Mandatory Apprenticeship Training in Firms*

Santiago Caicedo†, Miguel Espinosa‡ and Arthur Seibold§

June 16, 2020

Abstract

We quantify the effect of training apprentices on firms and aggregate welfare, exploiting a reform to apprenticeship regulation in Colombia. The reform mandates training in firms by setting minimum and maximum apprentice quotas that vary discontinuously in the number of full-time workers. We document strongly heterogeneous firm responses across sectors, revealing differences in the net cost of training. In sectors with high skill requirements, firms decrease their size and bunch just below the regulation thresholds to avoid training apprentices. In contrast, firms in low-skill sectors increase their size to qualify for more apprentices. Guided by these reduced-form findings, we develop a structural model and find small static effects on aggregate output despite the sizeable labor input responses. Yet, our results indicate large benefits to both firms and apprentices when training increases the future supply of productive workers. Finally, we show that counterfactual policies that consider heterogeneity across sectors can deliver similar benefits from training while inducing fewer distortions in the firm size distribution and in the allocation of resources across sectors.

Key words: Training, Apprentices, Firm-size distortions
JEL Codes: E24, J21, J24.

*This project benefited from discussions with Ufuk Akcigit, Benito Arruñada, Leonardo Bursztyn, Jonathan Dingel, Christian Dustmann, Jan Eeckhout, Rosa Ferrer, Luis Garicano, Alessandro Gavazza, Maria Guadalupe, Ali Hortacsu, Greg Kaplan, Camille Landais, Bob Lucas, Rocco Macchiavello, Sandra McNally, Magne Mogstad, Joan Monras, Massimo Motta, Luis Rayo, Mar Reguant, Esteban Rossi-Hansberg, Uta Schönberg, Sandro Shelegia, Alexandra Spitz-Oener, Claudia Steinwender, Nancy Stokey, John Sutton, Chad Syverson and numerous conference and seminar participants as well as SENA authorities that solved many of the questions we had during the process of writing this article.

†University of Chicago and Universidad de los Andes, scaiceso@gmail.com, scaicedosoler@uchicago.edu

‡Universitat Pompeu Fabra & Barcelona GSE, miguel.espinosa@upf.edu

§University of Mannheim, seibold@uni-mannheim.de

1 Introduction

Apprenticeships are gaining popularity around the world. These training programs combine formal education, usually at vocational education institutions, and training *within* firms.[1] Probably the most well-known examples are the dual systems in German-speaking countries, where there is a long-standing tradition of vocational education and apprenticeships. In Germany, for instance, more than two thirds of young individuals complete an apprenticeship (Steedman 2010, 2012). Both in developed and developing countries worldwide, there is growing interest in expanding apprenticeship training. In the US, a country with historically little vocational training, there have been more than 500,000 new apprentices since the Presidential Executive Order "Expanding Apprenticeships in America" in 2017.[2] Apprenticeship programs typically aim at improving labor market outcomes for young, inexperienced workers, in particular those without a college education. Apprenticeship policies have been effective in tackling some of the most prevalent issues these groups face in labor markets, such as youth unemployment and high informal employment rates (Ryan 2001, Zimmermann et al. 2013, Fazio et al. 2016).

Firms play a crucial role in these apprenticeships, providing training and facilitating knowledge transfer between workers and apprentices. Yet, firms 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. Apprenticeship programs can mandate firms to train apprentices, but also lower the net training costs to firms, for instance by improving the skills of potential apprentices during the formal education phase or by providing financial incentives to firms. Despite the ubiquity of apprenticeship programs, there is surprisingly limited evidence on their effects. The existing literature on the topic mainly studies the benefits of training to apprentices (Fersterer et al. 2008, Göggel and Zwick 2012), but there is scarce evidence of the impact of apprenticeships on firms and aggregate outcomes.

In this paper, we aim at filling this gap by quantifying the effects of training apprentices on firms and implications for aggregate welfare. To do so, we exploit a change in apprentice regulation in Colombia that requires all firms (with at least 15 full-time workers) to train apprentices. This reform provides ideal variation to capture firms' willingness to train apprentices, altering their incentives in three ways. First, it establishes a minimum and a maximum number of apprentices as a function of the firm's number of full-time employees. Second, it reduces apprentices' minimum wage, making it cheaper for firms to have apprentices. Third, it allows firms to "buy themselves out" of training by paying a per-apprentice fee that is nominally larger than the apprentices' minimum

[1] Wolter and Ryan (2011) define 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).

[2] For information on recent efforts to support apprenticeships in the US see https://www.apprenticeship.gov/.

wage. We combine reduced-form evidence on firm responses to the policy and a structural model to uncover the net costs of training apprentices for firms in different sectors. Based on the estimated model, we quantify the effects of the policy on aggregate outcomes, the welfare of apprentices and incumbent workers, and firm profits. Finally, our framework allows us to analyze counterfactual policies, highlighting the potential benefits of apprenticeship policies that take firm heterogeneity into account.

We begin by showing reduced-form evidence of the effects of the policy. 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 present three empirical facts that demonstrate 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 have 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. In particular, the third fact implies that apprentices have negative marginal productivity in many high-skill sector firms, which disciplines the production function in the theoretical model.

Guided by these empirical results, we develop a simple model of firm production featuring heterogeneous costs of training apprentices. With this model we recover the unobserved training cost distribution for firms, allowing us to quantify welfare effects of the policy. Inspired by Lucas (1978), we consider a two-sector economy with heterogeneous firms characterized by their managerial ability and net training costs. Firms in each sector 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. The net cost of training is a combination of the time it takes a worker to train them and the apprentices' relative productivity after training. This theoretical framework is able to capture the differential firm responses to the reform via heterogeneity in training costs across sectors. As in the data, high-skill sector firms with high costs of training avoid apprentices, while low-skill sector firms with low costs of training seek them. Moreover, we show that high-skill sector firms may prefer to pay fees, even if they are nominally larger than the apprentices' minimum wage, as apprentices can have negative marginal productivity when the opportunity cost of workers' time is high.

We estimate the structural parameters of the model via a three-step procedure using a combination of the pre-reform data, the policy specification, and firm responses to the reform. First, we estimate the output elasticity of full-time labor by sector using the pre-reform data and the model's production function. Second, we back out the parameters of the productivity distribution by matching the pre-reform firm size distribution by sector. Third, we estimate the non-parametric training cost distribution for each sector by simulated method of moments (SMM), using the policy specification and targeting moments that correspond to the key firm responses from the reduced-form analysis: the fraction of firms that choose the maximum apprentice quota, the fraction paying the fee, and the firm size distribution after the reform with the excess mass of firms bunching below and at the regulation thresholds.[3] From these moments, we can identify the structural parameters of the model and infer the net training cost distribution for each sector. As expected from the reduced-form results, we find that firms in low-skill sectors have lower average training costs than those in high-skill sectors.

Next, we use the estimated model to quantify the effects of the regulation on aggregate outcomes and welfare for three scenarios: i) a partial equilibrium scenario, where wages are fixed, ii) a general equilibrium scenario, where wages adjust and displaced workers are absorbed by labor markets, and iii) a dynamic scenario, where trained apprentices increase the supply of workers in future periods. In the partial equilibrium scenario, there are small effects of the policy on aggregate output. Intuitively, firms re-optimize and substitute labor from workers to apprentices. At the margin, although there is a significant number of displaced workers, total labor input (apprentices plus workers) does not change much, so output is relatively stable. The aggregate production of low-skill sector firms increases by around 0.4%, while for high-skill sector firms production decreases by 0.4%. We provide some additional reduced-form evidence of the reform on firm outcomes that supports these partial-equilibrium quantitative results. Using a difference-in-difference type methodology to compare firms above and below the regulation thresholds, we show that despite large labor substitution effects, the impact of the policy on firm-level output is indeed modest.

[3]For a discussion of SMM, see Gourieroux et al. (1996).

Moreover, the model allows us to quantify general equilibrium and dynamic effects. In these scenarios, the impact on output can become substantial. In general equilibrium, displaced workers are absorbed by the labor market, which drives wages down by 1.2%. Aggregate production in low-skill sectors increases by 1.3%, but also in high-skill sectors the initial negative effects of the policy on output are reversed. Moreover, these positive effects on output are amplified when trained apprentices increase the future supply of workers. Such dynamic effects lead to a total increase of 3% in aggregate production, 6.2% in low-skill sectors and 1.5% in high-skill sectors.

Hence, the policy has a positive aggregate impact, but further unpacking the effects suggests that it creates winners and losers. 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, with a more than 2% decrease in profits for more 45% of firms. These negative effects vanish when considering general equilibrium and dynamic effects however, where profits in both low-skill and high-skill sectors increase. Next, trained apprentices are among the winners of the regulation, as they earn higher wages in the future. If apprentices' outside option is sufficiently low, they may even benefit in the short-run.[4] The main losers from the regulation are incumbent trained workers, especially in low-skill sectors where most apprentices are trained. In the short term, around 1.2% of regular workers are displaced by the newly hired apprentices, while in general equilibrium they see a decrease in wages.

Finally, we use our model to study policy counterfactuals. First, we decompose the effects of the policy into its three main components: minimum and maximum quotas, the decrease in apprentices' minimum wages, and the possibility of paying a fee. Results indicate that each component plays an important role. Mandating a minimum quota pushes firms in all sectors to train apprentices. Without the minimum quota, high-skill sector firms would train less than half of the apprentices of the full regulation, even when apprentices' minimum wages are cut. 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. Without this maximum, the low-skill sector firms would hire six times more apprentices. The possibility of paying fees in turn reduces the negative effect on firms with very high training costs and generates revenue for the government, but naturally limits training in high-skill sectors.

Second, we study whether alternative policies can deliver similar benefits from training but induce less distortions in firm sizes and sectoral allocation. In particular, we focus on two policies that remove the discontinuous firm-size based quotas and instead resort to price-based instruments. To discipline these exercises, we fix the number of apprentices trained in order to achieve the same overall benefits from training. The first alternative policy is a subsidy on training costs financed

[4]The outside option may be unemployment (without insurance, as they have typically not contributed to social insurance yet) or working in the informal sector at a low wage (Gaviria and Mendez 2003). The excess supply of apprentices reported by SENA (2018) supports the view that individuals perceive benefits from apprenticeships.

by payroll taxes, similar to the UK “Apprenticeship Levy”.[5] The subsidy policy indeed avoids firm size distortions, but it still induce the sectoral distortions and actually reduces aggregate welfare by 0.1%. Intuitively, the subsidy only partially covers expenses for firms with high training costs, which means that most of the high-skill sector firms end up paying the tax without training apprentices, effectively cross-subsidizing the low-skill sector. As a second counterfactual exercise, we consider a sector-specific minimum wage 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. This policy also avoids firm size distortions, and additionally reduces the sectoral reallocation, harming high-skill sector firms less. There is an aggregate welfare gain of 0.1%, the highest among all policies considered. Overall, we conclude that policies that takes account of the heterogeneity in training costs across sectors can further increase the positive welfare effects of apprenticeship regulation.

Related Literature

This paper contributes mainly to three strands of literature. First, a rich literature empirically analyzes the effects of training and apprenticeships, focusing mostly on the benefits to apprentices.[6] In developed countries, studies tend to find positive effects on labor market outcomes of apprentices. For instance, Krueger and Pischke (1995), Fersterer and Winter-Ebmer (2003) and Fersterer et al. (2008) estimate the returns of apprenticeship programs in Germany and Austria and find them to be similar to those of other types of schooling. However, these effects depend to some extent on the quality and specific characteristics of apprenticeship programs (Soskice 1994) and vary across occupations (Göggel and Zwick 2012). There are few studies on apprenticeship training in developing countries, with the exception of Corseuil et al. (2019) who show that apprenticeships increase the probability of formal employment in Brazil.[7]

However, there is little evidence of the effect of apprenticeships on firm outcomes. Mohrenweiser and Zwick (2009) and Cappellari et al. (2012) study the impact of apprentices on firm performance, with somewhat mixed results. Other studies try to infer training costs directly from firm surveys (e.g. Dionisius et al. 2009, Wolter et al. 2006). The most closely related work to this paper is by Ospino (2018) who also studies the Colombian apprenticeship regulation using survey data. He

[5]The Apprenticeship Levy requires UK employers with an annual payroll of more than £3 million to pay a 0.5% payroll tax. Using these funds, plus a 10% contribution from the government, firms are reimbursed for apprenticeship training costs.

[6]See Ryan (1998, 2001) and Wolter and Ryan (2011) for detailed reviews of the literature on apprenticeships.

[7]Other work studying the effect on apprentices includes Bonnal et al. (2002), Groot et al. (1998), McIntosh (2007), Adda et al. (2010), Riphahn and Zibrowius (2016), Parey (2016) and Albanese et al. (2017). More broadly, there is a range of papers evaluating the returns to various forms of vocational training, finding mixed results depending on the quality of training and institutional arrangements (e.g. Attanasio et al. 2011, Bertrand et al. 2019).

finds small positive effects of the policy on firms' productivity.[8] We make several novel contributions to this literature. First, we leverage high-quality administrative data and a unique institutional setting to document strong heterogeneity in firm responses to apprenticeship regulation. Second, we structurally estimate the training cost distribution and argue that the heterogeneity in these costs is key to interpret firm responses. Third, we are able to quantify aggregate welfare effects of actual and counterfactual apprenticeship policies.

Second, our paper relates to the theoretical literature on training and human capital accumulation in firms. In his seminal work on human capital, Becker (1964) argues that firms are not willing to invest in general training whose returns they cannot appropriate. Acemoglu and Pischke (1998, 1999) show that under imperfect information and imperfect competition, firms may have incentives to provide some general training, but the level of training can still be inefficient.[9] Our empirical findings suggest that labor market restrictions, such as high minimum wages for apprentices, can be an important additional source of inefficient training. In our model, we highlight the role of heterogeneity in training costs, and we quantify welfare gains of apprenticeship programs that modify labor market regulation for apprentices.

Finally, 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 welfare and counterfactual analysis, we go beyond quantifying distortions and study the trade-off between the gains from increased training and such 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, section 5 shows the model estimation, and finally section 6 concludes.

2 Data and Institutional Context

Data

We use a novel administrative data set provided by the Colombian National Service of Vocational Education (henceforth SENA). SENA is the governmental institution overseeing technical and vocational training, including the implementation and enforcement of the apprenticeship program. More than 80% of apprentices receive the class-room portion of their training directly at SENA,

[8] Several studies analyze the vocational education system in Colombia more broadly, including Gaviria and Mendez (2003), Barrera-Osorio and Corchuelo (2003), Attanasio et al. (2015) and Bettinger et al. (2019).

[9] Garicano and Rayo (2017) and Fudenberg and Rayo (2017) consider other additional sources of training inefficiencies. They show that skilled workers can have incentives to transfer knowledge inefficiently slowly to apprentices in a dynamic framework.

where around 400,000 students per year are trained (SENA 2018). In addition, SENA supervises compliance with the apprenticeship regulation by firms and oversees vocational education programs offered by other institutions.

Our data covers the universe of manufacturing firms with at least 10 workers from 1995 to 2009.[10] 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. 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).[11] The survey data includes additional information on workers, production inputs, wages, sales, output, and costs.

Institutional Context and Apprenticeship Regulation

Before the 2003 reform, firms were required to train apprentices, but there was no minimum apprentice quota and the regulation was weakly enforced. Only a maximum number of apprentices of no more than 5% of the firm's total labor force was stated. Since 1999, SENA was responsible for determining the apprenticeship contract conditions, including the number of apprentices firms should have, the characteristics of the training program and sanctions for not complying. However, the regulation was hardly enforced in practice. The most prevalent way of complying was assigning regular workers to evening courses taught by SENA, without actually training new apprentices (Ospino 2018). As we show below, prior to 2003 barely any firms trained apprentices.

With the 2003 reform, a new apprenticeship regulation came into effect. This *firm-size* based apprenticeship policy was part of major labor reform with the aim of reducing the high levels of informal employment.[12] Among all the reform measures, only the apprenticeship policy depends on firm size. The reform establishes a dual vocational training system with two phases, the *teaching* phase where apprentices are taught full-time in a formal education institution, and the *productive* phase where they receive training in the firm.[13] Any individual with basic secondary education (8th grade) can apply for apprenticeships.

The post-reform apprenticeship regulation has three main components:

1. **Apprentice Quotas:** First, *apprentice quotas* depending on the number of full-time workers

[10] SENA 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.

[11] The data is collected by DANE at the establishment level and then aggregated at the firm level.

[12] See law 789 of 2002.

[13] The teaching phase lasts from one year (1760 hours) to 1.5 years (2640 hours) depending on the occupation. The productive phase lasts between 6 months and 2 years.

Table 1: Summary Statistics

	Full Sample			*Threshold Sample*
	All Sectors	Low-Skill	High-Skill	
Apprentices	1.08	1.72	0.50	0.74
	(4.12)	(5.58)	(1.93)	(1.80)
Workers	49.08	49.54	48.68	29.82
	(94.72)	(101.26)	(88.49)	(30.56)
Workers (Survey)	54.58	56.22	53.12	31.23
	(127.33)	(139.16)	(115.77)	(34.52)
Fraction Professionals	0.08	0.06	0.10	0.07
	(0.14)	(0.12)	(0.15)	(0.12)
Fraction Admin Wrks	0.37	0.38	0.36	0.33
	(0.26)	(0.27)	(0.25)	(0.21)
Fraction Production Wrks	0.55	0.56	0.54	0.60
	(0.27)	(0.28)	(0.27)	(0.24)
Output	11,031,021.10	10,679,501.60	11,343,706.21	5,579,873.15
	(28,153,949.13)	(27,074,390.85)	(29,077,246.56)	(12,491,211.57)
Value Added	4,755,892.13	4,412,604.88	5,057,618.25	2,243,634.14
	(12,739,041.21)	(12,081,762.97)	(13,282,717.54)	(5,575,936.32)
Profits	3,113,408.54	2,893,555.07	3,306,644.76	1,456,770.80
	(9,359,584.31)	(8,868,685.64)	(9,766,703.28)	(4,365,585.76)
Output per Worker	215,891.13	232,204.13	201,965.00	178,345.72
	(389,911.61)	(423,867.00)	(357,804.05)	(326,948.69)
Wage bill (perm. wrks.)	1,240,527.16	1,115,178.08	1,352,028.20	590,290.50
	(3,048,647.20)	(2,839,862.38)	(3,218,945.32)	(1,078,796.53)
Total wage bill	1,609,439.22	1,463,753.22	1,739,030.45	773,079.80
	(3,702,964.12)	(3,430,131.11)	(3,925,257.51)	(1,354,138.75)
Wage p.w. (perm. wrks.)	19,732.88	18,383.78	20,884.58	17,104.21
	(17,763.45)	(17,649.47)	(17,779.53)	(13,058.81)
Capital/Output	0.66	0.64	0.67	0.69
	(0.88)	(0.91)	(0.85)	(0.91)
Intermediates/Output	0.54	0.56	0.53	0.55
	(0.18)	(0.19)	(0.18)	(0.18)
Observations	108,385	51,024	57,361	14,848
Firms	14,586	7,403	7,986	2,018

Note: All monetary variables in 2009 thousands of pesos.

at a firm are established. The quota sets 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 required number for firms with more than 15 workers.

2. **Apprentices' Minimum Wage:** The reform also lowers the minimum wage for apprentices as an incentive for firms to train. Apprentices can be paid at least 50% of the regular minimum wage during the teaching phase and 75% during the productive phase.[14]

3. **Fee:** Finally, firms can pay a fee instead of hiring the minimum required apprentices. The total amount paid is proportional to the difference between the minimum apprentice quota and the number actually hired by the firm. Nominally it amounts to 100% of the minimum wage per missing apprentice.

The new apprenticeship regulation is strictly enforced by SENA. Whenever firms are found non-compliant with the apprentice regulation, 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 to SENA the total number of hours worked by regular workers every 6 months. Using this information, the equivalent number of full-time workers is calculated by averaging the hours of these workers.[15] 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. During the training, apprentices report their satisfaction with the apprenticeship program twice a month, and apprentices can be reallocated if they are dissatisfied with the quality of the training received at the firm or the firm does not find satisfying the performance of the apprentice.

The database used in this paper does not contain much information about apprentices' characteristics. To learn more about them, we have used different datasets that contain information of workers as well as their current or past apprenticeships status. Unfortunately, these datasets do not coincide with our period of analysis as one dataset is for 2013 and 2015, while the other is for

[14] Apprentices who get education from a university, as opposed to vocational education institutions, are paid a full minimum wage in both phases. The regulation also specifies that if the unemployment rate falls below 10%, then all apprentices have to be paid a full minimum wage. This only happened after 2013.

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

2015-2016. With these new datasets we show that apprentices are young, mid-high socioeconomic status, female low-paid workers.[16] They are in average 24 years old, 57% female and earn around the minimum wage required for apprentices during this period (equal to 100% minimum wage minus some payroll taxes). They tend to be trained by firms with larger labor force, larger fraction of female workers that pay larger wages and tend to be enrolled or have finished technical degrees in social sciences and business education.

The same month of graduation their income increases to 1.4 minimum-wages, a similar average income than a group of workers with the same socio-demographic characteristics has. Most of the apprentices are trained in a low-skill sector firm. After graduation, 75% of them, remain in the same firm. When they leave the training firm, they stay in the same skill sector 60% of the time. After graduation, their average wage is the largest when they are hired by firms different from the ones that provide the training but belong to the high-skill sector. Finally, consistent to a story of an excess supply of apprentices, workers' and not apprentices' earnings increase with age and firm size.

We have build an appendix with this and more information about the characteristics of the apprentices. The reader can find more details on this analysis in the *Online Appendix: Characteristics of the Apprentices*.

3 Empirical Analysis

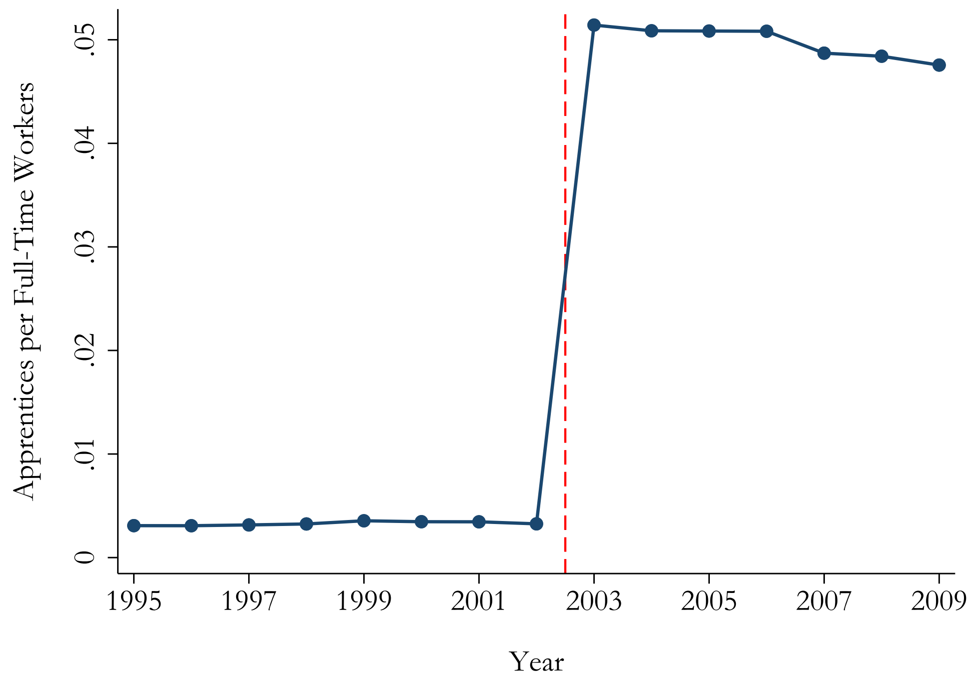
In this section we show the aggregate effects of the policy on the number of apprentices and the distribution of full-time workers. These results focus on the heterogeneity in firm responses across sectors, splitting the data into high-skill and low-skill sectors.

3.1 Aggregate Effects: Number of Apprentices and Firm Size Distribution

The primary objective of the apprentice regulation is to increase the extent of apprenticeship training. Figure 1 shows the policy is successful in this dimension, 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.

[16] This is aligned to the findings of Gaviria and Mendez 2003.

Figure 1: Average Number of Apprentices per Full-Time Worker



Notes: Full sample, all manufacturing firms with at least 10 workers. *Source*: own computations using data compiled by SENA.

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 2 shows that such distortions in the post-reform number of full-time workers are sizeable.[17] In the pre-reform years 1995 to 2002 the firm size distribution is relatively smooth. In contrast, the distribution becomes rugged in the post-reform years 2003 to 2009, with pronounced spikes around the regulation thresholds marked by the dashed vertical lines, and holes or "missing mass" on both sides of the thresholds. The figure provides first visual evidence that some firms change their labor inputs in response to the policy. Moreover, the fact 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.

[17] In Appendix A.1 we present year-by-year distribution, showing the same patterns as the pooled distributions in Figure 2.

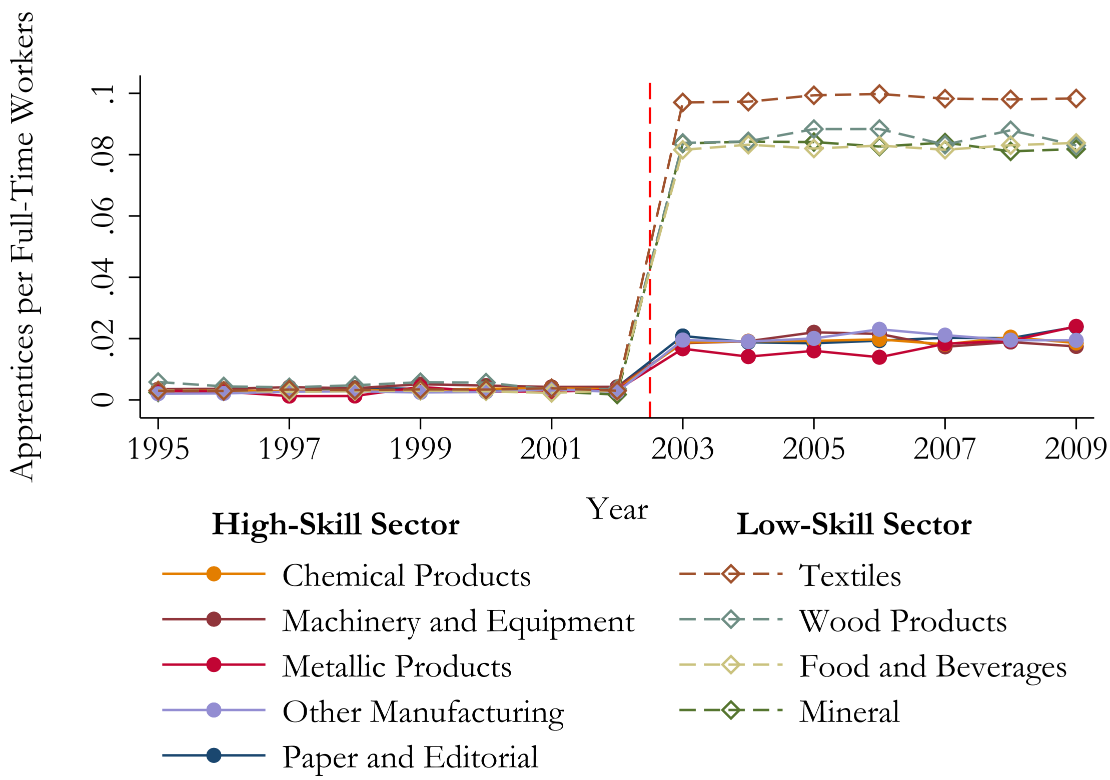
Figure 2: Firm Size Distribution Before and After Regulation

Notes: Distribution of full-time workers for all firms before (1995-2002) and after the reform (2003-2009), using a bin size of one. *Source*: own computations using data compiled by SENA.

To study the heterogeneity in firm responses, we divide firms into nine two-digit sectors using the Colombian industrial classification.[18] In Figure 3 we plot the number of apprentices per full-time worker in each sector. The figure suggests a sharp division in apprentice intake across sectors. In the wood products, textiles, food and beverage, and the 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.

[18] We use DANE's industrial classification, CIIU 3 A.C, which is adapted from the International Standard Industrial Classification (ISIC).

Figure 3: Number of Apprentices By Sector



Notes: Ratio of apprentices to full-time workers 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. *Source*: own computations using data compiled by SENA.

To understand these differences in apprentice intake, we rank sectors by the fraction of professional workers with tertiary education out of total 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 *low-skill sectors* as those below the median and *high-skill sectors* as those above the median. 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 the ones with the most apprentices in figure 3, whereas the five sectors classified as high-skill (paper and editorial, other manufacturing, machinery and equipment, metallic products, and chemical products) are the ones that take fewer apprentices.

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. This is reflected in substantially higher labor productivity measured by output per worker. High-skill sector firms pay almost 20% higher average wages, and they use more capital and less intermediate inputs relative to low-skill sector firms. In addition, table A.1 in Appendix A.2 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 variables that may be naturally related to skill requirements imply a similar ranking of sectors to the one from 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 ratio of capital to output. The individual ranking of sectors using these alternative measures remains relatively stable, with the exception of the mineral non-metallic sector that has relatively few observations.

3.2 Reduced-Form Evidence: Differential Responses Across Sectors

In this section, we provide reduced-form 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 that pay the fee in each sector.

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

Figure 4 presents the firm size distribution around the first five thresholds for firms in high-skill sectors (panel 4a) and in the low-skill sectors (panel 4b), pooled across 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 and 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.[19] 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 2. 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 4 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 particular.

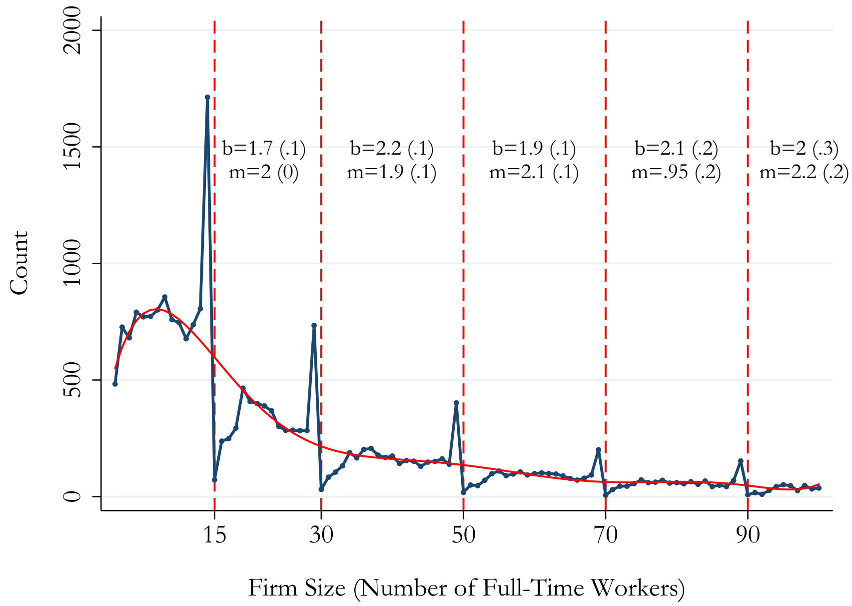
Fact 2: Firms in high-skill sectors tend to choose the minimum number of apprentices; firms in low-skill sectors tend to choose the maximum.

Figure 5 shows the number of apprentices by firm size in both 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, having an average number of apprentices below the minimum quota –some of these firms must be paying the fee or not complying with the regulation–, while firms in low-skill sectors tend to have an average number of apprentices close to the maximum quota.[20]

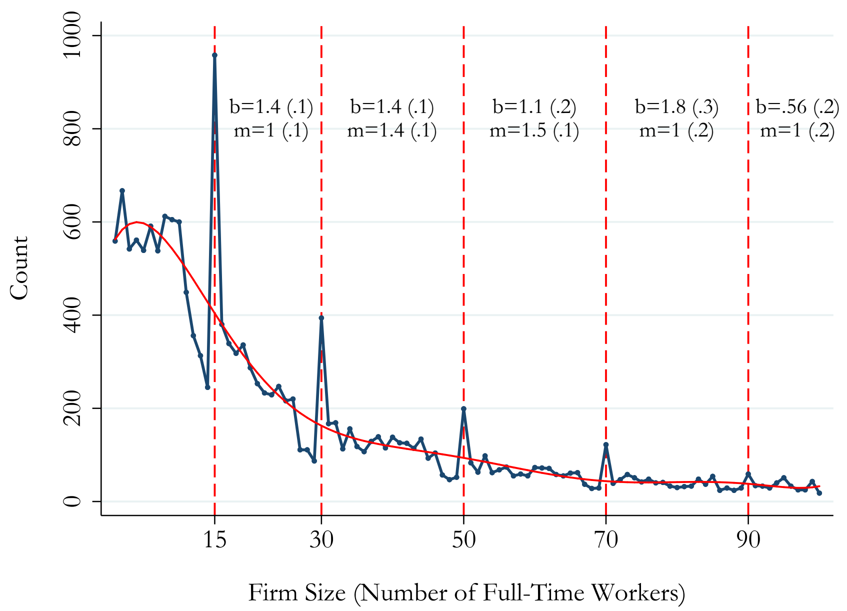
[19] Due to sharp bunching response, the bunching region is chosen as one bin below the threshold in panel (a) and the bin at the threshold in panel (b). The missing mass region is the threshold plus three bins to the right in panel (a), and the three bins below the threshold in panel (b). Results are very similar when using the convergence method of Kleven and Waseem (2013).

[20] In Appendix A.4 we show that the pre-reform number of apprentices by firm size is low and similar across all sectors. This suggests that the change in the apprentices' minimum wage is important to understand the differential responses across sectors.

Figure 4: Fact 1 - Bunching Responses in High-skill and Low-skill Sectors



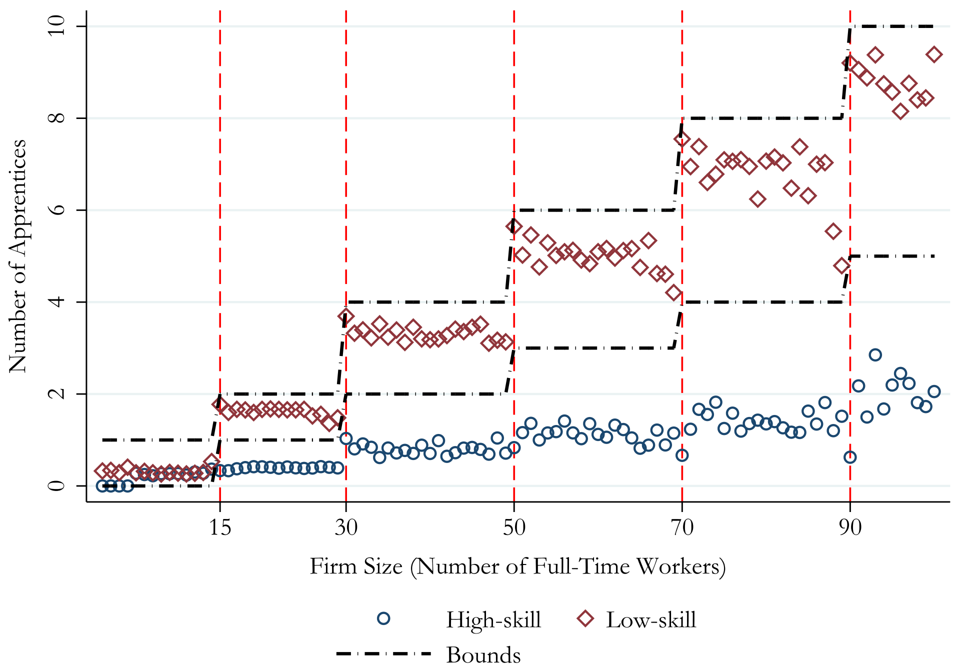
(a) High-skill Sectors



(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. *Source*: own computations using data compiled by SENA.

Figure 5: Fact 2 - Number of Apprentices by Firm Size



Notes: Average number of apprentices by firm size bin for high-skill and low-skill sector firms, post-reform (2003-2009). The dashed vertical lines denote the regulation thresholds. *Source*: own computations using data compiled by SENA.

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 selection and causal effects of the policy. 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. Those 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.

In Table A.2 in Appendix A.2 we study these responses in more detail by looking at the proportion of firms that choose 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, aligned to the results of this section, responses strongly differ across high-skill and low-skill sectors.

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 increase of the apprentice 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 for many firms it is now worthwhile training 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: High-skill sector firms more likely to pay the fee.

Figure 6 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 nominal fee (100% of the minimum wage). Thus, the responses of high-skill sector firms indicate that training costs must be high in these industries. In fact, apprentices must create an overall negative marginal benefit for these firms as they are unwilling to hire apprentices even though it is nominally cheaper than not hiring apprentices at all. Later on, this discussion will discipline the production functions used in our 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 Appendix A.3, we also show that very few firms end up paying fines. The vast majority of firms comply with the regulation by either taking the required number of apprentices, which is consistent with the strict enforcement described in section 2.

Appendix table A.3 shows regressions of indicators for the different types of responses to the regulation on firm characteristics. In order to not confound the effects with endogenous changes in these characteristics, the post-reform responses are regressed on the pre-reform firm-level average of each characteristic. This 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 sector. The correlation of responses with pre-reform firm size is very small in magnitude,

Figure 6: Fact 3 - Share of Firms Paying Fees by Firm Size

Notes: Fraction of firms that pay the fee by firm size bin in high-skill and low-skill sectors, post-reform (2003-2009). The dashed vertical lines denote the regulation thresholds. *Source*: own computations using data compiled by SENA.

suggesting that responses are similar across the firm size distribution.

Finally, appendix table A.4 shows the correlation of bunching behavior with the different types of responses to the regulation. In the post-reform period, there over 4000 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, on the other hand, tends to coincide with choosing the maximum number of apprentices, which 72% of bunchers above do.

4 Model: Heterogeneous Training Costs

We develop a model with heterogeneous firms that rationalizes our empirical findings and allows us to quantify the effect of the policy on firms and welfare. We suppose that firms differ in two dimensions: their net costs of training apprentices and their managerial ability. The net training costs come from a sector-specific distribution and explain the differential firm responses of our reduced-form evidence, while managerial ability gives rise to the firm size distribution in each sector. We consider an economy with multiple sectors and periods, where dynamics can stem from trained apprentices increasing the future supply of workers, but for simplicity we assume that firms' labor input choice is static.

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 net training costs t_a^k and managerial ability z^k, which come from sector-specific distributions $\mathcal{T}^k$ and $\mathcal{Z}^k$. We suppose these characteristics are invariant across time for all firms. Managerial ability z^k is an idiosyncratic characteristic of the firm that can also be interpreted as technological differences or other factors that affect a firm's productivity. Firms produce $y_t^k(z, t_a)$ units of good k in period t using labor, which is supplied either by workers l_t^k or by *apprentices* $l_{a,t}^k$.[21] We assume 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 the average time a worker spends training apprentices in sector k. Apprentices potentially have different productivity in the tasks they perform across firms and sectors. Differences in the training cost distribution across sectors reflect

[21] The model could be readily extended to include capital or other production inputs. Here we suppose labor is the only input to simplify the analysis and emphasize the role of apprentices. In Appendix B.3 we extend the model to allow for other endogenous inputs, showing that the results of the baseline model are qualitatively the same.

the differences in the time required to train apprentices and the relative productivity of workers and trained apprentices.[22] A sector whose technology requires simple menial tasks probably requires less of the workers' time to train apprentices. On the contrary, a sector with highly specialized tasks requires more training. Firms in different sectors also differ in their production technology $f^k(l_t^k, l_{a,t}^k; z^k)$. The production combines a firm's managerial ability z^k with the total labor supplied by both types of workers given the net training costs t_a^k. We suppose that the production function is increasing in these labor inputs $(l_t^k, l_{a,t}^k)$ and in the managerial ability of the firm. Note that depending on the way the production function combines both types of labor, the labor input from workers and apprentices can be complements or substitutes.

Firms maximize profits by choosing the number of workers and apprentices in each period. If a firm in sector k 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 $l_{a,t}^k := \zeta_a^k n_{a,t}^k$. ζ_a^k denotes the productivity of trained apprentices in sector k relative to the productivity of workers in that sector. Firms in each sector take as given the price of the good they produce p_t^k, the wage of workers w_t^k and that of apprentices $w_{a,t}^k$.

A firm in sector k with managerial ability z^k and net training costs t_a^k solves

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

The constraints $t_a^k n_{a,t}^k \leq n_t^k$ ensure that the firm has to hire enough workers to train the chosen number of apprentices in every period. We can substitute l_t^k and $l_{a,t}^k$ to write an equivalent (and perhaps more familiar) optimization problem

$$\max_{(l_t^k)_t, (l_{a,t}^k)_t \geq 0} \sum_{t=0}^{\infty} p_t^k f^k(l_t^k, l_{a,t}^k; z^k) - w_t^k l_{a,t}^k - \frac{(w_{a,t}^k + t_a^k w_t^k)}{\zeta_a^k} l_{a,t}^k.$$

Note that in this case the firm's decision is static and symmetric across all sectors. We therefore drop the sector and time superscripts and subscripts to avoid cluttered notation whenever there are no ambiguities.

We make some simplifying assumptions on f to further characterize the optimal number of workers and apprentices, guaranteeing the existence and uniqueness of the solution.

Assumption 1. ***(Production Function)*** *Suppose $f : \mathbb{R}_+^3 \to \mathbb{R}_+$ is twice continuously differentiable and*

[22] These heterogeneous training costs can reflect how complex the task a particular firm solves is, the way the firm is organized, search costs in terms of time to find apprentices, or differences in the teaching experience of workers, among other things. Here we do not attempt to distinguish between these stories, but rather to understand the consequences of heterogeneous net training costs across firms and sectors.

(i) homogeneous of degree $\gamma \in (0,1)$ *in* (l, l_a).

(ii) the Inada conditions hold for workers' labor, $\lim_{l \to 0} \frac{\partial f}{\partial l} = \infty$.

(iii) has non-negative cross derivatives with respect to z, *i.e.* $\frac{\partial^2 f}{\partial l \partial z} \geq 0$ *and* $\frac{\partial^2 f}{\partial l_a \partial z} \geq 0$.

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. All proofs are in Appendix B.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* $\frac{\partial n_a^*(z)}{\partial z} \geq 0$.

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

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

Intuitively 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. In their optimization, firms compare the marginal product of an additional apprentice to the marginal cost $w_a + t_a w$. 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 net 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.

Labor Markets

We consider a setting in which 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 k, $w_t^k \geq w_{a,t}^k$ $\forall t$. In addition, the minimum wage could be binding. So together both constraints imply $w_t^k \geq w_{a,t}^k \geq w_{min} \geq 0$, $\forall t, k$.

Let $N_t := \int\int n_t^*(z, t_a) d\mathcal{Z}(z) d\mathcal{T}(t_a)$ and $N_{a,t} := \int\int n_{a,t}^*(z, t_a) d\mathcal{Z}(z) d\mathcal{T}(t_a)$ denote aggregate demand for workers and aggregate demand for apprentices in period t, respectively. The labor market clears separately in each sector.[23] The market clearing conditions are

$$N_t^k + U_t^k = L_t^k \quad , \quad \sum_k N_{a,t}^k + U_{a,t} = L_{a,t}, \tag{2}$$

where $U_t, U_{a,t} \geq 0$ denote unemployed workers in sector k and untrained apprentices in period t. $N_{a,t}^k$ denotes the number of apprentices trained in period t in sector k. Trained apprentices increase the supply of workers in the sector where they were trained. 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[24]

$$L_{t+1}^k = L_t^k + N_{a,t}^k \tag{3}$$

Having separate labor markets for workers in each sector allows us to account for wage differences across sectors as observed in the data. Effectively this amounts to an assumption that workers' skills are sector-specific but equally useful for firms within the same sector. Becker (1964) emphasizes the distinction between general and specific human capital fundamental to firms' training incentives. In competitive markets, if training is not firm-specific, individuals still have incentives to pay for it, for instance by accepting lower wages. In the perfectly competitive case, as long there are no labor market restrictions or liquidity constraints, training will be efficient.

In our model, the minimum wage floor prevents wages to fully compensate for firms' training costs. As discussed in section 3.2, our reduced-form results indicate that this is indeed an important source of inefficiency in real-world labor markets. When the minimum wage for apprentices is set below the minimum wage for regular workers, this can increase training for individuals with low initial productivity, enabling apprentices to effectively pay for at least a part of their training.

Preferences

To close the model we suppose a simple preference structure. Individuals have homogeneous preferences $u(\cdot)$ over the goods produced by each sector. They choose a consumption bundle

[23] In Appendix B.4 we consider the case with multiple types of workers, where markets clear by occupation instead of sector. We show that results are similar to the ones in the baseline model. However, quantitatively if we want to trace cost distribution by type of worker, we need information on the number of apprentices trained in each occupation. Unfortunately, we do not have access to this data.

[24] We can extend the model to include imperfect training measuring the effective units of labor supplied by apprentices in future periods, for instance, $L_{t+1}^k = L_t^k + \xi_a N_{a,t}^k$, with $\xi_a \leq 1$. This may be an interesting direction, but we would need data on apprentices' outcomes and training to discipline the model for the quantitative exercises.

$c_t = (c_t^1, ..., c_t^K)$, taking prices in each sector p_t^k as given. Individuals have different incomes depending on their sector and on whether they are trained or untrained apprentices, workers or firm owners/managers. Suppose untrained apprentices earn a subsistence income $0 \leq \underline{w} \leq w_{min}$[25], trained apprentices earn wage $w_{a,t}$, sector k workers wage w_t^k, and firm owners earn profits $\pi(z^k, t_a^k)$ in sector k.

Suppose individuals are infinitely lived. An individual i maximizes their lifetime utility

$$\max_{(c_t)_t} \sum_{t=1}^{\infty} \beta^t u(c_t) \quad \text{s.t} \quad \sum_{k=1}^{K} p_t^k c_t^k = I_t^i \quad \forall t, \tag{4}$$

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, $\frac{\partial u_t / \partial c_t^k}{\partial u_t / \partial c_t^k} = \frac{p_t^k}{p_t^j}$ and $\sum_{k=1}^K p_t^k c_t^k = I_t^i \ \forall k, t$. Assuming $u(\cdot)$ is quasi-concave, let $c_t^*(I_t^i; p_t) = \left(c_t^{1*}(I_t^i; p_t), ..., c_t^{K*}(I_t^i; p_t)\right)$ 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 $p_t = (p_t^1, ..., p_t^K)$

$$C_t^k(p_t) + C_{a,t}^k(p_t) + C_{f,t}^k(p_t) = Y_t^k(p_t) \ \forall k, \tag{5}$$

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.

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

Equilibrium

A competitive equilibrium is defined as the set of wages and prices for 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 good markets clear. Formally,

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),

[25] This income represents the outside option of apprentices. It can be interpreted as an unemployment benefit.

(ii) *wage restrictions are satisfied,* $w_t^{*k} \geq w_{a,t}^{*k} \geq w_{min}$ *and labor markets clear* (2) *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* (3)*,*

(iv) *individuals maximize utility* (4) *and,*

(iv) *the good markets clear* (5) *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 $w_a^* = w_{min}$.

4.2 Equilibrium with Regulation

In this section, we describe the firm-size based apprenticeship regulation using 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 $n_a \in [\underline{n}_a^j, \overline{n}_a^j]$, $\forall j \geq 1$. Second, it reduces the minimum wage of apprentices to w_{min}^a, below the minimum wage for workers w_{min}. Alternatively, 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

$$\begin{aligned} \max_{n, n_a \geq 0,\ d_f \in \{0,1\}} \quad & f(n - t_a n_a, n_a; z) - wn - w_a n_a - d_f \mathcal{F}_a(n, n_a) \quad s.t \quad t_a n_a \leq n \\ & (n, n_a, d_f) \in \bigcup_j [N_{j-1}, N_j) \times [\underline{n}_a^j, \overline{n}_a^j] \times \{0\} \ , \quad \text{or} \\ & (n, n_a, d_f) \in \bigcup_j [N_{j-1}, N_j) \times [0, \underline{n}_a^j] \times \{1\} \quad \text{and} \quad \mathcal{F}_a(n, n_a) = \phi_a(\underline{n}_a^j - n_a)^+. \end{aligned} \tag{6}$$

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 never allowed to exceed the 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 (6) depending on 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 $n_a^*(z,t_a) > \overline{n}_a^j$, when it is between the bounds $n_a^*(z,t_a) \in \left[\underline{n}_a^j, \overline{n}_a^j\right]$ 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 summarizes these results.

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 net 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 (6) (with regulation).*

i. If $n_a^*(z,t_a) > \overline{n}_a^j$, then either $n^r(z,t_a) = N_k$ for $k \geq j+1$ and $n_a^r(z,t_a) > \overline{n}_a^j$ (increase size to get more apprentices) or $n_a^r(z,t_a) = \overline{n}_a^j$ and $n^r(z,t_a) < n^*(z,t_a)$ (bounded by maximum quota).

ii. If $n_a^*(z,t_a) \in \left[\underline{n}_a^j, \overline{n}_a^j\right]$, 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 either $n^r(z,t_a) = N_k - \varepsilon$ for $k \leq j$ (with $\varepsilon \to 0$) and $n_a^r(z,t_a) < \overline{n}_a^j$ (reduce size to avoid apprentices); $n^r(z,t_a) \geq N_j$ and $n_a^r(z,t_a) < \overline{n}_a^j$ 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).

Let us focus on the case where the optimal number of apprentices without regulation is a fixed proportion of the labor force, $n_a^* = Bn^*$ with $B \in \mathbb{R}_+$.[26] We show that if relative wages of apprentices are low enough $\frac{w_a}{w} \to 0$ and the net training cost are also low $t_a \to 0$, firms want to train as many apprentices as possible. In this case, the optimal number of apprentices is above the maximum quota for firms larger than N_1. Firms bunch at the thresholds N_j, with missing mass on the left and never pay the fee. In contrast, when $\frac{w_a}{w} \to \infty$ or $t_a \to \infty$, the optimal number of apprentices converges to zero, below the minimum quota for firms larger than N_1. 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 (6). Then,*

[26] This holds for example under Assumption 1 and when $\partial_l f / \partial_{l_a} f$ is homogeneous of degree zero in z. See Lemma 4 in the Appendix B.1 for a full characterization of these solutions.

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

i. *the number of apprentices without regulation is* $n_a^* = B_u n^*$ *and lies above the maximum quota,* $n_a^*(n) > \overline{n}_a^j$*.*

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_{j+1}*, so there is missing mass on the left of the thresholds.*

iii. *firms choose the maximum number of apprentices* $n_a^r = \overline{n}_a^j$*.*

iv. *firms never pay the fee.*

Case 2: *there exist* $\underline{\frac{w_a}{w}}, \underline{t_a}$ *such that for* $\frac{w_a}{w} \geq \underline{\frac{w_a}{w}}$ *or* $\underline{t}_a \geq \bar{t}_a$*,*

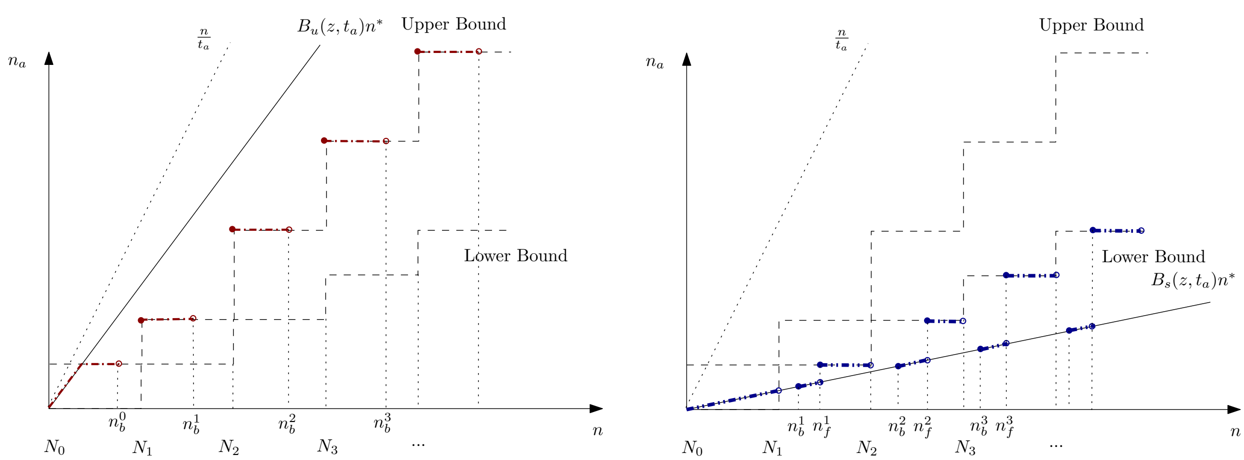
i. *the number of apprentices without regulation is* $n_a^* = B_s n^*$ *and lies below the minimum quota of the regulation,* $n_a^*(n) < \underline{n}_a^j$*.*

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 the threshold* N_j*.*

iii. *firms that increase the number of apprentices choose the minimum number* $\underline{n}_a^j$*.*

iv. *there exist* $\overline{\phi}_a > 0$ *such that for* $\phi_a \leq \overline{\phi}_a$*, there is an additional cutoff* z_f^j *such that firms* $z \in (z_r^j, z_f^j]$ *choose to pay the fee.*

Figure 7: Apprentices After Regulation

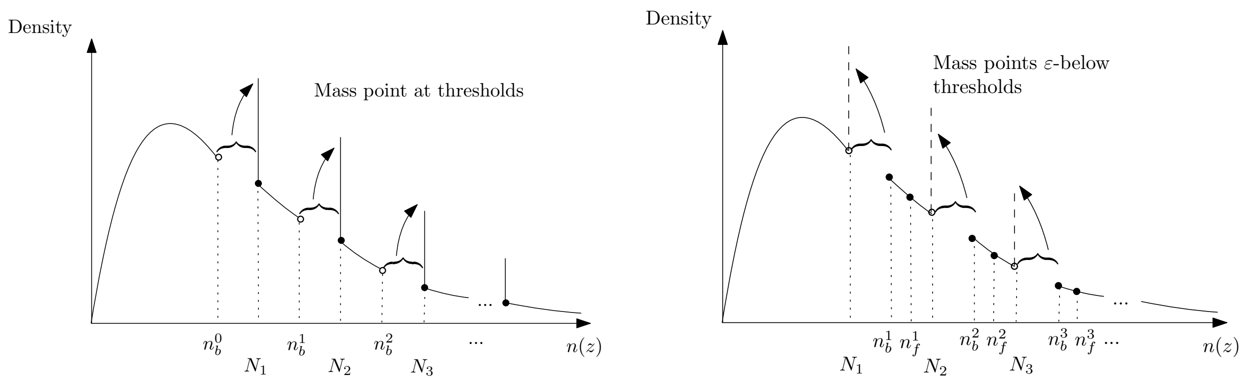


(a) Case 1: Low $\frac{w_a}{w}$ and low t_a

(b) Case 2: High $\frac{w_a}{w}$ and high t_a

Figure 7 illustrates the two cases in Proposition 1. Panel 7a depicts Case 1, where both $\frac{w_a}{w}$ and t_a are small. Firms seek apprentices as they are relatively cheap in terms of wages and training costs. In this case, the optimal number of apprentices without regulation lies above the maximum quota. This implies that some firms increase their size and bunch at the thresholds, leaving a missing mass of firms to the left. All of these firms choose to train at the maximum quota. Panel 7b depicts Case 2, where $\frac{w_a}{w}$ and/or t_a is large. In this case, training apprentices is too expensive relative to workers. The optimal number of apprentices lies below the minimum quota of the regulation. Firms bunch just below the threshold to avoid having to train extra apprentices. There is a missing mass of firms to the right of the thresholds. If the fee is sufficiently small, some firms prefer to pay the fee and choose the optimal number of apprentices and workers. Firms that do not pay fees choose the minimum quota of the regulation.

Figure 8: Firm Size Distribution



(a) Case 1: Low $\frac{w_a}{w}$ and low t_a

(b) Case 2: High $\frac{w_a}{w}$

Figure 8 shows the implications of the policy for the firm size distribution in the two cases from Proposition 1. Panel 8a depicts the case where firms increase their size to train more apprentices. They bunch at each of the thresholds $\{N_j\}_j$ and there is a missing mass of firms between $[n_b^j, N_{j+1})$. Panel 8b illustrates Case 2, where firms either reduce their size or pay the fee to avoid training more apprentices. Firms bunch below each threshold, leaving a missing mass of firms 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.

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 either decreasing their size and bunching just below the thresholds or by paying the fee.

Negative Marginal Productivity of Apprentices

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 apprentices' minimum wage $\phi_a > w_a^{min}$, 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 $\pi_f(z) > \pi_a(z) \Rightarrow \frac{\partial f}{\partial n_a} < 0$.*

In other words, if apprentices had positive productivity and it was 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 by adding training costs. 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 whenever there are high net training costs and/or the marginal productivity of labor f_l is high. This 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). Similarly to this literature, apprentices can only be productive to the firm after they have acquired knowledge from other workers. However, some firms may find it too costly to train apprentices as it is too costly in terms of workers' time and prefer to pay the fee.

4.3 Aggregate Effects and Welfare

To compute aggregate and welfare effects, we measure the change in total aggregate output, and we study changes in the welfare of different groups of agents, namely apprentices, workers, and firm owners/managers. Welfare changes for each of these groups are reflected by the effect of the policy on the number of trained apprentices, the number of employed workers, the wages of apprentices and workers, and profits received by firm owners, respectively.[27] As one may expect, the policy has differential effects by sector and by group of agents, creating winners and losers.

More formally, regulation $\mathcal{R}$ consists of a set of thresholds $(N_j)_{j=0}^{\infty}$, a set of minimum and maximum quotas $\underline{n}_a^j, \overline{n}_a^j$, the minimum wage for apprentices w_{min}^a and a fee function $\mathcal{F}_a : \mathbb{R}_+^2 \to \mathbb{R}_+$,

$$\mathcal{R} := \left\{ \{N_j, \underline{n}_a^j, \overline{n}_a^j\}_j, w_{min}^a, \mathcal{F}_a \right\}.$$

Let $n_{a,t}^{kr}(z, t_a)$ denote the equilibrium number of apprentices trained by a firm (z, t_a) in sector k and period t, with the regulation. We can add these up to obtain the total trained apprentices in each sector,

$$N_{a,t}^k = \int \int n_{a,t}^k(z, t_a) d\mathcal{Z}^k(z) d\mathcal{T}^k(t_a).$$

We compute the change in the number of trained apprentices as the difference with vs. without regulation, $\Delta N_a^k = N_a^{kr} - N_a^{k*}$. Similarly, we compute the change in the number of workers in each sector, $\Delta N^k = N^{kr} - N^{k*}$ and the change in aggregate output, $\Delta Y_t = Y_t^r - Y_t^*$.

The welfare effect for each group is given by the change in aggregate utility for each type of agent $j \in \{\text{Apprentices, Workers, Firm Owners}\}$ in each sector k, $\Delta \mathcal{U}_j^k = \mathcal{U}_j^{kr} - \mathcal{U}_j^k$. As we assume that all agents have preferences according to (4), when prices are fixed the differences in their utility come solely from differences in income. Thus, we track the change in apprentices' wages w_a, workers' wages w^k and firm owners' profits $\Pi^k(z, t_a)$. In addition, we have to track the number of workers and apprentices employed by firms, as any unemployed workers and untrained apprentices only receive their outside option.

5 Quantitative Exercises

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

[27] A benevolent social planner may want to weigh the effect on these three groups through some social welfare function. Here we do not attempt to spell out this welfare function, but rather quantify the effect on the income of each of these groups.

5.1 Parameter Estimation and Moments

Production Function

For our quantitative exercises, we parameterize the production function of firms $f^k(l, l_a; z, t_a)$ as a Cobb-Douglas function featuring managerial ability and labor input, allowing its parameters to vary across sectors $k \in \{u, s\}$.[28] We assume that labor input combines labor from workers and apprentices in a linear function.

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

Linearity implies that apprentices are substitutes for workers in the sector where they are trained. ζ_a^k denotes marginal productivity of a trained apprentice relative to workers in sector k. In this case 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.

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 proposition (2), a necessary condition for this to hold is that apprentices must have negative marginal productivity for firms. The production function with linear labor input features apprentices with negative marginal productivity whenever net training costs are larger than the productivity of trained apprentices, i.e. $(\zeta_a^k - t_a^k) < 0$. Second, as shown in Table A.2 in Appendix A.2, we document that most firms choose corner solutions: only very few choose apprentices between the bounds, and whenever they pay the fee they choose to have zero apprentices. This supports the view that the linear labor input function is a good approximation of firms' behavior.

Since in this case apprentices and workers are substitutes, the optimal number of workers and apprentices is given by corner solutions. If workers' wages are relatively low compared to the effective cost of apprentices, $w^k < \frac{w_a^k + t_a^k w^k}{\zeta_a^k}$, then it is optimal to only hire workers, $n^* = \left(\frac{\gamma^k}{w^k}\right)^{1/(1-\gamma^k)} z$ and $n_a^* = 0$. On the contrary, if $w^k > \frac{w_a^k + t_a^k w^k}{\zeta_a^k}$, firms want to train as many apprentices as possible, subject to having enough workers to train them, such that $n^* = t_a n_a^*$ and $n_a^* = \left(\frac{\gamma^k}{w^k t_a + w_a^k}\right)^{1/(1-\gamma^k)} z$. Finally, in the case where the cost of hiring workers equals the effective cost of training apprentices $w^k = \frac{w_a^k + t_a^k w^k}{\zeta_a^k}$, firms are indifferent between hiring workers or apprentices. In general equilibrium, binding minimum wages for apprentices prevent this last case from happening. We can use the market clearing conditions (2) in combination with the minimum wages to find equilibrium wages

[28] In Appendix B.5 we describe a more general parametrization of the production function. However, as we argue below, the data suggests the linearity assumption is a good approximation given the tendency of firms to choose corner solutions.

$(w_t^{k*}, w_{a,t}^{k*})$ and the number of trained apprentices $N_{a,t}^k$ for each sector k and in each period t.

Training Costs and Managerial Ability Distributions

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 fact, we use sector-specific training cost distributions. We estimate a 'non-parametric' distribution $\mathcal{T}^k(\cdot)$ for each sector. We identify 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.[29]

Additionally 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.[30]

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 control function methods. 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.

We estimate four parameters and $n_{\mathcal{T}} = 13$ points of the training cost distribution for each sector $k \in \{u, s\}$: the output elasticity of labor γ^k, the managerial ability distribution parameters $\{\lambda^k, \theta^k, \xi^k\}$ and the points $(t_{a,i}^k, \varrho_i^k)$ of the training costs CDF. We identify these parameters by targeting the pre-reform data on firm size and production, and key post-reform moments stemming from the firm responses described above.

[29] In Appendix C.3 we show that we obtain a similar but slightly worse fit for a calibrated truncated normal or uniform distribution.

[30] In Appendix C.1 we show 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.

As a first step, we estimate production functions in each sector 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 in our model full-time labor is the only production input, we estimate this elasticity by regressing output of each firm with respect to the number of full-time workers, controlling for time and firm-level fixed effects. The estimated elasticity is upward biased if the other production inputs are gross complements of full-time labor, hence we may overstate the importance of full-time labor in production. This means the costs and benefits of the effects we find in our quantitative analysis are an upper-bound of the true effect. In Appendix C.2 we follow the procedure of Levinsohn and Petrin (2003) with the Ackerberg et al. (2015) correction to estimate the output elasticity of labor γ^k for each sector.[31] Results using these estimated elasticities are quantitatively similar to the ones from our benchmark specification.

Next, we match the pre-reform firm size distribution using a maximum likelihood estimation procedure. The functional form of 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 for 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 of our reduced form analysis. In each sector we use in total 43 moments: 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 $(t_{a,i}^k, \varrho_i^k)$ minimizing the weighted sum $\mathcal{L}\left((t_{a,i}^k, \varrho_i^k)_{i=1}^{n_\mathcal{T}}\right)$ of the absolute difference between the model-implied moments and the data moments,

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

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.[32]

[31] We follow the code *prodest* developed by Mollisi and Rovigatti (2017) to estimate the production function using control function methods. In the appendix we compare different production function estimation methods including the one developed by Olley and Pakes (1996), Levinsohn and Petrin (2003) and Wooldridge (2009).

[32] See Appendix C.4 for the precise definition of these weights.

We use the regulation specifications for the remaining parameters,

$$\chi_a = \{\{N_j, \underline{n}_a^j, \overline{n}_a^j\}_j, w_a^{min}, \phi_a\},$$

where $\{N_j, \underline{n}_a^j, \overline{n}_a^j\}_j$ denote the apprentice thresholds $\{N_j\}_j = \{15, 30, 50, ..., 20(j-1)+10, ...\}$ and quotas establishing a minimum $\underline{n}_a^j$ and a maximum number of apprentices $\overline{n}_a^j$. w_a^{min} is the minimum wage for apprentices which we set to 75% of the minimum wage, $w_a = 0.75 w_{min}$,[33] and $\phi_a = w_{min}$ is the fee parameter that is proportional to the difference between required and actual apprentices.

Intuitively, the firm responses identify points in 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 $Pr\{t_a \geq \zeta_a^k + \frac{\phi_a - w_a}{w^k}\} = \mathcal{T}\left(\zeta_a^k + \frac{\psi_a - w_a}{w^k}\right)$. The minimization algorithm first chooses these three points of the CDF, matching these three probabilities. We determine the remaining $n_{\mathcal{T}} - 3$ points choosing percentiles and estimating the $t_{a,i}^k$ that minimize 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.[34] These points of the CDF identify the training cost distribution. Note that using only firm-level data, we cannot identify ζ_a separately from t_a. To anchor the training cost distribution, we normalize $\zeta_a = 1$ in both sectors. Therefore the scale of the estimated training costs should be interpreted as costs in terms of workers' time.

Table 2 summarizes the estimated parameters. We report confidence intervals at the 95% level for each parameter estimate, using 1000 bootstrap samples. Both sectors have similar labor intensity, but the high-skill sector has higher managerial ability. The managerial ability distribution has a larger location λ and scale parameter θ in high-skill sectors, but a smaller shape parameter ξ.[35] The production function estimation yields similar labor share of $\gamma^s = 0.61$ in high-skill sectors and in the low-skill sectors, $\gamma^u = 0.58$.

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

[33] 75% of the minimum wage is the apprentices' wage during the productive phase, which corresponds more closely to the model.

[34] We use a *pattern search* algorithm which is useful for solving non-smooth optimization problems with simulations. The additional $n_{\mathcal{T}} - 3$ percentiles ϱ_i are chosen as Chevyshev nodes around the anchored points. Choosing them uniformly spread yield similar results.

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

Table 2: Estimated Parameters

Parameter	Description	High-Skill	Low-Skill
γ^k	Labor share in output.	0.61	0.58
		(0.42, 0.64)	(0.26, 0.53)
λ^k	Location parameter of productivity distribution $\mathcal{Z}(z)$.	1916	939
		(724, 2511)	(444, 747)
θ^k	Scale parameter of productivity distribution $\mathcal{Z}(z)$.	1978	1014
		(745, 2582)	(478, 805)
ξ^k	Shape parameter of productivity distribution $\mathcal{Z}(z)$.	0.85	0.88
		(0.83, 0.88)	(0.85, 0.9)

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. On average, there are a similar 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. Using the estimated firm size distribution, we compute the implied initial supply of workers.

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 wages from the EAM survey data, 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 the low-skill sector 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 below.[36] 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 only half of the students enrolled in technical and vocational institutions can find an apprenticeship (SENA 2018).

[36] An alternative assumption may be to have different types of workers. In Appendix B.4 we consider an extension of the model where there are multiple types of workers in each sector with possibly different wages. However, estimating this model requires additional data on the type of workers and the occupations apprentices are trained in.

5.2 Goodness of Fit

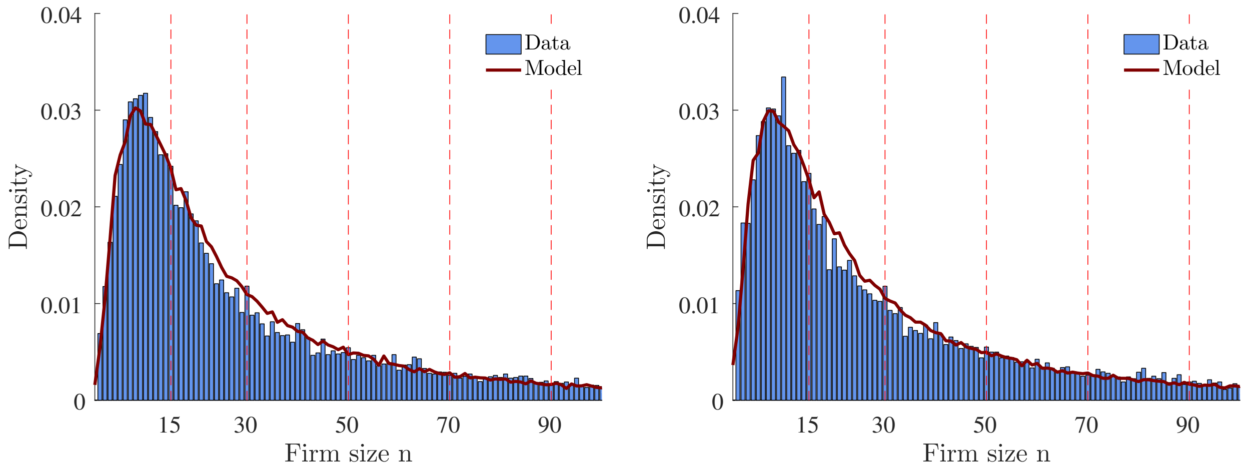
Table 3 reports the targeted moments in the data and the corresponding values we obtain from the estimated model. In cases where we match a full function a reference is provided to the relevant figure showing the fit. Our estimated model closely resembles the data. We match, almost exactly, the fraction of firms that choose the maximum quota and that paying the fee by sector. Moreover, Figure 9 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 3: Targeted Moments

Moment Description	**Model**	**Data**
High-Skill		
Pre-reform distribution of full-time	see Fig. 9a	see Fig. 9a
Fraction of firms in upper-bound pre-reform (≥ 15 workers)	1.5%	1.5%
Fraction of firms in upper-bound post-reform (≥ 15 workers)	1.5%	1.5%
Fraction of firms that pay the fee (≥ 15 workers)	58.2%	58.2%
Bunching and missing mass	see Fig. C.7c	see Fig. C.7c
Low-Skill		
Pre-reform distribution of full-time	see Fig. 9b	see Fig. 9b
Fraction of firms in upper-bound pre-reform (≥ 15 workers)	1.5%	1.5%
Fraction of firms in upper-bound post-reform (≥ 15 workers)	65.4%	65.4%
Fraction of firms that pay the fee (≥ 15 workers)	0.5%	0.5%
Bunching and missing mass	see Fig. C.7d	see Fig. C.7d

Figure 10 presents the fit of the estimated model to our three empirical facts. Panels C.7c and C.7d show the observed and simulated firm size distribution after the change in regulation, our 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 overestimates the fraction of firms bunching in the high-skill sector. This happens 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 potentially "smooth" firm

Figure 9: Fitted Pre-Reform Distribution of Full-Time Workers



(a) Pre-Reform High-skill Size Distribution

(b) Pre-Reform Low-skill Size Distribution

Notes: Distribution of full-time workers by sector for the pre-reform period (1995-2002). *Source*: own computations using simulations and data compiled by SENA.

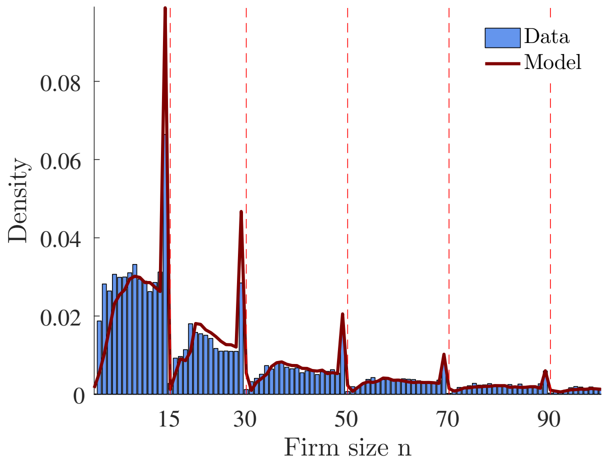
responses. For instance, if other inputs such as capital, material or other types of labor substitute for full-time workers, this may reduce the fraction of firms bunching. Note that such over-bunching induces upward bias in quantifying the costs of training apprentices in high training cost firms, as well as in the benefits of training apprentices for low training cost firms. Hence, our quantitative exercises can be interpreted as an upper bound on the aggregate effects of training apprentices.

Panel C.7e shows the fit of the model to empirical fact 2. High-skill sector firms choose an average number of apprentices below the minimum quota, while the average number of apprentices 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 threshold choose the maximum quota of the regulation, while the fraction of firms choosing this maximum decreases in firm size between thresholds. Finally, Panel C.7f 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 paying the fee is relatively flat in firm size.

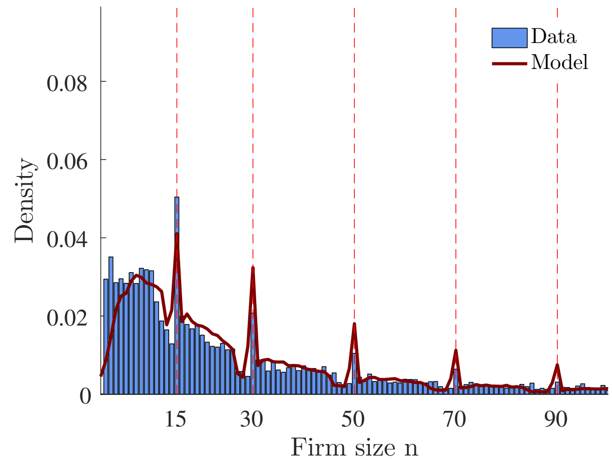
5.3 Quantitative Results

In this section, we present the main quantitative results. First, we show and interpret the net training cost distribution for each sector. Then we use the estimated training costs to compute the aggregate and welfare effects of the policy in three scenarios: (i) the short-term or *partial equilibrium* effect where wages are constant, (ii) the *general equilibrium* effect where displaced workers are

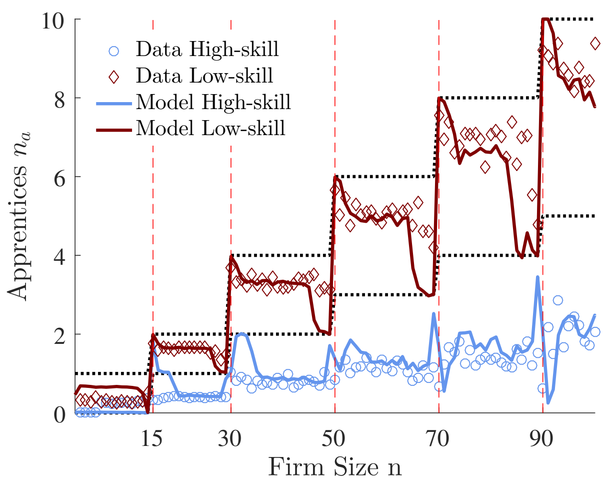
Figure 10: Fit of Empirical Facts



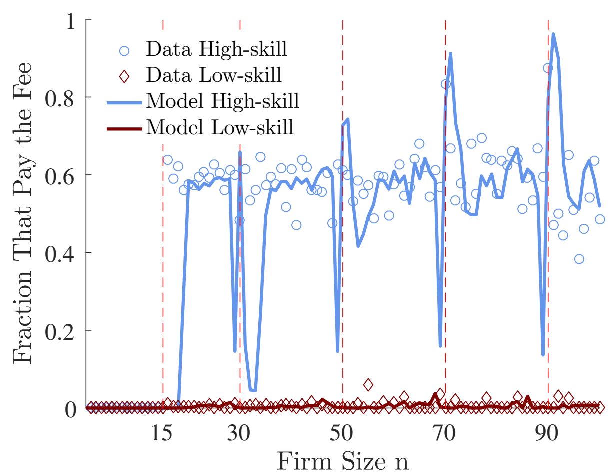
(a) Post-Reform High-skill Size Distribution



(b) Post-Reform Low-skill Size Distribution



(c) Number of Apprentices



(d) Fraction that Pay the Fee

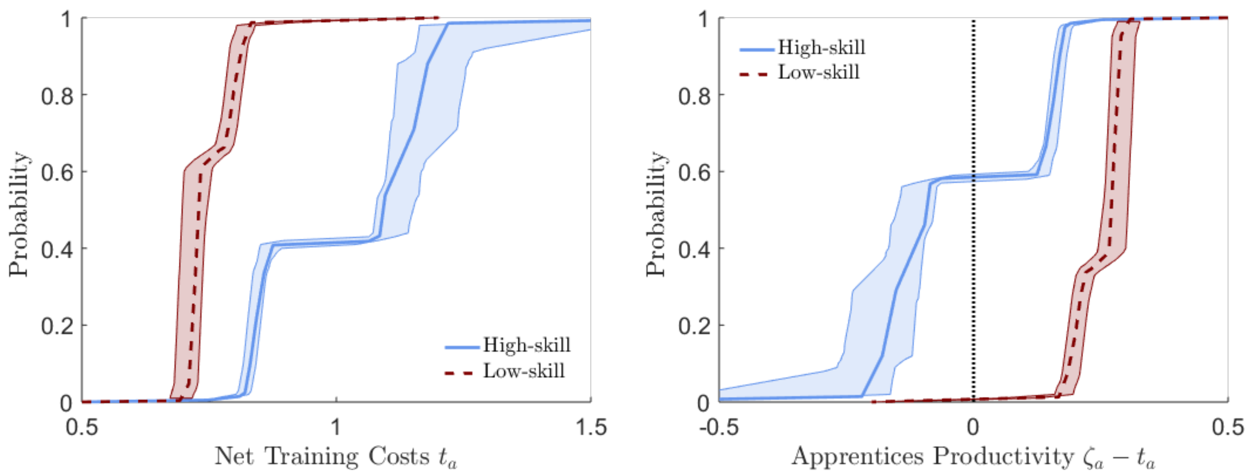
Notes: Panel (a) and (b) show the distribution of full-time workers by sector for the post-reform period (2003-2009) and the model fit. Panel (c) the number of apprentices in each sector by the firm size (full-time workers). Panel (d) the fraction of firms that pay the fee by firm size. *Source*: own computations using simulations and data compiled by SENA.

absorbed back into the labor market, and (iii) the *dynamic* effect where trained apprentices increase the future supply of workers.

Training Cost Distribution

Figure 11 shows the estimated training cost distribution, normalizing $\zeta_a = 1$. Average training costs are lower for firms in low-skill sectors ($\mathbb{E}(t_a^u) = 0.75$) than in high-skill sectors ($\mathbb{E}(t_a^s) = 1.05$), making them more likely to train apprentices. The training costs in high-skill sector are also estimated to be more dispersed ($Var(t_a^s) = 0.053$) than in low-skill sectors ($Var(t_a^u) = 0.0025$).

Figure 11: Training Cost Distribution and Marginal Productivity



(a) Training Costs Distribution

(b) Apprentices' Marginal Productivity

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

An interpretation of these heterogeneous 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, that is $\zeta_a - t_a$. Panel 11b shows this marginal productivity in both sectors. The units tell us how productive an untrained apprentices is relative to a regular worker. If $\zeta_a - t_a = 0.25$, for instance, then the marginal productivity of training an apprentice is equal to 25% of the marginal productivity of the average worker in that sector. 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 4.2, this ties in well with the empirical observation that many firms in the high-skill sector choose to pay fees.

Partial Equilibrium, General Equilibrium and Dynamic Effects

Wages are key to quantifying the effects of the policy on aggregate outcomes and welfare. In Table 4, we show wages in each sector (relative to the minimum wage) 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 percent. 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 more (6.2%) than in high-skill sectors (2%).

Table 4: 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 are computed using the EAM data for the last year before the reform (2002). General equilibrium and dynamic effects wages come from the estimated model.

Aggregate Effects

In this section, we quantify the partial equilibrium, general equilibrium and dynamic effects of the policy on aggregate variables. The effect on aggregate output reflects the change in efficiency induced by the policy. The changes in the total number of workers and apprentices are informative of the mechanisms that affect output.

Table 5: Effect on Aggregate Variables

	Workers	% Workers	Apprentices	% Output
A. ***Partial Equilibrium***				
High-skill Sectors	-1964	-0.92	4527	-0.39
Low-skill Sectors	-3304	-1.57	17622	0.36
Total	-5268	-1.24	22149	-0.06
B. ***General Equilibrium***				
High-skill Sectors	0	0.00	4557	0.18
Low-skill Sectors	0	0.00	17898	1.28
Total	0	0.00	22454	0.67
C. ***Dynamic Effects*** $(t+1)$				
High-skill Sectors	4602	2.16	4642	1.50
Low-skill Sectors	17922	8.49	19207	6.17
Total	22524	5.31	23849	3.57

Notes: Column (1) shows 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.

Table 5 shows the effects under each scenario by sector. First, we document that despite the large labor input responses, the static effects on aggregate output are relatively small. Panel 5A 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 six apprentices. Moreover, the first two columns show that the percentage of displaced workers is approximately 0.9%, with more workers laid off in the low-skilled sector. Despite the sizable increase in training, there are relatively small effects on output. Column (4) shows that high-skill sector firms decrease their production by 0.39% and low-skill sector firms increase production by 0.36%.

In contrast, Panels 5B and 5C 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 1.2% for firms in low-skill sectors, while the initial output losses are reverted in high-skill sectors. Dynamically, these positive effects on output can be even stronger, resulting in an overall increase in output of around 3.1% mostly driven by an increase in production in low-skill sectors.

Winners and Losers

The policy 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. However, we need additional assumptions on apprentices and workers to quantify changes in their welfare. We assume that all agents have the same homothetic preferences, so their utility is linear in income.

For apprentices, we normalize their outside-option income to zero,[37] and to compute dynamic effects of the policy we assume apprentices learn the skills necessary to replace an average worker in the sector where they receive training. These assumptions probably overestimate the welfare effects of the policy on apprentices, such that this benchmark provides an upper bound on welfare consequences. In order to quantify more precisely the benefits of the policy for apprentices we would need additional data on outside options and the productivity of apprentices before and after training. Unfortunately, we do not have such data. 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 become unemployed and they earn no wages. Again, this likely represents an upper bound on the negative effects on displaced 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 aggregate welfare for each group of agents in each scenario. $\Delta\mathcal{U}_j$ denotes the change in the aggregate utility of group $j \in \{\text{Apprentices, Workers, Firms}\}$, comparing the respective scenario to a situation without regulation. The change is divided by the sum of the utilities across all these agents before the reform, $\mathcal{U}^*$, such that the sum of the first three columns is equal to the percentage change in aggregate total utility in each row.

The first column of the table shows that apprentices gain from the policy under all scenarios. In the short term, their welfare improves as the wages they earn as apprentices are higher than their outside option. Dynamically, these effects are magnified as they earn the wages of regular workers once they are trained. Since the low-skill sectors train more apprentices, the positive effects are larger there, although wages in high-skill sectors are higher.

In contrast, column (2) shows that incumbent workers are harmed by the policy. 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

[37] The 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 for this type of training suggests they are getting benefits from the apprenticeship program (SENA 2018).

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 those in the low-skill sector.

Table 6: Change in Aggregate Welfare

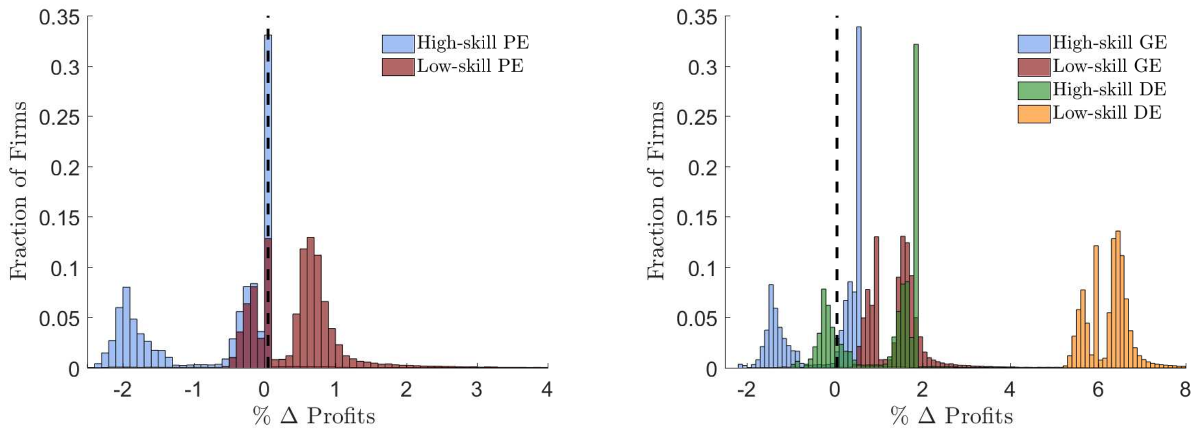
	Apprentices $\Delta\mathcal{U}_a/\mathcal{U}^*$	Workers $\Delta\mathcal{U}_w/\mathcal{U}^*$	Firms $\Delta\mathcal{U}_f/\mathcal{U}^*$	Total $\Delta\mathcal{U}/\mathcal{U}^*$
A. *Partial Equilibrium*				
High-skill Sectors	0.25	-0.56	-0.45	-0.77
Low-skill Sectors	1.12	-0.91	0.14	0.35
Total	0.64	-0.72	-0.19	-0.27
B. *General Equilibrium*				
High-skill Sectors	0.25	-0.22	-0.24	-0.21
Low-skill Sectors	1.14	-0.40	0.54	1.28
Total	0.64	-0.30	0.11	0.45
C. *Dynamic Effects* $(t+1)$				
High-skill Sectors	1.29	-0.70	0.24	0.83
Low-skill Sectors	4.65	-2.36	2.58	4.87
Total	2.78	-1.44	1.28	2.62

Notes: The table shows the change in aggregate utility of apprentices, incumbent workers and firm owners/managers, relative to the no-regulation total aggregate utility, $\mathcal{U}^*$. For each row, the sum of first three columns add up to column (4), the percentage change in total aggregate utility, $\Delta\mathcal{U}/\mathcal{U}^*$.

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. The general equilibrium scenario dampens some of the negative effects on high-skill sector firms and reinforces the gains of low-skill sectors, 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 $t+1$. Finally, the fourth column of the table sums utility changes across groups of agents. The total utility change is similar to the total output effect from table 5 in each scenario. Note that the total utility change tends to be slightly smaller, as any fees paid to the government are included in output but reduce profits and hence firm owners' utility.

Figure 12 further illustrates the heterogeneous effects of the policy on the profits of firms across sectors. As in table 6, low-skill sector firms gain on average, but high-skill sector firms lose. Due to within-sector heterogeneity in training costs, these costs vary substantially across firms. In Panel 12a, the distribution of the percentage change in profits for high-skill sector firms is bimodal, with one peak showing losses in profits of around 2%, while the other reflects that one-third of the firms are unaffected. In Panel 12b, the general equilibrium boost in profits is not enough to revert the losses of high-skill sector firms. In contrast, the majority of firms in low skill-sectors enjoy gains from the policy as they benefit from training apprentices. When we consider dynamic effects, the gains for low-skill sector firms increase further and almost all high-skill sector firms now also see gains from the regulation.

Figure 12: Heterogeneous Change in Profits



(a) Partial Equilibrium

(b) General Equilibrium and Dynamic Effects

Notes: Panel (a) shows de distribution of the percentage change in profits by sector for the partial equilibrium scenario, panel (b) for the general equilibrium and dynamic effects. We consider the partial equilibrium (PE), general equilibrium (GE) and dynamic effects (DE) of the policy.

Overall, our quantitative exercises show relatively small static effects of the policy on total output, as firms substitute workers for apprentices. There is some reallocation across sectors: production increases in low-skill sectors, but decreases in high-skill sectors. Across groups of agents, the policy also creates winners and losers. Trained apprentices benefit from the policy as they earn higher wages in the future and might even reap some gains in the short-run if their outside option is sufficiently low. Firms in low-skill sectors also gain as they can hire productive apprentices at relatively low wages, while most firms in high-skill sectors lose, some incurring sizable losses as training apprentices comes at a net cost to them. In general equilibrium and dynamically, there are increasing positive effects on profits as the supply of productive workers rises. Finally, incumbent workers can be harmed by the policy. In the short term, some are displaced by apprentices, and in general equilibrium they see a fall in wages.

5.4 Model Implications and Reduced-Form Evidence

To test the empirical plausibility of these quantitative findings, we next present some additional reduced-form evidence of the effect of the regulation on firm outcomes. Concretely, the quantitative exercises suggest that despite substantial labor input adjustments, there are small effects on output. To gauge the causal effect of the regulation on these outcomes in a reduced-form way, 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 large effects on the number of apprentices and workers, but small, insignificant on output.

We estimate the following difference-in-difference specification:

$$Y_{it} = \alpha_i + \delta_t + \beta Above_i \cdot Post_t + \epsilon_{it} \tag{7}$$

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 focuses on 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.[38]

The specification compares firms on the two sides of the thresholds, where they are subject to different apprentices' quotas. As discussed in section 3.2, firms' actual post-reform firm size is subject to manipulation, such that 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.

Table 7 shows results from the regression, focusing on four outcome variables, namely the number of apprentices, the 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 firms decrease the number of workers by 1.8. In high-skill sectors, there is a smaller and insignificant

[38] To be precise, we consider firms with a yearly change in full-time workers of no more than 50, and we restrict the analysis to firms with less than 300 full-time workers in 2002. In Appendix C.5 we show that without these restrictions, results are qualitatively similar but the effect on the number of workers becomes implausibly large, which seems to be driven by large firms jumping across multiple brackets.

point estimate of 0.8. Finally, the effect on log output and profit rates are small and insignificant in all sectors. Appendix table A.1 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 any other inputs, including capital and intermediate inputs, as well as on wages.

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

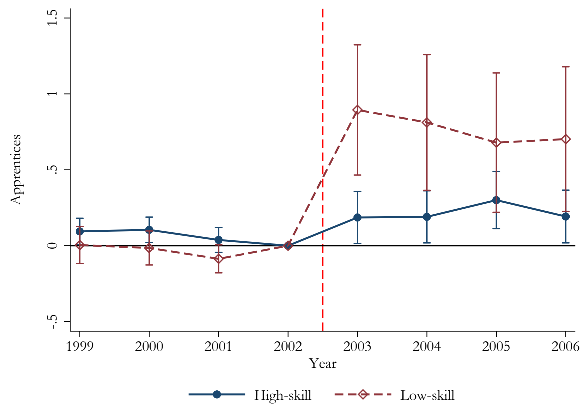
	Apprentices		Workers		Log Output		Profit Rate Π/Y	
Sector	High	Low	High	Low	High	Low	High	Low
Above*Post	0.159***	0.802***	-0.855	-1.781*	-0.0302	-0.00697	-0.0135	-0.00174
	(0.0576)	(0.224)	(0.863)	(1.036)	(0.0355)	(0.0472)	(0.0113)	(0.0107)
Mean (Pre-reform)	0.161	0.135	30.46	30.30	14.31	14.46	0.212	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

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

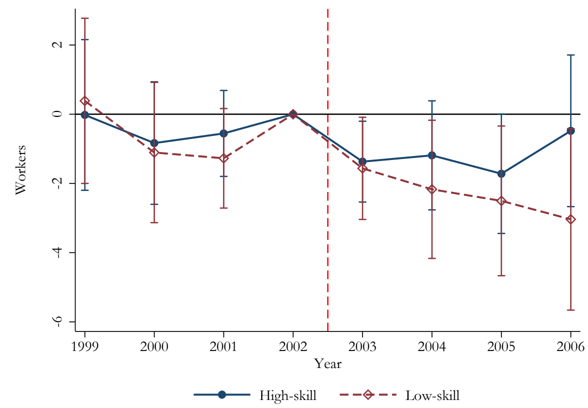
In addition, Figure 13 plots yearly difference-in-difference coefficients around the reform, comparing treated firms (above the thresholds) to control firms (below the thresholds) separately in high-skill and low-skill sectors. In both high-skill and low-skill sectors, pre-trends are flat for the number 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, Figure 14 shows a comparison of the reduced-form results from Table 7 to the same specifications from the model. For each outcome, we plot the difference-in-difference coefficient on $Above_i * Post_t$ in the diamonds with 95% confidence intervals. The hollow circles show the corresponding effects from the estimated model. As predicted by the model, the effect on the number of apprentices is stronger in low-skill sectors, although the model 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 effect on output is small and negative both in the model and the data. Profit rates are negatively affected both in model and

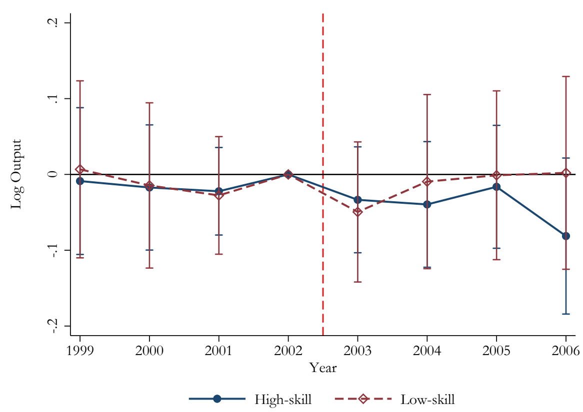
Figure 13: Reduced-Form Effects of the Policy, Yearly Coefficients



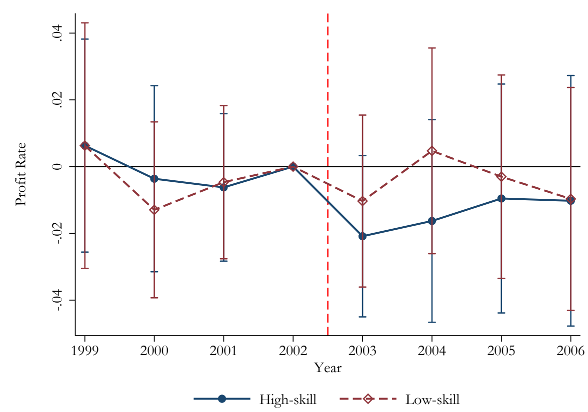
(a) Number of Apprentices



(b) Full-Time Workers



(c) Log Output

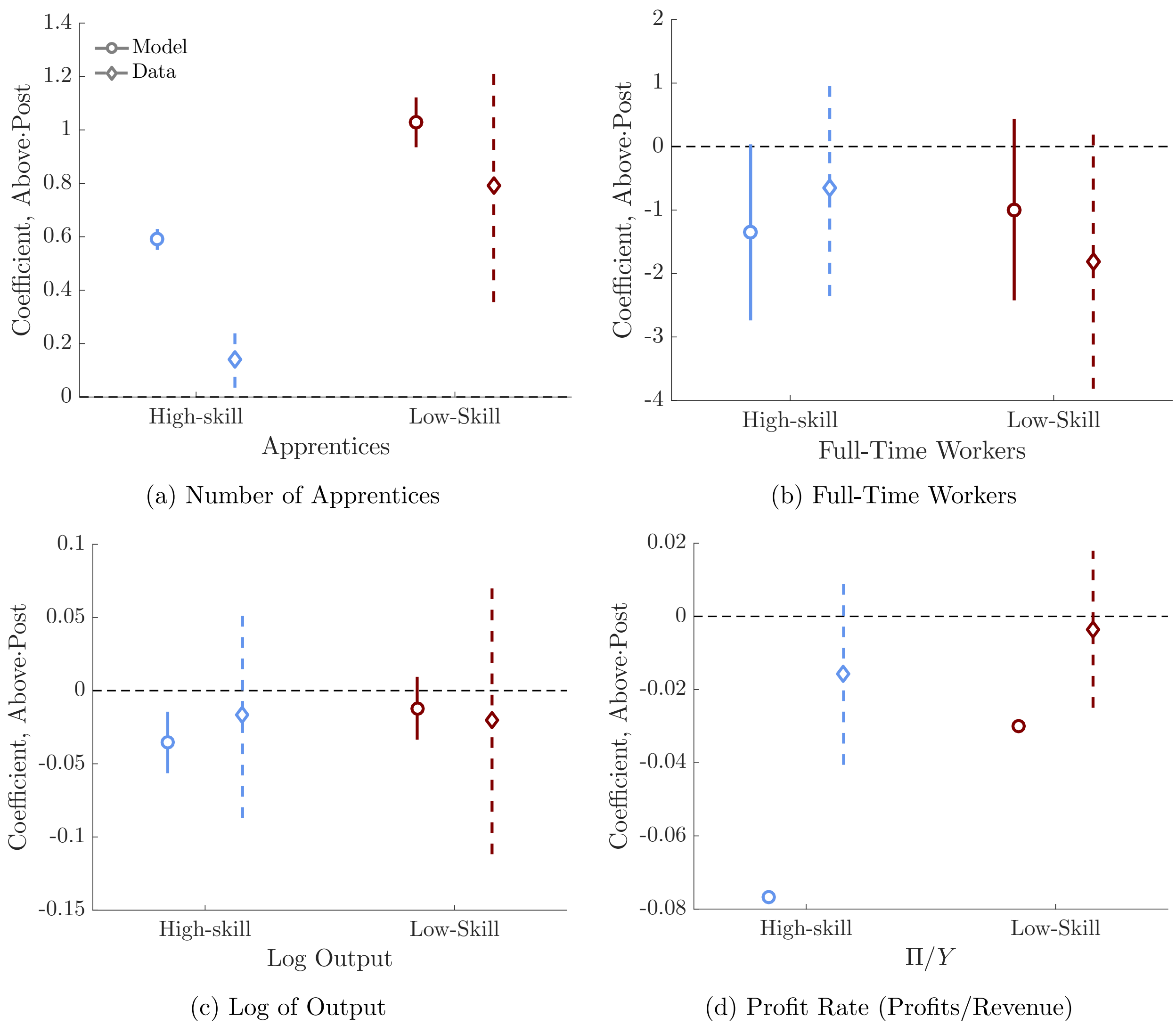


(d) Profit Rate (Profits/Revenue)

Notes: Yearly difference-in-difference coefficients for the period 1999-2006 using 2002 as base year.

data, but the model seems to overestimate these negative effects. The latter difference is probably related to the overestimation of bunching responses, such that the quantitative results become an upper bound. Overall, these results confirm that the estimated model not only closely replicates the empirical facts, but also produces predictions consistent with the local effects on firm outcomes around the thresholds.

Figure 14: Reduced-Form Coefficients vs. Model Prediction



Notes: Plots of $Above_i * Post_t$ for the model and the data in each sector, with 95% confidence intervals. *Source*: Data from SENA and EAM.

5.5 Policy Decomposition

Next, we can use the estimated model to study the separate effects of each of the components of the policy, namely the apprentice quotas, the decrease in the apprentice minimum wage, and the possibility of paying the fee. Here we present this decomposition in partial equilibrium and study the effects on the welfare of each group of agents. In Appendix C.6, we include a detailed description of the decomposition in terms of the effects on aggregate variables.

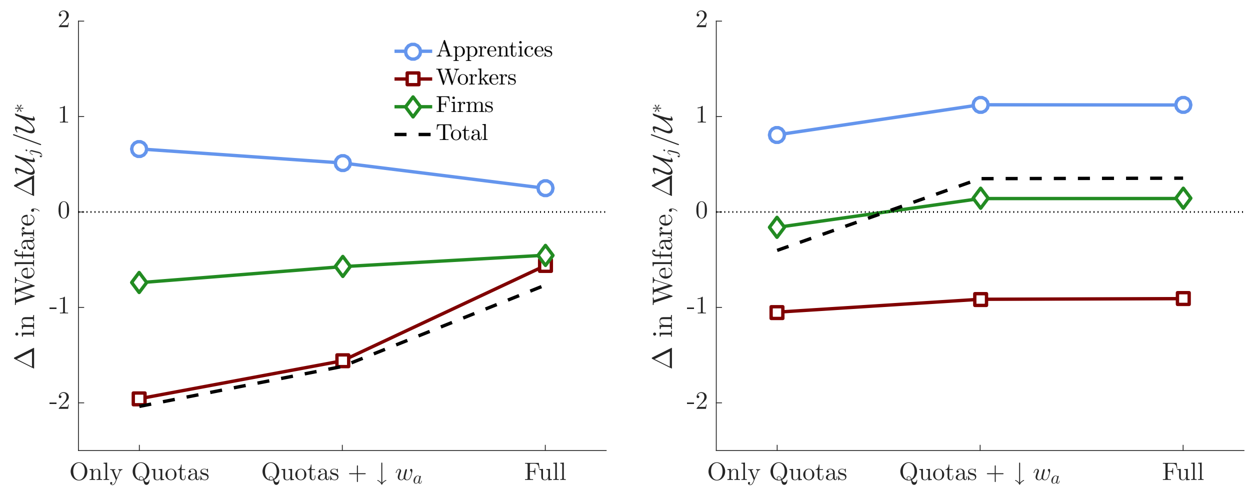
Figure 15 plots the change in aggregate and group-level welfare, considering three scenarios: (i) only quotas, (ii) quotas and the decrease in the apprentices' minimum wage, and (iii) the full regulation. As in Table 6 we compute the change in welfare as the change in aggregate income $\Delta\mathcal{U}_j$ for each group of agents $j \in$\{Apprentices, Workers, Firms\}, divided by total income $\mathcal{U}^*$ before the policy.

Panel 15a shows results for high-skill sectors, where the full regulation generates welfare gains for apprentices, but losses for both 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, welfare improves for workers and firms as firms can take advantage of lower apprentice wages and displace fewer workers. A few high-skill sector firms train more apprentices, but apprentices are now paid less, so the welfare gain of apprentices is slightly smaller. Finally, when the possibility of paying the fee is added and we consider the full policy, firms and workers again see smaller losses as those firms with the highest training cost now pay the fee and hence fire fewer workers. Fewer apprentices are trained, such that the welfare gain of apprentices decreases.

Panel 15b shows corresponding results for low-skill sectors. Here, apprentices and firms gain from the full regulation, but workers lose. Again, the decrease in the apprentice minimum wages reinforces gains for firms relative to a scenario with only quotas, but naturally lowers income gains of apprentices. In low-skill sectors, adding the possibility of paying fees has close to no effect, as those firms only rarely choose to pay fees.

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 substitute a significant part of their labor force 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 fewer apprentices are trained in those sectors. With these results in mind, we next propose alternative policies that explicitly take into account heterogeneity in training costs across sectors.

Figure 15: Policy Decomposition



(a) High-Skill

(b) Low-Skill

Notes: The figure shows the change in aggregate utility of apprentices, incumbent workers and firm owners/managers, relative to the no-regulation total aggregate utility, $\mathcal{U}^*$.

5.6 Alternative Apprenticeship Policies

In this section we study two alternative apprenticeship policies. First, we consider a policy that subsidizes training. In the spirit of the recently implemented Apprenticeship Levy in the UK, we examine a subsidy on training costs financed by payroll taxes. Second, we consider sector-specific minimum wages for apprentices. This policy resembles the situation in some European countries including Germany where there are no general restrictions on apprentices' wages, but there may be sector-specific agreements (Steedman 2012). To make the policies comparable to the benchmark Colombian regulation, we restrict the number of trained apprentices to be the same across the different scenarios. Both counterfactual policies remove firm-size distortions by disposing of the discontinuous apprentice quotas.

These exercises not only show that reducing these firm-size distortions increase welfare, but also that policies that take into account sector heterogeneity magnify these welfare gains.

Subsidizing Training in Firms

First, consider a policy that subsidizes training apprentices. We focus on budget-balanced policies, where firms are taxed to finance the subsidy. Firms pay taxes on their payroll at a rate τ, and these funds can be used to their training expenses. The government subsidizes an additional part of training by contributing ς% on top of the funds collected via taxes.

In this case the firm (z, t_a) solves

$$\max_{n_a, n, \mathcal{S} \geq 0} pz^{1-\gamma} \left((n - t_a n_a) + n_a \right)^{\gamma} - w(1+\tau)n - w_a n_a + \mathcal{S}$$
$$\text{s.t} \quad t_a n_a \leq n \;\; \text{and} \;\; \mathcal{S} \leq \min\{w t_a (1+\tau) n_a, w\tau(1+\varsigma)n\}, \tag{8}$$

where $\mathcal{S}$ denotes the subsidy the firm gets to cover part of its training expenses. This subsidy $\mathcal{S}$ can cover at most the firm's total training costs $wt_a(1+\tau)n_a$, and cannot exceed the total amount of tax paid plus the government's contribution, $w\tau(1+\varsigma)n$. To solve (8), firms always choose the largest subsidy $\mathcal{S}$ possible, such that the second constraint becomes an equality. 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' wages is large.[39]

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 5). This yields a tax rate of $\tau = 0.11\%$ and a large subsidy of $\varsigma = 25.6$. This large subsidy reflects levy funds only partially cover the costs of training apprentices. This provides a rationale to the observed surplus funds observed for the UK apprenticeship policy, where as of January of 2020 less than 35% of the levy-paying firms used their monthly funds to train apprentices.[40]

Table 8 shows total tax revenue and subsidy payments by sector. Firms in high-skill sectors pay more taxes than those in low-skill sectors, but most of the subsidy goes to low-skill sector firms, where training costs are lower. Hence, high-skill sector firms effectively subsidize low-skill sector firms. Intuitively, if firms in both sectors face the same tax and subsidy rates, it is still disproportionately more attractive to firms with low training costs to train apprentices, which benefits the low-skill sector. The policy attempts to cover the training costs, but the budget balanced subsidy is not large enough to cover the full costs of training (apprentices' wages and training costs) for many high-skill sector firms.

[39] See Lemma 7 in Appendix C.7.1 for a full characterization of the solution. In Appendix C.7.2 we show that the analysis and results are analogous if the government subsidizes the nominal cost of hiring apprentices $w_a n_a$ instead.

[40] Data from Freedom of Information request response (16 Jan 2020) from the Dept. for Education reported by FEnews article.

Table 8: Budget Balance, Apprenticeship Levy

	Tax Revenue	Total Subsidy	Net Balance
High-skill Sectors	907	192	715
Low-skill Sectors	717	1432	-715
Total	1624	1624	-0

Notes: Tax revenue and total subsidies paid to firms to train the same number of apprentices as in the benchmark policy.

Table 9: Aggregate Variables, Apprenticeship Levy

	Workers	% Workers	Apprentices	% Output
High-skill Sectors	-750	-0.35	1058	-0.14
Low-skill Sectors	-6403	-3.04	21083	-0.06
Total	-7152	-1.69	22141	-0.10

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.

Accordingly, Table 9 shows that even though the apprenticeship levy eliminates the firm-size distortions of the benchmark policy, it implies a larger decrease in aggregate output. Despite training the same total number of apprentices, the policy also displaces more workers. The benefits from the government subsidies concentrate only in a few firms and are not enough to compensate the decrease in production by the non-training firms. The percentage of firms with at least one apprentice decreases from around 60% of firms to only 5%. Firms that seek apprentices often choose as many as possible, hiring only as many workers necessary to train these apprentices. Overall, the training subsidy avoids firm size distortions as it does not feature quotas based on firm size, but it actually leads to stronger reallocation of resources towards low-skill sectors than the benchmark policy. This counterfactual highlights the importance of minimum and maximum apprentice quotas when regulation is the same for all firms, as quotas induce a more even distribution of apprenticeship training across firms both within and across sectors.[41]

[41] Note that we do not argue here that subsidizing apprenticeship training lowers welfare compared to no regulation. Our comparison focuses on the subsidy vs. the actual regulation in Colombia.

Sector-Specific Apprentice Minimum Wage

Next, we consider an alternative policy that explicitly takes into account heterogeneity in training costs in order to address this issue. The second counterfactual allows apprentices' minimum wages to differ across sectors. To discipline this exercise, we compute the minimum wage such that in each sector the number of trained apprentices is again the same as in the benchmark regulation. As in the first counterfactual, there are no quotas or fees. First, we find the sector-specific minimum wage that clears the labor market for apprentices.[42] The apprentice minimum wages is $w_a^{*s} = 0.74$ times the minimum wage for regular workers in high-skill sectors and $w_a^{*u} = 0.94$ in low-skill sectors.

Table 10 shows the effects of the sector-specific minimum wage.[43] In high-skill sectors, fewer workers are displaced, but in low-skill sectors there is more displacement. In contrast to the benchmark policy, output increases in high-skill sectors *and* in low-skill sectors. Total output increases more strongly than in the benchmark policy, and the reallocation of resources towards low-skill sectors is substantially diminished.

Table 10: Aggregate Variables, Apprentice Minimum Wage by Sector w_a^k

	Workers	% Workers	Apprentices	% Output
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: 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.

Comparing Apprenticeship Policies

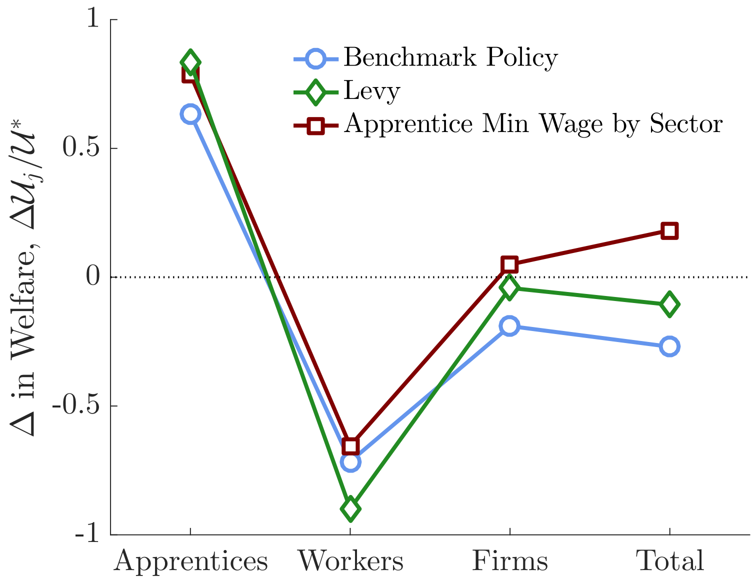
Finally, we compare the effects of the two counterfactual policies to the benchmark policy on the welfare of the different groups of agents. Figure 16 summarizes the results. Again, the change in welfare is measured as the variation in aggregate utility for apprentices, workers and firms, divided by the total aggregate utility without any apprenticeship policy. The figure shows that in terms of total welfare, the sector-specific apprentice minimum wage outperforms the other two. It is the only policy that produces overall welfare gains already in partial equilibrium. The levy implies training concentrates only on few firms, while displacing more workers than in the other two

[42] See Appendix C.7.3 for the analytical details.

[43] The total number of apprentices in tables 9 and 10 exhibit a very small difference due to slight approximation error in the simulation.

policies. Moreover, even without the firm size distortions, aggregate firm profits are still negative, as the effect of the levy taxes outweighs the positive effect of the training subsidy. In contrast, sector-specific minimum wages displace less workers, increasing output, profits and welfare across all agents.

Figure 16: Comparing Apprenticeship Policies: Change in Welfare



Notes: The figure shows the change in aggregate utility of apprentices, incumbent workers and firm owners/managers, relative to the no-regulation total aggregate utility, $\mathcal{U}^*$.

6 Final Comments

Policymakers in both developed and developing countries believe that well-designed apprenticeship policies have the potential to reduce youth unemployment and enhance the quality of jobs in the economy. Such policies can improve labor market outcomes of individuals with low initial productivity, who are often confined to informal labor markets in developing countries. Apprenticeships can also help narrow the mismatch between the demand and supply of skills, easing school-to-work transitions of young people. Since these training programs often yield external benefits not internalized by firms, there is space for government intervention. Combining reduced-form evidence based on a unique policy change and a structural model, our paper shows that widening the scope of training in firms can have potentially large positive effects on apprentices, firms, and aggregate output. In addition, we emphasize the importance of considering heterogeneity across firms and sectors when designing and implementing apprenticeship policies.

Two important implicit assumptions we make about labor markets and the quality of apprenticeship training are worth highlighting. First, we assume that workers displaced by apprentices are absorbed back into the formal labor market eventually. In reality, labor market frictions might attenuate or slow down these general equilibrium effects. If some displaced workers remain unemployed or end up in informal work, wages of formal workers would respond less, dampening some of the positive effects of the regulation. Second, the quality and content of the apprenticeship training also affects welfare quantifications (Hanushek et al. 2017). In successful apprenticeships programs, firms have incentives to provide high-quality training to apprentices because they can recoup their investment during the apprenticeship period, or because future turnover and recruitment costs are reduced. If firms facilitate knowledge transfer to apprentices, this ensures an adequate future supply of well-trained workers. However, if employers use apprentices only as "cheap labor", there may be more limited gains from these apprenticeship policies. Measuring the labor market outcomes of apprentices could be informative of the quality and content of their training. Extending the analysis in such directions can be fruitful areas for future research.

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.

Adda, J., Dustmann, C., Meghir, C., and Robin, J.-M. (2010). Career progression and formal versus on-the-job training. Technical report, IFS working papers.

Albanese, A., Cappellari, L., and Leonardi, M. (2017). The effects of youth labor market reforms: evidence from italian apprenticeships. *Luxembourg Institute of Socio-Economic Research (LISER) Working Paper*, (2017-13).

Attanasio, O., Guarín, A., Medina, C., and Meghir, C. (2015). Long term impacts of vouchers for vocational training: Experimental evidence for colombia. Technical report, National Bureau of Economic Research.

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.

Barrera-Osorio, F. and Corchuelo, A. (2003). Senas returns: A re-evaluation.

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.

Bertrand, M., Mogstad, M., and Mountjoy, J. (2019). Improving educational pathways to social mobility: Evidence from norways reform 94. Technical report, National Bureau of Economic Research.

Besley, T. and Burgess, R. (2004). Can labor regulation hinder economic performance? evidence from india. *The Quarterly journal of economics*, 119(1):91–134.

Bettinger, E., Kremer, M., Kugler, M., Medina-Durango, C. A., Posso-Suárez, C. M., Saavedra, J. E., and Medina, C. (2019). School vouchers, labor markets and vocational education. *Borradores de Economía; No. 1087.*

Bonnal, L., Mendes, S., and Sofer, C. (2002). School-to-work transition: Apprenticeship versus vocational school in france. *International Journal of Manpower*, 23(5):426–442.

Cappellari, L., DellAringa, C., and Leonardi, M. (2012). Temporary employment, job flows and productivity: A tale of two reforms. *The Economic Journal*, 122(562):F188–F215.

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.

Corseuil, C. H., Foguel, M. N., and Gonzaga, G. (2019). Apprenticeship as a stepping stone to better jobs: Evidence from brazilian matched employer-employee data. *Labour Economics*, 57:177–194.

Dabla-Norris, M. E., Jaramillo, L., Lima, F., and Sollaci, A. (2018). *Size Dependent Policies, Informality and Misallocation*. International Monetary Fund.

Dionisius, R., Muehlemann, S., Pfeifer, H., Walden, G., Wenzelmann, F., Wolter, S. C., et al. (2009). Costs and benefits of apprenticeship training. a comparison of germany and switzerland. *Applied Economics Quarterly (formerly: Konjunkturpolitik)*, 55(1):7–37.

Fazio, M. V., Fernández-Coto, R., and Ripani, L. (2016). Apprenticeships for the xxi century: A model for latin america and the caribbean?

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.

Fersterer, J. and Winter-Ebmer, R. (2003). Are austrian returns to education falling over time? *Labour Economics*, 10(1):73–89.

Fudenberg, D. and Rayo, L. (2017). Training and effort dynamics in apprenticeship.

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 Mendez, J. A. N. (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.

Gourieroux, M., Gourieroux, C., Monfort, A., Monfort, D. A., et al. (1996). *Simulation-based econometric methods*. Oxford university press.

Groot, W., Plug, E., et al. (1998). Apprenticeship versus vocational education: Exemplified by the dutch situation.

Guner, N., Ventura, G., and Xu, Y. (2008). Macroeconomic implications of size-dependent policies. *Review of Economic Dynamics*, 11(4):721–744.

Hanushek, E. A., Schwerdt, G., Woessmann, L., and Zhang, L. (2017). General education, vocational education, and labor-market outcomes over the lifecycle. *Journal of Human Resources*, 52(1):48–87.

Kleven, H. J. (2016). Bunching. *Annual Review of Economics*, 8:435–464.

Kleven, H. J. and Waseem, M. (2013). Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan. *The Quarterly Journal of Economics*, 128(2):669–723.

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.

Levinsohn, J. and Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies*, 70(2):317–341.

Lucas, R. E. (1978). On the size distribution of business firms. *The Bell Journal of Economics*, pages 508–523.

McIntosh, S. (2007). *A Cost-Benefit Analysis of Apprenticeships and Other Vocational Qualifications*, volume 834. Department for Education and Skills Nottingham.

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.

Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica: Journal of the Econometric Society*, 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*.

Parey, M. (2016). Vocational schooling versus apprenticeship training. evidence from vacancy data.

Riphahn, R. T. and Zibrowius, M. (2016). Apprenticeship, vocational training, and early labor market outcomes–evidence from east and west germany. *Education Economics*, 24(1):33–57.

Ryan, P. (1998). Is apprenticeship better? a review of the economic evidence. *Journal of Vocational Education & Training*, 50(2):289–325.

Ryan, P. (2001). The school-to-work transition: a cross-national perspective. *Journal of Economic Literature*, 39(1):34–92.

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.

Soskice, D. (1994). Reconciling markets and institutions: The german apprenticeship system. In *Training and the Private Sector: International Comparisons*, pages 25–60. University of Chicago Press.

Steedman, H. (2010). The state of apprenticeship in 2010: International comparisons-australia, austria, england, france, germany, ireland, sweden, switzerland. *Apprenticeship Ambassadors Network*.

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.

Zimmermann, K. F., Biavaschi, C., Eichhorst, W., Giulietti, C., Kendzia, M. J., Muravyev, A., Pieters, J., Rodríguez-Planas, N., Schmidl, R., et al. (2013). Youth unemployment and vocational training. *Foundations and Trends® in Microeconomics*, 9(1–2):1–157.

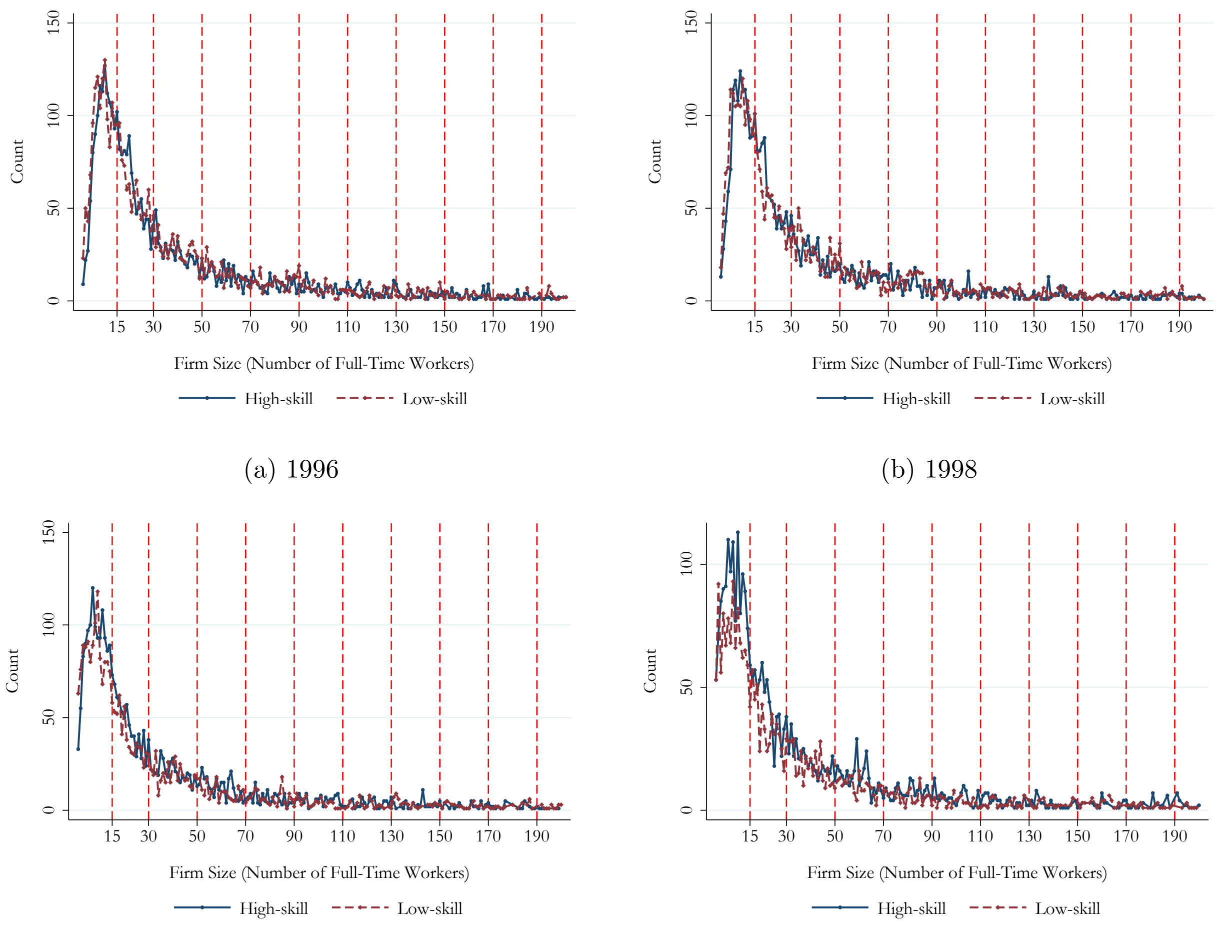
Appendix

A Additional Figures and Tables Empirical Analysis

A.1 Firm Size Distribution by Year

Figure A.1 shows the firm size distribution for high-skill and low-skill sectors is smooth for pre-reform years (1996, 1998, 2000 and 2002). Although the plots are more coarse given there are less observations, there are no visible bunching patterns around the thresholds.

Figure A.1: Firm Size Distribution Pre-Reform



(a) 1996

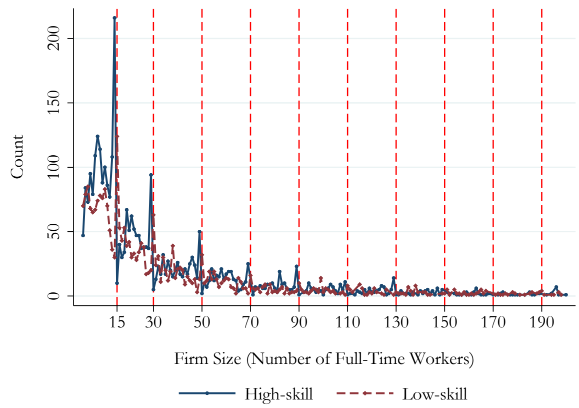
(b) 1998

(c) 2000

(d) 2002

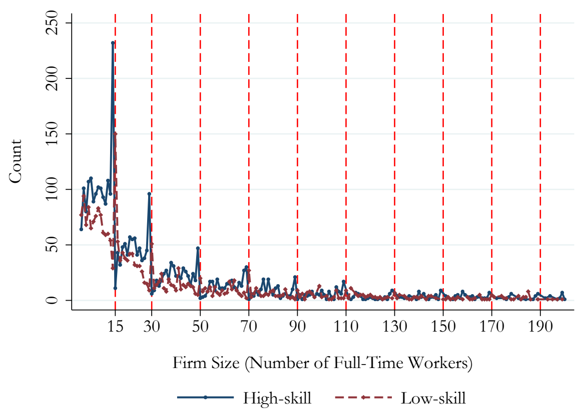
The firm size distribution in both sectors is remarkably similar. In contrast, if we look at the post reform years (Figure A.2) we see evidence of firms in high-skill sectors bunching below and firms in low-skill sectors at the threshold. This behavior is similar across the post reform years.

Figure A.2: Firm Size Distribution Post-Reform

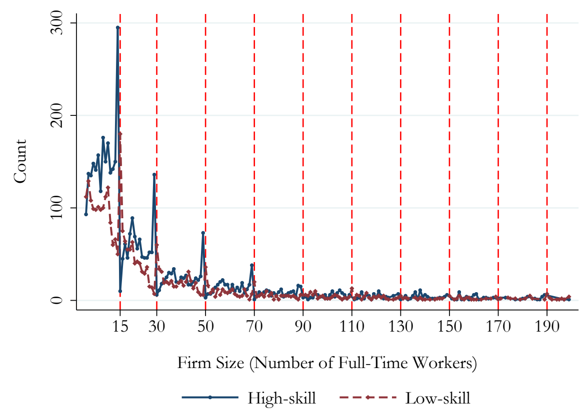


(a) 2003

(b) 2005



(c) 2007



(d) 2009

Table A.1: Reduced-Form Effects of the Policy on Firm Outcomes

Panel A: Workers by type

	High-Skill	Low-Skill	High-Skill	Low-Skill	High-Skill	Low-Skill	High-Skill	Low-Skill
	Workers		Professionals		Admin wrks		Production wrks	
Above*Post	-0.855	-1.781*	-0.0556	-0.0649	-0.199	-0.956**	-0.813	-0.898
	(0.863)	(1.036)	(0.224)	(0.185)	(0.359)	(0.445)	(0.652)	(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 by type (survey)

	High-Skill	Low-Skill	High-Skill	Low-Skill	High-Skill	Low-Skill	High-Skill	Low-Skill
	Workers		Workers (Survey)		Temporary		Outsourced	
Above*Post	-0.855	-1.781*	-0.757	-2.372*	0.604	0.392	-0.131	2.134
	(0.863)	(1.036)	(0.891)	(1.304)	(0.643)	(1.805)	(1.137)	(2.247)
Mean (Pre-reform)	30.46	30.30	31.46	32.33	3.329	4.831	4.719	7.299
Observations	8491	6357	8491	6357	8491	6357	8491	6357
R-squared	0.904	0.894	0.907	0.899	0.687	0.597	0.752	0.880

Panel C: Output, profits, value added

	High-Skill	Low-Skill	High-Skill	Low-Skill	High-Skill	Low-Skill	High-Skill	Low-Skill
	Log Output		Profit Rate		Log Sales		Log Value Added	
Above*Post	-0.0302	-0.00697	-0.0135	-0.00174	-0.0293	-0.0206	-0.0429	0.0190
	(0.0355)	(0.0472)	(0.0113)	(0.0107)	(0.0384)	(0.0512)	(0.0423)	(0.0553)
Mean (Pre-reform)	14.31	14.46	0.212	0.193	14.59	14.79	13.52	13.49
Observations	8491	6357	8491	6357	8422	6159	8491	6357
R-squared	0.866	0.862	0.527	0.488	0.842	0.852	0.814	0.786

Panel D: Other inputs

	High-Skill	Low-Skill	High-Skill	Low-Skill	High-Skill	Low-Skill	High-Skill	Low-Skill
	Log Capital		Log Intermediates		Log Energy		Log Water	
Above*Post	0.00349	0.0272	-0.0268	-0.00554	-0.0228	-0.0112	-0.0304	0.0298
	(0.0483)	(0.0689)	(0.0382)	(0.0502)	(0.0433)	(0.0497)	(0.0562)	(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

Table A.1: Reduced-Form Effects of the Policy on Firm Outcomes (continued)

Panel E: Wages

	High-Skill	Low-Skill	High-Skill	Low-Skill	High-Skill	Low-Skill	High-Skill	Low-Skill
	Log Wage bill (perm.)		Wage p.w. (perm.)		Log Wage bill		Wage p.w.	
Above*Post	-0.0306	-0.0368	465.3	-12.70	-0.00279	0.0207	-54.36	-32.91
	(0.0339)	(0.0442)	(455.1)	(572.3)	(0.0310)	(0.0391)	(261.2)	(247.4)
Mean (Pre-reform)	12.67	12.57	16574.2	14845.2	12.85	12.75	15558.9	12953.0
Observations	8473	6312	8376	6211	8491	6357	8490	6356
R-squared	0.840	0.814	0.802	0.766	0.863	0.843	0.770	0.749

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

A.2 Low-skill and High-skill Sectors Pre-Reform

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

	Observations	% Prof	Average Yearly Wages All Workers	Average Yearly Wages Prod. Workers	% Perm	K/Y
Low-skill						
Wood	2,543	3.5	11,161	10,210	84	30.66
		(8.0)	(5,657)	(4,406)	(25)	(44.32)
Textiles	12,354	4.9	10,955	10,581	73	56.60
		(12.5)	(4,938)	(5,879)	(36)	(131.13)
Food and Beverages	11,159	5.3	13,846	12,676	81	58.13
		(10.0)	(8,638)	(7,538)	(26)	(100.52)
Mineral Non-Metallic	2,689	7.0	16,398	14,664	82	62.96
		(11.5)	(11,237)	(9,087)	(27)	(112.06)
Total Low-skill	28,745	5.2	12,604	11,815	78	55.50
		(11.1)	(7,575)	(7,001)	(31)	(112.80)
High-skill						
Other Manufacturing	2,543	7.3	14,931	14,075	77	52.79
		(13.6)	(10,876)	(10,178)	(31)	(106.16)
Paper and Editorial	12,354	7.4	16,108	14,642	87	52.73
		(13.7)	(8,704)	(7,930)	(22)	(106.64)
Metallic	11,159	7.4	19,494	17,351	77	70.93
		(10.4)	(11,996)	(10,181)	(29)	(136.63)
Machinery	2,689	8.4	15,084	13,046	83	46.99
		(13.6)	(7,908)	(6,872)	(26)	(77.63)
Chemical	5,118	8.6	18,893	15,049	85	65.84
		(11.8)	(12,667)	(9,836)	(24)	(102.49)
Total High-skill	28,949	8.0	16,392	14,158	83	54.79
		(13.0)	(10,331)	(8,704)	(26)	(97.21)

Note: Data divided into two digit manufacturing industries using the CIIU Revision 3 codes from DANE. Low-skill and high-skill sectors are defined according to ranking of fraction of professionals out of the total workforce. Wages in 2009 pesos, units in thousands. Standard deviations in parentheses. *Source*: Data from SENA and EAM.

Table A.2: Summary of Responses to Regulation

Compliance Groups	Sector					
	Low-Skill		High-Skill		Total	
	No.	%	No.	%	No.	%
Upper Bound	7,727	65.4%	212	1.5%	7,939	30.2%
Lower Bound	3,185	26.9%	4,615	31.9%	7,800	29.7%
Between Bounds	832	7.0%	446	3.1%	1,278	4.9%
Below Lower Bound	40	0.3%	0	0.0%	40	0.2%
Above Upper Bound	5	0.0%	0	0.0%	5	0.0%
No Apprentices	35	0.3%	9,197	63.6%	9,232	35.1%
Total	11,824	100.0%	14,470	100.0%	26,294	100.0%

Note: Number of observations (firm-year) in the post-reform period (2003-2009). Only firms with more than 14 full-time workers. *Source*: own computations using data compiled by SENA.

Table A.3: Correlates of Responses to Regulation

	Choose max. quota	Choose min. quota	Between min./max.	Pay fee to avoid app.
High-Skill Sector	-0.63***	0.051***	-0.045***	0.58***
	(0.0100)	(0.011)	(0.0035)	(0.0050)
Number of Workers	0.00022***	-0.00018***	-0.000022	-0.0000021
	(0.000061)	(0.000057)	(0.000031)	(0.000025)
Fraction Professionals	0.046	-0.051	0.029*	-0.068**
	(0.041)	(0.050)	(0.017)	(0.032)
Fraction Production Workers	-0.068***	0.068***	0.017**	-0.032**
	(0.023)	(0.025)	(0.0078)	(0.013)
Wage p.w.	0.0013***	-0.0012**	-0.000026	0.000046
	(0.00045)	(0.00050)	(0.00015)	(0.00021)
Log Output	-0.018***	0.00081	0.017***	-0.00098
	(0.0064)	(0.0066)	(0.0026)	(0.0034)
Output p.w.	-0.00013***	0.00015***	-0.000029***	-0.000018
	(0.000031)	(0.000033)	(0.0000088)	(0.000015)
Profit Rate	-0.17***	0.24***	-0.044***	-0.036
	(0.048)	(0.053)	(0.016)	(0.027)
Capital/Output	-0.054***	0.049***	0.0026	-0.0039
	(0.0081)	(0.0085)	(0.0029)	(0.0039)
Intermediates/Output	-0.20***	0.25***	-0.057***	0.0088
	(0.045)	(0.049)	(0.017)	(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

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

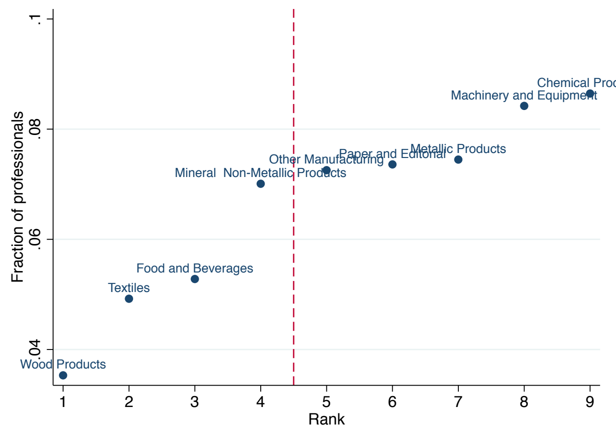
Table A.4: Correlates of Bunching Behavior

	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

Includes only post-reform years 2003 to 2009.

Figure A.3 shows the fraction of professionals for the ranked manufacturing sectors. The dotted line represents the of the nine sectors. We defined the sectors below this median low-skill sectors and sectors above, high-skill sectors. Using this definition the order ranking of sectors is: four low-skill sectors (Wood Products, Textiles, Food and Beverage, Mineral Non-Metallic Products) and five high-skill sectors (Paper and Editorial, Other Manufacturing, Metallic Products,Machinery and Equipment, Chemical Products).

Figure A.3: Fraction of Professional Workers by Sector

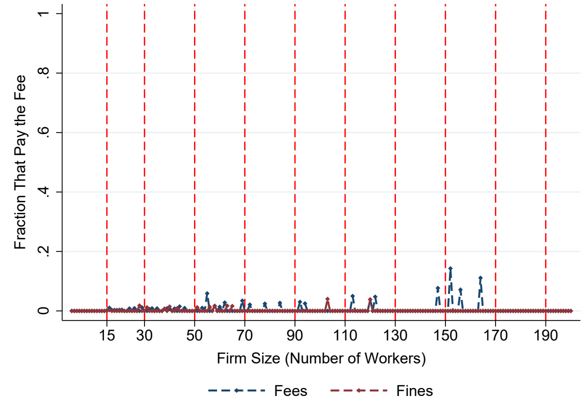


A.3 Fees and Fines

Figure A.4 shows the fraction of firms that pay fees and fines in each sector. Only a small fraction of firms pay the fines in either sector. This high compliance can be explained by the additional information SENA has on the number of workers in each firm, given they manage part of the parafiscal payments.

Figure A.4: Firm Size Distribution Pre-Reform

(a) High-skill

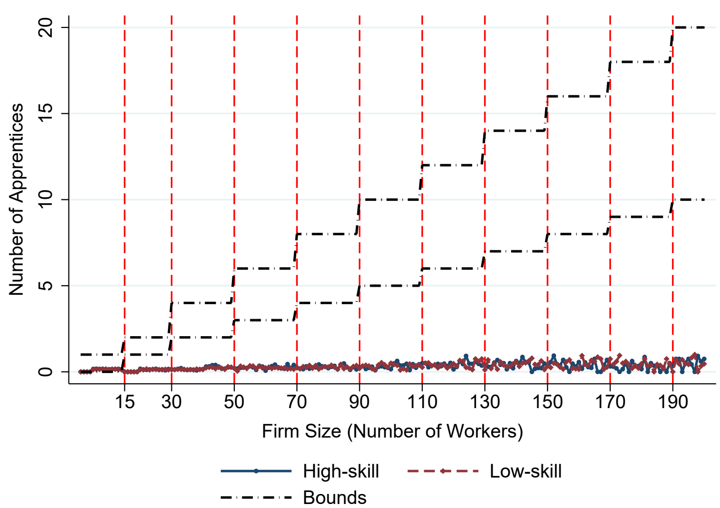


(b) Low-skill

A.4 Apprentices Pre-Reform

Figure A.5 shows the average number of apprentices for firms of different size both in low-skill and high-skill sectors. This average number of apprentices is significantly lower than in the period after the regulation. The figure shows a mildly increasing relation between number of workers and apprentices. Remarkably there aren't any clear differences between firms in the low-skill and the high-skill sectors before the change in regulation.

Figure A.5: Apprentices Pre-Reform by Skill



B Model Proofs and Extensions

B.1 Proofs

Lemma 1. *Assumptions 1 imply there are unique labor demands $n^*(z), n_a^*(z) > 0$, with $t_a n_a^* < n^*$ 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, $\frac{\partial n^*(z)}{\partial z}, \frac{\partial n_a^*(z)}{\partial z} \geq 0$. □

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

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

Intuitively the marginal cost of an apprentice is not only its wage w_a but also t_a units of time of a worker that earns wage w. The firm optimizes where the marginal product of an additional apprentice is equal to this marginal cost, $w_a + t_a w$. In equilibrium, the marginal rate of substitution between the two types of labor is equal to the ratio of marginal labor costs,

$$-\frac{\frac{\partial f}{\partial l}}{\frac{\partial f}{\partial l_a}} = -\frac{w\zeta_a}{w_a + t_a w}. \tag{9}$$

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

□

Lemma 3. *Let $(n^*(z,t_a), n_a^*(z,t_a))$ denote the optimal number of workers and apprentices a firm (z,t_a) hires when solving the maximization problem (1) (without the regulation).*

i. If $n_a^*(z,t_a) > \overline{n}_a^j$, then either $n^r(z,t_a) = N_k$ for $k \geq j+1$ and $n_a^r(z,t_a) > \overline{n}_a^j$ (increase size to get more apprentices) or $n_a^r(z,t_a) = \overline{n}_a^j$ and $n^r(z,t_a) < n^*(z,t_a)$ (bounded by maximum quota).

ii. If $n_a^*(z,t_a) \in \left[\underline{n}_a^j, \overline{n}_a^j\right]$, 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 either $n^r(z,t_a) = N_k - \varepsilon$ for $k \leq j$ and $n_a^r(z,t_a) < \overline{n}_a^j$ (reduce size to avoid apprentices); $n^r(z,t_a) \geq N_j$ and $n_a^r(z,t_a) < \overline{n}_a^j$ and $d_f = 1$ (pay the fee to avoid apprentices) or $n_a^r(z,t_a) = \underline{n}_a^j$ (bounded by minimum quota).

Proof. Choose any firm $z > 0$ and $t_a \in (0,1)$. 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,

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

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, $\frac{\partial \pi(N_a)}{\partial N_a} = pf_{l_a}\zeta_a - w_a$. We can differentiate again this expression with respect to N_a to obtain,

$$\frac{\partial^2 \pi(N_a)}{\partial N_a^2} = p\left(f_{l_a l} \frac{dl^r}{dN_a} + f_{l_a l_a}\zeta_a\right). \tag{10}$$

Where l^r solves the FOC of the fixed N_a optimization problem, $pf_l(l^r, N_a) = w$. Assumptions (1) imply the existence and uniqueness of the solution l^r. Totally differentiating this FOC wrt l^r and N_a, implies $\frac{dl^r}{dN_a} = -\frac{f_{ll_a}}{f_{ll}}\zeta_a \geq 0$, the inequality given $f_{ll_a} \geq 0$ and $f_{ll} \leq 0$. Replacing this derivative in (10),

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

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

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

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

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

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

Considering the FOC and totally differentiating,

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

Substituting in the previous equation,

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

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 $\overline{n}_a^{j+1}$, 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 $n_a^r \in [\overline{n}_a^{k-1}, \overline{n}_a^k]$. The optimal number of apprentices is in this case $n_a^r(z, t_a) > \overline{n}_a^j$.

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,

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

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 $\lim_{\phi_a \to 0} \pi^*(\phi_a) = \pi^*$. So there exist $\tilde{\phi}_a > 0$ small enough, such that $\pi^j(\tilde{\phi}_a) \geq \pi(N)$ and $\pi^j(\tilde{\phi}_a) \geq \pi(N_a)$.

□

Proposition 1. *Suppose Assumptions 1 holds and firms solve the maximization problem with regulation (6). Then,*

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

- *i. the number of apprentices without regulation is $n_a^* = B_u n^*$ and lays above the maximum quota, $n_a^*(z, t_a) > \overline{n}_a^j$.*
- *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_{j+1}, so there is missing mass to the left of the thresholds.*
- *iii. firms choose the upper-bound of the regulation $n_a^r = \overline{n}_a^j$.*
- *iv. firms never pay the fee.*

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

- *i. the number of apprentices without regulation is $n_a^* = B_s n^*$ and lays 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 the threshold N_j.*
- *iii. firms that choose to increase apprentices, choose the minimum number $\underline{n}_a^j$.*
- *iv. there exist $\overline{\phi}_a > 0$ such that for $\phi_a \leq \overline{\phi}_a$, 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. Use Lemma (2) for comparative statics, then Lemma (3) for cases. □

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,

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

the first line 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, $\pi_f > \pi_a \Rightarrow \frac{\partial f}{\partial n_a} < 0$. □

B.2 More Results

Lemma 4. *Under Assumptions 1, for each firm z there exists $A(z) > 0$ such that $l_a^* = A(z)l^*$.*

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

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

iii If $A'(z) < 0$, the parametric mapping $(n^*(z), n^*(z))$ 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) $\frac{dn^*}{dz} > 0$, 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.

□

B.3 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,

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

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.

B.4 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 net 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,

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

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 $A = \frac{w^i}{w^j}\frac{\gamma_j}{\gamma_i}$. 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 $n_a^j = 0 \;\; \forall j \neq i$. 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

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

B.5 CES Labor Input

More generally we suppose the production function of a firm (z, t_a) in sector k combines workers' and apprentices' labor in a CES function,

$$f^k(l, l_a; z, t_a) = z^{1-\gamma^k} \left[\eta^k l^{\rho^k} + (1-\eta^k) l_a^{\rho^k} \right]^{\frac{\gamma^k}{\rho^k}},$$

where $\gamma^k \in (0,1)$, $\eta^k \in (0,1)$ and $\frac{1}{1-\rho^k}$ is the elasticity of substitution between the two labor inputs. The production function in each sector can be different, so all of these parameters are sector specific.[44]

The FOC for this CES case imply the ratio of apprentices to workers is constant,

$$\frac{l_a}{l} := \frac{n_a}{n - t_a n_a} = \left(\frac{1-\eta^k}{\eta^k} \frac{w^k}{(w^k t_a + w_a^k)} \right)^{\frac{1}{1-\rho^k}} =: A_k \in \mathbb{R}_+ \Rightarrow \frac{n_a}{n} = \frac{A_k}{1 + A_k t_a^k} =: B_k$$

This ratio $\frac{n_a}{n}$ is decreasing in t_a and $\frac{w_a^k}{w^k}$ as Lemma 2 predicts.[45]

[44] If $\rho^k = 1$, then both type of are perfect substitutes once the apprentices are trained, if $\rho^k \to 0$ the total labor input is Cobb-Douglas and if $\rho^k \to -\infty$ labor inputs are perfect complements.

[45] We can explicitly solve for the labor demand functions, $n_a^*(z) = \left[\frac{\gamma^k \left[\eta^k A_k^{-\rho} + (1-\eta^k) \right]^{\frac{\gamma^k - \rho^k}{\rho^k}} \eta^k A_k^{1-\rho^k}}{w^k} \right]^{\frac{1}{1-\gamma^k}} z$, $n^*(z) = \frac{n_a^*(z)}{B_k}$.

B.6 Endogenous Training

The function $g : [0,1] \to [0, \zeta_a]$ denotes the training technology and can be interpreted as how difficult it is to teach/learn a particular task.

Firms might endogenously choose how much to train apprentices. Function $g(\cdot)$ reflects how difficult it is to train these apprentices. Suppose more training time makes apprentices more productive $g'(t_a) > 0$, but there are decreasing return to training $g''(t_a) < 0$. To solve for endogenous training, we add to firm z's original optimization problem (1), the choice of $t_a \in [0,1]$.

The FOC with respect to t_a implies,

$$f_l n_a = g'(t_a) f_{l_a} n_a,$$

so the marginal cost of training an apprentice is equalized to the marginal improvement in the apprentice abilities.

If $n_a^* \neq 0$ and using the FOCs of (1),

$$\frac{g'(t_a)}{g(t_a)} = \frac{w}{w_a + t_a w}. \tag{11}$$

Additionally suppose, $\lim_{t_a \to 0} g'(t_a) = \infty$ and $\lim_{t_a \to 1} g'(t_a) = 0$. The first condition states that supplying a small amount of training significantly improves apprentices productivity. The second that once workers allocate most of their unit of time to training apprentices, more training won't increase productivity too much. These two conditions together with the concavity of g, guarantee a unique interior solution $t_a^* \in (0,1)$ that solves the firms training decision.

Note that this training decision doesn't depend on the firm's managerial ability z. This means training per apprentice won't change for firms of different size or with different number of apprentices. Optimal training is only affected by changes in the training technology $g(\cdot)$ or wages. In particular, an increase in the apprentice wage w_a increases the optimal amount of training. Intuitively, hiring more apprentices becomes more costly so firms choose to train more each of the apprentices they hire. Conversely, an increase in workers wages w decreases the amount of training each apprentice gets. As the opportunity cost of training is the worker's wage, this opportunity cost increases as the wage of workers rise. Lemma 6 summarizes these results.

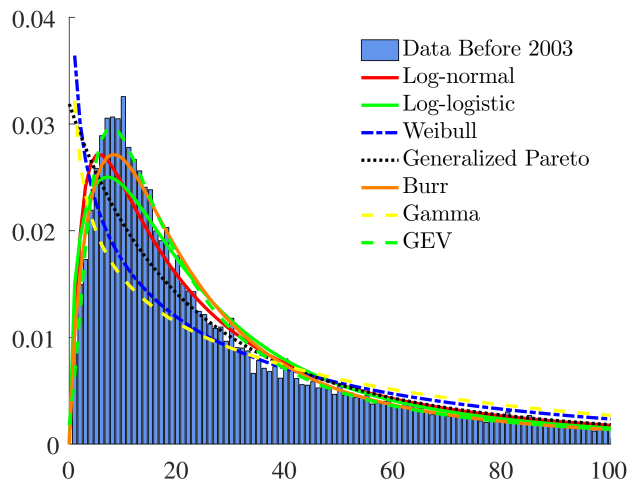
Lemma 6. *Suppose* $g'(t_a) > 0$, $g''(t_a) < 0$, $\lim_{t_a \to 0} g'(t_a) = \infty$ *and* $\lim_{t_a \to 1} g'(t_a) = 0$. *Then there exist a unique* $t_a^* \in (0,1)$ *that solves* (11). *Moreover,* $\frac{\partial t_a^*}{\partial w_a} > 0$ *and* $\frac{\partial t_a^*}{\partial w_a} < 0$.

C Quantitative Appendix

C.1 Fit Pre-Reform Firm Size distribution

Figure C.6 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 C.6: Fitting the Pre-Reform Distribution



C.2 Production Function Estimation

Table C.1 shows the estimated labor share by sector γ^k using six different methodologies. We suppose the production function depends on capital K, full-time labor l and other intermediate inputs m.

Table C.1: Labor Share γ^k, Full-Time Workers

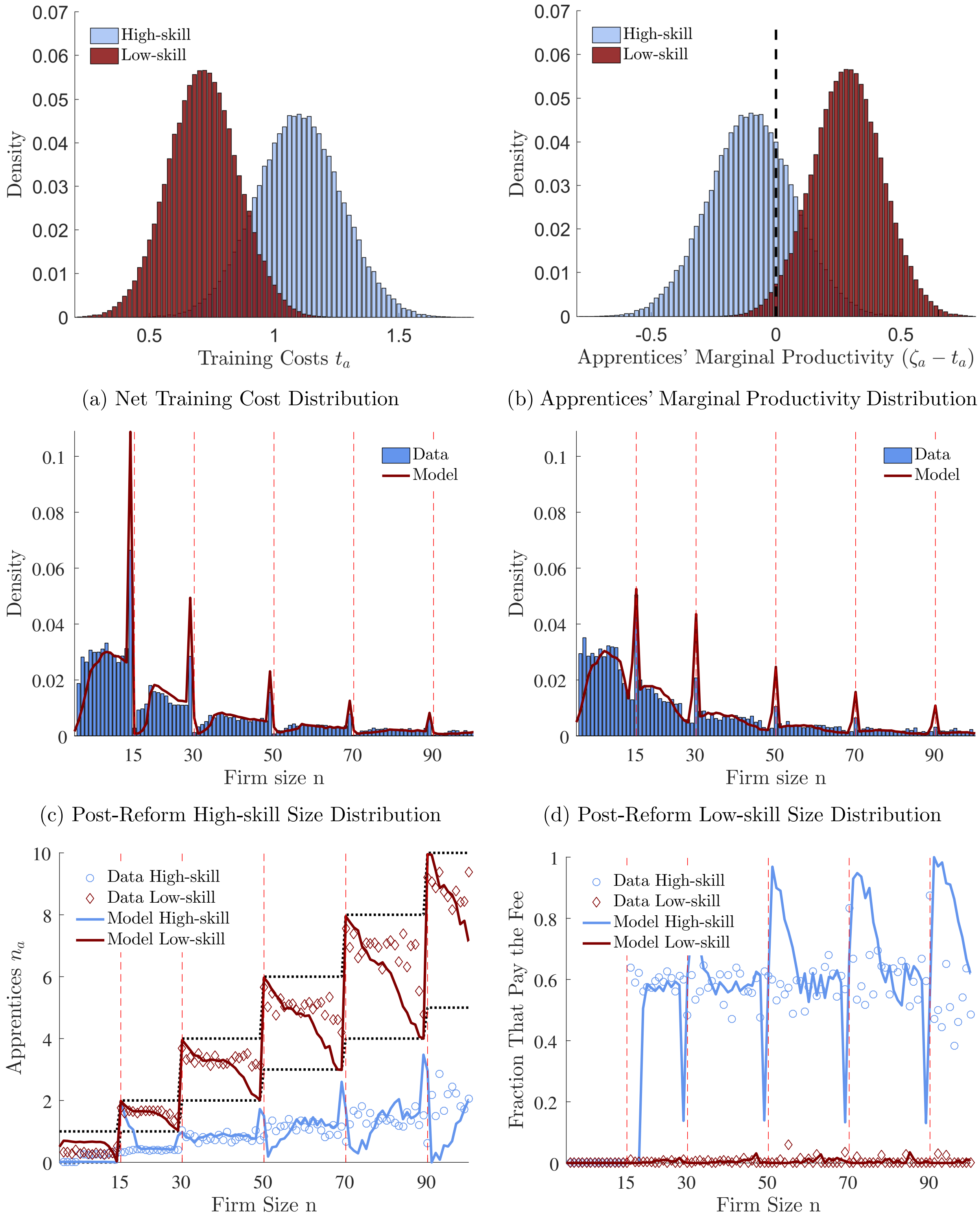
	OLS	FE	OP	LP	W	LP-ACF
High-Skill	0.70	0.64	0.51	0.42	0.23	0.49
Low-Skill	0.63	0.57	0.49	0.35	0.28	0.45

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

C.3 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 C.8) or a uniform (figure C.8) training cost distribution.

Figure C.7: Truncated Normal Training Distribution and Facts



(a) Net Training Cost Distribution

(b) Apprentices' Marginal Productivity Distribution

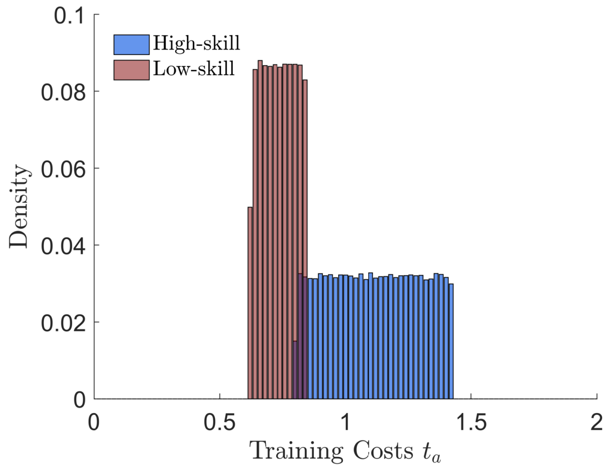
(c) Post-Reform High-skill Size Distribution

(d) Post-Reform Low-skill Size Distribution

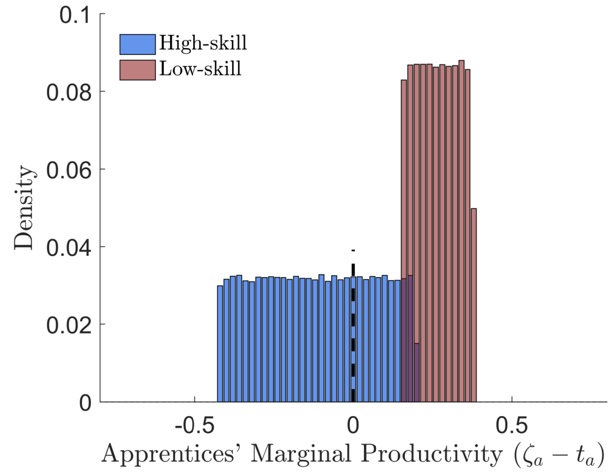
(e) Number of Apprentices

(f) Fraction that Pay the Fee

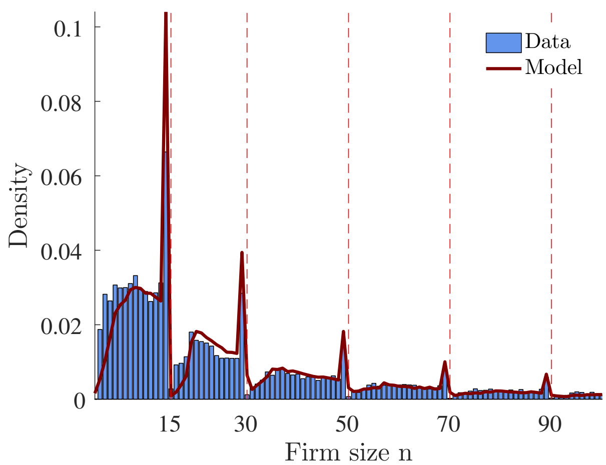
Figure C.8: Uniform Training Distribution and Facts



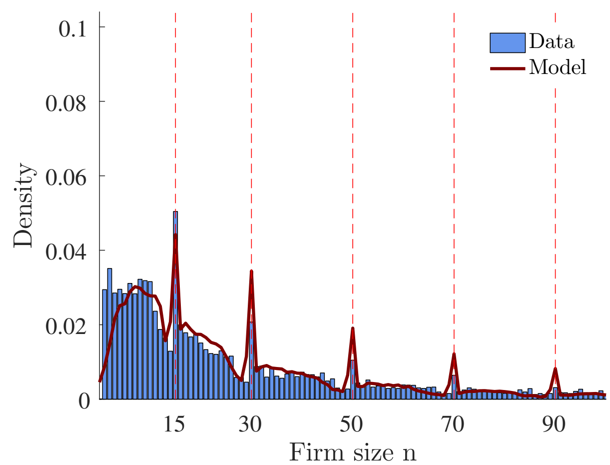
(a) Net Training Cost Distribution



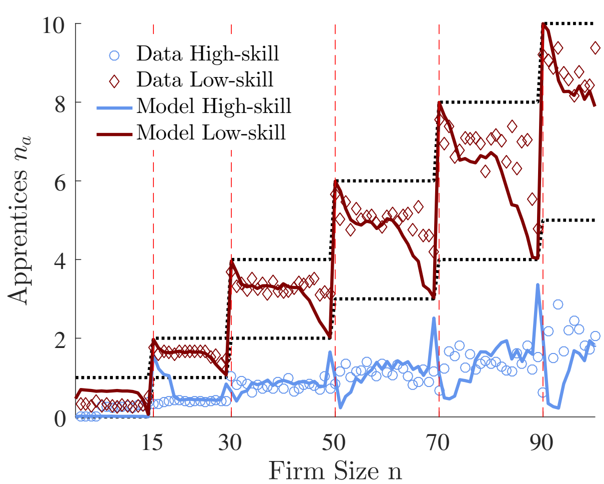
(b) Apprentices' Marginal Productivity Distribution



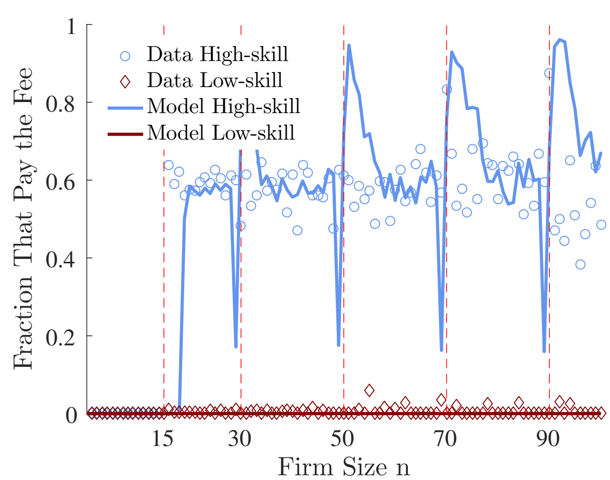
(c) Post-Reform High-Skill Size Distribution



(d) Post-Reform Low-Skill Size Distribution



(e) Number of Apprentices



(f) Fraction that Pay the Fee

C.4 Moment Weights

Table C.2 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^s_{14}. 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, $\omega^k_j = \frac{1}{4}\frac{1}{2}$ and that the fraction of firms that pay the fee, $\omega^k_j = 1/4$.

Table C.2: Moment Weights

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

C.5 DD Results: Full Sample

Table C.3 shows the results for the full sample. The results are qualitatively similar to those in our main specification. Firms above the threshold have more apprentices and less full-time workers. The negative coefficient on full-time workers is more negative than before. Output and profits are still small and not significant. However the point estimates are now positive. The coefficient plots in Figure C.9 show also show similar trends as in the main specification.

Table C.3: Effect Around the Thresholds: Both Sectors (Full Sample)

	Apprentices	Workers	Log of Output	Profit Rate Π/Y
Above*Post	1.030***	-4.318**	-0.0312	-0.00813
	(0.273)	(2.120)	(0.0279)	(0.00670)
Mean (Pre-reform)	0.197	60.77	7.560	0.218
Observations	20453	20453	20453	20453
R-squared	0.474	0.793	0.868	0.505

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.4: Effect Around the Thresholds: High-Skill Sectors (Full Sample)

	Apprentices	Workers	Log of Output	Profit Rate Π/Y
Above*Post	0.341***	-4.241*	-0.0417	-0.0124
	(0.110)	(2.444)	(0.0358)	(0.00981)
Mean (Pre-reform)	0.207	59.15	7.511	0.227
Observations	11289	11289	11289	11289
R-squared	0.282	0.808	0.878	0.528

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

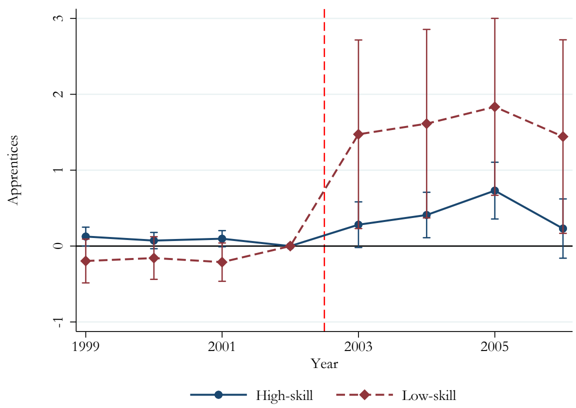
Table C.5: Effect Around the Thresholds: Low-Skill Sectors (Full Sample)

	Apprentices	Workers	Log of Output	Profit Rate Π/Y
Above*Post	1.728***	-4.998	-0.00662	-0.000404
	(0.577)	(3.565)	(0.0439)	(0.00893)
Mean (Pre-reform)	0.185	62.69	7.618	0.207
Observations	9164	9164	9164	9164
R-squared	0.525	0.797	0.871	0.498

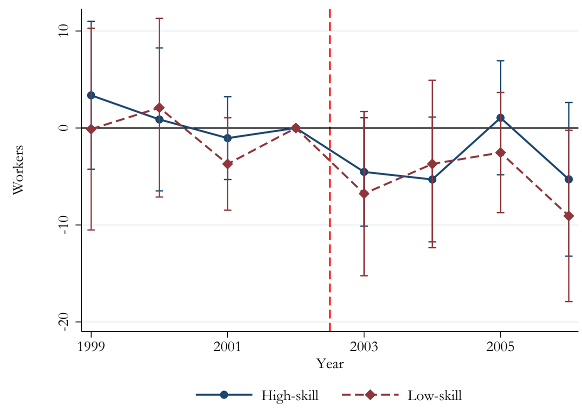
Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

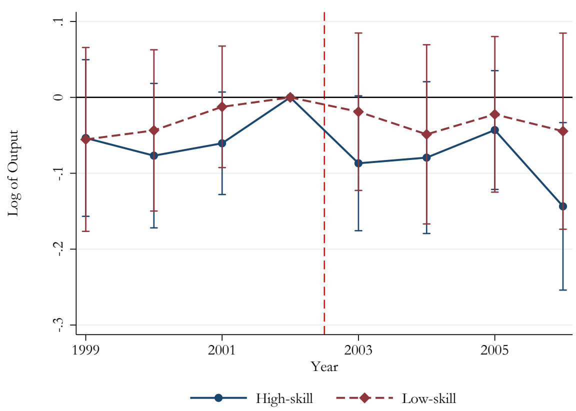
Figure C.9: Effect Around the Thresholds: Coefficient Plots (Full Sample)



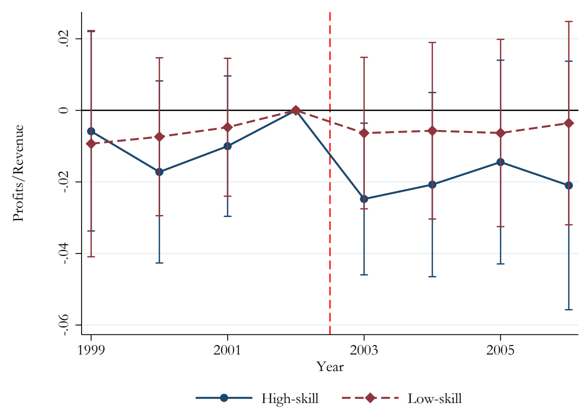
(a) Number of Apprentices



(b) Full-Time Workers



(c) Log of Output



(d) Profit Rate (Profits/Revenue)

C.6 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 C.6 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 apprentices 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 C.6A 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 C.6B 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 C.6C 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 C.6A), 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 C.6D 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 C.6: Policy Decomposition: Aggregate Variables

	ΔN	% ΔN	ΔN_a	% ΔY
A. Only Quotas				
High-skill	-7289	-1.90	17533	-1.25
Low-skill	-8625	-2.06	31434	0.19
Total	-15914	-1.98	48967	-0.61
B. Only $\downarrow w_a$				
High-skill	-647	-0.17	4620	0.06
Low-skill	-48059	-11.47	380064	9.46
Total	-48706	-6.07	384684	4.23
C. Quotas + $\downarrow w_a$				
High-skill	-6085	-1.59	17848	-1.07
Low-skill	-4101	-0.98	35456	0.81
Total	-10186	-1.27	53304	-0.24
D. Full Regulation				
High-skill	-2702	-0.71	9379	-0.37
Low-skill	-4083	-0.97	35405	0.81
Total	-6786	-0.85	44784	0.15

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.

C.7 Additional Results on Counterfactual Exercises

C.7.1 Subsidizing Training Costs $w(1+\tau)t_a$

Suppose the government covers the training costs $wt_a(1+\tau)n_a$, so long as it does not exceed the subsidized tax rebate, $w\tau(1+\varsigma)n$. Lemma 7 formalizes the results of subsidizing training,

Lemma 7. *Suppose a firm (t_a, z) solves (8),*

i. If $w_a + w(1+\tau)t_a - w\tau(1+\varsigma)t_a > w(1+\tau)$, then the firm avoids apprentices choosing, $n^* = \left(\frac{p\gamma}{w(1+\tau)}\right)^{1/(1-\gamma)} z$, $n_a^* = 0$, $\mathcal{S}^* = 0$.

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

a. If $1 \geq \tau\varsigma$, then $n^* = t_a n_a^*$, $n_a^* = \left(\frac{p\gamma}{\tilde{w}}\right)^{1/(1-\gamma)} z$, $\mathcal{S}^* = wt_a(1+\tau)n_a^*$.

b. If $1 < \tau\varsigma$, then $n^* = t_a n_a^*$, $n_a^* = \left(\frac{p\gamma}{\tilde{w}}\right)^{1/(1-\gamma)} z$, $\mathcal{S}^* = w\tau(1+\varsigma)t_a n_a^*$,
where $\tilde{w} := w_a + w(1+\tau)t_a - \min\{wt_a(1+\tau), w\tau(1+\varsigma)t_a\}$.

Case (i) reflects when the total 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. Consequently these firms reduce their labor demand to $n^* = \left(\frac{p\gamma}{w(1+\tau)}\right)^{1/(1-\gamma)} z$. 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 7. 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, $Rev(\tau,\varsigma) - Sub(\tau,\varsigma) = 0$.

C.7.2 Subsidizing Nominal Costs of Training w_a

We can prove an analogous result in case the government subsidize the nominal cost of training apprentices, $w_a n_a$.

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

$$\max_{n_a, n, \mathcal{S} \geq 0} pz^{1-\gamma}\left((n - t_a n_a) + n_a\right)^{\gamma} - w(1+\tau)n - w_a n_a + \mathcal{S}$$
$$\text{s.t} \quad t_a n_a \leq n \ \text{ and } \ \mathcal{S} \leq \min\{w_a n_a, w\tau(1+\varsigma)n\}, \tag{12}$$

where $\mathcal{S}$ is the subsidy the firm gets to pay for the apprentices training. Note we assume the government only covers the monetary costs of training apprentices, a maximum amount of $w_a n_a$.

To solve (12) firms choose the largest refund $\mathcal{S}$ possible. The linear labor inputs again imply corner solutions. Intuitively whenever the tax τ or the subsidy ς is large enough it will incentive firms to hire more apprentices. The incentives are stronger for firms with low training costs or in sectors where there is a large difference between average wages and the apprentice wages. We formalize this discussion in the following lemma.

Lemma 8. *Suppose a firm (t_a, z) solves (12),*

i. If $w_a > w(1-t_a)+w\tau(1+\varsigma t_a)$, then the firm avoids apprentices choosing, $n^* = \left(\frac{p\gamma}{w(1+\tau)}\right)^{1/(1-\gamma)} z$, $n_a^* = 0$, $\mathcal{S}^* = 0$.

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

a. If $w_a < w\tau(1+\varsigma)t_a$, then $n^* = t_a n_a^*$, $n_a^* = \left(\frac{p\gamma}{w(1+\tau)t_a}\right)^{1/(1-\gamma)} z$, $\mathcal{S}^* = w_a n_a^*$.

b. If $w_a > w\tau(1+\varsigma)t_a$, then $n^* = t_a n_a^*$, $n_a^* = \left(\frac{p\gamma}{w_a + w t_a(1-\tau\varsigma)}\right)^{1/(1-\gamma)} z$, $\mathcal{S}^* = w\tau(1+\varsigma)t_a n_a^*$.

C.7.3 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,

$$N_a^{*k} = \int\int n_a^{k,min}(z, t_a; w_a^{*k}, w^k)\ d\mathcal{Z}^k(z)\ d\mathcal{T}^k(t_a), \tag{13}$$

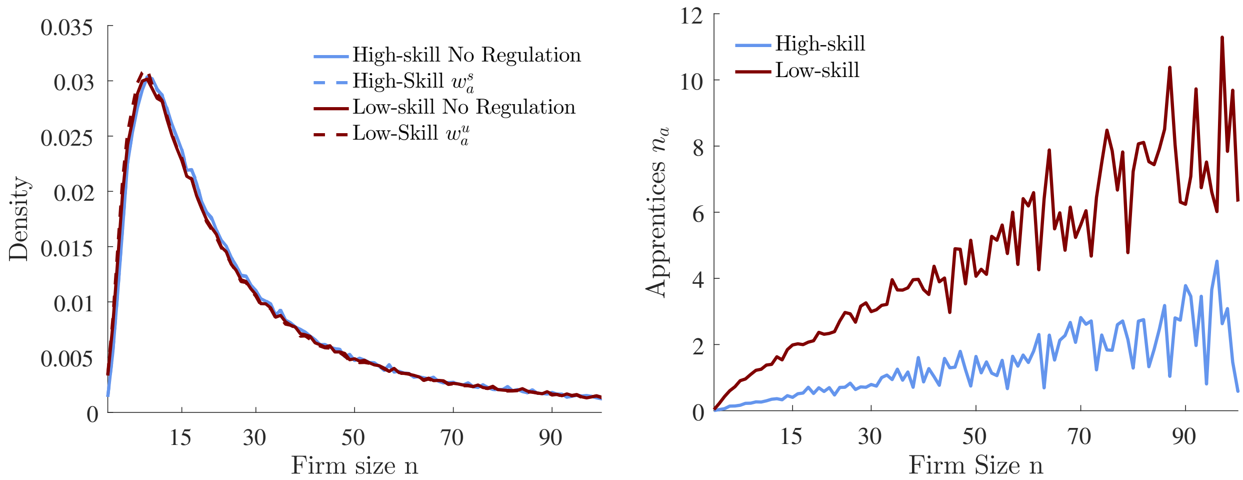
where N_a^{*k} denotes the aggregate number of apprentices trained in sector k when implementing the original policy. We assume each sector has the same worker wages as before and take the structural parameters from our estimated benchmark model.

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 (13) numerically and get an apprentice wage of $w_a^{*s} = 0.74$ for the high-skill sector and of $w_a^{*u} = 94$ for the low-skill sector.

Figure C.10: Apprentice Minimum Wage by Sector w_a^k



(a) Distribution of Full-Time Workers

(b) Average Number of Apprentices

Figure C.10 plots the distribution of full-time workers before and after the policy, and the average number of workers by firm size. Panel C.10(a) shows that this policy has a small effect on the distribution of full-time workers. This contrasts with the visible distortions documented in the firm bunching and responses to the original policy. The average number of apprentices also differs across the two policies. Without the quotas, the optimal number of apprentices grows linearly in $1/t_a$. The figure plots the average behavior of firms given the sector-specific apprentice wages w_a^k. Panel C.10(b) shows low-skill firms have on average more apprentices than high-skill sector firms and that this difference increase with firm size.