Who benefits from reducing the cost of formality? Quantile regression discontinuity analysis*

Tommaso Gabrieli
University of Reading

Antonio F. Galvao, Jr.
University of Iowa

Gabriel V. Montes-Rojas†
City University London

November 23, 2011

Abstract

This paper studies the effect of increasing formality via tax reduction and simplification schemes on micro-firm performance. We develop a simple theoretical model that yields two intuitive results. First, low- and high- ability entrepreneurs are unlikely to be affected by a tax reduction and therefore, the reduction has an impact only on a segment of the micro-firm population. Second, the benefits to such reduction, as measured by profits and revenues, are increasing in the entrepreneur's ability. Then, we estimate the effect of formality on the entire conditional distribution (quantiles) of revenues using the 1996 Brazilian SIMPLES program and a rich survey of formal and informal micro-firms. The econometric approach compares eligible and non-eligible firms, born before and after SIMPLES in a local interval about the introduction of SIMPLES. We develop an estimator that combines both quantile regression and the regression discontinuity design. The econometric results corroborate the positive effect of formality on micro-firms' performance and produce a clear characterization of who benefits from these programs.

Key Words: Formality; Micro-firms; Quantile regression; Regression discontinuity

JEL Classification: J23; L25

*The authors would like to express their appreciation to two anonymous referees, Tiziano Razzolini, Kostantinos Tatsiramos, Hartmut Lehmann, Scott Adams, John Heywood, William Maloney, Blaise Melly, and participants at the IZA/World Bank Workshop on Institutions and Informal Employment in Emerging and Transition Economies, Bonn, the 2011 Noth American Summer Meeting of the Econometric Society, St Louis, the 2011 Conference of the Royal Economic Society, Royal Holloway, and seminars at City University London, Queen Mary University London and Universidad Autónoma de Barcelona for helpful comments and discussions. All the remaining errors are ours.

†Corresponding author: Department of Economics, City University London, 10 Northampton Square, London EC1V 0HB, UK. E-mail: Gabriel.Montes-Rojas.1@city.ac.uk

"*Lo que pasa es que acá si vos queres abrir un negocio te matan a papeles, y después te controlan, y los impuestos te revientan.*" [What happens here is that when you try to open a business they kill you on paperwork (red tape), then they control you, and taxes are unbearable.] Martín Caparrós, *El Interior*, a book on interviews and anecdotes from the poor countryside in Argentina.

1 Introduction

Formality is broadly defined as participation in societal and governmental institutions, such as paying taxes, being registered with the authorities, etc. (see Gerxhani, 2004; Maloney, 2004, for a survey). Firms' inability to become formal is thought to have deleterious effects on performance. As examples, formality offers the firm access to risk pooling mechanisms that may attract more educated paid workers and engage them in a longer relationship with the firm, which in turn makes training and capital goods acquisition more profitable; formality may be a requirement for access to formal credit markets or Government provided business development services or as Paula and Scheinkman (2007, 2010) have argued, for subcontracting relations with formal firms. Moreover, to the extent that formality increases the ability of micro-entrepreneurs to establish property rights over their investments and reduces the risk of being fined by Government inspectors, it creates incentives for operating out of fixed locations rather than in an ambulatory fashion (see de Soto, 1989).

The high costs of complying with government regulations and institutions have often been seen as largely responsible for the presence of large informal sectors in developing countries. The perceived onerous cost of formality was tackled by several Latin American governments by introducing tax reductions and simplifications. Examples of such programs are the Monotributo[1] in Argentina, SARE[2] in Mexico, and the SIMPLES[3] in Brazil. Avail-

[1] Régimen Simplificado para Pequeños Contribuyentes, see González (2006).

[2] SARE stands for "Sistema de Apertura Rápida de Empresas." It was implemented in selected municipalities and consolidated in single local offices all the federal, state and municipal procedures needed to register a firm, reducing the total duration of the process to at most 48 hours.

[3] SIMPLES stands for "Sistema Integrado de Pagamento de Impostos e Contribuições as Microempresas e Empresas de Pequeno Porte". See Section 4 for a detailed description of the program.

able evidence shows that these programs had a positive effect on formality. See Kaplan, Piedra, and Seira (2006) for SARE; and Monteiro and Assunção (2006) and Fajnzylber, Maloney, and Montes-Rojas (2011) for SIMPLES.

We contribute to this literature by answering three questions: First, what is the effect of formality on firm performance? Second, which firms benefit from tax reduction and simplification schemes? Third, is there heterogeneity on the effect of formality on firm performance? These questions have very important policy implications. In a Ricardian setting, tax reductions imply a redistribution of wealth, and therefore, it is important to quantify which firms are really benefiting from these programs. In particular, if tax reductions only benefit already well-off formal firms, then the program did not accomplish the task of broadening the scope of formality. We focus on the micro-firm sector, defined as own-account workers and firms with a maximum of 5 paid employees, that constitutes the majority of firms in developing countries.[4] Within this sector three groups can be distinguished. First, high-ability entrepreneurs with substantial growth prospects may have self-selected into formality with the old (high) tax system, as the perceived benefits of being formal offset the cost of formality. Then, this segment benefits only from the tax reduction. Second, some micro-entrepreneurs are in the informal sector as a subsistence strategy as predicted in the Harris and Todaro (1970) dual labor market hypothesis (see Maloney, 1999, 2004; Mandelman and Montes-Rojas, 2009, for a discussion). These are low-ability entrepreneurs and they will not value future gains from becoming formal and, therefore, tax reductions will not affect them. Third, in between those segments there are micro-firms that may become formal only when the cost of formality is low enough. These micro-firms receive the gains from being formal but have to pay taxes as a result. We call this segment the *target* group and it corresponds to medium-ability entrepreneurs. These are the firms that should benefit from

[4] This is the definition adopted in Fajnzylber, Maloney, and Montes-Rojas (2009, 2011).

the tax reduction programs and change their formality status.

We begin our analysis by developing a theoretical model motivated by the work of Rauch (1991) and Paula and Scheinkman (2007, 2010), with emphasis on the effect of a reduction in taxes. This model yields two intuitive results. First, low- and high-ability entrepreneurs are unlikely to be affected by a tax reduction policy reform and therefore, the reform has an impact only on a segment of the micro-firm population, defined by default as medium-ability entrepreneurs. Second, the benefits of such reform, as measured by profits and revenues, are increasing in the entrepreneur's ability.

Empirically, our goal is to quantify the impact of formality on the conditional distribution (quantiles) of micro-firm's revenues, and the size of the *target* group (i.e. which firms benefit from the tax reduction). Two problems arise in our empirical set-up. First, formality is endogenous, and in particular, correlated with the unobserved entrepreneurial ability. Second, we might not be able to identify the effect of formality for all firms.

To solve the first problem, the identification strategy makes use of the SIMPLES program in Brazil, that offers an exogenous change in legislation that can be used to control for self- selection and endogeneity. Thus, our paper builds on the work of Monteiro and Assunção (2006) and Fajnzylber, Maloney, and Montes-Rojas (2011) by analyzing the SIMPLES program in Brazil that offers an exogenous change in legislation that can be used to control for self-selection and endogeneity. We use the same unique dataset for micro-firms, the ECINF 1997. Moreover, following those authors we use a difference-in-differences approach with the age of the firm and with ineligible firms as a control group to identify the effect of formality on firm performance. Monteiro and Assunção (2006) study the effect of SIMPLES on having a government issued license, which constitutes a necessary requirement for further formalization (such as paying taxes of social security), and they find an increase in formal licensing among retail firms of 13 percentage points, but no effect on eligible firms from other sectors (construction, manufacturing, transportation and other services). In addition, using SIMPLES as an instrumental variable (IV) for formality, they show that the latter significantly increases access to credit, and alters the amount and composition of investment toward larger and longer-term projects. Fajnzylber, Maloney, and Montes-Rojas (2011) show that SIMPLES has only a *local* effect on licensing rates for firms born just after the introduction of the program. Using a regression discontinuity design (see Hahn, Todd, and van der Klaauw, 2001; van der Klaauw, 2002, for a discussion about regression discontinuity estimators), with weights given by time-in-business and its distance to the introduction of SIMPLES, they find a significant effect on licensing, tax registration, tax payments and social security contributions. When more firms were taken into consideration, the statistical significance of these effects decreases monotonically with the sample average time-distance to the introduction of SIMPLES. We build on their analysis and extend it to a quantile regression (QR) discontinuity analysis.

In order to address estimation of the distributional effects of formality, we make use of the heterogeneity in the conditional distribution of revenue applying QR techniques, which will prove an indispensable tool for the problem in question. QR methods offer the advantage of describing not only averages of possible outcomes but also their entire distribution. Thus, QR techniques provide a systematic method to analyze differences in covariates effects (see Koenker and Hallock, 2001; Koenker, 2005), a framework for robust estimation and inference, and most importantly allow exploring a range of conditional quantiles exposing conditional heterogeneity. For the present problem, the micro-firm heterogeneity given by unobserved characteristics (entrepreneurial ability) can be analyzed along the single dimensional conditional quantiles of the firm revenues. Along this dimension, high quantiles correspond to *high-ability entrepreneurs* and low quantiles to *low-ability entrepreneurs*. Chesher (2005) studies identification under discrete variation and shows that the identifying intervals can be estimated using QR methods. Thus, as argued in Chesher (2005), the identification through QR strategy may work for some quantiles (in our case *target* entrepreneurs) but not for others (in our case the low- and high-ability entrepreneurs). We face a similar situation where the SIMPLES program can be used for identification only for medium-ability entrepreneurs but not for low- and high-ability ones.

Our proposed estimation strategy thus combines the regression discontinuity approach and the QR framework. In this paper, we employ the linear instrumental variables quantile regression (IVQR) estimator proposed by Chernozhukov and Hansen (2006, 2008) applied to estimate a *fuzzy* regression discontinuity design model. The model is semiparametric in the sense that the functional form of the conditional distribution of the response variable given the regressors is left unspecified. The use of IVQR in a regression discontinuity design appeared in Guiteras (2008) motivated by an empirical application to the returns to compulsory schooling, and Pereda-Fernandez (2010) estimating the effects of class size on scholastic achievement. Frolich and Melly (2008) propose a nonparametric identification of the quantile treatment effects in the regression discontinuity design and they propose an uniformly consistent estimator for the potential outcome distributions and for the function-valued effects of the policy. Frandsen (2008) introduces a procedure to nonparametrically estimate local quantile treatment effects in a regression discontinuity design with binary treatment.

The rest of the paper is organized as follows. Section 2 develops a theoretical model. Section 3 describes the ECINF micro-firm survey. Section 4 describes the SIMPLES program and the identification strategy. Section 5 develops the quantile regression discontinuity estimator. Section 6 presents the econometric results. Section 7 concludes.

2 Taxes and the informal sector

In this section, we present a simple model that generates a segmentation characterized by salaried workers, informal and formal micro-entrepreneurs. The model shows that an individual becomes an informal entrepreneur, rather than being a salaried worker, if her individual ability is higher than a certain threshold, and becomes a formal entrepreneur, rather than being an informal one, if her individual ability is higher than an even higher threshold. The higher is the cost of formality the higher is the threshold value of ability to become a formal entrepreneur. This simple model builds on the models of Rauch (1991) and Paula and Scheinkman (2007, 2010). The model will then be used to analyze the impact of SIMPLES on formality.

We consider a continuum of agents, each denoted by i and characterized by entrepreneurial ability θ_i, which is distributed according to a probability density function $g(\cdot)$. Agents choose between working for an existing firm and earning a wage of w independent of their ability[5], thus becoming a salaried worker, operating a firm in the informal sector or operating a firm in the formal sector. The last two options correspond to the entrepreneurial sector. An entrepreneur produces quantity y_i of an homogeneous good using capital k_i and labor l_i as inputs. In order to maintain tractability we consider a Cobb-Douglas technology $y_i = \theta_i k_i^{\alpha} l_i^{\beta}$, with $\alpha, \beta > 0$ and $\alpha + \beta < 1$.[6]

We normalize the price of the homogeneous good to 1. The unit costs of k and l are respectively r and w, where r and w are given. We distinguish between formal and informal entrepreneurs. A formal entrepreneur pays an *ad valorem* tax ϕ. An informal entrepreneur *cheats* the system and pays no taxes, but if detected is out of business. We assume that the probability of detection p increases with the size of the firm and that $p(k) = 0$ if $k \leq k^*$ and

[5] Ability is thus only relevant when managing a firm. Modeling the salaried sector exceeds the scope of this paper.

[6] The results of the model would still apply with any concave production function.

$p(k) = 1$ if $k > k^*$, that is, an informal entrepreneur cannot employ more than k^* but is able to evade taxes.[7]

The profit functions for an entrepreneur of ability θ_i who chooses to be respectively informal or formal follow:

$$\begin{aligned} \pi_i^I &= \max_{l_i, k_i \le k^*} \{\theta_i k_i^\alpha l_i^\beta - r k_i - w l_i\}, \\ \pi_i^F &= \max_{l_i, k_i} \{(1-\phi)\theta_i k_i^\alpha l_i^\beta - r k_i - w l_i\}. \end{aligned} \tag{1}$$

The maximization of (1) gives the optimal quantity of production factors which are respectively used by an informal and a formal entrepreneur, given her ability θ_i:

$$k_i^I = \min\{\theta_i^{\frac{1}{1-\alpha-\beta}} (\frac{\alpha}{r})^{\frac{1-\beta}{1-\alpha-\beta}} (\frac{\beta}{w})^{\frac{\beta}{1-\alpha-\beta}}, k^*\},\ l_i^I = \min\{\theta_i^{\frac{1}{1-\alpha-\beta}} (\frac{\alpha}{r})^{\frac{\alpha}{1-\alpha-\beta}} (\frac{\beta}{w})^{\frac{1-\alpha}{1-\alpha-\beta}}, (\frac{\beta\theta^i k^{*\alpha}}{w})^{\frac{1}{1-\beta}}\},$$

$$k_i^F = ((1-\phi)\theta_i)^{\frac{1}{1-\alpha-\beta}} (\frac{\alpha}{r})^{\frac{1-\beta}{1-\alpha-\beta}} (\frac{\beta}{w})^{\frac{\beta}{1-\alpha-\beta}},\ l_i^F = ((1-\phi)\theta_i)^{\frac{1}{1-\alpha-\beta}} (\frac{\alpha}{r})^{\frac{\alpha}{1-\alpha-\beta}} (\frac{\beta}{w})^{\frac{1-\alpha}{1-\alpha-\beta}}.$$

When is it optimal for an entrepreneur to become formal? In choosing whether to become formal or not micro-entrepreneurs trade-off the gains of employing more than k^* with the cost of paying the tax ϕ. On one hand formality decreases productivity as it decreases the marginal products of the factors of production and such effect shows that informality can work as a device to enhance flexibility and productivity. On the other hand formality allows firms to grow bigger as it increases the production set. It is the extent of the trade-off between the two effects that determines which entrepreneurs find it optimal to become formal rather than remaining informal.

As shown by Paula and Scheinkman (2007), the convexity of the profit functions (1) in θ implies that there is a unique threshold level of ability above which entrepreneurs become formal. The following proposition formally establishes this result and finds an analytical expression for the threshold level of ability. The proof is given in the Appendix.

[7] The functional form of the probability of detection could be more general: Paula and Scheinkman (2007) show that as long as p is an increasing function of k there is still a threshold level of ability such that entrepreneurs go from informal to formal and therefore the same conclusions hold.

Proposition 1 *There exists a threshold level of ability $\bar{\theta}$ such that an entrepreneur i will decide to be formal if and only if her ability θ_i is greater than $\bar{\theta}$. $\bar{\theta}$ increases in ϕ.*

This result is driven by the fact that productivity increases in θ_i and therefore agents with higher θ_i can afford to trade-off a decrease (measured by ϕ) in the marginal product of factors for an increase of the production set. [8]

Define an ability threshold $\hat{\theta}$ such the individual with ability $\hat{\theta}$ is indifferent between becoming a salaried worker or an informal entrepreneur, hence $w = \pi^I(\hat{\theta})$. Plugging the first order conditions into (1) we find that $\hat{\theta} = (1-\alpha-\beta)^{(\alpha+\beta-1)}(r/\alpha)^{\alpha}(1/\beta)^{\beta}w^{1-\alpha}$. Therefore we have that:

if $\theta_i \leq \hat{\theta}$, then i is a salaried worker;

if $\theta_i \in (\hat{\theta}, \bar{\theta}]$, then i is an informal entrepreneur;

if $\theta_i > \bar{\theta}$, then i is a formal entrepreneur.

Effect of a policy change

If the salaried wage is fixed[9], the fact that $\bar{\theta}$ increases in ϕ (Proposition 1) implies the following corollary.

Corollary 1 *The greater the tax ϕ, the greater the cut-off level of ability $\bar{\theta}$ and the smaller the formal sector (and vice-versa).*

[8] As in Rauch (1991) and Paula and Scheinkman (2007, 2010) the weakly monotonic relationship between exogenous ability and optimal level of formality is implied by the standard assumption of convex technology. Non-convex profit functions could imply more than one crossing point, hence a non monotonic relationship over a certain range of ability, but the relationship would still be monotonic for high levels of ability if formality constraints the production set. Moreover it could be an interesting avenue for future research to analyze the possibility that ability is not exogenous but is affected by the formality/informality decision, for instance by learning dynamics.

[9] As the tax rate ϕ changes, the equilibrium wage may in principle change. *Ceteris paribus*, a decrease in the tax fosters a larger formal sector, but this effect increases in turn the demand for labor. We abstract from the possibility of a change in the salaried wage.

It is interesting to note that, those who gain the most out of a reduction in the cost of formalization from ϕ to ϕ' are the more able individuals. The following proposition shows this result, and proof is relegated to the Appendix. As we will remark, this result is due to the convexity of the technology.

Proposition 2 *The greater the individual ability θ_i is the greater is the increase in the profit $\pi(\theta_i)$ and revenue $y_i(\theta_i)$ for a decrease in the tax rate from ϕ to ϕ'.*

We illustrate the results from the propositions above using diagrams. Figure 1 illustrates the informal entrepreneurs' profit function (thick line) and the formal entrepreneurs' profit and revenue function before and after a reduction in the tax (thin and dash lines). From the figure it is possible to notice the results of Propositions 1 and 2.[10] Moreover, from the figure, it is also evident that the result of Proposition 2 would not apply to a different model in which $\pi^F(\phi')$ is not always convex for $\theta > \bar{\theta}'$.[11]

The model can be extended to the case of a lump-sum tax. In this case, the profit function of a formal entrepreneur is the following: $\pi_i^F = \max_{l_i,k_i}\{\theta_i k_i^\alpha l_i^\beta - rk_i - wl_i - \phi\}$, where ϕ now represents a lump-sum tax. In such case all the previous conclusions still hold. Figure 2 illustrates the profit function plot for this case of a lump-sum tax change.[12]

[10] We use $\alpha = 0.2, \beta = 0.7, r = 3, w = 5$. Then, it can be computed that $k^* = 3.123$ and $\theta^* = 10$. Figure 1 shows the informal entrepreneurs' profit (thick line) and those of formal entrepreneurs given $\phi = 0.2$ (thin line) and given $\phi' = 0.1$ (dash line). It can be computed that the threshold value of ability is $\bar{\theta} = 16.1$ for $\phi = 0.2$ and decreases to $\bar{\theta} = 13.2$ for $\phi' = 0.1$.

[11] These would be the case with the non-convexities described in McKenzie and Woodruff (2006), where the return to capital is higher for low-capital firms.

[12] Given values $\alpha = 0.2, \beta = 0.7, r = 3, w = 5$, it can be computed that $k^* = 3.123$ and $\theta^* = 10$. Figure 2 shows a plot of the informal entrepreneurs' profit (thick line) and those of formal entrepreneurs given $\phi = 500$ (thin line) and given $\phi' = 250$ (dash line). It can be computed that the threshold value of ability is $\bar{\theta} = 16$ for $\phi = 500$ and decreases to $\bar{\theta} = 14.5$ for $\phi' = 250$.

3 Data and descriptive statistics

We employ the Brazilian Survey of the Urban Informal Sector (Pesquisa Economia Informal Urbana, ECINF) collected in October 1997 (11 months after the introduction of the SIMPLES) by the Brazilian Statistical Institute (IBGE, Instituto Brasileiro de Geografia e Estadística). This survey is a cross-section representative of all the urban self-employed and micro-firm owners with at most five paid employees, excluding domestic workers. The stratified sampling design (in two stages) allows studying a population of units which are rare, heterogeneous and hard to detect in standard household surveys. Geographically, it covers all of the 26 Brazilian states, as well as the Federal District, and also each of the 10 Metropolitan Areas (Belém, Fortaleza, Recife, Salvador, Belo Horizonte, Vitória, Rio de Janeiro, São Paulo, Curitiba and Porto Alegre) and the municipality of Goiânia. In each of its two waves, ECINF interviewed roughly 50,000 households among which it found more than 40,000 individuals which reported owning a micro-enterprise.

We analyze firms with a government issued license as our measure of formality. Only 23.2% of all micro-firms have a license which increases to 31.1% for micro-firms with at least one paid employee.

Within the Brazilian micro-entrepreneur sector, the most frequent sectors of activity are retail trade (26% of micro-firms) and personal services (20%), followed by construction (15%), technical and professional services (11%) and manufacturing (11%). Respectively 8% and 7% of micro-firms belong to the sectors of hotels and restaurants, and transportation. Most firms are very small both in terms of revenues and employment: the average and median monthly revenues of Brazilian micro-firms were $US 1,083 and $US 600, respectively. We find that 87% of all Brazilian micro-firms have no paid employees, and 79% have no employees or partners at all, 10% of the surveyed micro-firms have one or two paid employees, and only 3%

have between 3 and 5 paid workers. In those firms with at least one paid employee, roughly 22% of all workers are family members, almost two thirds of paid workers are non-registered (sem carteira assinada) and only 35% pay social security contributions.

The ECINF asks whether respondents started their firms themselves or became owners at a later date. The survey then collects data on the number of years and months since respondents respectively started the firm or became owners-partners. We use this information to construct our time-in-business variable. For firms that were not started by their current owners, our time-in-business variable reflects the time since the current owner joined in as a partner, which is not necessarily the actual age of the firm. This problem, however, affects only 8% of firms (92% of respondents report having started their own firms) and it does not appear to have a significant impact on our main conclusions. Given that the IV strategy relies heavily on the validity of this measure we will also consider separately the subsample of micro-firms where the firm was started by the current owner.

4 The SIMPLES program and identification strategy

In November 1996, the Brazilian Government implemented a new unanticipated simplified tax system for micro and small firms, the SIMPLES. The new national system consolidated several federal taxes and social security contributions. Basically, the SIMPLES abridged procedures for the verification and payment of federal, state and municipal taxes. At the Federal level, the system allowed eligible firms to combine six different types of federal taxes and five different social security contributions into a one single monthly payment, varying from 3% to 5% of gross revenues for micro-enterprises, and from 5.4% to 7% of revenues for small firms. One important aspect of the new system is that it allowed substituting a fixed (and relatively low) percentage of total invoicing for the standard payroll contribution, which led to a substantial reduction in labor costs and hence created a strong incentive to hire new employees and/or legalize already existing labor relationships. The motivation behind these reductions in direct and indirect taxes was to enable small, unskilled labor-intensive firms to compete more effectively with larger enterprises, for which high tax burdens are more manageable due to scale economies. Moreover, while value added taxes collected at the state and municipal levels - the Imposto Sobre Circulação de Mercadorias e Prestação de Serviços (ICMS) and the Imposto Sobre Serviços (ISS) - were initially not included in SIMPLES, States and Municipalities could enter into agreements with the Federal Government to transfer to the latter the collection of the corresponding taxes through an increase in the SIMPLES rates. As a result, SIMPLES permitted an overall reduction of up to 8 percentage points in the tax burden faced by eligible firms (Monteiro and Assunção, 2006). SIMPLES, however, explicitly excluded from program eligibility all activities that by law require the employment of professionals in regulated occupations. Examples of ineligible activities include the manufacturing of chemical products, machinery and equipment, as well as education, health, accounting, insurance and financial services, among others.[13]

Given the previous model, firms' output or revenues $y_i = \theta_i k_i^\alpha l_i^\beta$ can be re-expressed as a function of formality (which can be thought of as an indicator variable with 0 and 1 and labeled with d), and entrepreneurial ability θ_i:

$$y_i = f(d_i, \theta_i).$$

As the previous section showed formality affects output through the quantity of capital as formal entrepreneurs can employ a quantity $k_i > k^*$. Net of the effect of costs of formality ϕ, an entrepreneur i would employ $k_i > k^*$ if and only if $\theta_i > \theta^*$. Therefore $f(1, \theta_i) - f(0, \theta_i) > 0, \theta_i > \theta^*$ (return to formality) and $\frac{\partial f(\cdot,\cdot)}{\partial \theta_i} = k_i^\alpha l_i^\beta > 0$ (return to ability).

As we have shown, there exists a cut-off value of ability, $\bar{\theta}$, and firms with ability above that threshold will select into formality. SIMPLES can be conceived of as a reduction in

[13]This corresponds to the indicator variable $ELIG$ below.

the cost of formalization to $\phi' < \phi$ (albeit across many margins: registration costs, labor costs etc.) that will change the cut-off value of ability from $\bar{\theta}$ to $\bar{\theta}'$ (Corollary 1). Firms that change their formality status because of SIMPLES are those with $\theta \in (\bar{\theta}', \bar{\theta}]$. This also implies that there will be a subset of firms who will not change their formality status: some will remain formal (*best* entrepreneurs), others will remain informal (*worst* entrepreneurs).

The introduction of SIMPLES by unanticipated administrative decree can be seen as an exogenous policy change that significantly altered the incentives to become formal and hence is useful to solve the endogeneity problem. The theoretical model developed above predicts that only for a segment of firms we will be able to identify the effect of formality. The reason is that we will only observe an effect of SIMPLES on those firms with $\theta \in (\bar{\theta}', \bar{\theta}]$. This is the group of firms that have a large enough θ such that the SIMPLES tax reduction makes them to re-evaluate their formality status, but not so large as to make the change in ϕ irrelevant to their formality decision. This segment contains firms that will become formal only after the reduction in taxes, and therefore we can identify β_1 by using the regression discontinuity approach described above. Note that this does not mean that for firms with $\theta < \bar{\theta}'$ or $\bar{\theta} < \theta$ formality has no effect on the firm performance variable. Rather that we cannot identify the effect of formality for those firms.

Monteiro and Assunção (2006) argue that for relatively young firms (i.e. less than 2 years old) when the firm was started clearly differentiates firms that benefit from SIMPLES from those that did not. Although all firms could benefit from SIMPLES, firms born after SIMPLES show a much higher propensity to have a license than those born before. Overall this suggests a dual process for formalization: first, a firm's decision to formalize is primarily taken at the time of its creation; second, the likelihood of becoming formal increases with time-in-business.[14] The ECINF provides some evidence on this: only 1 out of 4 licensed

[14] See the analysis for micro-firms in Mexico and other evidence for Latin American countries in Fajnzylber, Maloney, and Montes-Rojas (2009).

business made no attempt at regularizing at the time of starting up compared to 4 out of 5 non-licensed business. Thus, the decision of whether to operate formally or informally appears to be made in most cases at the time of start-up. This could be due either to costly and/or complex registration procedures, to high tax rates, or to a limited demand among very small businesses for the government services or the expanded access to markets that are associated with formality at any price. While the data do not allow us to distinguish among these different two possible explanations, 72% of the firms that do attempt to register report having no difficulties in the process.

Monteiro and Assunção (2006) exploit the first process, that is, the differential effect on licensing caused by the introduction of SIMPLES for firms born before and after it. Let $AFTER$ be an indicator for whether a firm was created before or after the SIMPLES was implemented (such that $AFTER_i = 1$ if $t_i \leq \bar{t}$ and $AFTER_i = 0$ otherwise, where firms that have been in business for at most $\bar{t}$ months were created after SIMPLES) and $ELIG$ an indicator for the eligibility status of the firm. Monteiro and Assunção (2006) the interaction of eligible/non-eligible and before/after indicators, i.e. $AFTER \times ELIG$, as an IV difference-in-differences to measure the impact of formality on firm performance.

Figure 3 plots licensing rates for firms with different dates of creation (see Section 3 for a description of the database of micro-firms used). The first two graphs plot separately eligibles and non-eligibles for all firms; the last two take only the sample of entrepreneurs that started as owners of the firm. The figures show that there is a significant jump in licensing rates for eligible firms, but no change for non-eligible firms. Moreover, the jump is observed only for firms born about the time of the introduction of SIMPLES. Then, as argued in Fajnzylber, Maloney, and Montes-Rojas (2011), the validity of $AFTER \times ELIG$ as an IV for formality crucially depends on comparing firms that were born just after and before than $\bar{t}$, i.e. $|t_i - \bar{t}| < \epsilon$ for ϵ small enough.

The regression discontinuity literature (see Hahn, Todd, and van der Klaauw, 2001; van der Klaauw, 2002) argues that an unbiased estimate of the treatment impact can be obtained by giving heavier weights to observations arbitrarily close to a discontinuity. If, conditional on a set of exogenous covariates, we assume very similar distributions of unobservable characteristics of firms born immediately before and after SIMPLES implementation, the discontinuity that the introduction of SIMPLES introduces in the factors determining formality can be exploited to provide unbiased estimates of the local average treatment effect of the program. Using this argument, Fajnzylber, Maloney, and Montes-Rojas (2011) show that the regression coefficient of $AFTER \times ELIG$ is dependent on the weighting scheme. Following these authors we will implement a *fuzzy* regression discontinuity design, where on a small enough interval about the introduction of SIMPLES, identification can be achieved by comparing firms born just before and just after the SIMPLES introduction.

The validity of the estimates of the effect of formality on revenues relies on the validity of SIMPLES as an IV. In particular, if self-selection into treatment occurred this would produce biased estimates, and the direction of the bias would depend on the correlation between those that benefit from SIMPLES treatment and unobservables. The first concern is that some firms might have strategically delayed their creation after the introduction of SIMPLES, thus changing the composition of firms before and after. Monteiro and Assunção (2006) show that SIMPLES did not produce any change in the number of starting firms as compared to similar months before (i.e. SIMPLES produced no "rush" to start a firm) and it only affected formality of eligible firms. Moreover, Fajnzylber, Maloney, and Montes-Rojas (2011) compare firms born before and after together with eligible and noneligible firms on several observable characteristics (education, age, gender, location) and find that there are no statistically significant differences. While this still does not rule out differences in unobservables, these characteristics are likely to be correlated with unobservables, and therefore they provide indirect evidence for the validity of SIMPLES as an IV. Finally, Monteiro and Assunção (2006) show that the SIMPLES effect is not due to seasonal effects (they repeat their analysis one and two years later as if SIMPLES had been introduced in November 1995 and 1994, respectively, and they found no effect) which shows that there are no intrinsic differences between firms born before and after about the November cut-off in other years.

The second concern is that SIMPLES might have changed the composition of eligible and non-eligible firms.[15] First, changes in market conditions might produce that low skilled entrepreneurs are pushed out by new entrants and excluded from the survey (which is a retrospective survey, taken one year after SIMPLES, see Section 3). Although we cannot control for potential attrition bias and sample selection, sectoral transition studies (see Fajnzylber, Maloney, and Montes-Rojas, 2006; Maloney, 1999) suggests that micro-entrepreneurs will remain within the micro-firm sector and will not become salaried workers or unemployed, hence that the micro-entrepreneurs sector will not change its overall composition. Second, firms might have strategically changed the industry or sector to become eligible. However, given that the definition of non-eligibility mostly applies to regulated and professional occupations, for an entrepreneur to change from the non-eligible to the eligible sector would require a substantial change in the goods or services offered, a possibility which seems unlikely in the short run.[16]

To summarize, our identification strategy allow us to estimate the effect of formality on firm performance for firms with $\theta \in (\bar{\theta}', \bar{\theta}]$ and born near the introduction of SIMPLES, i.e. $|t_i - \bar{t}| < \epsilon$ for ϵ small enough. This strategy requires the use of both quantile regression (to model θ) and regression discontinuity designs (to amplify the effect of SIMPLES at the time

[15] We thank Tiziano Razzolini and an anonymous referee for pointing this out.

[16] A formal analysis of the choice non-eligible vs. eligible sector and of the general equilibrium effects of a reduction in the cost of formality (see footnote 9) goes beyond the scope of the present paper.

of its introduction).

5 Quantile regression discontinuity

In order to find the threshold values $\bar{\theta}'$ and $\bar{\theta}$ we will consider the single dimensional conditional quantiles, indexed by $\tau \in (0,1)$, of the firm's revenues, y,

$$Q_y(\tau|d, x, |t_i - \bar{t}| < \epsilon) = \beta_1(\tau)d_i + \beta_2(\tau)t_i + \beta_3(\tau)x_i, \tag{2}$$

where i denotes the firm, d is a binary formality indicator (licensing), t denotes time-in-business and x is a set of exogenous covariates. If we assume that for all $\theta_1 \leq \theta_2$ there exists $0 < \tau_1 \leq \tau_2 < 1$, then this conditional quantile function can be used to find $\bar{\tau}'$ and $\bar{\tau}$ that match $\bar{\theta}'$ and $\bar{\theta}$, respectively. With the proposed identification we can estimate $\beta_1(\tau)$ for $0 < \bar{\tau}' < \tau \leq \bar{\tau} < 1$. This case was discussed by Chesher (2003) where he argued about "the possibility of identification of a structural derivative evaluated at some quantile probabilities but not at others"(p.1411).

It should be emphasized that $\beta_1(\tau)$ measures the difference in revenues due to the effect of licensing (i.e. being formal) and that the conditioning on a small interval about the introduction of SIMPLES , i.e. $|t_i - \bar{t}| < \epsilon$, does not imply this effect occurred in a given interval in time. These differences are the result of potentially multiple simultaneous effects, such as hiring more labor, capital, access to credit, operating in a fixed location, etc.[17] We only focus on the quantile heterogeneity in total revenues.

As argued in the previous section we use $z = (AFTER \times ELIG)$ as a valid instrument for d. This identification condition is discussed in Monteiro and Assunção (2006) and Fajnzylber, Maloney, and Montes-Rojas (2011). The instrumental variables quantile regression (IVQR) estimation method may be viewed as an appropriate quantile regression analog of the two

[17]We thank an anonymous referee for pointing this out.

stage least squares (2SLS) that makes use of a valid exclusion restriction. More formally, and following Chernozhukov and Hansen (2006, 2008), from the availability of an IV, z, we consider estimators defined as:

$$\hat{\beta}_1(\tau) = argmin_{\beta_1} \|\hat{\gamma}(\beta_1, \tau)\|_A, \tag{3}$$

where $\hat{\gamma}(\beta_1, \tau)$ is obtained from

$$argmin_{\beta_2,\beta_3,\gamma} \sum_{i=1}^{N} \omega(|t_i - \bar{t}|)\rho_\tau \left(y_i - \beta_1 d_i - \beta_2 t_i - \beta_3 X_i - \gamma z_i\right), \tag{4}$$

with $\omega(\cdot)$ a weighting function that is monotonically decreasing in $|t_i - \bar{t}|$, $\rho_\tau(\cdot)$ the τ-quantile regression check function, $\|x\|_A = \sqrt{x'Ax}$ and A is a positive definite matrix.[18] Differently to IV least-squares, however, it does not have a first-stage.

The asymptotic properties of the estimator are described in Chernozhukov and Hansen (2006, 2008). In particular asymptotic normality holds,

$$\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) \xrightarrow{d} N(0, J(\tau)^{-1}S(\tau)J(\tau)^{-1})$$

where $\beta = (\beta_1, \beta_2, \beta_3)'$, $J(\tau) = E[f_{\epsilon(\tau)}(0|d,t,x,z)(t,X,z)(d,t,x)']$ with $\epsilon(\tau) = y_i - \beta_1 d_i - \beta_2 t_i - \beta_3 x_i - \gamma z_i$, $f_{\epsilon(\tau)}(\cdot)$ the density function, and $S(\tau) = (\min(\tau,\tau') - \tau\tau')E[(d,t,x)(t,x,z)']$.

We refer the reader to Chernozhukov and Hansen (2005, 2006) for a more detailed discussion on the assumptions used for identification and the asymptotic results of the IVQR estimator. One important assumption for identification of the IVQR is rank invariance. This implies that, conditional on all other variables, a common unobserved factor, such as unobserved ability, determines the ranking in the outcome conditional distribution of a given subject across treatment states.[19] In our application, a firm considers a binary formality

[18] As discussed in Chernozhukov and Hansen (2006), the exact form of A is irrelevant when the model is exactly identified, but it is desirable to set A equal to the asymptotic variance-covariance matrix of $\hat{\gamma}(\alpha(\tau), \tau)$ otherwise.

[19] Chernozhukov and Hansen (2005) show that it is possible to achieve identification with IVQR using a weaker assumption called rank similarity. Rank similarity relaxes exact rank invariance by allowing unsystematic deviations, "slippages" in one's rank away from some common level.

variable, $d \in \{0,1\}$. The potential outcome under each level is given by the firm's earnings under the different licensing $\{y_d, d = 0,1\}$. We assume that the potential revenue outcomes, conditional on $X = (x,t)$, are given by (2), $Q_{y_d}(U|d,x,t) = \beta_1(U)d + \beta_2(U)t + \beta_3(U)x$, where rank $U \sim \mathrm{U}(0,1)$ indexes the unobserved heterogeneity, $\mathrm{U}(0,1)$ denotes the standard Uniform distribution, and $Q_{y_d}(U|d,x,t)$ is increasing in U. Thus, the distribution of potential outcome y_d is characterized by the quantile functions $Q_{y_d}(U|d,x,t)$. The rank variable U is assumed to be determined by entrepreneurial ability and other unobserved factors that do not vary with d. Moreover, in this model, the independence condition only requires that U is independent of the instruments z, conditional on X. Finally, the rank variable U (entrepreneurial ability) is assumed invariant to d, which ascribes an important role to conditioning on covariates X. Having a rich set of covariates makes rank invariance a more plausible approximation.

6 Econometric results

Our main goal is the estimation of eq. (2), that is, the conditional quantiles of the logarithm of total revenues. In order to implement this we follow the strategy described in Section 4 where $AFTER \times ELIG$ is used as an IV for having a license.[20] We increase the power of the instrument by interacting it with gender and age of the entrepreneur. Moreover, we use the same weighting scheme as in Fajnzylber, Maloney, and Montes-Rojas (2011) with $\omega(|t_i - \bar{t}|) = f(0, |t_i - \bar{t}|)$, where $f(0,\sigma)$ is the normal density of a standard Gaussian random variable with mean 0 and standard deviation σ.

Our measure of firm performance y is the logarithm of total monthly revenues. Unfortunately, we cannot apply the same analysis to profits, because this would need additional instruments for both capital and labor, which are endogenous and affected by SIMPLES.

[20]The implied first-stage regression is $License_i = \alpha_1 AFTER_i + \alpha_2 ELIG_i + \alpha_3(AFTER_i \times ELIG_i) + \alpha_4 x_i + e_i$.

Moreover, there may be measurement errors in the cost of capital and imputation of the owner's salary. These are potentially large in micro-firms surveys. Therefore, the return to formality is the ultimate effect on revenues arising from several channels: hiring both more labor and capital, higher productivity, more business opportunities, access to credit, etc. This effect may also include changes in the composition of clients as in Paula and Scheinkman (2007) model. As additional control variables x we use the $AFTER$, $ELIG$, gender (dummy for female), age and education of the entrepreneur (the latter as categorical dummies, base category: no formal education), number of members in the household, a set of dummy variables for the reasons to become an entrepreneur, time in business (interacted with $AFTER$ and as a square polynomial), and dummy variables by industry and state.

Tables 1 and 2 present the 2SLS and IVQR estimates of the conditional mean and quantiles (selected quantiles) of firm revenues for the selected weighting scheme described above for all and for those entrepreneurs that started as owners, respectively. Figures 4 and 5 summarizes the effect of licensing on firm revenues.

The figures show that the effect of licensing is not statistically significant for $\tau < 0.10$ and $\tau > 0.60$ ($\tau > 0.50$ for the sample of original owners). This suggest that, in terms of the characterization proposed in the this paper, $\bar{\theta}' = 0.10$ and that therefore, 10% of the sample corresponds to the entrepreneurs that did not benefit from SIMPLES because they opted out of formality even after the tax reduction. Moreover $\bar{\theta} = 0.50(0.60)$, and then the upper 50% (40%) of the sample were already considering that the cost of formality was not very high. For these segments, we cannot identify the effect of formality through the introduction of SIMPLES. Taking the complement of those groups, we define the *target* population given by $0.10 \leq \tau \leq 0.50$ or $0.10 \leq \tau \leq 0.60$ depending on the sample. Note that for this group the effect is roughly similar to the 2SLS estimate.

Note however that the point estimates being non statistically significant does not imply that the instruments are not working and that the effect of licensing cannot be identified. In fact, this cannot be a priori be distinguished from it being statistically equal to zero. The lack of a first stage does not allow us to use the OLS techniques for evaluating the IV performance. Therefore, we propose a new procedure based on the Chernozhukov and Hansen (2006, 2008) estimator. If the identification strategy using the IV works well, then $\hat{\gamma}(\beta_1, \tau)$, based on eq. (3), should have a clear global minimum. If, however, the IV is not appropriate, it should not have a clear minimum. We thus plot several graphs of $(\hat{\gamma}(\tau), \beta_1)$ for different quantiles τ and analyze them. Figures 6 and 7 report these for both samples and $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$. From the graphs it can be noted that only for $\tau \in \{0.25, 0.50\}$ the function is convex almost everywhere with a clear minimum, but it is less so for the remaining quantiles. This implies that the lack of significance in $\hat{\beta}_1$ is associated with an IV that does not satisfy the Chernozhukov and Hansen (2006, 2008) identification criterion.

The 2SLS point estimate is 3.40 (std.err. 1.04) for all firms and 3.23 (std.err. 0.97) for the owners subsample. Note that the subsample of firms whose current entrepreneur was the original owner has higher standard errors. These high and rather imprecise estimates are similar in magnitude to those in Monteiro and Assunção (2006) and Fajnzylber, Maloney, and Montes-Rojas (2011). Moreover, although not reported, similar point estimates are obtained in levels if we compute the correponding percentage increment. As a result the large log estimates appear because of the fact that firms have in fact low levels of revenues. Overall, they clearly point out that formality (licensing) has a positive effect on firms' revenues. In fact, these high positive effects are observed for all quantiles, although as mentioned above the effect is statistically significant only for the *target* population.

To examine the heterogeneity associated with the IVQR estimates we perform diagnosis tests using Kolmogorov-Smirnov tests.[21] First, we test the hypothesis of a zero constant coefficient for the IVQR estimates across quantiles, that is, we test the hypothesis that $H_0 : \beta_1(\tau) = 0$. In order to implement the test, we estimate the model for $\tau \in [0.1, 0.9]$, compute the Wald statistic for each particular quantile and take the maximum over the corresponding quantiles. The results for the test statistics are 27.83 and 21.74 for the all micro-firms and owners samples, respectively. These results strongly reject the null hypothesis at the 1% level of significance (the critical values are: 12.69 at 1% level of significance, 9.31 at 5% level of significance, and 7.63 at 10% level of significance). Thus, there exists strong evidence to reject the hypothesis of zero or negative impact of licensing on log revenues.

Secondly, we test the hypothesis of a constant given effect of SIMPLES on revenues, that is, $H_0 : \beta_1(\tau) = \bar{\beta}$, where we set $\bar{\beta}$ as the 2SLS estimate. The results for the tests statistics are 9.43 and 6.53 for all micro-firm and owners samples respectively, such that we reject the null at 5% level of significance for the first case. Thus, although the confidence interval of the IVQR contains the point estimate of 2SLS, for various intermediate quantiles, the evidence suggests that the effect of SIMPLES on revenues is heterogeneous. However, in the second sample the wide confidence intervals made the 2SLS estimate to remain inside the bands and we cannot reject the null hypothesis.

Finally, we apply the latter test, $H_0 : \beta_1(\tau) = \bar{\beta}$, only over the selected quantiles where we have evidence of identification of the parameters of interest, that is, for $\tau \in [0.10, 0.60]$ ($\tau \in [0.10, 0.50]$ for the sample of original owners).[22] In this case, the results for the test statistics are 11.08 and 7.57 for all micro-firms and owners subsamples, respectively, such that we reject the null at 5% level of significance for the first case, and at 10% for the second case. This shows that there is heterogeneity within the target group segment. In fact, we

[21] Kolmogorov-Smirnov test in QR are discussed in Chernozhukov and Hansen (2006) and Koenker (2005).

[22] In general the index used for Kolmogorov-Smirnov tests in QR is symmetric of the form $[\epsilon, 1-\epsilon]$. However, in some situations it is desirable to restrict the interval of estimation to to a subinterval, as $[\tau_0, \tau_1] \in (0, 1)$. As Koenker (2005) discusses, this can be easily accommodated by using a renormalized statistic.

observe that the effect is actually decreasing on τ for this range. This result contradicts that in Proposition 2 and could be due to the non-convexities described in McKenzie and Woodruff (2006), where the return to capital is higher for low-capital firms. Overall, this suggests that, over the range of identified quantiles, the formality treatment has a bigger impact on low quantiles than in high quantiles.

The study of the covariate effects is of independent interest too. The negative coefficient of Female reflects the fact that women engage in less profitable activities, possibly due to household commitments or outright gender discrimination.[23] There is no clear pattern across quantiles, which determines that the gender effect applies uniformly to all types of firms. Education is non-monotonic for the conditional mean model and for low quantiles. In those cases, incomplete secondary education has the highest effect in both subsamples. However, education becomes monotonically increasing for $\tau \geq 0.5$. This determines that for firms in the low conditional quantiles, higher education is not necessarily associated with higher revenues, but it is with outstanding firms. Finally, the reasons to become entrepreneur show interesting variability across quantiles. Reasons such as "Accumulated experience", "Be independent", "Make a good deal" and "Profitable business" which may be associated with entrepreneurs with high ability are larger for high quantiles, while reasons for low ability entrepreneurs (such as "To help family income") are larger for the low quantiles.

We also implement the method of Frolich and Melly (2008) and Frandsen (2008) for comparison reasons. This estimator differs in several aspects to the one proposed here. First, it corresponds to a standard regression discontinuity design and is not designed to be used in a difference-in-differences fashion. In our set-up we implement this estimator by comparing only treated (born after SIMPLES) and non-treated (born before SIMPLES) considering a discontinuity in age of the firm. Second, as a nonparametric estimator, it posses difficulties

[23]However, as argued by an anonymous referee, it is also the case that women engage in less risky activities, and it is not necessarily the case that more risk is optimal.

with a large set of covariates. Thus, we implement the estimator without covariates and then, following Frolich and Melly (2008), we use an alternative parametric specification using the propensity score ($Prob[t_i \geq \bar{t}|d, x, |t_i - \bar{t}| < \epsilon]$) as a unique conditioning variable. Third, standard errors are available only for the case without covariates, and therefore only point estimates are provided for the case with covariates. Finally, the choice of bandwidth is always an important concern in nonparametric and semiparametric estimation, and estimates may have large variation depending on the bandwidth. We therefore use 3 different choices of bandwidth.

We estimate the model using the subsample of all micro-firms.[24] The results for both estimators, with and without covariates, are presented in Table 3. Regarding the case with no covariates, there are only a few quantiles where the point estimates are statistically different from zero. The point estimates for the bandwidths 2 and 3 are somehow similar to the IVQR estimates, while those for a bandwidth of 4 are negative and are not statistically different from zero, evidencing the sensitivity to the bandwidth choice. When covariates are used through the propensity score, the point estimates are reduced to 1.1 on average. These point estimates provide additional evidence on formality having a positive effect on revenues. As mentioned above, the lack of a measure of dispersion precludes us to provide any inference on these estimates. Thus, we are not able to statistically analyze the question posed in the paper regarding which firms benefit from the reduction in formality costs. However, given the large standard errors for the IV estimates presented in Tables 1 and 2, in most cases, these nonparametric estimates are included in the 95% confidence intervals of the estimates discussed above.

[24] Similar results are obtained for the owners-only subsample. Results are available from the Authors upon request.

7 Conclusion and policy implications

The econometric results are summarized as follows. First, the results show positive point estimates evidencing that formality has a positive effect on revenues. Overall this confirms the effect of formality on firm performance is positive and suggests that formality gains are potentially large. From a policy perspective this implies that improving institutions to increase participation benefits the micro-firm sector. Reducing the cost of formality allows firms to approach the steady state size dictated by their intrinsic entrepreneurial ability.

Second, the answer to the question "which firms benefit from the tax reduction and simplification?" is given by the estimates from the empirical exercise showing that the *target* population corresponds to τ quantiles in $0.10 \leq \tau \leq 0.50$ or $0.10 \leq \tau \leq 0.60$ depending on the sample. This means that SIMPLES had a potential effect on 40% to 50% of the micro-entrepreneur population, mostly concentrated on low ability firms. Note that this corresponds to benefits in terms of changing formality status (i.e. becoming formal) not on the overall effect of SIMPLES, because SIMPLES also had benefits for those already formal that would face lower taxes. The theoretical model also shows that the larger is the tax reduction, the larger will be the segment of firms that will change their formality status.

Third, for the *target* group where the effect of formality can be identified, we find evidence of heterogeneity across quantiles on the impact of license on the conditional distribution of revenues. These estimates suggest that reducing the cost of formality might significantly benefit low ability firms more. However, these effects can only be studied for the quantiles where the effect of formality can be identified, and therefore, we cannot offer a complete analysis of the heterogeneity in the effect of formality on revenues.

References

Chernozhukov, V., and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261.

——— (2006): "Instrumental Quantile Regression Inference for Structural and Treatment Effects Models," *Journal of Econometrics*, 132, 491–525.

——— (2008): "Instrumental Variable Quantile Regression: A Robust Inference Approach," *Journal of Econometrics*, 142, 379–398.

Chesher, A. (2003): "Identification in Nonseparable Models," *Econometrica*, 71, 1405–1441.

——— (2005): "Nonparametric Identification under Discrete Variation," *Econometrica*, 73, 1525–1550.

de Soto, H. (1989): *The Other Path: The Invisible Revolution in the Third World.* Harper and Row, New York, NY.

Fajnzylber, P., W. F. Maloney, and G. Montes-Rojas (2006): "Microenterprise Dynamics in Developing Countries: How Similar are They to Those in the Industrialized World? Evidence from Mexico," *World Bank Economic Review*, 20, 389–419.

——— (2009): "Releasing Constraints to Growth or Pushing on a String? Policies and Performance of Mexican Micro-firms," *Journal of Development Studies*, 45, 1027–1047.

——— (2011): "Does Formality Improve Micro-firm Performance? Evidence from the Brazilian SIMPLES Program," *Journal of Development Economics*, 94, 262–276.

Frandsen, B. R. (2008): "A Nonparametric Estimator for Local Quantile Treatment Effects in the Regression Discontinuity Design," MIT, mimeo.

Frolich, M., and B. Melly (2008): "Quantile Treatment Effects in the Regression Discontinuity Design," IZA DP No. 3638.

Gerxhani, K. (2004): "The Informal Sector in Developed and Less Developed Countries: A Literature Survey," *Public Choice*, 120, 267–300.

González, D. (2006): "Regímenes Especiales de Tributación para Pequeños Contribuyentes en América Latina," Banco Interamericano de Desarrollo.

Guiteras, R. (2008): "Estimating Quantile Treatment Effects in a Regression Discontinuity Design," University of Maryland, mimeo.

Hahn, J., P. Todd, and W. van der Klaauw (2001): "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 69, 201–209.

Harris, J., and M. Todaro (1970): "Migration, Unemployment and Development: A Two-Sector Analysis," *American Economic Review*, 40, 126–142.

Kaplan, D., E. Piedra, and E. Seira (2006): "Are Burdensome Registration Procedures an Important Barrier on Firm Creation? Evidence from Mexico," SIEPR Discussion Paper 06-13.

Koenker, R. (2005): *Quantile Regression*. Cambridge University Press, Cambridge.

Koenker, R., and K. Hallock (2001): "Quantile Regression," *Journal of Economic Perspectives*, 15(1), 143–156.

Maloney, W. (1999): "Does Informality Imply Segmentation in Urban Labor Markets? Evidence from Sectoral Transitions in Mexico," *World Bank Economic Review*, 13, 279–302.

——— (2004): "Informality Revisited," *World Development*, 32, 1159–1178.

Mandelman, F., and G. Montes-Rojas (2009): "Is Self-Employment and Microentrepreneurship a Desired Outcome?," *World Development*, 37, 1914–1925.

McKenzie, D., and C. Woodruff (2006): "Do Entry Costs Provide an Empirical Basis for Poverty Traps? Evidence from Mexican Microenterprises," *Economic Development and Cultural Change*, 55, 3–42.

Monteiro, J. C. M., and J. J. Assunção (2006): "Outgoing the Shadows: Estimating the Impact of Bureaucracy Simplification and Tax Cut on Formality and Investment," European Meeting of the Econometric Society, Vienna.

Paula, A., and J. Scheinkman (2007): "The informal sector," NBER Working Paper 13486.

——— (2010): "Value Added Taxes, Chain Effects and Informality," *American Economic Journal: Macroeconomics*, 2, 195–221.

Pereda-Fernandez, S. (2010): "Quantile Regression Discontinuity: Estimating the Effect of Class size on Scholastic Achievement," Master Thesis CEMFI No. 1002.

Rauch, J. (1991): "Modelling the Informal Sector Formally," *Journal of Development Economics*, 35, 33–47.

van der Klaauw, W. (2002): "Estimating the effect of financial aid offers on college enrollment: A regression-discontinuity approach," *International Economic Review*, 43, 1249–1287.

Appendix

Proof of proposition 1

An entrepreneur with ability $\theta_i \leq \theta^*$ always finds optimal to be informal. An entrepreneur with ability $\theta_i > \theta^*$ finds optimal to become formal if and only if $\pi_i^F \geq \pi_i^I$. Plugging the first order conditions into (1) we obtain that

$$\pi^I(\theta^*) = (1-\alpha-\beta)\theta^{*\frac{1}{1-\alpha-\beta}}(\frac{\alpha}{r})^{\frac{\alpha}{1-\alpha-\beta}}(\frac{\beta}{w})^{\frac{\beta}{1-\alpha-\beta}}$$

and

$$\pi^F(\theta_i) = (1-\alpha-\beta)((1-\phi)\theta_i)^{\frac{1}{1-\alpha-\beta}}(\frac{\alpha}{r})^{\frac{\alpha}{1-\alpha-\beta}}(\frac{\beta}{w})^{\frac{\beta}{1-\alpha-\beta}}.$$

An entrepreneur with ability $\theta_i > \theta^*$ who decides to be informal will choose capital k^* and labor $l^I(k^*, \theta^i) = (\frac{\beta\theta^i k^{*\alpha}}{w})^{\frac{1}{1-\beta}}$. Defining $\gamma_i \equiv \theta_i/\theta^* - 1$ we can re-express $\theta_i = (1+\gamma_i)\theta^*$ and $l^I(k^*, \theta^i) = (1+\gamma_i)l^*$. Plugging k^* and $l^I(k^*, \theta_i)$ into the expression for the profit of a formal entrepreneur we obtain that $\pi^I(\theta_i) = (1+\gamma_i)^{\frac{1}{1-\beta}}(1-\alpha/(1+\gamma_i)^{\frac{1}{1-\beta}} - \beta)\theta^{*\frac{1}{1-\alpha-\beta}}(\frac{\alpha}{r})^{\frac{\alpha}{1-\alpha-\beta}}(\frac{\beta}{w})^{\frac{\beta}{1-\alpha-\beta}}$.

Therefore we obtain that $\pi^I(\theta_i) > \pi^F(\theta_i)$ if and only if $\frac{(1+\gamma_i)^{\frac{\alpha}{(1+\beta)(1-\alpha-\beta)}}}{1-\alpha/(1+\gamma_i)^{\frac{1}{1-\beta}}-\beta} < \frac{1}{(1-\alpha-\beta)(1-\phi)^{\frac{1}{1-\alpha-\beta}}}$. The left hand side

$$\frac{(1+\gamma_i)^{\frac{\alpha}{(1+\beta)(1-\alpha-\beta)}}}{1-\alpha/(1+\gamma_i)^{\frac{1}{1-\beta}}-\beta} \tag{5}$$

of the inequality above increases in γ_i as the derivative of (5) $d(\cdot)/d\gamma_i = \left(\frac{\alpha(1-x)}{(1-\alpha-\beta)}x^{\frac{\alpha}{(1-\beta)(1-\alpha-\beta)}-1}\right)/D^2$, where $D \equiv$ denominator of (5), $x \equiv (1+\gamma)^{-\frac{1}{1-\beta}}$ and $0 < x < 1$.

Define $\bar{\gamma}$ such that the condition above is satisfied with equality. This condition identifies a threshold level of ability $\bar{\theta} = (1+\bar{\gamma})\theta^*$ such that an entrepreneur i decides to become formal if and only if $\theta_i > \bar{\theta}$.

Notice that the right hand side of the inequality increases in ϕ therefore $\bar{\gamma}$ and $\bar{\theta}$ increase in ϕ. *QED*

Proof of proposition 2

The second cross-derivative $\frac{d^2\pi^F(\cdot)}{d\theta d\phi}$ is negative. Therefore the difference $(\pi^F(\phi') - \pi^F(\phi))$, where $\phi' < \phi$, increases in θ. This proves the proposition for formal entrepreneurs. $\pi^F(\phi')$ increases in θ at a faster rate than $\pi^F(\phi)$ as $\frac{d^2\pi^F(\cdot)}{(d\theta)^2}$ is decreasing in ϕ. The result of proposition 1 (single crossing between π^F and π^I) implies that $\pi^F(\phi)$ increases at a faster rate than π^I for $\theta < \bar{\theta}$. Therefore it must be the case that $(\pi^F(\phi')$ increases at a faster rate than π^I for $\theta \in [\bar{\theta}', \bar{\theta}]$, where $\bar{\theta}'$ is the new cut-off level of ability given ϕ'. Therefore this proves the proposition also for those entrepreneurs that change their status from informal to formal as a result of the policy change.

Plugging the first order conditions into the expression for output/revenue $y_i = \theta_i k_i^\alpha l_i^\beta$ we obtain that

$$y^I(\theta^*) = \theta^{*\frac{1}{1-\alpha-\beta}} (\frac{\alpha}{r})^{\frac{\alpha}{1-\alpha-\beta}} (\frac{\beta}{w})^{\frac{\beta}{1-\alpha-\beta}}$$

and

$$y^F(\theta_i) = ((1-\phi)\theta_i)^{\frac{1}{1-\alpha-\beta}} (\frac{\alpha}{r})^{\frac{\alpha}{1-\alpha-\beta}} (\frac{\beta}{w})^{\frac{\beta}{1-\alpha-\beta}}.$$

represent respectively revenues for informal and formal entrepreneurs. It is immediate to notice that the revenue functions behave exactly as the profit functions. *QED*

Figures and tables

Figure 1: Ad-valorem tax. Profit functions: informal (thick line), formal (thin line), formal after decrease in tax (dash line)

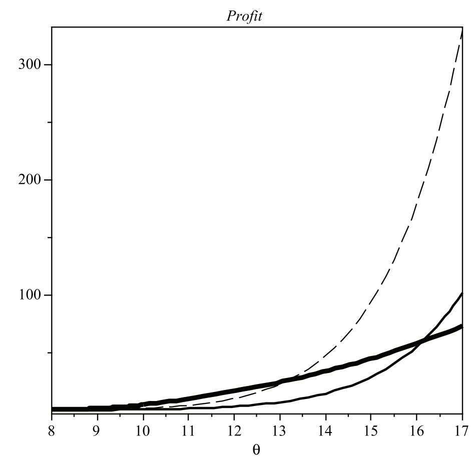


Figure 2: Lump-sum tax. Profit functions: informal (thick line), formal (thin line), formal after decrease in tax (dash line)

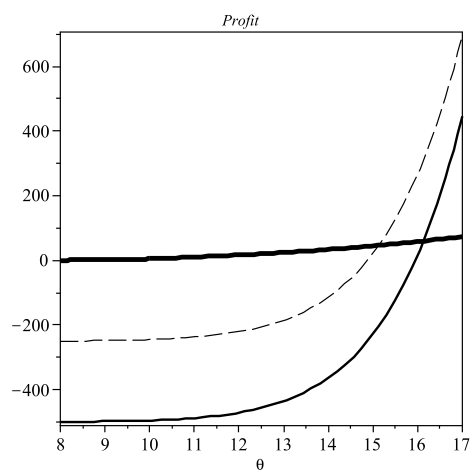
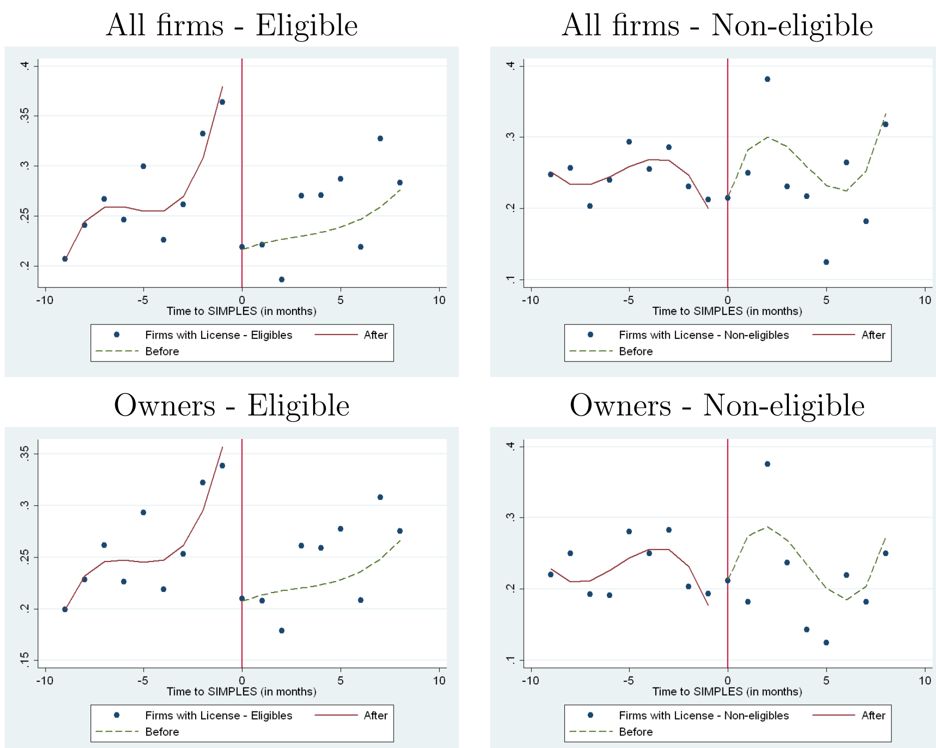
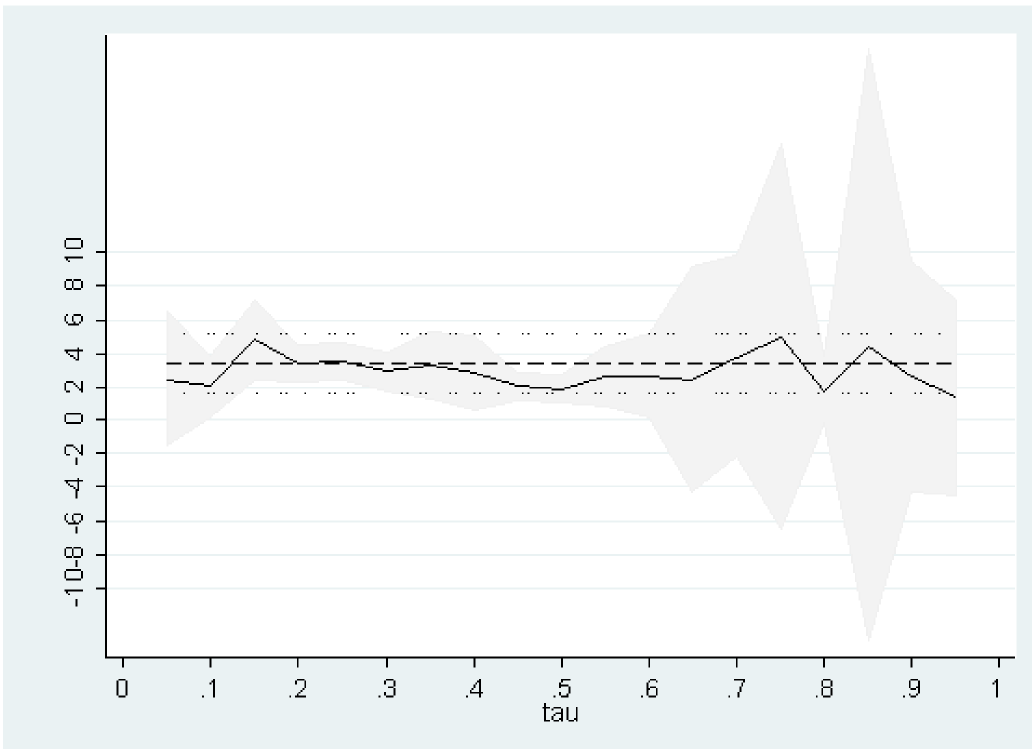


Figure 3: Average licensing rates by month of firm creation



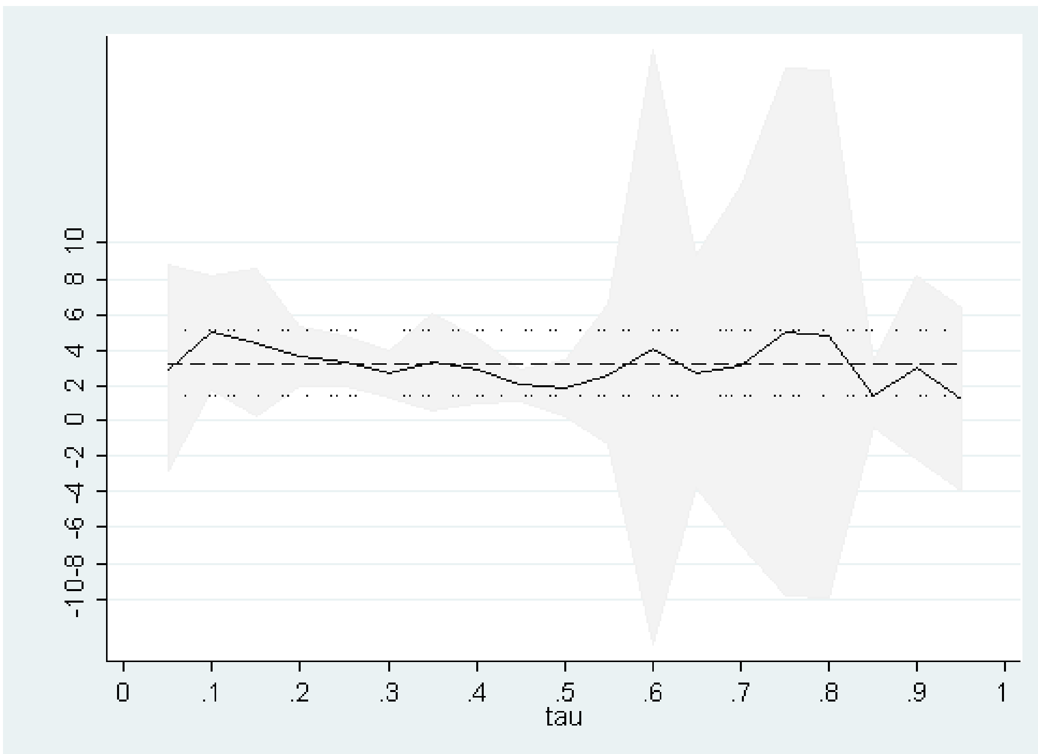
Notes: Average licensing rates by reported month of firm creation. Owners: Original owners of the micro-firm.

Figure 4: Quantile regression, all micro-firms



Notes: Plot for 2SLS and IVQR estimates with their corresponding 95% confidence intervals. Y-axis contains the coefficient estimates and X-axis the quantiles. The dashed horizontal line is the 2SLS estimate, and the dotted lines the corresponding confidence interval. The solid line is the IVQR estimate, and the shadow its corresponding confidence interval.

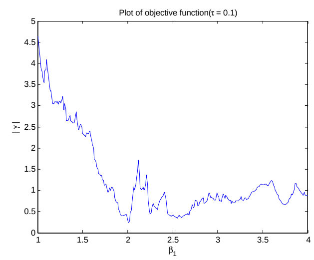
Figure 5: Quantile regression, started firm as owner

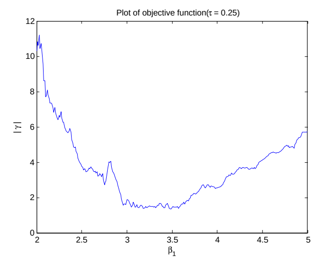


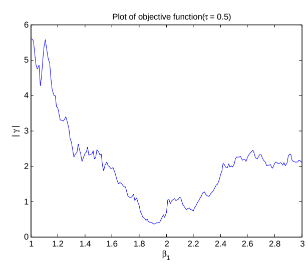
Notes: Plot for 2SLS and IVQR estimates with their corresponding 95% confidence intervals. Y-axis contains the coefficient estimates and X-axis the quantiles. The dashed horizontal line is the 2SLS estimate, and the dotted lines the corresponding confidence interval. The solid line is the IVQR estimate, and the shadow its corresponding confidence interval.

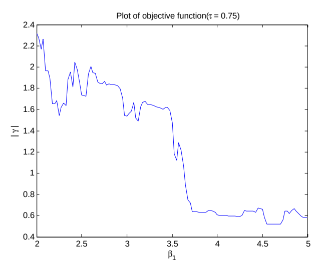
Figure 6: Validity of the IV

Notes to Figure 6: Plot of the function $\|\hat{\gamma}\|$ – All micro-firms. Y-axis contains the estimates of $\|\hat{\gamma}\|$ and X-axis β_1. Selected quantiles $\tau = \{0.10, 0.25, 0.50, 0.75, 0.90\}$.









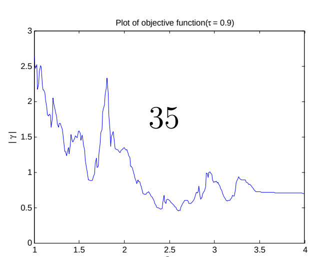
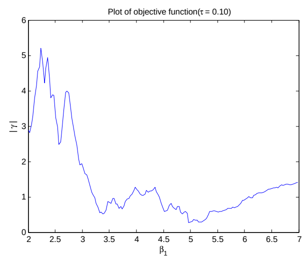
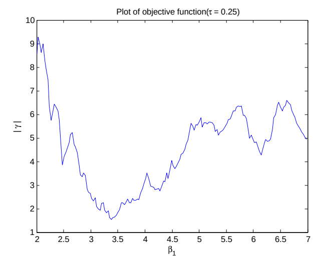
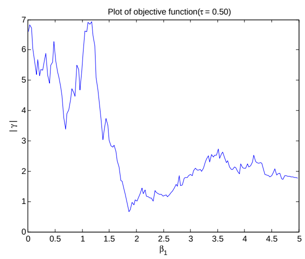


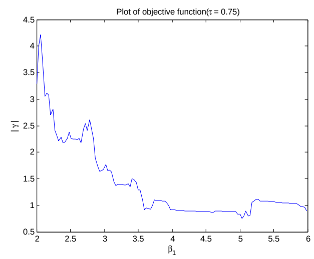
Figure 7: Validity of the IV

Notes to Figure 7: Plot of the function $\|\hat{\gamma}\|$ – Owners. Y-axis contains the estimates of $\|\hat{\gamma}\|$ and X-axis β_1. Selected quantiles $\tau = \{0.10, 0.25, 0.50, 0.75, 0.90\}$.









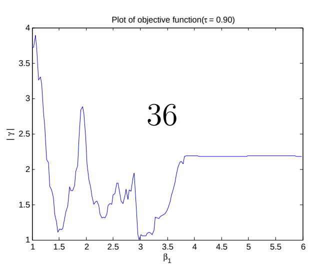


Table 1: Quantile Regression Discontinuity Analysis - All micro-firms

	IV Least-squares regression	IV Quantile regression $\tau = 0.1$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.9$
License	3.40***	2.03*	3.48***	1.90***	4.92	2.60
	(1.04)	(1.09)	(0.66)	(0.49)	(6.93)	(4.15)
Female	-0.546***	-0.676***	-0.292	-0.587***	-0.474***	-0.538***
	(0.075)	(0.162)	(0.200)	(0.100)	(0.111)	(0.120)
Age	0.0039**	-0.0021	0.021	0.016***	0.018	0.030**
	(0.020)	(0.0059)	(0.006)	(0.004)	(0.015)	(0.14)
Education categories (base: no formal education)						
Primary inc	0.334***	0.195	0.425	0.672***	0.988**	1.24***
	(0.090)	(0.253)	(0.296)	(0.136)	(0.414)	(0.16)
Primary comp	0.411***	0.135	0.555*	0.918*	1.19**	1.52***
	(0.119)	(0.388)	(0.329)	(0.49)	(0.47)	(0.37)
Secondary inc	0.735 ***	0.562**	1.15***	1.16***	1.37***	1.66***
	(0.111)	(0.313)	(0.36)	(0.16)	(0.46)	(0.21)
Secondary comp	0.591***	0.632**	0.633*	1.21***	1.39**	1.90***
	(0.196)	(0.306)	(0.351)	(0.17)	(0.58)	(0.23)
College inc	0.573*	0.717	0.764*	1.41***	1.75***	2.08***
	(0.301)	(0.492)	(0.455)	(0.47)	(0.57)	(0.50)
Reasons to become entrepreneur (base: *Did not find a job*)						
Profitable business	0.402*	0.968**	-0.103	0.513	1.136**	1.64**
	(0.287)	(0.441)	(0.614)	(0.441)	(0.454)	(0.65)
Flexible hours	0.227*	-0.022	0.397	0.127	0.369	0.476
	(0.132)	(0.338)	(0.496)	(0.184)	(0.386)	(0.445)
Be independent	0.127	0.350	0.048	0.409***	0.390**	0.472
	(0.165)	(0.286)	(0.268)	(0.118)	(0.165)	(0.322)
Family tradition	-0.230	-0.526	0.030	0.494**	0.334	0.689
	(0.302)	(1.225)	(0.354)	(0.214)	(0.427)	(1.304)
To help family income	-0.204***	-0.469**	-0.152	-0.171*	-0.023	-0.029
	(0.060)	(0.211)	(0.203)	(0.110)	(0.120)	(0.156)
Accumulated experience	0.330**	0.530**	0.447**	0.422***	0.407	0.909
	(0.151)	(0.230)	(0.244)	(0.158)	(0.519)	(0.912)
Make good deal	0.090	-0.070	0.061**	0.409**	0.558***	0.395
	(0.136)	(0.470)	(0.301)	(0.153)	(0.211)	(0.405)
As a secondary job	0.558***	1.013***	0.886**	0.380	0.968**	0.768**
	(0.178)	(0.413)	(0.495)	(0.338)	(0.431)	(0.353)

Notes: 6741 observations. Standard errors in parenthesis. Instrumental variables: $AFTER \times ELIG$ interacted with gender and age of the entrepreneur. See text for additional details. *** Significant at the 1% level; **Significant at the 5% level; *Significant at the 10% level.

Table 2: Quantile Regression Discontinuity Analysis - Owners

	IV Least-squares regression	IV Quantile regression $\tau = 0.1$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.9$
License	3.23***	4.97***	3.37***	1.87**	5.00	2.98
	(0.97)	(1.61)	(0.73)	(0.82)	(7.53)	(2.65)
Female	-0.549***	-0.034	-0.317*	-0.577***	-0.482***	-0.421***
	(0.077)	(0.382)	(0.176)	(0.095)	(0.112)	(0.135)
Age	0.0043**	0.015	0.021***	0.015***	0.019	0.027**
	(0.019)	(0.012)	(0.006)	(0.004)	(0.017)	(0.11)
Education categories (base: no formal education)						
Primary inc	0.294***	-0.364	0.291	0.606***	0.968*	1.17***
	(0.095)	(0.686)	(0.258)	(0.164)	(0.459)	(0.20)
Primary comp	0.391***	-0.058	0.480	0.863***	1.17	1.37***
	(0.121)	(0.772)	(0.293)	(0.157)	(0.51)	(0.32)
Secondary inc	0.718***	0.307	1.05	1.09***	1.42***	1.66***
	(0.111)	(0.883)	(0.30)	(0.17)	(0.52)	(0.24)
Secondary comp	0.553***	-0.014	0.570	1.14***	1.36**	1.74***
	(0.201)	(1.054)	(0.320)	(0.18)	(0.64)	(0.24)
College inc	0.647**	0.487	0.728	1.52***	1.88***	2.04***
	(0.278)	(1.013)	(0.512)	(0.45)	(0.66)	(0.37)
Reasons to become entrepreneur (base: *Did not find a job*)						
Profitable business	0.222	-0.201	-0.106	0.685	0.863	1.71***
	(0.300)	(0.961)	(0.747)	(0.690)	(0.742)	(0.36)
Flexible hours	0.387***	0.853	0.325	0.177	0.369	0.770
	(0.140)	(0.690)	(0.400)	(0.208)	(0.366)	(0.478)
Be independent	0.182	-0.257	0.089	0.445***	0.384**	0.367*
	(0.146)	(0.433)	(0.258)	(0.120)	(0.158)	(0.226)
Family tradition	0.172	-0.618	0.189	0.688***	0.486	1.00**
	(0.262)	(1.257)	(0.342)	(0.255)	(0.387)	(0.496)
To help family income	-0.224***	-0.104	-0.208	-0.210**	-0.062	-0.063
	(0.058)	(0.301)	(0.205)	(0.113)	(0.132)	(0.174)
Accumulated experience	0.323**	-0.017	0.393*	0.426**	0.395	0.944*
	(0.148)	(0.675)	(0.246)	(0.197)	(0.555)	(0.592)
Make good deal	0.084	-0.452	0.050	0.448**	0.526***	0.370*
	(0.132)	(0.437)	(0.298)	(0.193)	(0.203)	(0.193)
As a secondary job	0.657***	1.58***	1.03***	0.478	1.00**	0.569**
	(0.194)	(0.64)	(0.337)	(0.311)	(0.411)	(0.228)

Notes: 6300 observations. Standard errors in parenthesis. Instrumental variables: $AFTER \times ELIG$ interacted with gender and age of the entrepreneur. See text for additional details. *** Significant at the 1% level; **Significant at the 5% level; *Significant at the 10% level.

Table 3: Nonparametric analysis without and with covariates – All micro-firms

	Without Covariates			With Covariates		
quantiles	Band=2	Band=3	Band=4	Band=2	Band=3	Band=4
$\tau = 0.1$	5.586	3.832	-3.011	1.194	1.281	1.281
	(5.92)	(3.74)	(2.30)	–	–	–
$\tau = 0.2$	4.500	3.832	-2.606	1.099	1.099	1.099
	(7.16)	(3.70)	(2.99)	–	–	–
$\tau = 0.3$	4.605	3.817	-2.548	1.066	0.971	1.012
	(6.17)	(5.16)	(1.50)	–	–	–
$\tau = 0.4$	4.700	4.209	-2.534*	1.130	1.003	1.099
	(11.94)	(3.83)	(1.36)	–	–	–
$\tau = 0.5$	4.423*	4.081	-2.485**	1.110	1.110	1.099
	(2.30)	(3.89)	(1.24)	–	–	–
$\tau = 0.6$	4.423*	4.159	-2.659	1.099	1.012	1.107
	(2.40)	(4.27)	(2.05)	–	–	–
$\tau = 0.7$	4.423*	4.338	-3.079	1.163	1.163	1.163
	(2.49)	(4.65)	(2.69)	–	–	–
$\tau = 0.8$	4.423*	4.232	-3.344*	1.139	1.139	1.281
	(2.59)	(3.48)	(1.77)	–	–	–
$\tau = 0.9$	4.605	4.232	-2.784	1.124	1.046	1.225
	(9.23)	(3.47)	(1.78)	–	–	–

Notes: 6741 observations. Band = Bandwidth. Standard errors in parenthesis. *** Significant at the asymptotic 1% level; **Significant at the asymptotic 5% level; *Significant at the asymptotic 10% level.