Do Lower Taxes Reduce Informality?
Evidence from Brazil[*]

Rudi Rocha
FGV EAESP

Gabriel Ulyssea
University of Oxford

Laísa Rachter
FGV EPGE

March 26, 2018

Abstract

We exploit a unique, large-scale formalization program in Brazil to estimate the effects of reducing the costs of formality on firm formalization. We rely on both firm-level administrative data and individual panel data to show that reducing taxes once registration costs have already been eliminated reduces firm informality. This effect comes from the formalization of existing informal firms, and not from the creation of new formal businesses nor greater formal firm survival. The implied formalization elasticity is otherwise low, and our cost-benefit analysis indicates that the program led to net losses in tax revenues.

Key words: Taxes, informality, firm creation.
JEL code: O17, H32, H26.

1 Introduction

The vast majority of firms in developing countries are small, having at most one or two employees (Bruhn and McKenzie, 2014; Hsieh and Olken, 2014). Informality is very high among these firms, which means that they do not pay taxes nor comply with the relevant laws and regulations.[1] These high informality levels might have important consequences for development for at least two reasons. First, they reduce the tax base thus hindering fiscal capacity, which is an important constraint on growth in developing countries (e.g. Besley and Persson, 2013). Second, informality might imply substantial misallocation of resources, as these small informal firms face *de facto* lower costs of production relatively to formal firms operating in the same industry (De Mel et al., 2013). Not surprisingly, the question of how to induce informal firms to formalize has motivated a large literature as well as an intense policy debate.

The existence of high registration costs has long been pointed out as a major constraint to firm entry and formalization (e.g. De Soto, 1989; Djankov et al., 2002). Indeed, policymakers around the world have devoted substantial efforts to reduce them.[2] The existing literature, however, shows little empirical support for the effectiveness of such reforms (La Porta and Shleifer, 2014). An alternative explanation is that the ongoing costs of formality, mainly tax liabilities, constitute the relevant constraint. Even though this conjecture has been suggested in previous work (Kaplan et al., 2011; De Mel et al., 2013), there is scant empirical evidence on the impacts of reducing the tax burden on entrepreneurs' decisions to formalize or to create a new formal business.

This paper estimates the effects of reducing the ongoing costs of formality on firm formalization. To get to this question, we estimate the impacts of a large-scale formalization program implemented in Brazil, the *Individual Micro-Entrepreneur Program* (henceforth IMP).[3] Introduced in 2009, the program targeted entrepreneurs with at most one employee, and was designed to reduce both dimensions of formality costs: (i) registration (entry) costs; and (ii) the costs of remaining formal, by reducing monthly taxes and red tape. Unintentionally, the program had two distinct phases. The first drastically reduced registration costs but failed to reduce tax rates for most eligible entrepreneurs.[4] The second phase *de facto* reduced taxes by halving monthly tax expenditures, when entry costs were already zero. This staggered implementation, coupled with plausibly exoge-

---

[1]In Brazil, for example, 89% of all entrepreneurs have at most one employee and nearly 3/4 of them are informal (data from the National Household Survey – PNAD 2012).

[2]The Doing Business project of the World Bank reports 368 reforms that took place in 149 economies between 2003 and 2012 (Bruhn and McKenzie, 2014).

[3]*Programa do Microempreendedor Individual - MEI*, in Portuguese.

[4]This occurred because IMP introduced a relatively high monthly lump-sum tax, so only high-income individuals perceived a lower tax rate in comparison to the pre-existing flat rate.

nous variation in access to the program, provides an unique empirical setting to assess the effects of reducing the tax burden faced by small formal firms.

Our empirical strategy relies on the exogeneity of the program's eligibility criterion, which was defined at the 7-digit industry level. We argue that the list of eligible industries followed the one used by the pre-existing tax framework for micro, small and medium enterprises, which was introduced more than a decade prior to the program. Moreover, there is no indication that the government hand-picked specific industries based on their past performance or observable characteristics. We perform several empirical tests that provide compelling evidence in support of our identification assumption.

We use two main data sources in our empirical analysis. The first is an administrative microdata set from the Ministry of Labor that contains the universe of formal firms in Brazil (*Registro Anual de Informações Sociais*, RAIS). This data set allows us to directly assess the impact of IMP on the number of small formal firms, but it does not contain information on informal firms and workers, nor on individual transitions between different statuses in the labor market. We thus turn to individual level data from the Brazilian Monthly Employment Survey (PME). We exploit PME's panel structure to examine whether the effects on the number of formal firms, if any, come from the formalization of existing informal entrepreneurs, the creation of new formal businesses or greater survival of existing formal firms. In both data sets, variation in eligibility and access to the program allows for a treatment/control strategy that contrasts the evolution of a given outcome for individuals in different industries and regions to identify the contribution of IMP to the observed changes.

The results from our preferred specification show that industries eligible to the tax reduction experienced an increase of 4.3% in the number of formal firms. The individual-level regressions show that this increase came from the formalization of existing informal businesses, and not from the creation of new formal firms nor greater survival of existing formal businesses. The estimates indicate that halving monthly taxes led to a 1.9 percentage points increase in the average entrepreneur's registration rate, from a baseline rate of 20%. Several falsification and robustness tests confirm this result. This average effect is entirely driven by high-income entrepreneurs, which is consistent with the regressive tax rate schedule introduced by IMP.[5] We show that these formalization effects are transitory, peaking at the sixth month and reverting back to zero thereafter. Given that the implied formalization elasticity is low and the effects are transitory, a simple back-of-the-envelope exercise indicates that the program led to losses in tax revenues.

To the best of our knowledge, this paper provides the first attempt to test whether

---

[5]This result is also consistent with the idea that a large fraction of low-productivity informal firms perceive very small or no benefits at all from formality (e.g. De Mel et al., 2013; Ulyssea, 2017).

the costs of remaining formal are indeed a relevant constraint for firms to become formal, net of potential confounding effects and irrespective of registration costs. The available evidence is very scarce and the existing studies face relevant limitations (e.g. Fajnzylber et al., 2011; Monteiro and Assunção, 2012). First, they rely on a cross-section of firms, so they cannot disentangle program effects from confounding factors such as industry shocks. Second, they cannot assess the full effect of a tax reduction net of the influence of entry costs.[6] Moreover, contrary to previous papers we are able to identify whether the observed formalization effects come from new formal businesses being created, informal firms moving into formality or greater survival of formal firms.

Due to the program's staggered implementation, we can also exploit its first phase to assess whether reducing registration costs is a sufficient condition to induce firms to formalize. Similarly to the previous literature, we find no effect of reducing entry costs on firm informality. The existing non-experimental studies estimate the effects of reducing registration costs in settings where the ongoing costs of formality remain unchanged (Bertrand and Kramarz, 2002; Kaplan et al., 2011; Bruhn, 2011). A second stream of papers relies on randomized experiments and focuses on the formalization of existing informal firms (e.g. De Mel et al., 2013; De Andrade et al., 2013; De Giorgi and Rahman, 2013). As in our context, the results in this literature indicate very limited or no effects on formalization. Differently from previous studies, however, we show that this is the case even in a context where the costs of remaining formal are also reduced (i.e. amongst high-income entrepreneurs). We are able to establish this fact both at the aggregate level and by examining individual level transitions into formal entrepreneurship.

The remainder of this paper is organized as follows. Section 2 describes the institutional setting. Section 3 describes the data, while Section 4 describes our empirical strategy. Section 5 presents the results and robustness checks. In this section we also present a cost-benefit analysis of the program and further discuss the implications of our results. Section 6 concludes.

2 Institutional Setting

The first major initiative to reduce the tax burden of micro, small and medium enterprises in Brazil dates back to 1996, when the federal government introduced the SIMPLES program. This program created a differentiated tax system for these firms, which reduced

---

[6]Monteiro and Assunção (2012) and Fajnzylber et al. (2011) rely on a cross-section of firms to estimate the effects of SIMPLES, a program that reduced taxes and red tape associated with tax payments in Brazil but left registration costs (monetary and non-monetary) unchanged. As Monteiro and Assunção (2012) recognize, identification concerns leave open the question of whether the results are driven by confounding effects, such as sectoral shocks and seasonal effects.

red tape, consolidated several taxes and social security contributions into a single payment, and reduced the overall tax burden. However, it did so without changing the procedures to register firms, which remained cumbersome and costly (Monteiro and Assunção, 2012).[7] Eligibility to the program was defined at the 7-digit industry level and was oriented towards manufacturing, transportation, and other services not subject to specific regulations.[8] Additionally, a cap on annual revenue limited participation of larger firms. Over the years the revenue cap was substantially expanded and the SIMPLES became the relevant institutional framework for most firms in Brazil.

In July 2009, the federal government introduced a new tax and business registration framework specifically designed for micro entrepreneurs with up to one employee. The *Individual Micro-Entrepreneur Program* (henceforth IMP) was launched as a nation-wide program and was widely advertised in the media.[9] The IMP had three main goals: to foster entrepreneurship and the creation of new formal businesses; to increase tax registration and compliance of existing informal micro entrepreneurs; and to increase contributions to the social security system. In fact, the program locked these two margins of formalization together: when registering their business through IMP, entrepreneurs were also automatically enrolled in the social security system.[10] IMP substantially reduced formalization costs in two dimensions: (i) registration costs, both monetary and non-monetary (e.g. the number of procedures required to register a business); and (ii) taxes and red tape associated with tax payments. Hence, the program reduced both the costs of entering the formal sector and the costs of remaining formal.

Even though these two sets of incentives (lower entry costs and taxes) were intended to be introduced simultaneously at the time the program was created, this was not the case. We show that in fact the program had two distinct phases, which we name IMP1 and IMP2. In the first (IMP1), the program eliminated monetary entry costs and introduced a web platform for online business registration, which consolidates all procedures required by agencies at national and sub-national levels into a few online steps. Due to technical constraints, this platform was not available to all states simultaneously. There was a 8-month window before all states had access to it, which occurred only in February 2010 (Figure 1).[11] Since firm registration within the program must be made through the web

---

[7]Starting a business in Brazil has required about 16 formal procedures and 120 days. It has cost, on average, nearly US$770 (IFC, 2009).

[8]For instance, the SIMPLES eligibility rule excluded the financial sector, real estate companies, developers, cleaning service firms, and activities with regulated occupations, such as physicians, architects, and musicians.

[9]Google trends indicate that searches for the topic peaked right after the launch of the program in July 2009 and, despite oscillations, remained around the same level over the entire period of our analysis.

[10]In Brazil, a firm is formally registered if it holds a CNPJ, which is an acronym for *Cadastro Nacional da Pessoa Jurídica*. This is an identification number for firms that is issued by the federal government.

[11]Access was first made available in July 1st 2009 for entrepreneurs located in the Federal District; it

platform – and firms cannot register in one state and operate in another – this generated variation in access to the program across states and time during IMP1.

Figure 1: IMP Implementation Timeline
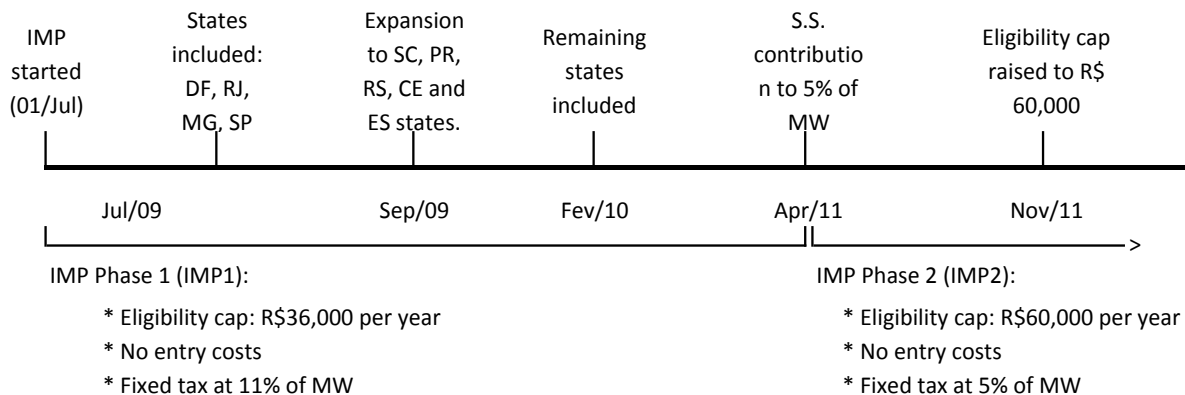
During IMP1 there was also an attempt to reduce the ongoing costs of formality. Federal taxes were eliminated, while state and municipality taxes were fixed at negligible values. The major tax component became the monthly social security contribution, which was fixed at 11% of the minimum wage and corresponded to more than 90% of total monthly tax expenditures. At the time of implementation, IMP1's monthly tax burden was nearly R$50 (around US$25).[12] By fixing monthly tax expenditures, the program automatically introduced a regressive tax rate schedule for eligible entrepreneurs. Moreover, monthly taxes were fixed at high values relatively to eligible entrepreneurs' income. As a consequence, only entrepreneurs at the highest percentiles of the income distribution perceived an actual reduction in their tax rate in comparison to the pre-existing rate under the SIMPLES framework (a flat rate of around 4%).[13] Figure 2 illustrates this point by plotting the tax rate schedules under IMP1, IMP2 and SIMPLES,

---

was then expanded to the states of Minas Gerais, Rio de Janeiro and São Paulo in July 24th; to Santa Catarina, Paraná, Rio Grande do Sul, Ceará and Espírito Santo in September 2009. The remainder states had access to the platform in February 2010.

[12]The conversion to US$ considers the exchange rate as of July 1st 2009. The minimum wage was R$465 at the time of implementation and the corresponding social security contribution was R$51 (approximately US$25.5). State and municipality taxes were fixed at R$1 and R$5 per month, respectively (roughly 50 cents and 2.5 US dollars, respectively). IMP1 thus implied an overall tax burden of either R$52 or R$56, depending on the industry (US$26 or US$28, respectively).

[13]This corresponds to the tax rate alone. Other variable costs, such as social security contribution and accountant fees are excluded from our calculations, since these depend on entrepreneurs' choice.

together with the quartiles of entrepreneurs earnings distribution. The figure shows that only entrepreneurs above the 75th percentile perceived a lower tax rate under IMP1.

In April 2011 the program entered its second phase (IMP2), when social security contribution was reduced from 11% to 5% of the minimum wage. As shown in Figure 2, it was only then that most eligible entrepreneurs (those above the 25th percentile of the earnings distribution) perceived an actual reduction in tax rates in comparison to that under the SIMPLES framework. In sum, the first phase reduced entry costs but failed to effectively reduce tax rates; the second phase maintained the benefits from IMP1 and *de facto* reduced tax rates, thus reducing the cost of remaining formal for most eligible entrepreneurs.

Figure 2: Tax rates under SIMPLES, IMP1 and IMP2
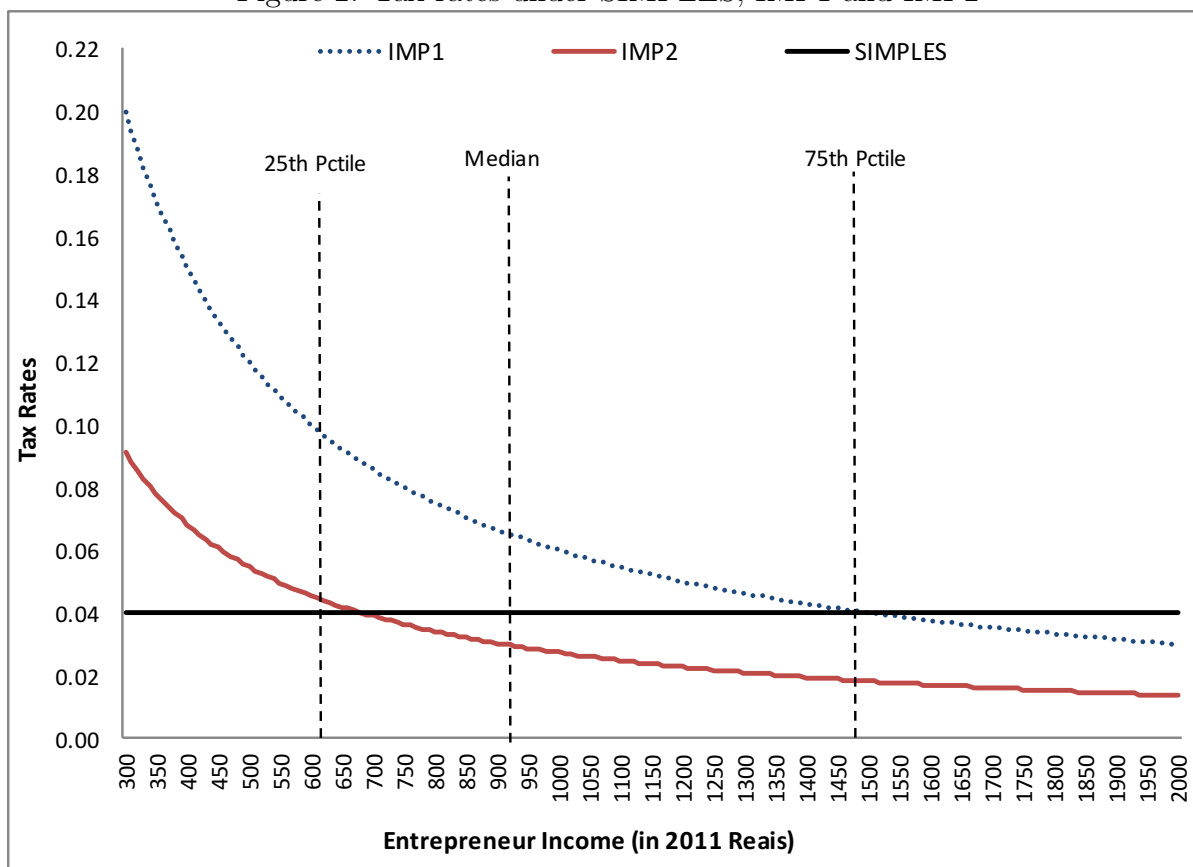
Note: Tax rates under the SIMPLES, IMP1 and IMP2 using income in 2011 Reais. The tax rates (TR) are computed using program rules, as follows: $\text{TR}_{IMP1} = \frac{11\% MW}{Y_i}$ and $\text{TR}_{IMP2} = \frac{5\% MW}{Y_i}$, where MW denotes the minimum wage and Y_i entrepreneur's income. The 25th, median and 75th percentiles refer to the income distribution among entrepreneurs with at most one employee in April 2011, the month of IMP2 implementation.

Eligibility to the program is based on three criteria. The first is firm size measured by the number of employees, as the program is targeted at entrepreneurs with at most one employee. Second, eligibility is defined by annual revenue, with a cap initially set at

6

R$36,000 per year (around US$18,000), and further expanded to R$60,000 in November 2011. Third, eligibility was defined by economic activity at the 7-digit industry level.

IMP's list of eligible industries comprised nearly five hundred industries, which in 2009 (just before the program was fully implemented) encompassed 46.5% of all entrepreneurs in Brazil, and 52% of those with at most one employee.[14] The definition of eligible and non-eligible industries followed the previous configuration already in place under SIM-PLES, which had been introduced more than a decade before IMP implementation. The list of eligible industries was concentrated on manufacturing, transportation, and services not subject to specific regulations. There were minor changes in this list over time, but with no significant variation in the total number of eligible industries (or entrepreneurs) actually observed in the data. We provide descriptive statistics and further details on industry eligibility in the following sections.

3   Data

We use three data sources in our analysis. The first one is the *Registro Anual de Informações Sociais* (RAIS), an administrative data set from the Ministry of Labor. It is a matched employer-employee data set that contains the universe of formal workers and firms in Brazil. It provides detailed information on formal firms at an annual frequency, and also makes available data on labor contracts (at the worker level) at a higher frequency. We use the micro data at the worker level, from January 2006 to December 2012, to create a quarterly panel at the industry-by-region level containing our first outcome variable, the logarithm of the number of formal firms with up to one employee. The RAIS data contain 670 industries, 8 of which experienced changes in eligibility status over time. We exclude these industries in order to avoid potentially endogenous selection into/out of treatment. Our final data set has a total of 500,472 industry-state-trimester observations (662 industries x 28 quarters x 27 states). Analogously, we construct a panel of industry-region-trimester cells at the metropolitan region level that has 111,216 observations (662 industries x 28 quarters x 6 regions), which enables us to mimic the samples based on the PME data, our second data set. In these balanced panels of industry-state-trimester cells, the computation of our dependent variable considers the logarithm of the total number of firms plus 1, so that zeros are included in the final sample as $\log(1)$.[15]

---

[14]Data from the National Household Survey (PNAD), own tabulation.

[15]The results hold whether we include the cells with zero firms or not, and whether we exclude industries that experienced changes in eligibility status or not. We discuss these robustness checks in Appendix B. We also note that, in principle, it would be possible to construct a monthly panel. However, this would likely increase measurement error in the outcome variable since firms might inaccurately report their workers' contract dates. The aggregation at the quarter level alleviates this concern.

RAIS constitutes a census of all formal firms in Brazil, so it allows us to directly assess the impact of IMP on formal firm creation. However, we cannot observe informal firms or workers, transitions into formality, and formalization rates. We thus use a second data set, the Brazilian Monthly Employment Survey (PME), a rotating panel at the individual level that covers the six largest metropolitan areas in Brazil. It tracks each household for 4 consecutive months, they rest for 8 months, and are then re-interviewed for 4 additional months, allowing to follow individuals for up to 16 months. We use data from January 2006 to August 2012. The PME data allow us to conduct the analysis both at the individual level, and aggregated at the industry-by-region level. At the individual level, we observe all transitions between occupational statuses, including transitions into formal and informal entrepreneurship. At the aggregate level, we are able to compute the share of formal and informal entrepreneurs, unemployment rates, and other labor market outcomes at the industry-by-region level. We are also able to compute a series of control variables, such as the mean monthly wage, the mean number of hours of work, the log of the total employment, the share of males, the average number of years of schooling and the average age of individuals in each industry-by-region cell. The sample is restricted to individuals aged between 21 and 65 in their first interview. Among the employed, we considered those who work more than 20 hours per week, so not to include occasional work. We also exclude individuals in agriculture, public sector and domestic activities.

Despite its many advantages, PME only contains information on individuals' industry at the 2-digit level, while program eligibility is defined at a more disaggregated level. The main consequence is that we cannot directly define an eligibility dummy in the PME data. We overcome this by using auxiliary information available in our third data set, the National Household Survey (PNAD). The PNAD data is an annual cross section representative of the entire country, containing information on industry classification at the same level of disaggregation used to define program eligibility.[16] We thus use PNAD data to construct a *potential of treatment intensity* variable, which is the *pre-program* share of entrepreneurs in eligible industries within each 2-digit industry. Formally:

$$Intense_s = \frac{\sum_{k \in s} \left(\mathbf{I} \left[\text{Sub-Industry } k \text{ is eligible} \right] \times N_{ks} \right)}{\sum_{k \in s} N_{ks}} \tag{1}$$

where N_{ks} denotes the number of entrepreneurs that belong to sub-industry k within

---

[16]The National Bureau of Statistics (IBGE) uses two different industry classification systems. One for the National Accounts, which is the same used to define program eligibility (at the 7-digit level), and another for Household Surveys, which includes the PNAD (at the 5-digit level). These correspond to the same level of disaggregation, and there is a direct correspondence between them provided by IBGE. We use this correspondence to link industry classification in the PNAD to the one used to define industry eligibility in the IMP.

the 2-digit industry s in 2009.[17] The variable $Intense_s$, calculated with PNAD data, is then merged at the industry s level with our monthly panel of individuals from PME. Analogously to the RAIS data, we exclude industries that experienced changes in their eligibility status, which correspond to 1.5% of the sample.

Table 1 presents the descriptive statistics from both the RAIS and PME samples at the baseline.[18] We define as formal entrepreneurs those who contribute to the social security system, which is a commonly used measure of formalization in the literature. As discussed in the previous section, the program locked together social security contribution and business registration decisions, which makes contribution to the social security system a particularly good formality measure in the present context. We observe on Table 1 that contribution to social security is much less frequent amongst small entrepreneurs in comparison to the full sample average (20% vs 64%, respectively).

---

[17]IMP was first available in Rio de Janeiro, Minas Gerais and São Paulo by August 2009, just a few weeks before the 2009 PNAD's month of reference. Since the share of entrepreneurs in eligible industries within each 2-digit industry remained quite stable between 2008 and 2009, we opt to base our calculations in 2009, which is closer to IMP1 introduction.

[18]Descriptive statistics computed for all observations between January 2006 and December 2008 – prior to 2009, the year when IMP is introduced.

Table 1: Descriptive Statistics at Baseline

	Obs	Mean	SD	Min	Max
PME Data[†]					
Schooling (yrs.)	999,697	9.15	3.73	0	15
Age	999,697	37.2	11.1	17	67
Male	999,697	0.61	0.49	0	1
Unemployed	999,697	0.09	0.29	0	1
Formal Employee	999,697	0.48	0.50	0	1
Informal Employee	999,697	0.13	0.33	0	1
Entrepreneur	999,697	0.28	0.45	0	1
Small Entrepreneur	999,697	0.24	0.43	0	1
Small Formal Entrepreneur	999,697	0.05	0.21	0	1
$\log(Wage)$	856,439	6.95	0.78	0.3	9.5
$Intense_s$	954,162	0.47	0.41	0	1
Contributes to S.S.	906,227	0.64	0.48	0	1
Contributes to S.S. if Small Entrepreneur	243,222	0.20	0.40	0	1
RAIS Data[††]					
# Formal Firms ($\leq$ 1 employee) - All Firms	296,810	13.8	49.1	0	1,261
# Formal Firms ($\leq$ 1 employee) - PME Sample	62,980	3.8	6.6	0	65

Notes: All tabulations refer to to pre-program years (2006–2008). [†]Authors' own tabulations from the PME. Entrepreneurs include all self-employed or employers; small entrepreneurs are those with up to one employee; and small formal entrepreneurs are are those who contribute to social security.[††]Authors' own tabulations from the RAIS micro data aggregated at the region-by-industry level, quarterly frequency.

4 Empirical Strategy

We take advantage of the staggered implementation of IMP and its eligibility criteria to design a difference-in-differences strategy. We estimate two sets of regressions. The first estimates aggregate policy effects at the industry-by-region level, using both the RAIS or the PME data sets. We use RAIS data to identify whether IMP had any impacts on the number of formal firms with up to one employee. We extend this analysis with PME data to assess the effects on the: (i) share of small formal entrepreneurs; (ii) share of formal employees; (iii) share of informal employees; and (iv) unemployment rate.

Once we examine the impacts of IMP on aggregate outcomes, we ask whether the effects come from the formalization of existing informal firms, the creation of new formal

businesses, or simply greater survival of previously existing formal firms. In order to answer this question, we estimate our second set of regressions, at the individual level, to examine individual transitions from different occupational statuses into formal entrepreneurship. In what follows, we describe our econometric specifications for each set of regressions, separately by data sets.

4.1 Aggregate Effects

In order to identify aggregate effects of IMP on the total number of formal firms, we use RAIS data to estimate the following regression:

$$
\begin{aligned}
Y_{srt} \;=\;\; & \beta_1(Post1_{rt} \times Eligible_s) + \beta_2(Post2_t \times Eligible_s) \qquad\qquad\qquad (2)\\
& + \gamma_1 Post1_{rt} + (Post1_t \times Eligible_s)'\gamma_2 + \gamma_3 Eligible_s \times \xi_r + \lambda_t + \nu_s + \xi_r + \epsilon_{srt}
\end{aligned}
$$

where Y_{srt} denotes the log of the number of formal firms in industry s, region r and trimester t. The term $Post1_{rt}$ is a dummy that indicates whether access to IMP1 was available in region r at time t, while $Post2_t$ is the analogous variable for IMP2, which does not vary across regions. The variable $Eligible_s$ denotes the dummy for eligible industries. $Post1_t$ denotes a set of two indicator variables for the different start dates of IMP1 across regions, i.e., $1[t \geq 3^{rd}\ trim\ 2009]$ and $1[t \geq 1^{st}\ trim\ 2010]$, while the variable $Post1_{rt}$ denotes the interaction between the regional dummies and the time dummies in $Post1_t$. The terms λ_t, ν_s, and ξ_r denote, respectively, time, industry and region fixed-effects.

Our empirical strategy thus contrasts the evolution of the log of the number of formal firms across regions and between eligible and non-eligible industries. The triple-interaction between industry, time and region, $Post1_{rt} \times Eligible_s$, is our variable of interest when estimating IMP1 effects. The double-interaction between industry and time, $Post2_t \times Eligible_s$, is our variable of interest when estimating the effects of IMP2. The focus of our analysis lies on the β_2, which captures the effects of reducing the tax burden for small formal firms, while the β_1 coefficient captures potential effects from the combination of incentives provided in the first phase of the program.

We use PME data to complement the previous analysis by assessing IMP effects on the share of small formal entrepreneurs, small informal entrepreneurs, formal employees, informal employees, and the unemployment rate. In order to do so, we estimate the following regression:

$$
\begin{aligned}
Y_{srt} \;=\; & \beta_1 (Post1_{rt} \times Intense_s) + \beta_2 (Post2_t \times Intense_s) && (3)\\
& + \gamma_1 Post1_{rt} + (Post1_t \times Intense_s)' \gamma_2 + \gamma_3 Intense_s \times \xi_r + \lambda_t + \nu_s + \xi_r \\
& + Trends'_{srt} \theta + \epsilon_{srt}
\end{aligned}
$$

where Y_{srt} denotes the outcome variable for industry s, region r, at month t. Again, $Post1_{rt}$ is a dummy that indicates whether access to IMP1 was available in region r at time t, and $Post2_t$ is the analogous variable for IMP2. The variable $Intense_s$ is defined by equation 1, and denotes the *potential of treatment intensity* of industry s. The term $Post1_t$ now denotes a set of three indicator variables for the different start dates of IMP1, which can be defined more precisely in the PME data due to its monthly frequency: $1[t \geq Aug09]$, $1[t \geq Sept09]$ and $1[t \geq Feb10]$; while the terms λ_t, ν_s, and ξ_r denote, respectively, time, industry and region fixed-effects. As the PME is a richer data set, we include the term $Trends_{srt}$ in equation 3, which contains time-varying economic variables at the industry-by-metropolitan region level, as well as their square, cubic and quartic. More specifically, the economic variables are the mean monthly wage, the mean number of hours of work, and the log of the total employment for each region, industry and month. These variables help absorbing the influence of the economic cycle and heterogeneous shocks across industries and regions. The term $Trends_{srt}$ also includes the share of males as well as the average number of years of schooling, and the average age of individuals at the month-industry-region level. Our PME data contain a total of 40 different 2-digit industries, so there is a total of 18,960 industry-region-month cells in our dataset (40 x 79 months x 6 metropolitan regions). However, our analysis considers a sample of 16,676 observations, which excludes industry-region-month cells that have zero observations and those industries whose eligibility status changed over time.[19]

---

[19]Exactly 2 out of 40 industries experienced changes in eligibility status over time in the PME data. We do not input zero to the dependent variables when they are missing because these are shares computed from a survey, as opposed to being a headcount of firms from a census (which is the case of RAIS). In the latter, it makes sense to assume that missing observations mean zero firms in that particular cell, whereas this may not be the case in survey data. Also, zero firms would necessarily imply missing shares. In Table B.2 of Appendix Section B, we provide evidence that our benchmark estimates are robust to this sample definition.

4.2 Individual Transitions

Our benchmark specification at the individual level is analogous to equation 3, and is given by the following expression:

$$
\begin{aligned}
Y_{isrt} = {} & \beta_1(Post1_{rt} \times Intense_s) + \beta_2(Post2_t \times Intense_s) \\
& + \gamma_1 Post1_{rt} + (Post1_t \times Intense_s)'\gamma_2 + \lambda_t + \phi_i \\
& + Trends'_{srt}\theta + X'_{isrt}\psi + \epsilon_{isrt}
\end{aligned}
\tag{4}
$$

where Y_{isrt} is a dummy that equals one if individual i, in industry s, metropolitan region r, at month t is a small formal entrepreneur. The remaining terms indexed by s, r, and t are analogous to those defined in equation 3. The term ϕ_i refers to individual fixed-effects, which in our setting absorb both region and industry fixed effects. The term X_{isrt} includes time-varying observables at the individual-level (age squared and years of schooling). Coefficient β_2 is again the main parameter of interest, and captures the effects of reducing taxes on transitions into formal entrepreneurship. Our individual regressions are based on a shorter time span, in which we have less power to absorb potential trends. In addition to the variables already discussed in equation 3, we also include controls for convergence across industries and regions in extended specifications. More specifically, we include interaction terms between the dependent variable computed at the baseline (the average of the dependent variable at the industry-region cell over the period of January 2006 to December 2006) and linear, square, cubic and quartic time trends. These four additional time-varying controls should absorb the potential confounding convergence effects across industries, within states, over time.

We analyze transitions from distinct occupational statuses into formal entrepreneurship by running different regressions restricting the sample of individuals according to their occupational status in the first interview (e.g. individuals who are informal entrepreneurs). In our benchmark specification we only include those who have at least one observation before and one after the start dates of the first and second phases of the program (July 2009 and April 2011, respectively).[20] This resembles an event study type of specification, as we observe all individuals at least once before and after either IMP1 or IMP2 start dates. Importantly, since there is a long time interval between start dates, the samples used for estimating the effects of both phases do not overlap, i.e., individuals included in the estimation of IMP1 effects are not in the sample used to estimate IMP2 effects, and vice-versa.[21]

---

[20]In order to cluster standard errors at the industry level, we restrict the sample to those entrepreneurs who do not change industries between interviews. This restriction has no effect on the estimated results.

[21]We perform several robustness tests, among which we include regressions that do not impose any

Finally, we note that since we do not have a dichotomous treatment variable, the empirical strategy underlying regressions 3 and 4 (using PME data) is based on the fact that individuals located in industries with higher potential of treatment are more likely to benefit from the program than those located in industries with a lower potential.[22] Our approach will be closer to a standard diff-in-diff the more polarized the distribution of the variable $Intense_s$ is. Indeed, Table A.1 of Appendix A shows that $Intense_s$ is a discrete variable (it assumes 15 different values) and is very concentrated around zero and one.

4.3   Identification

In the experimental ideal, in which regions and industries are randomly selected into treatment, coefficients β_1 and β_2 in equations 2 and 3 identify the variation in a given outcome caused by the introduction of IMP. The same reasoning at the individual level is valid in equation 4. In our non-experimental context, however, other unobservable determinants may be correlated with the variables of interest. The validity of our strategy relies on two main conditions. First, industry eligibility should not be correlated with pre-trends in formalization. If that was the case, our estimates would be spuriously driven by trends that are specific to eligible *versus* non-eligible industries. Second, economic shocks should impact different industries uniformly throughout the period being considered in the analysis.

Regarding the first condition, our identification assumption is that industry eligibility is exogenously determined. As discussed in the previous section, eligibility under IMP was based on the SIMPLES's list of eligible industries, and therefore defined a decade prior to IMP. Moreover, there is no indication that the government hand-picked specific industries based on their past performance or observable characteristics. In order to assess the validity of our identification assumption, we start by plotting in Figure 3 the evolution of the monthly share of formal entrepreneurs with up to one employee in industries with potential of treatment below the median ($Intense_s < 0.5$) and above the median ($Intense_s > 0.5$).[23] As the figure shows, there is no clear difference in pre-trends before IMP.

Table A.2 of Appendix A provides a further test for the influence of pre-trends in determining industry eligibility. We examine the association between $Intense_s$ and previous variation in formality rates and average income at the industry level. We consider

---

time restriction on our sample. The results remains unaltered (see Section 5).

[22]We also assume that elasticities to formalization should be similar within groups.

[23]It is worth clarifying that, coincidentally, the value of $Intense_s = 0.5$ is both the midpoint of the domain and the approximate median of the distribution of this discrete variable.

Figure 3: Monthly Share of Small Formal Entrepreneurs in Industries with Potential of Treatment Below and Above the Median
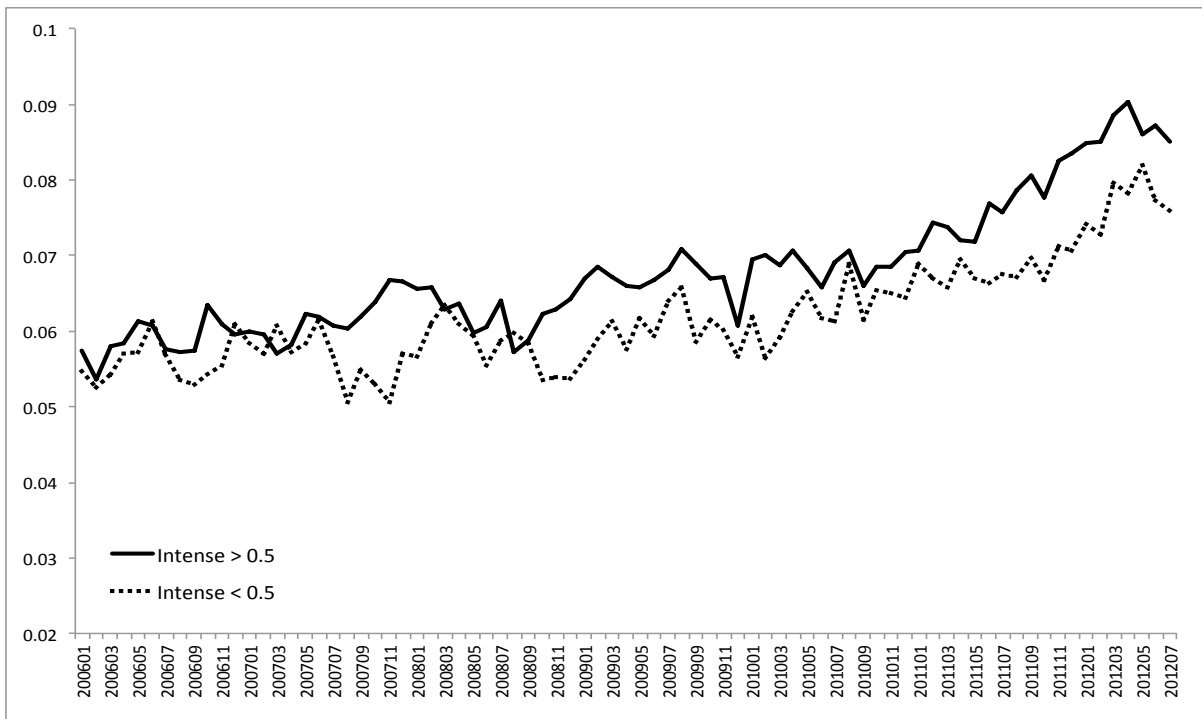
Note: Authors' own calculation from the Monthly Employment Survey (PME). Monthly formalization rates for each month correspond to the share of formal entrepreneurs with up to one employee. These rates are computed separately for entrepreneurs in industries with low potential of treatment ($Intense_s < Median$) and high potential of treatment ($Intense_s > Median$).

two periods. The mid-run, which considers the variation in formality rates and average income between June 2006 and June 2009, just before the introduction of IMP; and the short run, from June 2008 to June 2009. All regressions control for metropolitan region fixed-effects. We find no association between $Intense_s$ and pre-trends in formalization rates nor in aggregate income. This result holds whenever we regress them separately (columns 1–4) or jointly (columns 5–6). We further assess whether industry eligibility is uncorrelated with pre-trends in formalization in a series of robustness checks reported in Section 5.5. The results provide compelling evidence in support of this assumption.

As for the second condition, there is no substantial economic fluctuation around IMP2's implementation, which is the focus of our analysis. This reduces our concerns about the potential influence of heterogeneous economic shocks across industries. Even so, in order to absorb potential confounding effects, and ensure that our results are robust to shocks, we add time-varying economic variables at the region-by-industry level, and their square, cubic and quartic. As we show in Section 5, the estimates remain remarkably stable. As for IMP1, its introduction coincides with the impact of the great recession, which were felt more strongly in Brazil in 2009. Because of the staggered implementation of IMP1, however, we are able to exploit a triple-interaction between time, region and industry, which minimizes this concern. The identification assumption here is that there is no reason other than IMP1 for the difference of formalization between industries to increase more in regions that were treated earlier when compared to those treated later.

4.4 A Simple Framework to Interpret the Estimated Effects

In this section we introduce a very simple model of entrepreneurs' decision to register their business. The goal here is to use this conceptual framework to help us interpreting the effects we identify with the empirical strategy outlined above. Even though the model refers to informal entrepreneurs' decisions to become formal, the main implications are also valid to the analysis of heterogenous employees deciding whether to start a formal business.

Let y denote entrepreneurs productivity and, to make notation simple, assume that entrepreneurs are infinitely lived and have discount rate β.[24] Since our empirical analysis focuses on entrepreneurs with up to one employee, we simply assume that production equals entrepreneur's productivity. We normalize the price of the final good to one, so that revenues are given by $r(y) = y$. Entrepreneurs can choose between being formal (IMP1 and IMP2) or informal (INF). Additionally, it is useful to characterize the previous

---

[24]It is straightforward to modify this problem to consider entrepreneurs who work for T periods and then retire.

relevant framework for formal entrepreneurs, which was given by the SIMPLES (SIM). The value function of each state can be described as follows:

$$V^{INF}(y) = \sum_{t=0}^{\infty} \beta^t \Pi^{INF} = \frac{\Pi^{INF}}{1-\beta} \equiv \frac{y - c(y)}{1 - \beta}$$

$$V^j(y) = \sum_{t=0}^{\infty} \beta^t \Pi^j = \frac{\Pi^j}{1-\beta} \equiv \frac{y - T_j}{1 - \beta}, \quad j = IMP1, IMP2$$

$$V^{SIM}(y) = \sum_{t=0}^{\infty} \beta^t \Pi^{SIM} - C_F = \frac{\Pi^{SIM}}{1-\beta} - C_F \equiv \frac{(1-\tau)\, y}{1 - \beta} - C_F$$

where Π^k denotes monthly profits in state $k = INF, IMP1, IMP2, SIM$; $c(\cdot)$ is a function that summarizes the costs of being informal, where $c', c'' > 0$;[25] T_{IMP1} and T_{IMP2} denote the fixed monthly tax expenditures in IMP1 and IMP2, respectively; τ and C_F denote the monthly tax rate and the registration costs under SIMPLES, respectively.

With this simple structure, it is possible to characterize entrepreneurs' decisions in terms of productivity thresholds, as shown in Figure 4.[26] Previously to the introduction of IMP1 (under the SIMPLES), an entrepreneur would decide to be formal if $V^{SIM}(y) \geq V^{INF}(y) \iff c(y) - \tau y \geq (1-\beta)C_F$, or equivalently if $y \geq y_4$ in Figure 4. After IMP1 is implemented, entrepreneurs formalize if and only if $V^{IMP1}(y) \geq V^{INF}(y) \iff c(y) \geq T_{IMP1}$, or equivalently iff $y \geq y_3$ in Figure 4. The magnitude of formalization effects after the introduction of IMP1 thus depends on the mass of entrepreneurs with productivity such that $y \in [y_3, y_4]$. For higher values of the fixed monthly tax expenditures (T_{IMP1}), we will have that $y_3 \to y_4$. Even though IMP1 eliminated entry costs ($C_F = 0$) it raised tax expenditures for a substantial share of entrepreneurs. Due to the regressive structure of IMP's taxation, high-income entrepreneurs are the only ones who perceived a decline in formalization costs under IMP1, which correspond to those whose productivity is such that $\tau y + (1-\beta)C_F \geq T_{IMP1}$, or equivalently $y \geq y_2$ in Figure 4. Finally, under IMP2 the costs of remaining formal, T_{IMP2}, were further reduced, which led to a *de facto* reduction in formalization costs for most eligible entrepreneurs. Under IMP2, entrepreneurs formalize if and only if $V^{IMP2}(y) \geq V^{INF}(y) \iff c(y) \geq T_{IMP2}$, or equivalently iff $y \geq y_1$ in Figure 4.

In sum, this simple analysis shows that estimating IMP1's average effect does not identify the effect of reducing either registration costs nor taxes. This is so because the

---

[25]This cost function can be motivated as the expected cost of being caught by the government, or any other opportunity cost of being informal (e.g. not having property rights, access to credit, among others). This is a common formulation in the literature, see for example de Paula and Scheinkman (2010) and Ulyssea (2017), among others.

[26]For this graph, we parameterize $c(y) = ay^2$, where a is a parameter. We set $\tau = 0.04$ and $T_{IMP2} = 0.5 \times T_{IMP1}$.

Figure 4: Graphic Representation of Entrepreneurs' Decision to Formalize
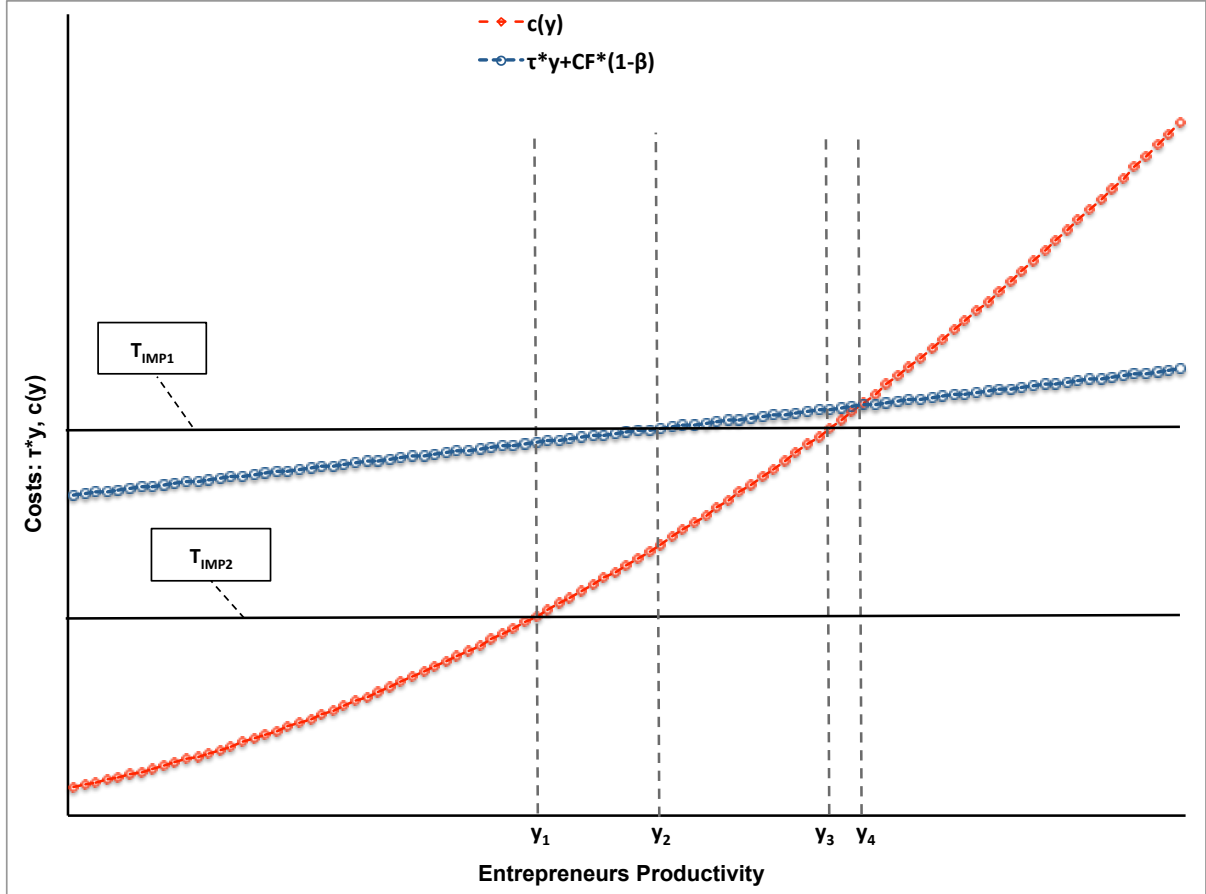
Note: For this graph, we parameterize $c(y) = ay^2$. The thresholds are determined as follows: (i) $V^{IMP2}(y_1) = V^{INF}(y_1) \iff c(y_1) = T_{IMP2}$; (ii) $\tau y_2 + (1 - \beta)C_F = T_{IMP1}$; (iii) $V^{IMP1}(y_3) = V^{INF}(y_3) \iff c(y_3) = T_{IMP1}$; (iv) $V^{SIM}(y_4) = V^{INF}(y_4) \iff c(y_4) - \tau y_4 = (1 - \beta)C_F$.

program eliminated entry costs but simultaneously increased tax expenditures for a substantial share of eligible entrepreneurs. Hence, overall formalization costs increased for some entrepreneurs (for those with $y < y_2$ in Figure 4) but decreased for others. The average IMP1 effect will be influenced by both groups, which implies that the net effect captured by the estimator will be ambiguous. Nevertheless, estimating IMP1's effects for different income levels can identify the impact of eliminating registration costs, as tax rates remained nearly constant or were slightly reduced for higher income entrepreneurs. Thus, finding a positive effect on formalization for higher income informal entrepreneurs can be interpreted as evidence that entry costs are indeed a binding restriction to formalization. Similarly, finding no effect provides evidence that reducing entry costs is not a sufficient condition to induce informal firms to formalize.

Under IMP2 most entrepreneurs perceived an actual reduction in tax rates. Therefore, estimating IMP2's average effect should identify the impact of reducing the costs of remaining formal $(T_{IMP1} \rightarrow T_{IMP2})$, *given that entry costs were already eliminated* $(C_f = 0)$. Moreover, if the ongoing costs of being formal are the main factor driving entrepreneurs' decisions, we should observe formalization effects that are increasing in entrepreneurs' income, as the relative cost of formalization is declining in income under IMP.

5 Results

5.1 Aggregate Effects

We start by using RAIS data to estimate regression 2. The dependent variable is the logarithm of the number of formal firms with up to one employee at the industry-by-state level. In the first column we use data for the entire country, while in the second column we restrict the sample to the metropolitan regions available in the PME. The results in Table 2 show that IMP2 had a positive, sizable and statistically significant effect on the number of formal businesses. The point estimate in column 1 indicates that eligible industries had an increase of 4.3% in the number of small formal firms as a result of halving tax expenditures in the second phase of IMP. This result remains robust when we restrict the sample to the same metropolitan regions available in the PME data. As for IMP1, we find no statistically significant effect on the creation of new small formal businesses, regardless of the sample used.[27]

---

[27]Appendix Section B and Appendix Table B.1 provide further discussion on inference and robustness tests for different specifications of clustering and samples. This result is also robust once we further restrict the sample to the 2009-2011, which is the time span used in our benchmark analysis of individual transitions (see Section 5.2 below); results available upon request.

Table 2: Effects on the Number of Formal Firms: Administrative Data (RAIS)

Dependent Variable: Log (# Formal Firms)		
	All Regions	PME Regions
	(1)	(2)
IMP1	-0.011	0.010
	(0.007)	(0.016)
IMP2	0.043	0.032
	(0.006)***	(0.009)***
Observations	500,472	111,216
R-squared	0.845	0.821
Sample	≤ 1 Employee	≤ 1 Employees

Notes: Robust standard errors clustered at the industry level. Significance: ***1 percent, **5 percent, and *10 percent levels. Specifications in both columns 1 and 2 correspond to regression 2, detailed in Section 4.1. Column 1 uses the full sample including all regions in the country, column 2 restricts the sample to the six main metropolitan regions in Brazil, which constitute the sample used by the Monthly Employment Survey (PME): Porto Alegre, Rio de Janeiro, Recife, Salvador, Belo Horizonte and São Paulo. IMP1 and IMP2 correspond to the estimated coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ in regression 2, respectively.

The RAIS data only includes formal firms and employees. We thus complement the analysis by estimating regression 3 based on PME data aggregated monthly at the industry-by-metropolitan region level. Table 3 shows that IMP2 did not have any significant effects on the shares of formal/informal workers, on the unemployment rate, nor on the overall share of entrepreneurs. However, it increased the share of small formal firms. The point estimate in column 5 – evaluated at the average value of the potential of treatment variable ($Intense_s$) – indicates that the IMP2 increased the share of formal entrepreneurs with up to one employee by 0.58 percentage points ($= 0.012 \times 0.49 \times 100$), which corresponds to an increase of around 11% of the baseline rate of 5%. Again, Table 3 shows that IMP1 did not have any significant effects, irrespectively of the outcome being considered.

Taken together, the results of Tables 2 and 3 show that IMP2 did increase the number of formal firms, and this increase seems to have come from the formalization of existing small informal businesses rather than from the creation of new formal firms by individuals who were previously unemployed or employees. In the following sections we directly investigate this conjecture by using PME data at the individual level.

Table 3: Effects at the Industry-Region Level: Monthly Employment Survey (PME)

Dep. Var.: Share of	Formal Employees	Informal Employees	Unemployed	Entrepren.	Small Formal Entrepren.
	(1)	(2)	(3)	(4)	(5)
IMP1	-0.026	0.011	0.007	0.009	-0.010
	(0.030)	(0.015)	(0.013)	(0.022)	(0.010)
IMP2	-0.011	0.008	-0.002	0.005	0.012
	(0.014)	(0.009)	(0.004)	(0.010)	(0.004)***
Observations	16,676	16,676	16,676	16,676	16,676
R-squared	0.671	0.242	0.138	0.707	0.375

Notes: Robust standard errors clustered at the industry level. Significance: ***1 percent, **5 percent, and *10 percent levels. Specifications in all columns correspond to regression 3, as detailed in Section 4.1. IMP1 and IMP2 correspond to the estimated coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ in regression 3, respectively.

5.2 Formalization of Informal Firms

In this section we investigate whether the results in Section 5.1 come from the formalization of existing small informal entrepreneurs. For that, we exploit the panel structure of the PME data to estimate equation 4. As discussed in Section 4.2, our sample includes all individuals who have at least one observation before and one after the start dates of IMP1 or IMP2. Given that we are interested in the effects of formalization of the *stock of informal entrepreneurs*, we restrict our sample to those who report being informal entrepreneurs with up to one employee in their first interview. The dependent variable is a dummy for being a formal entrepreneur with up to one employee.

Table 4 presents the results. The first column includes time, industry and individual fixed-effects, while the second column adds controls for fluctuation in economic variables. In the third column we control for convergence in formalization rates across industries and regions. More specifically, we add a fourth degree polynomial in time interacted with the share of small formal entrepreneurs at the region-by-industry level in 2006, in order to control for potential convergence across regions and industries in the dependent variable. Standard errors are clustered at the industry level in all regressions.

The results show positive, sizable and significant effects for IMP2 across all specifications. The point estimate remains remarkably stable whenever we control for economic fluctuations in column 2. This pattern reduces our concerns about the potential confounding influence of heterogeneous economic shocks across industries. In column 3, as

expected, the point estimate falls once we control for convergence in formalization rates. In this specification, IMP's effects are partially absorbed by the influence of rather flexible industry-specific time trends estimated over the entire period. Nevertheless, IMP2 effect remains robust and economically significant. Considering the average of the potential of treatment variable for this sample ($E\left[Intense_s\right] = 0.43$), the point estimate in column 3 implies that IMP2 led to an increase of 1.9 percentage points in entrepreneurs' formalization rate ($= 0.044 \times 0.43$), for a baseline rate of 20%. In line with the results from the previous section, we find no effects of IMP1 on formalization of informal entrepreneurs.[28]

Table 4: Effects on Formalization of Informal Entrepreneurs

Dependent Variable: Formal Entrepreneur (0/1)			
	(1)	(2)	(3)
IMP1	0.003	0.005	-0.001
	(0.034)	(0.032)	(0.026)
IMP2	0.067	0.068	0.044
	(0.032)**	(0.032)**	(0.012)***
Number of individuals	5,151	5,151	5,151
R-squared	0.058	0.059	0.069
Basic Dif-in-Dif Specification	Yes	Yes	Yes
Control Economic Fluctuation	No	Yes	Yes
Control Convergence	No	No	Yes

 Notes: Robust standard errors clustered at the industry level. Significance: ***1 percent, **5 percent, and *10 percent levels. Specifications in all columns correspond to regression 4, detailed in Section 4.2. IMP1 and IMP2 correspond to the estimated coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ in regression 4, respectively. The different specifications are the following: column 1 includes time and entrepreneur fixed effects, time-varying observables at the entrepreneur level (age squared and years of schooling) as well as the interaction terms $Post1_{rt}$ and $Post1_t \times Intense_s$, as described in regression 4; column 2 adds controls for fluctuations in economic activity (the mean monthly wage, the mean number of hours of work, and the log of the total employment for each region, industry and month, as well as their square, cubic and quartic) and the share of males, the average number of years of schooling, and the average age of individuals for each region, industry and month; column 3 adds controls for convergence in formalization rates (interaction terms between the dependent variable computed at the baseline - the average of the dependent variable at the industry-region cell over the period of January 2006 to December 2006 - and linear, square, cubic and quartic time trends).

---

[28]We also find a statistically insignificant coefficient whenever IMP1 effects are estimated on a double-difference specification, analogous to the double-difference estimation of IMP2 effects in column 3 of Table 4 (we find a point estimate of 0.044, standard error of 0.011).

As discussed in Section 4.4, IMP2's coefficient captures the effect of reducing taxes on formalization, given that entry costs had been previously eliminated. As for IMP1, its average effect is a combination of the effects of eliminating registration costs with decreasing tax rates for higher income entrepreneurs and increasing tax rates for lower income ones. In order to disentangle the impacts of reducing entry costs from a reduction in taxes, we explore heterogeneous effects across entrepreneurs' income levels. In particular, if reducing entry costs is not a sufficient condition to induce firms to formalize, then we should find no effect from IMP1 on higher-income entrepreneurs. Similarly, if all formalization effects observed in IMP2 come from lowering the costs of remaining formal, the effects across income levels should follow the pattern implied by IMP's tax structure. That is, we should observe stronger formalization effects for upper income entrepreneurs, and lower or zero effects for lower income ones.

To examine these conjectures, we divide the sample into quartiles of entrepreneurs' annual income, and estimate our most complete specification from Table 4 (column 3) separately for each quartile.[29] Table 5 presents the results. We observe that IMP1 did not have any significant effect on formalization, irrespective of the income quartile considered. In particular, we observe no effect for entrepreneurs in the fourth quartile, for whom tax rates remained roughly constant or were slightly reduced. This result thus implies that reducing entry costs is not a sufficient condition to induce firms to register, otherwise we should observe positive effects of IMP1 on formalization amongst higher income entrepreneurs.

Turning to IMP2, we find that formalization effects are increasing in income. We observe small and statistically insignificant coefficients for the bottom half of the income distribution (columns 2 and 3), and a modest and marginally significant effect for the third quartile. It is in the fourth quartile that we find more substantial and statistically significant effects. These results show that the positive average effect (first column) is entirely driven by entrepreneurs in the higher income quartiles. Considering the fourth quartile, the average of the potential of treatment variable is 0.46, and therefore the point estimate implies that IMP2 increased by 6.1 percentage points the average formalization rate ($= 0.133 \times 0.46$), for a baseline formalization rate of 55.3%.

---

[29]To compute the income quartiles, we use only pre-program (before Jul/2009) observations and regress log-wages onto time, region and industry dummies to remove any systematic variation across time, industries and regions. We use the residuals from this regression to compute the income quartiles and classify individuals accordingly.

Table 5: Formalization of Informal Entrepreneurs: Heterogeneity by Income Level

Dependent Variable: Formal Entrepreneur (0/1)					
	(1)	(2)	(3)	(4)	(5)
IMP1	0.001	-0.049	0.021	-0.087	0.047
	(0.026)	(0.052)	(0.054)	(0.052)	(0.105)
IMP2	0.044	-0.008	0.001	0.039	0.132
	(0.012)***	(0.023)	(0.028)	(0.021)*	(0.027)***
Number of Individuals	5,151	1,288	1,287	1,289	1,287
R-squared	0.069	0.082	0.048	0.093	0.109
Mean($Intense_s$)	0.428	0.468	0.437	0.380	0.433
Sample	Benchmark	1st Quartile	2nd Quartile	3rd Quartile	4th Quartile

 Notes: Robust standard errors clustered at the industry level. Significance: ***1 percent, **5 percent, and *10 percent levels. Specifications in all columns correspond to regression 4, as detailed in Section 4.2. IMP1 and IMP2 correspond to the estimated coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ in regression 4, respectively. The benchmark specification in column 1 corresponds to the third column in Table 4. It includes time and entrepreneur fixed effects, time-varying observables at the entrepreneur level as well as the interaction terms $Post1_{rt}$ and $Post1_t \times Intense_s$, as described in regression 4; controls for fluctuations in economic activity; controls for convergence in formalization rates. Columns 2 to 5 use the same specification in sub-samples of different income quartiles, computed from the residuals of a regression of log-wages onto time, region and industry dummies estimated using only pre-program data. The median income values of each quartile are the following: R$600.98 (1st quart.), R$723.96 (2nd quart.), R$1,082.65 (3rd quart.) and R$1,947.43 (4rd quart.). The median income of the unconditional distribution is R$ 989.39.

5.3 Transitions into Entrepreneurship

We now examine whether the program fostered transitions from employment (both formal and informal) or unemployment into formal entrepreneurship.[30] We also assess whether the program increased the likelihood of survival for existing formal entrepreneurs. This can be particularly relevant, at it seems reasonable to expect that lowering tax rates could have a positive effect on small formal firms' survival. Once again we focus on our most complete specification, and define the sample such that all individuals have at least one observation before and one after IMP1 or IMP2. We sequentially restrict the sample to individuals who report being formal employee (column 1), informal employee (column 2), unemployed (column 3), and formal entrepreneur with up to one employee (column 4) in their first interview. Table 6 shows that reducing taxes did not have any statistically significant effect on transitions into entrepreneurship (columns 1–3), nor on the survival

---

[30]We need individuals' industry affiliation in their first interview in order to construct the treatment variable, which cannot be done for those out of the labor force. For the unemployed we can recover this information from their previous job.

of small formal entrepreneurs (column 4).[31]

Table 6: Transitions into Formal Entrepreneurship and Firm Survival

Dependent Variable: Formal Entrepreneur (0/1)

	(1)	(2)	(3)	(4)
IMP1	-0.005	-0.006	0.022	0.008
	(0.003)	(0.027)	(0.029)	(0.154)
IMP2	0.001	-0.005	0.012	-0.036
	(0.002)	(0.022)	(0.009)	(0.032)
Number of Individuals	15,324	3,297	1,217	2,132
R-squared	0.003	0.025	0.040	0.094
Transition from:	Formal Employee	Informal Employee	Unemployed	Formal Entrepreneur

Notes: Robust standard errors clustered at the industry level. Significance: ***1 percent, **5 percent, and *10 percent levels. Specifications in all columns correspond to regression 4, as detailed in Section 4.2. IMP1 and IMP2 correspond to the estimated coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ in regression 4, respectively. All regressions include time and entrepreneur fixed effects, time-varying observables at the entrepreneur level as well as the interaction terms $Post1_{rt}$ and $Post1_t \times Intense_s$, as described in regression 4; controls for fluctuations in economic activity; controls for convergence in formalization rates. The sample is restricted as follows: (1) formal employee in the first interview; (2) informal employee in the first interview; (3) unemployed in the first interview; (4) formal entrepreneur in the first interview.

These results thus indicate that the sharp reduction in the costs of being a small formal entrepreneur was not a sufficiently strong incentive to induce individuals to move out of employment or unemployment and start a formal business. The results also point against a potential unintended consequence of the program: in order to avoid the costs from labor regulations, larger firms could in principle substitute a formal labor contract for a firm-to-firm contract with the same employee who would now be a "micro entrepreneur". This effect would show up in the data as transitions from employment to entrepreneurship, and we would not be able to separate these competing stories. We find no evidence in support of either interpretation.

---

[31]It should be noted that the estimation does not capture individuals who change industries when starting a formal businesses, as our sample is restricted to individuals who remain in the same industry. This could potentially lead to an underestimation of the effects of IMP2.

5.4 Timing of Formalization Effects

The previous sections have established that IMP increased both the number and the share of small formal businesses, and that this effect seems to have come exclusively from the formalization of the stock of informal firms. In this section we investigate the timing of this latter effect. More specifically, we re-estimate equation 4, but replacing the terms $Post1_{rt} \times Intense_s$ and $Post2_t \times Intense_s$ by interactions between $Intense_s$ and dummies indicating the $d \in \{0, \ldots, 12\}$ months before and after IMP1 and IMP2 implementation, respectively, where $d = 0$ indicates the respective start dates. Thus, by estimating these regressions we are not only investigating the timing of the effects but also examining the presence of pre-trends in the 12 months prior to both IMP1 and IMP2 in a flexible way.

More concretely, for the 12 months prior to the introduction of IMP1 and IMP2, the interaction terms are given by: $\sum_{d=1}^{12} \beta_{1,-d} D_{1,r,-d} \times Intense_s$ and $\sum_{d=1}^{12} \beta_{2,-d} D_{2,-d} \times Intense_s$, respectively. The term $D_{1,r,-d}$ denotes a dummy for the d-th month before IMP1 started; analogously, $D_{2,-d}$ is a dummy for the d-th month before IMP2 started. Note that the first varies across regions because IMP1 start date varied across regions, while the same is not true for IMP2. For the 12 months after the introduction of either IMP1 or IMP2, the interactions are defined analogously: $\sum_{d=0}^{12} \beta_{1,d} D_{1,r,d} \times Intense_s$ and $\sum_{d=0}^{12} \beta_{2,d} D_{2,d} \times Intense_s$, respectively, where $d = 0$ denotes the exact month the two phases started. We estimate the regressions for IMP1 and IMP2 separately, truncating the sample at $d \leq 12$, that is, we do not go further than 12 months after the respective start dates.[32]

Figure 5 plots the timing of the effects for IMP1 (upper panel) and IMP2 (lower panel). The figure confirms the null effect of IMP1, irrespective of which month one considers. Even though the standard errors increase substantially after the 6th month, the point estimates are typically small in magnitude. As for IMP2, the lower panel of Figure 5 shows that its effect was transitory. It kicked in around the fifth month, increased up to sixth month, and then started declining towards zero. By the eight month the effect was statistically zero and by eleventh month the point estimate of the program's effect had already vanished.

---

[32]For the IMP1 regression we use all pre-program observations, starting in 2006. For IMP2, we use the same window around the start date used in the estimations in the previous subsections. We do so in order to completely separate the samples used for IMP1 and IMP2.

Figure 5: Timing of Formalization Effects
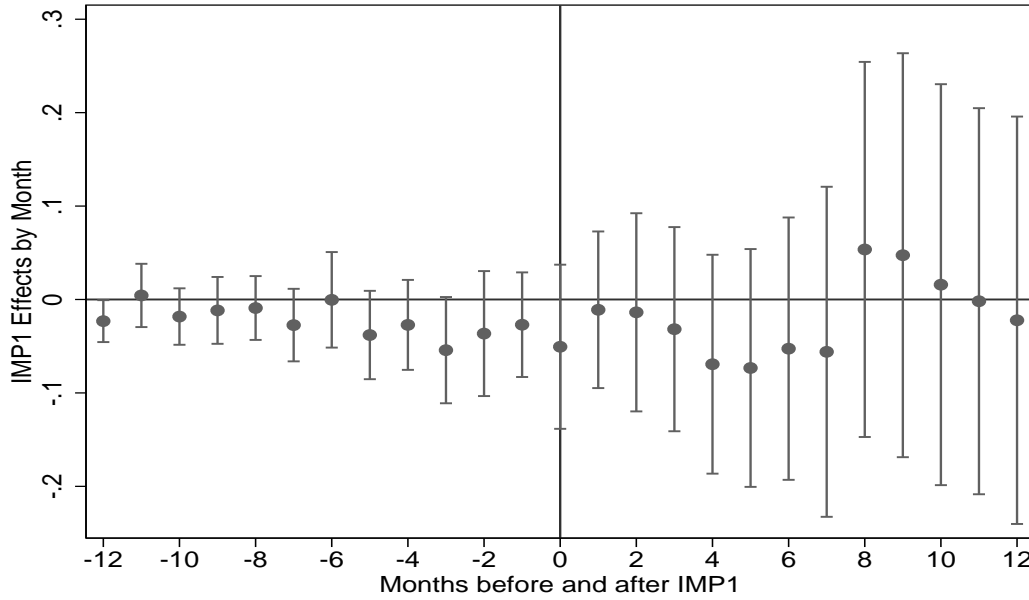
(a) **IMP1 Effects**. Values of the estimated coefficients, $\beta_{1,d}$'s, and their respective standard errors from the interaction terms $\sum_{d=-12}^{12} \beta_{1,d} D_{1,r,d} \times Intense_s$.
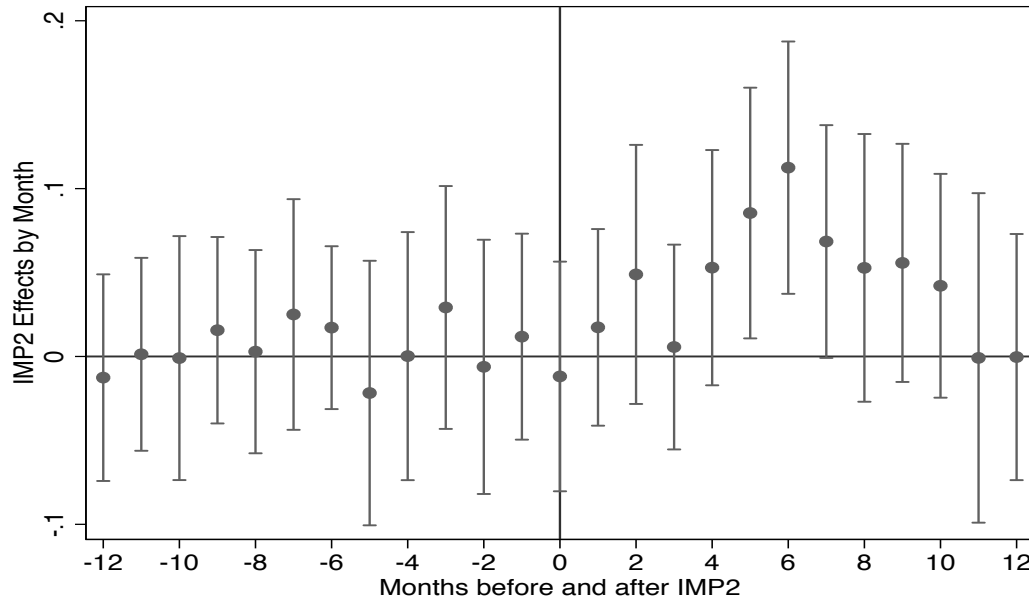
(b) **IMP2 Effects**. Values of the estimated coefficients, $\beta_{2,d}$'s, and their respective standard errors from the interaction terms $\sum_{d=-12}^{12} \beta_{2,d} D_{2,d} \times Intense_s$.

27

5.5 Robustness checks

We start our robustness analysis by performing additional falsification tests of our estimates regarding the timing of the formalization of existing informal businesses. For that, we artificially change the timing of IMP implementation, considering lags of one, two and three years relatively to the actual start date. For each false implementation date, we re-define the time window accordingly, using the same procedure as in our benchmark model. For example, in the first column of Table 7, Panel A, we present the coefficient of a triple-interaction between the potential of treatment intensity, time and region eligibility lagged one year relatively to IMP1's starting date. Panel B reports analogous estimates for IMP2, although in this case we do not have a clear placebo test since the coefficients are also partially capturing the effects of IMP1 over 2010. The second and third columns report the results from analogous regressions considering two and three years lags, respectively.[33] The higher-order lags allow for a true placebo test for IMP2, as the relevant time window in both cases is defined before the actual start date of the program. As Table 7 shows, we find no statistically significant effect in any of these regressions. Thus, there is no association between either IMP1 or IMP2 and formalization rates before the actual implementation of the program. Importantly, the result regarding the first lag of IMP2 indicates that the sample used to identify IMP2's effects is not contaminated by the first phase of the program. This result also reduces our concerns about the influence of unobserved trends.

We run several additional robustness checks, but for the sake of conciseness we briefly comment the results here and include the respective tables in Appendix B. In Subsection B.1 we assess the robustness of our results relatively to different choices of sample and clustering of standard errors. Table B.1 shows that the results from aggregate regressions based on RAIS data (Table 2) hold irrespectively of sample choice and clustering. Table B.2 shows that the results from aggregate regressions based on PME data (Table 3) remain similar once we use a balanced sample of month-industry-region cells. At the individual level, we re-estimate the results regarding formalization of informal entrepreneurs (Table 4) and transitions into formal entrepreneurship (Table 6) without restricting our sample to individuals who have at least one observation before and one after either IMP1 or IMP2. Put differently, we pool all observations over the entire period (January 2006 through August 2012) and re-estimate Tables 4 and 6. As Tables B.3 and B.4 show, the results remain generally similar.

We then focus on our individual-level results on formalization of informal entrepreneurs,

---

[33]We are not able to build the placebo for IMP1 lagged three years because our sample begins in January 2006. It should be also noted that there are fewer observations in Table 7 than in Table 4 because in placebo tests we use only pre-IMP data.

Table 7: Placebo Tests on Formalization of Informal Entrepreneurs

Dependent Variable: Formal Entrepreneur (0/1)

	Lag 1 Year	Lag 2 Years	Lag 3 Years
	(1)	(2)	(3)
IMP1	0.042	0.021	–
	(0.026)	(0.028)	–
R-squared	0.040	0.036	–
Number of individuals	2,875	3,049	–
IMP2	-0.007	0.010	-0.004
	(0.011)	(0.012)	(0.011)
R-squared	0.058	0.048	0.039
Number of individuals	2,573	2,791	2,716

Notes: Robust standard errors clustered at the industry level. Significance: ***1 percent, **5 percent, and *10 percent levels. Specifications in all columns correspond to regression 4, as detailed in Section 4.2. IMP1 and IMP2 correspond to the estimated coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ in regression 4, respectively, considering lags of one, two and three years relatively to their actual start dates. All regressions use our preferred specification used in the third column of Table 4. The controls include time and entrepreneur fixed effects, time-varying observables at the entrepreneur level as well as the interaction terms $Post1_{rt}$ and $Post1_t \times Intense_s$, as described in regression 4; controls for fluctuations in economic activity; controls for convergence in formalization rates.

and examine whether the estimates of Table 4 are affected if one applies the income threshold for eligibility, thus only considering entrepreneurs with annual income below the eligibility threshold (up to R\$36,000 until November 2011, and up to R\$60,000 after that). It is important to emphasize that we only observe earnings and not revenues in the data, so this is an imperfect approximation of the eligibility criteria. Moreover, we should not expect the income threshold to be binding since firms can easily under-report revenues. With these caveats in mind, Table B.5 shows the same patterns obtained before in our main specifications, albeit the point estimates are now smaller than before. In the most complete specification, the coefficient decreases to 0.037 against 0.044 in our benchmark sample. This is expected, as we excluded from the sample higher income entrepreneurs, who are the ones who benefited the most from the IMP's regressive tax scheme. We also re-estimate our regressions excluding the observations after the increase in the eligibility cap, which occurred in November 2011. Table B.6 shows that the results

remain largely unchanged.

In Table B.7 we assess whether the program affected firms with more than one employee. Even though this is a (potentially) sharp eligibility criterion, many small formal firms hire informal workers (Ulyssea, 2017), who are invisible to the government. It is therefore possible that firms registered in IMP have more than one employee if they hire informal workers as well. We thus run the same specifications as before, but considering entrepreneurs with at most five employees. As Table B.7 shows, the same patterns arise and the point estimates are remarkably similar to those obtained in our main results. In the last robustness exercise of Subsection B.1, we adopt an alternative specification that does not condition on individuals' initial status. As Table B.8 shows, the results remain qualitatively the same.

In Subsection B.2 we assess the sensitivity of our results to the definition of the treatment variable. For that, instead of considering the continuous variable $Intense_s$, we define two alternative binary variables:

$$T_i^{50} = \mathbf{1}\left[Intense_{s,i} \geq 0.5\right]$$

$$T_i^{90} = \begin{cases} 1 \text{ if } Intense_{s,i} \geq 0.9 \\ 0 \text{ if } Intense_{s,i} \leq 0.1 \end{cases}$$

The first, T^{50}, considers as treated all entrepreneurs in industries above the median value of $Intense_s$. The second one, T^{90} considers the two extremes of the distribution of $Intense_s$ (see Table A.1 of Appendix A). We consider as treated those with $Intense_s \geq 0.9$ and as control those with $Intense_s \leq 0.1$, while the remaining observations are discarded. As can be seen in Tables B.9 and B.10, the results remain statistically stable. The same is true whenever we combine these two alternative treatment definitions with the unrestricted sample discussed above (Table B.11).

Finally, in Subsection B.3 we estimate two additional placebo exercises. The first exploits the staggered implementation of the program to estimate its effects on the sample of states that were not eligible during the first phase of IMP. The second exercise uses the RAIS data to investigate whether the program had any effect on medium and large firms (more than 5 formal employees), which were not eligible to the program. As Tables B.12–B.15 show, we find no significant results, which reinforces the validity of our identification strategy.

5.6   Cost-Effectiveness

IMP introduced a substantial tax cut, and also a *de facto* subsidy to social security contribution – which remained at 11% of the minimum wage for self-employed and entrepreneurs with up to one employee outside IMP. Given its limited and transitory effects (see Figure 5), it is likely that the program did not lead to higher tax revenues. To assess whether the program was cost effective, we take the zero effect estimated for IMP1 literally and perform a very simple back of the envelope calculation for the impact on tax revenues implied by IMP2. We take the stock of formal and informal entrepreneurs with up to one employee in the month before IMP2 started, which we denote by $\bar{N}^f$ and $\bar{N}^i$, and we assume that the only thing that affects these stocks is the policy (IMP2).[34] We compute tax revenues with and without IMP2 for the $t = 0, \dots, 12$ months after IMP2 was implemented (where $t = 0$ is the month IMP2 started). To do that, we use the monthly average effects computed from the point estimates reported in Figure 5, panel (b) as follows:

$$N_t^i = \left(1 - \tilde{\beta}_{2,t}\right) N_{t-1}^i$$
$$N_t^f = \tilde{\beta}_{2,t} N_{t-1}^i + N_{t-1}^f$$

where $t = 0, \dots, 12$; $N_{-1}^j \equiv \bar{N}^j$; $j = i, f$; $\tilde{\beta}_{2,t} = \hat{\beta}_{2,t} \times E\left[Intense_s\right]$, and $\hat{\beta}_{2,t}$ denotes the point estimates reported in Figure 5, panel (b). It is worth noting that we use the point estimates irrespectively of their statistical significance.

Figure 6 shows the results. There is an immediate drop in tax revenues because the tax cut is effective immediately, while the formalization effect is very small in the first three months. Starting in the fourth month, the formalization effect from IMP2 starts increasing and reaches a peak in the 6th month and reverting back to zero by the 10th month (Figure 5). As a result, tax revenues increase up until the 10th month but not enough to revert all of the losses, and they remain at 80% of their original value before IMP2 was implemented.

It is worth noting that this simple calculation does not account for potential formalization effects that go beyond the direct formalization of the target group. For example, it does not account for potential cascade formalization effects along the lines analyzed by de Paula and Scheinkman (2010). If such effects are sizable, our cost-benefit calculation should be interpreted as an upper bound for the tax revenue losses. However, despite these potential indirect formalization effects, the monthly losses shown in Figure

---

[34]We use the PME individual weights to expand the sample of formal and informal entrepreneurs with up to one employee. By doing that, we obtain 2.05 millions informal entrepreneurs and 772.4 thousand formal entrepreneurs with up to one employee.

6 are likely a *lower bound* because we ignore imperfect compliance with tax payments once entrepreneurs formalize (Bosch et al., 2014, show that non-payment is substantial in IMP).

Figure 6: Impacts on Monthly Tax Revenues
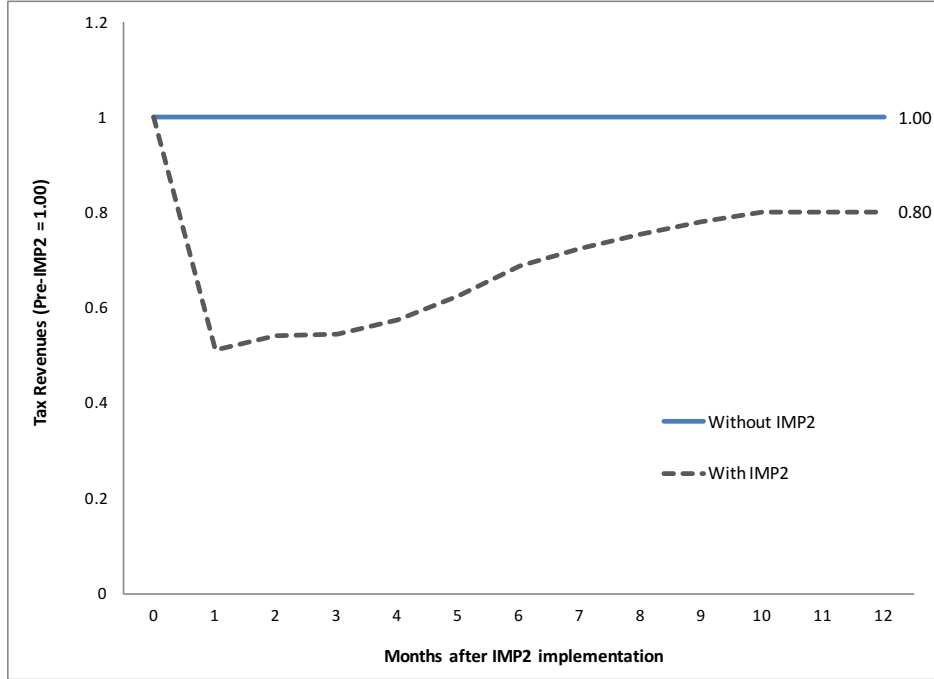
Notes: The black dotted line (with IMP2) is computed using the actual mandated tax expenditures defined in IMP2, times the counterfactual number of formal entrepreneurs, N_t^f, as described in the text: $N_t^f = \hat{\beta}_t N_{t-1}^i + N_{t-1}^f$, where we use all point estimates $\hat{\beta}_t$, regardless of their statistical significance. The blue solid line (without IMP2) is computed holding the pre-IMP2 stock of formal entrepreneurs constant at $\bar{N}^f$ and using the mandated tax expenditures defined in IMP1.

5.7   The Short Run Effects of Formalization on Income

Given the strong set of incentives provided by IMP, a relevant question is why the program did not induce greater formalization. Part of the literature argues that the perceived benefits of formalization are very low for small-scale entrepreneurs (e.g. Bruhn and McKenzie, 2014). If this is the case, one should not expect sizable formalization effects from policies that reduce the costs of formality alone. We assess this conjecture by estimating the effect of formalization on entrepreneurs' income using a 2SLS approach, where access to the program is an instrument for entrepreneurs' formality status. We also estimate the reduced-form effects of IMP2 on entrepreneurs' income. In both cases, we use our preferred specification (column 3 of Table 4) and the log of entrepreneurs' income as the dependent variable.

As Table 8 shows, we do not find any significant effect of IMP2 on income neither when we consider our baseline sample of informal entrepreneurs (column 1), nor when we estimate it separately by income quartiles (columns 2–5). In the sixth column, we estimate an instrumental variable model using IMP2 as an instrument for formalization. The IV model gives a sizable but statistically insignificant point estimate. Thus, the results are consistent with the conjecture that perceived benefits from formalization are low for small firms. It is worth noting, however, that our results refer to the very short term, as we observe individuals at most 14 months after the program was implemented. If formalization effects can only be felt in the mid to long run, then one should not expect to find any effects in these regressions to begin with. This will be the case, for example, if the short run effects of formalization are concentrated on greater access to the financial system and credit lines, a larger customer base, increased access to capital and so on, which in turn could translate into higher income in the future (e.g. Fajnzylber et al., 2011). Unfortunately, the employment survey does not have detailed information about entrepreneurs' businesses, so we cannot test for the presence of these channels.

Table 8: Effects on Income

Dep. Variable: Entrepreneur's Log(Income)						
	(1)	(2)	(3)	(4)	(5)	(6)
IMP2	0.008	-0.015	-0.016	-0.013	0.002	–
	(0.025)	(0.057)	(0.032)	(0.032)	(0.040)	–
Formal (Dummy)	–	–	–	–	–	0.122
	–	–	–	–	–	(0.350)
First Stage F-Stat	–	–	–	–	–	12.5
Observations	12,931	2,774	3,253	3,382	3,522	12,931
R-squared	0.027	0.180	0.070	0.074	0.078	
Number of Indiv.	2,592	633	655	651	653	2,592
Sample	All	1st Quart.	2nd Quart.	3rd Quart.	4th Quart.	All
Model	OLS	OLS	OLS	OLS	OLS	2SLS

Notes: Robust standard errors clustered at the industry level. Significance: ***1 percent, **5 percent, and *10 percent levels. Specifications in all columns correspond to regression 4, as detailed in Section 4.1. IMP1 and IMP2 correspond to the estimated coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ in regression 4, respectively. All regressions include time and entrepreneur fixed effects, time-varying observables at the entrepreneur level as well as the interaction terms $Post1_{rt}$ and $Post1_t \times Intense_s$, as described in regression 4; controls for fluctuations in economic activity; controls for convergence in formalization rates. Columns 1-5 report estimates of the reduced-form regression of log income on IMP2 for the entire sample. Column 1 uses the entire sample and columns 2-5 the 1st through the 4th quartiles, respectively. The quartiles were defined as described in Table 5. Column 6 reports the 2SLS regression of log income on a formal dummy instrumented by IMP2.

6 Final Remarks

This paper assesses whether entry regulation and taxes hinder firm formalization. To this end, we estimate the effects of a large scale formalization program in Brazil, the *Individual Micro-Entrepreneur Program* (IMP). We use both administrative firm-level data and a panel of entrepreneurs to exploit time, inter-industry and inter-region variations in access to the program to separately estimate the effect of reducing registration costs and of reducing taxes once entry costs have already been eliminated.

Our results indicate that while reducing registration costs has no effect on firm informality, reducing the tax burden does increase formalization. This effect comes from existing informal firms becoming formal and not from the creation of new formal businesses nor greater survival of existing formal firms. We also find no evidence that those entrepreneurs who formalized due to the policy observed an increase in their income in

the short run, which is consistent with the view that small firms perceive little benefits from formalization. For example, De Mel et al. (2013) find (in a experimental setting) that eliminating formal sector's entry costs alone does not induce firms to formalize; only when provided with substantial monetary payments some entrepreneurs decide to register their business. Accordingly, our results indicate that the main restriction to formalization are not entry costs, but the costs of remaining formal. In terms of cost-effectiveness, the formalization effects from the tax reduction are too modest to generate a net gain in terms of tax revenues and our back of the envelope calculation suggests a net monthly loss in tax revenues.

References

Bertrand, M. and F. Kramarz (2002). Does entry regulation hinder job creation? evidence from the french retail industry. *The Quarterly Journal of Economics 117*(4), pp. 1369–1413.

Besley, T. and T. Persson (2013). Taxation and development. In A. J. Auerbach, R. Chetty, M. Feldstein, and E. Saez (Eds.), *Handbook of Public Economics*, Volume 5, pp. 51–110. Elsevier.

Bosch, M., D. Fernandes, and J. M. Villa (2014). Nudging the self-employed into contributing to social security: Evidence from at scale quasi experiment in brazil. Mimeo.

Bruhn, M. (2011). License to sell: The effect of business registration reform on entrepreneurial activity in Mexico. *Review of Economics and Statistics 93*(1), 382–386.

Bruhn, M. and D. McKenzie (2014). Entry regulation and the formalization of microenterprises in developing countries. *The World Bank Research Observer 29*(2), 186–201.

De Andrade, G. H., M. Bruhn, and D. McKenzie (2013). A helping hand or the long arm of the law? Experimental evidence on what governments can do to formalize firms. Mimeo.

De Giorgi, G. and A. Rahman (2013). SME's registration: Evidence from an RCT in Bangladesh. *Economics Letters 120*(3), 573 – 578.

De Mel, S., D. McKenzie, and C. Woodruff (2013). The demand for, and consequences of, formalization among informal firms in Sri Lanka. *American Economic Journal: Applied Economics 5*(2), 122–50.

de Paula, A. and J. A. Scheinkman (2010). Value-added taxes, chain effects, and informality. *American Economic Journal: Macroeconomics 2*(4), 195–221.

De Soto, H. (1989). *The Other Path.* Harper e Row, New York.

Djankov, S., R. L. Porta, F. Lopez-De-Silanes, and A. Shleifer (2002, February). The regulation of entry. *The Quarterly Journal of Economics 117*(1), 1–37.

Fajnzylber, P., W. F. Maloney, and G. V. Montes-Rojas (2011). Does formality improve micro-firm performance? Evidence from the Brazilian SIMPLES program. *Journal of Development Economics 94*(2), 262 – 276.

Hsieh, C.-T. and B. Olken (2014). The missing "missing middle". *Journal of Economic Perspectives 28*(3), 89–108.

IFC (2009). Doing business 2010: Reforming through difficult times. International Finance Corporation, Washington, DC.

Kaplan, D. S., E. Piedra, and E. Seira (2011). Entry regulation and business start-ups: Evidence from Mexico. *Journal of Public Economics 95*(11), 1501–1515.

La Porta, R. and A. Shleifer (2014). Informality and development. *The Journal of Economic Perspectives 28*(3), 109–126.

Monteiro, J. and J. J. Assunção (2012). Coming out of the shadows? Estimating the impact of bureaucracy simplification and tax cut on formality in Brazilian microenterprises. *Journal of Development Economics 99*(1), 105–115.

Ulyssea, G. (2017). Firms, informality and development: Theory and evidence from Brazil. Mimeo.

Appendix

A   Intensity of Treatment and Pre-Trends

Table A.1: Variable Intense: Tabulation for Different Samples

Var.	Frequencies for Different Sub-samples			
Intense	Informal Entrepren.	Inf. Employee	Form. Employee	All
0.0000	0.221	0.215	0.224	0.216
0.0922	0.203	0.176	0.180	0.205
0.1303	0.045	0.039	0.054	0.047
0.3587	0.018	0.011	0.027	0.017
0.3827	0.014	0.017	0.014	0.015
0.3834	0.009	0.010	0.011	0.009
0.6141	0.057	0.060	0.061	0.060
0.7301	0.038	0.039	0.041	0.037
0.8223	0.111	0.109	0.133	0.114
0.8858	0.041	0.060	0.036	0.041
0.8968	0.035	0.052	0.011	0.036
0.9281	0.003	0.003	0.003	0.003
1.0000	0.204	0.209	0.205	0.200

Notes: First column contains the values observed for the variable $Intense_s = \frac{\sum_{k \in s}(\mathbf{I}[\text{Sub-Industry } k \text{ is eligible}] \times N_{ks})}{\sum_{k \in s} N_{ks}}$. The remainder columns contain the frequencies for each sub-sample used in regressions: informal entrepreneur in the first interview (column 2), informal employee in the first interview (column 3), formal employee in the first interview (column 4), and the entire sample (column 5).

Table A.2: Pre-Trends and Industry Eligibility

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dependent variable: $Intense_s$						
$\Delta_{2009-2006}$(Formality Rate)	-0.021				-0.145		-0.049
	(0.188)				(0.238)		(0.294)
$\Delta_{2009-2008}$(Formality Rate)		0.234			0.299		0.297
		(0.329)			(0.339)		(0.296)
$\Delta_{2009-2006}$(Log Income)			-0.082			-0.105	-0.113
			(0.101)			(0.086)	(0.110)
$\Delta_{2009-2008}$(Log Income)				0.003		0.059	0.043
				(0.124)		(0.115)	(0.088)
Observations	216	216	216	216	216	216	216

Notes: Robust standard errors clustered at the metropolitan region level. Significant at the *** 1 percent, **5 percent, and *10 percent levels. Data from PME data collapsed at the industry and metropolitan region level. In the first column, the independent variable is the change in formality rate between June 2006 and June 2009, while in the second column it is defined as the change between June 2008 and June 2009. In the remaining columns the change in the logarithm of average income is defined analogously. All specifications include metropolitan region fixed-effects.

B   Additional Robustness Checks

B.1   Sample choice and clustering

We first provide further robustness tests using different samples and choices of clustering. Table B.1 presents the results. All specifications follow regression 2. We report standard errors clustered at the industry level in parentheses (our benchmark), and clustered at the state level in brackets. Columns 1 and 2 report specifications based on unbalanced samples of industry-state-trimester cells, respectively for the entire country and restricted to the metropolitan regions available in the PME. In this case, we drop cells with missing observations (for which we do not observe any firm), thus avoiding the inclusion of zeros for these missing observations that potentially reflect the fact that small-scale firms are unlikely to exist in these industries or regions for structural/technological reasons (e.g. agricultural industries in metropolitan regions). In these samples, we also include the industries that experienced variation in the eligibility status over time. Columns 3 and 4 report specifications based on the balanced samples of industry-state-trimester cells, while in columns 5 and 6 we report our benchmark specifications (balanced sample, without industries whose eligibility status varied over time). Overall, we observe a slightly decrease in the point estimates going from columns 1 and 2 to columns 3 and 4, although the differences are not statistically significant. The results remain remarkably stable as we move from columns 3 and 4, to columns 5 and 6. As for the computation of standard errors, in general we observe they tend to remain relatively stable, slightly decreasing as me move from clustering at the industry to the state level.

Table B.1: Effects on the Number of Formal Firms Using Administrative Data (RAIS): Testing Different Samples and Estimates of Standard Errors

Dependent Variable: Log(# Formal Firms)

	All Regions	PME	All Regions	PME	All Regions	PME
	(1)	(2)	(3)	(4)	(5)	(6)
IMP1	-0.007	0.017	-0.010	0.010	-0.011	0.010
	(0.008)	(0.021)	(0.007)	(0.015)	(0.007)	(0.016)
	[0.007]	[0.009]	[0.006]	[0.009]	[0.006]*	[0.010]
IMP2	0.048	0.035	0.043	0.032	0.043	0.032
	(0.007)***	(0.011)***	(0.006)***	(0.009)***	(0.006)***	(0.009)***
	[0.006]***	[0.006]***	[0.005]***	[0.008]***	[0.005]***	[0.008]***
Balanced	No	No	Yes	Yes	Yes	Yes
Exclude ΔEligibility	No	No	No	No	Yes	Yes
Observations	379,030	81,252	506,520	112,560	500,472	111,216
R-squared	0.831	0.792	0.844	0.820	0.845	0.821

Notes: Robust standard errors clustered at the industry level in parentheses, and at the state level in brackets. Significance: ***1 percent, **5 percent, and *10 percent levels. All specifications follow regression 2, detailed in Section 4.1. Columns 1, 3 and 5 consider the sample including all regions in the country, while in the remainder we restrict the sample to the six main metropolitan regions in Brazil, which constitute the sample used by the Monthly Employment Survey (PME): Porto Alegre, Rio de Janeiro, Recife, Salvador, Belo Horizonte and São Paulo. IMP1 and IMP2 correspond to the estimated coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ in regression 2, respectively. Columns 1 and 2 report specifications based on unbalanced samples of industry-state-trimester cells and include those industries that experienced variation in the eligibility status over time. Columns 3 and 4 report specifications based on the balanced samples of industry-state-trimester cells, while columns 5 and 6 report specifications based on balanced samples, without industries whose eligibility status varied over time.

Table B.2 shows that the results of Table 3 hold irrespectively of sample selection. We report results based on our benchmark specification, but ran on a balanced sample of industry-region-month cells. In order to build this balanced panel, we identify the missing observations in a given industry-region-month and replace them by their average of the respective year and industry-region cell (we keep the cell missing if they are missing within the entire year). Overall, the results are similar to those of Table 3.

Table B.2: Effects at the Industry-Region Level using the Monthly Employment Survey (PME): Robustness Checks

Dep. Var.: Share of	Formal Employees	Informal Employees	Unemployed	Entrepren.	Small Formal Entrepren.
	(1)	(2)	(3)	(4)	(5)
IMP1	0.036	-0.029	-0.011	0.004	0.002
	(0.035)	(0.030)	(0.008)	(0.009)	(0.004)
IMP2	-0.012	0.009	-0.003	0.006	0.013
	(0.014)	(0.008)	(0.004)	(0.009)	(0.004)***
Observations	17,976	17,976	17,976	17,976	17,976
R-squared	0.623	0.211	0.067	0.675	0.353

Notes: Robust standard errors clustered at the industry level. Significance: ***1 percent, **5 percent, and *10 percent levels. Specifications in all columns correspond to a regression analogous to expression 3, which is estimated on a balanced sample of industry-region-month cells, and excludes time-varying controls (see text for details).

In Tables B.3 and B.4 we do not restrict the sample to individuals who are observed at least once before and after either IMP1 or IMP2. This implies that some individuals might have all their observations before or after either IMP1 or IMP2 were implemented. In Table B.5 we use the annual income eligibility threshold to restrict our sample, while in Table B.6 we limit our time horizon to the period prior to the increase in the revenue eligibility cap. In Table B.7 we extend our sample to the entrepreneurs with up to five employees

Table B.3: Formalization of Informal Entrepreneurs: Full Sample

Dependent Variable: Formal Entrepreneur (0/1)

	(1)	(2)	(3)
IMP1	-0.010	-0.012	-0.007
	(0.021)	(0.021)	(0.020)
IMP2	0.058	0.059	0.036
	(0.031)*	(0.031)*	(0.011)***
Number of individuals	59,976	59,976	59,976
R-squared	0.028	0.028	0.034
Basic Dif-in-Dif Specification	Yes	Yes	Yes
Control Economic Fluctuation	No	Yes	Yes
Control Convergence	No	No	Yes

Notes: Robust standard errors clustered at the industry level. Significance: ***1 percent, **5 percent, and *10 percent levels. Specifications in all columns correspond to regression 4, detailed in Section 4.2. IMP1 and IMP2 correspond to the estimated coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ in regression 4, respectively. The sample is not restricted to individuals who are observed at least once before and after either IMP1 or IMP2. The different specifications are the following: column 1 includes time and entrepreneur fixed effects, time-varying observables at the entrepreneur level (age squared and years of schooling) as well as the interaction terms $Post1_{rt}$ and $Post1_t \times Intense_s$, as described in regression 4; column 2 adds controls for fluctuations in economic activity (the mean monthly wage, the mean number of hours of work, and the log of the total employment for each region, industry and month, as well as their square, cubic and quartic) and the share of males, the average number of years of schooling, and the average age of individuals for each region, industry and month; column 3 adds controls for convergence in formalization rates (interaction terms between the dependent variable computed at the baseline - the average of the dependent variable at the industry-region cell over the period of January 2006 to December 2006 - and linear, square, cubic and quartic time trends).

Table B.4: Transitions into Formal Entrepreneurship and Firm Survival: Full Sample

Dependent Variable: Formal Entrepreneur (0/1)				
	(1)	(2)	(3)	(4)
IMP1	-0.004	-0.031	0.026	-0.109
	(0.004)	(0.015)**	(0.021)	(0.073)
IMP2	0.001	-0.006	0.015	-0.038
	(0.002)	(0.022)	(0.010)	(0.034)
Observations	649,656	147,017	40,249	75,956
R-squared	0.001	0.011	0.013	0.067
Number of individuals	163,740	37,927	12,344	18,113
Transition from:	Formal Employee	Informal Employee	Unemployed	Formal Entrepren.

Notes: Robust standard errors clustered at the industry level. Significance: ***1 percent, **5 percent, and *10 percent levels. All results correspond to the specification described in regression 4. The sample is not restricted to individuals who are observed at least once before and after either IMP1 or IMP2. All regressions include time and entrepreneur fixed effects, time-varying observables at the entrepreneur level as well as the interaction terms $Post1_{rt}$ and $Post1_t \times Intense_s$, as described in regression 4; controls for fluctuations in economic activity and for convergence in the formalization rates. The sample is restricted as follows: (1) formal employee in the first interview; (2) informal employee in the first interview; (3) unemployed in the first interview; (4) formal entrepreneur in the first interview.

Table B.5: Formalization of Informal Entrepreneurs - Eligibility using Annual Income

Dependent Variable: Formal Entrepreneur (0/1)			
	(1)	(2)	(3)
IMP1	-0.001	0.001	-0.002
	(0.025)	(0.026)	(0.027)
IMP2	0.054	0.056	0.037
	(0.024)**	(0.025)**	(0.010)***
R-squared	0.049	0.050	0.054
Number of individuals	4,889	4,889	4,889
Basic Dif-in-Dif Specification	Yes	Yes	Yes
Control Economic Fluctuation	No	Yes	Yes
Control Convergence	No	No	Yes

Notes: Robust standard errors clustered at the industry level. Significance: ***1 percent, **5 percent, and *10 percent levels. Specifications in all columns correspond to regression 4, detailed in Section 4.2. IMP1 and IMP2 correspond to the estimated coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ in regression 4, respectively. We use the annual income eligibility threshold to further restrict our sample. The different specifications are the following: column 1 includes time and entrepreneur fixed effects, time-varying observables at the entrepreneur level (age squared and years of schooling) as well as the interaction terms $Post1_{rt}$ and $Post1_t \times Intense_s$, as described in regression 4; column 2 adds controls for fluctuations in economic activity (the mean monthly wage, the mean number of hours of work, and the log of the total employment for each region, industry and month, as well as their square, cubic and quartic) and the share of males, the average number of years of schooling, and the average age of individuals for each region, industry and month; column 3 adds controls for convergence in formalization rates (interaction terms between the dependent variable computed at the baseline - the average of the dependent variable at the industry-region cell over the period of January 2006 to December 2006 - and linear, square, cubic and quartic time trends).

Finally, instead of selecting the sample according to individuals' initial status (as done in Table 4 onwards), we use the entire sample and include their status in $t-1$ as a control in a regression that has the dummy for small formal entrepreneur as the dependent variable. More concretely, we use a specification analogous to that in expression 4 but instead of estimating one regression for each sub-sample defined by individuals' initial status, we include lagged dummies for formal employee, informal employee, small informal entrepreneur, small formal entrepreneur and unemployed. We fix these dummies so that they do not change after either IMP1 or IMP2 starts, otherwise these would be endogenous regressors, as individuals' statuses are potentially responding to the program.

This regression captures the average effect of IMP1 and IMP2 on the transition into formal entrepreneurship from all statuses, including formal entrepreneurship itself (which

Table B.6: Formalization of Informal Entrepreneurs – Prior to eligibility cap increase

Dependent Variable: Formal Entrepreneur (0/1)			
	(1)	(2)	(3)
IMP2	0.074	0.069	0.038
	(0.035)**	(0.036)*	(0.015)**
Number of individuals	2,389	2,389	2,389
R-squared	0.068	0.070	0.085
Basic Dif-in-Dif Specification	Yes	Yes	Yes
Control economic fluctuation	No	Yes	Yes
Control Convergence	No	No	Yes

Notes: Robust standard errors clustered at the industry level. Significance: ***1 percent, **5 percent, and *10 percent levels. Specifications in all columns correspond to regression 4, detailed in Section 4.2. IMP2 correspond to the estimated coefficients $\hat{\beta}_2$ in regression 4 prior to eligibility cap increase. The different specifications are the following: column 1 includes time and entrepreneur fixed effects, time-varying observables at the entrepreneur level (age squared and years of schooling) as well as the interaction terms $Post1_{rt}$ and $Post1_t \times Intense_s$, as described in regression 4; column 2 adds controls for fluctuations in economic activity (the mean monthly wage, the mean number of hours of work, and the log of the total employment for each region, industry and month, as well as their square, cubic and quartic) and the share of males, the average number of years of schooling, and the average age of individuals for each region, industry and month; column 3 adds controls for convergence in formalization rates (interaction terms between the dependent variable computed at the baseline - the average of the dependent variable at the industry-region cell over the period of January 2006 to December 2006 - and linear, square, cubic and quartic time trends).

corresponds to the survival rate in formal entrepreneurship). Thus, this specification combines the effects estimated in Tables 4 and 6. We report the results from this specification in the first column of Table B.8 below. Columns 2-6 reproduce the results from Table 4 (column 3) and Table 6 (columns 3-6), and are included for comparison purposes only. As Table B.8 shows, the same pattern from our main results arises: IMP1 has no effect whatsoever, while IMP2 has a positive and significant effect. The point estimate is lower than the one in column 2 exactly because it represents a combination of the effects estimated separately for different transitions displayed in columns 2–6. Since the only transition affected was the one from informal to formal entrepreneurship, it is not surprising that when we pool all possible origins together the point estimate decreases, as shown by column 1.

Table B.7: Effects on Formalization of Informal Entrepreneurs: Entrepreneurs with up to five employees

Dependent Variable: Formal Entrepreneur (0/1)			
	(1)	(2)	(3)
IMP1	-0.009	-0.004	-0.011
	(0.040)	(0.037)	(0.030)
IMP2	0.067	0.068	0.042
	(0.034)*	(0.035)*	(0.013)***
Number of ind	5,151	5,151	5,151
R-squared	0.060	0.060	0.072
Basic Dif-in-Dif Specification	Yes	Yes	Yes
Control Economic Fluctuation	No	Yes	Yes
Control Convergence	No	No	Yes

Notes: Robust standard errors clustered at the industry level. Significance: ***1 percent, **5 percent, and *10 percent levels. Specifications in all columns correspond to regression 4, detailed in Section 4.2 and including entrepreneurs with up to five employees. IMP1 and IMP2 correspond to the estimated coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ in regression 4, respectively. The different specifications are the following: column 1 includes time and entrepreneur fixed effects, time-varying observables at the entrepreneur level (age squared and years of schooling) as well as the interaction terms $Post1_{rt}$ and $Post1_t \times Intense_s$, as described in regression 4; column 2 adds controls for fluctuations in economic activity (the mean monthly wage, the mean number of hours of work, and the log of the total employment for each region, industry and month, as well as their square, cubic and quartic) and the share of males, the average number of years of schooling, and the average age of individuals for each region, industry and month; column 3 adds controls for convergence in formalization rates (interaction terms between the dependent variable computed at the baseline - the average of the dependent variable at the industry-region cell over the period of January 2006 to December 2006 - and linear, square, cubic and quartic time trends).

B.2 Alternative definitions of the treatment variable

In Tables B.9 and B.10 we assess the robustness of our results to changes in the definition of our treatment variable. Instead of considering the continuous variable $Intense_s$, we define two alternative dichotomous treatment variables:

$$T_i^{50} = \mathbf{1}\left[Intense_{s,i} \geq 0.5\right]$$

$$T_i^{90} = \begin{cases} 1 \text{ if } Intense_{s,i} \geq 0.9 \\ 0 \text{ if } Intense_{s,i} \leq 0.1 \end{cases}$$

In Table B.11 we use both alternative treatment variables combined with the unrestricted sample used in Tables B.3 and B.4.

Table B.8: Transitions into Formal Entrepreneurship Without Conditioning on Initial States

Dependent Variable: Formal Entrepreneur (0/1)

	(1)	(2)	(3)	(4)	(5)	(6)
IMP1	-0.002	-0.001	-0.005	-0.006	0.022	0.008
	(0.011)	(0.026)	(0.003)	(0.027)	(0.029)	(0.154)
IMP2	0.011	0.044	0.001	-0.005	0.012	-0.036
	(0.006)*	(0.012)***	(0.002)	(0.022)	(0.009)	(0.032)
R-squared	0.256	0.069	0.003	0.025	0.040	0.094
Number of Individuals	28,044	5,151	15,324	3,297	1,217	2,132
Transition from:	All States	Informal Entrep.	Formal Employee	Informal Employee	Unemployed	Formal Entrep.

Notes: Robust standard errors clustered at the industry level. Significant at the *** 1 percent, **5 percent, and *10 percent levels. All results correspond to the specification described in regression 4. Column (1) corresponds to the specification where we do not condition the sample to individuals' initial statuses but we control for their statuses up to the start dates of either IMP1 or IMP2; Column (2) corresponds to the last column of Table 4 and restricts the sample to those who were informal entrepreneurs in their first interview; and Columns (3)-(6) correspond to columns (1)-(4) in Table 6 and restrict the sample as follows: (1) formal employee in the first interview; (2) informal employee in the first interview; (3) unemployed in the first interview; (4) formal entrepreneur in the first interview. Columns (2)-(6) are included for comparison purposes.

Table B.9: Effects on Formalization of Informal Entrepreneurs: Alternative Treatment Definition (T^{50})

Dependent Variable: Formal Entrepreneur (0/1)			
	(1)	(2)	(3)
IMP1	0.011	0.015	0.006
	(0.028)	(0.027)	(0.021)
IMP2	0.059	0.060	0.038
	(0.027)**	(0.027)**	(0.010)***
R-squared	0.058	0.059	0.069
Number of individuals	5,151	5,151	5,151
Basic Dif-in-Dif Specification	Yes	Yes	Yes
Control Economic Fluctuation	No	Yes	Yes
Control Convergence	No	No	Yes

Notes: Robust standard errors clustered at the industry level. Significance: ***1 percent, **5 percent, and *10 percent levels. We change the definition of our treatment variable. Instead of considering the continuous variable $Intense_s$, we define an alternative dichotomous treatment variables: $T_i^{50} = \mathbf{1}\left[Intense_{s,i} \geq 0.5\right]$. Specifications in all columns correspond to regression 4, detailed in Section 4.2. IMP1 and IMP2 correspond to the estimated coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ in regression 4, respectively. The different specifications are the following: column 1 includes time and entrepreneur fixed effects, time-varying observables at the entrepreneur level (age squared and years of schooling) as well as the interaction terms $Post1_{rt}$ and $Post1_t \times Intense_s$, as described in regression 4; column 2 adds controls for fluctuations in economic activity (the mean monthly wage, the mean number of hours of work, and the log of the total employment for each region, industry and month, as well as their square, cubic and quartic) and the share of males, the average number of years of schooling, and the average age of individuals for each region, industry and month; column 3 adds controls for convergence in formalization rates (interaction terms between the dependent variable computed at the baseline - the average of the dependent variable at the industry-region cell over the period of January 2006 to December 2006 - and linear, square, cubic and quartic time trends).

Table B.10: Effects on Formalization of Informal Entrepreneurs: Alternative Treatment Definition (T^{90})

Dependent Variable: Formal Entrepreneur (0/1)			
	(1)	(2)	(3)
IMP1	-0.002	-0.002	0.006
	(0.032)	(0.036)	(0.035)
IMP2	0.031	0.032	0.030
	(0.016)*	(0.017)*	(0.012)**
R-squared	0.046	0.047	0.051
Number of individuals	3,432	3,432	3,432
Basic Dif-in-Dif Specification	Yes	Yes	Yes
Control Economic Fluctuation	No	Yes	Yes
Control Convergence	No	No	Yes

Notes: Robust standard errors clustered at the industry level. Significance: ***1 percent, **5 percent, and *10 percent levels. We change the definition of our treatment variable. Instead of considering the continuous variable $Intense_s$, we define an alternative dichotomous treatment variables: $T_i^{90} = \mathbf{1}\left[Intense_{s,i} \geq 0.9\right]$. Specifications in all columns correspond to regression 4, detailed in Section 4.2. IMP1 and IMP2 correspond to the estimated coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ in regression 4, respectively. The different specifications are the following: column 1 includes time and entrepreneur fixed effects, time-varying observables at the entrepreneur level (age squared and years of schooling) as well as the interaction terms $Post1_{rt}$ and $Post1_t \times Intense_s$, as described in regression 4; column 2 adds controls for fluctuations in economic activity (the mean monthly wage, the mean number of hours of work, and the log of the total employment for each region, industry and month, as well as their square, cubic and quartic) and the share of males, the average number of years of schooling, and the average age of individuals for each region, industry and month; column 3 adds controls for convergence in formalization rates (interaction terms between the dependent variable computed at the baseline - the average of the dependent variable at the industry-region cell over the period of January 2006 to December 2006 - and linear, square, cubic and quartic time trends).

Table B.11: Effects on Formalization of Informal Entrepreneurs: Alternative Treatment Definition - Full Sample

Dependent Variable: Formal Entrepreneur (0/1)						
	$Treatment = T^{50}$			$Treatment = T^{90}$		
	(1)	(2)	(3)	(4)	(5)	(6)
IMP1	-0.013	-0.015	-0.014	0.008	0.009	0.014
	(0.017)	(0.018)	(0.019)	(0.022)	(0.022)	(0.022)
IMP2	0.061	0.061	0.035	0.031	0.031	0.029
	(0.026)**	(0.026)**	(0.010)***	(0.016)*	(0.016)*	(0.013)**
Number of individuals	53,314	53,314	53,314	37,535	37,535	37,535
R-squared	0.029	0.030	0.036	0.023	0.023	0.024
Basic Dif-in-Dif Specification	Yes	Yes	Yes	Yes	Yes	Yes
Control Economic Fluctuation	No	Yes	Yes	No	Yes	Yes
Control Convergence	No	No	Yes	No	No	Yes

Notes: Robust standard errors clustered at the industry level. Significance: ***1 percent, **5 percent, and *10 percent levels. Specifications in all columns correspond to regression 4, detailed in Section 4.2. IMP1 and IMP2 correspond to the estimated coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ in regression 4, respectively. In this table we use both alternative treatment variables T_i^{50} and T_i^{90} combined with the unrestricted sample used in Tables B.3 and B.4. The different specifications are the following: columns 1 and 3 include time and entrepreneur fixed effects, time-varying observables at the entrepreneur level (age squared and years of schooling) as well as the interaction terms $Post1_{rt}$ and $Post1_t \times Intense_s$, as described in regression 4; columns 2 and 4 add controls for fluctuations in economic activity (the mean monthly wage, the mean number of hours of work, and the log of the total employment for each region, industry and month, as well as their square, cubic and quartic) and the share of males, the average number of years of schooling, and the average age of individuals for each region, industry and month; columns 3 and 6 add controls for convergence in formalization rates (interaction terms between the dependent variable computed at the baseline - the average of the dependent variable at the industry-region cell over the period of January 2006 to December 2006 - and linear, square, cubic and quartic time trends).

B.3 Additional placebo tests

We start by assessing the effects of the program when one restricts the sample to the states that were not eligible during the first phase of the program. More specifically, we restrict the time horizon up to February 2010, which is the moment where all states are included in the program. Between July 2009 and February 2010, we only include the states that were not eligible at each point in time (see Figure 1). It is worth emphasizing that despite its usefulness, this placebo exercise has three important limitations: (i) it relies on a very short time horizon; (ii) it is concentrated in the aftermath of a very adverse economic shock, the 2009 great recession; and (iii) does not allow for a triple difference approach in order to clean the potential confounding effects of this shock. In order to at least partially address these problems we rely on the PME data, which has monthly frequency and provides us with a larger set of controls for economic fluctuation. In Tables B.12, B.13 and B.14 we apply this placebo specification to re-estimate the aggregate regressions, transitions from informal to formal entrepreneurship and other transitions into formal entrepreneurship. As the tables show, we do not find any significant results in these regressions, except for a marginally significant (at 10%), negative effect in the transition from formal employment to formal entrepreneurship (Table B.14, column 1). We do not place much weight in this particular result because of the caveat made above that this specification does not allow us to use the triple difference specification. Therefore, it becomes harder to adequately clean the effect of the 2009 economic shock, which can be an important confounder.

Table B.12: Effects at the Industry-Region Level using Non-Eligible States (PME)

Dep. Var.: Share of	Formal Employees	Informal Employees	Unemployed	Entrepren.	Small Formal Entrepren.
	(1)	(2)	(3)	(4)	(5)
IMP1	0.0113	-0.0265	0.0063	0.0088	0.0168
	(0.0416)	(0.0168)	(0.0201)	(0.0336)	(0.0106)
Observations	9,655	9,655	9,655	9,655	9,655
R-squared	0,602	0,242	0,108	0,631	0,332

Notes: Robust standard errors clustered at the industry level. Significant at the *** 1 percent, **5 percent, and *10 percent levels. Regressions follow a specification analogous to that of expression 3: $Y_{rts} = \beta(Post1 \times Intense_s) + \gamma_1 Post1 + \lambda_t + \nu_s + \xi_r + \theta Trends_{srt} + X'_{srt}\delta + u_{rst}$. The term $Trends_{srt}$ is the same as described in Section 4.1.

Our final placebo exercise estimates the effects of IMP on medium and large firms that were not eligible to the program. This exercise is only possible if one uses the RAIS data set, as the PME only contains information about firm size for those who have at most 5 employees (formal or informal). As we show in Table B.7 above, exactly because these smaller firms can hire some informal workers, they are still able to benefit from the program and indeed we show that there is some spillover to these firms in the PME data. We therefore use the RAIS data set for this exercise, which contains information

Table B.13: Effects on Formalization of Informal Entrepreneurs in Non-Eligible States

Dependent Variable: Formal Entrepreneur (0/1)

	(1)	(2)	(3)
IMP1	-0.017	-0.013	-0.029
	(0.044)	(0.042)	(0.036)
Number of individuals	2,932	2,932	2,932
R-squared	0.022	0.025	0.036
Basic Dif-in-Dif Specification	Yes	Yes	Yes
Control Economic Fluctuation	No	Yes	Yes
Control Convergence	No	No	Yes

Notes: Robust standard errors clustered at the industry level. Significance: ***1 percent, **5 percent, and *10 percent levels. Specifications in all columns correspond to an specification analogous to regression 4, which is $Y_{isrt} = \beta_1(Post1_t \times Intense_s) + \beta_2(Post2_t \times Intense_s) + \lambda_t + \phi_i + Trends'_{srt}\theta + X'_{isrt}\psi + \epsilon_{isrt}$. The term $Trends_{srt}$ is the same as described in Section 4.2.

Table B.14: Transitions into Formal Entrepreneurship and Firm Survival in Non-Eligible States

Dependent Variable: Formal Entrepreneur (0/1)

	(1)	(2)	(3)	(4)
IMP1	-0.008	-0.030	-0.047	-0.122
	(0.005)*	(0.023)	(0.062)	(0.094)
Number of Individuals	7,886	1,814	168	1,085
R-squared	0.005	0.025	0.370	0.111
Transition from:	Formal Employee	Informal Employee	Unemployed	Formal Entrepreneur

Notes: Robust standard errors clustered at the industry level. Significance: ***1 percent, **5 percent, and *10 percent levels. Specifications in all columns correspond to an specification analogous to regression 4, which is $Y_{isrt} = \beta_1(Post1_t \times Intense_s) + \beta_2(Post2_t \times Intense_s) + \lambda_t + \phi_i + Trends'_{srt}\theta + X'_{isrt}\psi + \epsilon_{isrt}$. The term $Trends_{srt}$ is the same as described in Section 4.2. The sample is restricted as follows: (1) formal employee in the first interview; (2) informal employee in the first interview; (3) unemployed in the first interview; (4) formal entrepreneur in the first interview.

on all formal establishments in Brazil and their size (measured in number of formal employees). We use the same specification as in Table 2, namely: we use a balanced panel and exclude the industries that change their eligibility status (which correspond to 1.2% of our sample). Our dependent variable is the logarithm of the number of firms with

more than 5 formal employees. Table B.15 shows the results for all regions and for the 6 main metropolitan areas covered by the PME (columns 1 and 2, respectively). As the Table shows, we do not find any significant results from neither IMP1 nor IMP2, which reinforces the validity of our identification strategy.

Table B.15: Effects on the Number of Formal Firms with more than 5 Employees (RAIS)

Dependent Variable: Log (# Formal Firms)		
	All Regions	PME Regions
	(1)	(2)
IMP1	0.000	0.002
	(0.016)	(0.038)
IMP2	0.004	0.010
	(0.010)	(0.019)
Observations	500,472	111,216
R-squared	0.807	0.807
Sample	> 5 Employees	> 5 Employees

Notes: Robust standard errors clustered at the industry level. Significance: ***1 percent, **5 percent, and *10 percent levels. Specifications in both columns 1 and 2 correspond to regression 2, detailed in Section 4.1. Column 1 uses the full sample including all regions in the country, column 2 restricts the sample to the six main metropolitan regions in Brazil, which constitute the sample used by the Monthly Employment Survey (PME): Porto Alegre, Rio de Janeiro, Recife, Salvador, Belo Horizonte and São Paulo. IMP1 and IMP2 correspond to the estimated coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ in regression 2, respectively.