MANDATING TRAINING IN FIRMS: THEORY AND EVIDENCE FROM THE
COLOMBIAN APPRENTICESHIP PROGRAM *

SANTIAGO CAICEDO, MIGUEL ESPINOSA AND ARTHUR SEIBOLD

July 22, 2019

Preliminary and Incomplete (Please do not circulate)

Abstract

We study the effect of apprenticeship programs on firms and welfare, using novel administrative data on the universe Colombian manufacturing firms with at least ten workers, and a unique reform to apprenticeship regulation. The reform simultaneously establishes apprentice quotas that vary discontinuously in firm size, lowers apprentices’ wages and includes the possibility of paying a fee instead of hiring the required apprentices. We begin by documenting that the policy is successful in increasing the number of trained apprentices more than ten times the pre-reform levels. However, the reform also induces significant firm size distortions driven by heterogeneous firm responses. In sectors with high skill requirements, firms avoided hiring apprentices, decreasing their size, and bunching just below the regulatory thresholds. In contrast, firms in low-skill sectors increase their size and bunch at the regulatory thresholds to access more apprentices. We develop a simple theoretical model featuring heterogeneous training costs across sectors to rationalize and quantify these empirical findings. We find that despite the substantial firm-size distortions, there are potentially large welfare gains stemming from general equilibrium and dynamic effects; as training apprentices expands the supply of productive workers, lowering wages and increasing aggregate output.

Key words: Training, Apprentices, Firm-size distortions

JEL Codes: E24, J21, J24.

* Santiago Caicedo, University of Chicago and Universidad de los Andes, scaiceso@gmail.com, scaicedosoler@uchicago.edu; Miguel Espinosa, Universitat Pompeu Fabra, miguel.espinosa@upf.edu; Arthur Seibold, University of Mannheim, seibold@uni-mannheim.de. We would like to thank Ufuk Akcigit, Benito Arrunada, Leonardo Bursztyn, Jonathan Dingel, Rosa Ferrer, Luis Garicano, Alessandro Gavazza, Maria Guadalupe, Ali Hortacsu, Camille Landais, Bob Lucas, Sandra McNally, Joan Monras, Massimo Motta, Luis Rayo, Mar Reguant, Esteban Rossi-Hansberg, Sandro Shelegia, Alexandra Spitz-Oener, Claudia Steinwender, Nancy Stokey, John Sutton and participants of the UPF, University of Mannheim and Universidad de los Andes seminars for useful comments. All errors are ours.
1 Introduction

Many countries have adopted apprenticeship programs that combine both training in firms with formal education.¹ These programs can improve labor market outcomes for young, inexperienced workers; particularly those with a non-college education (Zimmermann et al. 2013). Firms play a crucial role in these apprenticeships, as they provide training and knowledge transfer from workers to apprentices. Absent external intervention, firms may lack the incentives to provide efficient levels of training. For instance, restrictions in the labor market, such as minimum wages, may dissuade them from training individuals with low initial productivity.² This lack of training then negatively affects workers labor market opportunities throughout their lives.

In response to these concerns, governments around the world have set up apprenticeship programs. For instance, Germany, Austria, and Switzerland have a long-standing tradition of vocational training that cemented successful apprenticeships (Steedman 2012). In Germany, two-thirds of firms and one-third of workers participate in the apprenticeship system. However, these programs are not limited to continental Europe. A growing list of developed and developing countries, including Australia, France, Sweden, Norway, Canada, US, UK, India, Turkey, Mexico, Brazil, Colombia, Peru, and Chile have active apprenticeships programs.³ Surprisingly, despite their ubiquity, there is limited evidence on the general effect of these programs. The scarce literature on the topic mainly studies the returns of training and benefits of apprenticeships to workers (Fersterer et al. 2008, Göggel and Zwick 2012). However, there is almost no literature on the impact on firms’ decisions and on aggregate welfare.⁴

In this paper, we aim at filling this gap by studying the effects of apprenticeship programs on firms decisions and their effect on aggregate welfare. To do so, we use a novel administrative data set on the universe Colombian manufacturing firms with at least ten employees, matched with rich firm-level census data between 1995 and 2009. We exploit a unique change in apprentice regulation that mandates firms to hire apprentices, imposing apprentice quotas changing discontinuously with firm size.⁵ These quotas include both minimum and maximum levels. At the same time, the regulation reduces the minimum wage for apprentices, encouraging firms to hire apprentices. In addition, the policy allows to “buy themselves out” of training apprentices by paying a fee larger than the apprentice’s wage. We study firms responses to this policy and develop a model to quantify the effect

¹We follow Wolter and Ryan (2011) definition of apprenticeships as: “programs that comprise both work-based training and formal education, in most countries at upper-secondary level, and lead to a qualification in an intermediate skill, not just to semi-skilled labor”.

²There are additional reasons for firms to provide inefficient levels training. There is a well-established theoretical literature exploring training by firms, starting with Becker (1964) who argues firms have the incentive to provide general training as long as the worker pays for it with lower wages. Acemoglu and Pischke (1998, 1999) study inefficient training under imperfect information and imperfect competition. More recently, Garicano and Rayo (2017) and Fudenberg and Rayo (2017) show in a dynamic framework that experts have the incentive to transfer knowledge inefficiently slow to apprentices. We discuss this literature below.

³For detailed comparisons of apprenticeship countries in G20 countries see Steedman (2010, 2012). For Latin American countries see Fazio et al. (2016).

⁴Mohrenweiser and Zwick (2009) and Ospino (2018) study the effect of apprenticeship programs on firm productivity, for Germany and Colombia respectively. There is also a related literature on the effects of work training on firm productivity and wages, see for instance Dearden et al. (2006).

⁵The firm size considered by the regulation is the total number of full time permanent workers. It does not include apprentices, temporary or outsourced workers.
of apprenticeship programs. To do so, we organize the paper in three parts. First, we show reduced-form evidence on the effect of the reform on firms’ outcomes and document strong heterogeneity in these responses across sectors. Second, we develop a model that rationalizes these responses and provides a framework to compute aggregate welfare effects. Third, we estimate the model to quantify these effects, decompose the different elements of the policy and study counterfactual apprenticeship regulations.

In our reduced-form evidence, we begin by showing that the reform is successful in increasing training as expected, with the average number of apprentices per worker increasing more than tenfold. This sharp rise is accompanied by sizeable changes in firms’ labor decisions. We show that these responses are strongly heterogeneous across firms in different sectors: firms in sectors with a large fraction of highly skilled workers (henceforth high-skill sector) avoid training apprentices, while firms in sectors with a low fraction of highly-skilled workers (henceforth low-skill sector) seek as many apprentices as possible. We organize these responses into three empirical facts. First, we use bunching methods to gauge firm size responses to the discontinuities in apprenticeship quotas. We find that firms in high-skill sectors reduce their size in order locate just below the regulation thresholds where they have to hire fewer apprentices, leaving missing mass just above the thresholds. On the contrary, firms in low-skill sectors bunch just above the regulation thresholds to increase the number of apprentices they can hire. Second, we show that conditional on their post-reform size, firms in high-skill sectors tend to hire the minimum number of apprentices required, while most firms in low-skill sectors hire the maximum number of apprentices possible. Third, many firms in high-high-skill sectors do not even satisfy the minimum apprentice quota, but instead pay fees to the government as a “buy-out” from the regulation. In contrast, this behavior is virtually nonexistent among firms in low-high-skill sectors. In summary, our results imply highly heterogeneous responses across firms in different sectors; there seems to be a net benefit to training apprentices for firms in low-skill sectors, but a net cost of hiring apprentices in high-skill sectors.

In the second part of the paper, we develop a simple model featuring heterogeneous training costs to rationalize these results, quantify the effect of the policy on welfare, and analyze counterfactual policies. Inspired by Lucas (1978), we consider an economy with heterogeneous firms characterized by their managerial ability. Firms produce a homogenous good using labor from workers and apprentices. Apprentices differ from workers in that they require training for production. Their productivity depends on the amount of training they get from workers. Training apprentices is costly to the firm as it requires the workers’ time. The net cost of training apprentices is a combination of the time it takes a worker to train them and the apprentice productivity after training. We consider differences in these net costs of training apprentices to capture the varied firm responses across sectors. On average, training apprentices in high-skill sectors is more costly than in low-skill sectors, parallel to the firm responses documented in our reduced-form exercises.

Using this theoretical framework, we study the Colombian apprentice regulation and show that firms’ optimal response to the policy is consistent with the three empirical facts. Concretely, we show that whenever apprentice wages are sufficiently low, firms in low-skill sectors those with low training costs - hire as many apprentices as possible. To do so, some of these firms increase their

---

6This could be interpreted as actual training time by workers or the additional time it takes for them to ‘fix’ mistakes made by apprentices. Although the formulation is equivalent, generally we will use the latter interpretation in our discussion and description below.
size to get more apprentices, inducing bunching at the thresholds and missing mass of firms to the left of this regulatory cutoffs. These firms also choose the maximum number of apprentices and consistent with the empirical results, they choose to not opt out of training apprentices by paying the fee. These effects are in sharp contrast with firms in high-skill sectors, where training is more time-consuming and/or apprentices are relatively less productive. We show that firms in these sectors avoid hiring apprentices by getting close to the minimum quota of the regulation or bunching just below the regulatory thresholds. If the fee is low enough, they even prefer to pay the fee instead of having a suboptimal number of workers and apprentices. Our theory predicts that if wages are fixed, the average firm in the low-skill sector benefits from the policy, increasing output and profits. In contrast, the reform seems to bring more costs than benefits to the average firm in the high-skill sector.

To quantify the total welfare consequences, we consider the general equilibrium and dynamic effects of the policy. We close the model, starting with a fixed supply of workers and apprentices. We suppose workers don’t move across sectors to allow for wages to be different across sectors. If an apprentice is trained in a sector, this increases the next period’s labor supply of that same sector. Wages are determined by the usual market clearing conditions in each sector. Importantly, we include in our analysis wage restrictions for workers and apprentices. Before the reform, wages for both workers and apprentices must be above the minimum wage. The reform lowers the minimum wage for apprentices below the workers’ minimum wage. If before the reform, the equilibrium wage for apprentices was below the minimum wage, some potential apprentices cannot be trained. At the minimum wage, firms are not adequately compensated for training apprentices and therefore choose to train less than the socially optimal number. The policy partially offsets these inefficiencies by lowering apprentices’ wages to induce firms to train more apprentices. There is also a positive dynamic effect on future aggregate output, as newly trained apprentices increase the supply workers in subsequent periods. With the model, we can quantify such general equilibrium and dynamic effects of the policy. We show that these effects are positive and significant in terms of aggregate output and profits. We conclude there are potential welfare gains from the policy despite the labor distortions it induces.

We estimate the model using our firm-level microdata. We target the moments in the data that correspond to the reduced-form results for each sector: the firm-size distribution before the reform, the bunching and missing mass after the policy, the fraction of firms choosing the upper and lower bound of the apprentices quotas, and the fraction of firms that pay the fee. First, we infer the parameters of the productivity distribution and the production function by matching the pre-reform firm size distribution and the elasticity of labor in each sector. Then, we fit the remaining parameters of the model, targeting the moments relative to the firm responses, and using the policy specifications. From this estimation procedure, we recover the training cost distribution by sector, finding that firms in the low-skill sectors have on average less training costs but higher variance. Lower mean implies that firms in the low-skill sectors want apprentices. A higher variance that their responses to the policy are more varied than for firms in the high-skill sectors, as observed in the data.

Next, we use the estimated model to quantify the welfare effects of the policy. Concretely, we

\footnote{In the model sector could choose not to train apprentices at all, either because training time is too high, or the apprentices’ initial productivity is too low.}
study the consequences of the policy on wages, aggregate output and profits. We divide our welfare analysis looking first at the partial equilibrium effects policy, then allowing for worker wages to adjust in general equilibrium, and finally, quantify the dynamic effect when apprentices increase the supply of workers in succeeding periods. For fixed wages, we find quantitatively small effects on output and profits. Intuitively, firms optimize and substitute labor from workers to apprentices. At the margin, although there is a significant number of displaced workers, output and profits are relatively stable. The aggregate production of low-skill sector firms increases by around 0.43%, almost canceling out the decrease in the total production for high-skill sector firms, -0.37%. Profits move even less. In contrast, there are substantial general equilibrium and dynamic effects on output and profits. Displaced workers absorbed by the labor market drive down wages by 2%. This decrease in wages benefits all firms, increasing the aggregate production in the low-skill sectors by 4%, while high-skill firms also benefit from the wage cut reverting the initial negative effects of the apprenticeship policy. Moreover, these positive effects on aggregate output and profits are amplified when trained apprentices increase the future supply of workers. These dynamic effects lead to a total increase of 5.7% of the aggregate production, 8.7% in the low-skill sectors, and 3.2% in the high-skill.

We complement these quantitative exercises by showing reduce-form evidence on the effects of the policy around the regulatory thresholds. We use a Difference-in-Difference type methodology for firms whose size is within 5 workers around these thresholds. We measure firm size as the number of permanent workers in 2002, the last year before the policy. We find that as predicted by our model, firms above the threshold have relatively more apprentices and less permanent workers. Also the effects on output and profits are small, negative and not significant, coinciding with our quantitative results.

Finally, we use the estimated model to decompose the effects of the policy into its three principal components. The minimum quota guarantees that at least some apprentices are trained in every sector. The maximum that firms don’t substitute too many workers for apprentices. Lower apprentice wages incentive firms to train these apprentices, increasing training mostly in the low-skill sectors, but also in the high-skill sectors. Fees reduce the negative effect on high-skill firms that try to avoid apprentices, but naturally decreases the number of apprentices trained in those sectors. Despite the negative distortions of the policy in terms of firm size and subsidies across sectors, we find positive impacts of aggregate output and profits. We plan to study alternative policies that deliver similar benefits without creating inefficient firm size distortions.

**Related Literature:** There is a rich theoretical literature examining training and human capital accumulation in firms. In his seminal work on human capital, Becker (1964) shows that contrary to conventional wisdom of the time, firms could still provide general human capital training to workers given these workers are willing to pay for it. Similarly, in our framework, if there are no minimum wage restrictions, wages would adjust to compensate for the training apprentices get. In this case, firms would provide socially optimal training. Of course, imposing a binding minimum wage prevents this from happening. Using our data, we quantify the welfare consequences of relaxing the minimum wage restriction for apprentices, as in the change in policy we study.

Firms might still provide socially inefficient training absent wage restrictions. Acemoglu and Pischke (1998, 1999) show that under imperfect information and imperfect competition, firms have incentives to provide some general training; however, training might still be suboptimal. More re-
recently, Garicano and Rayo (2017) show in a dynamic framework, that workers with skills have the incentive to transfer knowledge inefficiently slow to apprentices. Moreover, Fudenberg and Rayo (2017) argue that apprentices could end up doing menial tasks and exerting inefficiently high effort on those tasks. As a whole, this literature shows firms training might be insufficient so policies increasing training could be desirable. Our paper highlights that a well-intended apprenticeship policy that increases training can distort the labor market, affecting aggregate output and the composition of training. Using our rich microdata, we can quantify these effects over welfare-relevant variables.

Some part of the literature has focused on explaining quality differences in apprenticeship training. For instance, Soskice (1994) shows that larger firms pay more and therefore, can attract better apprentices. This discussion is relevant in our context because if differences in quality training exist between firms of high-skill and low-skill sectors, these differences can explain some of our results. For instance, the quality of training provided by high-skill sectors is lower, and therefore, students enrolled in apprentice systems choose to only apply to low-skill sectors. To exclude this possibility, we build on Göggel and Zwick (2012) who show that differences in apprenticeship quality are usually small between firms that differ by size, sector, and location.

Fersterer et al. (2008) study the effect of labor unions on training in the German economy. Unionized firms provide better wage contracts to their employees by providing higher wage floors, incentive their apprentices to perform better with the promise to become one of these employees. We provide empirical evidence that alternative channels to wage compression can also increase training in the economy. Most of the empirical literature on apprentices such as Krueger and Pischke (1995), Fersterer and Winter-Ebmer (2003) and Fersterer et al. (2008) focus on estimating the returns of apprenticeship programs concluding that the returns are similar to those of other types of schooling. We complement their results by studying the effects of apprentice programs on firms and not on workers.

Ospino (2018) is the only other paper that evaluates the change in the Colombian apprentice regulation. The author uses census data to study the effect of this policy over the performance of firms. Using a regression discontinuity approach, Ospino shows positive effects on output per worker, total factor productivity and share of exported sales, and negative effects on the wage bill of regular workers. We depart from Ospino’s paper in three important ways. First, all the empirical facts we document are new since our paper uses administrative data collected directly by Colombian National Service of Vocational Education (SENA) to implement the apprentice policy. With these data we are able to capture exactly the number of workers used by SENA to implement the policy, an the apprentices each firm train before the policy was implemented. Second, to understand the effects of the policy on the firm’s performance, we show it is vital to consider the heterogeneity in the responses of firms in high-skill and low-skill sectors. Third, given we develop a structural model, we can quantify the welfare consequences and analyze counterfactual policies.

Finally, our paper is related to the literature that analyze the effect of firm size distortions on the economy (Besley and Burgess 2004, Guner et al. 2008, Garicano et al. 2016). Garicano, Lelarge, and Van Reenen (2016) study how firm size policies in France affect labor allocation and the productivity distribution. They show how employment protection laws imposed on firms with at least 50 employees, implied sizeable welfare losses as high as 5% of GDP. Similar to their paper, we analyze a size-dependent regulation and find substantial labor and production responses
of firms. However, the apprentice policy we study, highlights a different mechanism to that of previous studies. Apprentices, in contrast to taxes and other labor regulations, are part of the firm’s productive inputs. We show how this difference is vital given not all firms change the labor decision to avoid the policy. Firms in low-skill sectors, where training costs are relatively small, benefit directly from the policy increasing their size and hiring more apprentices. These unintended consequences on the composition of trained apprentices enrich the welfare considerations for the case of these apprenticeship programs.

We organize the rest of the paper as follows. In the next section, we describe the data and introduce relevant institutional context. Then we present the reduced-form evidence, highlighting the three aforementioned empirical facts of the behavioral response of firms after the apprentice regulation. Next, we present a theoretical framework that rationalizes these empirical findings and provides a basis to quantify the welfare effects of the policy and study counterfactuals. Then, in the quantitative analysis, we estimate the model, recover the training costs distribution, quantify the welfare effects of the policy and study alternative counterfactual apprentice regulation. Finally, we conclude highlighting missing pieces and future work.

2 Data and Institutional Context

We use a novel administrative data set collected by the Colombian National Service of Vocational Education (SENA). SENA was created in 1957 with the objective of providing technical and vocational training. It currently provides more than 60% of the total technical and vocational education in the Colombia, with more than 400 thousand enrolled students annually (SENA 2018). SENA also supervises the compliance of the apprenticeship regulation and oversees the vocational education programs offered by other institutions.

Our administrative data covers the universe of manufacturing firms with at least 10 workers from 1995 to 2009.8 There is a total of 108,385 firm-year observations, averaging approximately 7000 firms per year. For each firm-year observation, the data includes the number of workers, the number of apprentices, and indicators for fees and fines paid by the firms. We match this data set to census data from the Colombian manufacturing survey (EAM). This is rich firm-level data collected by the National Department of Statistics (DANE).9 It includes additional information on workers divided into three production/skill layers,10 detailed production inputs, wages, sales, output, and production costs.

8SENA collects the information for all firms in the economy. We only use manufacturing sector data, as it is the only sector available in the census for years prior to the regulation.
9The data is collected by DANE at the establishment level and then aggregated at the firm level.
10The census data divide workers into: professional workers who are workers with tertiary education working on the production, production workers are the ones with less than tertiary education working on production and administrative workers which do not work directly in the production processes.
Apprentice Regulation

Before the change of regulation in 2003, an apprentice contract was a type of labor contract where apprentices provided a service to firms in exchange of professional formation on a given occupation. In paper, Law 188 of 1959 stated the initial wage of the apprentice could not be lower than 50% of the minimum wage and had to increase during the apprenticeship until it reached at least one full minimum wage. The apprentice was responsible of attending courses facilitated by the employer, while the employer was expected to provide adequate professional formation to the apprentice. The law also stipulated that employers should favor apprentices in case there were any open vacancies at the firms. Decree 2838 of 1960 established that firms with more than 20 workers, or capital above $100,000 pesos of the time, should hire apprentices that could not exceed 5% of the firm’s total labor force. The regulation did not establish a minimum number of apprentices. Decree 122 of 1999 updated the policy and specified SENA was responsible of determining the apprenticeship contract conditions including: the apprentice quota, the relevant occupations, characteristics of the training program and the sanctions in case of no compliance. However, in practice, the apprentice policy was weakly enforced. There were no clear consequences of not complying the regulation. As we show below, before 2003 hardly any firm had apprentices. The most prevalent way of complying was signing regular workers to night courses taught by SENA (Ospino 2018).

In December 2002 the Colombian congress approved a major labor reform which included a redefinition of the apprenticeship contract. The reform came into effect in June 2003. It defined two phases of the apprenticeships program: the teaching phase where apprentices are taught full time in a formal education institution, and the productive phase where they get training and on-the-job experience working at a firm. The teaching phase duration spans from one year (1760 hours) to 1.5 year (2640 hours) depending on the occupation chosen by the apprentice. The productive phase lasts at least 6 months and has a maximum duration of 2 years. Potential apprentices are individuals 14 years or older that have ‘semi-qualified’ formation. SENA oversees the apprentice regulation, gathering the relevant information on firms and apprentices, and supervising other vocational education providers. It also manages a centralized matching system where firms can hire apprentices.

All firms with more than 10 workers report twice a year the total number of hours worked by directly hired workers. To implement the regulation SENA computes the number of permanent workers as the equivalent full-time workers in each firm using the reported total hours per worker. They compare this information to monthly payroll tax data from the revenues SENA receives directly

---

11 For details on this Law see http://normograma.sena.edu.co/normograma/docs/ley_0188_1959.htm.
12 $100,000 pesos in 1960 is equivalent to around $15,000 2018 US dollars. Decree 266 of 2000 updated these cutoffs to firms with more than 10 workers, or capital above 200 monthly minimum wages.
13 This labor reform had other major changes in the labor regulation and was motivated in part to reduce the high levels of informality in the Colombian labor markets. None of the other policies depended on firm size. For more information on the labor reform see Law 789 of 2002 at http://normograma.sena.edu.co/normograma/docs/ley_0789_2002.htm
14 SENA classify the occupations into two categories, technician and technologists. Technician occupations usually require basic secondary education, while technologist occupations require finishing secondary education (SENA 2018).
15 Indirectly hired workers, such as some temporary or outsourced workers don’t count as permanent workers in the implementation of the regulation.
from each firm. Using the data on permanent workers, SENA determines the apprentice quota applicable to each firm. After receiving the notification on the apprentice quota from SENA, firms have 2 months to comply with the apprentice regulation. Additionally firms are obliged to report any changes in their workforce that would modify their apprentice quota. Failure to do so is subject to hefty fines, also enforced by SENA.

The apprentice regulation after 2003 has three important components. First, it establishes apprentice quotas depending on the number of permanent workers of a firm. These quotas set a minimum number of required apprentices: at least one apprentice for every 20 workers, with one more for each fraction of 10 workers. For firms with workers between 15 and 20 need to have at least one apprentice. So firms with 15 to 29 permanent workers, require at least one apprentice, ones with 30 to 49 at least 2, firms with 50 to 69 at least 3, and so on. The quota also sets an upper bound for apprentices. The maximum number of apprentices permitted is at most twice the minimum bound. Firms with less than 15 workers could have at most one apprentice. Second, to incentive firms to hire apprentices the regulation stipulated that apprentices could be paid at least 50% of minimum wage during the teaching phase and 75% of the minimum wage during the productive phase. Third, firms could pay a fee instead of hiring the minimum required apprentices. This fee is proportional to the difference between the minimum required apprentices and the ones hired by the firm. Whenever firms don’t comply with the apprentice regulation, SENA imposes a fine equivalent to two minimum wages per missing apprentice, plus moratory interest.

3 Number of Apprentices and Firm Size Distribution

The primary objective of the apprentice regulation was to increase the number of apprentices trained by firms. Figure 1 shows the policy was successful in this dimension, substantially increasing the number of trained apprentices per permanent worker from 3 every 1000 permanent workers before 2003, to around 50 every 1000 permanent workers after 2003. The total number of apprentices in the manufacturing sector increased from on average 1297 before the reform (1995-2002) to an average of 15,201 after the reform (2003-2009).

---

16 This tax corresponds to 2% of a firm’s payroll.
17 This upper bound was modified in 2009. Firms between 1-14 workers could have at most 50% of total workers, 15-50 workers up to 40% of total workers, 51 to 200 workers until 30% of total workers and more than 200 workers until 20% workers (see Decree 1779 of 2009 available in http://normograma.sena.edu.co/normograma/docs/decreto_1779_2009.htm).
18 Apprentices that get education from a university, as opposed to vocational education institutions, are paid a full minimum wage in both phases. The regulation also specified that if the unemployment rate falls below 10%, then all apprentices have to be paid a full minimum wage. This only happened after 2013.
However, the regulation also induced firm size distortions by setting apprentice quotas conditional on the number of permanent workers. Figure 2 shows these distortions were sizeable and affected the reported firm size distribution after 2002.\textsuperscript{19} Policy thresholds are represented by the vertical dotted lines. This Figure shows the size distribution of permanent workers before the policy (1995-2002) was relatively smooth. In contrast, after the regulation the firm size distribution is rugged, with a mass of firms ‘bunching’ before the regulatory thresholds. The notches of ‘missing mass’ around these thresholds suggests some firms changed their size as a response to the apprentice regulation. The fact that there is missing mass both below and above the threshold indicates heterogeneous responses of firms. The missing mass above the threshold together with the bunching right below the threshold, points to some firms avoiding apprentices by reducing the number of workers they hire. Conversely, the missing mass below the thresholds, suggests some firms increased their workers to get more apprentices.

\textsuperscript{19}In Appendix A we present the distribution every two years for the period before and after the policy. These year by year distributions have the same behavior as the pooled distributions in Figure 2.
To study these heterogeneous effects we first divide the manufacturing firms into nine two-digit sectors of the Colombian industrial classification. In Figure 3 we plot the number apprentices per permanent worker for each of these sectors. There is a stark division in the apprentice uptake for the different manufacturing sectors. In Wood Products, Textiles, Food and Beverage, Mineral Non-Metallic Products sectors there is between 80 to 100 apprentices for every 1000 permanent workers. In contrast, the Paper and Editorial, Other Manufacturing, Machinery and Equipment, Metallic Products, Chemical Products sectors, there are around 20 apprentices every 1000 permanent workers.

---

We use DANE’s industrial classification, CIIU 3 A.C, which is adapted from the International Standard Industrial Classification (ISIC) the industry classification system of the United Nations. This classification is closely related to the US Standard Industry Classification.
To understand these differences in the apprentices uptake across sectors, we rank industries by the fraction of professional workers (with tertiary education) over total workers. We interpret this fraction as a proxy for the skill requirements in each sector. We argue this measure is related to the net costs of hiring apprentices. Sectors with more complex technologies that solve difficult tasks, probably have a larger fraction of professional workers. It might be more costly to train apprentices in these sectors. Whereas, sectors with routine manual tasks will have a smaller fraction of professional workers, and apprentices can learn faster the skills required. We divide the manufacturing sectors in two groups using the median. We denote low-skill sectors those below the median and the high-skill sectors the ones above. Using this definition, the four sectors that take more apprentices (Wood Products, Textiles, Food and Beverage, Mineral Non-Metallic Products) are classified as low-skill and five sectors that take fewer apprentices as high-skill (Paper and Editorial, Other Manufacturing, Machinery and Equipment, Metallic Products, Chemical Products). This split is relatively balanced. Low-skill industries represent 50% of the firms in the pre-reform sample (44% in the post-reform period). Table 1 shows the summary statistics for the fraction of professionals by sector as well as for three additional variables related to the net costs of training apprentices: the total average wages for all workers, the average wages for production workers and the fraction of permanent workers. We find low-skill sectors have lower average wages for total workers and production workers, and a lower fraction of permanent workers. Moreover, the ranking of these

21 In Appendix B we plot the fraction of professional workers for each of the ranked sectors and show this way of splitting the data captures a natural break-point between sectors.
sectors across these measures is relatively stable, particularly for sectors with more observations.

Table 1: Low-skill and High-skill Sectors Pre-Reform

<table>
<thead>
<tr>
<th></th>
<th>Observations</th>
<th>% Prof</th>
<th>Wages</th>
<th>Wages</th>
<th>% Perm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>All Workers</td>
<td>Prod. Workers</td>
<td></td>
</tr>
<tr>
<td><strong>Low-skill</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wood</td>
<td>2,543</td>
<td>3.53</td>
<td>5,218</td>
<td>4,660</td>
<td>83.61</td>
</tr>
<tr>
<td>(8.00)</td>
<td>(3.427)</td>
<td>(2,520)</td>
<td>(25.08)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Textiles</td>
<td>12,354</td>
<td>4.92</td>
<td>5,602</td>
<td>5,469</td>
<td>73.10</td>
</tr>
<tr>
<td>(12.47)</td>
<td>(3,099)</td>
<td>(6,812)</td>
<td>(36.35)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Food and Beverages</td>
<td>11,159</td>
<td>5.28</td>
<td>7,223</td>
<td>6,563</td>
<td>81.37</td>
</tr>
<tr>
<td>(10.05)</td>
<td>(5,173)</td>
<td>(5,325)</td>
<td>(25.84)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mineral Non-Metallic</td>
<td>2,689</td>
<td>7.01</td>
<td>8,486</td>
<td>7,416</td>
<td>82.77</td>
</tr>
<tr>
<td>(11.52)</td>
<td>(7,151)</td>
<td>(5,726)</td>
<td>(26.63)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total Low-skill</strong></td>
<td>28,745</td>
<td>5.22</td>
<td>6,467</td>
<td>6,037</td>
<td>78.09</td>
</tr>
<tr>
<td>(11.14)</td>
<td>(4,626)</td>
<td>(5,869)</td>
<td>(31.12)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>High-skill</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other Manufacturing</td>
<td>2,543</td>
<td>7.25</td>
<td>8,034</td>
<td>7,662</td>
<td>76.76</td>
</tr>
<tr>
<td>(13.62)</td>
<td>(6,522)</td>
<td>(7,966)</td>
<td>(30.98)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paper and Editorial</td>
<td>12,354</td>
<td>7.36</td>
<td>8,443</td>
<td>7,487</td>
<td>86.72</td>
</tr>
<tr>
<td>(13.69)</td>
<td>(5,664)</td>
<td>(5,080)</td>
<td>(22.32)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metallic</td>
<td>11,159</td>
<td>7.45</td>
<td>10,001</td>
<td>8,845</td>
<td>77.19</td>
</tr>
<tr>
<td>(10.45)</td>
<td>(6,581)</td>
<td>(5,929)</td>
<td>(29.21)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Machinery</td>
<td>2,689</td>
<td>8.42</td>
<td>7,633</td>
<td>6,452</td>
<td>83.03</td>
</tr>
<tr>
<td>(13.57)</td>
<td>(4,800)</td>
<td>(4,047)</td>
<td>(25.65)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemical</td>
<td>5,118</td>
<td>8.65</td>
<td>9,968</td>
<td>7,796</td>
<td>84.56</td>
</tr>
<tr>
<td>(11.78)</td>
<td>(7,875)</td>
<td>(6,607)</td>
<td>(24.04)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total High-skill</strong></td>
<td>28,949</td>
<td>8.05</td>
<td>8,542</td>
<td>7,267</td>
<td>82.77</td>
</tr>
<tr>
<td>(13.05)</td>
<td>(6,339)</td>
<td>(5,870)</td>
<td>(26.07)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Data divided into two digit manufacturing industries using the CIIU Revision 3 codes from DANE. Low-skill and high-skill sectors are defined according to ranking of fraction of profession workers out of the total workforce. Source: SENA.
4 Reduced-Form Evidence

In this section, we provide reduced-form evidence of the heterogenous firm responses to the apprentice regulation across low-skill and high-skill sectors. We show that firms in high-skill sectors avoid training apprentices, while firms in low-skill sectors train as many apprentices as possible. We organize this evidence in three facts, first studying the firm size distribution by sector, then, the apprentices uptake by firm size, and finally the fraction of firms that pay the fee in each sector.

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

Figure 4 shows the firm size distribution for firms in high-skill sectors (Panel 4a) and in low-skill sector (Panel 4b). From this figure, it is clear that firms in the high-skill sectors bunch below the threshold, while firms in the low-skill sector bunch at the threshold. More formally we follow Saez (2010) to compute the bunching estimates. For each firm size distribution, we fit a 7th-degree polynomial excluding the bunching region to construct a smooth counterfactual distribution (solid line in plots). We then compute the excess mass \( b = B/h_o(\hat{n}) \) as the count of firms \( B \) at the bunching point \( \hat{n} \) over the counterfactual count \( h_o(\hat{n}) \) at that point. This excess mass indicates the marginal bunching firm in the high-skill sectors had around 2 more workers, \( b \approx \Delta n \approx 2 \). These estimates are relatively stable across thresholds (Panel 4a). In contrast, for low-skill sectors, the marginal bunching firm has between 0.56 and 1.79 fewer workers depending on the threshold (Panel 4b). Similarly, we can compute the missing mass as the ‘hole’ in the observed distribution relative to the counterfactual, \( m = M/h_o(\hat{n}) \). We compute this missing mass for a size bin of three workers. Note the missing mass has similar value to the bunching estimate suggesting those where the firms that bunched. Both the excess mass and the missing mass are statistically significant. Standard errors are presented in the plots in parenthesis and are computed by bootstrapping the firm size distribution 500 times.
Figure 4: Fact 1 - Bunching of Firms in High-skill and Low-skill Sectors

(a) High-skill Sectors

(b) Low-skill Sectors

Source: SENA.
Fact 2: Firms in high-skill sectors choose the lower bound of the regulation, firms in low-skill sectors choose the upper bound.

Figure 5 shows the lower and upper bounds of the apprentices regulation, as well as the average number of apprentices by firm size for both high-skill and low-skill sectors. Firms in high-skill sectors have an average number of apprentices below the regulation lower bound. This implies that some of these firms are either paying the fee or not complying with the regulation. In contrast, firms in the low-skill sectors have an average number of apprentices close to the upper bound. Within each regulation window, the average number of apprentices seems decreasing, specially for low-skill firms.\textsuperscript{22}

In Table 2 we study these responses in more detail by looking at the proportion of firms that choose the upper-bound, the lower-bound, apprentices between the bounds, apprentices below or above the bounds, and the ones that have zero apprentices. We consider the sample of firms with more than 15 workers.\textsuperscript{23} The last two columns, show most of the action comes from firms either choosing the lower or the upper bound, or no apprentices at all. Together these responses account for close to 95\% of the total observations. However, these responses are markedly different across the high-skill and low-skill sectors. While 63\% of firms in the high-skill sectors choose to have zero apprentices, almost all firms in the low-skill sectors (99.7\%) choose to train apprentices. Whenever high-skill sector firms have apprentices, they choose the minimum required. On the contrary, the majority of firms in the low-skill sector (65\%) choose the upper bound of the regulation. There are some that choose the lower bound of the regulation, but virtually none choose to have no apprentices.

Together this evidence implies most of the apprentices are trained by firms in the low-skill sectors. Even though 56\% of the observations after the reform are from firms in the high-skill sector, 77\% of the apprentices are trained in low-skill sector firms.

\textsuperscript{22}See Appendix D for the pre-reform number of apprentices by firm size.

\textsuperscript{23}Although the policy allows firms with less than 15 workers to have at least on apprentices, in practice SENA gave priority to firms with more than 15 workers to train the apprentices. In our data 83\% of these firms choose to have no apprentices. Interestingly this number doesn’t differ across low-skill and high-skill sectors.
Figure 5: Fact 2- Apprentices and Number of Workers

Source: SENA.

Table 2: Apprentices and Bounds

<table>
<thead>
<tr>
<th>Compliance Groups</th>
<th>Low-Skill</th>
<th></th>
<th>High-Skill</th>
<th></th>
<th>Total</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>%</td>
<td>No.</td>
<td>%</td>
<td>No.</td>
<td>%</td>
</tr>
<tr>
<td>Upper Bound</td>
<td>7,727.0</td>
<td>65.4</td>
<td>212.0</td>
<td>1.5</td>
<td>7,939.0</td>
<td>30.2</td>
</tr>
<tr>
<td>Lower Bound</td>
<td>3,185.0</td>
<td>26.9</td>
<td>4,615.0</td>
<td>31.9</td>
<td>7,800.0</td>
<td>29.7</td>
</tr>
<tr>
<td>Between Bounds</td>
<td>832.0</td>
<td>7.0</td>
<td>446.0</td>
<td>3.1</td>
<td>1,278.0</td>
<td>4.9</td>
</tr>
<tr>
<td>Below Lower Bound</td>
<td>40.0</td>
<td>0.3</td>
<td>0.0</td>
<td>0.0</td>
<td>40.0</td>
<td>0.2</td>
</tr>
<tr>
<td>Above Upper Bound</td>
<td>5.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>5.0</td>
<td>0.0</td>
</tr>
<tr>
<td>No Apprentices</td>
<td>35.0</td>
<td>0.3</td>
<td>9,197.0</td>
<td>63.6</td>
<td>9,232.0</td>
<td>35.1</td>
</tr>
<tr>
<td>Total</td>
<td>11,824.0</td>
<td>100.0</td>
<td>14,470.0</td>
<td>100.0</td>
<td>26,294.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Note: Number of observations in the post-reform period (2003-2009). Only firms with more than 14 permanent workers. Source: SENA.
Fact 3: High-skill sectors firms pay the fee.

Figure 6 shows around 60% of firms in the high-skill sectors comply by paying the fee instead of having the required apprentices. Interestingly, this fraction is relatively stable on firm size. This suggests having apprentices is a proportional cost to firms rather than a fix cost. In contrast, low-skill sector firms rarely pay the fee. In the Appendix C we show very few firms end up paying fines. So most of the firms comply either by having the required number of apprentices or paying the fees.\textsuperscript{24}

Figure 6: Fact 2- Paying the Fee

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig6}
\caption{Fraction That Pay the Fee}
\end{figure}

Source: SENA.

Together these three facts show that firms in the low-skill and the high-skill sectors responded radically different to the change in policy. Firms in the high-skill sectors tried to avoid training apprentices, while firms in low-skill sectors wanted as many apprentices as possible. Though the policy was successful in increasing the number of trained apprentices, it did so at the expense of inducing firm-size distortions. Moreover, most of the apprentices were trained by low-skill sector firms, something that isn’t necessarily desirable or intended by the policy.

\textsuperscript{24}As mentioned above SENA can easily verify the reported number of workers with the information of the firm payroll taxes.


5 Model

We present the baseline model that captures our empirical predictions and allow us to quantify the effect of the policy on welfare. We also plan to use the model to study counterfactual policies. For exposition purposes, we start describing a one sector, static economy. We then develop the simple extensions to multiple sectors and multiple periods.

5.1 Without Regulation

Suppose an economy with heterogeneous firms characterized by a managerial ability $z \sim \Phi(z)$, with a continuous density $\phi(z)$ (similar to Lucas 1978). This managerial ability is an idiosyncratic characteristic of the firm that can also be interpreted as technological differences or other factors that affect a firm’s productivity. Firms produce a homogenous final good $y(z)$ using labor, which is supplied either by workers $l$ or by apprentices $l_a$.\(^{25}\) Suppose both types of individuals have a unit of time which they supply inelastically. Production combines firm’s managerial ability $z$, with the total labor supplied by both type of workers, $f(l, l_a; z)$. More labor from workers or apprentices, or higher managerial ability produce more output, $f_l, l_a, f_z > 0$.\(^{26}\)

Apprentices are trained using workers’ time. Let $t_a \in [0, 1]$ denote the units of time a worker spends training apprentices. Once trained, an apprentice supplies $g(t_a)$ units of labor. The function $g : [0, 1] \rightarrow [0, 1]$ denotes the training technology and can be interpreted as how difficult it is to teach/learn a particular task. It also reflect the relative productivity of a trained apprentice compared to the productivity of a worker. First we assume $t_a$ is exogenous and set $g(t_a) := 1$. So $t_a$ denotes the net training costs to fully train an apprentice. These net costs captures both the time it takes for workers to train apprentices, as well as the productivity differences between workers and trained apprentices. Later we endogenize the firm’s training decision. Note that depending on the way the production function combines both types of labor, the labor inputs from workers and apprentices can be complements or substitutes in production. Apprentices could be trained in tasks that substitute workers’ labor or tasks that complement it.

Firms maximize profits choosing the number of workers and apprentices. If a firm hires $n$ workers and trains $n_a$ apprentices, the total labor supplied by the workers is $l := n - t_a n_a$ and by the apprentices $l_a := g(t_a) n_a$. Firms interact in a competitive market, taking the wage of workers $w$ and of apprentices $w_a$ as given. We normalize the price of the final good to 1.

A firm with managerial ability $z$ solves,

$$\begin{align*}
\max_{n, n_a \geq 0} & \quad f(n - t_a n_a, g(t_a) n_a; z) - w n - w_a n_a \quad \text{s.t} \quad t_a n_a \leq n.
\end{align*}$$

The constraint implies the number of workers has to be large enough to train the total number of apprentices.

\(^{25}\)The model could be readily extended to have capital or other production inputs. Here we suppose labor is the only input to simplify the model and emphasize the role of apprentices.

\(^{26}\)We use the notation $f_x$ to denote $\frac{\partial f}{\partial x}$. 

We can substitute $l$ and $l_a$ to write an equivalent (perhaps more familiar) optimization problem,

$$\max_{l,l_a \geq 0} f(l,l_a;z) - wl - \frac{w_a + t_a w}{g(t_a)} l_a.$$ 

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

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

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

(ii) the Inada conditions hold, $\lim_{x \rightarrow 0} f_x = \infty$ and $\lim_{x \rightarrow \infty} f_x = 0$ for $x \in \{l,l_a\}$.

(iii) has non-negative cross derivatives with respect to $z$, i.e. $f_{l,z} \geq 0$ and $f_{l_a,z} \geq 0$.

Condition (i) and (ii) imply the existence of a unique interior solution. Condition (iii) that the optimal number of workers $n^*(z)$ and apprentices $n_a^*(z)$ are non-decreasing in $z$. This means firms with higher managerial ability are larger. We formalize these claims in Lemma 1.

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

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

$$[n] : \frac{\partial f}{\partial l} = w, \quad [n_a] : \frac{\partial f}{\partial t_a} g(t_a) = w_a + t_a w.$$ 

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

$$- \frac{f_l}{f_{l_a} g(t_a)} = - \frac{w}{w_a + t_a w}.$$ 

We can use this equation to analyze how wages or the required training time, affect the optimal labor allocation decision. As usual, an increase in the relative wages of apprentices decreases their demand. In the case where $t_a$ is exogenous and $g(t_a) \equiv 1$, an increase in training time decreases the demand for apprentices.

**Lemma 2.** Suppose Assumption 1 holds and the training time is exogenous (with $g(t_a) \equiv 1$), then $\frac{n_a}{n}$ is decreasing in $w_a$ and $t_a$, and increasing in $w$. 

20
5.1.1 Equilibrium

We suppose a simple supply side to this economy. There is a fixed mass of workers $L$ and of apprentices $L_a$. Workers are able to do the work of apprentices but not the other way around. This implies that in equilibrium the apprentice wages has to be smaller or equal to the one that workers get, $w \geq w_a$. We consider a minimum wage that could be binding. So together both constraints imply, $w \geq w_a \geq w_{min} \geq 0$. 27

Let $N := \int n^*(z)\phi(z)dz$ denote the aggregate demand for workers and $N_a := \int n^*_a(z)\phi(z)dz$ the aggregate demand for apprentices. The market clearing conditions are,

$$N + U = L , \quad N_a + U_a = L_a,$$

where $U,U_a \geq 0$ denote the ‘unemployed' workers or apprentices.

Definition 1. A competitive equilibrium is given by wages $(w^*, w_a^*)$, unemployed workers and apprentices $(U^*, U_a^*)$ and labor demands $(n^*(z), n^*_a(z))$ for each firm $z$, such that,

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

(ii) wage restrictions are satisfied, $w^* \geq w_a^* \geq w_{min}$ and

(iii) labor markets clear (2) with $U^* \geq 0$ and $U_a^* \geq 0$.

Notice there is only unemployment when the wage restrictions are binding. If there are some unemployed workers $U^* > 0$, then $w^* = w_a^* = w_{min}$. Similarly if there are some unemployed apprentices $U_a^* > 0$ then $w_a^* = w_{min}$. These restriction and whether they are binding, affect the welfare consequences of the labor policy and the model predictions.

5.2 Regulation

In this section, we describe the firm-size policy in this theoretical framework. This policy intended to increase the number of apprentices trained in two ways. First, it reduced the effective lower bound firms pay for an apprentices to $w_a \geq w_{a min}$, where $w_{a min}$ is lower than the minimum wage $w_{min}$. 28 Second, it imposed apprentice quotas based on the firm-size (number of workers). If the number of workers is $n \in [N_{j-1}, N_j)$ then the number of apprentices has to be $n_a \in [u_j^a, \bar{n}_j^a] =: \phi_j$, $\forall j$. Alternatively, firms can pay a fee $f_a(n, n_a)$ instead of having the required apprentices. This fee is a function of the total number of workers $n$ and apprentices the firm chooses. It is proportional to the difference between the minimum number of required apprentices $u_j^a$ and the apprentices hired $n_a$.

27 The informal labor markets might absorb some of the workers displaced by the minimum wage regulations. In particular, in Colombia the informal labor market is large. This will be relevant for the welfare calculations when thinking about the outside options of workers when the minimum wage is binding. In our empirical analysis, we only observed the formal sector, formal firms that report formal workers.

28 For the period we consider, most apprentices get paid 50% of the minimum wage in the learning phase and 75% in the productive phase.
Firm $z$'s optimization problem when facing this regulation is,

$$\max_{n, n_a \geq 0, \; d_{fa} \in \{0, 1\}} f(n - t_a n_a, g(t_a) n_a; z) - wn - w_a n_a - d_{fa} f_a(n, n_a) \quad \text{s.t.} \quad t_a n_a \leq n$$

if $n \in [N_{j-1}, N_j)$ and $d_{fa} = 0$, then $n_a \in [\bar{n}_a^j; \bar{n}_a^j] \forall j$,

if $n \in [N_{j-1}, N_j)$ and $d_{fa} = 1$, then $f_a(n, n_a) = \psi_a (\bar{n}_a^j - n_a)^+$, $n_a \leq \bar{n}_a^j, \forall j$. (3)

where $d_{fa} \in \{0, 1\}$ is the decision of paying the fee or not and $\psi_a$ is a positive constant. When a firm decides to pay the fee, it can train fewer apprentices than the minimum required $\bar{n}_a^j$ and pay an amount proportional to the number of workers hired multiplied to the percentage difference between the quota's lower bound and the apprentices hired, $f_a(n, n_a) = \psi_a (\bar{n}_a^j - n_a)^+$. As in the actual regulation paying the fee doesn’t allow the firm to exceed the upper bound of the apprentice quota, $n_a \leq \bar{n}_a^j$. This fee function only takes the positive difference between the lower bound of the regulation and the number of apprentices trained by the firm. There is no fee for firms that choose apprentices above the lower bound, so $d_{fa} = 0$ for those firms.

Let us focus on the case where the optimal number of apprentices is a fix proportion of the labor force, $n_a^* = B n^*$ with $B \in \mathbb{R}_+$. We show that if relative wages of apprentices are low enough $\frac{w_a}{w} \to 0$ and the training cost is low $t_a \to 0$, firms want to get as many apprentices as possible. In this case, the optimal number of apprentices is above the regulation’s upper bound for firms larger than $N_1$. Firms bunch at the thresholds $N_j$’s, with missing mass to the left and never pay the fee. In contrast when $\frac{w_a}{w} \to \infty$, the optimal number of apprentices converges to zero, below the regulation’s lower bound for firms larger than $N_1$. This implies some firms to the right of the thresholds $N_j$’s reduce their size and bunch just below the thresholds to avoid having extra apprentices. Additionally, if the fee is low enough ($\psi_a$ small), then some firms prefer to pay the fee instead of having the additionally required apprentices. Proposition 1 formalizes and compiles these results.

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

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

i. the number of apprentices without regulation is $n_a^* = B_a n^*$ and lays above the regulation’s upper bound, $n_a^*(n) > \bar{n}_a^j$.

ii. there exist cutoffs $\{z_b^j, z_r^j\}_j$ such that firms $z \in [z_b^j, z_r^j]$ increase their size to the threshold $N_{j+1}$, so there is missing mass to the left of the thresholds.

iii. firms choose the upper-bound of the regulation $n_a^* = \bar{n}_a^j$.

iv. firms never pay the fee.

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

i. the number of apprentices without regulation is $n_a^* = B_a n^*$ and lays below the regulation’s lower bound, $n_a^*(n) < \bar{n}_a^j$.

---

29Note that the homogeneity of degree $\gamma$ allows for this case. For the linear relationship between apprentices and workers, we additionally need that $f_i/f_i$ is homogenous of degree zero in $z$. 

22
ii. there exist cutoffs \( \{z_j^b, z_j^r \} \) such that firms \( z \in [z_j^b, z_j^r] \), reduce their size \( \epsilon \) below the threshold \( N_j \).

iii. firms that choose to increase apprentices, choose the lower bounds \( n_j^a \).

iv. there exist \( \psi_a > 0 \) such that for \( \psi_a \leq \psi_a \), there is an additional cutoffs \( z_j^f \) such that firms \( z \in (z_j^f, z_j^r] \) choose to pay the fee.

Figure 7: Apprentices After Regulation

Figure 7 illustrates the two cases in Proposition 1. Panel 7a depicts Case 1, where both \( \frac{w_a}{w} \) and \( t_a \) are small. Firms want to have apprentices as they are relatively cheap in terms of wages and training costs. In this case, the optimal number of apprentices without regulation lays above the regulation’s upper bound. This implies that some firms increase their size and bunch at the thresholds, leaving a missing mass of firms to the left of them. All of these firms choose the upper bound of the regulation. Panel 7b depicts Case 2, where \( \frac{w_a}{w} \) is large. In this case, workers are more attractive than apprentices to firms. The optimal fraction of apprentices lays below the lower bound of the regulation. Firms bunch just below the threshold to avoid having extra apprentices from the policy. There is a missing mass of firms to the right of the thresholds. If the fee is relatively small, some firms prefer to pay the fee and choose the optimal number of apprentices and workers. Firms that comply the policy with the apprentice quota choose the lower bound of the regulation.
Figure 8 shows the implications of the policy on the firm size distribution for the two cases in Proposition 1. Panel 8a depicts the case where firms increase their size to get more apprentices. They bunch at each of the thresholds \( \{N_j\}_j \) and there is a missing mass of firms between \( [n_j^b, N_j] \). Panel 8b illustrates Case 2, where firms either reduce their size or pay the fee to avoid having more apprentices. Firms bunch below each threshold, leaving a missing mass of firms of size between \( [N_j, n_j^b] \). If the fee is low enough, firms of size \( [n_j^b, n_j^f] \) prefer to pay the fee instead of having the required apprentices.

5.3 Effects of the Regulation

Proposition 1 gives us a framework to understand our first three empirical findings. Reducing the apprentices minimum wage, allowed firms in low-skill sectors to find it profitable to hire as many apprentices as possible. This implies some firms in the low-skill sector bunch at the thresholds to get more apprentices, choose the upper bound of the regulation and never pay the fee. On the other hand, even though apprentice wages went down, it wasn’t enough for high-skill sector firms to demand more apprentices than the lower bound of the regulation. These firms would, therefore, avoid having additional apprentices. Some of them choose to decrease their size and bunch just below the thresholds. Others prefer to pay the fee instead of having the required apprentices.

However, this proposition doesn’t tell us about the effect of the regulation on firms’ output and profits. To understand these effects is useful to distinguish the three components of the regulation: i) the change in the apprentice minimum wage \( w_{a_{\text{min}}} \), ii) the apprentice quotas based on firm size and iii) the possibility of paying a fee. Each of these has a different effect on output and profits, implications we can later test and use empirically.

Apprentice Wages

If before the regulation the apprentice minimum wage is binding, \( w^*_a < w_{a_{\text{min}}} \), then a decrease in \( w_{a_{\text{min}}} \) increases output and profits. This follows immediately from the firm’s maximization problem
Let $\pi^*(z)$ denote the firm $z$’s profits and $y^*(z)$ its output from the solution of the optimization problem without regulation. Firms profits $\pi^*(z)$ and output $y^*(z)$, are decreasing in wages $w_a$ and $w$. This implies that a decrease in the binding minimum wage $w_{a_{\text{min}}}$ increases profits and output for all firms. Clearly, this is a Pareto improvement. Allowing firms to pay lower wages to apprentices improves efficiency.

**Apprentice Quotas**

Introducing apprentice quotas always decrease total profits and output. If firms want more apprentices than the quota, then the binding regulation will limit their output and lower their profits. If firms want fewer apprentices than the minimum bound required by the regulation, some would hire fewer workers to avoid the regulation reducing output. Firms’ profits decrease both for these firms and the ones that decide to hire the apprentices.

**Apprentice Fee**

The regulation also allows firms to pay a fee instead of having the required apprentices. Firms that decide to pay the fee have the same output as in the case without regulation, but lower profits. This is the case since once they pay the fee they optimally choose the same number of workers and apprentices as they would without the regulation.

Using our structural model we can quantify the net effect of each of these margins on the total change in output and profits observed for each sector. Moreover, making use of the estimated model we quantify the net welfare implications in the economy.

### 5.4 Endogenous Training

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

The FOC with respect to $t_a$ implies,

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

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

If $n_{a^*} \neq 0$ and using the FOCs of (1),

$$\frac{g'(t_a)}{g(t_a)} = \frac{w}{w_a + t_a w}.$$  \hspace{1cm} (4)
Additionally suppose, \( \lim_{t_a \to 0} g'(t_a) = \infty \) and \( \lim_{t_a \to 1} g'(t_a) = 0 \). The first condition states that supplying a small amount of training significantly improves apprentices productivity. The second that once workers allocate most of their unit of time to training apprentices, more training won’t increase productivity too much. These two conditions together with the concavity of \( g \), guarantee a unique interior solution \( t_a^* \in (0,1) \) that solves the firms training decision.

Note that this training decision doesn’t depend on the firm’s managerial ability \( z \). This means training per apprentice won’t change for firms of different size or with different number of apprentices. Optimal training is only affected by changes in the training technology \( g(\cdot) \) or wages. In particular, an increase in the apprentice wage \( w_a \) increases the optimal amount of training. Intuitively, hiring more apprentices becomes more costly so firms choose to train more each of the apprentices they hire. Conversely, an increase in workers wages \( w \) decreases the amount of training each apprentice gets. As the opportunity cost of training is the worker’s wage, this opportunity cost increases as the wage of workers rise. Lemma 3 summarizes these results.

**Lemma 3.** Suppose \( g'(t_a) > 0 \), \( g''(t_a) < 0 \), \( \lim_{t_a \to 0} g'(t_a) = \infty \) and \( \lim_{t_a \to 1} g'(t_a) = 0 \). Then there exist a unique \( t_a^* \in (0,1) \) that solves (4). Moreover, \( \frac{\partial t_a^*}{\partial w_a} > 0 \) and \( \frac{\partial t_a^*}{\partial w_a} < 0 \).

### 5.5 Multiple Sectors

The same analysis applies when we extend the model to \( K \) sectors. Suppose managerial ability is sector specific \( z_k \) and has an exogenous distribution \( \Phi_k \) with continuous density \( \phi_k \). Firms of different sectors differ in their production \( f^k(\cdot) \) and training technologies \( g^k(\cdot) \). For the case where the training is exogenous, we assume that the required training time \( t_a^k \) to fully train an apprentice \( g^k \equiv 1 \) varies across sectors. A sector whose technology requires simple menial tasks probably requires less of the workers time to train apprentices. On the contrary, a sector with highly specialized tasks requires more training.

A firm \( z_k \) in sector \( k \) solves the same problem as in (1),

\[
\max_{n_k, n_{a,k} \geq 0} \int f^k(n_k - t_a^k n_{a,k}, g^k(t_a^k n_{a,k}; z_k) - wn_k - w_a n_{a,k} \ s.t \ t_a^k n_{a,k} \leq n_k
\]

Let \( n_k^*(z_k) \) and \( n_{a,k}^*(z_k) \) denote the optimal labor demand for workers and apprentices in sector \( k \). If Assumptions 1 hold in every sector, Lemmas 1 and 2 follow.

The main difference when solving for the competitive equilibrium with multiple sectors is the market clearing conditions. We consider two possibilities, either there is a common pool of workers who can work in either sector or separate labor markets for workers in each sector. We study both of these cases below. For apprentices, we suppose there is a fixed mass of \( L_a \) apprentices that can be trained in any sector.

First, let us suppose there is a mass \( L \) of workers that can work in any sector. This implies in equilibrium both sectors will pay the same wage rate \( w \) to workers. Let \( N_k := \int n_k^*(z_k) \phi_k(z_k) dz_k \) denote the aggregate demand for workers in sector \( k \) and \( N_{a,k} := \int n_{a,k}(z_k) \phi_k(z_k) dz_k \) the aggregate demand for apprentices in sector \( k \). The market clearing conditions in the common worker case
are,
\[ \sum N_k + U = L, \quad \sum N_{a,k} + U_a = L_a. \]
In this case a competitive equilibrium is still defined as wages \( (w^*_a, w^*) \), unemployed workers \( (U^*, U^*_a) \) and labor demands for each sector \( k \), that satisfy Definition 1 in each sector.

In the second case, we suppose that the labor markets in each sector are perfectly segmented and there is a fixed supply of workers \( L_k \) in each sector. We use these specification in our current quantitative exercises to match the differential wages across sectors. In this case, there is a separate market clearing condition and equilibrium wage rate \( w^*_k \) for each sector \( k \),
\[ N_k + U_k = L_k, \quad \sum N_{a,k} + U_a = L_a. \]
In a competitive equilibrium there are now equilibrium wage in each sector \( \{w^*_k\}_k \) and unemployment of workers in each sector \( \{U^*_k\}_k \).

The same analysis of the regulation applies separately to each sector. Depending on the equilibrium wages \( (w^*_a, \{w^*_k\}_k) \) and the training technologies \( t^*_k \), different cases of Proposition 1 apply.

### 5.6 Change in Workers, Apprentices and Aggregate Output

We study the welfare consequences of the policy by considering its effect over aggregate output, the number of trained apprentices and the number of unemployed workers. A benevolent social planner weights these three variables through some social welfare function. Here we don’t attempt to spell out this welfare function, but rather quantify what happens to each of its ‘inputs’ when with the apprentice regulation. Notice that when we consider multiple sectors there are additional composition effects that could also be important for the social planner. For instance, it might not be socially optimal to train all apprentices in low-skill sectors. Therefore we also consider the composition consequences of the regulation.

Formally a regulation \( \mathcal{R} \) consists of a set thresholds \( N_j \)'s, a set of lower and upper bounds \( w^j_a, \bar{w}^j_a \), the minimum wage for apprentices \( w^a_{min} \) and a fee function \( f_a : \mathbb{R}_+^2 \to \mathbb{R}_+ \), \( \mathcal{R} := \{ \{N_j, \bar{w}^j_a\}_j, w^a_{min}, f_a \} \). Let \( y^k(z) \) denote the output a firm in sector \( k \) with managerial ability \( z \) produces when solving (3). Aggregate output is the sum of the output of all firms across every sector,

\[ Y = \sum_k \int y_k(z) \phi_k(z) dz. \]

We simply compute the change in aggregate output as the difference between aggregate output with regulation and without regulation, \( \Delta Y = Y^r - Y^* \).

Similarly, we compute the change in unemployed workers (if any) as \( \Delta U = U^r - U^* \) and unemployed apprentices as \( \Delta U_a = U^r_a - U^*_a \). We also calculate and report the change in trained apprentices by sector, \( \Delta N_{a,k} = N^r_{a,k} - N^*_a \).

Once we estimate the parameters of the model, we compute the changes in these variables and decompose the effect of each of the elements of the regulation. Moreover, we use the estimated model to compare these implications to counterfactual policies.
5.7 Multiple Periods: Benefits of Training Apprentices

To quantify the potential dynamic benefits of training apprentices, we suppose trained apprentices increase the supply of workers in future periods. As in the model above, in each period the firm chooses the number of workers and apprentices that maximizes that period profits. This means their decision is still static. However, the pool of future workers increases if firms train more apprentices. Hence we suppose that firms don’t internalize this effect given that the labor decision is period by period.

Suppose firms maximize profits for an infinite number of periods, taking prices and wages as given. They discount the future at a rate $\beta$. Firm $z$ solves,

$$\Pi(z) = \max_{\{n_t, n_{at}\}} \sum_{t=0}^{\infty} \beta^t \left[ f(n_t - t_\alpha n_{at}; z) - w_t n_t - w_{at} n_{at} \right] \quad s.t \quad t_\alpha n_{at} \leq n_{at} \quad \forall t$$

The solution to this problem is the same as in the static model above,

$$[n_t]: \frac{\partial f}{\partial l_t} = w_t, \quad [n_a]: \frac{\partial f}{\partial l_{at}} g(t_a) = w_{at} + t_\alpha w_t \quad \forall t.$$

Similarly, when firms are subject to the apprentice regulation, the same period by period solution described above applies as firms don’t internalize the benefits from training.

The difference comes from the market clearing conditions, where the supply of trained workers increase by the number of trained apprentices. Let $L_t$ and $L_{at}$ denote the supply of workers and apprentices in period $t$. If $N_{at}$ apprentices are trained in period $t$, then the supply of apprentices one period ahead $t+1$ is $L_{t+1} = L_t + N_{at}$.

The effect of trained apprentices on welfare comes from the increase in the future supply of workers, which lowers wages and increases aggregate output. So long as the minimum wage is not binding for workers, welfare will increase by training more apprentices.

6 Quantitative Exercises

In this section we estimate the parameters of the model to quantify the effect of the policy on welfare and to analyze policy counterfactuals. In this first estimation we use the simplest version of the model where the labor input of apprentices and workers is linear. In the Appendix E we describe a more general CES parametrization of the production function.
6.1 Parameter Estimation and Moments

Linear Case

Suppose a production function where the labor input from workers and apprentices enters linearly and the case of exogenous training costs ($g(t_a) = 1$),

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

so apprentices are perfect substitutes of workers once they are trained. In the linear case, it immediate to see the interpretation of $t_a$. It simultaneously captures the training costs in terms of workers’ time and the productivity differences between trained apprentices and workers.

Note that for this linear case the Inada conditions of the production function (Assumption 1(ii)) doesn’t hold. However the solution is still unique and has the same properties as in Proposition 1. Analyzing this limit case is useful as we can sharply characterize the solution to (1) with an active regulation.

Since in this case apprentices and workers are substitutes, the optimal number of workers and apprentices come from corner solutions. If workers’ wages are relatively low in comparison to the real cost of apprentices, $w < \frac{w_a}{(1-t_a)}$, then it is optimal to only hire workers, $n^* = \left( \frac{\gamma}{w} \right)^{1/(1-\gamma)} z$ and $n_a^* = 0$. On the contrary if $w > \frac{w_a}{(1-t_a)}$, firms would like to hire as many apprentices as possible. However a firm can’t hire zero workers, as there would be no one available to train the apprentices. Hence the corner solution is to hire as many apprentices as possible and sufficient workers to train them, $n^* = t_a n_a^*$ and $n_a^* = \left( \frac{\gamma}{w t_a + w_a} \right)^{1/(1-\gamma)} z$. Finally in the case where the cost of workers equals the cost of real cost of hiring apprentices, $w = \frac{w_a}{(1-t_a)}$, firms are indifferent between hiring workers or apprentices.

For the supply side, we suppose the supply of workers in each sector $L_k$ and apprentices $L_{a}$ is fixed. There is a minimum wage that is binding for apprentices in both sectors leaving all apprentices untrained. The change in regulation lowers the minimum wage for apprentices to a fraction $\omega \in (0, 1)$ of the workers’ minimum wage.

Training Costs Distribution

Training costs are not the same across firms, even if they belong to the same sector. To capture this unobserved variation, we introduce heterogenous training costs distributions. Each firm draws an idiosyncratic training cost parameter $t_a$ from a sector specific distribution. For our quantitative exercises we parameterize this distribution as a two parameter truncated normal distribution $t_a \sim \mathcal{T}N(\mu_k, \sigma_k)$ with support $[0, 1]$,

$$\mathcal{T}N(t_a; \mu_k, \sigma_k) = \frac{\psi \left( \frac{t_a - \mu_k}{\sigma_k} \right)}{\sigma \left( \psi \left( \frac{1 - \mu_k}{\sigma_k} \right) - \psi \left( \frac{-\mu_k}{\sigma_k} \right) \right)},$$
where $\Psi(\cdot)$ denotes cumulative standard normal distribution and $\psi(\cdot)$ its density. Heterogenous net training costs across firms not only add realism to the model, but also help us match imperfect bunching and the various endogenous responses documented in the aforementioned empirical facts.

**Estimation**

In addition to the previous parametrization, we suppose the managerial ability $z$ has a Log-Logistic distribution, $z \sim LL(\lambda, \theta)$, $\Phi(z) = \frac{1}{1 + \lambda z^{-1/\theta}}$, where $\lambda$ denotes the location parameter of the distribution and $\theta$ its tail parameter, where larger $\theta$ implies a fatter tail.\textsuperscript{30}

Table summarizes the functional form assumptions of the model,

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f(n, n_a; z) = z^{1-\gamma} ((n - t_a n_a) + n_a)^\gamma$</td>
<td>Linear production function.</td>
</tr>
<tr>
<td>$\Phi(z) = \frac{1}{1 + \lambda z^{-1/\theta}}$</td>
<td>Log-logistic managerial ability $z$ distribution.</td>
</tr>
<tr>
<td>$\mathcal{T}N(t_a; \mu_k, \sigma_k) = \frac{\psi(t_a - \mu_k/\sigma_k)}{\sigma \left( \psi \left( \frac{1-\mu_k}{\sigma_k} \right) - \psi \left( \frac{-\mu_k}{\sigma_k} \right) \right)}$</td>
<td>Truncated Normal training costs $t_a$ distribution in each sector $k$.</td>
</tr>
</tbody>
</table>

Using these functional forms we simulate the model for $n_{sim} = 100,000$ firms. We match key moments in the data that stem from the facts above, to the ones generated by the simulations.

Our model has five parameters for each sector $k \in \{u, s\}$,

$$\chi = \{\lambda_k, \theta_k, \gamma_k, \mu_k, \sigma_k\},$$

that we estimate these parameters targeting the pre-reform data on firm size and production, and the post-reform moments stemming from the three empirical facts described above. Concretely, we estimate the production function in each sector using data before the policy to calibrate the production function parameters. Then, we target three set of moments in the data,

1. **The firm size distribution** before and after the regulation (Fact 1),
   - (a) Before: use it to estimate the productivity distribution.
   - (b) After: bunching and missing mass around each threshold in each sector.

2. **The average number of apprentices** by firm size after the regulation and the fraction that choose the upper and the lower bound of the regulation (Fact 2).

\textsuperscript{30}In the Appendix we show that this distribution has the best fit to the pre-reform firm size distribution among two parameter distributions often used to model productivity.
3. The fraction of firms that pay the fee by sector (Fact 3).

Additionally, even though the regulation determines the nominal value of the fee is one minimum wage per missing apprentice, in practice the relative cost of paying the fee instead of hiring the apprentices might be different. We allow $\psi_a$ to vary in order to fit the fraction of fees that pay the fee. We use the regulation specifications to calibrate the remaining parameters,

$$\chi_a = \{w_a, \{N_j, \pi^j_a, \bar{\pi}^j_a\}_j\}.$$

Table 4 summarize the estimated/calibrated parameters,

Table 4: Estimated Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>High-skill</th>
<th>Low-skill</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_k$</td>
<td>Location parameter of productivity distribution $\Phi(z)$.</td>
<td>900.9</td>
<td>638.2</td>
</tr>
<tr>
<td>$\theta_k$</td>
<td>Tail parameter of productivity distribution $\Phi(z)$.</td>
<td>0.694</td>
<td>0.723</td>
</tr>
<tr>
<td>$\gamma_k$</td>
<td>Labor share in output.</td>
<td>0.80</td>
<td>0.56</td>
</tr>
<tr>
<td>$\mu_k$</td>
<td>Mean training distribution.</td>
<td>0.40</td>
<td>0.26</td>
</tr>
<tr>
<td>$\sigma_k$</td>
<td>Standard deviation training distribution.</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>$w_a$</td>
<td>Apprentice wages.</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>$\psi_a$</td>
<td>Fee parameter.</td>
<td>0.06</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Supply Side

To close the model and compute the welfare consequences of the policy, we calibrate the supply side of the economy. We use data on the number of firms in each sector before the policy and the estimated firm size distribution to get the initial supply of workers in each sector $L_k$. On average in the pre period there were 3,496 firms in the high-skill sectors and 3,362 in the low-skill sectors. Using the estimated firm size distribution, we compute the implied initial supply of workers. This yields $L_s = 209,262$ workers in the high-skill sectors and $L_u = 217,518$ workers in the low-skill sectors.$^{31}$

We normalize the minimum wage to $w_{min} = 1$, so all wages are expressed relative to this minimum wage. Using the wage data in the manufacturing census, we compute the average wages for production workers and administrative personnel in each sector before the policy. After taking account social security payments, payroll taxes and other expense related to the labor force, we find comparable workers in the high-skill sector earn on average 10% more than the minimum wage, whilst the ones in the low-skill sector are around 8% higher. This distance of the wages to the $^{31}$The model overestimates the number of workers in the economy, as the fitted distribution has a larger mass of medium and large size firms compared to the data. The average number of workers in the data for the pre-reform period (1995-2002) is 198,259 in the high-skill sectors and 199,405 for the low-skill sectors. However the model captures there are more workers in the low-skill sectors despite it has less firms.

31
minimum wage bound is crucial to identify the training distribution parameters as well as for the quantitative exercises below.

Finally we suppose the number of apprentices $L_a$ is larger than the aggregate demand for apprentices. The enrolment data after the policy suggest there is an excess supply of potential apprentices each year. Only around half of the students enroll in technical and vocational institutions are able to get an apprenticeship (SENA 2018). Therefore for the welfare calculations below, we assume there is enough supply of potential apprentices to cover the demand after the policy.

**Production Function**

First, we estimate the production function for each sector using a control function approach and data from before the reform. We follow the procedure of Levinsohn and Petrin (2003) with the Ackerberg, Caves, and Frazer (2015) correction to estimate the labor share in output $\gamma_k$. We find that the high-skill sector has a higher labor share $\gamma_s = 0.80$, compared to the low-skill sectors, $\gamma_u = 0.56$.

1. **Firm Size Distribution**

We match the firm size distribution before the policy using a maximum likelihood estimation procedure. Suppose before the policy change all firms choose not to have apprentices, $w < \frac{w_a}{1-t_a}$, given the minimum wage restriction $w_a \geq w_{min}$ for apprentices. This implies firm size is a linear function of the managerial ability, $n(z) = \left(\frac{2}{w_a}\right)^{1/(1-\gamma)} z$. As the managerial ability $z$ has a Log-logistic distribution, firm size $n(z)$ will also have a Log-logistic distribution with the same tail parameter $\theta$ but a location parameter $\lambda_n = \lambda \left(\frac{2}{w_a}\right)^{1/\theta(1-\gamma)}$, $n(z) \sim LL(\lambda_n, \theta)$.

Using maximum likelihood we estimate the parameters $\lambda_{n,k}$ and $\theta_k$ that best fit the observed distribution before the policy in each sector $k \in \{u, s\}$, pooling the data from the period before the reform (1995-2002). In Figure 9 we show that the Log-Logistic distribution resembles the main features of the empirical firm size distribution both sectors before the policy. Despite the close fit, it underestimates the mass of firms of size below 15, while it overestimates the mass of firms with number of workers between 15 and 40. Additionally, note that the firm size distribution before the policy is remarkably similar across sectors. This changes drastically after the reform.

---

32 We use the Stata function `prodest` from Mollisi and Rovigatti (2017) to estimate the production function. In Appendix we compare different production function estimation methods including the one developed by Olley and Pakes (1996), Levinsohn and Petrin (2003) and Wooldridge (2009). The estimated labor shares vary significantly across these methods. We plan to recompute the welfare calculations for a wide range of values of $\gamma_k$.

33 Note this holds if worker wages are close enough to the minimum wage. For instance, if the minimum wage is binding, $w_{min} = w = w_a$, the firm won’t hire apprentices for any $t_a \in [0, 1]$.
We use the estimated parameter of the productivity distribution to compute the location parameter of the skill distribution in each sector $\lambda_k$ as a function the other parameters of the model, $\lambda_k = \lambda_k \frac{\gamma_k}{\gamma_k^0} \left( \frac{w_k}{w^0_k} \right)^{1/(1-\gamma_k)\theta_k}$. This ensures we have the best fit of the pre-reform firm size distribution for any combination of wage rates $w_k$ and labor shares $\gamma_k$.

Now, for the periods after the reform, the remaining parameters help match the bunching firms and the missing mass observed in the firm size distribution. We set the apprentices’ wages to 75% of the minimum wage, $w_a = 0.75w_{min}$.\(^{34}\)

\(^{34}\)This corresponds to the apprentices’ wage during the productive phase, which corresponds more closely to the model.
Figure 10 shows the observed and simulated firm size distribution after the change in regulation. The model captures the incentives high-skill firms have to bunch one worker below the threshold. We replicate the excess mass at these bunching peaks, as well as the breath and depth of the missing mass notches. Similarly the model captures low-skill firms’ behavior. Most of these firms seek apprentices. To do so, some choose to bunch at the thresholds to access the upper-bound of the apprentice quota of a higher regulatory bracket, leaving notches below the thresholds. Note that these notches are less pronounced than in the high-skill sector case, leading to an estimation of more dispersed costs of training for firms in low-skill sectors.\textsuperscript{35}

2. Average Number of Apprentices

As we documented in our empirical fact 2, high-skill firms choose an average number of apprentices below the minimum quota, while the average number of apprentices is close to the upper bound of the quota for firms in low-skill sectors. Additionally the model shows that within a regulatory bracket there is an increasing pattern for high-skill firms and an increasing pattern for low-skill firms. For firms in high-skill sectors, to the right of the boundary there are some which prefer to pay the fee and have zero apprentices. This are the ones that preferred to pay the fee instead of reducing there firm size and bunch -below the threshold. The peeks at the thresholds come from some firms within the high-skill sectors having small training costs and trying to get apprentices, this increases the average number of apprentices at the threshold for the high-skill sector firms. On the contrary, low-skill firms just to the right of the boundary have an average number of apprentices closer to the upper-bound than at the end of the regulatory bracket. These firms that bunch at the threshold, do so to get as many apprentices as possible, this is why they choose the upper-bound. Beyond the thresholds there are some firms that have high training costs and avoid having apprentices. At the extreme of the regulatory bracket most firms that want to have apprentices choose to bunch at the threshold above, leaving behind only firms that want to have less apprentices.

\textsuperscript{35}Despite the good fit, our current estimation of the model underestimates the excess bunching mass at the 15 worker threshold and overestimates the excess mass at the other thresholds \{30, 50, 70, ...\}. This is directly related to the fitting disparities of the pre-reform distribution, discussed above.
3. Fraction Paying the Fee

To match the fraction of firms that pay the fee we estimate the cost of paying the fee to be \( \psi_a = 0.06w_{min} \). Figure 12 shows that the simulate model replicates empirical fact 3 where around 60% of firms in the high-skill sector with more than 15 worker decide to pay the fee instead of hiring the required apprentices. In contrast almost none of the firms in the low-skill sectors choose to pay the fee. In line with the theoretical model, firms in the high-skill sector that bunch below the threshold choose not to pay the fee, while a significant portion of the firms above the thresholds pay the fee.
6.2 Results

Training Cost Distribution

Figure 13 shows the estimated distribution of the net training costs. These costs reflect both the differences in productivity between workers and apprentices and the opportunity cost of the worker’s time to train the apprentices. We estimate that training costs are on average higher for firms in the high-skill sector $\mu_s = 0.38$, compared to firms in the low-skill sector, with $\mu_u = 0.25$. This implies that firms in the low-skill sectors are more likely to train apprentices than firms in the high-skill sectors. The training costs for the low-skill sectors are also estimated to be more disperse $\sigma_u = 0.04$, than the ones from high-skill sectors $\sigma_s = 0.025$. Lower $\sigma_k$ makes the notches of the post-reform firm size distribution more pronounced as there is less heterogeneity in training costs. The dispersion of the training costs also reflects the more diverse responses of the firms to the policy. For instance, as we document above, firms in the low-skill sectors sometimes choose the lower bound of the regulation; while nearly all high-skill firms either choose the lower bound of the regulation or pay the fee.
The main positive welfare effects of the policy stem from the change in wages. The policy induced firms to substitute from hiring workers to training apprentices. In general equilibrium the labor markets absorb this excess supply of workers lowering their wages. This will happen as long as the minimum wage restrictions are not binding. For our estimated model the effect of the policy lowers wages in both sectors. The response is stronger in the low-skill sectors where wages fall by 2.7 percentage points, in comparison to only a fall of 0.8 percentage point in the high-skill sector (Table 5). We refer to this decrease in wages as the *general equilibrium effect*.

<table>
<thead>
<tr>
<th></th>
<th>$w$ Before</th>
<th>$w$ After</th>
<th>$w$ Dynamic Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skilled Sectors</td>
<td>1.100</td>
<td>1.095</td>
<td>1.090</td>
</tr>
<tr>
<td>Unskilled Sectors</td>
<td>1.080</td>
<td>1.050</td>
<td>1.013</td>
</tr>
</tbody>
</table>

When trained apprentices enter the next period’s labor supply there is a further decrease in wages. Since the low-skill sectors train more apprentices, wages fall more sharply in that sector. Taking into account this *dynamic effect*, wages in period $t+1$ fall to 1.3% above the minimum wage for the low-skill sectors while the average wages in the high-skill sectors have a less pronounced fall to 9% above the minimum wage (Table 5).

In the next section we quantify and compare the partial equilibrium effects (PE), the general equilibrium effects (GE) and the dynamic effects (DE) of the policy in terms of aggregate labor variables, output and profits.
6.3 Welfare Effects

To quantify the welfare effects of the policy we study three scenarios: i) the partial equilibrium effects where wages don’t change, ii) the general equilibrium where wages adjust to absorb the displaced workers iii) the dynamic effects where trained apprentices become workers.

Table 6: Welfare Effects

<table>
<thead>
<tr>
<th>Scenario</th>
<th>ΔN</th>
<th>% ΔN</th>
<th>ΔNa</th>
<th>% ΔY</th>
<th>% ΔΠ</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Partial Equilibrium</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skilled Sectors</td>
<td>-4081</td>
<td>-1.79</td>
<td>4711</td>
<td>-0.37</td>
<td>-0.89</td>
</tr>
<tr>
<td>Unskilled Sectors</td>
<td>-15273</td>
<td>-6.32</td>
<td>22761</td>
<td>0.43</td>
<td>0.71</td>
</tr>
<tr>
<td>Total</td>
<td>-19354</td>
<td>-4.13</td>
<td>27472</td>
<td>-0.02</td>
<td>0.12</td>
</tr>
<tr>
<td><strong>General Equilibrium</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skilled Sectors</td>
<td>0</td>
<td>0.00</td>
<td>4493</td>
<td>1.00</td>
<td>0.44</td>
</tr>
<tr>
<td>Unskilled Sectors</td>
<td>0</td>
<td>0.00</td>
<td>22712</td>
<td>3.88</td>
<td>4.09</td>
</tr>
<tr>
<td>Total</td>
<td>0</td>
<td>0.00</td>
<td>27205</td>
<td>2.28</td>
<td>2.76</td>
</tr>
<tr>
<td><strong>Dynamic Effects</strong> (t + 1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skilled Sectors</td>
<td>5471</td>
<td>2.41</td>
<td>5519</td>
<td>3.15</td>
<td>2.58</td>
</tr>
<tr>
<td>Unskilled Sectors</td>
<td>22436</td>
<td>9.29</td>
<td>21245</td>
<td>8.66</td>
<td>8.76</td>
</tr>
<tr>
<td>Total</td>
<td>27907</td>
<td>5.95</td>
<td>26764</td>
<td>5.59</td>
<td>6.51</td>
</tr>
</tbody>
</table>

Table 6 shows the effect of the policy on labor variables, aggregate output and aggregate profits for each scenario. When wages are fixed the number of workers decrease by 4.1% after the policy. The substitution of workers is stronger in the low-skill sectors. In general equilibrium, wages adjust to absorb the displaced workers. As long as the minimum wage doesn’t bind these displaced workers get jobs but earn lower wages. In the next period, some apprentices are trained and become part of the labor force, there is an increase of 2.4% of workers in the high-skill sector and 9.3% in the low-skill sectors. The third column shows that in partial equilibrium, the increase in trained apprentices more than compensates for the fired workers. The reduction of the apprentices minimum wage makes it relatively cheaper to hire apprentices increasing the demand for them above compensating for the workers fired. Even in the high-skill sectors firms train more apprentices than the workers they substitute. When workers’ wages adjust, the change in workers is zero by construction, while the number of apprentices hired by both sectors is slightly lower than before given the lower wages for workers.

Note that dynamically, there are less incentives to hire apprentices as the wage of workers fall. This means the policy won’t have the same effect each period. As equilibrium wages get closer to the minimum wage the benefits of training apprentices shrink. In the limit, when wages are at the minimum, the labor market can’t adjust and trained apprentices will add to unemployment or the informal workforce, loosing some of the welfare gains from the policy.
The last two columns of Table 6 show the effect of the policy on aggregate output and profits. This effect is relatively small when wages are fixed. Aggregate output and profits increase in the low-skill sectors and decrease for firms in the high-skill sectors. This negative impact is modest, as firms optimally substitute between workers and apprentices after the regulation. The total labor input doesn’t change significantly and hence the change in output and profits is also small.

In contrast, general equilibrium effects are substantial, specially for the low-skill sector. The fall in wages triggers an increase output and profits for both sectors. Production increases by 3.9% in the low-skill sectors and by 1% in the high-skill sectors, reverting the initial negative effects on output. The percentage change in aggregate profits are of similar magnitude. The dynamic effects magnify the rise of output and profits. The gains are larger for firms in the low-skill sectors as they train more apprentices. Aggregate output increases by 5.6% in total, mostly coming from the 8.7% increase from the low-skill sector and the remaining from the 3.2% from the high-skill sectors. These large effects on output and profits come almost exclusively from the increase in supply of workers and its effect on workers’ wages.

Figure 14 shows the effect of the policy on the number of workers, output and profits, for firms of different size before the policy. It shows the effect of the policy for the three scenarios, compared to the counterfactual scenario where there wasn’t a change in policy. Panels 14 (a) and (b), show that in partial equilibrium, firms’ demand for workers decreased after the change in policy. This substitution is weaker for low-skill firms just below the thresholds, as they increase the number of permanent workers to access more apprentices. On the contrary, high-skill sector firms that were just above the threshold avoid apprentices firing relatively more permanent workers than other high-skill sector firms. Note these effects are attenuated when we consider the general equilibrium and dynamic effects. Once the displaced workers are absorbed by the labor market and their wages fall, the relative benefits of substituting workers for apprentices also go down. Once we include the dynamic effects, there is an increase in the workers demanded in both sectors. As wages fall more sharply in the low-skill sector, the effect is more visible for those firms.

Similarly, Panels 14 (c) and (d) show the effects of the policy on output. As predicted by our theory, low-skill sector firms below the threshold increase their output as they increase the number of permanent workers to access more apprentices. Conversely, high-skill sector firms above the threshold reduce production as they fire production workers to avoid training additional apprentices. Keeping wages fixed, firms away from the threshold won’t change production as they will substitute workers from apprentices maintaining their labor input constant. In general equilibrium, the decrease in wages lessens the response of firms around the thresholds and benefits all other firms inducing a positive effect on output. These positive outcomes for all firms are accentuated by the dynamics effects of the policy.

Finally, Panels 14 (e) and (f) show profits respond even less when wages are fixed. There is no visible effects around threshold. As firms optimize substituting workers for apprentices the effects on profits are negligible. Once wages adjust, all firm benefit from cheaper labor inputs and therefore profits increase. These positive effects are larger for the low-skill firms where wages drop more.
Figure 14: Effect of the Policy on Labor, Output and Profits

(a) Labor

(b) Labor Zoom

(c) Production

(d) Production Zoom

(e) Profits

(f) Profits Zoom
Reduced-Form Evidence

We contrast these quantitative results with direct evidence following a ‘difference-in-difference’ type specification around the regulatory thresholds:

\[ Y_{it} = \alpha_i + \delta_t + \beta Above_i Post_t + \epsilon_{it} \]  

(5)

where \( Above_i \) is an indicator for firms whose size is within 5 workers above the threshold in 2002, the last year pre-reform. We compare those firms to the ones within 5 workers below the thresholds in 2002. \( \alpha_i \) denotes firm fixed effects, \( \delta_t \) year fixed effects, \( Post_t \) is an indicator for the post-reform years from 2003 onwards and \( \epsilon_{it} \) is an error term.

We can compare this regression specification directly to the plots in Figure 14 as we define \( Above_i \) based on pre-reform size. This comes with the additional advantage that the pre-reform firm size distribution is not subject to manipulation, so assignment to this group is exogenous. We focus on four outcome variables: the number of apprentices, the number of permanent workers, output and profits. For our main specification we restrict the sample to firms that stay within two adjacent regulatory thresholds between years, and have less than 300 permanent workers in 2002.36

Our quantitative analysis implies that after the policy, firms above threshold take more apprentices and have relatively less permanent workers. Consistent with these predictions, Table 11 shows a strong ‘first-stage’ where firms above the threshold increase by 0.6 the number of apprentices. These firms also reduced the number of permanent workers by 3.5 workers after the policy. Figure 14 also shows that the quantitative effect on output and profits is small across thresholds. Our regression results coincide with these observations, exhibiting small negative and not significant coefficients for these variables (Table 11).

Table 7: Apprentices and Workers

<table>
<thead>
<tr>
<th></th>
<th>Apprentices</th>
<th>Workers</th>
<th>Output</th>
<th>Profits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Above*Post</td>
<td>0.594***</td>
<td>-3.527***</td>
<td>-35.45</td>
<td>-15.53</td>
</tr>
<tr>
<td></td>
<td>(0.125)</td>
<td>(1.305)</td>
<td>(385.1)</td>
<td>(140.3)</td>
</tr>
<tr>
<td>Mean (Pre-reform)</td>
<td>0.175</td>
<td>42.92</td>
<td>5596.3</td>
<td>1612.2</td>
</tr>
<tr>
<td>Observations</td>
<td>18229</td>
<td>18229</td>
<td>18229</td>
<td>18229</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.542</td>
<td>0.821</td>
<td>0.795</td>
<td>0.713</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

* \( p < 0.10 \), ** \( p < 0.05 \), *** \( p < 0.01 \)

36Concretely we consider firms with a yearly change in permanent workers smaller than 50 permanent workers. We restrict the analysis to firms with less than 300 permanent workers, as the number of firm above and below the thresholds becomes unbalance after this firm size. In the Appendix H we show these results are qualitatively similar to the ones for the full sample, but the coefficient of permanent workers becomes more negative.
Figure 15 plots the ‘leads and lags’ of these regressions for four years before and after the reform (1999-2006). It shows that the parallel trend assumptions hold for these variables. The effect is relatively constant for the number of apprentices and for permanent workers. Most of the coefficients for output and profits are not significant, although there is a small decrease in both these variables the year following the policy. These results coincide with our quantitative findings.

6.4 Policy Decomposition

In Table 8 shows the decomposition of the apprenticeship policy into its three main components: the decrease in the apprentice wage, the quotas and the possibility of paying the fee. We present this decomposition for the partial equilibrium scenario to study the number of displaced workers, the number of trained apprentices and the direct effect on output and profits, without any wage adjustments.
Panel 8A shows that if only the apprentice wage is lowered there is a strong substitution between workers and apprentices. The demand for workers falls more than 34%. This strong substitution comes from the corner solutions in the linear model, as firms will only hire the necessary workers to train the optimal number of apprentices. These effects come mainly from the low-skill sectors, where there is a 67% drop of the labor force. The effects for high-skill sector firms are mild, with only a handful of firms voluntarily training apprentices. Given these labor responses, it is not surprising output and profits increase substantially for low-skill sector firms.

In contrast, when we introduce only quotas (Panel 8B) all firms in the economy are harmed. High-skill sector firms try to avoid more the apprentices low-skill sector firms, as training is more costly for them. In both sectors, firms train less apprentices that the workers they displace, resulting in decrease in aggregate output. Profits also decline for both sectors, but particularly for firms in the high-skill sectors.

Panel C shows that combining quotas with apprentice wage reductions ameliorates some of the negative effects of the quotas, but training a larger number of apprentices. The number of trained apprentice rise, particularly in the low-skill sectors that have more appetite for apprentices. The number of apprentices trained in both sectors exceeds the dislodged workers. Firms in the high-skill sectors have less incentives to reduce the number of workers to avoid the policy as the apprentices became relatively more attractive. There is a small positive effect on output and profits for the low-skill sectors, and a 1% loss for the high-skill sectors.

In Panel D we reintroduce the possibility of paying the fee. This cuts down by half the number of apprentices trained and the associated displaced workers. The decrease in aggregate production is now of 0.44%, while profits get a similar hit as before. As almost no firm in the low-skill sectors pay the fee, all the variables of this sector stay virtually the same.

This decomposition illustrates the role of each of the three components of the policy. The quotas set limits to the number of apprentices a firm trains. The minimum guarantees that apprentices are trained in most firms, the maximum that firms don’t substitute most of their labor force for apprentices. To incentive firms to train apprentices, the regulation also lowers the apprentice wages. In practice lower apprentice wages benefit mostly firms in the low-skill sectors, although it does induce more training for some of the high-skill sectors firms. Finally, fees partially undo the harm to firms that really want to avoid apprentices. This reduces some of the negative impact of the policy, but also decreases the positive general equilibrium and dynamic effects as less apprentices are trained in those sectors.
Table 8: Policy Decomposition

<table>
<thead>
<tr>
<th></th>
<th>$\Delta N$</th>
<th>$% \Delta N$</th>
<th>$\Delta N_a$</th>
<th>$% \Delta Y$</th>
<th>$% \Delta \Pi$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Only $\downarrow w_a$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-skill</td>
<td>-1123</td>
<td>-0.54</td>
<td>1788</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Low-skill</td>
<td>-145107</td>
<td>-66.71</td>
<td>234308</td>
<td>8.09</td>
<td>8.09</td>
</tr>
<tr>
<td>Total</td>
<td>-146230</td>
<td>-34.26</td>
<td>236096</td>
<td>3.61</td>
<td>5.17</td>
</tr>
<tr>
<td><strong>B. Only Quotas</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-skill</td>
<td>-15195</td>
<td>-7.26</td>
<td>8211</td>
<td>-3.92</td>
<td>-4.80</td>
</tr>
<tr>
<td>Low-skill</td>
<td>-10225</td>
<td>-4.70</td>
<td>9323</td>
<td>-0.84</td>
<td>-0.98</td>
</tr>
<tr>
<td>Total</td>
<td>-25421</td>
<td>-5.96</td>
<td>17535</td>
<td>-2.55</td>
<td>-2.37</td>
</tr>
<tr>
<td><strong>C. Quotas $\uparrow \downarrow w_a$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-skill</td>
<td>-8298</td>
<td>-3.97</td>
<td>9110</td>
<td>-1.01</td>
<td>-1.08</td>
</tr>
<tr>
<td>Low-skill</td>
<td>-13684</td>
<td>-6.29</td>
<td>20709</td>
<td>0.49</td>
<td>0.71</td>
</tr>
<tr>
<td>Total</td>
<td>-21982</td>
<td>-5.15</td>
<td>29819</td>
<td>-0.35</td>
<td>0.06</td>
</tr>
<tr>
<td><strong>D. Full Regulation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-skill</td>
<td>-3701</td>
<td>-1.77</td>
<td>4092</td>
<td>-0.41</td>
<td>-0.88</td>
</tr>
<tr>
<td>Low-skill</td>
<td>-13674</td>
<td>-6.29</td>
<td>20697</td>
<td>0.49</td>
<td>0.71</td>
</tr>
<tr>
<td>Total</td>
<td>-17375</td>
<td>-4.07</td>
<td>24789</td>
<td>-0.01</td>
<td>0.13</td>
</tr>
</tbody>
</table>

With these results in mind, we propose alternative policies that acknowledge the heterogeneous training costs across sectors and avoids some of the firm size distortions created by the quotas in the Colombian reform. To discipline these exercises, we compare the welfare consequences of alternative policies keeping a balance budget and training the same number of apprentices in each sector. Then, we discuss the conditions where increasing the number of trained apprentices is desirable.

### 6.5 Other Counterfactual Policies

To be completed.

### 7 Final Comments

In this paper, we study the effect of apprenticeships programs on firms’ outcomes. Using a rich administrative data on Colombian manufacturing firms we document strong heterogeneous firm responses to the policy change. Firms in sectors with higher skill requirements avoided training apprentices, while firms in sectors with low skill requirements benefited from the policy and tried
to get as many apprentices as possible. We develop a simple model to rationalize these empirical findings and quantify the welfare consequences of the policy.

We find, despite the visible firm-size distortions, small partial equilibrium effects of the policy. We argue these small effects are related to the substitution between workers and apprentices after the policy, which in the end had minor consequences in the total labor input of firms. However, we show the apprenticeship policy may have substantial general equilibrium and dynamic effects. If there is room for the labor market to absorb the newly trained apprentices, output and profits increase. This comes at the expense of the welfare of old workers, who before benefited from higher wages when their labor was scarcer.

These positive consequences depend on the slackness of the labor market and the quality of apprentices training. We plan to complement our quantitative analysis with direct evidence on the effects of training. To do so, will study the labor outcomes and wages for apprentices trained in different sectors. Additionally we want to evaluate the general equilibrium predictions of our model looking at wage data across sectors after the policy and how slack is the labor market across sectors and occupations. This allows us to assess the limits of the welfare gains from the policy.

Finally we want to study alternative apprenticeship policies that can deliver the same benefits, with less firm-size and sectoral distortions. Within the studied policy, we can find the optimal combination of the three components for each sector, including sector specific wages for apprentices. More broadly, we will look at alternative policies, such as subsidizing the number of trained apprentices or rebating the training costs to firms, discuss the welfare effects of such policies and the practical issues behind their implementation.
References


46


Appendix

A Firm Size Distribution by Year

Figure A.1 shows the firm size distribution for high-skill and low-skill sectors is smooth for pre-reform years (1996, 1998, 2000 and 2002). Although the plots are more coarse given there are less observations, there aren’t any visible bunching patterns around the thresholds. The firm size distribution in both sectors is remarkably similar. In contrast, if we look at the post reform years (Figure A.2) we see evidence of firms in high-skill sectors bunching below and firms in low-skill sectors at the threshold. This behavior is similar across years.

Figure A.1: Firm Size Distribution Pre-Reform

(a) 1996  (b) 1998

(c) 2000  (d) 2002
B High-skill and Low-skill Sectors

Figure B.1 shows the fraction of professionals for the ranked manufacturing sectors. The dotted line represents the median of the nine sectors. We defined the sectors below this median as low-skill sectors and sectors above as high-skill sectors. Using this definition, the order ranking of sectors is: four low-skill sectors (Wood Products, Textiles, Food and Beverage, Mineral Non-Metallic Products) and five high-skill sectors (Paper and Editorial, Other Manufacturing, Metallic Products, Machinery and Equipment, Chemical Products).
C Fees and Fines

Figure C.1 shows the fraction of firms that pay fees and fines in each sector. Only a small fraction of firms pay the fines in either sector. This high compliance can be explained by the additional information SENA has on the number of workers in each firm, given they manage part of the parafiscal payments.
D Apprentices Pre-Reform

Figure D.2 shows the average number of apprentices for firms of different size both in low-skill and high-skill sectors. This average number of apprentices is significantly lower than in the period after the regulation. The figure shows a mildly increasing relation between number of workers and apprentices. Remarkably there aren’t any clear differences between firms in the low-skill and the high-skill sectors before the change in regulation.

Source: SENA.

E CES Parametrization

For our quantitative exercises, we parameterize the production function as a CES function. In particular, suppose,

\[ f(l, l_a; z) = \left[ \eta(z)^{\rho} + \eta_a(z)^{\rho} \right]^\frac{1}{\rho}, \]

where \( \gamma \in (0, 1) \) and \( \frac{1}{1-\gamma} \) is the elasticity of substitution between the two labor inputs.\(^{37}\) The functions \( \eta(z) \) and \( \eta_a(z) \) allow some flexibility on how the managerial ability affects each type of labor. To have a linear relationship between the optimal number of apprentices and the number of workers, we assume the ratio of \( \frac{\eta_a(z)}{\eta(z)} \) doesn’t depend on \( z \).

The FOC for the CES case (with \( g(t_a) = 1 \)) imply,

\(^{37}\)If \( \rho = 1 \), then both type of are perfect substitutes, if \( \rho \to 0 \) the total labor input is Cobb-Douglas and if \( \rho \to -\infty \) labor inputs are perfect complements.
\[
\frac{l_a}{T} := \frac{n_a}{n - t_a n_a} = \left( \frac{\eta_a(z)}{\eta(z)} \frac{w}{(w_{t_a} + w_a)} \right)^\frac{1}{1 - \rho} =: A \in \mathbb{R}_+.
\]

So the ratio of apprentices to workers is,

\[
\frac{n_a}{n} = \frac{A}{1 + A t_a} =: B.
\]

This ratio \( \frac{n_a}{n} \) is decreasing in \( t_a \) and \( \frac{w_a}{w} \) as Lemma 2 predicts.

We can explicitly solve for the labor demand functions,

\[
n_a^*(z) = \left[ \frac{\gamma [\eta(z) A^{-\rho} + \eta_a(z)]^{\frac{\gamma - \rho}{\rho}} \eta(z) A^{1 - \rho}}{w} \right]^\frac{1}{1 - \rho}, \quad n^*(z) = \frac{n_a^*(z)}{B}.
\]

From Lemma 1 we know \( \frac{\partial n_a^*(z)}{\partial z} \geq 0 \) and \( \frac{\partial n^*(z)}{\partial z} \geq 0 \). Suppose these conditions hold with strict inequality such that we can invert the optimal number of apprentices and workers. Let \( n_a^{-1}(n_a) \) and \( n^{-1}(n) \) denote these inverse functions. Then the firm size distribution is, \( \chi(n) = \Phi(n^{-1}(n)) \).

Using the market clearing conditions (2) and the wage restriction, to can find the equilibrium wages \( (w^*, w_{a}^*) \) and unemployment \( (U^*, U_{a}^*) \). For the multi-sector case, we use the more general market clearing conditions.

### F Fit Pre-Reform Firm Size distribution

Figure F.3 shows the fit of various parametric distributions. Out of the two parameter distributions the Log-logistic better fits the data. Out of all the distributions the Generalized Extreme Value distribution does the best job.
G Production Function Estimation

Table 10 shows the estimated labor share by sector $\gamma_k$ using six different methodologies. We suppose the production function depends on capital $K$, permanent labor $l$ and other intermediate inputs $m$.

Table 9: Labor Share $\gamma_k$

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>FE</th>
<th>OP</th>
<th>LP</th>
<th>W</th>
<th>LP-ACF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skilled</td>
<td>0.98</td>
<td>1.05</td>
<td>0.76</td>
<td>0.63</td>
<td>0.44</td>
<td>0.80</td>
</tr>
<tr>
<td>Unskilled</td>
<td>0.82</td>
<td>0.87</td>
<td>0.64</td>
<td>0.48</td>
<td>0.34</td>
<td>0.55</td>
</tr>
</tbody>
</table>

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

Table 10: Labor Share $\gamma_k$, Only Permanent Workers

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>FE</th>
<th>OP</th>
<th>LP</th>
<th>W</th>
<th>LP-ACF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skilled</td>
<td>0.70</td>
<td>0.64</td>
<td>0.51</td>
<td>0.42</td>
<td>0.23</td>
<td>0.48</td>
</tr>
<tr>
<td>Unskilled</td>
<td>0.63</td>
<td>0.57</td>
<td>0.49</td>
<td>0.35</td>
<td>0.28</td>
<td>0.44</td>
</tr>
</tbody>
</table>

*Note:* The last four columns are computed using Stata user-written program *prodest* from Mollisi and Rovigatti (2017).
Table 11 shows the results for the full sample. The results are qualitatively similar to those in our main specification. Firms above the threshold have more apprentices and less permanent workers. The negative coefficient on permanent workers is more negative than before. Output and profits are still small and not significant. However, the point estimates are now positive. The coefficient plots in Figure H.4 show also show similar trends as in the main specification.

Table 11: Apprentices and Workers

<table>
<thead>
<tr>
<th></th>
<th>Apprentices</th>
<th>Workers</th>
<th>Output</th>
<th>Profits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Above*Post</td>
<td>1.030***</td>
<td>-4.318**</td>
<td>184.5</td>
<td>128.1</td>
</tr>
<tr>
<td></td>
<td>(0.273)</td>
<td>(2.120)</td>
<td>(493.6)</td>
<td>(187.9)</td>
</tr>
<tr>
<td>Mean (Pre-reform)</td>
<td>0.197</td>
<td>60.77</td>
<td>8816.7</td>
<td>2599.2</td>
</tr>
<tr>
<td>Observations</td>
<td>20453</td>
<td>20453</td>
<td>20453</td>
<td>20453</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.474</td>
<td>0.793</td>
<td>0.780</td>
<td>0.711</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

* p < 0.10, ** p < 0.05, *** p < 0.01
Figure H.4: Coefficient Plots

(a) Number of Apprentices

(b) Permanent Workers

(c) Output

(d) Profits