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

Carlos Ospino*‡

*Universidad de los Andes*

July 12, 2016

**Abstract**

In this paper I evaluate 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 reform it went through in 2002 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, this regulation has been in place for more than a decade but has not been evaluated. Results using a regression discontinuity design (RDD) which compares small firms subject to the regulation and those that are not, shows positive effects on output per worker (10 log points) and total factor productivity (20 log 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 their workers through higher wages.

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

*This paper is 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 the CEDLAS' Grant for Graduate Thesis competition 2015. I thank an anonymous referee from CEDLAS for her helpful comments and suggestions to an earlier draft.

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

# 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 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

[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

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

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

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 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

[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.

[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).

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 are in no obligation to do so. 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 three quarters of the apprenticeship contract'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 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.

[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.

## 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 $\delta$ 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 constructed using the production function estimates obtained by Eslava et al. (2004) for the Colombian manufacturing sector. These authors estimate the Cobb-Douglas production function, $f(k, l, e, m) = Ak^{\alpha}l^{\beta}e^{\gamma}m^{\sigma}$. Where $k$ is capital, $l$ is labor, $e$ is electrical energy consumption measured in Kilowatts and $m$ is intermediate materials. Thus, $\log(TFP)[\log(A)] = \log(Output) - \log(f(k, l, e, m))$[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 was used. Each parameter is estimated as the average factor share, across firms and industries during the period 2000-2002. In this case, the parametric Cobb-Douglas production function only uses capital, labor and intermediate materials. The two estimates are highly correlated. The results using this measure are available upon request

*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.

### 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. 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. The main difference is its compulsory quota but it is a useful regulation for the analysis of how apprenticeship contracts affect firm performance. Apprenticeship contract regulations are a topic of regional interest[11].

To evaluate the impact of being subject to the apprenticeship contract, a suited 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. Unfortunately, the empirical analysis in section 5.2 of the Appendix shows that the assumptions necessary for the DD estimation do not hold. In particular I found evidence of anticipation effects and trends did not follow a common growth pattern. Given these findings and to take advantage of the regulation threshold which separates similar firms from

[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 this dissertation which restricts the analysis to the first threshold, the committee recommended that the other thresholds be studied in another paper.

[11] See http://blogs.iadb.org/trabajo/category/aprendices/ for a series of blogs on the subject by The Inter-American Development Bank

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. 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 likely that firms had no incentives to change their directly hired labor demand at the threshold of compliance with the regulation before the reform was in effect. If firms cannot *perfectly* modify the number of directly hired workers to avoid compliance, then having to comply with 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 imperfect capacity of firms to 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 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.

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 $\beta_0$ 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]. $D_{ij} \times m(N_{dij}, p)$ allows polynomial slopes to differ for firms below and above the threshold of compliance.

$\beta_1$ 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 the subject of 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.

# 4 Results

## 4.1 Assumptions and validity tests

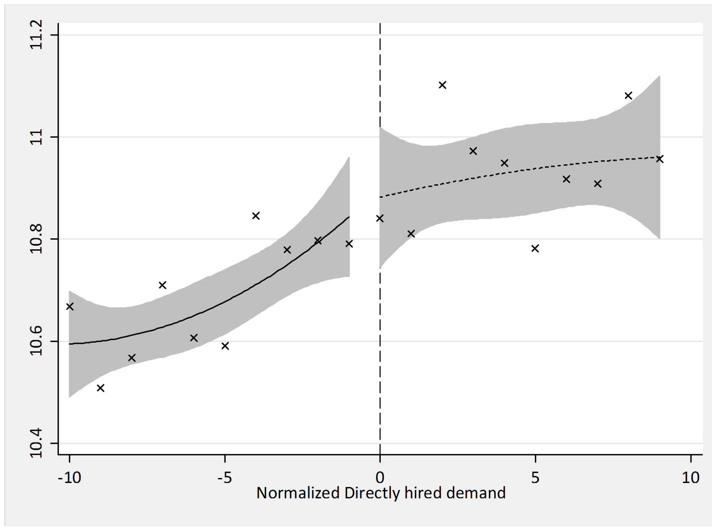
In this section I carry out standard assumptions and validity test for RDDs to make sure the approximation is appropriate for the current evaluation.

**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 1-7 show quadratic non parametric fit graphics 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 quadratic functions 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).

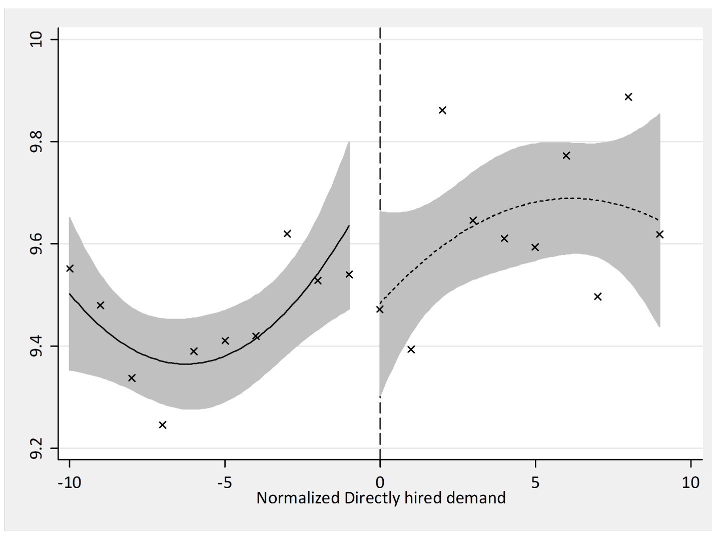
[12] A second degree polynomial was used since it provides sufficient functional flexibility and Gelman and Imbens (2014) warn against the use of higher order polynomials in RDD estimations.

Figure 1: Average Log output per total number of workers



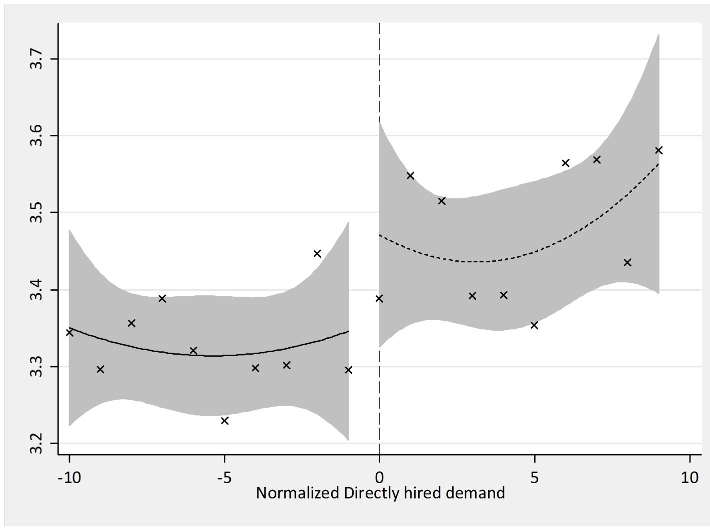

Source: EAM 2004. Figure plots quadratic fits of the outcome variable for treatment and control firms. x's represent the average value of the outcome for each level of the assignment variable. Shaded areas contain a 95% confidence interval for the estimation.

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



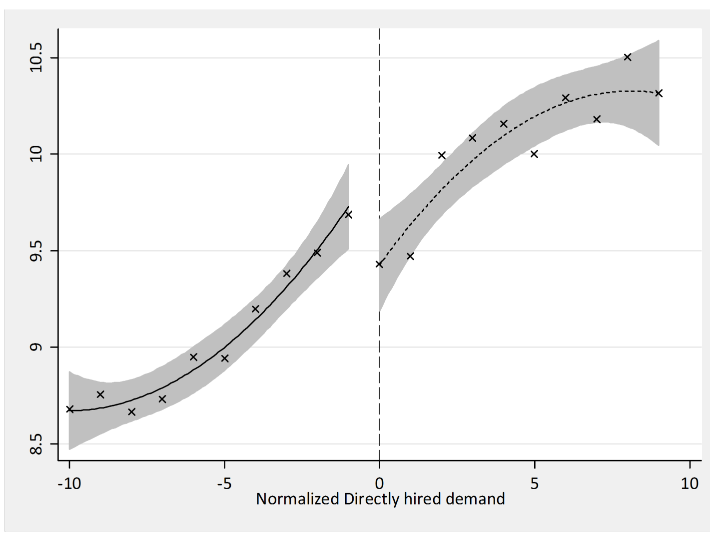

Source: EAM 2004. Figure plots quadratic fits of the outcome variable for treatment and control firms. x's represent the average value of the outcome for each level of the assignment variable. Shaded areas contain a 95% confidence interval for the estimation.

Figure 3: Average Log Total Factor Productivity



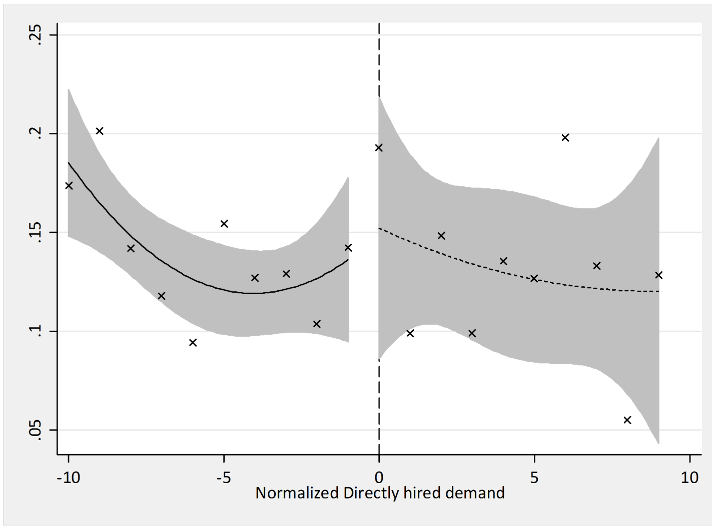

Source: EAM 2004. Figure plots quadratic fits of the outcome variable for treatment and control firms. x's represent the average value of the outcome for each level of the assignment variable. Shaded areas contain a 95% confidence interval for the estimation.

Figure 4: Average Log Investment



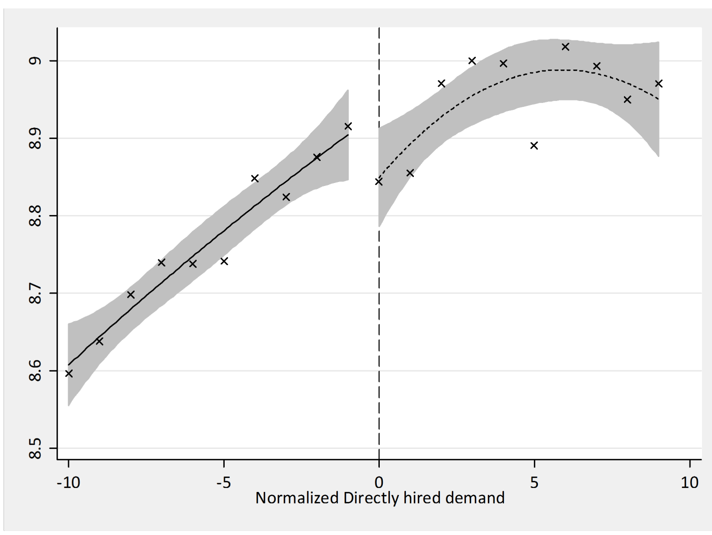

Source: EAM 2004. Figure plots quadratic fits of the outcome variable for treatment and control firms. x's represent the average value of the outcome for each level of the assignment variable. Shaded areas contain a 95% confidence interval for the estimation.

Figure 5: Average Skilled/Unskilled ratio



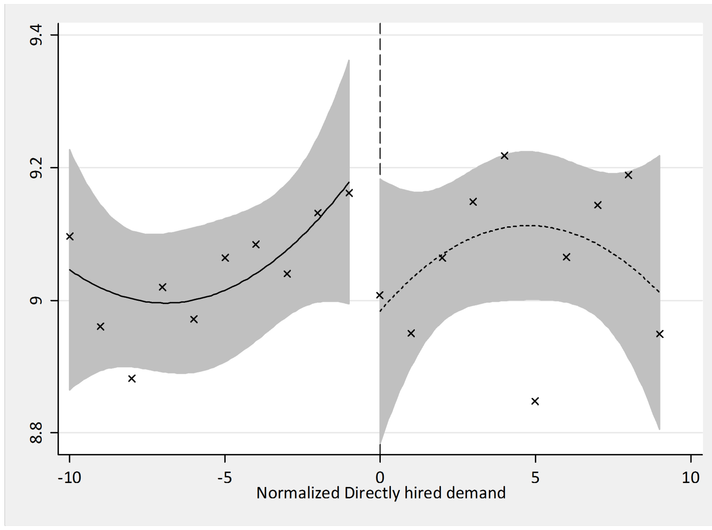

Source: EAM 2004. Figure plots quadratic fits of the outcome variable for treatment and control firms. x's represent the average value of the outcome for each level of the assignment variable. Shaded areas contain a 95% confidence interval for the estimation.

Figure 6: Average Log Directly hired wage bill




Source: EAM 2004. Figure plots quadratic fits of the outcome variable for treatment and control firms. x's represent the average value of the outcome for each level of the assignment variable. Shaded areas contain a 95% confidence interval for the estimation.

Figure 7: Average Log Outsourced hired wage bill




Source: EAM 2004. Figure plots quadratic fits of the outcome variable for treatment and control firms. x's represent the average value of the outcome for each level of the assignment variable. 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 are presented in the Appendix[13].

[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 founded a year before firms that would not be subject to it.

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

| VARIABLES | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| $D_{15} = I(N_d \geq 0)$ | -0.148*** | -0.125 | -0.252** | -0.011 |
| | [0.042] | [0.088] | [0.108] | [0.046] |
| $N_d$ =Direct labor 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.

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 can not be rejected and provides evidence of no manipulation of the running variable. In section 5.1 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.

## 4.2 Estimation results

Table 2 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 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 20 log points as a result of the apprenticeship contract for firms subject to the regulation. 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.

Table 2: Parametric RDD estimation results

Panel A: Log Output per worker

| VARIABLES | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $I(N_d \geq 15 \mid t = 2002)$ | -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) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $I(N_d \geq 15 \mid t = 2002)$ | -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) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $I(N_d \geq 15 \mid t = 2002)$ | 0.109 | 0.048 | 0.157 | 0.205 | 0.111 | 0.037 | 0.093 | 0.102 | 0.177** | 0.141* | 0.128 | 0.203*** |
| | [0.107] | [0.124] | [0.089] | [0.119] | [0.078] | [0.090] | [0.073] | [0.091] | [0.084] | [0.073] | [0.077] | [0.055] |
| Observations | 2,846 | 1,711 | 1,421 | 1,102 | 2,638 | 1,660 | 1,391 | 1,076 | 2,638 | 1,660 | 1,391 | 1,076 |
| 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) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $I(N_d \geq 15 \mid t = 2002)$ | -0.549*** | -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 |

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 2: Parametric RDD estimation results (Continued)

Panel E: Skilled/Unskilled Ratio

| VARIABLES | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $I(N_d \geq 15 \vert t = 2002)$ | 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 |
| 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) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $I(N_d \geq 15 \vert t = 2002)$ | -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-Outsorced labor)

| VARIABLES | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $I(N_d \geq 15 \vert t = 2002)$ | -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 |

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 section 5.1 of 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. 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 5.1 in Appendix shows statistically significant effects for capital per worker (72.6 log points), investment (50 log points), and the skilled/unskilled ratio (12 log points). The fact that the effects of capital and investment are larger than the anticipation effects, suggest the policy might have affected these variables in 2003.

## 5 Discussion

In this paper I have used regression discontinuity design (RDD) methods to evaluate the impact of the apprenticeship contract on Colombian firm dynamics. Firms showed statistical significant differences at the moment when the regulation changed and most outcome variables do not followed 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 RDD 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, capital per worker, total factor productivity and negative effects on the direct labor average wage bill. While testing for anticipation effects, capital per worker showed a statistically significant increase in 2002. Therefore, the estimated effect for this variable should not fully be attributed to the apprenticeship contract. The increase in productivity measured by output per worker and total factor productivity suggests that the policy might have forced firms around the threshold of compliance with the regulation to adjust their labor force size more efficiently. 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 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 a positive effect on productivity for this group of firms. 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 that determine whether firms are subject to the apprenticeship contract regulation 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 Pischke, J.-S. (1999). Beyond becker: Training in imperfect labour markets. *The Economic Journal*, 109(453):112–142.

Angrist, J. D. and Lavy, V. (1999). Using maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics*, 114(2):533–575.

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

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

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

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

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

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

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

Imbens, G. W. and Lemieux, T. (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 Vanormelingen, S. (2010). The impact of training on productivity and wages: Firm level evidence.

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

Maloney, W. and Mendez, J. (2004). Measuring the impact of minimum wages. evidence from latin america. In *Law and Employment: Lessons from Latin America and the Caribbean*, pages 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 Zwick, T. (2009). Why do firms train apprentices? the net cost puzzle reconsidered. *Labour Economics*, 16(6):631–637.

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

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

## 5.1 Appendix

### Parametric McCrary (2008) tests

Table 3: 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 4: Parametric McCrary (2008) test of no manipulation. Year 2004

| VARIABLES | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| $D_{15} = I(N_d \geq 0)$ | -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 5: 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.

## Local continuity of baseline covariates

Table 6: 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.

Table 7: 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.

Table 8: 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.

## Robustness Checks: Anticipation Effects

Table 9: Anticipation Effect: Log Output per worker

| VARIABLES | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $I(N_d \geq 15 \| t = 2002)$ | 0.022 | 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 |

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 10: Anticipation Effect: 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 | 3,152 | 1,864 | 1,541 | 1,191 | 2,852 | 1,802 | 1,507 | 1,163 | 2,852 | 1,802 | 1,507 | 1,163 |
| 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 |

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: Anticipation Effect: Log Total Factor Productivity

| VARIABLES | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $I(N_d \geq 15 \mid t = 2002)$ | 0.077 | 0.134** | 0.171*** | 0.251*** | 0.068 | 0.041* | 0.021 | -0.021 | 0.110 | 0.110 | 0.057 | -0.001 |
| | [0.063] | [0.059] | [0.052] | [0.047] | [0.054] | [0.022] | [0.033] | [0.037] | [0.070] | [0.077] | [0.085] | [0.104] |
| Observations | 3,137 | 1,823 | 1,499 | 1,156 | 2,863 | 1,781 | 1,482 | 1,142 | 2,863 | 1,781 | 1,482 | 1,142 |
| 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 |

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 12: Anticipation Effect: 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 |

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 13: Anticipation Effect: 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 |
| 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 |

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 14: Anticipation Effect: Log Wage bill per worker–Direct labor

| VARIABLES | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $I(N_d \geq 15 \mid t = 2002)$ | -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 |

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 15: Anticipation Effect: Log Wage bill per worker–Outsourced labor

| VARIABLES | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $I(N_d \geq 15 \mid t = 2002)$ | -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 |

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.

## RDD for outcome variables in the year 2003

Table 16: 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) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $I(N_d \geq 15 \mid t = 2002)$ | 0.092 | 0.123 | 0.230*** | 0.301*** | 0.076 | 0.059 | 0.090* | -0.005 | 0.149 | 0.166** | 0.167 | 0.078 |
| | [0.089] | [0.073] | [0.042] | [0.029] | [0.092] | [0.037] | [0.046] | [0.037] | [0.099] | [0.073] | [0.092] | [0.133] |
| Observations | 2,930 | 1,745 | 1,441 | 1,109 | 2,728 | 1,707 | 1,424 | 1,095 | 2,728 | 1,707 | 1,424 | 1,095 |
| 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 |

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 17: Parametric RDD estimation results. Sample year 2003 (Continued)

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 |
| 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-Outsourced 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 |

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.

## 5.2 Difference-in-Difference validity tests

**Balance of covariates**

It's not unusual for natural experiments to generate treatment and control groups which are not balanced in covariates. I tested for mean differences at baseline in terms of firm age, the share of management workers, electricity consumption, and the number of skilled workers. Table 18 shows that firms do not differ in terms of their average age, but treatments have a lower share of managers, consume more electricity, and have a higher number of skilled workers.

Table 18: Sample mean difference test of some baseline covariates

| Variable | Treatment | Control | Difference | P-value |
|---|---|---|---|---|
| Age | 18,98 | 18,84 | 0,14 | 0,818 |
| Share of admin workers | 0,29 | 0,34 | -0,04 | 0,000 |
| Log. Energy consumption | 10,72 | 9,83 | 0,89 | 0,000 |
| No. Of Skilled Workers | 6,60 | 2,75 | 3,85 | 0,000 |

Treatment: Firms with 15-29 direct workers in 2002. Control: Firms with less than 15 workers in 2002. The number of observations varied between tests.

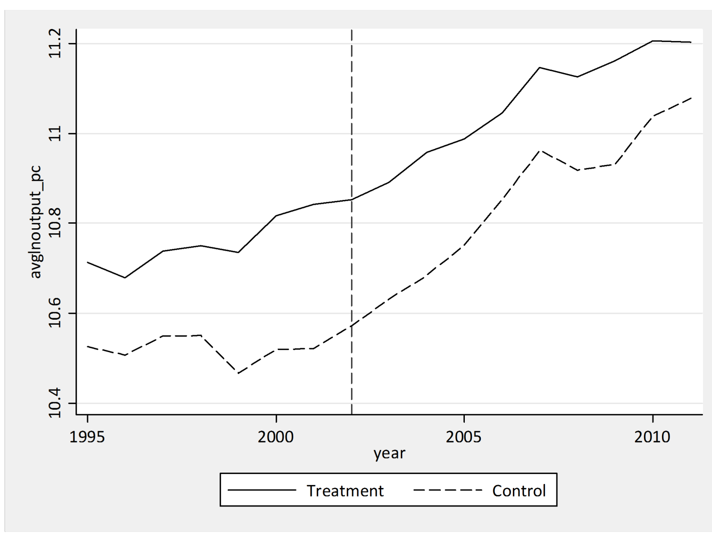
**Common trends assumption**
As a test of the validity of the DD estimator I fitted a model of leads and lags as specified in equation (2).

$$y_{it} = \alpha_i + \lambda_t + \sum_{\tau=1995}^{2011} \beta_\tau D_i(I[N_d \geq 15|t = 2002]) \times \ \lambda_\tau + \mu_{it} \tag{2}$$

The variable $D_i$ is defined as the treatment status firm $i$ had in the year 2002 and is therefore the time invariant intent to treat. The model thus formally tests the DD assumption of no difference in trends previous to treatment by testing $\beta_\tau = 0 \quad \forall \quad \tau < 2003$. The test also serves as a way to provide support for the interpretation of the DD estimators as identifying causal effects.

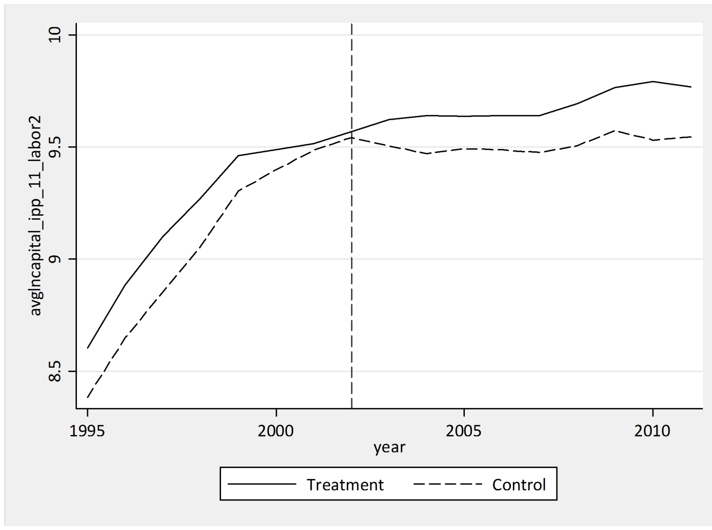
The key identifying assumption of the DD estimator is that in absence of treatment both groups would follow the same trend in outcome variables. In line with figures 8 to 14, table 19 showed significant effects for all outcomes except the skilled/unskilled ratio, suggesting anticipation effects of the change in regulation. The result implies that the common trend assumption does not hold for the evaluation of the apprenticeship contract. The failure to meet the DD assumptions invalidates this approach and was therefore not pursued.

Figure 8: Log Output per worker



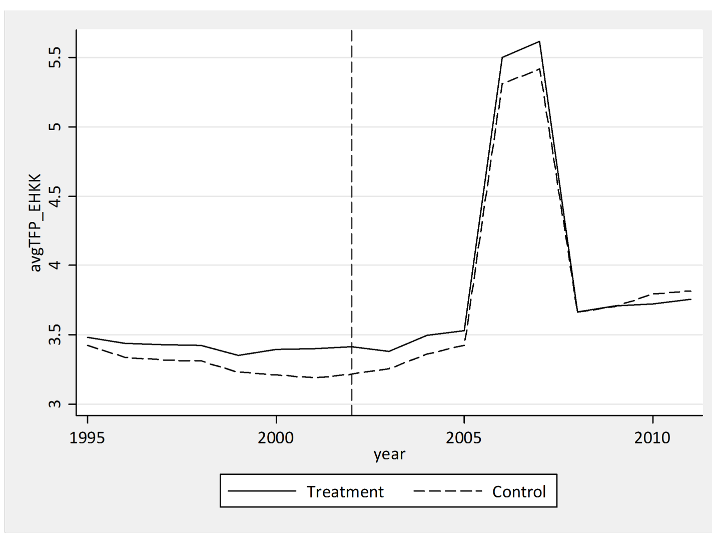

Source: EAM 1995-2011. Treatment is defined in 2002.

Figure 9: Log Capital per worker



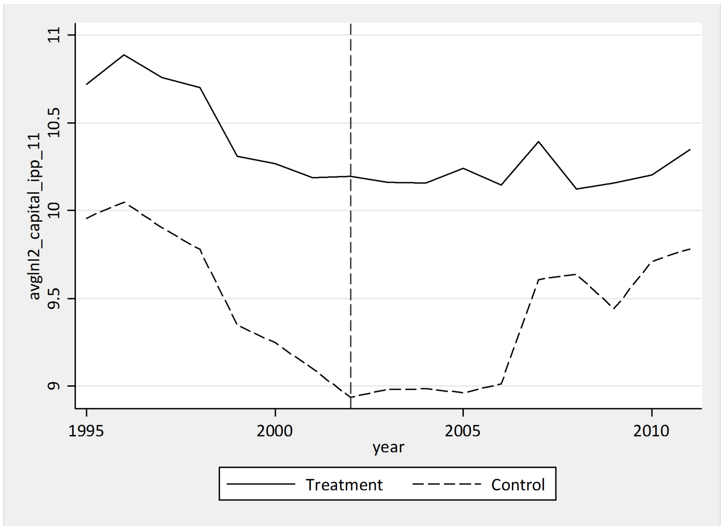

Source: EAM 1995-2011. Treatment is defined in 2002.

Figure 10: Log Total Factor Productivity



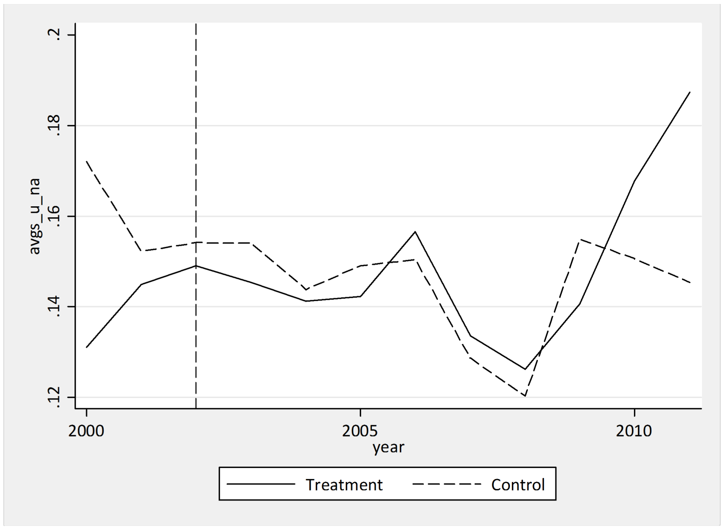

Source: EAM 1995-2011. Treatment is defined in 2002.

Figure 11: Log Investment



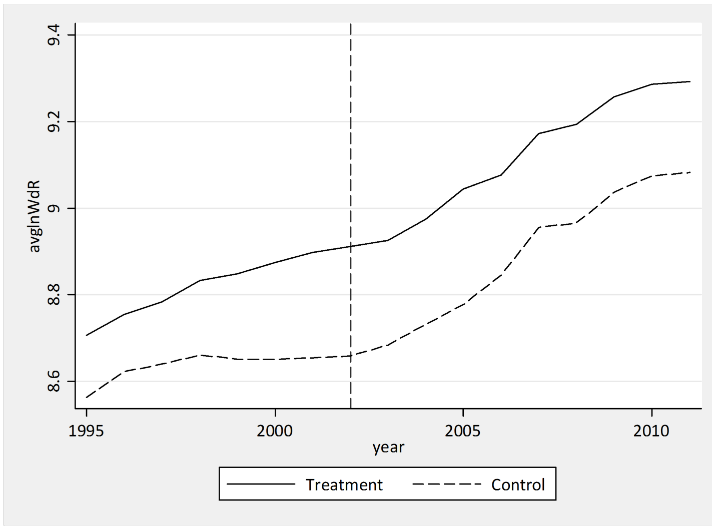

Source: EAM 1995-2011. Treatment is defined in 2002.

Figure 12: Skilled/Unskilled Ratio



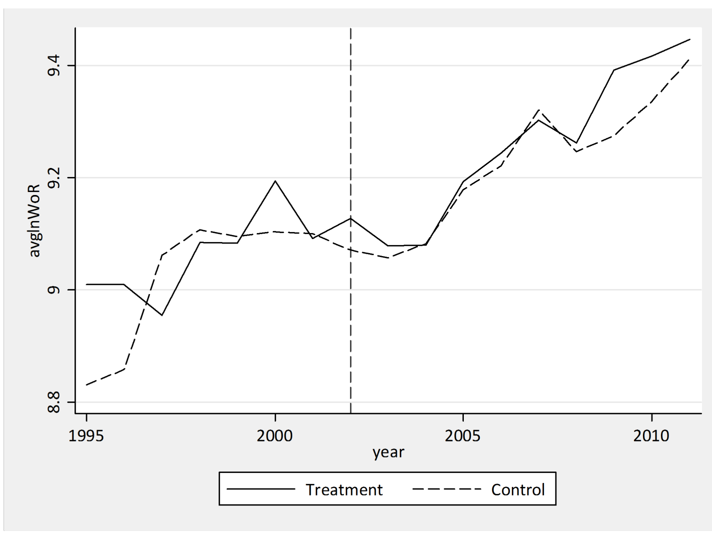

Source: EAM 1995-2011. Treatment is defined in 2002.

Figure 13: Log Wage bill per worker–Direct labor




Source: EAM 1995-2011. Treatment is defined in 2002.

Figure 14: Log Wage bill per worker–Outsourced labor




Source: EAM 1995-2011. Treatment is defined in 2002.

Table 19: Leads and Lags estimation

| VARIABLES | (1) Log(Output/ worker) | (2) Log(Capital/ worker) | (3) Log(TFP) | (4) Log(Investm ent) | (5) Skilled/ Unskilled ratio | (6) Log(Avg. Payroll) | (7) Log(Avg. Direct Payroll) | (8) Log(Avg. Outsourced Payroll) |
|---|---|---|---|---|---|---|---|---|
| $I(N_d \geq 15 \vert t = 2002) \times \lambda_{1996}$ | 0.006 | -0.033 | 0.069** | 0.030 | -0.000 | -0.066* | -0.004 | -0.093 |
| | [0.023] | [0.029] | [0.034] | [0.055] | [0.004] | [0.035] | [0.013] | [0.136] |
| $I(N_d \geq 15 \vert t = 2002) \times \lambda_{1997}$ | 0.018 | 0.009 | 0.082** | 0.064 | 0.001 | -0.073* | -0.017 | -0.306** |
| | [0.028] | [0.035] | [0.037] | [0.059] | [0.006] | [0.039] | [0.014] | [0.133] |
| $I(N_d \geq 15 \vert t = 2002) \times \lambda_{1998}$ | 0.036 | -0.015 | 0.120*** | 0.213*** | -0.003 | -0.075* | 0.018 | -0.208 |
| | [0.031] | [0.039] | [0.039] | [0.060] | [0.008] | [0.042] | [0.016] | [0.154] |
| $I(N_d \geq 15 \vert t = 2002) \times \lambda_{1999}$ | 0.106*** | -0.019 | 0.108*** | 0.248*** | -0.001 | -0.023 | 0.035** | -0.211 |
| | [0.033] | [0.041] | [0.040] | [0.064] | [0.010] | [0.405] | [0.016] | [0.142] |
| $I(N_d \geq 15 \vert t = 2002) \times \lambda_{2000}$ | 0.128*** | -0.082** | 0.186*** | 0.318*** | 0.023 | 0.033 | 0.046*** | -0.156 |
| | [0.034] | [0.042] | [0.040] | [0.065] | [0.025] | [0.036] | [0.017] | [0.135] |
| $I(N_d \geq 15 \vert t = 2002) \times \lambda_{2001}$ | 0.139*** | -0.092** | 0.209*** | 0.335*** | 0.038 | 0.019 | 0.047*** | -0.270* |
| | [0.035] | [0.042] | [0.041] | [0.068] | [0.026] | [0.036] | [0.018] | [0.144] |
| $I(N_d \geq 15 \vert t = 2002) \times \lambda_{2002}$ | 0.047 | -0.147*** | 0.210*** | 0.425*** | 0.025 | 0.005 | 0.028 | -0.310** |
| | [0.036] | [0.044] | [0.042] | [0.072] | [0.029] | [0.037] | [0.018] | [0.137] |
| $I(N_d \geq 15 \vert t = 2002) \times \lambda_{2003}$ | 0.100*** | -0.004 | 0.170*** | 0.373*** | 0.036 | 0.035 | 0.060*** | -0.114 |
| | [0.038] | [0.046] | [0.043] | [0.074] | [0.030] | [0.037] | [0.019] | [0.140] |
| $I(N_d \geq 15 \vert t = 2002) \times \lambda_{2004}$ | 0.094** | 0.069 | 0.132*** | 0.335*** | 0.019 | 0.038 | 0.063*** | -0.164 |
| | [0.038] | [0.047] | [0.045] | [0.076] | [0.030] | [0.038] | [0.020] | [0.142] |
| $I(N_d \geq 15 \vert t = 2002) \times \lambda_{2005}$ | 0.099*** | 0.086* | 0.134*** | 0.395*** | 0.033 | 0.030 | 0.056*** | -0.187 |
| | [0.038] | [0.049] | [0.046] | [0.078] | [0.036] | [0.038] | [0.020] | [0.146] |
| $I(N_d \geq 15 \vert t = 2002) \times \lambda_{2006}$ | 0.086** | 0.092* | 0.144*** | 0.334*** | 0.014 | 0.029 | 0.057*** | -0.195 |
| | [0.038] | [0.050] | [0.051] | [0.086] | [0.026] | [0.039] | [0.021] | [0.143] |
| $I(N_d \geq 15 \vert t = 2002) \times \lambda_{2007}$ | 0.105*** | 0.085 | 0.163*** | 0.147 | 0.041 | 0.021 | 0.050** | -0.138 |
| | [0.040] | [0.053] | [0.052] | [0.134] | [0.025] | [0.039] | [0.022] | [0.151] |
| $I(N_d \geq 15 \vert t = 2002) \times \lambda_{2008}$ | 0.085** | 0.097* | -0.006 | -0.228 | 0.029 | 0.016 | 0.060*** | -0.163 |
| | [0.041] | [0.055] | [0.053] | [0.142] | [0.023] | [0.039] | [0.023] | [0.149] |
| $I(N_d \geq 15 \vert t = 2002) \times \lambda_{2009}$ | 0.101** | 0.089 | 0.036 | 0.049 | 0.021 | 0.018 | 0.048** | -0.103 |
| | [0.043] | [0.056] | [0.056] | [0.170] | [0.027] | [0.040] | [0.024] | [0.152] |
| $I(N_d \geq 15 \vert t = 2002) \times \lambda_{2010}$ | 0.051 | 0.142** | -0.050 | -0.146 | 0.042 | 0.022 | 0.059** | -0.121 |
| | [0.044] | [0.058] | [0.062] | [0.164] | [0.034] | [0.040] | [0.025] | [0.157] |
| $I(N_d \geq 15 \vert t = 2002) \times \lambda_{2011}$ | 0.035 | 0.121** | -0.072 | -0.066 | 0.028 | 0.009 | 0.055** | -0.295* |
| | [0.047] | [0.061] | [0.064] | [0.159] | [0.033] | [0.041] | [0.025] | [0.164] |
| Constant | 10.676*** | 8.498*** | 3.561*** | 10.321*** | -0.007 | 6.476*** | 8.647*** | 8.832*** |
| | [0.014] | [0.018] | [0.017] | [0.026] | [0.009] | [0.016] | [0.007] | [0.062] |
| Observations | 30,038 | 29,910 | 29,931 | 24,953 | 30,301 | 28,019 | 30,643 | 2,890 |
| Number of firms | 3,874 | 3,826 | 3,871 | 3,810 | 3,937 | 3,859 | 3,918 | 778 |
| model | fe | fe | fe | fe | fe | fe | fe | fe |

Source: EAM 1995-2011. Robust standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Treatment is defined in 2002. Table shows estimates of parameters $\beta_\tau$ in model (2).