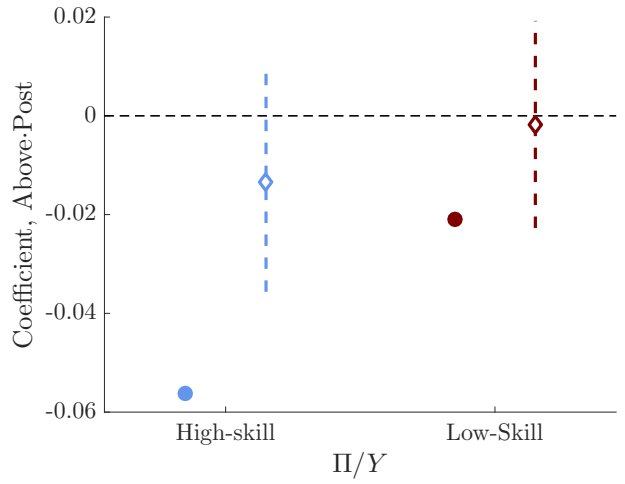
(a) Number of Apprentices



(b) Full-Time Workers



(c) Log of Output



(d) Profit Rate (Profits/Revenue)

Notes: The diamonds denote difference-in-difference coefficient with 95% confidence intervals, and the hollow circles show the corresponding effects based on the estimated model.

Table A.1: Sector-Level Summary Statistics (Pre-Reform)

	(1)	(2)	(3)	(4)	(5)	(6)
	Obs.	Fraction of Professional Workers	Wage per Worker All Workers	Production Workers	Fraction of Permanent Workers	Capital/ Output
<i>Low-Skill Sectors</i>						
Wood	2,543	0.035 (0.08)	11,161 (5,657)	10,210 (4,406)	0.84 (0.25)	0.31 (0.44)
Textiles	12,354	0.049 (0.13)	10,955 (4,938)	10,581 (5,879)	0.73 (0.36)	0.57 (1.31)
Food and Beverages	11,159	0.053 (0.10)	13,846 (8,638)	12,676 (7,538)	0.81 (0.26)	0.58 (1.00)
Mineral Non-Metallic	2,689	0.070 (0.12)	16,398 (11,237)	14,664 (9,087)	0.82 (0.27)	0.63 (1.12)
All	28,745	0.052 (0.11)	12,604 (7,575)	11,815 (7,001)	0.78 (0.31)	0.56 (1.13)
<i>High-Skill Sectors</i>						
Other Manufacturing	2,543	0.073 (0.14)	14,931 (10,876)	14,075 (10,178)	0.77 (0.31)	0.53 (1.06)
Paper and Editorial	12,354	0.074 (0.14)	16,108 (8,704)	14,642 (7,930)	0.87 (0.22)	0.53 (1.07)
Metallic	11,159	0.074 (0.10)	19,494 (11,996)	17,351 (10,181)	0.77 (0.29)	0.71 (1.37)
Machinery	2,689	0.084 (0.14)	15,084 (7,908)	13,046 (6,872)	0.83 (0.26)	0.47 (0.78)
Chemical	5,118	0.086 (0.12)	18,893 (12,667)	15,049 (9,836)	0.85 (0.24)	0.66 (1.02)
All	28,949	0.080 (0.13)	16,392 (10,331)	14,158 (8,704)	0.83 (0.26)	0.55 (0.97)

Notes: The table shows summary statistics by 2-digit sectors, using the CIIU R3 classification. Low-skill and high-skill sectors are defined according to ranking of fraction of professionals as explained in Section 3.1. Wages are in 2009 thousands of pesos. Standard deviations in parentheses.

Table A.2: Correlates of Responses to Regulation

	(1) Choose maximum quota	(2) Choose minimum quota	(3) Between minimum and maximum	(4) Pay fee to avoid apprentices
High-Skill Sector	-0.63*** (0.0100)	0.051*** (0.011)	-0.045*** (0.0035)	0.58*** (0.0050)
Number of Workers	0.00022*** (0.000061)	-0.00018*** (0.000057)	-0.000022 (0.000031)	-0.0000021 (0.000025)
Fraction Professionals	0.046 (0.041)	-0.051 (0.050)	0.029* (0.017)	-0.068** (0.032)
Fraction Production Workers	-0.068*** (0.023)	0.068*** (0.025)	0.017** (0.0078)	-0.032** (0.013)
Wage per worker	0.0013*** (0.00045)	-0.0012** (0.00050)	-0.000026 (0.00015)	0.000046 (0.00021)
Log Output	-0.018*** (0.0064)	0.00081 (0.0066)	0.017*** (0.0026)	-0.00098 (0.0034)
Output per worker	-0.00013*** (0.000031)	0.00015*** (0.000033)	-0.000029*** (0.0000088)	-0.000018 (0.000015)
Profit Rate	-0.17*** (0.048)	0.24*** (0.053)	-0.044*** (0.016)	-0.036 (0.027)
Capital/Output	-0.054*** (0.0081)	0.049*** (0.0085)	0.0026 (0.0029)	-0.0039 (0.0039)
Intermediates/Output	-0.20*** (0.045)	0.25*** (0.049)	-0.057*** (0.017)	0.0088 (0.023)
Mean dep. var.	0.30	0.30	0.049	0.32
Observations	21036	21036	21036	21036
R-squared	0.48	0.020	0.016	0.38

Notes: All monetary variables in 2009 pesos, in units of thousands. All regressions include year FE. Standard errors clustered by firm in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.3: Correlates of Bunching Behavior

	Bunchers Above	Bunchers Below	All Firms
Fraction of all firms	0.04	0.08	1.00
Share in High-Skill Sector	0.07	0.88	0.56
Mean number of Workers	61.04	51.14	42.19
Choose maximum quota	0.72	0.18	0.21
Choose minimum quota	0.17	0.49	0.56
Between min./max.	0.06	0.02	0.03
Pay fee to avoid apprentices	0.05	0.27	0.17
Observations	2,167	4,154	50,691
Firms	1,468	2,624	10,740

Notes: Includes only post-reform years 2003 to 2009.

Table A.4: Reduced-Form Effects of the Policy on Additional Firm Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Workers by type								
	High-Skill Workers	Low-Skill Workers	High-Skill Professionals	Low-Skill Professionals	High-Skill Admin workers	Low-Skill Admin workers	High-Skill Production workers	Low-Skill Production workers
Above*Post	-0.855 (0.863)	-1.781* (1.036)	-0.0556 (0.224)	-0.0649 (0.185)	-0.199 (0.359)	-0.956** (0.445)	-0.813 (0.652)	-0.898 (0.781)
Mean (Pre-reform)	30.46	30.30	2.581	1.500	9.393	10.08	19.10	19.08
Observations	8491	6357	7471	5587	8491	6357	8491	6357
R-squared	0.904	0.894	0.767	0.867	0.887	0.859	0.874	0.883
Panel B: Workers (survey), sales, value added								
	High-Skill Workers (Survey)	Low-Skill Workers (Survey)	High-Skill Temp. workers (survey)	Low-Skill Temp. workers (survey)	High-Skill Log Sales	Low-Skill Log Sales	High-Skill Log Value Added	Low-Skill Log Value Added
Above*Post	-0.757 (0.891)	-2.372* (1.304)	0.604 (0.643)	0.392 (1.805)	-0.0293 (0.0384)	-0.0206 (0.0512)	-0.0429 (0.0423)	0.0190 (0.0553)
Mean (Pre-reform)	31.46	32.33	3.329	4.831	14.59	14.79	13.52	13.49
Observations	8491	6357	8491	6357	8422	6159	8491	6357
R-squared	0.907	0.899	0.687	0.597	0.842	0.852	0.814	0.786
Panel C: Other inputs								
	High-Skill Log Capital	Low-Skill Log Capital	High-Skill Log Intermediates	Low-Skill Log Intermediates	High-Skill Log Energy	Low-Skill Log Energy	High-Skill Log Water	Low-Skill Log Water
Above*Post	0.00349 (0.0483)	0.0272 (0.0689)	-0.0268 (0.0382)	-0.00554 (0.0502)	-0.0228 (0.0433)	-0.0112 (0.0497)	-0.0304 (0.0562)	0.0298 (0.0760)
Mean (Pre-reform)	13.44	13.34	13.58	13.82	11.67	11.98	9.978	9.837
Observations	8491	6357	8491	6357	8491	6357	8407	6288
R-squared	0.834	0.850	0.861	0.871	0.876	0.882	0.787	0.753

Notes: All regressions include year and firm FE. Standard errors clustered by firm. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B Additional Information on Apprentices

This appendix presents information on apprentices, using two additional data sources. The first dataset is a survey of school-to-work transitions (ETET) focused on young individuals aged between 14 and 29 years. The survey was conducted by the Colombian government in 2013 and 2015. The second dataset are monthly administrative records from the social security system (PILA) gathered by the Colombian Ministry of Health and Social Protection and processed by the Colombian National Statistical Agency (DANE). This data is available for 2015 and 2016. In both data sets, we pool the available years.

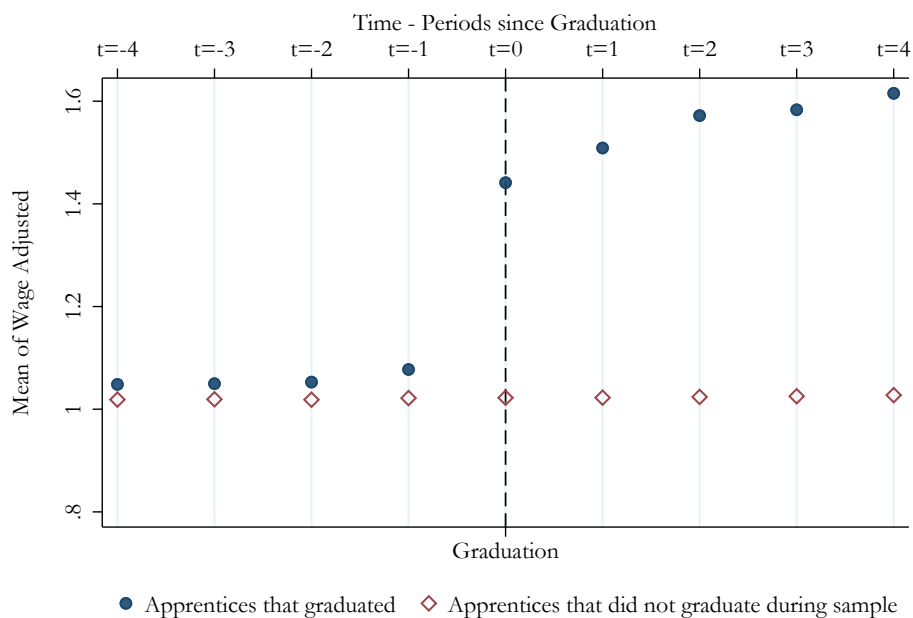
Table B.1 shows characteristics of apprentices and other workers based on the two data sources. Figure B.1 and Table B.2 focus on apprentices in the PILA data who complete their apprenticeship during the sample period. Figure B.1 shows wages of this group around the month of graduation, compared to a control group of apprentices that do not graduate throughout this period. Table B.2 shows transition probabilities between firms and sectors and wages of apprentices after they complete their apprenticeship.

Table B.1: Characteristics of Apprentices

	(1)	(2)	(3)	(4)
	Apprentices (ETET)	Other Young Workers (ETET)	Apprentices (PILA)	Other Workers (PILA)
Fraction female	68.4%	44.7%	57%	38%
Age	20.94	23.35	23.92	38.21
Fraction 14 to 19	35.7%	18.4%	11.8%	0.5%
Fraction 20 to 24	51.0%	39.1%	58.5%	9.3%
Fraction 25 to 29	13.3%	42.5%	17.7%	15.7%
Education				
Primary or less	0.0%	5.4%		
Some Secondary	2.0%	14.0%		
High School	18.4%	32.7%		
Technical/vocational	49.0%	24.6%		
College	30.6%	23.3%		
Last monthly wage	513,367	808,381		
Wage relative to minimum wage			0.97	2.77
Field of training/occupation				
Humanities & arts	4.1%	5.8%		
Social sciences and business	35.7%	24.3%		
Sciences	3.1%	3.6%		
Engineering, industry and construction	23.5%	27.5%		
Agriculture	0.0%	1.3%		
Health and social services	13.3%	17.0%		
Education	1.0%	6.6%		
General Services	9.2%	5.3%		
Other or not specified	10.2%	8.6%		
Socioeconomic status				
Low	44.8%	66.3%		
Medium	52.1%	31.9%		
High	3.1%	1.8%		
Satisfied with job/apprenticeship	91.8%	81.2%		
Want to move job/apprenticeship	29.6%	42.5%		
Number of observations	98	6,467	18,273	646,811

Notes: The table summarizes characteristics of apprentices and workers from ETET data (Columns 1 and 2) and PILA data (Columns 3 and 4). “Other young workers” in ETET are 14 to 29 years old.

Figure B.1: Wages before and after Completing the Apprenticeship



Notes: The figure shows wages of apprentices in the PILA data who graduate during the sample period vs. a control group of apprentices who do not graduate during this time. Time in months relative to time of completing apprenticeship. Wages are shown in units relative to minimum wage.

Table B.2: Firm and Sector Transitions after Completing the Apprenticeship

		(1)	(2)
		Trained in High-Skill Sector	Trained in Low-Skill Sector
Stay in Same Firm	Probability	74.35%	75.14%
	Wage	1.47 (0.97)	1.35 (0.64)
Move to Different Firm in High-Skill Sector	Probability	15.56%	7.50%
	Wage	1.55 (1.19)	1.44 (0.74)
Move to Different Firm in Low-Skill Sector	Probability	10.10%	17.36%
	Wage	1.45 (0.74)	1.33 (0.64)
Observations		9,032	12,344

Notes: The table shows the probabilities of transitions between firms and sectors for apprentices in the PILA data who graduate during the sample period. Wages are shown in units relative to the minimum wage with standard deviations in parentheses.

C Model Proofs and Extensions

C.1 Equilibrium and Proofs

Here we characterize the equilibrium with and without regulation, show some additional theoretical results and give a proof to Proposition 1.

Assumption 1 on f allow us to further characterize the optimal number of workers and apprentices, guaranteeing the existence and uniqueness of the solution. Condition (i) and (ii) ensure the existence of a unique solution with $l > 0$. Condition (iii) implies that the optimal number of workers $n^*(z, t_a)$ and apprentices $n_a^*(z, t_a)$ are non-decreasing in z . In other words, firms with higher managerial ability are larger. We formalize these claims in Lemma 1.

Lemma 1. *Assumption 1 implies there are unique labor demands $n^*(z) > 0$ and $n_a^*(z) \geq 0$, with $t_a n_a^* \leq n^*$ solving the firm z 's optimization problem (1). Moreover, these labor demands are non-decreasing in the firm's managerial ability, $\frac{\partial n^*(z)}{\partial z} \geq 0$ and $\frac{\partial n_a^*(z)}{\partial z} \geq 0$.*

Proof. Similar to standard production theory, homogeneity of degree γ implies concavity (and hence quasiconcavity) of production function. Since we assumed $\gamma \in (0, 1)$, the solution $l^*(z), l_a^*(z)$ exists. Additionally the Inada condition on l guarantees the solution is unique. From these labor demands we can back out the optimal number of workers n^* and apprentices n_a^* , $n^* = l^* + t_a n_a^*$ and $n_a^* = \frac{l_a^*}{\zeta_a}$. Since $l^*(z), l_a^*(z) > 0$, then $t_a n_a^* < n^*$ and $n_a^* > 0$.

Now, since the cross-derivatives are non-negative, monotone comparative statics imply $\frac{\partial l^*(z)}{\partial z} \geq 0$ and $\frac{\partial l_a^*(z)}{\partial z} \geq 0$. This immediately implies, $\frac{\partial n^*(z)}{\partial z}, \frac{\partial n_a^*(z)}{\partial z} \geq 0$. \square

At an interior solution, the marginal rate of substitution between the two types of labor is equal to the ratio of marginal labor costs,

$$-\frac{\frac{\partial f}{\partial l}}{\frac{\partial f}{\partial l_a}} = -\frac{w\zeta_a}{w_a + t_a w}.$$

We can use the FOCs to analyze how wages or the required training time affect the optimal labor allocation decision. As usual, an increase in the relative wage of apprentices lowers their demand. Similarly, an increase in training costs decreases the demand for apprentices.

Lemma 2. *Suppose that Assumption 1 holds, then $\frac{n_a}{n}$ is weakly decreasing in w_a and t_a , and weakly increasing in w .*

Proof. From the firm's optimization problem,

$$\frac{dl_a}{dl} = -\frac{\frac{\partial f}{\partial l}}{\frac{\partial f}{\partial l_a}} = -\frac{w\zeta_a}{w_a + t_a w}.$$

Let $W = \frac{w\zeta_a}{w_a + t_a w}$ denote the ratio of the price of workers' and apprentices' labor. In equilibrium if W increases then $\frac{dl_a}{dl}$ decreases. Since f is homogenous of degree $\gamma \in (0, 1)$, this means l_a^*/l^* increases.

Now, $\frac{n_a^*}{n^*} = \frac{1}{\zeta_a \frac{l^*}{l_a} + t_a}$. All the comparative static results follow from this equation and the previous observations.

If w_a increases then W decreases, so l_a^*/l decreases, implying n_a^*/n^* also decreases. Similarly, an increase in t_a , implies W and l_a^*/l^* decrease. Now this increase in t_a decreases n_a^*/n^* directly and indirectly through l_a^*/l^* , so n_a^*/n^* decreases.

The analogous logic applies to w . If w increases, W and l_a^*/l^* also rise, implying n_a^*/n^* increases. \square

Equilibrium

All individuals in the economy are infinitely lived and are endowed with a unit of labor that they supply inelastically. An individual i maximizes their lifetime utility

$$\max_{(c_t)_t} \sum_{t=1}^{\infty} \beta^t u(c_t) \quad \text{s.t.} \quad \sum_{k=1}^K p_t^k c_t^k = I_t^i \quad \forall t, \quad (6)$$

where $I_{i,t}$ denotes individual i 's income in period t . Note that as in the firms' case, individual decisions are static. Solving this problem implies the usual optimality conditions, $\frac{\partial u_t / \partial c_t^k}{\partial u_t / \partial c_t^j} = \frac{p_t^k}{p_t^j}$ and $\sum_{k=1}^K p_t^k c_t^k = I_t^i \quad \forall k, t$. Assuming $u(\cdot)$ is quasiconcave, let $c_t^*(I_t^i; p_t) = (c_t^{1*}(I_t^i; p_t), \dots, c_t^{K*}(I_t^i; p_t))$ be the solution to individual i 's optimization problem in period t . We use the market clearing conditions in the goods market for each sector to determine the sectoral prices $p_t = (p_t^1, \dots, p_t^K)$,

$$C_t^k(p_t) + C_{a,t}^k(p_t) + C_{f,t}^k(p_t) = Y_t^k(p_t) \quad \forall k,$$

where $C_t^k(p_t)$ is workers' aggregate demand for good k , $C_{a,t}^k$ is apprentices' aggregate demand, $C_{f,t}^k$ is firm owners' aggregate demand and $Y_t^k(p_t)$ the aggregate production of good k .

$$C_t^k(p_t) := \sum_{j=1}^K L_t^j c_t^{k*}(w_t^j; p_t), \quad C_{a,t}^k := \sum_{j=1}^K L_{a,t}^j c_t^{k*}(w_{a,t}^j; p_t),$$

$$C_{f,t}^k := \sum_{j=1}^K F_t^j \int \int c_t^{k*}(\pi_t^j(z, t_a); p_t) d\mathcal{Z}(z) d\mathcal{T}(t_a), \quad Y_t^k(p_t) := \int \int y_t^{k*}(p_t; z, t_a) d\mathcal{Z}(z) d\mathcal{T}(t_a).$$

Definition 1. A competitive equilibrium is given by wages $((w_t^{k*})_k, w_{a,t}^*)_t$ and prices p_t^k for each sector k and each period t ; and quantities of unemployed workers and untrained apprentices $((U_t^{k*})_k, U_{a,t}^*)_t$, labor demands $(n_t^{k*}(z, t_a), n_{a,t}^{k*}(z, t_a))$ for each firm (z, t_a) and consumption c_t^{k*} such that

(i) firms solve the optimization problem (1),

(ii) wage restrictions are satisfied, $w_t^{*k} \geq w_{a,t}^{*k} \geq w_{\min}$ and labor markets clear with $U_t^{k*} \geq 0$ and $U_{a,t}^* \geq 0 \quad \forall t, k$.

(iii) apprentices increase labor in each period for all sectors, as in (2),

(iv) individuals maximize utility (6) and,

(iv) the goods market clear for each sector.

Note that there is unemployment and untrained apprentices whenever the wage restrictions are binding. If there are some unemployed workers $U^* > 0$, then $w^* = w_a^* = w_{min}$. Similarly if there are some untrained apprentices $U_a^* > 0$ then $w_a^* = w_{min}$.

Equilibrium With Regulation

In the Lemma 3 we characterize the solution to the optimization problem with regulation, relative to the problem without regulation. We use this characterization for the proof of Proposition 1.

Lemma 3. *Let $(n^*(z, t_a), n_a^*(z, t_a))$ denote the optimal number of workers and apprentices a firm with managerial ability z and training costs t_a hires when solving the maximization problem (1) (without regulation). Let $n^r(z, t_a)$ and $n_a^r(z, t_a)$ denote the optimal number of workers and apprentices the firm hires when solving (3) (with regulation). For all $j \geq 1$ let \underline{n}_a^j and \bar{n}_a^j be the maximum and the minimum quota between thresholds N_{j-1} and N_j .*

i. If $n_a^*(z, t_a) > \bar{n}_a^j$, then

– $\exists k \geq j+1$ such that $n^r(z, t_a) = N_k$ and $n_a^r(z, t_a) > \bar{n}_a^j$ (increase size to get more apprentices),

or

– $\forall k \geq j$ $n_a^r(z, t_a) = \bar{n}_a^k$ and $n^r(z, t_a) < n^*(z, t_a)$ (bounded by maximum quota).

ii. If $n_a^*(z, t_a) \in [\underline{n}_a^j, \bar{n}_a^j]$, then $n^r(z, t_a) = n^*(z, t_a)$ and $n_a^r(z, t_a) = n_a^*(z, t_a)$.

iii. If $n_a^*(z, t_a) < \underline{n}_a^j$, then

– $\exists k \leq j$ such that $n^r(z, t_a) = N_k - \varepsilon$ (with $\varepsilon \rightarrow 0$) and $n_a^r(z, t_a) < \bar{n}_a^j$ (reduce size to avoid apprentices),

– $n^r(z, t_a) \geq n^*(z, t_a)$ and $n_a^r(z, t_a) < \underline{n}_a^j$ and $d_{f_a} = 1$ (pay the fee to avoid apprentices) **or**

– $n_a^r(z, t_a) = \underline{n}_a^j$ (bounded by the minimum quota).

Proof. Choose any firm $z > 0$ and $t_a \geq 0$. Denote by $\pi(N)$ the maximum profit function when the number of workers is fixed to N and $\pi(N_a)$ when the number of apprentices is fixed to N_a ,

$$\pi(N) = \max_{l_a \geq 0} pf(N - \frac{t_a}{\zeta_a} l_a, l_a) - wL - \frac{w_a + t_a w}{\zeta_a} l_a, \quad \pi(N_a) = \max_{l \geq 0} pf(l, N_a) - wl - (w_a + t_a w)N_a.$$

To simplify the notation define $\tilde{w}_a = \frac{w_a + t_a w}{\zeta_a}$ and $\tilde{t}_a = t_a / \zeta_a$. We also use the subindex notation of partial derivatives to economize on the writing, $f_x := \frac{\partial f}{\partial x}$.

First we show that $\pi(N)$ and $\pi(N_a)$ are concave.

Lets start with $\pi(N_a)$. Using the envelope theorem, $\frac{\partial \pi(N_a)}{\partial N_a} = p f_{l_a} \zeta_a - w_a$. We can differentiate again this expression with respect to N_a to obtain,

$$\frac{\partial^2 \pi(N_a)}{\partial N_a^2} = p \left(f_{l_a l} \frac{dl^r}{dN_a} + f_{l_a l_a} \zeta_a \right). \quad (7)$$

Where l^r solves the FOC of the fixed N_a optimization problem, $pf_l(l^r, N_a) = w$. Assumption (1) imply the existence and uniqueness of the solution l^r . Totally differentiating this FOC wrt l^r and N_a , implies $\frac{dl^r}{dN_a} = -\frac{f_{l_a l_a}}{f_{l_a}} \zeta_a \geq 0$, the inequality given $f_{l_a l_a} \geq 0$ and $f_{l_a} \leq 0$. Replacing this derivative in (7),

$$\frac{\partial^2 \pi(N_a)}{\partial N_a^2} = p \left(f_{l_a l_a} - \frac{f_{l_a l_a}}{f_{l_a}} \zeta_a + f_{l_a l_a} \zeta_a \right) = \frac{p \zeta_a}{f_{l_a}} (f_{l_a} f_{l_a l_a} - f_{l_a l_a}^2) \leq 0,$$

since Assumption (1) imply f is concave in l, l_a , so $f_{l_a}, f_{l_a l_a} \leq 0$ and $(f_{l_a} f_{l_a l_a} - f_{l_a l_a}^2) \geq 0$. This means $\pi(N_a)$ is concave in N_a .

Importantly this function is maximized at (n^*, n_a^*) . If we choose n_a away from n_a^* , profits will decrease.

So in Case (i), if $n_a^* > \bar{n}_a^j$, whenever the firm stays in the same j th regulation bracket, it chooses the feasible number of apprentices that is closest to n_a^* . This means the upper-bound is binding $n_a^r = \bar{n}_a^j$. Moreover, since $\frac{dl^r}{dN_a} \geq 0$ we know $l^r < l^*$ given $n_a^* > \bar{n}_a^j$. This implies $n^r = l^r + t_a n^r = l^r + t_a \bar{n}_a^j < n^*$.

Similarly we can show $\pi(N)$ is concave. In this case,

$$\frac{\partial^2 \pi(N)}{\partial N^2} = p f_{l_a} \left(1 - \tilde{t}_a \frac{dl_a^r}{dN} \right) + p f_{l_a} \frac{dl_a^r}{dN}.$$

Considering the FOC and totally differentiating,

$$\frac{dl_a}{dN} = -\frac{f_{l_a l_a} - f_{l_a} \tilde{t}_a}{f_{l_a} \tilde{t}_a^2 - 2 f_{l_a} \tilde{t}_a + f_{l_a l_a}} \geq 0.$$

Substituting in the previous equation,

$$\frac{\partial^2 \pi(N)}{\partial N^2} = p \frac{(f_{l_a} f_{l_a l_a} - f_{l_a}^2)}{f_{l_a} \tilde{t}_a^2 - 2 f_{l_a} \tilde{t}_a + f_{l_a l_a}} \leq 0,$$

this last inequality again from the concavity and cross-partial derivative of f .

This proves $\pi(N)$ is concave. Using a similar argument as before, the firm wants to get as close as possible the optimal labor demands (n^*, n_a^*) . However, now we also have to compare subsequent thresholds N_k for $k \geq j+1$, as $n_a^r(a, t_a)$ might still be larger than \bar{n}_a^{j+1} , so a firm might want to jump multiple thresholds to get a higher number of apprentices. In all these cases, the firm chooses the number of worker at a threshold N_k as it is the closet to the optimal number of workers that allows the firm to get $n_a^r \in [\bar{n}_a^{k-1}, \bar{n}_a^k]$. The optimal number of apprentices is in this case $n_a^r(z, t_a) > \bar{n}_a^j$.

Case (ii) is immediate as the unconstrained optimum is within the regulation bounds, so the firm won't change its optimal decision.

Case (iii), whenever firms choose to bunch just below a threshold or choose the minimum quota of the regulation is analogous to the proof of Case (i). It remains to show that for ϕ_a relatively low, some firms prefer to pay the fee instead of hiring the minimum number of required apprentices.

To see this, suppose $n_a^* < \underline{n}_a^j$ and define, $\pi^*(\phi_a) := pf(l^*, l_a^*) - wl^* - \tilde{w}_a l_a^* - \phi_a (\underline{n}_a^j - n_a^*)$. Note that

the optimal choice of workers and apprentices when paying the fee, implies larger or equal profits,

$$\pi^j(\phi_a) := \max_{l_a, l \geq 0} pf(l, l_a) - wl - wl_a - \phi_a \left(\frac{n_a^j}{\zeta_a} - \frac{l_a}{\zeta_a} \right) \geq \pi^*(\phi_a).$$

Now, we know $\pi^* := pf(l^*, l_a^*) - wl^* - \tilde{w}_a l_a^* \geq \pi(N)$ and $\pi^* \geq \pi(N_a)$, for any $N, N_a \geq 0$. Also since $\pi^j(\phi_a)$ is continuous in ϕ_a , and $\lim_{\phi_a \rightarrow 0} \pi^*(\phi_a) = \pi^*$. So there exist $\tilde{\phi}_a > 0$ small enough, such that $\pi^j(\tilde{\phi}_a) \geq \pi(N)$ and $\pi^j(\tilde{\phi}_a) \geq \pi(N_a)$. \square

Lemma 4. *Suppose Assumption 1 except f does not necessarily have CRS in (l, l_a, z) . For each firm z there exists $A(z) > 0$ such that $l_a^* = A(z)l^*$.*

i If $A'(z) > 0$, the parametric mapping $(n_a^*(z), n^*(z))$ is strictly convex.

ii If $A'(z) = 0$, the parametric mapping $(n_a^*(z), n^*(z))$ is linear.

iii If $A'(z) < 0$, the parametric mapping $(n_a^*(z), n^*(z))$ is strictly concave.

Proof. Take any firm $z > 0$. First lets show that $l_a^* = A(z)l^*$. Since f is homogenous of degree γ , then $\frac{\partial f}{\partial l}$ and $\frac{\partial f}{\partial l_a}$ are homogenous of degree $\gamma - 1$. This means that for any constant $k > 0$,

$$\frac{\frac{\partial f}{\partial l}(kl, kl_a; z)}{\frac{\partial f}{\partial l_a}(kl, kl_a; z)} = \frac{k^{\gamma-1} \frac{\partial f}{\partial l}(l, l_a; z)}{k^{\gamma-1} \frac{\partial f}{\partial l_a}(l, l_a; z)} = \frac{\frac{\partial f}{\partial l}(l, l_a; z)}{\frac{\partial f}{\partial l_a}(l, l_a; z)}.$$

So the derivative of the isoquants are constant along any ray starting from the origin. Since $\gamma \in (0, 1)$ implies the production function is quasiconcave, then there is only one point (l^*, l_a^*) such that $-\frac{\frac{\partial f}{\partial l}(l^*, l_a^*)}{\frac{\partial f}{\partial l_a}(l^*, l_a^*)} = -\frac{w\zeta_a}{w_a + t_a w}$. Together this means l_a/l is constant whenever the derivative of the isoquant is the same. So, $\frac{l_a^*}{l^*} = A(z)$ for some $A(z) > 0$.

Now note that since, $l = n - t_a n_a$ and $l_a = \zeta_a n_a$, then $\frac{n_a^*}{n^*} = \frac{1}{\zeta_a A(z)^{-1} + t_a}$, call this last term $B(z)$. Hence, $A'(z) > 0 \iff B'(z) > 0$.

From the equation above, $\frac{dn_a^*}{dn^*} = B(z)$, $\forall z$. This means that if the $B'(z)$ is increasing in z , $\frac{dn_a^*}{dn^*}$ is increasing in z . From Lemma (2) $\frac{dn_a^*}{dz} > 0$, implying the parametric mapping $n_a^*(n^*)$ is convex. Similarly if $A'(z) = 0 \Rightarrow B'(z) = 0$ and so the derivative is constant for any z , $\frac{dn_a^*}{dn^*} = B \in \mathbb{R}_+ \forall z$. This means the parametric mapping is linear. Finally if $A'(z) < 0 \Rightarrow B'(z) < 0$, then $\frac{dn_a^*}{dn^*}$ is decreasing in z and hence $n_a^*(n^*)$ is concave. \square

Corollary 1. *Under Assumption 1, $n_a^* = Bn^*$ where B doesn't depend on z*

Proof. The additional CRS of f on (l, l_a, z) implies $\frac{(\partial f / \partial l)}{(\partial f / \partial l_a)}$ is homogenous of degree 0 in z . So case (ii) in Lemma 4 applies, $l_a^* = Al^*$ for some $A \geq 0$ independent of z . Using the same argument in Lemma 4, $n_a^* = Bn^*$, for $B = \frac{1}{\zeta_a A^{-1} + t_a}$. \square

Proposition 1. *Suppose Assumption 1 hold and firms solve the maximization problem with regulation (3). Then,*

Case 1: there exist $(\frac{\bar{w}_a}{w}, \bar{t}_a)$ such that for $\frac{w_a}{w} \leq \frac{\bar{w}_a}{w}$ and $t_a \leq \bar{t}_a$,

- i. the number of apprentices without regulation is $n_a^* = B_u n^*$ and is above the maximum quota, $n_a^*(z, t_a) > \bar{n}_a^j$.
- ii. there exist cutoffs $\{z_b^j, z_r^j\}_j$ such that firms $z \in [z_b^j, z_r^j]$, increase their size to the threshold N_k with $k > j$, so there is missing mass to the left of the thresholds.
- iii. firms choose the upper-bound of the regulation $n_a^r = \bar{n}_a^j$.
- iv. firms never pay the fee.

Case 2: there exist $\frac{\underline{w}_a}{w}$ such that for $\frac{w_a}{w} \geq \frac{\underline{w}_a}{w}$,

- i. the number of apprentices without regulation is $n_a^* = B_s n^*$ and is below the minimum quota, $n_a^*(n) < \underline{n}_a^j$.
- ii. there exist cutoffs $\{z_b^j, z_r^j\}_j$ such that firms $z \in [z_b^j, z_r^j]$, reduce their size ϵ below a lower threshold N_k for $k \leq j$.
- iii. firms that choose to increase apprentices, choose the minimum number \underline{n}_a^j .
- iv. there exist $\bar{\phi}_a > 0$ such that for $\phi_a \leq \bar{\phi}_a$, there is an additional cutoffs z_f^j such that firms $z \in (z_r^j, z_f^j]$ choose to pay the fee.

Proof. Let $j \geq 1$. Assumption 1 and Corollary 1 imply the solution without regulation is $n_a^* = Bn^*$ for some $B \geq 0$ that doesn't depend on z . Lemma 2 implies B is a continuous non-increasing function of $\frac{w_a}{w}$ and t_a .

Let's start with Case 1. When $\frac{w_a}{w}$ and t_a go to 0, then the optimal relative number of apprentices n_a^*/n^* is unbounded. This means below some threshold $(\frac{\bar{w}_a}{w}, \bar{t}_a)$, $n_a^* > \bar{n}_a^j$, part (i). Now, for any $t_a < \bar{t}_a$, $w_a/w \leq \frac{\bar{w}_a}{w}$, Lemma 1 implies n_a^* and n^* are increasing and continuous in z . After some threshold z_b^j , by Lemma 3 ($n_a(N_k), N_k$) for some $k > j$ is closer to the optimal labor input (n_a^*, n^*) , hence firm (z, t_a) , with $z \geq z_b^j$ will bunch at this threshold N_k above the j -th level. Let z_r^j be the productivity such that N^k is the optimal number workers without regulation for firm (z_r^j, t_a) , then firms beyond $z > z_r^j$ won't increase their size, relative to the equilibrium with no regulation, completing the proof of part (ii). Since $n_a^* > \bar{n}_a^j$ by Lemma 3 firms will choose $n_a^r = \bar{n}_a^j$, part (iii). And finally, firms never pay the fee as the number of desired apprentices is above the maximum quota.

The proof of Case 2 (i)-(iii) is analogous. Note that for $\phi \rightarrow 0$, we can use the same argument as in the proof of Lemma 3, to prove firms prefer to pay the fee instead of hiring the minimum quota of apprentices. \square

Negative Marginal Productivity of Apprentices

In this section we show that the production function has to allow for apprentices to have negative marginal productivity for the firms that choose to pay this fee.

To see this, consider a standard production function $f(n, n_a; z)$ combining managerial ability z with labor input from workers and apprentices. Let us compare two scenarios based on the elements

of the policy. First, suppose firms are required to train at least \underline{n}_a apprentices paying a wage rate of w_a^{min} . Alternatively, firms can pay a fee ϕ_a for each of the required apprentices.

When a firm z chooses to train the apprentices it solves

$$\pi_a(z) := \max_{n, n_a} pf(n, n_a; z) - wn - w_a^{min} n_a, \quad n_a \geq \underline{n}_a.$$

Alternatively, the firm could choose to pay the fee and solve

$$\pi_f(z) := \max_n pf(n, 0; z) - wn - \phi_a \underline{n}_a.$$

In proposition 2 we show that if $\phi_a > w_a^{min}$, then firms choose to pay the fee only if the marginal productivity of apprentices is negative. Formally,

Proposition 2. *If $\phi_a > w_a^{min}$, then $\pi_f(z) > \pi_a(z) \Rightarrow \frac{\partial f}{\partial n_a} < 0$.*

Proof. By way of contradiction suppose $\frac{\partial f}{\partial n_a} \geq 0$. In this case,

$$\pi_a(z) \geq f(n_0^*, \underline{n}_a; z) - wn_0^* - w_a \underline{n}_a \geq f(n_0^*, 0; z) - wn_0^* - w_a \underline{n}_a > f(n_0^*, 0; z) - wn_0^* - \phi_a \underline{n}_a = \pi_f(z),$$

the first two inequalities as $\pi_a(z)$ is the maximum profit function and $\frac{\partial f}{\partial n_a} \geq 0$, and the last inequality given $\phi_a > w_a$. This contradicts $\pi_f(z) > \pi_a(z)$. Therefore, $\pi_f > \pi_a \Rightarrow \frac{\partial f}{\partial n_a} < 0$. \square

In other words, if apprentices have positive productivity and it is cheaper to hire an apprentice than paying the fee, firms would naturally choose to hire the apprentice. In the model, we allow for a negative marginal revenue product of apprentices simply by adding training costs.

C.2 Additional Inputs

In this section we describe an extension of the model when we add other inputs. We discuss a simple example to illustrate the results based on the benchmark model used in our quantitative exercises.

Suppose we have an additional input x with price w_x that firms choose in each period. First we consider the firm problem without regulation. Consider a simple Cobb-Douglas specification,

$$\max_{n, n_a, x} pz^{1-\gamma} (n - t_a n_a + n_a)^{\gamma_l} x^{\gamma_x} - wn - w_a n_a - w_x x \quad s.t \quad t_a n_a \leq n$$

where γ_l is the output elasticity of labor, and γ_x the output elasticity of input x . Suppose the production function has constant returns to scale on (z, n, n_a, x) , so $\gamma_l + \gamma_x = \gamma$. As in the benchmark model, linearity of the labor input implies there are corner solutions. A firm (z, t_a) avoids apprentices whenever $w < t_a w + w_a$. In that case, from the FOC, $x = \frac{w}{w_x} \frac{\gamma_x}{\gamma_l} n =: \chi n$.

So the optimal demand for inputs are, $n_a^* = 0$, $n^* = \left(\frac{\gamma_l \chi^{\gamma_x}}{w} \right)^{\frac{1}{1-\gamma}} z$, $x^* = \chi n^*$.

The corresponding output and profits,

$$y^* = \left(\frac{\gamma_l}{w} \right)^{\gamma/(1-\gamma)} \chi^{\frac{\gamma_x}{1-\gamma}} z, \quad \pi^* = \left(\frac{\gamma_l}{w} \right)^{\gamma/(1-\gamma)} \chi^{\frac{\gamma_x}{1-\gamma}} (1-\gamma)z.$$

On the other hand, if $w > t_a w + w_a$ the firm seeks apprentices so, $x = \frac{w_a + t_a w}{w_x} \frac{\gamma_x}{\gamma_l} n =: \chi_a n_a$. So the optimal demand for inputs are $n_a^* = \left(\frac{\gamma_l \chi_a^{\gamma_x}}{w_a + t_a w} \right)^{\frac{1}{1-\gamma}} z$, $n^* = t_a n_a^*$, $x^* = \chi_a n_a^*$. With corresponding output and profits,

$$y^* = \left(\frac{\gamma_l}{w_a + t_a w} \right)^{\gamma/(1-\gamma)} \chi_a^{\frac{\gamma_x}{1-\gamma}} z, \quad \pi^* = \left(\frac{\gamma_l}{w_a + t_a w} \right)^{\gamma/(1-\gamma)} \chi_a^{\frac{\gamma_x}{1-\gamma}} (1-\gamma) z.$$

Now, with regulation lets study the case of a particular threshold. Firms have the option of bunch at the threshold N to avoid the policy, comply with the apprenticeship quota hiring the required number of apprentices \underline{n}_a or pay the fee without hiring any apprentices.

Suppose firm bunches at N to avoid the policy,

$$n^r = N, \quad n_a^r = 0, \quad x^r = \left(\frac{\gamma_x N^{\gamma_l}}{w_x} \right)^{1/(1-\gamma_x)} z^{(1-\gamma)/(1-\gamma_x)}$$

$$y^r = \left(\frac{\gamma_x}{w_x} \right)^{\gamma_x/(1-\gamma_x)} N^{(\gamma-\gamma_x)/(1-\gamma_x)} z^{(1-\gamma)/(1-\gamma_x)}, \quad \pi^r = \left(\frac{\gamma_x}{w_x} \right)^{\gamma_x/(1-\gamma_x)} (1-\gamma_x) N^{(\gamma-\gamma_x)/(1-\gamma_x)} z^{(1-\gamma)/(1-\gamma_x)},$$

Instead if the firm has to take n_a apprentices, the same analysis apply,

$$n^r = \left(\frac{\gamma_l \chi^{\gamma_x}}{w} \right)^{\frac{1}{1-\gamma}} z - (1-t_a)n_a, \quad x^r = \chi n^r.$$

Suppose the firm pays the fee,

$$n_a^r = 0, \quad n^r = \left(\frac{\gamma_l \chi^{\gamma_x}}{w} \right)^{\frac{1}{1-\gamma}} z, \quad x^r = \chi n^r.$$

The corresponding output and profits,

$$y^r = \left(\frac{\gamma_l}{w} \right)^{\gamma/(1-\gamma)} \chi^{\frac{\gamma_x}{1-\gamma}} z, \quad \pi^r = \left(\frac{\gamma_l}{w} \right)^{\gamma/(1-\gamma)} \chi^{\frac{\gamma_x}{1-\gamma}} (1-\gamma) z - \psi_a \underline{n}_a.$$

From the equations above, we see that the effect of adding other inputs is that firms have additional margins of substitution. Qualitatively, there are no differences with the benchmark model used throughout the paper. However quantitatively, this affects the magnitude of firm responses. In terms of the estimation exercises, the fit of the firm-size distribution would be similar, but now we need information on the share of these other input in production to identify all the parameters of the production function. Now, if γ_x is higher, then firms would respond less to the policy as the output elasticity with respect to labor decreases. This could help lessen the overestimation of bunching we had when estimating the benchmark model.

C.3 Multiple Types of Workers

In this section, we describe the extension of the model including multiple types of workers. For clarity of exposition let us suppose there are two types of workers, unskilled u and skilled s . We characterize the equilibrium for the linear labor input case, combining these types of workers in a Cobb-Douglas function.

Suppose firms are characterized by a managerial ability z and the training costs for each type of worker, t^i with $i \in \{u, s\}$.

First we study the case without regulation. A firm (z, t_a^u, t_a^s) solves,

$$\max_{n^i, n_a^i} z^{1-\gamma} (n^u + (1-t_a^u)n_a^u)^{\gamma_u} (n^s + (1-t_a^s)n_a^s)^{\gamma_s} - \sum_{i=u}^s (w^i n^i + w_a n_a^i) \quad s.t. \quad t_a^i n_a^i \leq n^i, \quad \forall i,$$

where $\sum_i \gamma_i = \gamma$.

As before we have corner solutions. Firms avoid apprentices of type i , if $w^i < w^i t_a^i + w_a$: $n_a^i = 0$ and $n^i = \left(\frac{\gamma_i A^{\gamma_j}}{w^i}\right)^{1/(1-\gamma)} z$, where $A = \frac{w^i \gamma_j}{w^j \gamma_i}$. Firms want apprentices of type i , if $w^i > w^i t_a^i + w_a$: $n_a^i = \left(\frac{\gamma_i A^{\gamma_j}}{w^i t_a^i + w_a}\right)^{1/(1-\gamma)} z$ and $n^i = t_a^i n_a^i$.

Now let us consider the firm decision when the regulation applies. For the quotas, suppose the firm has to train n_a apprentices. First, we show firms generically choose to train apprentices only in one occupation (by only one type of worker), depending on which one is relatively cheaper,

Lemma 5. *Firm chooses to train apprentices in occupation, $i^* = \arg \max_i w^i (1 - t_a^i)$.*

Lemma 5 implies we only have to compare the corner solutions to the choice of apprentices. Suppose the firm optimally chooses to training apprentices in occupation i . Let $x_i^* = \left(\left(\frac{\gamma_i}{w^i}\right)^{1-\gamma_j} \left(\frac{\gamma_j}{w^j}\right)^{\gamma_j}\right)^{1/(1-\gamma)} z$, then $n_i^r = x_i^* - (1-t_a^i)n_a^i$. Using the Lemma 5 we only have to consider when $n_a^i > 0$ and $n_a^j = 0 \quad \forall j \neq i$. The regulation quota is determine over $n^r = \sum_j n_j^r$.

For the quantitative exercises we follow a similar procedure as in the benchmark model. First, we estimate the output elasticity of each type of worker, γ_i using the pre-reform production data. Then, we match the pre-reform size distribution, using the firms that do not hire apprentices and the total number of workers,

$$n = \sum_i n^i = \sum_i \left(\left(\frac{\gamma_i}{w^i}\right)^{1-\gamma_j} \left(\frac{\gamma_j}{w^j}\right)^{\gamma_j}\right)^{1/(1-\gamma)} z.$$

In this case, to estimate the t_a^i distributions we need information on the type of worker (the occupation) for which the apprentice gets training. We can use the firm responses to policy to infer these net training cost distributions. Use the proportion of apprentices trained by each type of worker occupation to calibrate different thresholds. For instance, for both types, we can use the probability of choosing the maximum quota and paying the fee. The theoretical equivalents are

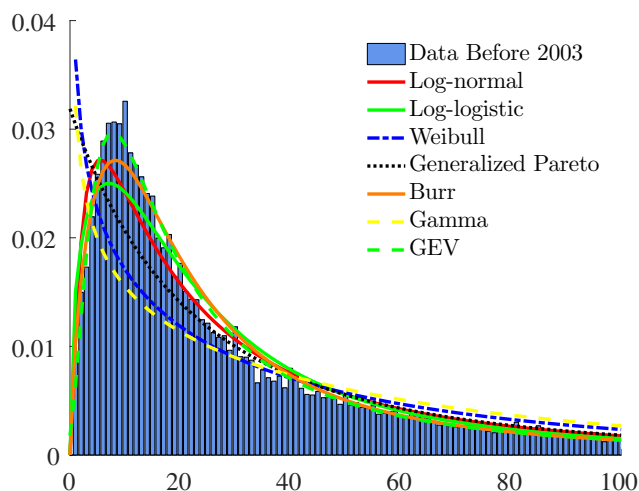
$$P_{ub}^i = Pr \left\{ t_a^i < 1 - \frac{w_a}{w^i} \right\}, \quad P_{fee}^i = Pr \left\{ t_a^i > 1 + \frac{\psi_a - w_a}{w^i}, \quad \forall i \right\}.$$

D Details of Quantitative Exercises

D.1 Fit Pre-Reform Firm Size distribution

Figure D.2 shows the fit of various parametric distributions. Out of the two parameter distributions the Log-logistic better fits the data. Out of all the distributions the Generalized Extreme Value distribution does the best job.

Figure D.2: Fitting the Pre-Reform Distribution



D.2 Production Function Estimation

Table D.3 shows the estimated labor share by sector γ^k using six different methodologies: using our baseline regression with time and firm fixed effects (FE), running an OLS regression, [Olley and Pakes \(1996\)](#) (OP), [Levinsohn and Petrin \(2003\)](#) (LP), [Wooldridge \(2009\)](#) (W) and [Ackerberg et al. \(2015\)](#) (LP-ACF). For the last four specifications we suppose the production function depends on capital K , full-time labor l and other intermediate inputs m which we approximate with energy, water and intermediate product expenditures.

Table D.3: Labor Share γ^k , Full-Time Workers

	FE (Baseline)	OLS	OP	LP	W	LP-ACF
High-Skill	0.61	0.70	0.51	0.42	0.23	0.49
Low-Skill	0.58	0.63	0.49	0.35	0.28	0.45

Note: The last four columns are computed using Stata user-written program *prodest* from [Mollisi and Rovigatti \(2017\)](#).

D.3 Moment Weights

Table D.4 details the weight for each group of moments. For the bunching and missing mass moments we use weights that correspond to the observed pre-reform fraction of firms at each bunching or missing mass point. For instance for the first bunching point for high-skill sector firms we weight the bunching mass at 14 by the fraction of high-skill sector firms of size 14 using the pre-reform data, h_{14}^s . Additionally we divide the moments for the missing mass by the bin size of 3 that we consider to make it comparable to the bunching weights. Finally, we equally weight our four group of moments. So the fraction of firms that choose the upper bound of the number of apprentices before and after the policy get weight, $\omega_j^k = \frac{1}{4} \frac{1}{2}$ and that the fraction of firms that pay the fee, $\omega_j^k = 1/4$.

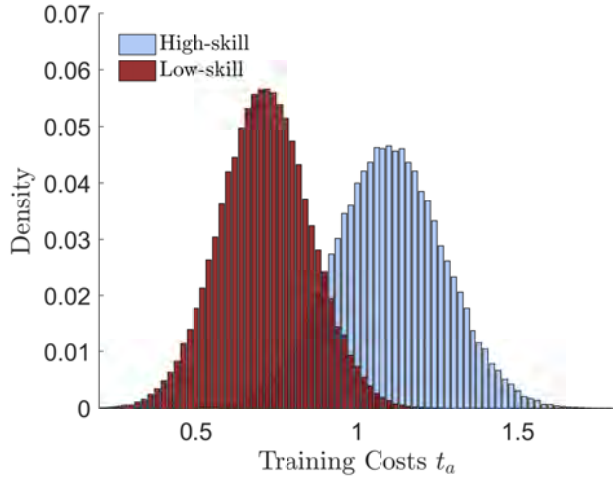
Table D.4: Moment Weights

Weight	Moment Description
$\omega_j^k = \frac{1}{4} h_{b(j)}$	Bunching points weighted by relative probability of pre-reform firm size distribution.
$\omega_j^k = \frac{1}{4} \frac{1}{3} h_{m(j)}$	Missing mass points divided by bin size 3.
$\omega_j^k = \frac{1}{4} \frac{1}{2}$	Fraction of firms choosing upper bound of apprentices before and after the policy.
$\omega_j^k = \frac{1}{4}$	Fraction of firms paying the fee.

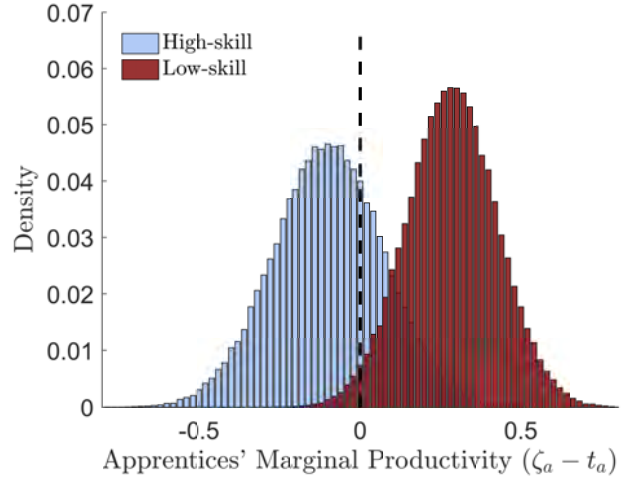
D.4 Truncated Normal and Uniform Training Cost Distribution

In this section we show the goodness of fit of the estimated model supposing a truncated normal (Figure D.4) or a uniform (Figure D.4) training cost distribution.

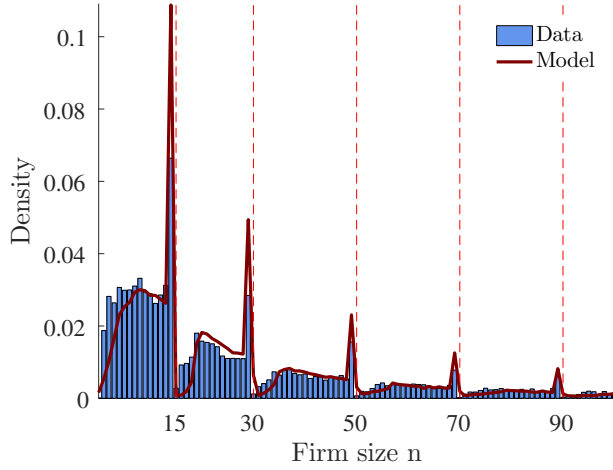
Figure D.3: Truncated Normal Training Distribution and Facts



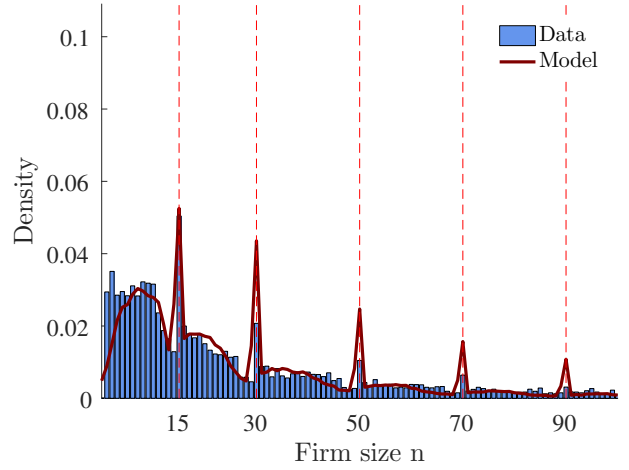
(a) Net Training Cost Distribution



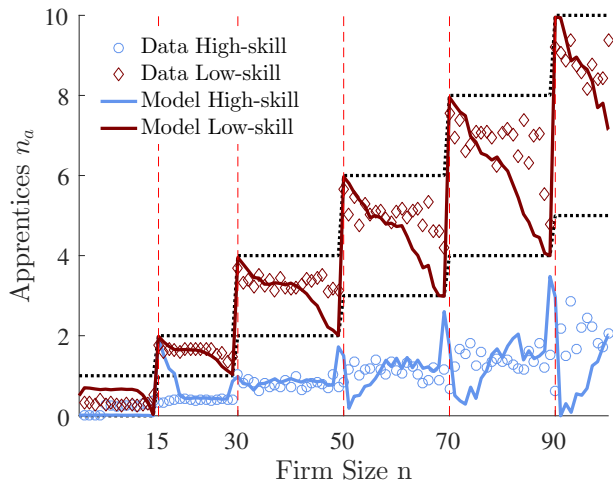
(b) Apprentices' Marginal Productivity Distribution



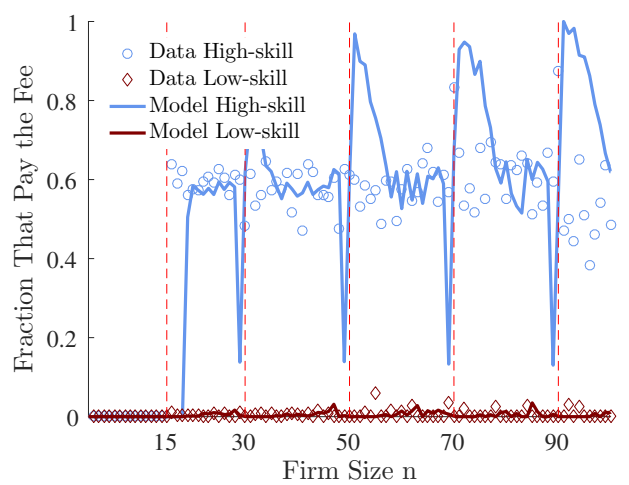
(c) Post-Reform High-skill Size Distribution



(d) Post-Reform Low-skill Size Distribution

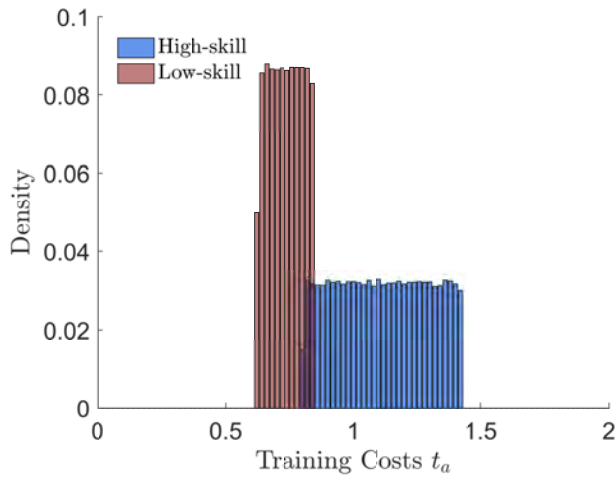


(e) Number of Apprentices

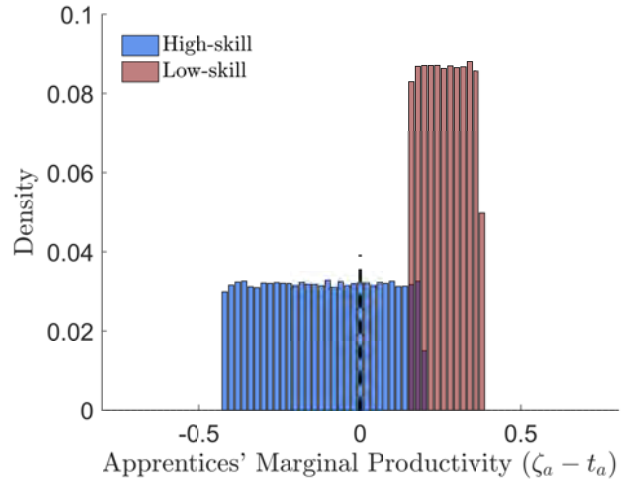


(f) Fraction that Pay the Fee

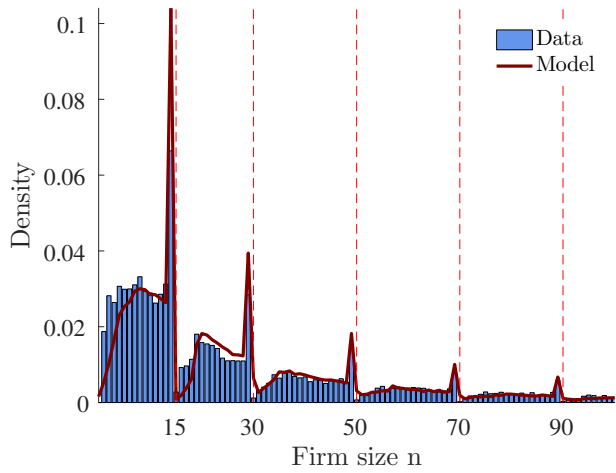
Figure D.4: Uniform Training Distribution and Facts



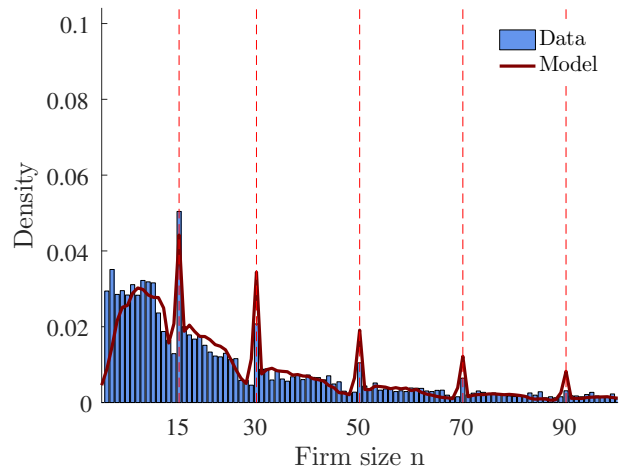
(a) Net Training Cost Distribution



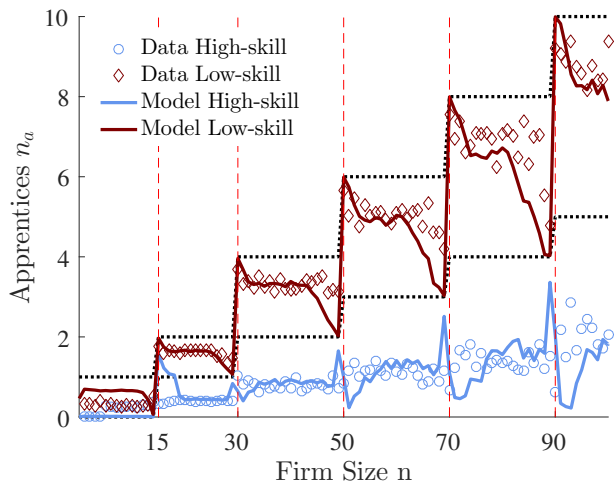
(b) Apprentices' Marginal Productivity Distribution



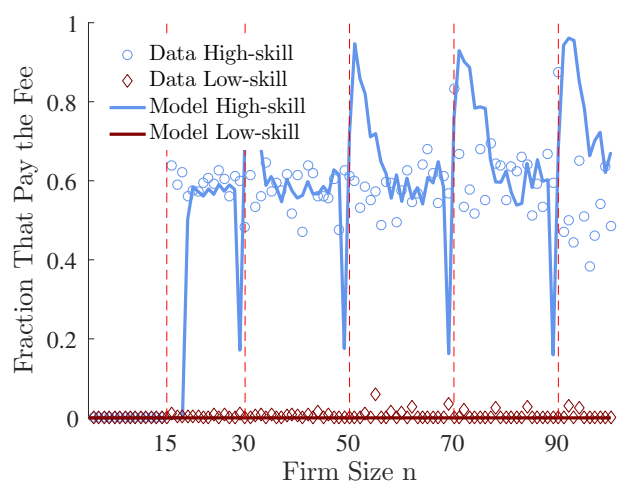
(c) Post-Reform High-Skill Size Distribution



(d) Post-Reform Low-Skill Size Distribution



(e) Number of Apprentices



(f) Fraction that Pay the Fee

D.5 Detailed Effect of Decomposition on Aggregate Variables

In this appendix we show the effects of each of the elements of the policy on aggregate variables.

Table D.1 reports this decomposition by looking at four combinations of the policy main components. First, when there are only apprentice quotas. Second, when there is only a reduction in apprentice wages. Third, when there is both a reduction in wages and there are apprentice quotas. And finally, when we have the when we consider the “full” policy that reduce wages, have apprentice quotas and allow the possibility of paying the fee. We show that with only the reduction in wages, firms with low training costs substitute many of their workers using labor from apprentices. Adding the quotas attenuate the displacement of workers by establishing a maximum bound in the number of trained apprentices. The minimum quota, on the other hand, mandates firms in the high-skill sectors to also train apprentices. Finally, the possibility of paying the fee lessens the negative effects for those firms with very high net training costs. However, as we emphasize above the apprenticeship policy induces labor and sector distortions that might be undesirable, suggesting there is room for improvement.

In Panel D.1A we show the effect with a policy of only the firm-size based quotas. These quotas harm all firms in the economy as it limits their choice of the number of apprentices. High-skill sector firms try to avoid apprentices more than low-skill sector firms, as training is more costly for them. Even though firms in both sectors train more apprentices than the workers they displace, there is a decrease in aggregate output due to the cost of training these apprentices. Profits also decline for both sectors, but the fall is more pronounced for the high-skill sectors.

Panel D.1B shows that, as expected, if only the apprentice wage is lowered, firms substitute a substantial number of workers for apprentices. The demand for workers falls more than 6% when aggregating the effect across both sectors. Low-skill sectors exhibit a larger substitution, lowering the number of workers by 11.7% and sharply increasing the apprentice intake more than six times the number of displaced workers. This strong substitution comes from the corner solutions in the linear model, as firms that benefit from lower wages only hire the necessary workers to train the optimal number of apprentices. In contrast, the effects for high-skill sector firms are mild, with only a handful of firms voluntarily training apprentices. Consistent with these labor responses, output and profits increase significantly for low-skill sector firms, while they barely change for the high-skill sectors.

Panel D.1C shows that combining quotas with the reduction in apprentice wage lessens some of the negative effects of the quotas on firms, but reduces the training of apprentices. The number of trained apprentices rises with respect to the case with only quotas (Panel D.1A), particularly in the low-skill sectors, where firms have more appetite for apprentices. The number of apprentices trained in both sectors exceeds the number of dislodged workers. Firms in the high-skill sectors have fewer incentives to reduce the number of workers to avoid the policy as the apprentices become relatively more attractive relative to the case of only quotas. There is a small positive effect on output for the low-skill sectors, but a loss in aggregate production for the high-skill sectors.

In Panel D.1D we introduce the possibility of paying the fee. This cuts down by half the number

of apprentices trained in the high-skill sectors. The high-skill sector firms experience smaller losses relative to the policy with no fees. As almost no firms in the low-skill sectors pay the fee, all the variables of this sector stay virtually the same than in the case without the fee.

Table D.1: Policy Decomposition: Aggregate Variables

	Workers	% Workers	Apprentices	% Output
A. Only Quotas				
High-skill	-6845	-3.21	9080	-2.04
Low-skill	-3819	-1.81	9518	-0.40
Total	-10664	-2.51	18598	-1.32
B. Only $\downarrow w_a$				
High-skill	-239	-0.11	4329	0.26
Low-skill	-26124	-12.38	153747	4.46
Total	-26362	-6.21	158076	2.12
C. Quotas + $\downarrow w_a$				
High-skill	-5455	-2.56	9413	-1.62
Low-skill	-3328	-1.58	17659	0.35
Total	-8783	-2.07	27073	-0.75
D. Full Regulation				
High-skill	-1964	-0.92	4527	-0.39
Low-skill	-3304	-1.57	17622	0.36
Total	-5268	-1.24	22149	-0.06

Notes: Column (1) is the change in the number of workers, Column (2) the percentage change of workers, Column (3) the change in the number of trained apprentices and Column (4) the percentage change in aggregate output.

D.6 Additional Results on Counterfactual Exercises

Subsidizing Apprentices' Wages $(1 - \varsigma)w_a$

The linearity in the model implies corner solutions. Lemma 6 characterize this solution,

Lemma 6. *Suppose a firm (t_a, z) solves (5),*

i. If $(1 - \varsigma)w_a + w(1 + \tau)t_a > w(1 + \tau)$, then the firm avoids apprentices choosing,

$$n^* = \left(\frac{p\gamma}{w(1 + \tau)} \right)^{1/(1-\gamma)} z, \quad n_a^* = 0.$$

ii. If $(1 - \varsigma)w_a + w(1 + \tau)t_a < w(1 + \tau)$, then the firm trains apprentices choosing,

$$n^* = t_a n_a^*, \quad n_a^* = \left(\frac{p\gamma}{\tilde{w}} \right)^{1/(1-\gamma)} z, \quad \text{where } \tilde{w} := w_a(1 - \varsigma) + w(1 + \tau)t_a.$$

Case (i) reflects when the cost to train apprentices is higher than the total cost of hiring workers, once the tax and the subsidy are taken into account. This tax scheme harms firms with high t_a given they cannot write-off those expenses. On the other hand, if the tax and the subsidy are high enough it incentive firms to train apprentices. This is case (ii) in Lemma 6. The demand for apprentices depends on whether the cost of training covers all the monetary training expenses.

Additionally we want to ensure the policy has a balanced budget. The total revenue coming from the payroll taxes is $Rev(\tau, \varsigma) := \sum_k \tau w^k N^k(\tau, \varsigma)$, where k denotes the sectors and N^k is the aggregate number of workers. The total subsidy paid by the government to firms is equal to $Sub(\tau, \varsigma) := \sum_k \int \int \mathcal{S}^*(t_a, z; \tau, \sigma) d\mathcal{Z}^k(z) d\mathcal{T}^k(t_a)$, where $\mathcal{S}^*(t_a, z; \tau, \sigma)$ denotes the optimal subsidy decision of firm (t_a, z) when facing taxes and subsidies (τ, σ) . For the exercises below, we consider the set of policies (τ, ς) that are budget balanced, $Rev(\tau, \varsigma) - Sub(\tau, \varsigma) = 0$.

Sector-Specific Apprentice Minimum Wage

In this section we describe the details of computing the sector-specific minimum wage counterfactual policy. Let $n_a^{k,min}(z, t_a; w_a^k, w^k)$ denote the solution to the firm maximization problem when the wage of workers is w^k and the apprentice wage is w_a^k . For each sector k , we compute the minimum wage for apprentices w_a^{*k} that solves,

$$N_a^{*k} = \int \int n_a^{k,min}(z, t_a; w_a^{*k}, w^k) d\mathcal{Z}^k(z) d\mathcal{T}^k(t_a), \quad (8)$$

where N_a^{*k} denotes the aggregate number of apprentices trained in sector k when implementing the original policy. We assume that each sector has the same worker wages as before and take the structural parameters from our estimated benchmark model.

For the linear labor input, a firm (t_a, z) trains apprentices if the total costs of training is smaller than hiring workers, $n_a^{k,min} = \left(\frac{\gamma^k}{w_a + w^k t_a}\right)^{1/(1-\gamma^k)} z$ and $n^{k,min} = t_a n_a^{k,min}$.

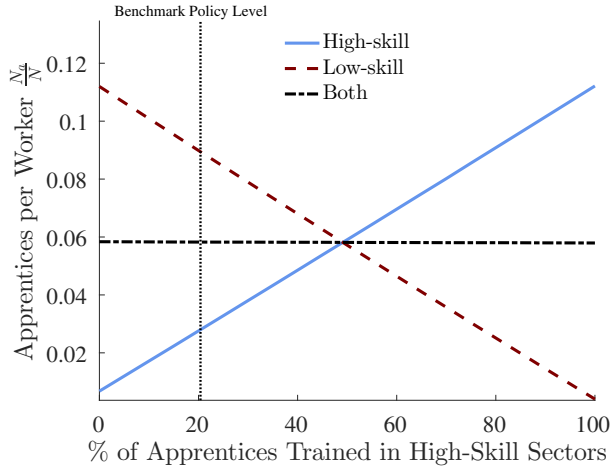
Conversely, the firm chooses not to train apprentices if $w_a + w t_a > w$, $n_a^{k,min} = 0$ and $n^{k,min} = \left(\frac{\gamma^k}{w}\right)^{1/(1-\gamma^k)} z$.

We solve equation (8) numerically and get an apprentice wage of $w_a^{*s} = 0.72$ for high-skill sectors and of $w_a^{*u} = 94$ for low-skill sectors.

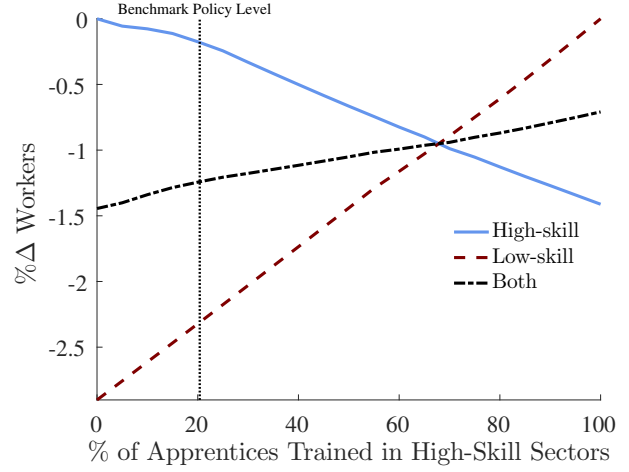
Sector-Specific Apprentice Minimum Wage and Additional Outcome Variables

In Figure D.5 we target the same total number of apprentices as in the baseline policy, but change the composition of the apprentices across sectors. We show the effect of changing the fraction of high-skill apprentices on the number of apprentices per worker, the displaced workers, total output and total welfare.

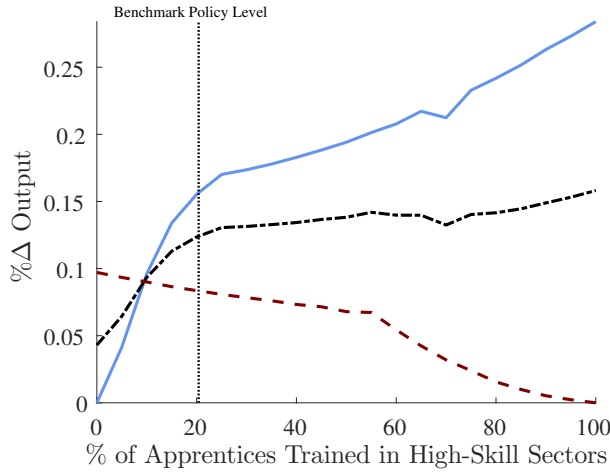
Figure D.5: Varying the Share of Training in High-Skill Sectors



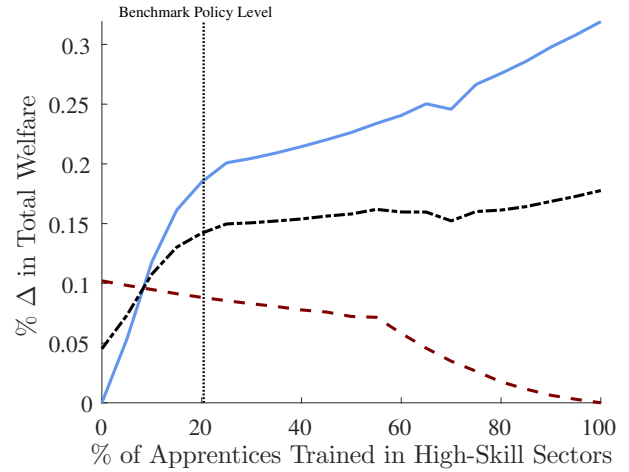
(a) Apprentices per Worker



(b) Change in Workers



(c) Change in Output



(d) Change in Welfare

Notes: Outcome variables when targeting different fractions of apprentices trained in high-skill sectors, holding the total number of apprentices fixed. Panels (a) shows the number of apprentices per worker, Panel (b) shows the percentage change in total workers, Panel (c) shows the percentage change in aggregate output and Panel (d) the change in aggregate welfare measured as the change in total aggregate utility relative to pre-reform utility, $\Delta U/U^*$. The solid line is for high-skill sectors, the dashed line for low-skill sectors.