CEDLAS
Facultad de Ciencias Económicas
UNIVERSIDAD NACIONAL DE LA PLATA

DOCUMENTOS DE TRABAJO

The Effects of Being Subject to the Colombian Apprenticeship Contract on Manufacturing Firm Performance

Carlos Ospino

Documento de Trabajo Nro. 230
Julio, 2018
ISSN 1853-0168

www.cedlas.econo.unlp.edu.ar

The effects of being subject to the Colombian apprenticeship contract on manufacturing firm performance.

Carlos Ospino*‡

Universidad de los Andes

February 14, 2018

Abstract

This paper evaluates the intent to treat local average treatment effects of the Colombian apprenticeship contract on manufacturing firm dynamics taking advantage of an exogenous variation generated by the 2002 labor reform and the regulation design. This evaluation is appealing because very little is known about the effects of apprenticeship policies on firm dynamics in developing countries. Moreover, although this regulation has been in place for years it has not been evaluated. Results using a regression discontinuity design which compares small firms subject to the regulation to those that are not, shows positive effects on output per worker (10 log points), total factor productivity (3 log points) and the share of exported sales (2 percentage points). It also shows a negative effect on the average wage bill of directly hired workers (9 log points). These results suggest that small firms which became subject to the regulation adjusted their labor force more efficiently, thus increasing productivity but did not share these gains with workers through higher wages.

Keywords: Apprenticeships, firm productivity, regression discontinuity design.
JEL Classification: C21, D22, O47

1 Introduction

The use of apprenticeship contracts is widespread in Latin America[1]. These regulations frequently link the use of apprenticeship contracts to firm size, either by limiting the maximum number of apprentices, or even by imposing quotas based on the number of regular workers. As such, regulations on apprenticeship contracts frequently fall within the category of size-dependent policies.

The impact of size-dependent policies on the efficiency of an economy has received increasing attention in the growth literature. Guner et al. (2008) have studied the effect of such policies for

*This paper incorporates comments by the CEDLAS' Grant for Graduate Thesis competition 2015 to the second chapter of my doctoral dissertation. I thank Marcela Eslava for her valuable mentorship during this project. All errors are my own.

†I acknowledge funding for this research project from CEDLAS. I thank anonymous referees from CEDLAS for their helpful comments and suggestions.

‡I thank my dissertation committe for their valuable comments and suggestions. They are: Juan Esteban Carranza, Pablo Lavado, Carlos Medina and Andrés Zambrano.

[1] ILO's CEINTERFOR reports the existence of regulation for many Latin American countries including: Argentina, Brazil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Honduras, Peru, Panama, and Uruguay. http://www.oitcinterfor.org/jovenes/contratos-aprendizaje

the size distribution of firms, productivity and output. In the first chapter I showed that labor substitution as a response to the apprenticeship contract can affect the allocation and composition of labor among firms (Ospino, 2016). Understanding the effects of the apprenticeship contract on firm productivity, wages and capital accumulation is vital to designing policies that consider how firms are affected, since most evaluations only consider the effects of policies on those that receive training.

In this paper I take advantage of the exogenous variation in the apprenticeship contract in Colombia which made small firms subject to this regulation. I exploit the regulation's design to identify the effects on small firm outcomes of the use of apprenticeship contracts. This regulation and the relevant features for the analysis are discussed in section 2. The paper is related to at least three different branches of the economics literature. A set of studies focuses on the impact of size-dependent policies on firm outcomes. This literature is both theoretical and empirical and finds that restrictions on the use of capital or labor which are conditional on firm size can have important effects on aggregate productivity. Such policies can explain the emergence of smaller firms which shift the size distribution to the left (Guner et al., 2008; Braguinsky et al., 2011; Garicano et al., 2013).

In their model, Guner et al. (2008) find that restrictions on labor use have larger effects on output, firm size and productivity, than restrictions on capital use because there are general equilibrium effects, where the most important are lower wages and the creation of smaller firms. The mechanism is the following. Higher labor costs reduce total labor demand which lowers wages. Lower wages make less productive firms profitable and induces the emergence of smaller firms. In this sense, Braguinsky et al. (2011) and Garicano et al. (2013) provide evidence on size-dependent labor regulation in Portugal and France respectively, which affects labor allocation and productivity. Braguinsky et al. (2011) argue that the level of employment protection in Portugal, which they consider among the highest in OECD countries, explains a size distribution of firms shifted to the left with respect to countries with less restrictive regulations. Their argument is that employment protection regulation affects disproportionately larger firms than smaller ones, thus providing incentives to reduce size. Finally, Garicano et al. (2013) argue that employment protection laws that affect firms with at least 50 workers in France, explain important dead-weight losses that can be as high as 5% of GDP. The current paper contributes to this literature by providing empirical evidence of how a size-dependent regulation in a developing country affects the allocation of labor and capital around the threshold where the regulation kicks in.

A second group of studies focuses on the impact of apprenticeship contracts and other forms of training on firm labor productivity. The main message from this literature is that evaluating training policies by just focusing on the wages of trainees is insufficient to capture all the benefits of training, since it ignores increases in labor productivity. It finds that labor productivity gains can be twice as much as wages gains, but these gains differ across economic sectors (Dearden et al., 2006; Mohrenweiser and Zwick, 2009; Konings and Vanormelingen, 2010). Dearden et al. (2006) find that the impact of training on wages is half the impact on firm labor productivity (0.35 and 0.60, respectively), thus providing evidence of the underestimation of the impacts of training when using wages alone. They also show that this result is driven by sectors with low wages, suggesting that the monopsony power by firms in these sectors allows for the difference between wages and productivity gains. Mohrenweiser and Zwick (2009) use matched employer-employee level data to estimate the impacts of training apprentices on productivity measures of German firms between 1997 and 2002. The paper's main contribution is testing whether all sectors face costs of increasing the share of apprentices, something that was taken for granted in the literature; they do this by considering apprentices in different types of occupations: manufacturing, craft and construction, and commercial occupations, within firms. In particular, only in the manufacturing sector an increase in the share of apprentices reduces net profits, but has no effect on labor productivity (measured as value-added per worker). Konings and Vanormelingen (2010) study how voluntary job training undertaken by Belgian firms affects firm productivity. They use a panel of firms that report detailed information about training expenditures, the intensity of training and the share of trained workers. They find that the productivity premium for trained workers relative to those that did not receive training is 23% while the wage premium is 12%. They conclude that in this context it is optimal for firms to provide training since productivity rises by a higher factor than the rise in wages. This effect is known as wage compression (Acemoglu and Pischke, 1999). While this literature has assessed the effect of voluntary training on workers and firms, in Colombia apprenticeship contracts are mandatory. This type of regulation has not been assessed and constitutes a relevant contribution to the training literature.

Finally, the paper is related to the literature about the effects of labor regulation reforms on the manufacturing sector's performance (Besley and Burgess, 2004; Eslava et al., 2004). Besley and Burgess (2004) exploit state level variation over time in amendments to the Industrial Dispute Act (IDA) of 1947, using data from 1958 to 1992 and find that pro-worker legislation in India had a negative effect on investment, employment, productivity and output of formal manufacturing firms. It also increased informal manufacturing activity. Eslava et al. (2004) studied the role of factor allocation and demand shocks in explaining changes in productivity after several reforms took place in Colombia in the early 1990's. While their focus is not exclusively on labor reforms, they find that after the reforms in the 90's the allocation of production towards more productive firms increased total factor productivity. The current paper provides evidence of how a particular size-dependent regulation reform to a flexible form of contracting that took place in 2002 and has not yet been evaluated affected manufacturing firms' performance in an institutional context which is different from India but that nonetheless shares some similarities. For example, Ospino (2016) shows that the change in the apprenticeship contract regulation is associated with an increase in the use of outsourced labor contracts in Colombia. Bertrand et al. (2015) show that the IDA, a size-dependent regulation, is associated with the increased use of contract labor from staffing companies by Indian firms with more than 100 workers. Bertrand et al. (2015) find that the availability of a flexible form of contracting allowed firms to invest in risky projects, increase total labor demand and cope with demand shocks in spite of the tight labor regulation they are subject to. Therefore, I will test whether the observed labor outsourcing by Colombian small manufacturing firms as a response to the apprenticeship contract is associated with capital investment decisions by firms.

The paper is structured as follows. The first section is this introduction. In section 2 I look at the relevant features of the regulation which are important for its evaluation. In section 3 I explain the empirical approximation to evaluate the effects of this policy on firm performance. In section 4 I discuss validity tests for the empirical approximation as well as the results of the econometric exercises. Finally section 5 ends with a discussion of the main findings and its implications for public policy. In the Appendix I provide additional results and robustness checks for the main exercise.

2 Regulatory framework

2.1 Regulation before 2003.

Law 188 of 1959 established the nature of the apprenticeship contract as a labor contract. Apprentices were employees of the firm and the labor code regulated this working relationship. Their salary could not be lower than 50% of the minimum wage and it had to increase as the apprentice gained knowledge in her craft until it reached at least a full minimum wage. Decree 2838 of 1960 established that employers with more than US$15,000[2] in capital or more than 20 permanent workers, had the obligation of hiring apprentices. The number of apprentices could not exceed 5% of firm personnel. Given that apprenticeships could only be hired for occupations defined by the labor ministry based on recommendations by SENA (Colombia's vocational and training institution), only students of programs offered or recognized by SENA could be hired using apprenticeship contracts. The prevalent form of compliance with the regulation was modifying regular workers labor contracts and providing time for training. This practice allowed employers to train their workforce at SENA's nocturnal programs while these continued to work in their regular daily shift. Such an alternative allowed firms to comply with the regulation without affecting their labor force or reducing production. Accordance 007 of 2000 established that the regulated quota would be determined using the number of skilled workers at the firm. The amendment defined skilled worker as those in the list occupations for which an apprenticeship contract could be signed. In practice [3] the apprenticeship quota was calculated using only the number of non-production regular workers which was known as the "administrative staff" at the firm.

2.2 Regulation after 2003.

Law 789 of 2002 was a major labor reform approved on December 27 of that year, which overhauled among other things, the apprenticeship contract regulation. For example, article 30 changed the legal nature of the apprenticeship contract from a regular labor contract to a special form of hiring which no longer implied an employer-employee relationship. The law limited the duration of each contract to a maximum of 2 years and stated that apprentices must receive a monetary stipend. Thus, starting in January of 2003 apprentices were no longer considered firm employees. This same article established that compensation will be as follows: 50% of a minimum wage during the classroom training phase and 75% for the duration of the on-the-job training phase (100% of a minimum wage in the case university students[4].).

Article 32 states that firms which hire at least 15 workers in any sector, except in construction, are obligated to hire apprentices for the occupations related to their economic activity. Article 33 defined the regulated quota (RQ) as the minimum number of apprentices the firms must hire. Firms subject to regulation must hire one apprentice for every 20 regular workers, and they must hire an additional apprentice if the number of workers is a multiple of 10. Therefore firms between 15 and 29 direct workers must hire one apprentice, firms between 30 and 49 workers must hire two, those between 50 and 69 workers must have three apprentices and so on[5]. This same article states that if the apprenticeship contract were to end for any reason, the firm must replace the apprentice so that

[2] $100,000 Colombian pesos of the time, converted using the exchange rate provided by Colombia's central bank.

[3] I thank Lizeth Cortés for providing helpful information to understand the details of the previous regulation

[4] University students can only be hired during the on-the-job training phase.

[5] University students can only be hired to fulfill maximum 25% of the regulated quota.

it always fulfills its RQ. Article 34 established an alternative way of complying with the RQ, which is called "monetizing". Under this option firms must pay a monthly fee to SENA. It is calculated by multiplying 5% of their labor force size, excluding contractors and temporary workers, times the minimum wage. Article 35 established that the apprentice selection process will be carried out by firms, but current or past employees can't be hired under apprenticeship contracts. Apprenticeship contracts cannot be renewed once they've expired which implies that the same person can not be an apprentice more than once while obtaining a degree.

Apprenticeship contracts require firms to incur in other costs. In addition to an apprentice's compensation, decree 933 (Signed in April) of 2003 established that firms must pay health and professional risk insurance for apprentices as if they earned a full minimum wage[6]. Given that apprentices are not considered firm workers they are less costly than a minimum wage worker during their productive phase. (Health costs amount to 8.5% while professional risk insurance range from 0.348% to 8.7%, depending on the economic activity of the firm. See footnote 6). A transitory paragraph in article 11 established that firms for which SENA had not established the RQ must do so themselves within 2 months of the decree's publication. Therefore in practice, all firms must have complied with the new regulation by June of 2003. Paragraph 1 allowed firms with less than 15 regular workers to voluntarily have one apprentice even though these firms were in no obligation to do so. This option was initially only allowed for firms with less than 10 workers in Law 789 of 2002. Paragraph 2 of article 11 allows the firm to split the RQ among its different plants according to its needs. Paragraph 3 allows firms to hire up to twice its RQ as long as the firm does not reduce the number of regular employees used to calculate the quota. Article 14 established the sanctions for not complying with the RQ in the amount of a full minimum wage for every apprentice not hired or monetized, in addition to the amount due, including interest.

In short, the current regulation applies to a broader group of firms (especially small firms) than before the reform took place since the quota is calculated based on the total number of regular workers and not just non-production staff. In practical terms this modification more than doubled the number of apprenticeship contracts between 2002 and 2003 from 33,000 to 72,000 per year across all economic sectors. However, it isn't clear whether the regulation generated more costs than benefits for firms. While apprentices cost less during their productive phase than minimum wage workers[7], firms are required to pay for them even during their classroom training which becomes a net cost since classroom training can be as long as 75% of the apprenticeship's duration.

Finally, firms subject to the regulation face other administrative costs which are not easy to quantify. For example, in July and December of every year firms must fill out forms informing SENA whether the number of workers hired during the past semester changed in a way which affects its RQ. In these forms firms must detail the number of workers in each occupation and the number of hours they work in a typical week. Once the form is filled out, firms must wait for the expedition of a legal document (Resolución) which determines the new official quota. Further, selection and interview of apprentices must be performed exclusively from the pool of candidates SENA lists in its website and the firm must incur in the affiliation costs of apprentices to social security and professional risks insurance. These administrative costs are more likely to be important for firms around the first threshold of compliance with the apprenticeship contract since it implies incurring in the learning

[6] In Colombia the minimum wage is high and binding (Maloney and Mendez, 2004). Non-wage labor costs in Colombia include severance payments, health and pension contributions, payroll taxes, two annual bonuses, vacation compensation, and a transportation subsidy, all of which amounts to 66.6% for minimum wage workers (Mondragón-Vélez et al., 2010).

[7] The apprenticeship contract regulation states that if the national unemployment rate falls below 10% apprentices must be paid a full minimum wage. The unemployment rate didn't reach single digit levels in Colombia until 2010.

costs of complying with a new regulation.

3 Empirical approximation

I now describe the data, the variables of interest and the econometric models to be used in estimating the impact of apprenticeship contracts on firm dynamics.

3.1 Data

The data for the main analysis comes from Annual Manufacturing Survey (*EAM* by its initials in Spanish) for the years 2001-2004[8]. Rather than a survey as its name suggests, EAM is a census of all formal manufacturing Colombian firms who hire at least 10 employees or generate an output value of at least 35,000 USD. The number of manufacturing establishments range from 7.909 in 1995 to 9.809 in 2011. It has very detailed information on output, sales, asset investments and intermediate materials consumption, as well as labor demands broken down by different worker categories (e.g. temporary, permanent, men, women, skilled, managerial, production, non-production.) This information allows the estimation of production functions from which TFP is recovered. The data are proprietary, administered by the National Statistical agency (DANE) and must be accessed on-site at DANE's External Special Processing Room (*SPEE* for its initials in spanish).

3.2 Construction of outcome and control variables

All monetary variables are expressed in 2011 prices using DANE's producer price index (IPP). The IPP varies by industry class at the two digit ISIC code (CIIU Revisión 3 AC).

Output.–It is measured as the wholesale value of all goods manufactured by the establishment net of indirect taxes.

Investment.–It is constructed as the net purchases of assets, excluding buildings and land.

Capital.–It is constructed using the iterative equation $K_t = K_{t-1} * (1 - \delta) + I_t$. Where δ is the depreciation rate (which was set to 5%) and I_t is asset investment by firms. The capital measure also excludes buildings and land purchases or sales.

TFP.–Total factor productivity was estimated using the methods by De Loecker and Warzynski (2012) and De Loecker (2013). These authors estimate a parametric Cobb-Douglas production function, $f(k, l, m) = Ak^{\alpha}l^{\beta}m^{\sigma}$. Where k is capital, l is labor, and m is intermediate materials. Total factor productivity follows an order 1 autoregressive process which also a function of past exporting status[9].

Skilled and Unskilled labor.–Skilled labor is defined as production professionals and technicians, while unskilled labor is defined as production laborers and operators. Both categories exclude non production workers and apprentices.

[8] I also constructed a longer longitudinal version of the dataset for the period 1995-2011 following the methods proposed by Eslava and Melendez (2011). This latter dataset was used to test the validity of using a difference-in-difference approximation for the current analysis.

[9] A second method to estimate production function parameters used the parameters estimated by Eslava et al. (2004) for the period 1982-1998 to construct TFP.

Wage bill.–Direct labor wage bill, includes wages of permanent and temporary workers directly hired by the firm but excludes social security payments and benefits of these workers. Outsourced labor wage bill includes the payments of all production workers hired through temporary third party agencies.

Exports.–Two variables were used. The first variable indicates whether the firm is an exporter or not. Exporters are defined as any firm reporting positive amounts of the share of sales exported. A second variable measures the share of sales exported, in levels.

3.3 Empirical strategy

An empirical evaluation of the impact of the Colombian apprenticeship contract on firm outcomes is appealing for several reasons. The natural experiment generated by the reform allows the estimation of the causal effect of this regulation on measures of firm productivity such as total factor productivity (TFP) and output per worker. Given that the empirical strategy to identify the parameter of the production function from which TFP is recovered, rests on the assumption that TFP evolves conditional on the exporting status of firms, it makes sense to explore whether the regulation also had an impact on the likelihood and the levels of exporting. It also allows to test whether this regulation had an impact on the substitution of capital for labor and on firm investment. While Ospino (2016) showed that the apprenticeship contract is associated with a reduction in total labor demand and the substitution between direct and outsourced labor, his model did not incorporate capital and thus was not able to answer whether firms also substituted labor for capital as a response to the policy. Finally, In addition to its negative effects on labor demand this regulation could have affected worker wages. Therefore, an evaluation of its effects on the average expenditures on workers wages will be carried out.

The 2002 reform to the apprenticeship contract regulation affected firms in two ways: 1) Many small firms that were not subject to the apprenticeship contract regulation before 2002 were required to do so starting in 2003. 2) The regulation that is currently in place generates heterogeneity in the share of apprentices that firms must hire. These shares change discontinuously at specific thresholds. This paper exploits the first feature and leaves the second one to be addressed in a subsequent paper[10]. Finally, the Colombian apprenticeship contract is similar to the ones used in other Latin American countries and apprenticeship contract regulations are a topic of regional interest. See Fazio et al. (2016) and `http://blogs.iadb.org/trabajo/category/aprendices/` for a series of blogs on the subject by The Inter-American Development Bank' labor markets division. The key difference in Colombia is its compulsory quota but it is, nevertheless, a useful regulation to understand how apprenticeship contracts can affect firm performance.

To evaluate the impact of being subject to the apprenticeship contract, an appealing methodological approach is a Difference-in-Difference (DD) estimation because it exploits the fact that before the regulation changed some small firms were not subject to the regulation and in spite of the change they are still not required to comply with it while other firms of similar size are. A deeper analysis of the data showed that the assumptions necessary for the DD estimation did not hold. In particular I found evidence of anticipation effects and trends did not follow a common-trend pattern[11]. Given these findings, to take advantage of the regulation threshold which separates near identical firms

[10] An earlier version of this paper estimated the impact of the share of apprentices on the outcomes of interest for several thresholds. For consistency with the first chapter in my doctoral dissertation which restricts the analysis to the first threshold, the committee recommended that the other thresholds be studied in another paper.

[11] These exercises are available upon request.

from having to comply with the apprenticeship contract, a regression discontinuity design was preferred.

As discussed, the changes introduced by the 2002 reform can be exploited under a regression discontinuity design (RDD) to determine the effects of the apprenticeship contract on firm performance for firms around the regulation threshold. This methodology relies on the similarities of the groups which are being compared. Firms locating before the threshold of compliance with this regulation were likely to be similar to those which have to comply with it before the regulation changed. Notice first that the 2002 regulation changed the threshold level of compliance with the regulation from 20 to 15 workers. And second, it changed the type of workers considered to determine the mandatory quota from management staff to all directly hired workers. Therefore it's very unlikely that firms had incentives to change their directly hired labor demand at the threshold of compliance with the regulation before the reform was in effect. If firms had no incentives, and did not systematically modify the number of directly hired workers in 2002 to avoid compliance once the regulation was in place, then treatment assignment into the regulation is "as good as" randomly assigned at the threshold, and a RDD may be valid (Lee and Lemieux, 2009).

The validity of the RDD rests on the limited capacity of firms to perfectly control the assignment variable (the number of directly hired workers in the apprenticeship contract regulation), therefore such limited capacity is assumed. If it's costly for firms in the short run to adjust the number of directly hired workers, then these firms will not systematically fire workers in order to avoid being subject to the regulation. This could be due to a number of legal factors such as severance payments and contractual clauses, or economic factors such as positive demand shocks and technological requirements in the production process. Figure 10 shows that in 2002 and 2003 a couple of new firms hiring 15 direct workers appear in the data whereas in 2004 and 2005 no new firms are located exactly at this regulation threshold. This suggests that either firms did not anticipate the changes introduced by the regulation at the threshold, or that they did not have incentives to avoid locating at this threshold before the regulation was in full effect.

As a technical point, implementing the RDD to evaluate the apprenticeship contract must consider the fact that the assignment variable, the number of directly hired workers at the firm, is discrete. In this case the limit of the expected value of the outcome of interest as we get arbitrarily close to the threshold from either side does not exist and thus the only way to identify the model parameter of interest is through a parametric estimation (Lee and Card, 2008; Gelman and Imbens, 2014). I will follow the standard practice of clustering standard errors at each level of the assignment variable as suggested by Lee and Card (2008). Doing this takes into account the correlation of firms that have the same number of directly hired workers.

$$y_{ij} = \beta_0 + \beta_1 D_{ij} + m(N_{dij}, p)\gamma_p + D_{ij} \times m(N_{dij}, p)\alpha_p + \mu_{ij} \tag{1}$$

The model to be estimated is given by equation (1). The assignment variable N_{dij} is the number of directly hired workers by each firm in 2002. It has been normalized so that it takes the value of zero at the threshold cut-off value (15 directly hired workers). i indexes firms and j indexes each discrete value of directly hired workers by firms, since standard errors are clustered at this level. The advantage of this normalization is interpreting β_0 as the expected value of the outcome at the threshold for firms not subject to the regulation. $D_{ij} = 1[N_{dij} \geq 0]$ is an indicator variable that identifies firms that hired between 15 and 24 directly hired workers in 2002. $m(N_{dij}, p)$ is a row vector of a second degree polynomial of N_{dij} which is also defined in 2002[12]. $D_{ij} \times m(N_{dij}, p)$

[12]A second degree polynomial was used since it provides sufficient functional flexibility and Gelman and Imbens

allows polynomial slopes to differ for firms below and above the threshold of compliance.

β_1 is the parameter of interest and captures the intent-to-treat (ITT) impact of the apprenticeship contract on firm outcome y since $E[y|N_{dij} = 0, D_{ij} = 1] - E[y|N_{dij} = 0, D_{ij} = 0] \equiv \beta_1$. It estimates the ITT parameter since outcomes are measured in the year 2004, but treatment and assignment variable polynomials are defined using the observed direct labor demand in 2002. The analysis takes the year 2004 as the main estimation sample for two reasons: 1) Ospino (2016) shows that firm labor demand responded to the change in the apprenticeship contract regulation starting in 2004. 2003 appeared to be a transition year since as discussed amendments to the regulation were introduced as far as June of that year. 2) In the year 2003 the number of apprentices hired was not reported in the data and these must be subtracted from total employment as apprentices are not considered, by regulation, firm workers. Moreover, since having an apprentice implies being subject to the regulation, per worker variables would by construction be lower at the threshold for treated firms. For these reasons D_{ij} determines whether firms should have been subject to the apprenticeship contract given their direct labor demand in 2002.

A possible concern to the identification strategy may be that firms not subject to the regulation could voluntarily hire apprentices. As discussed in the regulation section, this exception was initially allowed for firms with 10 or less workers when the regulation changed in December of 2002, while the modifying decree changed this restriction in April of 2003. My preferred specification will compare firms hiring between 11 and 19 workers, thus control firms were not allowed to hire apprenticeship at the moment the assignment variable is measured.

4 Results

4.1 Assumptions and validity tests

In this section I carry out standard assumptions and validity test for RDDs to make sure the approximation is appropriate for the current evaluation.

Changes in outcomes at the threshold

As Imbens and Lemieux (2008) suggest, graphical analysis is an integral part of the RDD. In this section I show non parametric estimations of the outcome variable around the cutoff value which should provide insights for whether there are any effects of the regulation. Figures 1-9 show local linear polynomial estimations that plot the relationship between outcomes in the year 2004 and direct labor demand in 2002. These figures show that in all cases the slopes of the functions fitting the data appear to be different for treatment and control firms which provides empirical support for estimating different slopes in equation (1). While point estimates appear to differ at the threshold, the 95% confidence intervals are so wide that one can not reject the null hypothesis of being equal. This second point suggests the inclusion of baseline covariates which may help reduce the sample variability of the estimators (Lee and Lemieux, 2009).

(2014) warn against the use of higher order polynomials in RDD estimations.

Figure 1: Average Log output per total number of workers

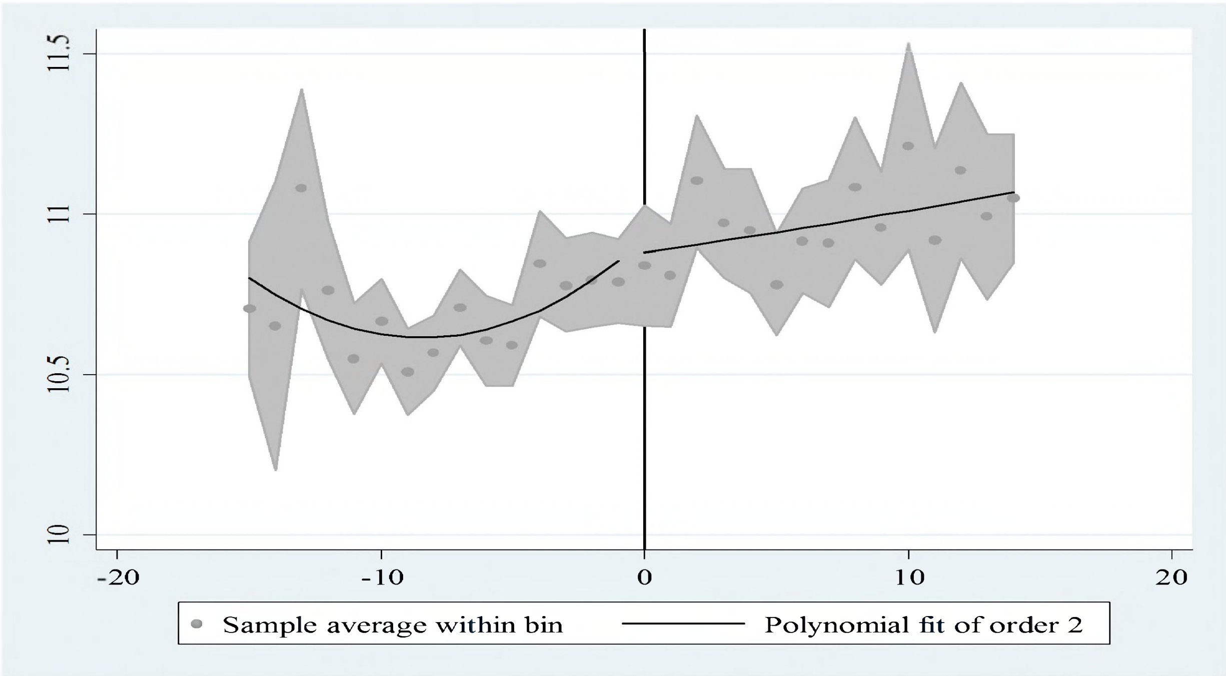
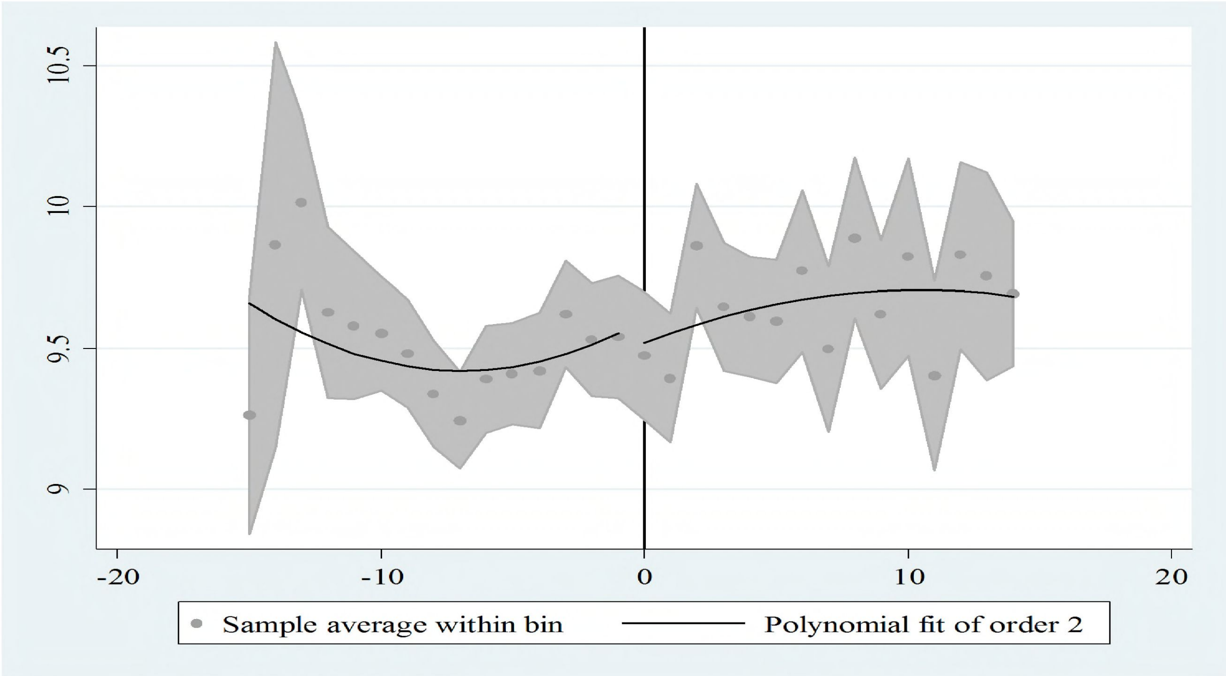


Figure 2: Average of Log capital per total number of workers



Source: EAM 2004. Figure plots local linear polynomial of order 2 of the outcome variable for treatment and control firms. Dots represent the average value of the outcome at each estimation bin. Shaded areas contain a 95% confidence interval for the estimation.

Figure 3: Average Log Total Factor Productivity

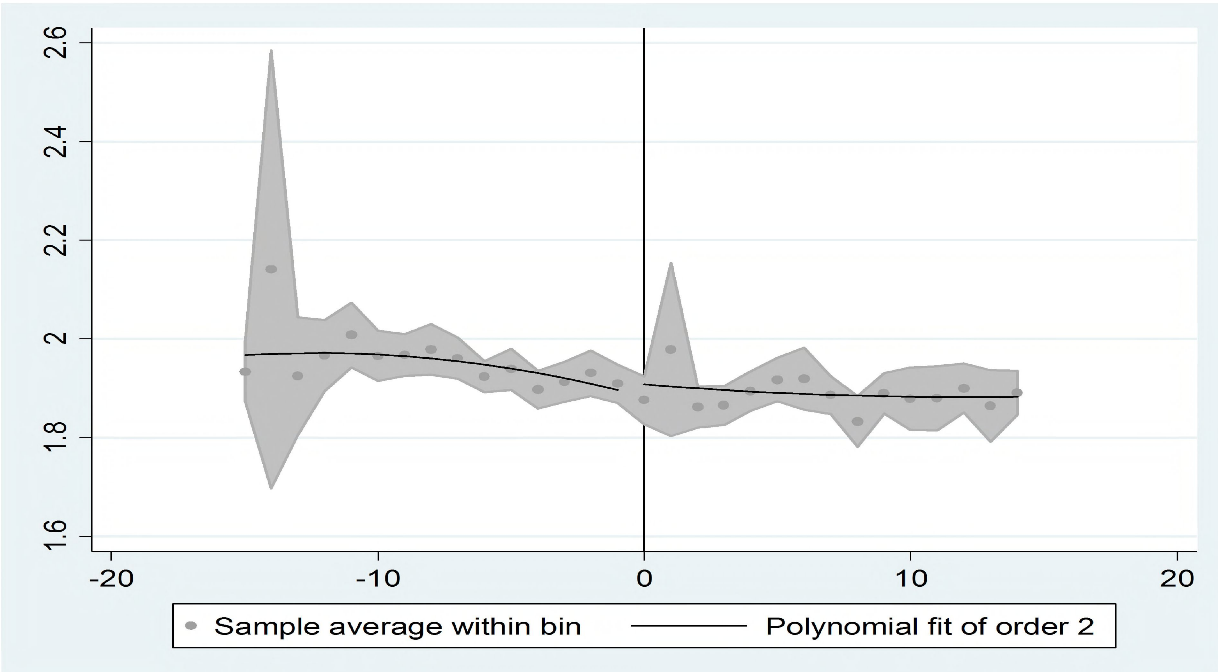
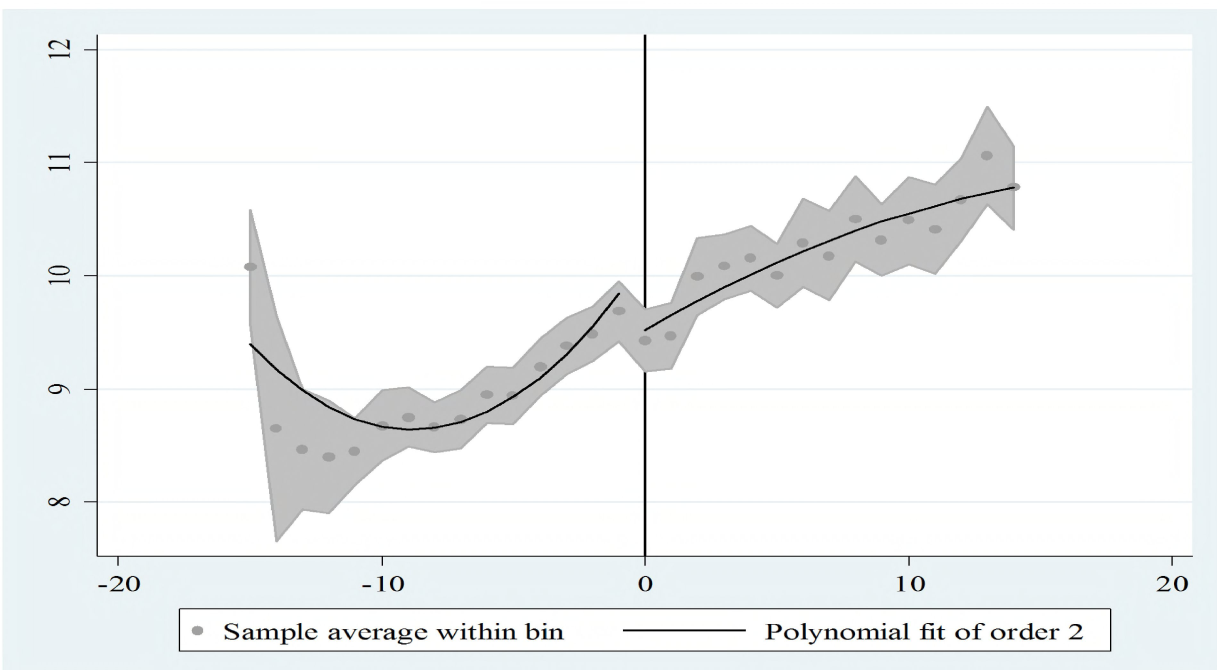


Figure 4: Average Log Investment



Source: EAM 2004. Figure plots local linear polynomial of order 2 of the outcome variable for treatment and control firms. Dots represent the average value of the outcome at each estimation bin. Shaded areas contain a 95% confidence interval for the estimation.

Figure 5: Average Skilled/Unskilled ratio

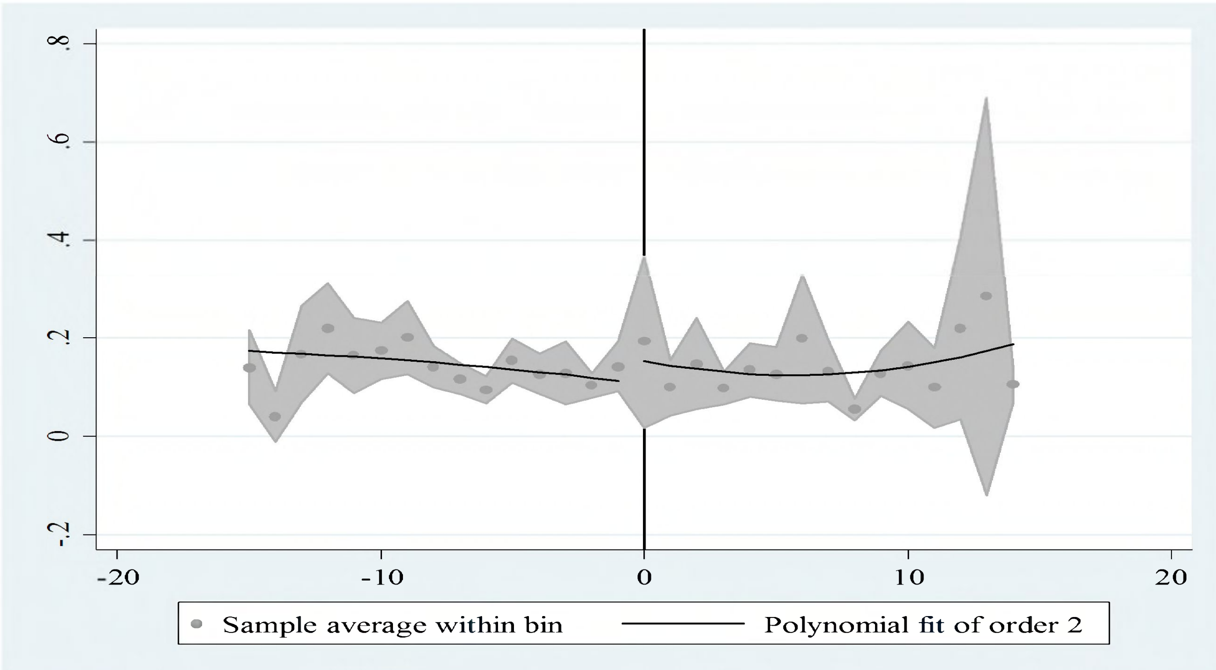
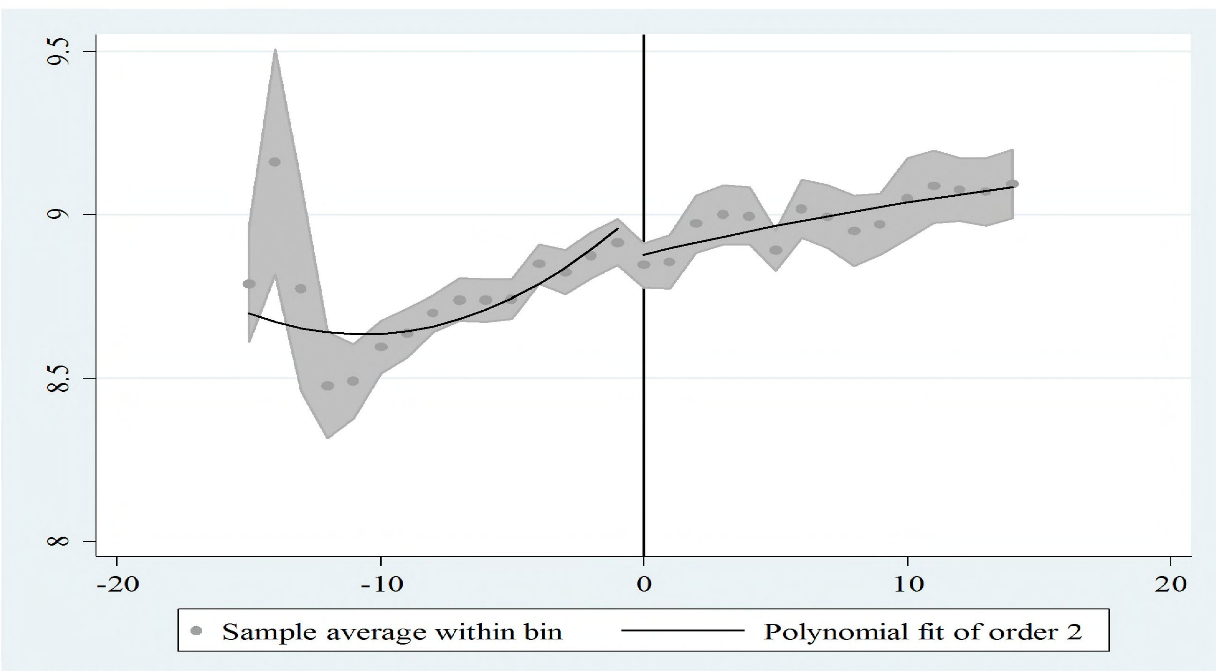


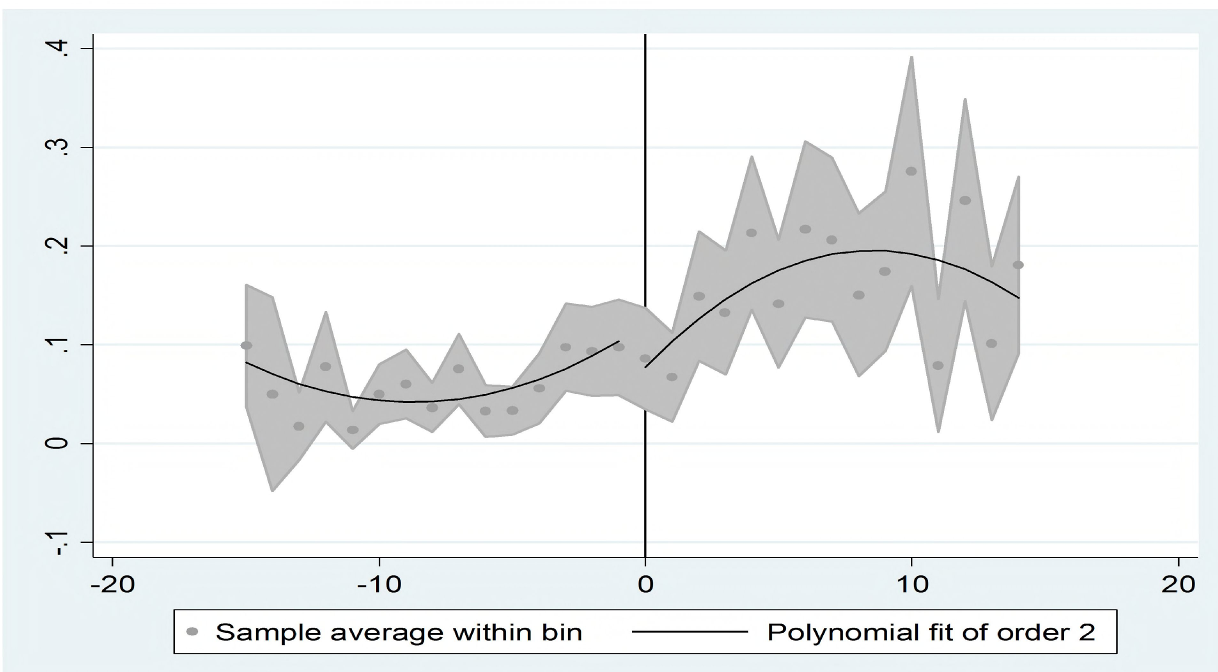
Figure 6: Average Log Directly hired wage bill



Source: EAM 2004. Figure plots local linear polynomial of order 2 of the outcome variable for treatment and control firms. Dots represent the average value of the outcome at each estimation bin. Shaded areas contain a 95% confidence interval for the estimation.

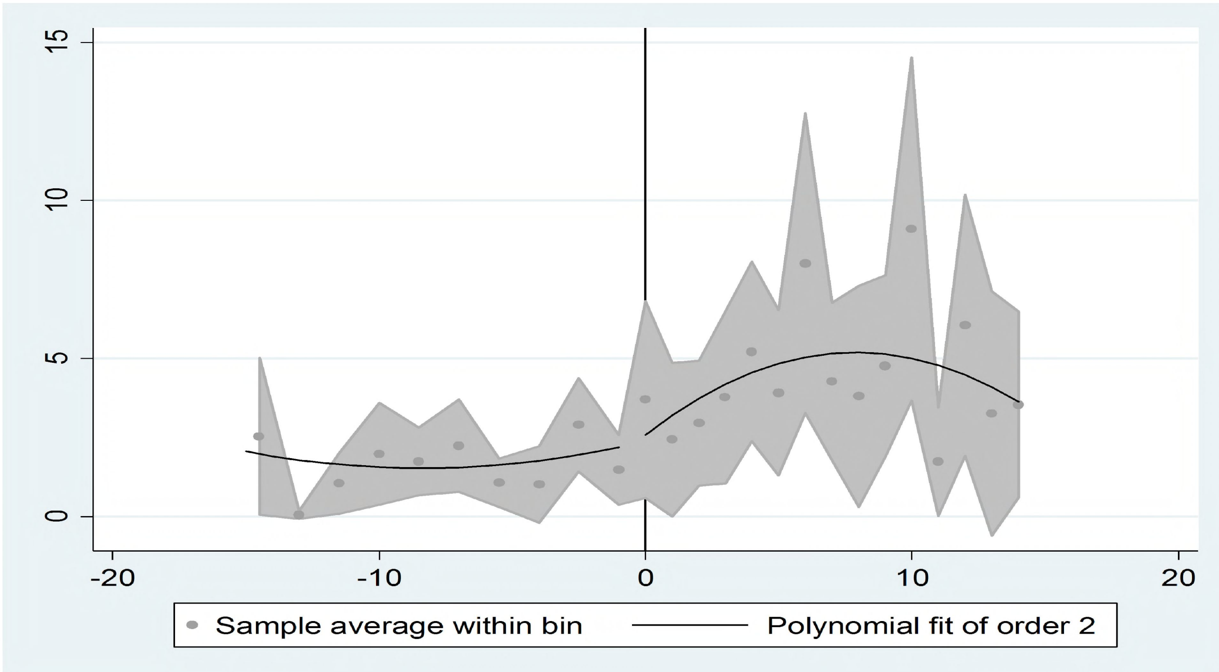
Figure 7: Average Log Outsourced hired wage bill

Figure 8: Average Share of exporters



Source: EAM 2004. Figure plots local linear polynomial of order 2 of the outcome variable for treatment and control firms. Dots represent the average value of the outcome at each estimation bin. Shaded areas contain a 95% confidence interval for the estimation.

Figure 9: Average Fraction of sales exported



Source: EAM 2004. Figure plots local linear polynomial of order 2 of the outcome variable for treatment and control firms. Dots represent the average value of the outcome at each estimation bin. Shaded areas contain a 95% confidence interval for the estimation.

No manipulation and local continuity

In this section I show the results of the McCrary (2008) test of no manipulation. Local continuity tests of baseline covariates, namely intermediate inputs, energy consumption and firm age are presented in the Appendix[13].

Table 1 shows the results of performing a parametric version of McCrary (2008) test of no manipulation by estimating model (1) on a sample of the average values of each variable for each level of $N_d \in (5, 24)$. The dependent variable is the Log number of firms at each level of direct labor demand. The parameter of interest is D_{15} which test whether the (log) number of firms is statistically different before and after the regulation threshold. The coefficient for the treatment variable is not significant for the $+/-6$ and $+/-4$ samples. This implies that the hypothesis that the assignment variable is continuous at the threshold of compliance with the regulation cannot be rejected and provides evidence of no manipulation of the running variable. In section 6 of the Appendix I show that the test is robust to using the number of firms at each level of direct labor demand, and that the result holds for the $+/-4$ sample when D_{15} is defined using the observed direct labor demand in the year 2004. As a further robustness check, Table 4 shows that the no manipulation hypothesis is valid when the test is performed using the local polynomial density estimation introduced by the rddensity command (Cattaneo et al., 2016). This particular test fails to reject the hypothesis that the distribution of direct labor do not differ at the threshold for the years 2002 and 2004. This confirms our findings of the parametric McCrary (2008) test of no manipulation. However it rejects the null hypothesis of no manipulation for the year 2003. For the years 2002 and 2004 the test selected a data-driven bandwidths of [5.521, 5.527] and [5.978, 5.987] to the left and and right of

[13]Firm age showed a statistically significant difference of one year. Firms that in 2002 had a labor demand of directly hired workers that would make them subject to the apprenticeship contract regulation had been established a year before firms that would not be subject to it.

Table 1: Parametric McCrary (2008) test of no manipulation

VARIABLES VARIABLES	(1) LN_bin	(2) LN_bin	(3) LN_bin	(4) LN_bin
$D_{15} = I(N_d \geq 0)$	-0.148***	-0.125	-0.252**	-0.011
	[0.042]	[0.088]	[0.108]	[0.046]
N_d=Normalized Directly hired demand	-0.086***	-0.110	0.002	-0.263***
	[0.024]	[0.073]	[0.094]	[0.051]
$N_d^2 = N_d$ Squared	-0.003	-0.007	0.013	-0.044***
	[0.002]	[0.010]	[0.015]	[0.010]
$D_{15} \times N_d$	0.096**	0.098	0.010	0.291***
	[0.035]	[0.074]	[0.095]	[0.053]
$D_{15} \times N_d^2$	-0.003	0.008	-0.019	0.033**
	[0.004]	[0.010]	[0.015]	[0.011]
Constant	5.009***	4.985***	5.104***	4.859***
	[0.035]	[0.087]	[0.107]	[0.046]
Observations	3,304	1,919	1,582	1,225
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4
model	ols	ols	ols	ols

Robust standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Treatment: Firms with 15-29 direct workers in 2002. Control: Firms with less than 15 workers in 2002. Dependent variable is the Log number of firms at each size level. Regression includes a second degree polynomial and interactions with treatment variable. Standard errors clustered at the assignment variable level.

the threshold for each year respectively.

4.2 Estimation results

Table 2 shows descriptive statistics of outcome variables for the main estimation sample.

Table 3 shows the results of estimating equation (1) for the first threshold of compliance with the apprenticeship contract regulation. All columns include a second degree polynomial and interactions with the treatment variable, which is measured in the year 2002. Columns (1)-(4) do not include any controls, columns (5)-(8) include baseline controls and columns (9)-(12) add industry indicators. Baseline controls, measured in 2002, are: The log value of intermediate materials used in production, the log value of electrical energy consumption in production and firm age, measured as the number of years since it was created. Each group of regressions uses different samples around the threshold of compliance with the apprenticeship contract. My preferred specification is column (12) which controls for baseline covariates, industry of economic activity and uses the sample of firms between 11 and 18 directly hired workers in 2002 (the $+/-4$ sample). Recall that the estimated coefficients must be interpreted as intent-to-treat effects.

Results in Panel A, show that the apprenticeship contract had a positive effect on output per worker. It increased labor productivity by 10 log points. The increase in output per worker is consistent with the fact that firms subject to the regulation reduced their total number of workers (Ospino, 2016). Panel B shows evidence of substitution of labor for capital at the margin, which is consistent with the findings of theoretical size-dependent distortion models. These models predict

Table 2: Mean and standard deviations for outcome variables. +/- 4 sample.

Treatment Status	(log) Output per worker	(log) Capital per worker	TFP (De Loecker & Warzynski (2012))	TFP (Eslava et al (2004))	(log) Investment	Skilled /Unskilled Ratio	(log) Average Wage Bill	(log) Average Wage Bill-Direct labor	(log) Average Wage Bill-Outsourced labor	Exports	Share of Exports
Direct labor<15	10.804	9.528	1.913	3.334	9.431	0.125	9.293	8.863	9.109	0.086	2.109
	[0.949]	[1.283]	[0.259]	[1.106]	[1.500]	[0.306]	[0.434]	[0.435]	[0.425]	[0.280]	[11.213]
Direct labor≥15	10.930	9.591	1.896	3.461	9.736	0.134	9.342	8.916	9.060	0.108	3.217
	[0.999]	[1.236]	[0.524]	[0.983]	[1.553]	[0.563]	[0.431]	[0.451]	[0.546]	[0.311]	[14.204]
Total	10.857	9.555	1.906	3.388	9.561	0.129	9.314	8.885	9.083	0.095	2.573
	[0.972]	[1.263]	[0.393]	[1.058]	[1.529]	[0.433]	[0.433]	[0.443]	[0.494]	[0.294]	[12.558]

that at the margin firms are constrained in their labor demand and will substitute labor for capital (Guner et al., 2008) or other untaxed workers (Ospino, 2016). The estimated effect is an increase of 52 log points in capital per worker. Panel C shows a positive effect of the apprenticeship contract regulation on firm TFP. Total factor productivity increased by 2 log points as a result of the apprenticeship contract for firms subject to the regulation[14]. Panel D shows that once I control for industry, the negative effect of the apprenticeship contract on firms investment disappear. Panel E shows that being subject to the apprenticeship contract regulation did not affect the ratio of skilled to unskilled labor. This result provides evidence that labor substitution of direct for outsourced labor did not affect the skill composition of production workers. Panel F shows that the apprenticeship contract reduced the average wage bill of directly hired workers by 9 log points which suggests that productivity gains were not shared with workers through higher wages. In contrast, Panel G shows that the regulation did not have an effect on the average wage bill of outsourced labor. Finally Panels H and I show that the while the apprenticeship contract had a positive effect on TFP it did not increase the likelihood of a firm becoming and exporter, however, it increased the share of sales exported in 2 percentage points by firms already exporting.

The productivity and export impacts may seem small. However recall that the set of firms where these impacts were observed are those around the threshold of 15 directly hired workers. The fact that there's a positive effect for small firms of having to comply with this regulation suggests that expanding it to all firms would not have a negative effect on firm's performance and could further increase the productivity of the whole manufacturing sector and apprenticeship slots for more workers.

[14]The coefficient when using TFP estimated from the parameter obtained by Eslava et al. (2004) was 0.213 with a standard error of 0.056.

Table 3: Parametric RDD estimation results

Panel A: Log Output per worker

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(N_d \geq 15 \| t = 2002)$	-0.017	0.050	0.160	-0.044	-0.050	-0.056	0.103*	-0.014	-0.005	0.015	0.165**	0.101**
	[0.087]	[0.093]	[0.103]	[0.055]	[0.054]	[0.071]	[0.046]	[0.024]	[0.063]	[0.077]	[0.052]	[0.029]
Observations	2,855	1,716	1,423	1,104	2,642	1,662	1,392	1,077	2,642	1,662	1,392	1,077
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4
Baseline controls	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES
Industry controls	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
model	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols

Panel B: Log Capital per worker

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(N_d \geq 15 \| t = 2002)$	-0.265*	-0.124	-0.062	0.063	-0.227**	-0.130	-0.013	0.159	-0.097	0.128	0.303**	0.516***
	[0.129]	[0.099]	[0.116]	[0.131]	[0.102]	[0.105]	[0.119]	[0.136]	[0.128]	[0.130]	[0.131]	[0.138]
Observations	2,766	1,681	1,397	1,084	2,573	1,633	1,369	1,060	2,573	1,633	1,369	1,060
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4
Baseline controls	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES
Industry controls	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
model	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols

Panel C: Log Total Factor Productivity

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(N_d \geq 15 \| t = 2002)$	-0.001	-0.006	-0.028	0.012	-0.028**	-0.008	-0.020	0.001	-0.002	0.014	-0.009	0.028**
	[0.038]	[0.046]	[0.043]	[0.023]	[0.013]	[0.009]	[0.012]	[0.012]	[0.013]	[0.014]	[0.018]	[0.011]
Observations	2,684	1,635	1,358	1,052	2,535	1,614	1,353	1,048	2,535	1,614	1,353	1,048
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4
Baseline controls	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES
Industry controls	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
model	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols

Panel D: Log Investment

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(N_d \geq 15 \| t = 2002)$	-0.549***	-0.574***	-0.410***	-0.456***	-0.432***	-0.561***	-0.318***	-0.276***	-0.369***	-0.393**	-0.115	0.017
	[0.095]	[0.123]	[0.092]	[0.072]	[0.067]	[0.123]	[0.037]	[0.023]	[0.100]	[0.160]	[0.102]	[0.211]
Observations	2,415	1,462	1,206	946	2,261	1,419	1,182	925	2,261	1,419	1,182	925
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4
Baseline controls	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES
Industry controls	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
model	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols

Panel E: Skilled/Unskilled Ratio

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(N_d \geq 15 \| t = 2002)$	0.003	0.064	0.013	0.007	0.013	0.056	0.005	0.003	0.045	0.096**	0.034	0.062
	[0.038]	[0.041]	[0.022]	[0.029]	[0.037]	[0.037]	[0.021]	[0.025]	[0.037]	[0.042]	[0.044]	[0.069]
Observations	2,793	1,680	1,390	1,078	2,584	1,632	1,366	1,058	2,584	1,632	1,366	1,058
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4
Baseline controls	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES

Continued on next page

– continued from previous page

Industry controls	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
model	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols
Panel F: Log Average wage bill-Direct labor												
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(N_d \geq 15 \mid t = 2002)$	-0.084***	-0.132**	-0.095*	-0.178***	-0.098***	-0.144***	-0.107**	-0.183***	-0.057*	-0.078*	-0.015	-0.092**
	[0.027]	[0.044]	[0.046]	[0.027]	[0.028]	[0.032]	[0.035]	[0.004]	[0.028]	[0.039]	[0.041]	[0.034]
Observations	2,810	1,691	1,401	1,091	2,602	1,638	1,371	1,065	2,602	1,638	1,371	1,065
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4
Baseline controls	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES
Industry controls	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
model	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols
Panel G: Log Average Wage bill-Outsourced labor												
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(N_d \geq 15 \mid t = 2002)$	-0.262***	-0.246**	-0.261***	-0.269***	-0.280***	-0.310**	-0.228***	-0.273***	-0.296*	-0.161	-0.124	-0.121
	[0.054]	[0.087]	[0.044]	[0.055]	[0.068]	[0.100]	[0.065]	[0.067]	[0.142]	[0.239]	[0.248]	[0.326]
Observations	379	245	204	154	364	236	196	148	364	236	196	148
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4
Baseline controls	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES
Industry controls	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
model	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols
Panel H: Exporter												
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(N_d \geq 15 \mid t = 2002)$	-0.060**	-0.044	-0.003	0.009	-0.063**	-0.052*	-0.012	-0.001	-0.089***	-0.071**	-0.046**	-0.038
	[0.025]	[0.027]	[0.015]	[0.019]	[0.024]	[0.024]	[0.010]	[0.014]	[0.031]	[0.030]	[0.019]	[0.022]
Observations	2,862	1,715	1,421	1,103	2,638	1,658	1,389	1,075	2,638	1,658	1,389	1,075
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4
Baseline controls	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES
Industry controls	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
model	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols
Panel I: Share of exports												
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(N_d \geq 15 \mid t = 2002)$	0.065	1.420	3.918***	5.691***	0.079	0.991	3.398***	5.137***	-1.264	-0.653	1.132	1.984**
	[1.197]	[1.364]	[0.970]	[1.074]	[1.165]	[1.319]	[0.769]	[0.558]	[1.293]	[1.468]	[0.901]	[0.745]
Observations	2,862	1,715	1,421	1,103	2,638	1,658	1,389	1,075	2,638	1,658	1,389	1,075
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4
Baseline controls	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES
Industry controls	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
model	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols

Standard errors clustered at the assignment variable level shown in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
Treatment: Firms with 15-29 direct workers in 2002. Control: Firms with less than 15 workers in 2002.
Regression includes a second degree polynomial and interactions with treatment variable.

As a robustness check, in the Appendix, I show that most outcome variables do not show significant effects in the year 2002, for different bandwidth samples. The two exceptions are, the capital per worker ratio and firm investment. These two variables showed positive coefficients of 57.4 and 39.5 log points respectively, which suggests firms capital accumulation decisions might have been affected by other policies or that firms reacted in expectation to the regulation substituting capital for labor. For the other variables, the lack of anticipation effects provide further assurance that the results found appear to be the effect of the policy and not of firms decisions before the regulation was introduced. I also carried out the analysis using the sample for the year 2003. As discussed, given that labor substitution as a result of the policy did not take place until 2004 (Ospino, 2016), I did not expect to find significant effects. Section ?? in Appendix shows statistically significant effects for capital per worker (72.6 log points), investment (49.6 log points), the skilled/unskilled ratio (12.1 log points) and the share of exports (3.9 percentage points).

Finally, I estimated the effects following Cattaneo et al. (2016) using the rdrandinf package. These results can be found in Table 11 where I show estimations for a second degree polynomial for the same bandwidth as the main estimation (+4/-4) and the optimal selected bandwidth by the package (+2/-2). These results are consistent but of a higher magnitude for the average wage bill of direct workers and the share of exports. Exporting status and the share of skilled and unskilled workers are not significant as in the main results. However, output per worker and total factor productivity which show positive and statistically significant effects in the main results are not statistically significant in this estimation[15]. Another difference is that investment had a positive effect in the main results while in this estimation the effect is negative and statistically significant. Despite the differences my preferred results are the ones in Table 3. One reason is that I'm able to control for industry fixed effects and still obtain results that speak about the manufacturing sector as a whole. Second, in Tables 14-17 in the Appendix I provide evidence about the presence of heterogeneous effects by industry for the outcomes I can be confident about obtaining causal effects following Cattaneo et al. (2016). These results provide support for the empirical strategy where I control for industry fixed effects.

5 Discussion

In this paper I have used parametric regression discontinuity design (RDD) methods to evaluate the impact of the apprenticeship contract on Colombian small firm dynamics. Firms showed statistical significant differences at the moment when the regulation changed and most outcome variables do not follow a common trend before the regulation was reformed. Therefore, a difference-in-difference approximation which at first seemed appealing could not be used for this evaluation. Nevertheless, the assumptions for implementing a regression discontinuity design held, and therefore I proceeded to carry out the evaluation using an intent-to-treat regression discontinuity design.

Results showed positive effects on output per worker, total factor productivity, and the share of exports; it showed negative effects on the direct labor average wage bill. The increase in productivity measured by output per worker and total factor productivity suggests that the policy could have benefited firms by increasing the skill component of their labor force. The increase in sales exported points in the same direction. The absence of effects on the skill composition of labor suggests labor productivity did not increase by firing unskilled labor. However, the negative effect on the average

[15] TFP calculated using the coefficients from Eslava et al. (2004) showed statistically significant effects of a similar magnitude that when I used the parametric estimator in Equation 1.

wage bill of directly hired workers and the absence of effects on the average wage bill of outsourced workers suggests that such productivity gains were not shared by firm workers through higher wages.

The policy implication of this paper is that having small firms being subject to the apprenticeship contract regulation had on average a positive effect on productivity for this group of firms which reflected in higher output per worker and a higher share of exports. From the worker's perspective the effect is a negative one since firm's wage bill per directly hired worker decreased approximately by 9%. This result suggests that workers which determine whether firms are subject to the apprenticeship contract regulation or new hires could be the ones bearing the cost of the regulation design. A rigorous evaluation of this regulation for apprentices and regular workers employment status and wages is a pending subject in Colombia to have a thorough assessment of the costs and benefits of the apprenticeship contract regulation from the worker's perspective.

Bibliography

Acemoglu, D. and J.-S. Pischke (1999). Beyond becker: Training in imperfect labour markets. *The Economic Journal 109*(453), 112–142.

Bertrand, M., C. Hsieh, and N. Tsivanidis (2015). Contract labor and firm growth in india. *Becker Friedman Institute Working Paper Series*.

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

Braguinsky, S., L. G. Branstetter, and A. Regateiro (2011). The incredible shrinking portuguese firm. *National Bureau of Economic Research Working Paper Series No. 17265.*

Cattaneo, M. D., M. Jansson, and X. Ma (2016). rddensity : Manipulation Testing based on Density Discontinuity. *The Stata Journal* (ii), 1–18.

Cattaneo, M. D., R. Titiunik, and G. Vazquez-Bare (2016). Inference in Regression Discontinuity Designs under Local Randomization. *The Stata Journal 16*(2), 331–367.

De Loecker, J. (2013). Detecting learning by exporting. *American Economic Journal: Microeconomics 5*(3), 1–21.

De Loecker, J. and F. Warzynski (2012, oct). Markups and firm-level export status. *American Economic Review 102*(6), 2437–2471.

Dearden, L., H. Reed, and J. VanReenen (2006). The impact of training on productivity and wages: Evidence from british panel data. *Oxford Bulletin of Economics and Statistics 68*(4), 397–421.

Eslava, M., J. Haltiwanger, A. Kugler, and M. Kugler (2004). The effects of structural reforms on productivity and profitability enhancing reallocation: evidence from colombia. *Journal of Development Economics 75*(2), 333–371.

Fazio, M. V., R. Fernández-Coto, and L. Ripani (2016). *Apprenticeships for the XXI Century.* Washington, D.C.: Inter-American Development Bank.

Garicano, L., C. LeLarge, and J. Van Reenen (2013). Firm size distortions and the productivity distribution: Evidence from france. *National Bureau of Economic Research Working Paper Series No. 18841.*

Gelman, A. and G. Imbens (2014). Why high-order polynomials should not be used in regression discontinuity designs. Technical report, National Bureau of Economic Research.

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

Imbens, G. W. and T. Lemieux (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics 142*(2), 615 – 635. The regression discontinuity design: Theory and applications.

Konings, J. and S. Vanormelingen (2010). The impact of training on productivity and wages: Firm level evidence.

Lee, D. S. and D. Card (2008). Regression discontinuity inference with specification error. *Journal of Econometrics 142*(2), 655 – 674. The regression discontinuity design: Theory and applications.

Lee, D. S. and T. Lemieux (2009, February). Regression discontinuity designs in economics. Working Paper 14723, National Bureau of Economic Research.

Maloney, W. and J. Mendez (2004). Measuring the impact of minimum wages. evidence from latin america. In *Law and Employment: Lessons from Latin America and the Caribbean*, pp. 109–130. University of Chicago Press.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics 142*(2), 698 – 714. The regression discontinuity design: Theory and applications.

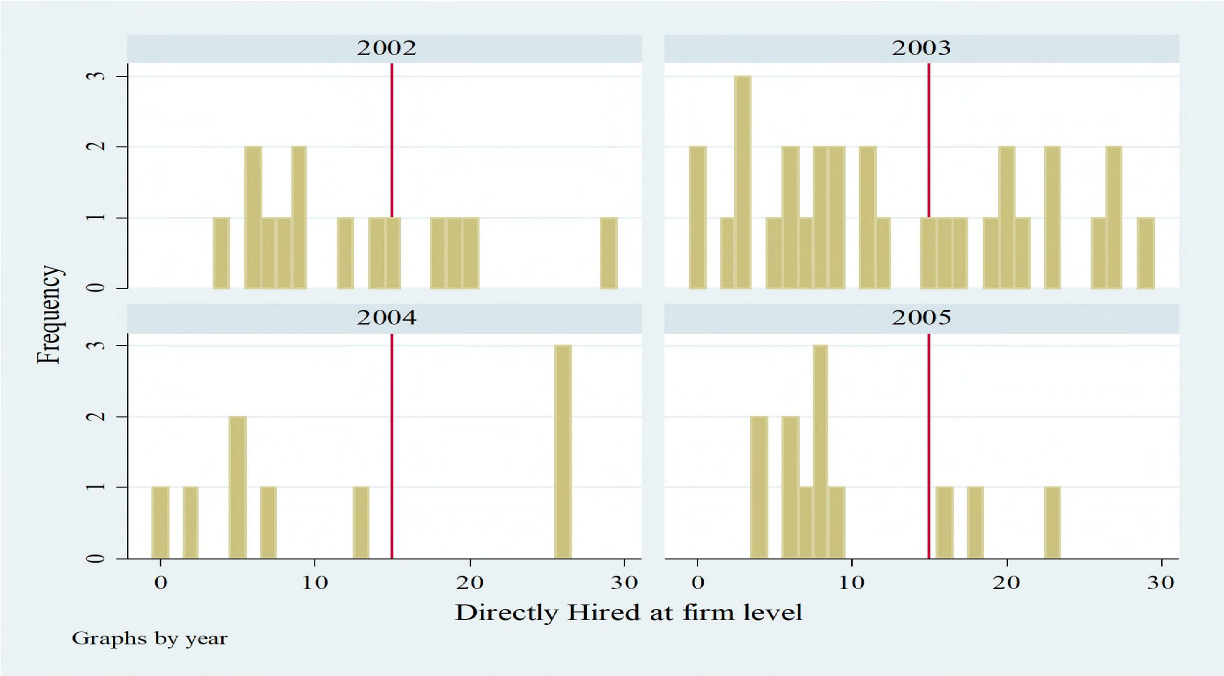
Mohrenweiser, J. and T. Zwick (2009). Why do firms train apprentices? the net cost puzzle reconsidered. *Labour Economics 16*(6), 631–637.

Mondragón-Vélez, C., X. Peña, D. Wills, and A. Kugler (2010). Labor market rigidities and informality in colombia [with comment]. *Economía*, 65–101.

Ospino, C. (2016). Size-dependent policies and labor substitution: Misallocation, productivity and firm size distribution.

6 Appendix

Figure 10: Distribution of firms according to number of directly hired workers



Source: EAM 2002-2005. Figure plots the distribution of new firms in the sample according to the number of directly hired workers. Bins are of size one.

Table 4: RD Manipulation Test using local polynomial density estimation

Year	Test Statistic	P>T
2002	-0.517	0.605
2003	-2.729	0.006
2004	-1.051	0.293

Note: Test carried out using rddensity Cattaneo et al. (2016).

Parametric McCrary (2008) tests

Local continuity of baseline covariates

Table 5: Parametric McCrary (2008) test of no manipulation. Year 2002

	(1)	(2)	(3)	(4)
$D_{15} = I(N_d \geq 0)$	-17.415**	-13.209	-41.486*	6.210
	[8.113]	[17.907]	[21.161]	[8.976]
N_d =Direct labor demand	-15.579***	-20.037	5.410	-46.733***
	[5.353]	[14.849]	[18.564]	[9.926]
$N_d^2 = N_d$ Squared	-0.366	-1.208	3.387	-7.909***
	[0.514]	[2.003]	[2.883]	[1.972]
$D_{15} \times N_d$	16.223**	18.599	-3.926	50.267***
	[5.987]	[14.923]	[18.603]	[10.093]
$D_{15} \times N_d^2$	-0.286	1.347	-4.110	6.395**
	[0.601]	[2.021]	[2.899]	[2.061]
Constant	146.599***	142.232***	169.472***	121.297***
	[7.637]	[17.783]	[21.131]	[8.955]
Observations	3,304	1,919	1,582	1,225
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4
model	ols	ols	ols	ols

Robust standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. $D_{15} = 1$: Firms with 15-29 direct workers in 2002. $D_{15} = 0$: Firms with less than 15 workers in 2002. Dependent variable is the number of firms at each level of direct labor demand. Regression includes a second degree polynomial and interactions with direct labor demand in 2002. Standard errors clustered at the assignment variable level.

Table 6: Parametric McCrary (2008) test of no manipulation. Year 2004

	(1)	(2)	(3)	(4)
VARIABLES				
$D_{15} = I(N_d \geq 0)$	-1.003***	-1.051***	-1.044**	0.049
	[0.281]	[0.240]	[0.390]	[0.467]
N_d =Direct labor demand	-0.031	-0.171***	-0.136**	-0.230**
	[0.034]	[0.019]	[0.056]	[0.069]
$N_d^2 = N_d$ Squared	-0.002	-0.023***	-0.016	-0.048**
	[0.003]	[0.003]	[0.010]	[0.016]
$D_{15} \times N_d$	0.113	0.442***	0.371**	0.187
	[0.069]	[0.070]	[0.124]	[0.160]
$D_{15} \times N_d^2$	-0.008	-0.003	-0.009	0.104*
	[0.006]	[0.009]	[0.021]	[0.052]
Constant	5.365***	5.331***	5.339***	4.839***
	[0.122]	[0.105]	[0.180]	[0.215]
Observations	3,015	1,798	1,487	1,152
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4
model	ols	ols	ols	ols

Robust standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. $D_{15} = 1$: Firms with 15-29 direct workers in 2004. $D_{15} = 0$: Firms with less than 15 workers in 2004. Dependent variable is the Log number of firms at each level of direct labor demand. Regression includes a second degree polynomial and interactions with direct labor demand in 2002. Standard errors clustered at the assignment variable level.

Table 7: Parametric McCrary (2008) test of no manipulation. Year 2004

VARIABLES	(1)	(2)	(3)	(4)
$D_{15} = I(N_d \geq 0)$	-132.406**	-177.880***	-168.462**	43.142
	[47.525]	[48.320]	[70.207]	[74.835]
N_d =Direct labor demand	-9.632*	-30.621***	-19.063*	-38.432**
	[5.338]	[4.592]	[9.338]	[11.409]
$N_d^2 = N_d$ Squared	-0.641	-3.962***	-1.739	-7.942**
	[0.455]	[0.763]	[1.726]	[2.674]
$D_{15} \times N_d$	24.298**	84.884***	59.580**	26.822
	[10.149]	[16.087]	[20.428]	[23.560]
$D_{15} \times N_d^2$	-0.726	-1.648	-3.321	18.081*
	[0.913]	[1.617]	[3.885]	[8.519]
Constant	194.730***	207.498***	206.376***	109.593**
	[20.145]	[21.393]	[32.283]	[34.863]
Observations	3,015	1,798	1,487	1,152
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4
model	ols	ols	ols	ols

Robust standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. $D_{15} = 1$: Firms with 15-29 direct workers in 2004. $D_{15} = 0$: Firms with less than 15 workers in 2004. Dependent variable is the number of firms at each level of direct labor demand. Regression includes a second degree polynomial and interactions with direct labor demand in 2002. Standard errors clustered at the assignment variable level.

Figure 11: Local Continuity Test: Log Value of raw materials

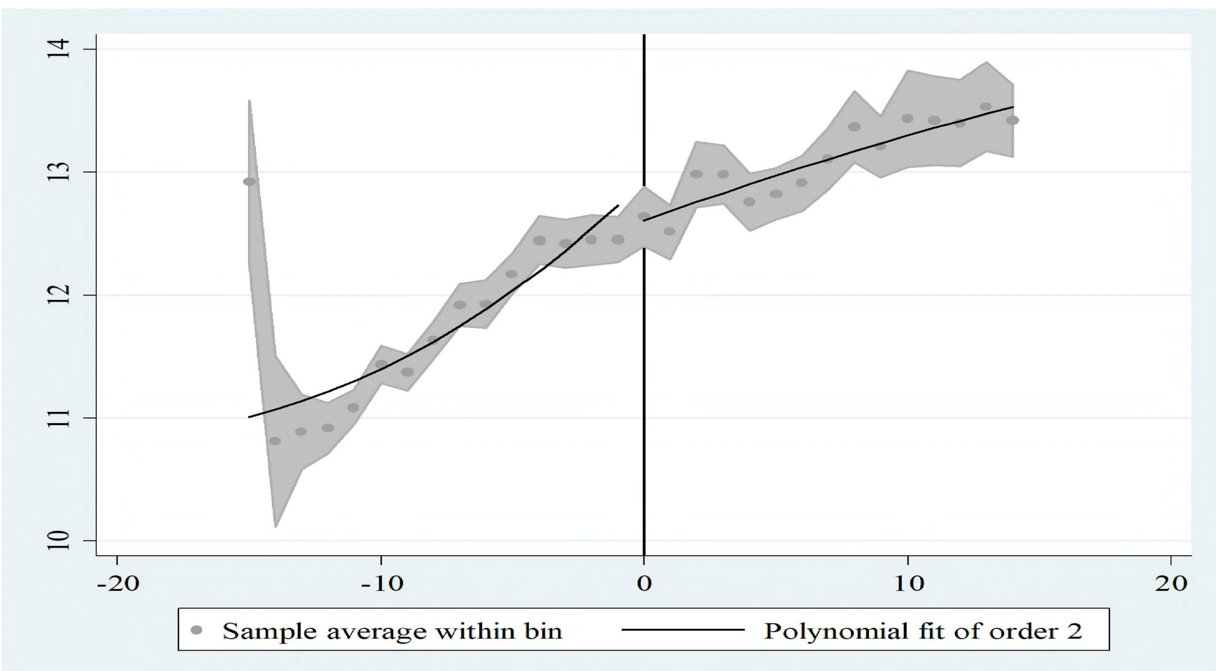


Table 8: Local Continuity Test: Log Value of raw materials

VARIABLES	(1)	(2)	(3)	(4)
$D_{15} = I(N_d \geq 0)$	0.060	0.264*	0.228	0.090
	[0.126]	[0.128]	[0.143]	[0.070]
N_d =Direct labor demand	0.041	-0.176***	-0.158**	0.046
	[0.044]	[0.054]	[0.069]	[0.025]
$N_d^2 = N_d$ Squared	-0.009*	-0.039***	-0.036***	0.008
	[0.004]	[0.007]	[0.011]	[0.005]
$D_{15} \times N_d$	0.006	0.382***	0.412**	0.001
	[0.059]	[0.099]	[0.146]	[0.197]
$D_{15} \times N_d^2$	0.011*	0.007	-0.010	0.026
	[0.006]	[0.017]	[0.034]	[0.064]
Constant	12.579***	12.286***	12.305***	12.493***
	[0.090]	[0.070]	[0.081]	[0.022]
Observations	3,197	1,869	1,540	1,189
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4
model	ols	ols	ols	ols

Robust standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Treatment: Firms with 15-29 direct workers in 2002. Control: Firms with less than 15 workers in 2002. Regression includes a second degree polynomial and interactions with treatment variable. Standard errors clustered at the assignment variable level shown in square brackets.

Figure 12: Local Continuity Test: Log Energy Consumption

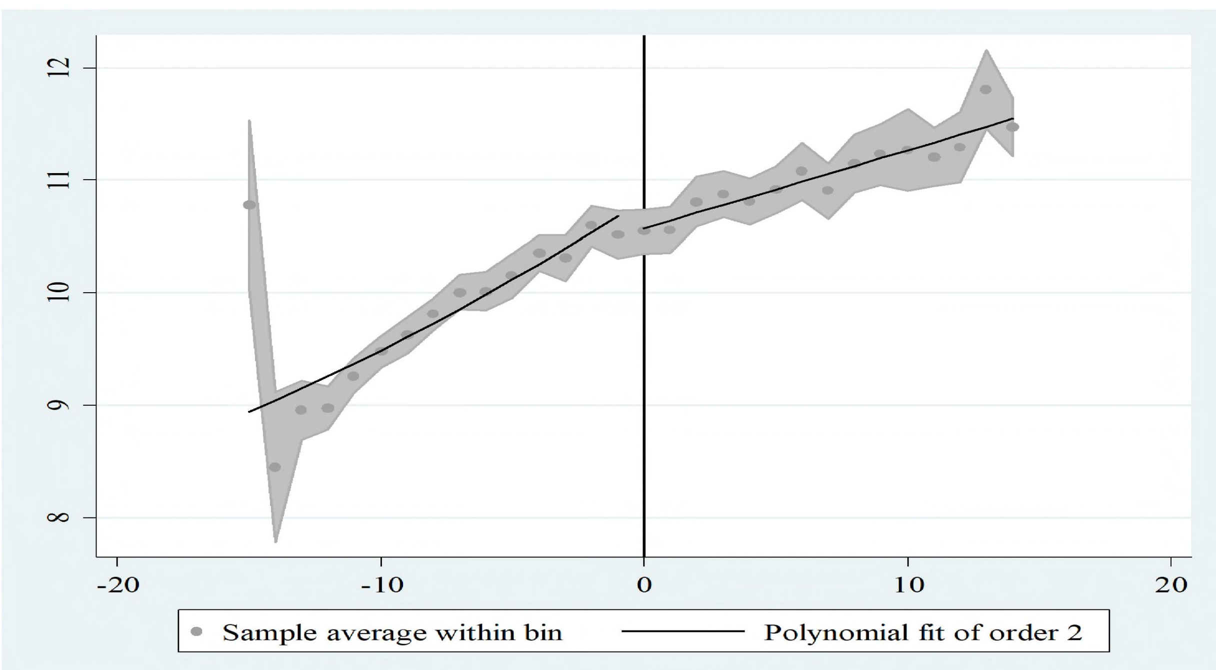


Table 9: Local Continuity Test: Log Energy Consumption

VARIABLES	(1)	(2)	(3)	(4)
$D_{15} = I(N_d \geq 0)$	-0.069	-0.077	-0.070	-0.090
	[0.086]	[0.093]	[0.120]	[0.177]
N_d =Direct labor demand	0.059**	0.024	0.006	0.053
	[0.026]	[0.061]	[0.106]	[0.195]
$N_d^2 = N_d$ Squared	-0.006***	-0.012	-0.015	-0.005
	[0.002]	[0.009]	[0.018]	[0.039]
$D_{15} \times N_d$	0.028	0.127*	0.176	0.031
	[0.032]	[0.070]	[0.118]	[0.208]
$D_{15} \times N_d^2$	0.004	-0.003	-0.009	0.019
	[0.003]	[0.011]	[0.022]	[0.045]
Constant	10.619***	10.586***	10.567***	10.610***
	[0.077]	[0.083]	[0.110]	[0.176]
Observations	3,293	1,914	1,577	1,220
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4
model	ols	ols	ols	ols

Robust standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Treatment: Firms with 15-29 direct workers in 2002. Control: Firms with less than 15 workers in 2002. Regression includes a second degree polynomial and interactions with treatment variable. Standard errors clustered at the assignment variable level shown in square brackets.

Figure 13: Local Continuity Test: Firm Age

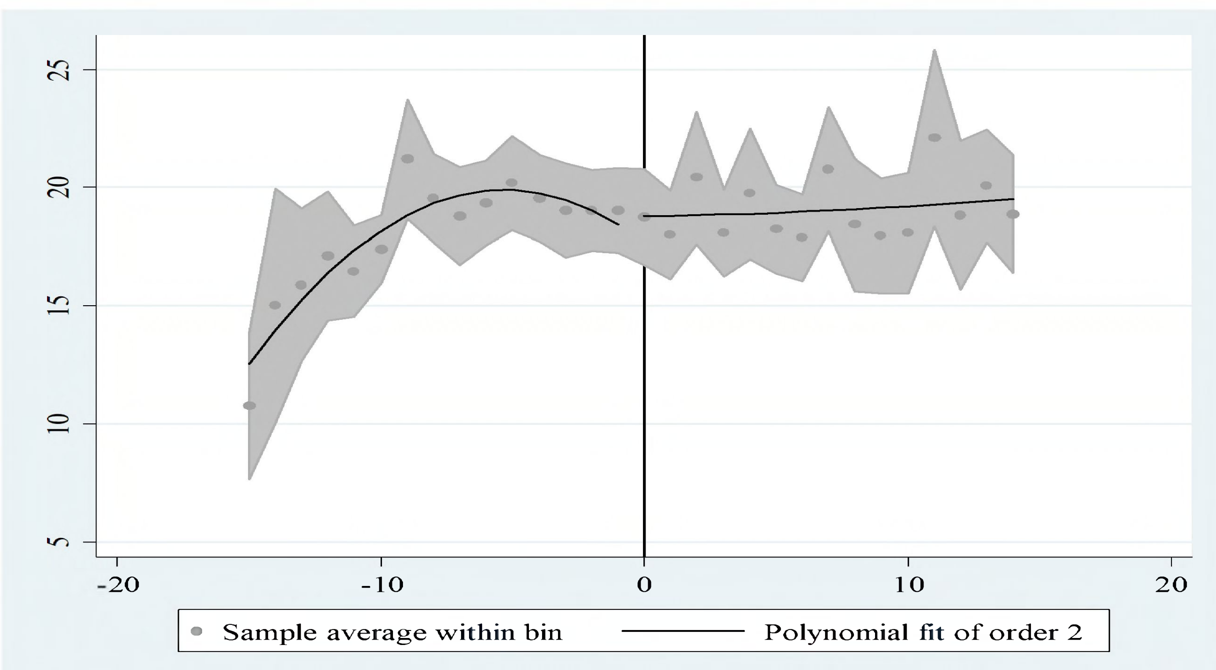


Table 10: Local Continuity Test: Firm Age

VARIABLES	(1)	(2)	(3)	(4)
$D_{15} = I(N_d \geq 0)$	0.366	-0.028	-0.833**	-1.068**
	[0.732]	[0.610]	[0.338]	[0.478]
N_d =Direct labor demand	-0.543*	-0.402	0.492***	0.496***
	[0.326]	[0.317]	[0.080]	[0.153]
$N_d^2 = N_d$ Squared	-0.050	-0.032	0.129***	0.130***
	[0.037]	[0.047]	[0.013]	[0.030]
$D_{15} \times N_d$	0.778*	1.016	-0.300	0.684
	[0.414]	[0.708]	[0.947]	[1.359]
$D_{15} \times N_d^2$	0.021	-0.088	-0.125	-0.507
	[0.045]	[0.133]	[0.231]	[0.443]
Constant	18.219***	18.456***	19.411***	19.415***
	[0.561]	[0.420]	[0.086]	[0.137]
Observations	3,011	1,887	1,577	1,220
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4
model	ols	ols	ols	ols

Robust standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Treatment: Firms with 15-29 direct workers in 2002. Control: Firms with less than 15 workers in 2002. Regression includes a second degree polynomial and interactions with treatment variable. Standard errors clustered at the assignment variable level shown in square brackets.

Table 11: Results using Cattaneo et al. (2016)

Bandwidth	[-4,4]		[-2,2]	
Variable	Effect	P-Value	Effect	P-Value
Log Output per worker	-0.047	0.424	-0.268	0.508
Log Capital per worker	0.058	0.410	-0.626	0.468
Log Total Factor Productivity	0.026	0.244	0.207	0.622
Log Total Factor Productivity (EHHK 2004)	0.223	0.000	0.442	0.065
Log Investment	-0.479	0.000	-0.934	0.014
Skilled/Unskilled Ratio	0.007	0.783	-0.131	0.395
Log Average wage bill-Direct labor	-0.186	0.000	-0.215	0.018
Log Average Wage bill-Outsorced labor	-0.282	0.000	-0.355	0.155
Exporter	0.013	0.461	-0.116	0.587
Share of exports	5.622	0.000	1.251	0.005
Number of Observations			7908	

Table 12: Robustness Checks: Anticipation Effects

Panel A: Log Output per worker

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(N_d \geq 15 \mid t = 2002)$	0.022	0.155	0.186	0.151*	-0.026	-0.014	-0.007	0.021	0.023	0.054	0.042	0.066
	[0.109]	[0.101]	[0.110]	[0.065]	[0.056]	[0.038]	[0.042]	[0.026]	[0.053]	[0.039]	[0.049]	[0.065]
Observations	3,161	1,837	1,510	1,164	2,884	1,794	1,493	1,150	2,884	1,794	1,493	1,150
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4
Baseline controls	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES
Industry controls	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
model	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols

Panel B: Log Capital per worker

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(N_d \geq 15 \mid t = 2002)$	-0.299*	-0.191	-0.046	0.229**	-0.322**	-0.243	-0.101	0.190*	-0.192	0.046	0.200	0.574***
	[0.150]	[0.164]	[0.154]	[0.080]	[0.114]	[0.144]	[0.138]	[0.098]	[0.149]	[0.169]	[0.178]	[0.118]
Observations	2,944	1,782	1,476	1,141	2,727	1,729	1,447	1,117	2,727	1,729	1,447	1,117
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4
Baseline controls	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES
Industry controls	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
model	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols

Panel C: Log Total Factor Productivity

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(N_d \geq 15 \mid t = 2002)$	0.000	-0.069***	-0.056***	-0.076***	0.002	-0.014	-0.018	-0.057**	0.027**	0.005	-0.000	-0.012
	[0.028]	[0.011]	[0.014]	[0.019]	[0.015]	[0.019]	[0.023]	[0.022]	[0.013]	[0.012]	[0.017]	[0.023]
Observations	3,002	1,772	1,458	1,125	2,778	1,749	1,457	1,124	2,778	1,749	1,457	1,124
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4
Baseline controls	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES
Industry controls	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
model	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols

Panel D: Log Investment

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(N_d \geq 15 \mid t = 2002)$	-0.200	-0.038	0.238	0.169	-0.209	-0.094	0.218*	0.215*	-0.130	0.088	0.336*	0.395**
	[0.215]	[0.209]	[0.180]	[0.126]	[0.182]	[0.177]	[0.118]	[0.102]	[0.184]	[0.171]	[0.168]	[0.162]
Observations	2,754	1,653	1,373	1,065	2,524	1,600	1,342	1,040	2,524	1,600	1,342	1,040
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4
Baseline controls	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES
Industry controls	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
model	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols

Panel E: Skilled/Unskilled Ratio

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(N_d \geq 15 \mid t = 2002)$	-0.009	0.013	0.020	0.035	-0.004	0.014	0.019	0.036	0.052	0.075*	0.056	0.107
	[0.041]	[0.027]	[0.028]	[0.035]	[0.041]	[0.027]	[0.028]	[0.033]	[0.048]	[0.042]	[0.047]	[0.063]
Observations	3,217	1,875	1,543	1,194	2,891	1,816	1,515	1,171	2,891	1,816	1,515	1,171
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4
Baseline controls	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES

Continued on next page

– continued from previous page

Industry controls	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
model	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols
Panel F: Log Average wage bill-Direct labor												
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(N_d \geq 15 \| t = 2002)$	-0.057*	-0.035	0.016	-0.038	-0.072***	-0.043**	-0.009	-0.057***	-0.028	0.005	0.062	0.033
	[0.028]	[0.036]	[0.034]	[0.032]	[0.018]	[0.019]	[0.020]	[0.007]	[0.022]	[0.029]	[0.038]	[0.052]
Observations	3,264	1,911	1,576	1,220	2,941	1,842	1,538	1,188	2,941	1,842	1,538	1,188
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4
Baseline controls	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES
Industry controls	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
model	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols
Panel G: Log Average Wage bill-Outsorced labor												
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(N_d \geq 15 \| t = 2002)$	-0.085	-0.013	-0.091	0.396***	-0.271	-0.070	-0.033	0.292*	-0.286	0.126	0.213	0.445
	[0.181]	[0.225]	[0.252]	[0.054]	[0.212]	[0.218]	[0.223]	[0.139]	[0.338]	[0.380]	[0.543]	[0.309]
Observations	257	168	144	104	246	163	140	102	246	163	140	102
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4
Baseline controls	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES
Industry controls	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
model	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols
Panel H: Exporter												
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(N_d \geq 15 \| t = 2002)$	0.029	0.051	0.091*	0.044	0.020	0.046	0.074*	0.024	-0.000	0.016	0.046	0.006
	[0.033]	[0.042]	[0.045]	[0.054]	[0.030]	[0.034]	[0.040]	[0.048]	[0.035]	[0.048]	[0.054]	[0.082]
Observations	3,185	1,839	1,509	1,164	2,881	1,792	1,491	1,149	2,881	1,792	1,491	1,149
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4
Baseline controls	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES
Industry controls	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
model	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols
Panel I: Share of exports												
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(N_d \geq 15 \| t = 2002)$	2.333**	4.505***	6.337***	3.608***	2.289**	4.641***	5.883***	2.966***	0.968	2.773**	3.918**	0.366
	[0.951]	[1.148]	[1.275]	[0.355]	[0.950]	[1.032]	[1.310]	[0.204]	[0.857]	[1.159]	[1.552]	[1.146]
Observations	3,185	1,839	1,509	1,164	2,881	1,792	1,491	1,149	2,881	1,792	1,491	1,149
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4
Baseline controls	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES
Industry controls	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
model	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols

Standard errors clustered at the assignment variable level shown in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
Treatment: Firms with 15-29 direct workers in 2002. Control: Firms with less than 15 workers in 2002.
Regression includes a second degree polynomial and interactions with treatment variable.

Table 13: Parametric RDD estimation results. Sample year 2003

Panel A: Log Output per worker

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(N_d \geq 15 \mid t = 2002)$	0.075	0.261**	0.356**	0.198**	0.050	0.062	0.131**	0.035	0.088	0.118**	0.174**	0.075
	[0.113]	[0.108]	[0.117]	[0.067]	[0.056]	[0.047]	[0.048]	[0.031]	[0.051]	[0.048]	[0.069]	[0.062]
Observations	2,950	1,758	1,451	1,118	2,746	1,718	1,434	1,104	2,746	1,718	1,434	1,104
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4
Baseline controls	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES
Industry controls	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
model	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols

Panel B: Log Capital per worker

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(N_d \geq 15 \mid t = 2002)$	-0.195	-0.026	0.154	0.421***	-0.174	-0.067	0.142	0.404***	-0.037	0.174	0.406**	0.726***
	[0.183]	[0.179]	[0.158]	[0.104]	[0.148]	[0.167]	[0.137]	[0.102]	[0.176]	[0.184]	[0.165]	[0.120]
Observations	2,944	1,782	1,476	1,141	2,727	1,729	1,447	1,117	2,727	1,729	1,447	1,117
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4
Baseline controls	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES
Industry controls	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
model	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols

Panel C: Log Total Factor Productivity

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(N_d \geq 15 \mid t = 2002)$	-0.034	-0.084***	-0.108***	-0.095***	-0.035**	-0.047**	-0.059***	-0.058**	-0.010	-0.026	-0.043*	-0.015
	[0.027]	[0.013]	[0.012]	[0.017]	[0.016]	[0.017]	[0.014]	[0.017]	[0.014]	[0.015]	[0.020]	[0.020]
Observations	2,757	1,663	1,382	1,061	2,601	1,643	1,379	1,059	2,601	1,643	1,379	1,059
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4
Baseline controls	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES
Industry controls	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
model	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols

Panel D: Log Investment

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(N_d \geq 15 \mid t = 2002)$	-0.402**	-0.086	0.164	0.180	-0.351**	-0.141	0.195*	0.266*	-0.277	0.108	0.457***	0.496**
	[0.192]	[0.142]	[0.133]	[0.161]	[0.166]	[0.160]	[0.099]	[0.125]	[0.188]	[0.160]	[0.133]	[0.153]
Observations	2,575	1,564	1,296	1,012	2,391	1,514	1,269	990	2,391	1,514	1,269	990
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4
Baseline controls	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES
Industry controls	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
model	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols

Panel E: Skilled/Unskilled Ratio

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(N_d \geq 15 \mid t = 2002)$	-0.004	0.031	0.006	0.065***	-0.007	0.025	-0.002	0.058***	0.032	0.053	0.020	0.121*
	[0.049]	[0.040]	[0.030]	[0.012]	[0.048]	[0.040]	[0.032]	[0.012]	[0.056]	[0.054]	[0.056]	[0.062]
Observations	2,979	1,785	1,475	1,139	2,744	1,732	1,450	1,118	2,744	1,732	1,450	1,118
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4
Baseline controls	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES

Continued on next page

– continued from previous page

Industry controls	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
model	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols
Panel F: Log Average wage bill-Direct labor												
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(N_d \geq 15 \mid t = 2002)$	-0.008	0.004	0.025	-0.038***	-0.022	-0.018	-0.007	-0.073***	0.009	0.034	0.073	0.003
	[0.017]	[0.035]	[0.035]	[0.008]	[0.025]	[0.031]	[0.033]	[0.016]	[0.034]	[0.043]	[0.059]	[0.058]
Observations	3,018	1,808	1,495	1,161	2,786	1,748	1,462	1,133	2,786	1,748	1,462	1,133
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4
Baseline controls	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES
Industry controls	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
model	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols
Panel G: Log Average Wage bill-Outsorced labor												
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(N_d \geq 15 \mid t = 2002)$	-0.129	-0.195	-0.247	0.050*	-0.223**	-0.314**	-0.216*	-0.012	-0.316**	-0.547***	-0.449	-0.113
	[0.120]	[0.118]	[0.151]	[0.022]	[0.082]	[0.118]	[0.100]	[0.074]	[0.140]	[0.175]	[0.253]	[0.236]
Observations	316	219	186	138	301	210	180	134	301	210	180	134
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4
Baseline controls	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES
Industry controls	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
model	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols
Panel H: Exporter												
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(N_d \geq 15 \mid t = 2002)$	-0.032	-0.051	-0.001	-0.006	-0.037	-0.058*	-0.020	-0.029	-0.047	-0.056	-0.025	-0.039
	[0.029]	[0.038]	[0.026]	[0.025]	[0.026]	[0.030]	[0.021]	[0.021]	[0.033]	[0.041]	[0.037]	[0.053]
Observations	2,912	1,730	1,431	1,099	2,698	1,688	1,414	1,085	2,698	1,688	1,414	1,085
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4
Baseline controls	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES
Industry controls	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
model	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols
Panel I: Share of exports												
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(N_d \geq 15 \mid t = 2002)$	1.700	3.097**	5.774***	5.907***	1.789	3.283***	5.467***	5.418***	1.011	2.420	4.567***	3.869***
	[1.491]	[1.332]	[0.553]	[0.189]	[1.434]	[0.984]	[0.477]	[0.497]	[1.592]	[1.485]	[0.881]	[0.764]
Observations	2,912	1,730	1,431	1,099	2,698	1,688	1,414	1,085	2,698	1,688	1,414	1,085
Bandwidth	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4	+/- 10	+/- 6	+/- 5	+/- 4
Baseline controls	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES
Industry controls	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
model	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols	ols

Standard errors clustered at the assignment variable level shown in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
Treatment: Firms with 15-29 direct workers in 2002. Control: Firms with less than 15 workers in 2002.
Regression includes a second degree polynomial and interactions with treatment variable.

Heterogeneous effects by industries of treatment in selected variables. In this section I show the results of estimating intent-to-treat regression discontinuity design estimations following Cattaneo et al. (2016) for all industries in the sample of main estimations.

Table 14: Log Output per worker

Bandwidth	[-4,4]		[-2,2]		
Industry	Effect	P-Value	Effect	P-Value	Observations
15	0.520	0.001	0.323	0.080	1662
17	-0.179	0.401	-1.382	0.406	379
18	-0.505	0.000	-1.083	0.224	887
19	-0.088	0.665	-0.242	0.955	400
20	0.043	0.826	-0.691	0.230	148
21	-0.455	0.004	-0.068	0.220	271
22	0.223	0.098	0.391	0.253	439
24	-0.403	0.133	0.821	0.302	630
25	0.301	0.048	0.020	0.311	583
26	-0.609	0.007	0.472	0.552	399
27	0.037	0.949	-1.591	0.965	148
28	0.084	0.660	-0.246	0.719	480
29	0.345	0.057	-1.469	0.980	409
31	-0.723	0.006	0.482	0.776	169
33	0.386	0.124	-1.012	0.046	57
34	1.511	0.000	1.132	0.088	194
36	0.146	0.276	0.090	0.071	518

15- Manufacture of food products and beverages; 17- Manufacture of textiles; 18- Manufacture of wearing apparel; dressing and dyeing of fur; 19- Tanning and dressing of leather; manufacture of luggage, handbags, saddlery, harness and footwear; 20- Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials; 21- Manufacture of paper and paper products; 22- Publishing, printing and reproduction of recorded media; 24- Manufacture of chemicals and chemical products; 25- Manufacture of rubber and plastics products; 26- Manufacture of other non-metallic mineral products; 27- Manufacture of basic metals; 28- Manufacture of fabricated metal products, except machinery and equipment; 29- Manufacture of machinery and equipment n.e.c.; 31- Manufacture of electrical machinery and apparatus n.e.c.; 33- Manufacture of medical, precision and optical instruments, watches and clocks; 34- Manufacture of motor vehicles, trailers and semi-trailers;36- Manufacture of furniture; manufacturing n.e.c.

Table 15: Log Total Factor Productivity

Bandwidth	[-4,4]		[-2,2]		
Industry	Effect	P-Value	Effect	P-Value	Observations
15	0.214	0.000	1.013	0.995	1662
17	-0.102	0.052	0.272	0.027	379
18	0.338	0.000	0.433	0.291	887
19	0.170	0.004	-0.055	0.201	400
20	-0.193	0.012	0.028	0.973	148
21	0.063	0.068	-0.130	0.245	271
22	-0.085	0.022	0.017	0.059	439
24	-0.133	0.006	-0.118	0.009	630
25	-0.182	0.000	-0.026	0.012	583
26	-0.200	0.003	-0.410	0.024	399
27	-0.153	0.094	-0.127	0.278	148
28	0.103	0.006	0.046	0.180	480
29	-0.157	0.001	0.351	0.212	409
31	0.333	0.005	0.269	0.911	169
33	0.048	0.716	1.065	0.046	57
34	-0.345	0.000	-0.160	0.419	194
36	-0.010	0.738	0.127	0.070	518

15- Manufacture of food products and beverages; 17- Manufacture of textiles; 18- Manufacture of wearing apparel; dressing and dyeing of fur; 19- Tanning and dressing of leather; manufacture of luggage, handbags, saddlery, harness and footwear; 20- Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials; 21- Manufacture of paper and paper products; 22- Publishing, printing and reproduction of recorded media; 24- Manufacture of chemicals and chemical products; 25- Manufacture of rubber and plastics products; 26- Manufacture of other non-metallic mineral products; 27- Manufacture of basic metals; 28- Manufacture of fabricated metal products, except machinery and equipment; 29- Manufacture of machinery and equipment n.e.c.; 31- Manufacture of electrical machinery and apparatus n.e.c.; 33- Manufacture of medical, precision and optical instruments, watches and clocks; 34- Manufacture of motor vehicles, trailers and semi-trailers;36- Manufacture of furniture; manufacturing n.e.c.

Table 16: Log Average wage bill-Direct labor

Bandwidth	[-4,4]		[-2,2]		
Industry	Effect	P-Value	Effect	P-Value	Observations
15	-0.087	0.090	0.070	0.771	1662
17	0.402	0.004	-0.823	0.264	379
18	-0.285	0.000	-0.486	0.136	887
19	-0.533	0.000	-0.378	0.011	400
20	-0.244	0.009	-0.895	0.085	148
21	0.131	0.241	-0.036	0.224	271
22	0.327	0.005	0.753	0.404	439
24	-0.376	0.014	-0.435	0.133	630
25	-0.592	0.000	-0.684	0.002	583
26	0.033	0.802	0.476	0.139	399
27	0.836	0.000	1.545	0.664	148
28	0.239	0.004	-0.042	0.983	480
29	-0.252	0.009	-0.522	0.847	409
31	-0.279	0.009	-0.256	0.346	169
33	-0.236	0.135	-0.303	0.229	57
34	0.573	0.000	-0.205	0.095	194
36	-0.472	0.000	0.085	0.001	518

15- Manufacture of food products and beverages; 17- Manufacture of textiles; 18- Manufacture of wearing apparel; dressing and dyeing of fur; 19- Tanning and dressing of leather; manufacture of luggage, handbags, saddlery, harness and footwear; 20- Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials; 21- Manufacture of paper and paper products; 22- Publishing, printing and reproduction of recorded media; 24- Manufacture of chemicals and chemical products; 25- Manufacture of rubber and plastics products; 26- Manufacture of other non-metallic mineral products; 27- Manufacture of basic metals; 28- Manufacture of fabricated metal products, except machinery and equipment; 29- Manufacture of machinery and equipment n.e.c.; 31- Manufacture of electrical machinery and apparatus n.e.c.; 33- Manufacture of medical, precision and optical instruments, watches and clocks; 34- Manufacture of motor vehicles, trailers and semi-trailers;36- Manufacture of furniture; manufacturing n.e.c.

Table 17: Share of exports

Bandwidth	[-4,4]		[-2,2]		
Industry	Effect	P-Value	Effect	P-Value	Observations
15	18.562	0.000	8.700	0.000	1662
17	29.541	0.000	1.226	0.991	379
18	3.974	0.217	-6.990	0.726	887
19	-2.308	0.660	-3.937	0.860	400
20	0.000	1.000	0.000	1.000	148
21	-0.031	0.991	-0.402	1.000	271
22	1.101	0.530	-3.089	0.889	439
24	-10.436	0.000	-5.966	0.000	630
25	-8.066	0.000	-7.297	0.000	583
26	1.847	0.009	0.541	0.511	399
27	-4.515	0.423	-4.927	0.000	148
28	6.032	0.000	-2.537	0.028	480
29	8.013	0.001	7.756	0.277	409
31	7.635	0.001	1.300	1.000	169
33	0.258	0.787	2.061	0.046	57
34	0.000	1.000	0.000	1.000	194
36	-0.061	0.995	0.852	1.000	518

15- Manufacture of food products and beverages; 17- Manufacture of textiles; 18- Manufacture of wearing apparel; dressing and dyeing of fur; 19- Tanning and dressing of leather; manufacture of luggage, handbags, saddlery, harness and footwear; 20- Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials; 21- Manufacture of paper and paper products; 22- Publishing, printing and reproduction of recorded media; 24- Manufacture of chemicals and chemical products; 25- Manufacture of rubber and plastics products; 26- Manufacture of other non-metallic mineral products; 27- Manufacture of basic metals; 28- Manufacture of fabricated metal products, except machinery and equipment; 29- Manufacture of machinery and equipment n.e.c.; 31- Manufacture of electrical machinery and apparatus n.e.c.; 33- Manufacture of medical, precision and optical instruments, watches and clocks; 34- Manufacture of motor vehicles, trailers and semi-trailers;36- Manufacture of furniture; manufacturing n.e.c.