

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

Santiago Caicedo Miguel Espinosa Arthur Seibold[†]

July 2021

Abstract

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

JEL Codes: E24, J21, J24, M5

*We thank Ufuk Akcigit, Benito Arruñada, Leonardo Bursztyn, Jonathan Dingel, Christian Dustmann, Jan Eeckhout, Luis Garicano, Alessandro Gavazza, Maria Guadalupe, Ali Hortacsu, Greg Kaplan, Camille Landais, Bob Lucas, Rocco Macchiavello, Sandra McNally, Magne Mogstad, Massimo Motta, Carlos Ospino, Luis Rayo, Mar Reguant, Esteban Rossi-Hansberg, Uta Schönberg, Sandro Shelegia, Alexandra Spitz-Oener, Claudia Steinwender, Nancy Stokey, John Sutton, Chad Syverson, John van Reenen and numerous conference and seminar participants for helpful comments and discussions. We are grateful to the Colombian National Training Service (SENA) for their help and support of this project. Santiago Caicedo gratefully acknowledges financial support from Universidad de los Andes. Gabriela Díaz, David Guio, Ana María Patrón and Lukas Rodrian provided excellent research assistance.

[†]Caicedo: Clemson University, scaiced@clemson.edu; Espinosa: Universitat Pompeu Fabra and Barcelona GSE; and Bocconi University, Management and Technology Department, miguel.espinosa@unibocconi.it; Seibold: University of Mannheim, seibold@uni-mannheim.de

1 Introduction

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

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

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

The Colombian setting provides several advantages for our analysis. First, with youth unem-

¹Throughout this paper, we follow [Wolter and Ryan \(2011\)](#) and consider apprenticeships as “programs that comprise both work-based training and formal education, in most countries at upper-secondary level, and lead to a qualification in an intermediate skill” (p.523).

ployment rates of around 30% and youth informality rates above 70% in the early 2000s, young people in Colombia experienced labor market issues in common with many other low- and middle-income countries. Second, a reform in 2003 introducing a new apprenticeship regulation provides unique, large-scale variation in firms' incentives to train. The reform introduced apprentice quotas featuring multiple discontinuities in firm size, requiring firms to train apprentices between a minimum and maximum quota depending on their size. The reform also changed apprentices' minimum wages and introduced the possibility for firms to pay a costly fee as a "buy-out" from training apprentices. A third advantage is that high-quality administrative records at the firm level are available for the analysis, and they can be linked to a firm census containing additional rich survey information on firm characteristics.

We begin the analysis by showing reduced-form evidence of the effects of the apprenticeship regulation. On aggregate, the reform is highly successful in increasing the number of apprentices to more than fifteen times the pre-reform levels, allowing many more young individuals to receive training in the formal labor market. However, the reform also induces sizeable labor input responses and firm size distortions. At the firm level, we organize the results into three empirical facts documenting the strong heterogeneity of responses across sectors. First, we use bunching methods to gauge firm size responses to the discontinuities in apprentice 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 avoid the higher apprentice quota above the threshold. Meanwhile, firms in sectors with a small 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 firms reduce their size by two full-time workers in high-skill sectors, and increase their size by around 1.5 workers in low-skill sectors.

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 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. Around 60% of high-skill sector firms pay fees allowing them not to train any apprentices. In low-skill sectors, on the other hand, this behavior is virtually non-existent. Taken together, the reduced-form results imply that high-skill sector firms tend to avoid training apprentices, while low-skill

sector firms seek apprentices. Hence, most training occurs in low-skill sectors.

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

The structural parameters of the model can be transparently identified, exploiting moments that correspond to the key firm responses from the reduced-form analysis. First, we estimate parameters of the production function and productivity distribution by sector using pre-reform data. We then estimate non-parametric training cost distributions via simulated method of moments, targeting excess mass and missing mass in the post-reform firm size distribution, the number of apprentices by firm size, and the fraction of firms paying the fee. These observed responses are informative of the costs and benefits of training to firms. They allow us to uncover the distribution of firms' *net* training costs, a combination of the cost of training in terms of workers' time and the lower productivity of apprentices.

The estimated net training costs are high for many firms. Scaled in terms of equivalent time spent by workers, average training costs are 0.75 in low-skill sectors and 1.19 in high-skill sectors. This implies that for the average low-skill sector firm, apprentices are only 25% as productive as workers, and for the average high-skill sector firm apprentices even have negative productivity. These results are consistent with existing studies such as [Alfonsi et al. \(2020\)](#) who find that many firms are reluctant to train apprentices even when they are incentivized by wage subsidies. As expected from the reduced-form results, training costs are higher in high-skill sectors, where apprentices have negative marginal productivity for 60% of firms. The strong heterogeneity in training costs has another important implication: at any given size of training incentives, only the firms with the lowest training costs are willing to take apprentices. If training expenses by firms are positively correlated with the amount of skills acquired by apprentices, this may help explain why the literature tends to find small benefits of some types of firm-based training.

We argue that our training cost estimates are robust to allowing for different types of labor market frictions. On the one hand, we can extend the model to allow for frictions preventing apprentices to move between firms after training. In this case, firms receive additional dynamic benefits from training, as they can employ former apprentices at compressed wages. If anything, such frictions would imply a higher level of revealed training costs, and exacerbate heterogeneity since dynamic benefits tend to be higher in high-skill sectors. Similarly, allowing for hiring frictions faced by firms does not affect estimated training costs much.

Next, we use the model to quantify the effects of the regulation under three scenarios: i) in partial equilibrium, ii) in general equilibrium, where wages and prices adjust and any excess labor supply is absorbed, and iii) in a dynamic scenario, where trained apprentices increase the future supply of workers. In partial equilibrium, the apprenticeship regulation has relatively small aggregate effects. Firms re-optimize and substitute around 1.3% of regular workers for apprentices, such that total labor input does not change much and output is relatively stable. Yet, there is some reallocation across sectors: Output increases by 0.3% in low-skill sectors but decreases by 0.3% in high-skill sectors. These partial equilibrium effects are consistent with additional reduced-form evidence we provide on firm outcomes. Results from a difference-in-difference strategy suggest that firms who are induced by the reform to train more apprentices reduce the number of regular workers, but the effects on output and profits are small and insignificant.

The general equilibrium and dynamic scenarios then yield further insights beyond the scope of the reduced-form analysis. In these scenarios, the impact on output can become substantial. When all displaced workers re-enter employment in general equilibrium, production increases by 1.3% in low-skill sectors, but also in high-skill sectors the negative partial-equilibrium effects are reversed. Finally, the positive impact of the regulation can be amplified dynamically. Assuming that trained apprentices eventually become as productive as the average worker in a sector, we find that aggregate output increases by up to 3.7% in the dynamic scenario. Sensitivity analysis further indicates that the magnitude of these dynamic gains depends on the transferability of apprentices' acquired skills within sectors.

Unpacking the aggregate effects, we find that the apprenticeship regulation creates winners and losers. Most directly related to the goal of improving youth employment opportunities, apprentices' welfare increases in all sectors and under all scenarios, as more young individuals are formally trained and employed. On the firm side, the effects vary by sector: most low-skill sector firms see an increase in profits as they hire apprentices contributing to production at relatively low cost. In

high-skill sectors, many firms become less profitable in partial equilibrium. However, these negative effects can be overturned in general equilibrium, where all firms benefit from the increased supply of trained workers. The main losers from the expansion of apprenticeships are incumbent workers, especially in low-skill sectors. In partial equilibrium, some workers are displaced, while in general equilibrium they see a decrease in wages.

Finally, we conduct counterfactual simulations based on the estimated model. We begin by decomposing the effects of the different components of the apprenticeship regulation, each of which plays an important role. Reducing the apprentice minimum wage contributes significantly to expanding training. Mandating a minimum quota ensures that many firms, including in high-skill sectors, train at least some apprentices. On the other hand, the maximum quota restricts apprentice intake so that firms in low-skill sectors do not use too many apprentices as “cheap labor”, merely substituting regular workers. The possibility of paying the fee instead of training serves as a buffer, reducing the negative impact on firms with very high training costs.

We also simulate the effects of counterfactual apprenticeship policies. In particular, we study policies that incentivize firms to train without imposing quotas that would induce firm size distortions. The first alternative policy is a pure subsidy on training costs financed by payroll taxes. This scenario resembles real-world subsidy schemes, and is closely related to experimental studies that provide firms with wage subsidies for apprentices (e.g. [Crépon and Premand, 2019](#), [Alfonsi et al., 2020](#)). We find that a uniform training subsidy performs similarly to the Colombian benchmark regulation on aggregate, but leads to even more concentration of training in low-skill sectors. In fact, training is concentrated only in few firms with the very lowest training costs. As a second counterfactual, we consider sector-specific minimum wages for apprentices. Specifically, the minimum wage is allowed to be lower in high-skill sectors, where firms face higher training costs. We find that this type of policy can limit reallocation of production towards low-skill sectors and encourage more training in high-skill sectors. Overall, sector-specific minimum wages generate the largest welfare gain among the policies considered. We conclude that policies taking into account heterogeneity in training costs by allowing for some sector-specific aspects of apprenticeship regulation can further enhance welfare gains.

This paper contributes to the literature on training programs in developing countries. A number of recent studies provide evidence of the effects of training on apprentices or workers, including [Card et al. \(2011\)](#), [Attanasio et al. \(2011, 2017\)](#), [Hirshleifer et al. \(2014\)](#), [Kugler et al. \(2019\)](#) and [Alfonsi et al. \(2020\)](#). [McKenzie \(2017\)](#) and [Card et al. \(2018\)](#) provide reviews of this literature. Although

firm participation is crucial, only few studies report take-up rates of training by firms, which tend to be low (e.g. Galasso et al., 2004; Alfonsi et al., 2020). More generally, there is limited evidence on firm responses and how to effectively incentivize firms to train. Related work includes Crépon and Premand (2019) and Hardy and McCasland (2020) who study the effect of apprentices on firm outcomes and de Mel et al. (2019) who consider wage subsidies for unskilled workers. We make several contributions to this literature. First, the unique nature of the reform allows us to study firm responses to apprenticeship regulation along various margins, which are informative of their demand for apprentices.² In particular, we provide novel evidence of heterogeneity in responses across firms and sectors. Second, further exploiting this variation, we structurally estimate the underlying training costs of firms, which are key in explaining firm responses. Third, while most of the literature uses randomized experiments, we follow a complementary approach and study a large-scale policy faced by all Colombian firms. This allows us to quantify general equilibrium and dynamic effects of the apprenticeship regulation. The relatively small partial-equilibrium effects we find are consistent with some of the mixed evidence on the benefits of training programs in the literature. However, our general-equilibrium results support the conjecture that a substantial benefit of expanding apprenticeships is an increase in the overall number of trained workers.

Our analysis builds on a classic labor economics literature studying training by firms. This literature emphasizes various sources of inefficiency in the provision of firm-based training (Becker, 1964; Acemoglu and Pischke, 1998, 1999a,b; Dustmann and Schönberg, 2009, 2012). Yet, the returns to training for apprentices found in developed countries tend to be sizeable (Krueger and Pischke, 1995; Fersterer et al., 2008; Göggel and Zwick, 2012). Our main contribution to this literature lies in the estimation of firms’ training costs. Some studies including Acemoglu and Pischke (1998) and Konings and Vanormelingen (2015) use measures of training costs reported by firms in surveys.³ To our knowledge, this paper is the first to structurally estimate training costs based on observed firm responses to apprenticeship regulation. Our “revealed preferences” approach provides credible estimates of total net training costs including for those firms that do not train, while avoiding reporting biases that may be present in survey measures. Moreover, we document systematic heterogeneity in training costs related to skill requirements.

We organize the rest of the paper as follows. Section 2 describes the data and institutional

²Our reduced-form results are closely related to work by Ospino (2016, 2018) who studies the Colombian apprenticeship regulation using survey data and finds that firms substituted away from permanent workers and increased their productivity.

³Other studies reporting survey measures of training costs include Del Boca and Rota (1998), Wolter et al. (2006), Muehlemann et al. (2010), Jones et al. (2012) and Dostie (2013).

context, Section 3 presents the reduced-form evidence, Section 4 outlines the theoretical framework, Section 5 presents the quantitative results, Section 6 discusses the counterfactual simulations, and finally Section 7 concludes.

2 Institutional Context and Data

2.1 Institutional Context and Apprenticeship Regulation

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

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

The 2003 reform establishes a dual vocational training system with two phases, the teaching phase in a formal education institution, and the productive phase where they receive training in the firm.⁵ Training courses span a wide variety of occupations in all sectors (see Appendix Tables B1 and B2). Prior education requirements vary by field of training, with basic secondary education (9 years) being the most common. On average, vocational training lasts between one and two years.⁶ SENAs sets the vocational education curriculum and provides the classroom portion of training directly for more than 80% of apprentices (SENA, 2016). During the productive phase, firms are free to structure training according to their needs within SENAs guidelines. Upon completion of the apprenticeship, trained apprentices receive a certificate from SENAs or another accredited

⁴Own calculations based on data from the Colombian Household Survey (ECH).

⁵SENA asks firms to hire apprentices in both phases. However, less than one third of apprentices are already hired by firms in the teaching phase in practice (own calculations based on PILA social security data).

⁶The average duration of vocational training is 1760 hours in “technical” courses and 3640 hours in more advanced “technological” courses (SENA, 2016), corresponding to around one to two years. There is a legal minimum training duration of 768 hours in the teaching phase and of 672 hours in the productive phase.

institution where they received their classroom instruction, as well as a certificate from the training firm.⁷

The post-reform apprenticeship regulation has three main components:

1. **Apprentice Quotas:** First, apprentice quotas depending on the number of full-time workers in a firm are established. There is a minimum number of required apprentices: firms with 15 to 29 full-time workers must have at least one apprentice, increasing by one more apprentice in intervals of 20 workers. Thus, if the firm has 30 to 49 workers it must train at least two apprentices, between 50 and 69 it must train at least three, etc. In addition, the quota sets a maximum number of apprentices. The maximum is one apprentice for firms with less than 15 workers, and twice the minimum for firms with more than 15 workers.⁸
2. **Apprentices' Minimum Wage:** The reform also lowers the minimum wage of apprentices as an additional incentive for firms to train. While firms had to pay apprentices the full minimum wage during the productive phase before the reform, this is lowered to only 75% after the reform.⁹
3. **Fee:** Finally, firms can pay a fee instead of hiring the minimum required apprentices. This fee is proportional to the difference between the minimum quota and the number actually hired by the firm. The fee amount is 100% of the (regular) minimum wage per missing apprentice.

The reform also provided SENA with tools for stricter enforcement of the regulation. Whenever firms are found non-compliant with the above rules, a fine equivalent to two times the minimum wage per missing apprentice can be imposed. Firms have to report the total number of hours of regular workers every 6 months. Using this information, the number of equivalent full-time workers is calculated.¹⁰ SENA checks the reported information by comparing it to independent data from payroll tax records. Next, SENA determines apprentice quotas, and firms have two months to

⁷Apprentices are assessed jointly by SENA and the training firm based on tasks performed, skills and competences (SENA, 2016). Until now there is no automatic industry-wide recognition of training certificates (Fazio et al., 2016).

⁸The labor market reform of 2003 included some other elements aimed at increasing labor market flexibility, none of which were based on the number of workers or related to the firm size thresholds of the apprenticeship regulation. More generally, we are not aware of any other policy discontinuities around the same thresholds. The only exception is the threshold of 50 workers, above which some specific rules concerning mass layoffs and workers' resting times change. These rules are unlikely to affect our results, since the patterns we document are very similar around all regulation thresholds.

⁹During the teaching phase, apprentices have to be paid only 50% of the minimum wage (before and after the reform). Apprentices who receive education from a university, as opposed to vocational education institutions, are paid a full minimum wage in both phases. The regulation also specifies that all apprentices have to be paid a full minimum wage if the unemployment rate falls below 10%, which did not occur throughout our main sample period.

¹⁰Only regular employees of the firm are counted. This excludes indirectly-hired workers, such as temporary or out-sourced workers.

comply. To hire apprentices, firms can post vacancies with SENA or hire independently.¹¹ Besides directly managing the teaching phase, SENA also monitors apprentices' training in firms. Apprentices report their satisfaction with the training firm twice a month, and they can be reallocated if they are dissatisfied with the quality of training or if tasks assigned to them do not correspond to their field of training.

2.2 Data and Summary Statistics

We use a novel administrative data set provided by SENA, covering the universe of manufacturing firms with at least 10 workers between 1995 and 2009. For each firm-year observation, the data includes the number of workers, the number of apprentices, and indicators for fees and fines paid by 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).¹² The survey data includes additional information on workers divided into three occupational/skill layers (professional, production and administrative workers), as well as on output, sales, wages, inputs and costs. Table 1 shows summary statistics of the data. In the full sample described in Column (1), there are 108,385 firm-year observations, and 14,586 unique firms.

Additional Information on Apprentices

Our main data is at the firm level and does not contain individual-level characteristics of workers or apprentices, except some limited information on gender and broad occupational layers. Appendix B shows descriptive statistics on apprentices based on a combination of data sources. The first additional dataset is a small-scale survey on school-to-work transitions of young individuals (ETET). Second, we use individual-level administrative data from social insurance records (PILA), which has a larger sample but fewer observed characteristics. Unfortunately these data sets do not coincide with our period of analysis but are only available in 2013 to 2015 and 2015 to 2016, respectively.

Appendix Table B3 summarizes apprentices' characteristics. Apprentices tend to be in their early 20s and more than 50% are female. The majority have high-school or secondary (including vocational) education. Apprentices wages' are significantly lower than those of all other workers

¹¹The process of hiring apprentices did not change significantly with the reform itself. Only in 2008, SENA introduced an electronic matching system to facilitate apprentice hiring.

¹²SENA collects data for all firms. The data we use focuses on manufacturing firms with at least 10 workers, as only these firms are available in the census data. The census data is collected at the establishment level and then aggregated to the firm level.

as well as those of other young workers in the same age group. Most apprentices self-report their socioeconomic status as low or medium. It is also worth noting that more than 90% are satisfied with their current apprenticeship/job, and less than 30% want to move jobs, which reflects higher satisfaction than other young workers. Apprentices are trained in virtually all economic sectors, with around 25% of apprenticeships in manufacturing. Within manufacturing, 22% of apprentices are trained in highly-skilled professional occupations, 40% are trained as production workers, and 38% as administrative workers.

Appendix Figure B1 shows wages of apprentices (relative to the minimum wage) around the month they complete the apprenticeship. Immediately upon graduation, former apprentices experience a jump in wages of around 40%, illustrating the returns to apprenticeship training. The figure also shows a second series for apprentices who do not graduate in the sample period, whose wage stays around the minimum wage.¹³ Finally, Appendix Table B4 shows transition probabilities of graduating apprentices. The scope of this analysis is limited by the fact that only around one quarter of individuals are still observed in PILA after graduating.¹⁴ Conditional on remaining observed in the data, around three quarters of apprentices stay in the same firm after completing the apprenticeship, implying an overall probability of staying in the firm of around 18%.

3 Reduced-Form Results

3.1 Aggregate Effects: Number of Apprentices and Firm Size Distribution

The primary objective of the regulation is to expand apprenticeship training. Panel (a) of Figure 1 shows that the policy is successful in this regard, dramatically increasing the number of individuals hired as apprentices in the formal labor market. 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 workers 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.

However, Panel (b) of Figure 1 shows that the regulation also induces sizeable changes in the

¹³Two aspects are worth highlighting in relation to Appendix Figure B1. First, apprentices had to be paid the full minimum wage in 2015 and 2016, the PILA sample period. This difference to our main sample period arises because unemployment was below 10%, triggering an increase in the apprentice minimum wage (see footnote 10). Importantly, this implies that the jump in wages in the figure is not mechanical, but driven by firms paying above the minimum wage. Second, the return to apprenticeships is due to a combination of vocational education by SENA and training in the firm, which we cannot disentangle with the available data.

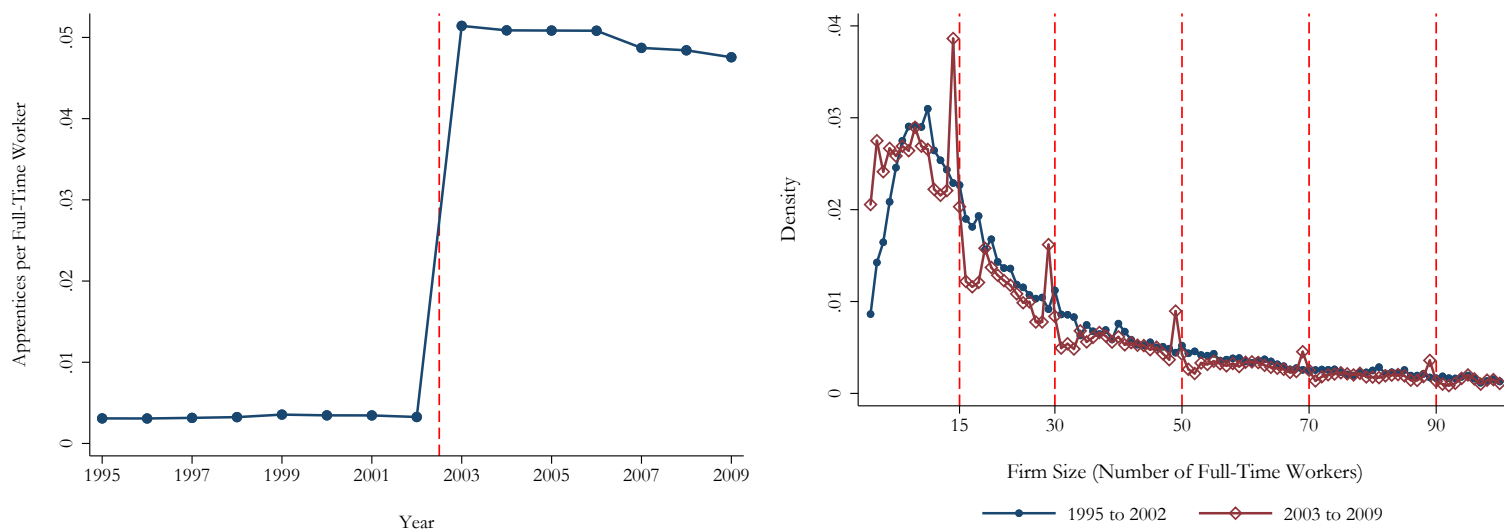
¹⁴Apprentices can exit the PILA data for three reasons. First, they may move to other manufacturing firms with less than 10 workers, which are not covered by the data. Second, they may work in a sector other than manufacturing. Third, they may become unemployed or take up informal work. According to SENA (2016), employment rates of graduated apprentices are between 51% and 70%. This suggests that the majority of unobserved former apprentices do not likely become unemployed.

Table 1: Summary Statistics

	(1)	(2)	(3)	(4)
	All Sectors	<i>Full Sample</i> Low-Skill	High-Skill	<i>Threshold Sample</i>
Apprentices	1.08 (4.12)	1.72 (5.58)	0.50 (1.93)	0.74 (1.80)
Workers	49.08 (94.72)	49.54 (101.26)	48.68 (88.49)	29.82 (30.56)
Workers (Survey)	54.58 (127.33)	56.22 (139.16)	53.12 (115.77)	31.23 (34.52)
Fraction Professionals	0.08 (0.14)	0.06 (0.12)	0.10 (0.15)	0.07 (0.12)
Fraction Production Workers	0.55 (0.27)	0.56 (0.28)	0.54 (0.27)	0.60 (0.24)
Fraction Admin Workers	0.37 (0.26)	0.38 (0.27)	0.36 (0.25)	0.33 (0.21)
Output	11,031,021 (28,153,949)	10,679,502 (27,074,391)	11,343,706 (29,077,247)	5,579,873 (12,491,212)
Value Added	4,755,892 (12,739,041)	4,412,605 (12,081,763)	5,057,618 (13,282,718)	2,243,634 (5,575,936)
Profits	3,113,409 (9,359,584)	2,893,555 (8,868,686)	3,306,645 (9,766,703)	1,456,771 (4,365,586)
Wage bill (permanent workers)	1,240,527 (3,048,647)	1,115,178 (2,839,862)	1,352,028 (3,218,945)	590,291 (1,078,797)
Total wage bill	1,609,439 (3,702,964)	1,463,753 (3,430,131)	1,739,030 (3,925,258)	773,080 (1,354,139)
Wage per worker (permanent workers)	19,733 (17,763)	18,384 (17,649)	20,885 (17,780)	17,104 (13,059)
Capital/Output	0.66 (0.88)	0.64 (0.91)	0.67 (0.85)	0.69 (0.91)
Intermediates/Output	0.54 (0.18)	0.56 (0.19)	0.53 (0.18)	0.55 (0.18)
Observations	108,385	51,024	57,361	14,848
Firms	14,586	7,403	7,986	2,018

Notes: Columns (1) to (3) of the table show summary statistics for the full sample in the years 1995 to 2009. Column (4) provides summary statistics for the threshold sample used for the difference-in-difference analysis in Section 5.5, which is restricted to firms around the first fifteen regulation thresholds in the years 1999 to 2006. All monetary variables in 2009 thousands of pesos. Standard deviations in parentheses.

Figure 1: Number of Apprentices and Firm Size Distribution



(a) Number of Apprentices over Time

(b) Firm Size Distribution Before and After the Reform

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

firm size distribution. In the pre-reform years 1995 to 2002, the distribution of the number of full-time workers is relatively smooth. In contrast, the post-reform distribution in the years 2003 to 2009 exhibits pronounced spikes around the regulation thresholds marked by the dashed vertical lines, and holes or “missing mass” on both sides of the thresholds.¹⁵ The figure provides first visual evidence of firm size distortions and changes in firms’ labor inputs. Moreover, the presence of missing mass on both sides of the thresholds gives a first indication of heterogeneous responses, as some firms seem to avoid being just below while others avoid being just above the thresholds.

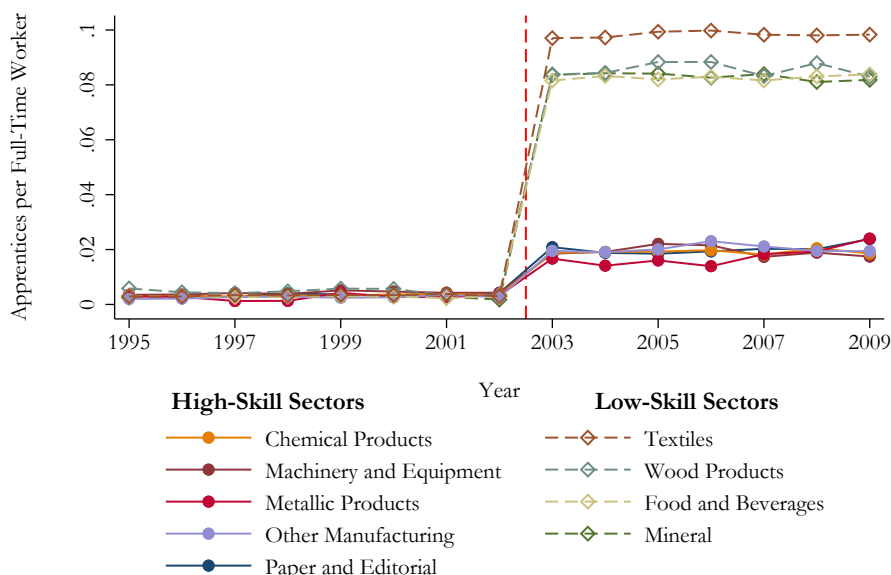
To study heterogeneity in firm responses, we divide firms into nine two-digit sectors using the Colombian industrial classification.¹⁶ Figure 2 plots the number of apprentices per full-time worker in each industry. The figure suggests clear heterogeneity in apprentice intake. In the wood products, textiles, food and beverage, and mineral non-metallic products industries, there are between eight and ten apprentices per 100 full-time workers (after the reform). In contrast, in the paper and editorial, machinery and equipment, metallic products, chemical products and other manufacturing industries, there are only around two apprentices per 100 full-time workers.

To make sense of these differences in apprentice intake, we rank industries by the fraction of professionals out of all workers, that is the fraction of workers in highly-skilled occupations that

¹⁵Appendix Figure A1 shows distributions year by year, exhibiting the same patterns as the pooled distributions.

¹⁶We use the industry classification CIU 3 A.C from DANE, which is adapted from the International Standard Industrial Classification (ISIC).

Figure 2: Number of Apprentices By Sector



Notes: The figure shows the number of apprentices per full-time worker by two-digit industry. The legend shows industries classified by our baseline skills proxy, the fraction of highly-skilled (professional) workers. “High-skill sectors” refers to industries ranked above the median share of professionals, and “Low-skill sectors” refers to industries below the median.

require tertiary education. This can be interpreted as a proxy for skill requirements in each industry, reflecting the difficulty or costs of training. Panel A of Appendix Table A1 shows the fraction of professionals by industry in the pre-reform years, which varies between 3.5% (wood products) to 8.6% (chemical products), and the implied sectoral ranking. We denote industries below the median as *low-skill sectors* and those above the median as *high-skill sectors*. This classification has remarkable power in explaining differences in the number of apprentices in the post-reform period. The four sectors classified as low-skill are precisely those with the most apprentices in Figure 2, whereas the five sectors classified as high-skill take significantly fewer apprentices. We revisit the issue of classifying sectors in Section 3.3, where we argue that our results are robust to using alternative skills proxies.

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 professional workers (10% vs. 6% in low-skill sectors), and they have fewer unskilled production workers and fewer administrative workers.¹⁷ Although the average high-skill sector firm has fewer workers than the average low-skill sector firm, output, value added and profits are higher. High-skill sector firms

¹⁷Our main data only contains information on these three broad occupational/skill layers. Appendix Table B5 shows the shares of more fine-grained occupational groups in high-skill vs. low-skill sectors obtained from the Colombian Household Survey (ECH).

pay higher wages per worker, and they use more capital and less intermediate inputs than low-skill sector firms. In addition, Appendix Table A2 shows summary statistics for the pre-reform years only, displaying similar patterns in these characteristics.

3.2 Heterogeneous Firm Responses Across Sectors

Next, we provide evidence of heterogeneous firm responses to the apprenticeship regulation. 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, and the share of firms paying fees.

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

Figure 3 presents the firm size distribution around the first five thresholds for firms in high-skill sectors (Panel a) and in low-skill sectors (Panel b), pooling the post-reform years 2003 to 2009. There are pronounced spikes in both panels, but the figure reveals a crucial difference in the location of bunching. Firms in high-skill sectors bunch *below* the thresholds, while firms in low-skill sectors bunch exactly *at* the thresholds. Moreover, in high-skill sectors there is missing mass above the thresholds, while in low-skill sectors the missing mass is below the thresholds. Taken together, this indicates that some firms in high-skill sectors reduce the number of regular workers to avoid the higher minimum apprentice quota that would apply above the thresholds, such that they have to train fewer apprentices. Low-skill sector firms, on the other hand, increase the number of 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.¹⁸ 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 Panel (b) of Figure 1. Similarly, we compute the missing mass as the hole in the observed distribution relative to the counterfactual,

¹⁸In line with the sharp bunching responses, the bunching region is chosen as one bin below the threshold in Panel (a) and the bin at the threshold in Panel (b). The missing mass region is the threshold plus three bins to the right in Panel (a), and the three bins below the threshold in Panel (b), which fits the observed density hole well. We do not explicitly restrict the missing mass to equal the bunching mass, but the two tend to line up well for our estimation region choices.

$m = M/h_o(\hat{n})$, where M is the firm count in the missing mass region. For both excess mass and missing mass estimates, bootstrapped standard errors are shown in parentheses.

The estimates in Figure 3 show sizeable and significant excess mass and missing mass in both panels. The fact that excess mass and missing mass are similar at each threshold confirms that bunching responses indeed originate from the neighborhood of the threshold. Comparing across sectors, bunching responses are somewhat larger in high-skill sectors, where the excess mass is between 1.7 and 2.2 across the five thresholds. This can be interpreted in terms of a firm size response: The marginal bunching firm is estimated to reduce the number of workers by around two. In low-skill sectors, on the other hand, the implied firm size response goes in the opposite direction, where the marginal bunching firm increases their firm size by between 0.6 and 1.8 workers. Finally, the patterns in the figure indicate little frictions that would prevent firms from adjusting their size. In both panels, 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.

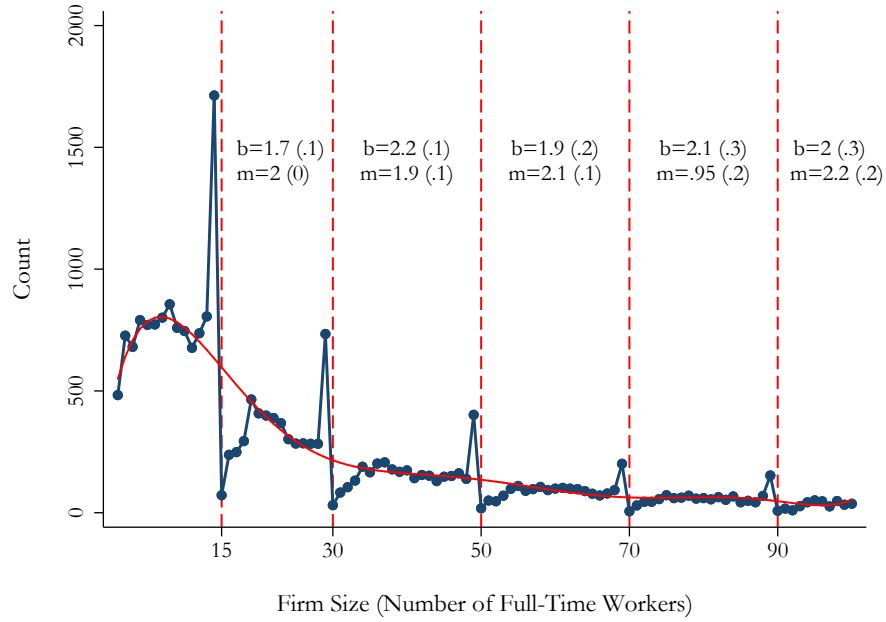
Fact 2: Firms in high-skill sectors tend to choose the minimum number of apprentices; firms in low-skill sectors tend to choose the maximum.

Panel (a) of Figure 4 shows the number of apprentices by firm size in high-skill and low-skill sectors, as well as the minimum and maximum apprentice quotas from the regulation. The figure suggests that firms in high-skill sectors generally try to avoid training apprentices while firms in low-skill sectors tend to train as many as possible. Firms in high-skill sectors have an average number of apprentices below the minimum quota of the regulation. Thus, some of these firms must be paying the fee or not comply with the regulation. In contrast, firms in low-skill sectors have an average number of apprentices close to the maximum quota, indicating that some of them may be constrained by this upper limit.¹⁹

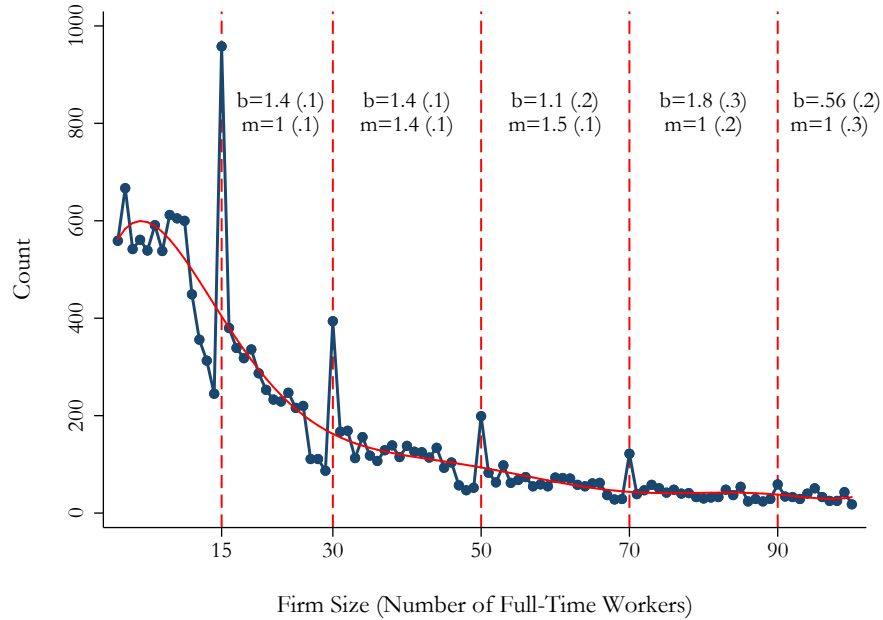
In line with the discontinuously increasing apprentice quotas, there are jumps in the average number of apprentices at the regulation thresholds. These jumps are particularly sharp in low-skill sectors. Moreover, within each bracket of the regulation, the average number of apprentices follows a decreasing pattern, which is again more marked in low-skill sectors. The number of apprentices is high just above the thresholds, relatively constant for a few bins to the right, and decreases in the highest bins of each bracket. Hence, the overall picture reflects a mixture of the causal effect of the policy and selection. Firms in brackets with higher apprentice quotas have to take more apprentices

¹⁹Panel (a) of Appendix Figure A2 shows that the pre-reform number of apprentices by firm size is low and similar across all sectors.

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



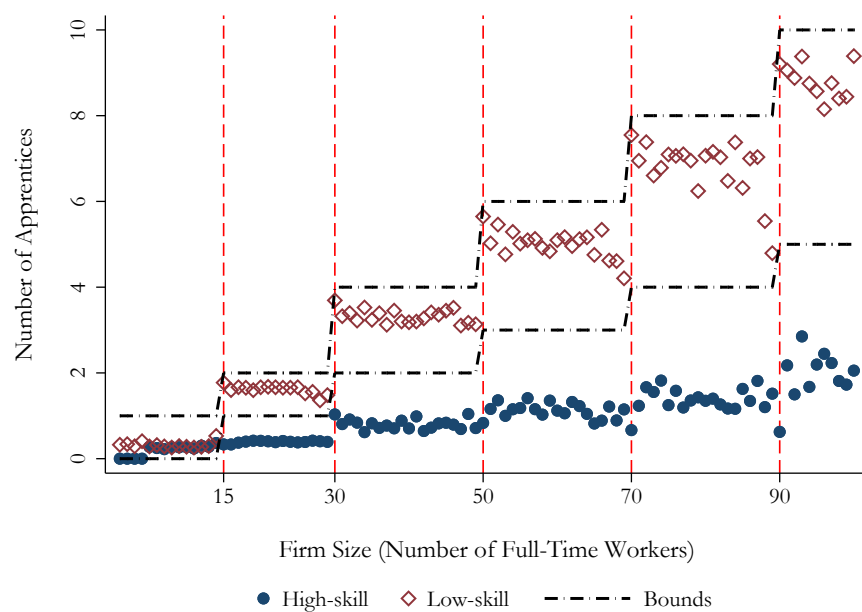
(a) High-skill Sectors



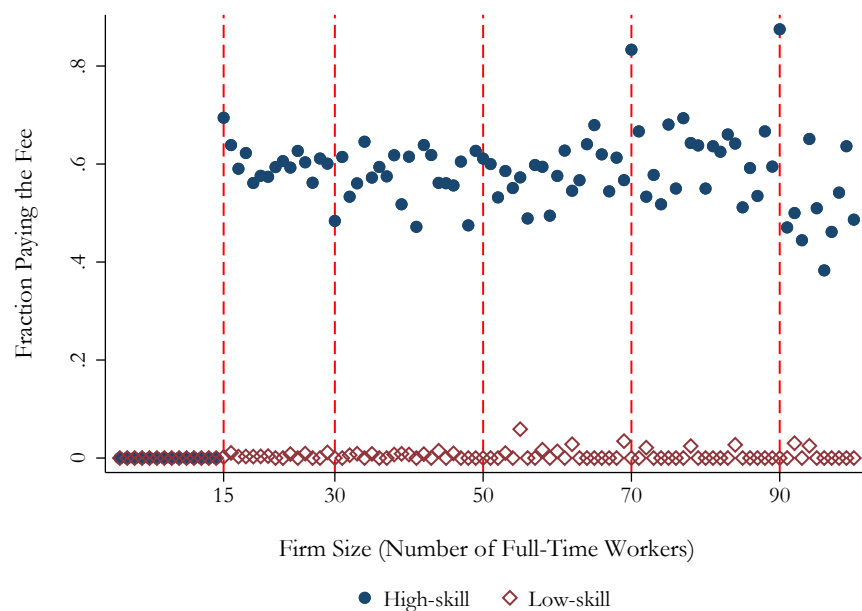
(b) Low-skill Sectors

Notes: The figure shows the 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 are reported at each threshold, with bootstrapped standard errors in parentheses.

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



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



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

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

as a result of the regulation, but locally firms can select into brackets via bunching responses. The firms bunching at or just above the thresholds are those who wish to hire many apprentices, while firms locating just below the thresholds are those who wish to hire few apprentices.

Table 2 summarizes these responses in more detail. The table shows the proportion of firms choosing the maximum number of apprentices, the minimum number, a number between the maximum and the minimum, below or above these bounds, and those that do not train any apprentices. The most common responses to the regulation are choosing exactly the minimum number, exactly the maximum number, or no apprentices at all. Together these responses account for close to 95% of observations across all sectors. However, responses differ strongly across high-skill and low-skill sectors. While more than 60% of firms in high-skill sectors train no apprentices at all, almost all firms in low-skill sectors (99.7%) train apprentices. Whenever high-skill sector firms have apprentices, they tend to choose exactly the minimum number required. On the other hand, almost two thirds of firms in low-skill sectors choose the maximum number of apprentices, and the remainder chooses a number at or somewhat above the minimum.

Table 2: Summary of Training Responses

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

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

These results have two additional important implications. First, most apprenticeship training happens in low-skill sectors. In the post-reform years, 77% of apprentices are trained in low-skill sector firms, although there are a similar number of firms in low-skill and high-skill sectors. The fact that low-skill sector firms increase their size but high-skill sector firms decrease their size further exacerbates the increasing employment share in low-skill sectors. Second, the results imply an important role of the apprentice minimum wage. In particular, the observation that many firms hire more than the minimum number of apprentices required suggests that it is now worthwhile to train extra apprentices. This cannot be explained by apprentice quotas alone, but it is consistent

with an additional effect of the lower minimum wage.

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

Panel (b) of Figure 4 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. Note that the nominal cost of hiring an apprentice is smaller (75% of the minimum wage) than the fee (100% of the minimum wage). Thus, the responses of high-skill sector firms indicate that their training costs must be high. In fact, apprentices must create an overall negative marginal benefit for many firms as they are unwilling to hire apprentices even though it is nominally cheaper than not hiring apprentices. Later on, this fact informs the specification of production functions in the theoretical framework.

Moreover, the fraction of high-skill sector firms paying the fee is relatively stable across firm sizes. This suggests that training apprentices entails a proportional cost to firms rather than a fixed cost, as in the case of fixed training costs one would expect that large firms subject to higher apprentice quotas should be less likely to pay the proportional fee. This informs the specification of training costs in the model. In contrast, low-skill sector firms rarely pay the fee. In addition, we show that very few firms are fined for non-compliance in Appendix Figure A2. The vast majority of firms comply with the regulation by either training the required number of apprentices or paying fees, which is consistent with the strict enforcement described in Section 2.

3.3 Further Discussion

Alternative Sector Classifications. Our baseline sector classification uses the fraction of highly-skilled professional workers as a proxy for skill levels. Appendix Table A1 shows alternative measures that may be naturally related to skill requirements, including additional variables from the main data in Panel B and some measures obtained from an independent data source, the Colombian household survey (ECH), in Panel C.²⁰ Overall, the alternative skill proxies line up well with the baseline sector classification. High-skill sectors are characterized by higher wages and higher education levels, both among all workers and among administrative and production workers, the positions in which the majority of apprentices work. Sector rankings implied by the alternative measures tend to be highly positively correlated with the baseline ranking, with rank correlation coefficients of around 70%. While most individual sectors are consistently classified as high-skill or low-skill based on all or most of the measures, the assignment of the mineral non-

²⁰See Appendix B for a description of the ECH data.

metallic and other manufacturing sectors appears somewhat less robust and depends on the skills proxy used. Appendix C replicates the reduced-form analysis excluding these two sectors. The results remain very similar, suggesting that the sectors with more ambiguous skill levels are not a major driving force behind our findings.

Other Correlates of Firms Responses. Firm responses vary strongly across sectors classified by our skills proxy based on the fraction of professionals. However, firms in different sectors may also differ in terms of other characteristics, which may confound this relationship. Appendix Table A3 shows regressions of indicators for the different firm responses on the skills proxy as well as other firm characteristics including firm size, wages, output, profits and inputs. In order to avoid endogenous changes in these characteristics, post-reform responses are regressed on their pre-reform firm-level averages. We find that the correlation of responses with the skills proxy is highly significant, even after including these controls. In line with the results from Table 2, choosing the maximum apprentice quota is negatively correlated with the skills proxy, whereas for choosing the minimum quota or paying the fee the correlation is positive. In addition, Appendix Table A4 shows the correlation of bunching behavior with the different types of responses to the regulation. The vast majority of firms bunching below the regulation thresholds are in high-skill sectors, and as expected, bunching below tends to coincide with avoiding apprentices. 49% of bunchers below choose the minimum number of apprentices and 27% pay the fee. Bunching above the thresholds, on the other hand, is concentrated in low-skill sectors and tends to coincide with choosing the maximum number of apprentices, which 72% of bunchers above do.

Bunching and Reporting Responses. Our data provides a precise measure of the number of formal workers, but some of the observed patterns might reflect reporting rather than real responses, an issue often encountered with administrative data (see e.g. Slemrod (2016)). In particular, we cannot fully exclude that some of the bunching around the thresholds is driven by reporting, i.e. changes in workers' formality. Yet, some of the observed responses are arguably hard to reconcile with pure reporting effects. For instance, many firms choose to pay costly fees, including firms close to the thresholds. This suggests that they may not be able to easily change their reported size in order to avoid training apprentices. Moreover, while reporting responses may readily explain reductions in firm size, they may be a less plausible explanation for bunching above the thresholds, as this would require all bunching firms to already have employed informal workers whom they can begin reporting formally. We return to this discussion in Section 5.5, where we

study the effect of the regulation on further firm outcomes.

Comparison to Other Settings. A few existing studies report take-up rates of training programs by firms, providing valuable points of comparison. Experimental evidence suggests that only a small fraction of firms are willing to hire apprentices at subsidized wages, ranging between 0.5% (Galasso et al., 2004) and 24% (Alfonsi et al., 2020). Even in settings where apprenticeships are common, the share of training firms tends to be limited. For instance, 27% of firms train apprentices in Germany (Dustmann and Schönberg, 2009) and 29% in Switzerland (Wolter et al., 2006). In our setting, the overall post-reform share of training firms is 41.6%, and Table 2 shows that only 35.1% of firms with at least 15 workers do not train. A number of factors could explain this relatively high share. First, the Colombian regulation is quite stringent, featuring mandatory apprentice quotas and penalties for non-training firms, as well as a low minimum wage. This combination appears to be effective at incentivizing many firms to train. Second, our data does not include small firms with less than 10 workers, which may be less likely to train. Third, some evidence in the literature suggests that manufacturing firms, on which we focus, are more likely to train.²¹

4 Model: Heterogeneous Training Costs

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

4.1 Model Setup and Equilibrium without Regulation

Consider an infinite-period economy composed of $K \geq 1$ sectors with a fixed number of heterogeneous firms in each sector.²² Firms in sector k are characterized by training costs t_a^k and managerial ability z^k , which follow sector-specific distributions \mathcal{T}^k and \mathcal{Z}^k . These firm characteristics are invariant across time. Managerial ability z^k is an idiosyncratic characteristic that can also be interpreted as technological differences or other factors that affect a firm's productivity. Firms

²¹For instance, Mohrenweiser and Zwick (2009), Konings and Vanormelingen (2015) and de Mel et al. (2019) report heterogeneity in training rates across broad sectors and find that manufacturing firms are more likely to train than those in non-manufacturing sectors.

²²To keep the model tractable, we do not allow for firm exit or entry.

produce y_t^k units of good k in period t using labor, which is supplied either by workers l_t^k or by apprentices $l_{a,t}^k$.²³ We suppose that workers are sector-specific, but apprentices can be trained in any sector. All individuals are endowed with a unit of time which they supply inelastically.

Apprentices have to be trained using workers' time in order to be able to contribute to production. Let $t_a^k \geq 0$ denote training costs in units of time workers have to spend training the apprentice. Across firms and sectors, apprentices potentially require different amounts of training time. A sector whose technology requires simple menial tasks probably requires less time to train apprentices compared to a sector with highly specialized tasks. In addition, the opportunity cost of workers who train apprentices can vary. Production combines a firm's managerial ability z^k with the total labor supplied by both types of workers given training costs t_a^k . If a firm hires n_t^k workers and trains $n_{a,t}^k$ apprentices, labor supplied by workers is $l_t^k := n_t^k - t_a^k n_{a,t}^k$ and labor supplied by apprentices is $l_{a,t}^k := \zeta_a^k n_{a,t}^k$. $\zeta_a^k \in [0, 1]$ is apprentices' productivity per unit of time they contribute to production (relative to workers), which may differ across sectors. Finally, firms in different sectors also differ in their production technology $f^k(l_t^k, l_{a,t}^k; z^k)$. We assume that the production function is increasing in labor inputs $(l_t^k, l_{a,t}^k)$ and in managerial ability.

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

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

The constraints $t_a^k n_{a,t}^k \leq n_t^k$ ensure that the firm must hire enough workers to train the chosen number of apprentices in every period. Note that firm choices are static and symmetric across sectors. We therefore drop time and sector subscripts and superscripts to simplify notation whenever there are no ambiguities.

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

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

The FOC for the number of apprentices intuitively shows how firms compare the marginal product of an additional apprentice to their marginal cost. The marginal product is scaled by ζ_a , reflecting that apprentices can be less productive than workers. The marginal cost of an apprentice is not only

²³The model could be readily extended to include capital or other production inputs. Here we suppose that labor is the only input to simplify the analysis and emphasize the role of apprentices. In Appendix D.2 we extend the model to allow for other endogenous inputs, showing that the results remain qualitatively the same.

their wage w_a , but also t_a units of time of a worker that could otherwise be used for production, valued at workers' wage rate w .

Our empirical findings discipline the family of production functions consistent with the observed firm responses. In particular, given that the policy imposes a fee that is nominally larger than the apprentice minimum wage and some firms choose to pay this fee, we show that the production function has to allow for apprentices to have negative marginal productivity. The production function in (1), $f(n - t_a n_a, \zeta_a n_a; z)$, implies that the marginal product of apprentices is $\frac{\partial f}{\partial n_a} = -f_l t_a + \zeta_a f_{l_a}$, which clearly can be negative when training costs are high and/or workers' marginal productivity f_l is high.²⁴ In Appendix D.1 we formalize this argument and show the existence and uniqueness of the solution under standard regularity conditions on the production function (see Assumption 1). In addition, the appendix shows that demand for apprentices decreases in their relative wages and in training costs.

Labor Markets, Preferences and Equilibrium

We consider a simple supply side, where the number of apprentices trained increases the number of workers in a 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 , respectively. Workers can perform the tasks of apprentices but not vice versa. This implies that in equilibrium, apprentices' wages are smaller or equal to those of workers. In addition, the minimum wage could be binding. Both constraints together imply $w_t^k \geq w_{a,t}^k \geq w_{min} \geq 0, \forall t, k$.

We suppose that the labor market clears separately by sector. The wage constraints imply that some workers could remain unemployed and potential apprentices untrained. Let N_t^k and $N_{a,t}^k$ denote aggregate demand for workers and apprentices, respectively. The market clearing conditions are $N_t^k + U_t^k = L_t^k$, $\sum_k N_{a,t}^k + U_{a,t} = L_{a,t}$, where $U_t, U_{a,t} \geq 0$ denote unemployed workers and untrained apprentices. Having separate labor markets by sector allows us to account for wage differences across sectors as observed in the data. Moreover, the assumption is supported by the results in Appendix Table B4, where the majority of apprentices observed after training remain in the same sector. In addition, we consider the case with multiple types of workers, where markets clear by occupation instead of sector in Appendix D.3, and show that theoretical results remain similar to the baseline model.

Trained apprentices increase the future supply of workers in their sector of training. This

²⁴This 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](#)).

component of the model captures potentially important dynamic benefits of the policy. We write the supply of workers in sector k one period ahead ($t + 1$) as

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

where χ_a^k denotes the effective units of labor a trained apprentice contributes to next-period labor supply. This parameter reflects how useful or transferable the skills acquired by apprentices are within a sector. Appendix Table B4 supports the notion of skill transferability, since the average wages of apprentices switching to another firm in the same sector are similar or even higher compared to those staying in the firm. When we estimate dynamic effects of the regulation in Section 5.4, we present sensitivity analysis with respect to different values of χ_a^k .

To close the model, we adopt a simple preference structure. Individuals have a common utility function $u(\cdot)$ over the goods produced by all sectors. They choose a consumption bundle $c_t = (c_t^1, \dots, c_t^K)$, taking prices p_t^k as given. Hence, changes in individuals' welfare come solely from their incomes that depend on whether they are trained or untrained apprentices, workers or firm owners. Apprentices earn wage $w_{a,t}^k$, workers earn wage w_t^k , firm owners earn profits $\pi(z^k, t_a^k)$, and untrained apprentices earn an outside-option income $\underline{w} \leq w_{min}$.²⁵ Individuals are infinitely lived and maximize discounted lifetime utility. A competitive equilibrium is defined as the set of wages and prices in each sector and period such that firms optimally choose the number of apprentices and workers, all individuals choose their optimal consumption bundles and labor and goods markets clear.²⁶

Inefficiency in Training Provision. Whether initial training levels are efficient is crucial for the welfare effects of the apprenticeship regulation. The very small training rates before the reform suggest that equilibrium training provision was likely inefficiently low in the absence of the policy intervention. The initial number of apprentices of around 0.3 per 100 full-time workers is much smaller than the fraction of individuals receiving training in countries with well-functioning apprenticeship systems. For instance, [Dustmann and Schönberg \(2009\)](#) report that the fraction of apprentices in German firms is around 4.9%, which is closer to the observed post-reform training levels in our setting (see Figure 1).

In the model, inefficiency in training provision arises due to a combination two factors. First, skills are transferable across firms within a sector and there is free labor mobility of apprentices

²⁵The outside option can be interpreted as the wage they would receive in the informal sector or an unemployment benefit.

²⁶See Appendix D.1 for the details and the formal definition of a competitive equilibrium.

after training. Thus, firms do not internalize some future benefits of training. Empirical support for this assumption is lent by the relatively low overall retention rate of trained apprentices of around 18% (see Appendix Table B4). Second, apprentices are unable to pay for their training, i.e. to fully compensate firms for training costs, due to the minimum wage floor. Given the estimated training costs we show in Section 5.3, apprentices would even have to accept negative wages in some firms where apprentices' marginal productivity is negative. As discussed in Section 3.2, the fact that many firms train more than the minimum quota indicates that the initially high minimum wage of apprentices is an important factor behind low training levels. Setting the apprentice minimum wage below the minimum wage for regular workers expands the scope of training of individuals with low initial productivity, as it enables apprentices to effectively pay part of the training costs.²⁷

Search and Screening Costs. In addition to the cost of training itself, firms may incur search or screening costs of finding suitable apprentices. While such costs almost surely exist to some extent, they are unlikely to account for the majority of net training costs in our view. Surveys from similar settings suggest that search costs are not the main factor preventing firms from taking in apprentices or other young workers (see McKenzie, 2017). Similarly, there is excess supply of potential apprentices in Colombia, as training courses are strongly oversubscribed (SENA, 2016). A related concern may be that learning about young workers' types is costly to firms, and such screening costs may differ across sectors. However, we do not find any evidence suggesting that the importance of screening varies across sectors. For instance, in Appendix Table B4, the probability of firms retaining apprentices after training is similar in high- and low-skill sectors. In addition, Appendix Figure B2 shows that the variance of former apprentices' wages does not differ much across sectors.

4.2 Equilibrium with Apprenticeship Regulation

Next, we incorporate the apprenticeship regulation into the theoretical framework. First, the regulation imposes apprentice quotas based on the number of workers. Let $(N_j)_{j=0}^{\infty}$ be a weakly increasing sequence of thresholds, where $N_0 = 0$. For a number of workers $n \in [N_{j-1}, N_j)$, the number of required apprentices is $n_a \in [\underline{n}_a^j, \bar{n}_a^j]$, with these minimum and maximum apprentice

²⁷Our modeling assumptions relate to a rich theoretical literature on sources of inefficiency in training provision, including the seminal work of Becker (1964) who emphasizes the role of general vs. firm-specific training and Acemoglu and Pischke (1998, 1999b) who focus on imperfect information and imperfect competition. Furthermore, Dustmann and Schönberg (2012) study the role of commitment and Garicano and Rayo (2017) and Fudenberg and Rayo (2019) argue that knowledge transfer to apprentices can be inefficiently slow. Note that incorporating additional sources of inefficiency in training provision into our model would further exacerbate the positive effects of policies aimed at increasing training.

quotas weakly increasing in j . Second, the regulation reduces the minimum wage of apprentices to w_{min}^a , below the minimum wage for workers w_{min} . Third, instead of training the required apprentices, firms can pay a fee $\mathcal{F}_a(n, n_a) = \phi_a(\underline{n}_a^j - n_a)^+$, which is proportional to the difference between the minimum quota and the apprentices hired.

A firm (z, t_a) facing this regulation solves

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

where $d_f \in \{0, 1\}$ is an indicator for paying the fee, and $\phi_a > 0$ is a positive constant. When a firm decides to pay the fee, it can train fewer apprentices than the minimum quota. As in the actual regulation, firms are not allowed to exceed the maximum quota.

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

We add more structure to the production function and focus on the case where the optimal number of apprentices without regulation is a fixed proportion of the labor force, $n_a^* = Bn^*$ with $B \in [0, \infty)$. Concretely, we suppose that the production function satisfies the regularity and homogeneity conditions of Assumption 1, guaranteeing the existence of this fixed proportion solution to the firm problem without regulation. We argue that production functions with these properties fit the data well in Section 5.1.²⁸

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

²⁸For completeness, Lemma 4 in Appendix D.1 characterizes all other solutions.

- (i) is homogeneous of degree $\gamma \in (0, 1)$ in (l, l_a) and has constant returns to scale (CRS) in (l, l_a, z) ,
- (ii) the Inada condition holds for the number of workers, $\lim_{n \rightarrow 0} \frac{\partial f}{\partial l} = \infty$,
- (iii) has non-negative cross-derivatives with respect to z , i.e. $\frac{\partial^2 f}{\partial l \partial z} \geq 0$ and $\frac{\partial^2 f}{\partial l_a \partial z} \geq 0$.

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

Proposition 1. *Suppose Assumption 1 holds and firms solve the maximization problem with regulation (3). Then $\forall z \geq 0$,*

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

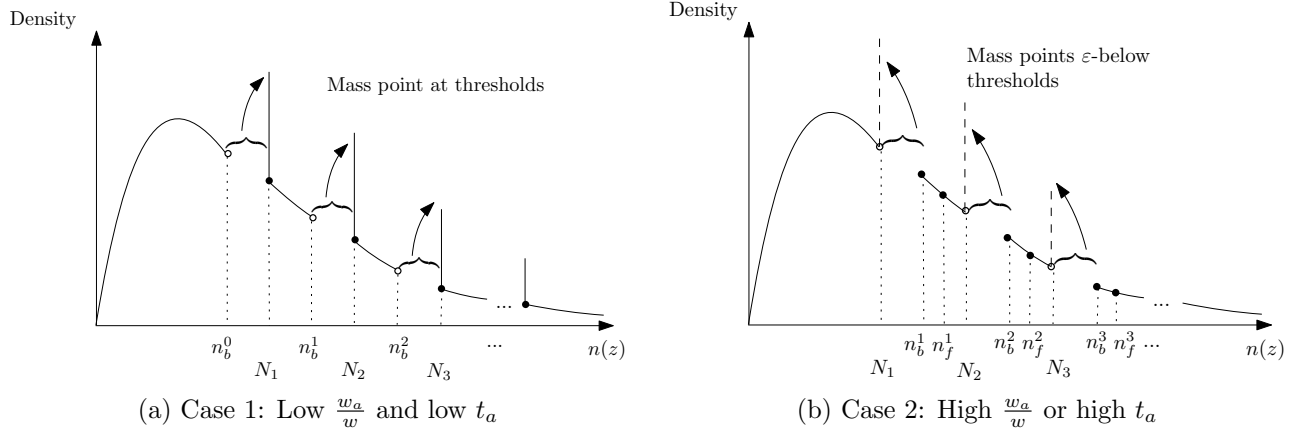
- i. *the number of apprentices without regulation is $n_a^* = B_u n^*$ and is above the maximum quota, $n_a^*(z, t_a) > \bar{n}_a^j$.*
- ii. *there exist cuto $s \{z_b^j, z_r^j\}$ such that firms $z \in [z_b^j, z_r^j]$ increase their size to threshold N_k with $k \geq j$.*
- iii. *firms choose the maximum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

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

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

Figure 5 shows the implications of the policy for the firm size distribution in the two cases from Proposition 1. Panel (a) depicts Case 1 where firms increase their size to train more apprentices. There is bunching at each of the thresholds $\{N_j\}_j$ and missing mass of firms on $[n_b^j, N_{j+1})$. Panel (b) illustrates Case 2, where firms either reduce their size or pay the fee to avoid training more

Figure 5: Firm Size Distribution



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

apprentices. Firms bunch below the thresholds, leaving missing mass on $[N_j, n_b^j)$. If the fee is low enough, firms of size $[n_b^j, n_f^j]$ prefer to pay the fee instead of training 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 decreasing their size and bunching below the thresholds, or by paying the fee.

5 Quantitative Exercises

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

5.1 Parameter Estimation and Moments

Training Costs, Production Function and Managerial Ability Distributions

For the quantitative exercises, we have to choose functional forms of the training cost distribution, the production function and the managerial ability distribution. The reduced-form results indicate

that training costs differ across sectors, but they also vary somewhat within sector, as for instance some high-skill sector firms pay fees while others train apprentices. To reflect this, we use sector-specific training cost distributions. We estimate a non-parametric distribution $\mathcal{T}^k(\cdot)$ for each sector $k \in \{u, s\}$, identifying points of these distributions using simulated method of moments (SMM). 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 derive the full distribution $\mathcal{T}^k(\cdot)$ as the linear interpolation anchored on these points.

Next, we parametrize firms' production function $f^k(l, l_a; z, t_a)$ as a Cobb-Douglas function featuring managerial ability and labor input, allowing for sector-specific parameters. We assume that labor input from workers and apprentices is combined linearly.

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

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

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

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

where $\lambda^k \geq 0$ denotes the location parameter of the distribution, $\theta^k > 0$ the scale parameter and $\xi^k > 0$ the shape parameter.²⁹

Estimation and Identification

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

²⁹Appendix E.1 shows that this distribution provides the best fit of the pre-reform firm size distribution among two and three-parameter distributions typically used to model productivity.

production function, and second, we estimate the parameters of the productivity distribution, both using pre-reform data. Third, we target post-reform firm responses to estimate the training cost distribution. For each sector k , we estimate four parameters, namely the output elasticity of labor γ^k and the ability distribution parameters $\{\lambda^k, \theta^k, \xi^k\}$, as well as $n_{\mathcal{T}} = 13$ points $(t_{a,i}^k, \varrho_i^k)$ of the training cost distribution.

In the first step, we estimate sector-specific production functions using the pre-reform data from 1995 to 2002. We use those firms that do not have apprentices before the reform to estimate the output elasticity of labor γ^k . Since labor is the only production input in the model, we estimate this elasticity by regressing log firm output on log number of full-time workers, controlling for time and firm fixed effects.³⁰

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

Finally, we use SMM to estimate the training cost distribution \mathcal{T}^k , targeting the key firm responses corresponding to the empirical facts from the reduced-form analysis. We use the following post-reform moments in each sector: the bunching mass points (one bin) and missing pass points (± 5 bins) at each of the first 10 thresholds, the number of apprentices by firm size, and the percentage of firms paying the fee. In addition, we target the fraction of firms training the maximum number of apprentices before the reform. We choose the values of $(t_{a,i}^k, \varrho_i^k)$ that minimize the weighted sum $\mathcal{L}((t_{a,i}^k, \varrho_i^k)_{i=1}^{n_{\mathcal{T}}})$ of the absolute difference between the model-implied moments and the empirical moments,

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

where ω_i^k is the weight assigned to moment i . We choose weights to give equal importance to the

³⁰The estimated elasticity would be upward-biased if other production inputs are gross complements of labor. In Appendix E.2, we allow for other inputs, following the procedure of [Levinsohn and Petrin \(2003\)](#) with the [Akerberg et al. \(2015\)](#) correction to estimate output elasticities, as well as some alternative estimation methods. Results are quantitatively similar to the baseline specification.

four groups of moments.³¹

The remaining parameters are implied by the regulation specification. $\{N_j, \underline{n}_a^j, \bar{n}_a^j\}_j$ denote the apprentice thresholds $\{N_j\}_j = \{15, 30, 50, \dots, 20(j-1) + 10, \dots\}$ and minimum and maximum apprentice quotas \underline{n}_a^j and \bar{n}_a^j . w_a^{min} is the minimum wage for apprentices set to 75% of the minimum wage, and $\phi_a = w_{min}$ is the fee parameter proportional to the difference between required and actual apprentices.

Intuitively, the observed firm responses identify different parts of the training cost distribution. For firms with low enough training costs $\tilde{t}_a^k < \zeta_a^k - \frac{w_a}{w^k}$, it is optimal to train as many apprentices as possible. Hence, the fraction of firms choosing the maximum number of apprentices corresponds to the point of the training cost distribution $Pr\{t_a \leq \zeta_a^k - \frac{w_a}{w^k}\} = \mathcal{T}\left(\zeta_a^k - \frac{w_a}{w^k}\right)$. Next, the bunching and missing mass patterns, in particular what share of bunching originates from each missing mass bin around the thresholds, are informative of the “middle” part of the training cost distribution. Finally, if training costs are large, the fraction of firms paying the fee identifies $Pr\{t_a \geq \zeta_a^k + \frac{\phi_a - w_a}{w^k}\} = \mathcal{T}\left(\zeta_a^k + \frac{\phi_a - w_a}{w^k}\right)$. The minimization algorithm first chooses three anchoring percentiles corresponding to these points of the CDF, and interpolates the remaining percentiles uniformly between the anchored points. We choose $n_{\mathcal{T}} - 3 = 10$ remaining percentiles, and estimate the $t_{a,i}^k$ that minimizes the objective function for each sector.³²

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

Table 3 summarizes the estimated parameters, with bootstrapped 95% confidence intervals. With 0.61 and 0.58, high-skill and low-skill sectors have similar labor shares. The managerial ability distribution has a larger location λ and scale parameter θ in high-skill sectors, but a smaller shape parameter ξ . This implies that high-skill sector firms have higher average managerial ability.³³

³¹The estimation results are robust to weighting moments by the inverse of their variance instead. See Appendix E.3 for details.

³²We use a *pattern search* algorithm. Spreading the additional $n_{\mathcal{T}} - 3$ percentiles as Chebyshev nodes between the anchored points yields results similar to the uniform interpolation.

³³The mean of a $GEV(\lambda, \theta, \xi)$ distribution is $\lambda + (1 - \Gamma(1 - \xi))^{\frac{\theta}{\xi}}$.

Table 3: Estimated Parameters

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

Notes: The table shows structural parameter estimates, with bootstrapped 95% confidence intervals in parentheses.

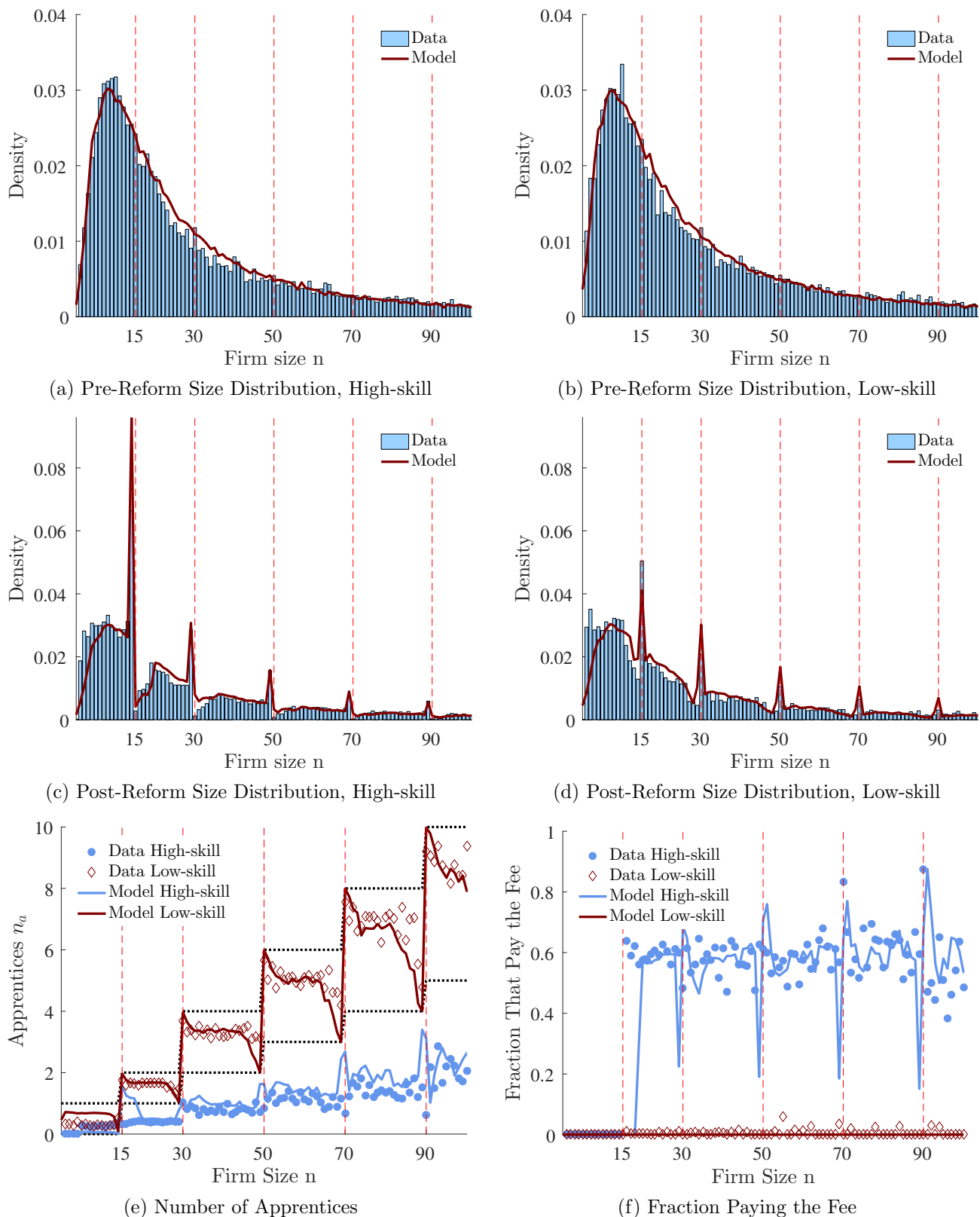
Supply Side

To close the model and compute the effect of the policy on aggregate outcomes, we have to calibrate the supply side of the economy. We use data on the number of firms before the reform and the estimated firm size distribution to obtain the initial supply of workers in each sector L^k . 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 the EAM data, we compute the average wages of workers by sector in 2002, the last year before the reform. Workers in high-skill sectors earn on average 3.95 times the minimum wage, whilst workers in low-skill sectors earn 3.25 times the minimum wage. Finally, we assume that the number of potential apprentices L_a exceeds aggregate demand for apprentices. As mentioned in Section 4.1, this assumption is consistent with the fact that training courses are strongly oversubscribed in Colombia.

5.2 Goodness of Fit

Table 4 reports the targeted moments in the data and the corresponding values we obtain from the estimated model. In cases where we match a full function, reference to the relevant figure showing the fit is provided. Overall, the estimated model closely resembles the data. We match, almost exactly, the fraction of firms choosing the maximum quota before the reform. Moreover, Panels (a) and (b) of Figure 6 show the targeted firm size distribution 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 workers and slightly overestimating the mass between 15 and 40 workers.

Figure 6: Goodness of Fit



Notes: The figure depicts the model fit to targeted moments. Panels (a) and (b) show the distribution of firm size (number of full-time workers) for pre-reform (1995-2002) and Panels (c) and (d) show the firm size distribution post-reform (2003-2009). Panel (e) shows the number of apprentices by firm size, and Panel (f) shows the fraction of firms paying the fee by firm size, both in the post-reform period.

Table 4: Targeted Moments

Moment Description	Model	Data
<i>High-Skill Sectors</i>		
Pre-reform firm-size distribution	see Fig. 6a	see Fig. 6a
Fraction choosing maximum apprentices pre-reform (≥ 15 workers)	1.5%	1.5%
Bunching and missing mass (Fact 1)	see Fig. 6c	see Fig. 6c
Average number of apprentices by firm size (Fact 2)	see Fig. 6e	see Fig. 6e
Fraction paying fee by firm size (Fact 3)	see Fig. 6f	see Fig. 6f
<i>Low-Skill Sectors</i>		
Pre-reform firm-size distribution	see Fig. 6b	see Fig. 6b
Fraction choosing maximum apprentices pre-reform (≥ 15 workers)	1.5%	1.5%
Bunching and missing mass (Fact 1)	see Fig. 6d	see Fig. 6d
Average number of apprentices by firm size (Fact 2)	see Fig. 6e	see Fig. 6e
Fraction paying fee by firm size (Fact 3)	see Fig. 6f	see Fig. 6f

Notes: The table compares targeted moments in the model and data. In case a full function is targeted, the table provides reference to the corresponding figure.

Figure 6 further illustrates how the model fits our three empirical facts very well. Panels (c) and (d) show the observed and simulated firm size distribution after the reform, corresponding to fact 1. The model captures the incentives of high-skill sector firms to bunch just below the threshold, and those of low-skill sector firms to bunch at the thresholds. Note that the model somewhat overestimates the fraction of bunching firms in high-skill sectors. This happens mainly for two reasons. First, as mentioned above, the number of firms between 15 and 40 workers is slightly overestimated. Second, the model abstracts from other margins of substitution that could “smooth” firm responses. For instance, if other inputs such as capital, material or other types of labor substitute for regular workers, this may reduce the fraction bunching.

Panel (e) shows the fit of the model to fact 2, the number of apprentices by firm size. High-skill sector firms choose an average number below the minimum quota, while the number is close to the maximum quota in low-skill sectors. Moreover, the model highlights that there is a decreasing pattern within each regulation bracket in low-skill sectors. Almost all firms at the thresholds choose the maximum number of apprentices, and the fraction of firms choosing the maximum decreases with firm size between thresholds. Finally, Panel (f) shows that the simulated model also replicates fact 3, with around 60% of firms in high-skill sectors with at least 15 workers paying the fee.³⁴

³⁴The estimated model somewhat underestimates the fraction of firms paying the fee just below the thresholds. This occurs because the solution algorithm allows firms to either bunch *or* pay the fee for simplicity, while some firms in the data do both.

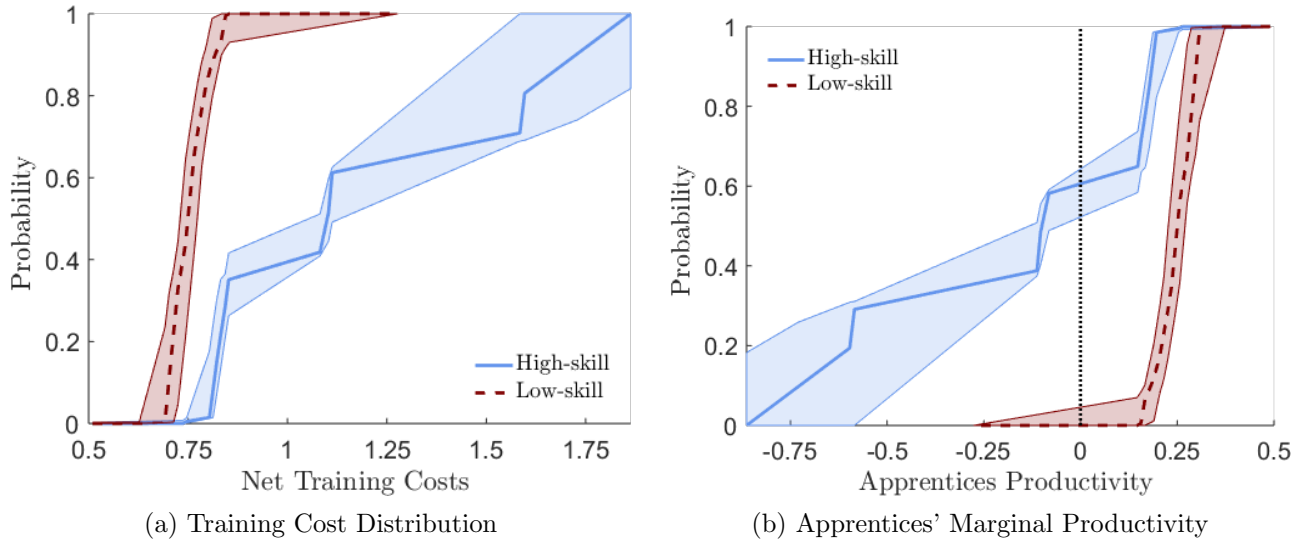
Meanwhile, as in the data, virtually no low-skill sector firms pay the fee.

Additionally, we conduct an out-of-sample fit exercise where we split the data set in half, run the estimation on one half and evaluate the fit using the other half not included in the estimation. Appendix Figure E4 shows that the resulting moments closely resemble the untargeted data, and Appendix Table E5 shows that the estimation errors and score function are similar to the baseline estimation.

5.3 Training Cost Distribution

As the first quantitative result, we present net training cost distributions, which are crucial in explaining firms' willingness to train apprentices. Panel (a) of Figure 7 shows that estimated training costs are substantial in general. The lower average training costs in low-skill sectors ($E(t_a^u) = 0.75$) compared to high-skill sectors ($E(t_a^s) = 1.19$) is consistent with most training happening in low-skill sectors. The training costs in high-skill sectors are also significantly more dispersed ($Var(t_a^s) = 0.13$) than in low-skill sectors ($Var(t_a^u) = 0.0016$). Within high-skill sectors, the higher variance and the plateaus at intermediate training cost levels reflect the more varied firm responses, where only some firms train apprentices while others pay fees.³⁵

Figure 7: Training Cost Distribution and Apprentices' Marginal Productivity



Notes: Panel (a) of the figure shows the estimated distribution of training cost t_a , and Panel (b) shows the implied marginal productivity of apprentices $\zeta_a - t_a$. The values in both panels are obtained by normalizing $\zeta_a = 1$. Shaded areas denote bootstrapped 95% confidence intervals.

An interesting additional interpretation of training costs stems from transforming them into

³⁵Appendix E.5 shows alternative parametric specifications of training costs, using a calibrated truncated normal or uniform distribution. These would yield a smoother shape of the training cost distribution, but result in a somewhat worse fit of the model.

apprentices' marginal productivity. As mentioned before, we are not able to separately identify t_a and ζ_a and we normalize ζ_a to 1. However, our model identifies the marginal productivity of apprentices $\zeta_a - t_a$ shown in Panel (b) of Figure 7. In low-skill sectors, $\zeta_a - t_a$ is around 25% on average, suggesting that the marginal productivity of an apprentice during training is equal to 25% of the marginal productivity of the average worker. Moreover, the figure shows that apprentices have negative marginal productivity in around 60% of high-skill sector firms, while apprentices' productivity is positive in most low-skill sector firms. As discussed in Section 4.1, this ties in well with the empirical observation that many firms in high-skill sectors pay fees to avoid training.

5.4 Quantitative Effects of the Apprenticeship Regulation

Partial Equilibrium, General Equilibrium and Dynamic Scenarios

We quantify the effects of the apprenticeship regulation under three scenarios: (i) the short-term or *partial equilibrium* effects where wages and prices are fixed, (ii) the *general equilibrium* effects where labor and goods markets clear in each sector, and (iii) the *dynamic* effects where trained apprentices increase the future supply of workers. In the baseline dynamic scenario, we assume that apprentices acquire the skills necessary to replace an average worker in their sector of training ($\chi_a = 1$), and we present sensitivity analysis with respect to this parameter at the end of this section.

Wages are key to quantifying aggregate and welfare effects. Table 5 shows wages (relative to the minimum wage) by sector in each scenario. In partial equilibrium, wages are fixed to 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 any previously displaced workers, exerting downward pressure on wages. Wages fall by 0.9% in low-skill sectors and by 0.5% in high-skill sectors. When trained apprentices become workers in the next period, there is a further decrease in wages. Since low-skill sectors train more apprentices, wages fall by more (4.3%) than in high-skill sectors (1.3%).

Aggregate Effects

We begin with the impact of the regulation on aggregate outcomes. In particular, the change in aggregate output reflects the change in overall efficiency, and changes in the total number of workers and apprentices are informative of the mechanisms affecting output. Table 6 shows the effects under each scenario in Columns (1) to (4). First, we find that despite the large labor input responses, the partial equilibrium effects on aggregate output are relatively small. Panel A shows

Table 5: Wages

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

Notes: The table shows wages of regular worker under the different scenarios. All wages are in units relative to the minimum wage. Partial equilibrium wages are based on EAM data in the last year before the reform (2002). General equilibrium and dynamic effects wages come from the estimated model and are deflated by production prices in each sector.

that firms respond to the policy by increasing the number of apprentices, which displaces some workers. In total, around one worker is displaced for every five apprentices. Moreover, the first two columns show that approximately 1.3% are displaced, with more layoffs in low-skill sectors. The overall effect on output is negative but close to zero. However, there is some reallocation of production across sectors: High-skill sector firms decrease production by 0.3% and low-skill sector firms increase production by 0.3%. In Section 5.5, we show additional reduced-form evidence consistent with these partial equilibrium effects, and discuss the relationship to existing literature.

In contrast, Panels B and C show that general equilibrium and dynamic effects can lead to substantial increases in aggregate output. In general equilibrium, aggregate output increases by 0.7%. There is higher growth in low-skill sectors and the initial output losses are reverted in high-skill sectors. The positive effects on output can be even stronger when apprentices become workers in the dynamic scenario, resulting in an overall output increase of around 3.7%. While output in high-skill sectors also increases in Panels B and C, some reallocation of production towards low-skill sectors persists as they see larger output growth under all scenarios.

The general equilibrium and dynamic effects are illustrative of the potential long-run benefits of expanding apprenticeship training. Importantly, we assume in these scenarios that any displaced workers are absorbed back into the labor market. If some of them remain unemployed in general equilibrium, this would temper the aggregate impact of the regulation, and effects may be closer to the partial equilibrium scenario even in the longer run.

Winners and Losers

The policy also has a distributional impact, as the effects on the welfare of apprentices, workers and firm owners vary. There are winners and losers from the regulation, both across these groups of agents and across sectors. Since we assume no entry or exit, the effect on firms can be readily seen as the change in profits. To quantify changes in the welfare of apprentices and workers,

Table 6: Quantitative Effects of the Apprentice Regulation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	<i>Aggregate Outcomes</i>				<i>Changes in Agents' Welfare $\Delta U_j/U^*$</i>			
	Workers	% Workers	Apprentices	% Output	Apprentices	Workers	Firms	Total
A. Partial Equilibrium								
High-skill Sectors	-1905	-0.89	4937	-0.34	0.27	-0.54	-0.46	-0.74
Low-skill Sectors	-3519	-1.67	17866	0.30	1.13	-0.97	0.13	0.30
Total	-5423	-1.28	22803	-0.06	0.65	-0.73	-0.20	-0.28
B. General Equilibrium								
High-skill Sectors	0	0.00	5123	0.22	0.27	0.07	-0.07	0.28
Low-skill Sectors	0	0.00	18098	1.27	1.14	-0.75	0.29	0.68
Total	0	0.00	23221	0.68	0.66	-0.29	0.09	0.46
C. Dynamic Effects ($\chi_a = 1$)								
High-skill Sectors	5123	2.40	5579	1.70	1.43	0.40	1.07	2.90
Low-skill Sectors	18098	8.58	18586	6.16	4.61	-3.66	1.50	2.45
Total	23221	5.47	24166	3.68	2.84	-1.40	1.26	2.70

Notes: Columns (1) to (4) of the table show the effects of the apprenticeship regulation on aggregate outcomes, namely on the number of workers, the number of apprentices and output. Columns (5) to (7) show the effect on the welfare of apprentices, workers and firm owners. Column (8) shows the sum of welfare effects across the groups of agents from Columns (5) to (7). Panel A shows partial equilibrium effects, Panel B shows general equilibrium effects, and Panel C shows dynamic effects under full skill transferability ($\chi_a = 1$).

we assume that all agents have the same homothetic preferences, such that their utility is linear in income. For apprentices, we normalize their outside-option income to zero.³⁶ For workers, we track the welfare of those who are working before the reform. In the partial equilibrium scenario, we suppose that the displaced workers earn no wage when they become unemployed. Taken together, these assumptions probably imply that we obtain an upper bound on the welfare consequences for apprentices and workers.

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

In Column (5), apprentices gain from the regulation under all scenarios. In the short-run, their

³⁶The outside option is important to determine the gains for apprentices. The alternatives could be unemployment or informality, both of which have low value. In Colombia, there is no unemployment insurance over the sample period. The average informal wage is below the minimum wage and more than 40% of informal workers earn less than half the minimum wage (Bernal, 2009). Moreover, the excess supply of potential apprentices is suggestive of the benefits young people receive from apprenticeships (SENA, 2016).

welfare improves as the wages they earn as apprentices exceed their outside option. In the dynamic scenario, these effects are magnified as they become employed as regular workers. Since low-skill sectors train more apprentices, the positive effects are larger there, although high-skill sectors pay higher wages. The positive impact on apprentices reflects the goal of the reform of improving youth employment opportunities.

In contrast, Column (6) shows that incumbent workers are harmed by the regulation in partial equilibrium, since some of them are displaced by apprentices. In general equilibrium and dynamically, workers re-enter employment, but wages and prices change. The negative effects on workers' welfare persist in low-skill sectors, where workers' real wages fall. However, in high-skill sectors the decrease in workers' welfare is reverted in general equilibrium. This is driven by higher demand for goods, which more than offsets the increase in labor supply, increasing the real wages of high-skill sector workers.³⁷ The dynamic scenario further magnifies these effects, substantially lowering the welfare of incumbent workers in low-skill sectors.

Firms are unequally affected by the policy. In partial equilibrium, firm profits fall in high-skill sectors, while they increase in low-skill sectors. In general equilibrium, high-skill sector firms experience smaller negative effects and the gains of low-skill sectors are reinforced, as firms benefit from higher overall labor supply at lower wages. In all sectors, profits increase more strongly when apprentices increase the future supply of productive workers. In addition, Appendix Figure A3 shows the distribution of firms' profit changes, suggesting substantial heterogeneity in the impact on firms, especially in high-skill sectors.

Transferability of Skills and Dynamic Effects

Table 6 shows large positive dynamic effects of the regulation. Recall that these are calculated under the assumption $\chi_a = 1$, reflecting the optimistic view that skills are transferable across firms within apprentices' sector of training, such that they are eventually able to provide the same effective units of labor as the average worker in that sector. In our view, this is a useful benchmark for the dynamic impact of training more young workers. Appendix Figure E7 provides sensitivity analysis with respect to this crucial parameter. When skills are not transferable at all ($\chi_a = 0$), the effects of the regulation correspond to the general equilibrium scenario, which does not take into account that apprentices can become future workers. Importantly, this implies that the effects on overall output, profits and welfare are positive for for any $\chi_a \in [0, 1]$. Moreover, the figure

³⁷Note that the wages shown in Table 5 are deflated by production prices in each sector. When taking into account consumer prices instead, real wages decrease in low-skill sectors but increase in high-skill sectors.

shows that dynamic gains increase linearly with skill transferability. We conclude that the positive dynamic effects of the regulation are robust to the choice of χ_a , but their magnitude naturally depends on the ability of apprentices to make productive use of the skills they acquired during training.

5.5 Reduced-Form Evidence on Firm Outcomes

To complement these model-based findings, we next present additional reduced-form evidence of the effects of the apprenticeship regulation on firm outcomes. Since all firms are subject to the new regulation, obtaining credible reduced-form estimates of its overall effects is difficult. Instead, we exploit that fact that firms are differentially affected based on their size and compare firms above vs. below the regulation thresholds who face different apprentice quotas.

We estimate the following difference-in-difference specification:

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

where $Above_i$ is an indicator for firm size above the threshold in the last year pre-reform year 2002, $Post_t$ is an indicator for the post-reform years 2003 onward, α_i is a firm fixed effect, δ_t is a year fixed effect and ϵ_{it} is an error term. As discussed in Section 3.2, firms' actual post-reform size is subject to endogenous changes, such that observed differences in outcomes across brackets reflect a mixture of the causal effect of the regulation and selection. Hence, we assign $Above_i$ based on *pre-reform* firm size, when the distribution is still smooth. We include firms within five size bins above and below the regulation thresholds, pooling across the first fifteen thresholds, and we restrict the sample to firms that stay within two adjacent regulation brackets across years.³⁸ Moreover, the analysis is limited to the years 1999 to 2006. Column (4) of Table 1 shows summary statistics of this *threshold sample*. Since only firms around the first thresholds are included, the main difference to the full sample is the smaller average firm size of around 30 workers, but other characteristics not related to size tend to be similar.

Table 7 shows results for four outcomes, the number of apprentices and full-time workers, log output and the profit rate. The first two columns show a strong “first stage”, where firms above the thresholds are induced to train significantly more apprentices. As expected, the causal effect

³⁸More precisely, in order to focus on the first fifteen regulation thresholds, we exclude firms with more than 300 workers in 2002. This restriction ensures that there are at least some firms on both sides of each threshold. Since the firm size distribution declines fast (see Figure 3 and Appendix Figure A1), this excludes only few large firms. Moreover, we exclude firms which jump across more than two adjacent regulation brackets, which is likely to be driven by factors other than the apprenticeship regulation. This is implemented by excluding firms with year-to-year changes in the number of workers of more than 50.

of being subject to a higher quota on the actual number of apprentices is stronger in low-skill sectors, with an increase of 0.80 compared to 0.16 in high-skill sectors. Next, the results indicate that firms training more apprentices due to the regulation decrease the number of regular workers. Again, the effects are larger in low-skill sectors, where the point estimate is -1.8. In high-skill sectors, there is a smaller and insignificant effect of -0.86. Finally, the effect on log output and profit rates are small and insignificant in all sectors. In addition, Appendix Figure A4 plots yearly difference-in-difference coefficients. In both high-skill and low-skill sectors, there are no significant pre-trends in all four outcomes. Coefficients on the number of apprentices become positive exactly at the time of the reform and stay relatively constant throughout the post-reform period. Similarly, there are clearly negative effects on the number of 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.

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

Sector	(1)		(2)		(3)		(4)		(5)		(6)		(7)		(8)	
	Apprentices		Workers		Log Output		Profit Rate Π/Y									
	High	Low	High	Low	High	Low	High	Low	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								

Notes: The table shows results from difference-in-difference regressions as described by equation (4). Regressions are run on the threshold sample, using years 1999 to 2006, and include year and firm fixed effects. Standard errors clustered at the firm level in parentheses.

Appendix Table A5 shows results for 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 regular workers reported in the survey data, where the estimates for both sectors are remarkably similar to those from the main administrative data. This may be indicative of real labor adjustments, as firms should have little incentive to change the reporting of workers in the survey.³⁹ Moreover, there are no significant effects on two types of “semi-formal” labor, namely temporary workers and outsourced workers, although the point estimate on outsourced

³⁹There are two additional differences between the measures of the number of workers in the administrative and survey data. First, the administrative measure is calculated in terms of equivalent full-time workers, whereas the survey simply asks about the number of workers. Second, the administrative measure has to be reported at specific times when SENA determines apprentice quotas, which is unlikely to coincide with the time of the survey.

workers in low-skill sectors is relatively large. In Panel C, we do not find significant changes in any other production inputs such as capital and intermediate inputs.

Reduced-Form vs. Model-Based Results. The difference-in-difference specification relies on a short-run, within-sector comparison of firms subject to different apprentice quotas. We argue that these results are most directly related to the partial equilibrium effects from the model, whereas obtaining credible reduced-form evidence to validate general equilibrium effects at the sector level or longer-run dynamic effects would be more difficult. Appendix Figure A5 shows a comparison of the reduced-form results from Table 7 to analogously simulated differential effects on firms above vs. below regulation thresholds from the partial-equilibrium model.⁴⁰ As predicted by the model, the effect on the number of apprentices is stronger in low-skill sectors. The effect on the number of workers is remarkably similar in model and data, and the partial-equilibrium effect on output and profits is small and negative in both. For high-skill sectors, the model seems to somewhat over-predict the apprentice intake as well as the negative impact on profits, which is likely related to the over-estimated bunching shown in Figure 6.

Relationship to the Literature. Our difference-in-difference analysis is closely related to Ospino (2016, 2018) who finds that the reform leads to labor substitution, but has positive effect on some measures of firm productivity. Like Ospino (2016), we find that higher apprentice quotas lead to reductions in the number of regular workers, although we do not find significant effects on outsourced workers. Moreover, the results in Panel D of Appendix Table A5 do not suggest a significant effect on firms’ productivity measured by output per worker and total factor productivity.⁴¹ More broadly, the results are related to recent studies that find mixed evidence of the effect of changes in labor input composition on firm outcomes such as output and profits (e.g. de Mel et al., 2019; Hardy and McCasland, 2020). An important difference to these experimental studies is the Colombian apprenticeship regulation “forces” all firms to take apprentices or pay costly fees, whereas they tend to offer wage subsidies or help with finding apprentices/workers. In addition, they study very small firms where initial labor constraints may be more likely to play a role. This could explain why we find less positive effects on firm outcomes in partial equilibrium. Moreover, existing studies do not tend to find clear displacement effects of training programs on

⁴⁰In order to obtain analogous effects from the model, we simulate firms’ outcomes with and without the apprenticeship regulation in partial equilibrium, and then compute a hypothetical difference-in-difference regression focusing on the same sample of firms as the reduced-form specification.

⁴¹These differences in results could be due to several factors. First, Ospino (2018) uses a regression discontinuity design, albeit exploiting similar variation in apprentice quotas around the regulation thresholds. Second, we use a different data set, including administrative data. Third, our specification includes more years and pools across multiple thresholds.

other workers (e.g. [Attanasio et al., 2017](#), [Alfonsi et al., 2020](#)). On the contrary, we find such effects in the short-run, especially in low-skill sectors. The model in [Section 4](#) naturally predicts some displacement when firms train more, as workers and apprentices are to some extent substitutable in production. The difference-in-difference results are in line with this prediction. Note that the reduced-form results capture the total effect on the number of workers in a firm, which includes both direct substitution between workers and apprentices and additional changes in the number of workers due to endogenous firm size responses. Hence, the difference-in-difference results may not directly correspond to a pure labor substitution effect.

5.6 Model Extensions: Adding Labor Market Frictions

In this section, we present model extensions considering two types of labor market frictions that could potentially explain some of the observed firm responses and affect the training cost estimation.

Dynamic Frictions and Wage Compression. One possibility is that apprentices may face labor market frictions after training which prevent them from moving to different firms. In this case, training firms can re-hire their apprentices at a wage below apprentices’ marginal productivity. In other words, there is *wage compression*. This could occur when firms have market power or due to asymmetric information, for instance ([Acemoglu and Pischke, 1998, 1999b](#); [Dustmann and Schönberg, 2009](#)).

To shed light on this mechanism, we consider a simple two-period extension of the model with sector-specific dynamic frictions.⁴² A fraction $\rho > 0$ of apprentices are unable to move after training, and firms can retain those at a discounted wage rate w_a^r . As is well known from the literature, a key result arising in this case is that firms have stronger incentives to train apprentices, since they can appropriate a larger share of the returns to training. We can obtain a modified FOC, which implies that a firm wants to train apprentices if

$$t_a \leq \zeta_a - \frac{w_a}{w_t} + \underbrace{\rho\beta \frac{(\zeta_a^r w_{t+1} - w_a^r)}{w_t}}_{\text{Additional benefits of training}}$$

where β is the discount factor and ζ_a^r the productivity of apprentices retained after training.

Larger additional benefits due to higher frictions would induce more firms to train apprentices. Importantly, if the extent of frictions differs across sectors, this may confound the heterogeneity

⁴²See [Appendix D.4](#) for details of the model extension and [Appendix E.8](#) for details of the calibration and quantitative results.

in training costs we find. To address this issue, we calibrate the parameters of the extended model using the retention probabilities from Appendix Table B4 and additional information on average wages by sector from the PILA data. The calibration implies dynamic benefits in high-skill sectors of $B^s = \beta \rho^s \frac{(c_a^{r,s} w_{t+1}^s - w_a^{r,s})}{w_t^s} = 0.40$, larger than $B^u = 0.33$ in low-skill sectors. If anything, this amplifies the differences in training cost between the two sectors: $\Delta t_a = E(t_a^s) - E(t_a^u) = 0.50$, compared to $\Delta t_a = 0.44$ in the baseline estimation.⁴³ Moreover, aggregate effects remain very similar when considering dynamic frictions. Firms hire more apprentices and displace fewer workers, but total effective units of labor remain relatively constant and the effects on output and total welfare are similar to the baseline scenario. Interestingly, the impact on profits remains similar too, since the higher estimated training cost revealed by given firm responses offset the additional benefits of training to firms.

Hiring Frictions. Another type of friction could be given by firms facing costs of adjusting their labor inputs, for instance due to asymmetric information in the process of recruiting workers. To gauge the potential importance of this mechanism, we consider a model extension allowing for costs firms incur when hiring new workers.⁴⁴ In particular, we re-estimate the model under two scenarios, one with a fixed cost of adjusting the number of workers upwards, and one with a variable cost per new hire. As one may expect, adding hiring costs attenuates firm size responses, and especially imply less bunching at the thresholds in low-skill sectors. Consequently, some firms train fewer apprentices as they face a smaller maximum quota. The estimated training cost distributions remain relatively unaffected, however. Moreover, allowing for this type of friction generally worsens the fit of the model to observed firm responses. This finding is consistent with the sharp bunching responses from Section 3.2, which indicate that many firms are not prevented by frictions from adjusting labor inputs.

6 Counterfactual Simulations

6.1 Decomposing the Effects of the Regulation

The Colombian apprenticeship regulation is a bundle of apprentice quotas, the decrease in the apprentice minimum wage, and the possibility of paying the fee. Our first counterfactual analysis studies the separate effects of these components. Here we present this decomposition in partial

⁴³A potential caveat of this extension is that we consider heterogeneity in frictions across sectors, but not within sectors. Investigating whether idiosyncratic firm-level frictions could explain some of the within-sector dispersion in estimated training costs is outside the scope of our analysis.

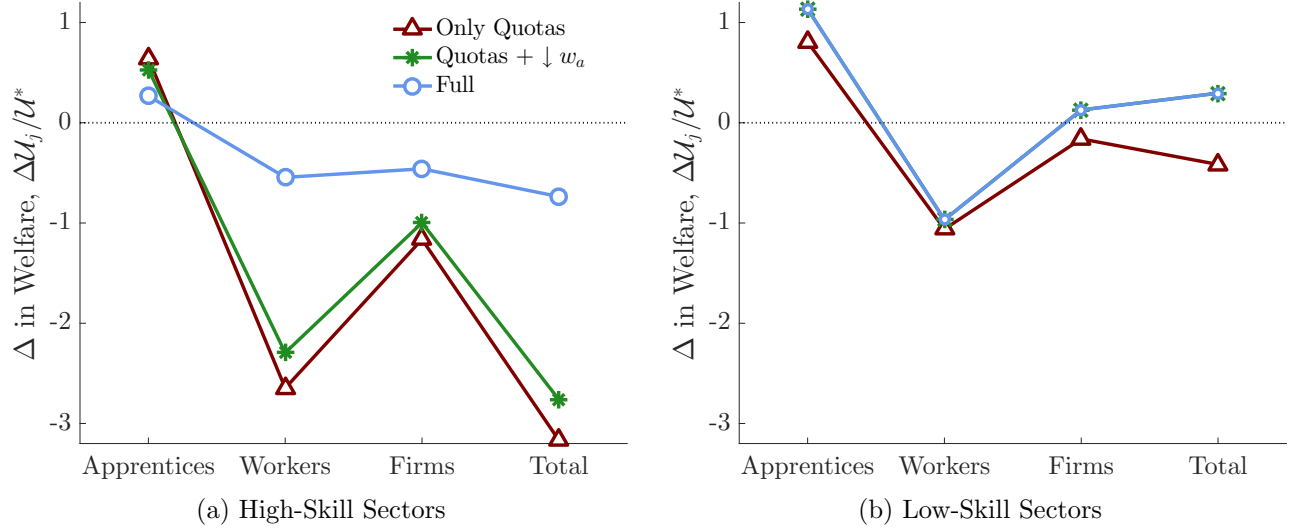
⁴⁴See Appendix D.5 for details of the model extension and Appendix E.8 for details of the quantitative results.

equilibrium and study the effects on the welfare of each group of agents. Appendix E.7 shows additional details, including effects on aggregate outcomes. We consider three scenarios: (i) only apprentice quotas, (ii) quotas and the decrease in the apprentices' minimum wage, and (iii) the full regulation.

Figure 8 plots the change in total and group-level welfare under each scenario calculated analogously to the results from Table 6. Panel (a) shows results for high-skill sectors. The scenario with only apprentice quotas generates gains for apprentices, but large losses for workers and firms. When the apprentice minimum wage falls in addition, the number of apprentices trained increases significantly. Losses for firms and workers as well as gains for apprentices only change slightly, though. Finally, when the possibility of paying the fee is added in the full regulation, losses for firms and their workers are significantly dampened, since those with the highest training cost now pay the fee instead of training. However, fewer apprentices are trained, such that the welfare gain of apprentices decreases. Panel (b) shows corresponding results for low-skill sectors. Here, apprentices and firms gain from the full regulation, but workers lose. Compared to quotas alone, the decrease in the apprentice minimum wage is crucial for firms to gain in partial equilibrium. Adding the possibility of paying fees has close to no effect (the respective points in the graph overlap almost perfectly). Interestingly, welfare of apprentices increases with the lower minimum wage due to the strongly increased number of apprentices hired by firms. Incumbent workers lose from the regulation in all scenarios.

This decomposition illustrates the role of each component of the policy. The minimum quota guarantees that many firms train. The maximum quota ensures that firms do not excessively substitute incumbent workers for apprentices, using apprentices as “cheap labor”. To incentivize firms to train, 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, the possibility of paying fees reduces the negative impact on some firms, in particular those whose training cost would be very high.

Figure 8: Policy Decomposition



Notes: The figure shows welfare effects of different combinations of policy instruments for each group of agents, namely apprentices, incumbent workers and firm owners. The effects are scaled relative to no-regulation total aggregate utility, U^* . *Total* denotes the sum of welfare changes across the three groups.

6.2 Alternative Apprenticeship Policies

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

Subsidizing Apprenticeship Training

First, we consider subsidizing apprenticeship training. We focus on budget-balanced policies, where firms pay payroll taxes at rate τ , and these funds are used to pay a subsidy of $\varsigma\%$ of apprentices' wages. This policy is similar to real-world training subsidies,⁴⁵ and relates to subsidy interventions analyzed in the literature (e.g. Crépon and Premand, 2019, Alfonsi et al., 2020).

Given the subsidy policy, a firm (z, t_a) solves

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

The linear labor inputs again imply corner solutions.⁴⁶ Whenever the tax τ or the subsidy ς is large enough, firms have incentives to train more apprentices. These incentives are stronger for

⁴⁵For instance, Latin American including Chile and Brazil have implemented training subsidies (see Fazio et al., 2016). Moreover, the UK “Apprenticeship Levy”, a subsidy on training costs financed by payroll taxes, closely resembles the policy we consider here.

⁴⁶See Appendix E.9 for a full characterization of the solution.

firms with low training costs or in sectors where the difference between the wages of workers and apprentices is large. We compute the policy (τ, ς) such that the government’s budget is balanced and the total number of apprentices trained is the same as under the benchmark policy (see Table 6). The solution yields a tax rate of $\tau = 0.05\%$ and a subsidy of $\varsigma = 3.0\%$.

Panel A of Table 8 shows total tax revenue and subsidy payments in Columns (1) to (3). Firms in high-skill sectors pay more taxes than those in low-skill sectors, but 94% of subsidy payments go to low-skill sector firms, where training costs are lower. Accordingly, Column (6) shows that 98% of apprentices are trained in low-skill sectors. Hence, high-skill sector firms effectively subsidize training in low-skill sectors. Intuitively, if firms in both sectors face the same tax and subsidy rates, it is disproportionately more attractive to firms with low training costs to train. In fact, the subsidy leads to an extreme concentration of training in firms with the lowest training costs even within low-skill sectors. The percentage of firms with at least one apprentice decreases from around 60% of firms to only 5%. Training firms often have as many apprentices as possible, hiring only the workers necessary to provide training. Moreover, the subsidy leads to more displacement of regular workers, but the effect of the subsidy on total output is slightly less negative than that of the benchmark regulation.

Table 8: Effect of Alternative Apprenticeship Policies

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	<i>Budget Balance</i>			<i>Aggregate Outcomes</i>			
	Tax Revenue	Subsidy Payments	Balance	Workers	% Workers	Apprentices	% Output
A. Training Subsidy							
High-skill Sectors	387	39	348	-317	-0.15	348	-0.07
Low-skill Sectors	306	654	-348	-6793	-3.21	22430	0.01
Total	693	693	-0	-7110	-1.67	22778	-0.03
B. Sector-Specific Apprentice Minimum Wage							
High-skill Sectors	-	-	-	-704	-0.33	4927	0.09
Low-skill Sectors	-	-	-	-5284	-2.50	17866	0.04
Total	-	-	-	-5988	-1.41	22793	0.07

Notes: The table shows aggregate effects of alternative apprenticeship policies. Columns (1) to (3) present budget balance variables: tax revenues, subsidy payments and net fiscal balance. Columns (4) to (7) show (partial equilibrium) effects on aggregate outcomes: the change in the number of workers, the percentage change of workers, the change in the number of trained apprentices and the percentage change in aggregate output. The total number of apprentices in Column (6) exhibits a small difference from Table 6 due to approximation error in the simulation.

This counterfactual highlights the importance of minimum and maximum apprentice quotas when regulation is uniform across all firms, as quotas induce a more even distribution of apprentice-

ships across firms both within and across sectors. With a uniform subsidy and no other regulation, training is concentrated only in those firms with the lowest training cost. However, if policymakers' main goal is increasing total output, the subsidy performs slightly better than the benchmark regulation.

Sector-Specific Apprenticeship Minimum Wage

As a second counterfactual, we consider an alternative policy that explicitly takes into account heterogeneity in training costs. Specifically, we allow apprentices' minimum wages to differ across sectors. This policy resembles the situation in some European countries including Germany, where apprentices' wages are governed by sector-specific agreements rather than general restrictions (Steedman, 2012). Again, there are no quotas or fees in the counterfactual. We find sector-specific minimum wages that clear the labor market such that each sector trains as many apprentices as in the benchmark regulation.⁴⁷ Relative to the minimum wage for regular workers, the apprentice minimum wage is $w_a^{*s} = 0.77$ in high-skill sectors and $w_a^{*u} = 0.98$ in low-skill sectors. Table 8 shows that the sector-specific minimum wage policy increases total output already in partial equilibrium, which is driven by growth both in low-skill *and* in high-skill sectors. Thus, the policy diminishes the reallocation of resources towards low-skill sectors observed under the benchmark regulation.

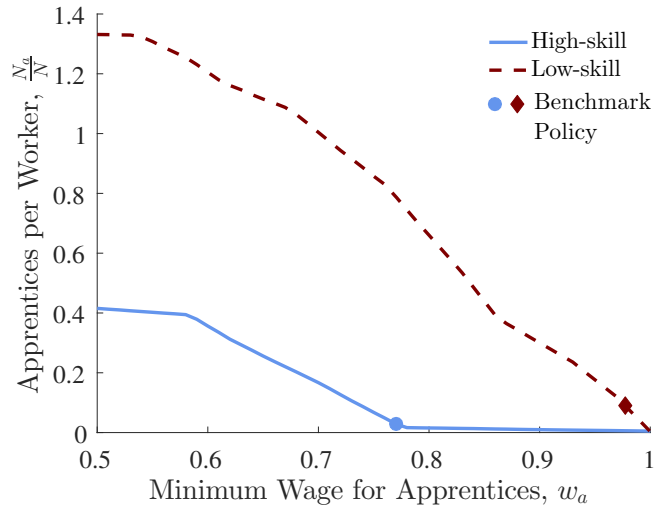
Under the sector-specific minimum wage, the same number of apprentices are trained in each sector without inducing firm size distortions. Thus, comparing the number of displaced workers between Table 8 (Column 4) and Table 6 (Column 1) is informative of the relative importance of firm size distortions vs. labor substitution under the benchmark regulation. Taking the difference at face value, around 64% of total displacement is due to firm size distortions in high-skill sectors where firms reduce their size to bunch below thresholds. In low-skill sectors, on the other hand, firms increase their size to bunch above thresholds, which counters labor substitution and reduces worker displacement by 33%.

This counterfactual is relatively conservative in the sense that it restricts the number of apprentices in each sector to be the same as in the benchmark regulation, where most training occurs in low-skill sectors. In fact, it would be possible to shift this imbalance and increase training in high-skill sectors by further adjusting apprentice minimum wages. This is illustrated in Figure 9, which shows the number of apprentices (per worker) by the level of the minimum wage in each sector. The dots mark the number of apprentices from the benchmark regulation targeted by the main counterfactual. Not surprisingly, low-skill sectors train more apprentices than high-skill

⁴⁷See Appendix E.9 for details.

sectors at any given level of the minimum wage. Furthermore, the figure implies that a higher number of apprentices in high-skill sectors could be achieved by reducing the sectoral minimum wage, enabling apprentices to compensate firms for the costlier training they receive. For instance, the same apprentice-to-worker ratio as in low-skill sectors under the benchmark regulation could be obtained by setting the high-skill sector apprentice minimum wage to around 0.73.

Figure 9: Effect of Minimum Wages on Apprenticeship Training



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

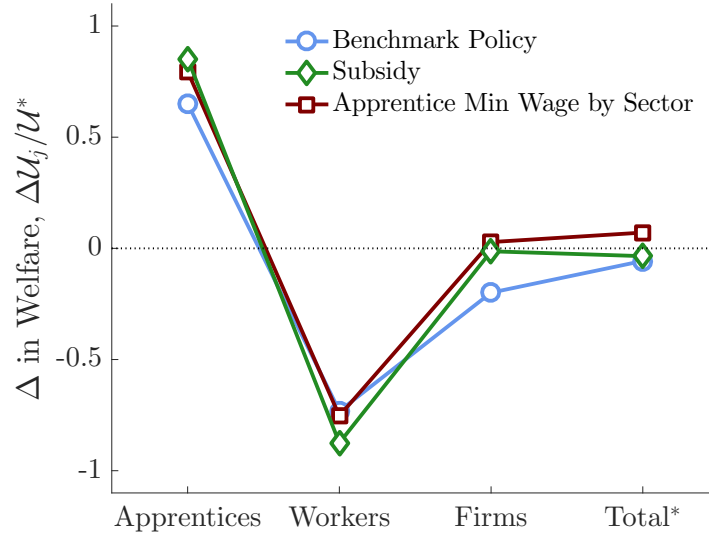
A potential caveat with setting a lower minimum wage is that apprentices may lack incentives to participate. The simulation does not take into account such supply side considerations for three reasons. First, there is excess supply of apprentices across all industries in Colombia (SENA, 2016). Second, the minimum wage levels we consider are close to the actual apprentice minimum wage in Colombia (see Section 2). Third, the minimum wage would have to be lower in high-skill sectors, where apprentices are likely to benefit more from training as future wages are higher. If apprentices value this, they should be willing to accept initially lower wages (Becker, 1964). Nevertheless, when decreasing the minimum wage even more, apprentices’ participation responses may become relevant.

Comparing Apprenticeship Policies

Finally, we compare the welfare effects of the counterfactual scenarios to the benchmark regulation in Figure 10. The overall welfare effect of the training subsidy is slightly above the benchmark regulation. This is mainly driven by larger gains of apprentices and a less negative effect on firms. Moreover, the figure shows that all groups of agents are better off under the sector-specific

minimum wages than under the benchmark regulation. In fact, this is the only policy under which firms gain and overall welfare improves already in partial equilibrium. We conclude that a policy taking into account heterogeneity across sectors, such as sector-specific apprentice minimum wages, can magnify the welfare gains of apprenticeship programs.⁴⁸

Figure 10: Welfare Effects of Different Apprenticeship Policies



Notes: The figure shows the change in welfare under different apprenticeship policies by group of agents, namely apprentices, incumbent workers and firm owners. The effects are scaled relative to no-regulation total aggregate utility, U^* . *Total** denotes the sum of welfare changes across the three groups plus government revenue.

7 Conclusion

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

These results have important policy implications. First, given the large training costs many

⁴⁸Recall that the main minimum wage counterfactual targets the same number of apprentices as the benchmark regulation *in each sector*. Appendix Figure E12 shows how aggregate outcomes and welfare would be affected by varying the fraction of apprentices trained in high-skill sectors via further adjustments to apprentice minimum wages, holding the total number of apprentices fixed. The figure suggests that further, albeit modest welfare improvements could be achieved when setting minimum wage levels such that the share of training of high-skill sectors increases.

firms face, policy interventions such as training subsidies would have to provide sizeable incentives to firms to achieve a broad increase in training. The Colombian regulation, featuring a mixture of apprentice quotas and lower minimum wages, does achieve a large increase in the number of apprentices, but it induces firm-size distortions and some reallocation towards low-skill sectors. Second, policies that take into account heterogeneity in training costs, such as sector-specific apprentice minimum wages, can deliver additional benefits. These lessons are likely to be relevant for other low- and middle-income countries suffering from labor market issues similar to Colombia in the early 2000s.

Some limitations of our analysis are worth highlighting. First, the lack of individual-level information in our main data limits the evidence we can gather on apprentices' outcomes. Future work could complement our study of firm responses and explore the impact of the regulation on employment and wage trajectories of young individuals. Second, while our structural approach yields estimates of total net training costs, gaining additional insights into the different components of training costs may be useful. A related avenue for future research could be to study the quality of training and how it may be correlated with training costs. Finally, our model abstracts from entry and exit of firms, and it may be worthwhile to investigate whether apprenticeship programs affect these firm dynamics, for instance by favoring firms with lower training costs.

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. (1999a). Beyond Becker: Training in Imperfect Labor Markets. *Economic Journal*, 109(453):112–142.
- Acemoglu, D. and Pischke, J.-S. (1999b). The Structure of Wages and Investment in General Training. *Journal of Political Economy*, 107(3):539–572.
- Akerberg, D. A., Caves, K., and Frazer, G. (2015). Identification Properties of Recent Production Function Estimators. *Econometrica*, 83(6):2411–2451.
- Alfonsi, L., Bandiera, O., Bassi, V., Burgess, R., Rasul, I., Sulaiman, M., and Vitali, A. (2020). Tackling Youth Unemployment: Evidence from a Labor Market Experiment in Uganda. *Econometrica*, 88(6):2369–2414.
- Attanasio, O., Guarín, A., Medina, C., and Meghir, C. (2017). Vocational Training for Disadvantaged Youth in Colombia: A Long-Term Follow-Up. *American Economic Journal: Applied Economics*, 9(2):131–143.
- Attanasio, O., Kugler, A., and Meghir, C. (2011). Subsidizing Vocational Training for Disadvantaged Youth in Colombia: Evidence from a Randomized Trial. *American Economic Journal:*

- Applied Economics*, 3(3):188–220.
- Becker, G. S. (1964). Human Capital Theory. *Columbia, New York*, 1964.
- Bernal, R. (2009). The Informal Labor Market in Colombia: Identification and Characterization. *Revista Desarrollo y Sociedad*, (63):145–208.
- Card, D., Ibarra, P., Regalia, F., Rosas-Shady, D., and Soares, Y. (2011). The Labor Market Impacts of Youth Training in the Dominican Republic. *Journal of Labor Economics*, 29(2):267–300.
- Card, D., Kluve, J., and Weber, A. (2018). What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations. *Journal of the European Economic Association*, 16(3):894–931.
- Chetty, R., Friedman, J. N., Olsen, T., and Pistaferri, L. (2011). Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records. *The Quarterly Journal of Economics*, 126(2):749–804.
- Crépon, B. and Premand, P. (2019). Direct and Indirect Effects of Subsidized Dual Apprenticeships. de Mel, S., McKenzie, D., and Woodruff, C. (2019). Labor Drops: Experimental Evidence on the Return to Additional Labor in Microenterprises. *American Economic Journal: Applied Economics*, 11(1):202–35.
- Del Boca, A. and Rota, P. (1998). How Much Does Hiring and Firing Cost? Survey Evidence from a Sample of Italian Firms. *Labour*, 12(3):427–449.
- Dostie, B. (2013). Estimating the Returns to Firm-Sponsored On-the-Job and Classroom Training. *Journal of Human Capital*, 7(2):161–189.
- Dustmann, C. and Schönberg, U. (2009). Training and Union Wages. *The Review of Economics and Statistics*, 91(2):363–376.
- Dustmann, C. and Schönberg, U. (2012). What Makes Firm-based Vocational Training Schemes Successful? The Role of Commitment. *American Economic Journal: Applied Economics*, 4(2):36–61.
- Fazio, M. V., Fernández-Coto, R., and Ripani, L. (2016). Apprenticeships for the XXI Century: A Model for Latin America and the Caribbean? *Working Paper*.
- Fersterer, J., Pischke, J.-S., and Winter-Ebmer, R. (2008). Returns to Apprenticeship Training in Austria: Evidence from Failed Firms. *Scandinavian Journal of Economics*, 110(4):733–753.
- Fudenberg, D. and Rayo, L. (2019). Training and Effort Dynamics in Apprenticeship. *American Economic Review*, 109(11):3780–3812.
- Galasso, E., Ravallion, M., and Salvia, A. (2004). Assisting the Transition from Workfare to Work: A Randomized Experiment. *Industrial and Labor Relations Review*, 58(1):128–142.
- Garicano, L. and Rayo, L. (2017). Relational Knowledge Transfers. *American Economic Review*, 107(9):2695–2730.
- Garicano, L. and Rossi-Hansberg, E. (2004). Inequality and the Organization of Knowledge. *American Economic Review*, 94(2):197–202.
- Garicano, L. and Rossi-Hansberg, E. (2006). Organization and Inequality in a Knowledge Economy.

- The Quarterly Journal of Economics*, 121(4):1383–1435.
- Gaviria, A. and Nuñez, J. A. (2003). Evaluating the Impact of SENA on Earnings and Employment. Technical report, Departamento Nacional de Planeación.
- Göggel, K. and Zwick, T. (2012). Heterogeneous Wage Effects of Apprenticeship Training. *The Scandinavian Journal of Economics*, 114(3):756–779.
- Groh, M., Krishnan, N., McKenzie, D., and Vishwanath, T. (2016). Do Wage Subsidies Provide a Stepping-Stone to Employment for Recent College Graduates? Evidence from a Randomized Experiment in Jordan. *The Review of Economics and Statistics*, 98(3):488–502.
- Hardy, M. and McCasland, J. (2020). Are Small Firms Labor Constrained? Experimental Evidence from Ghana. *Working Paper*.
- Hirshleifer, S., McKenzie, D., Almeida, R., and Ridao-Cano, C. (2014). The Impact of Vocational Training for the Unemployed: Experimental Evidence from Turkey. *The Economic Journal*, 126(597):2115–2146.
- ILO (2017). ILO Toolkit for Quality Apprenticeships, Volume 1: Guide for Policy Makers.
- Jones, D. C., Kalmi, P., and Kauhanen, A. (2012). The Effects of General and Firm-Specific Training on Wages and Performance: Evidence from Banking. *Oxford Economic Papers*, 64(1):151–175.
- Kleven, H. J. (2016). Bunching. *Annual Review of Economics*, 8:435–464.
- Konings, J. and Vanormelingen, S. (2015). The Impact of Training on Productivity and Wages: Firm-Level Evidence. *Review of Economics and Statistics*, 97(2):485–497.
- Kremer, M. (1993). The O-Ring Theory of Economic Development. *The Quarterly Journal of Economics*, 108(3):551–575.
- Krueger, A. B. and Pischke, J.-S. (1995). A Comparative Analysis of East and West German Labor Markets: Before and After Unification. In *Differences and Changes in Wage Structures*, pages 405–446. University of Chicago Press.
- Kuczera, M. (2017). Striking the Right Balance: Costs and Benefits of Apprenticeship. *Working Paper*.
- Kugler, A., Kugler, M., Saavedra, J., and Prada, L. O. H. (2019). Long-Term Educational Consequences of Vocational Training in Colombia: Impacts on Young Trainees and Their Relatives. *NBER working paper no. 21607*.
- Levinsohn, J. and Petrin, A. (2003). Estimating Production Functions Using Inputs to Control for Unobservables. *The Review of Economic Studies*, 70(2):317–341.
- McKenzie, D. (2017). How Effective Are Active Labor Market Policies in Developing Countries? A Critical Review of Recent Evidence. Policy Research Working Paper 8011, World Bank Group, Washington, D.C.
- Mohrenweiser, J. and Zwick, T. (2009). Why Do Firms Train Apprentices? The Net Cost Puzzle Reconsidered. *Labour Economics*, 16(6):631–637.
- Muehlemann, S., Pfeifer, H., Walden, G., Wenzelmann, F., and Wolter, S. C. (2010). The financing

of apprenticeship training in the light of labor market regulations. *Labour Economics*, 17(5):799–809.

Ospino, C. (2016). *The Effects of Apprenticeship Contracts on Labor Demand: Theory and Empirical Applications for Colombia*. PhD thesis, Universidad de los Andes.

Ospino, C. (2018). The Effects of Being Subject to the Colombian Apprenticeship Contract on Manufacturing Firm Performance. *Documentos de Trabajo del CEDLAS*.

Saez, E. (2010). Do Taxpayers Bunch at Kink Points? *American Economic Journal: Economic Policy*, 2(3):180–212.

SENA (2016). Data Report on Vocational Education and Training – Information and Analysis that Contribute to the Development of Colombian Vocational Education and Training.

Slemrod, J. (2016). Caveats to the Research Use of Tax-Return Administrative Data. *National Tax Journal*, 69(4):1003–1020.

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.