Size-based business taxation in a high-informality context

Cristiano C. Carvalho[†]

September, 2024

Click here to access the latest version

Abstract

This paper provides evidence on the effects of size-based taxation in a context of low enforcement and a large informal sector. I examine the effect of a Brazilian tax reform that applied to micro-businesses below a revenue cutoff. For affected companies, the reform replaced all business taxes with a low, fixed monthly fee. Evidence indicates that the intervention increased the number of formal firms by 43%. Because so many informal firms transitioned to the formal sector, total tax revenues increased, despite the tax cut that the reform offered firms that were already in the formal sector. The paper's model identifies sufficient statistics for welfare analysis of this type of reform, and its estimates suggest that the reform increased welfare, with compliance cost reduction—rather than tax reduction—being the primary driver of enrollment in the new regime.

[†]University of Michigan (crcarval@umich.edu)

1 Introduction:

Size-based regulation is a common feature of the tax regime in many countries. To alleviate the compliance cost and help business creation, governments usually either exempt or offer simpler taxation schemes for firms below a certain revenue cutoff. Even though these regulations create incentives for firms to stay below the threshold, the business creation fostered by a simpler tax alternative can more than compensate for size distortions. In developing countries, a large informal sector may exacerbate the importance of size-based taxation, as the possibility of small businesses exiting to informality makes the tax base more sensitive to tax liabilities. If the tax base elasticity is sufficiently high for this group, special tax regimes for small businesses can serve as an effective policy tool for enhancing formalization and increasing tax revenue. Despite its prevalence and increased significance in developing countries, there is a lack of evidence in the literature on the effects of size-based taxation in a context with high informality.

This paper measures the effects of a large-scale size-based taxation program implemented in Brazil on formal business creation, firm size distortion, and tax revenue. The Individual Micro-entrepreneurs (IME) program, introduced in 2009 by the Brazilian government, established a new tax regime designed to reduce compliance costs and monthly taxes for eligible firms, which were those with up to one employee and revenue below a maximum cap. Besides the maximum revenue cap, a key feature of the program is the replacement of all business taxes with a low fixed monthly tax. This policy further advanced the simplification agenda initiated by a prior size-based intervention known as *Simples*, which allowed firms to be taxed on their revenue rather than their profits. The IME fixed fee relieved participants of the burden of calculating their taxes monthly and the requirement to hire an accountant, which was mandatory under the *Simples*.

My identification strategy is based on the eligibility criteria for the program, which were determined using seven-digit industry codes. While multiple legal changes have influenced the industries permitted under the IME program over time, the initial set of eligible industries largely corresponded to a list defined by the *Simples* tax system, established over a decade earlier, providing credibility to its exogeneity. The IME program was integrated into industries where the *Simples* option was available and specifically targeted micro firms and self-employed individuals to enhance the formalization rate. Prior to its implementation, approximately 80% of entrepreneurs with up to one employee were unregistered with the tax authority and thus did not remit taxes in Brazil. I leverage the variation provided by this industry eligibility list and the timing of the program's implementation to conduct a difference-in-difference analysis focused on measuring the effects of tax simplification and reduction introduced by the IME intervention.

I provide a model to guide my empirical analysis, which also fills a gap in the public finance literature. While the fixed fee system has been implemented in several countries with low state capacity and discussed in policy reports (Engelschalk, 2007), there has been no formal analysis addressing the trade-offs of such a system or the sufficient statistics for welfare[1]. In my model, individuals decide whether to operate in the informal sector or choose between two formal tax options: (a) a presumptive output tax regime, or (b) a fixed fee tax regime, provided they report income below a specified cap. The government allocates tax revenue collected from the formal sector, net of administrative costs, to fund a public good, whereby total welfare is defined as the sum of the private utilities of entrepreneurs and the total utility they derive from the public good. I first demonstrate an important theoretical result: as long as the output taxation involves higher compliance costs, finding a fee and an income cap that improve welfare is always possible compared to a pure output tax system.

Second, I present two implications of my model that facilitate the interpretation of the empirical results. I first show that increasing tax revenue is a sufficient, but not necessary, condition for the program's implementation to be welfare-increasing. By revealed preference, entrepreneurs enroll in the IME program only if it optimizes their private utility. If the tax revenue, net of administrative costs, increases, greater welfare is generated through the provision of public goods, resulting in a Pareto improvement. Second, I derive sufficient statistics to optimally set a fixed fee tax regime, and identify two critical extensive margin responses. The optimal setting of parameters depends on their influence on the flow of informality into the IME and on how entrepreneurs already in the formal sector transition to the new program due to tax incentives. Specifically, since the fee does not distort the optimal income of IME participants, the primary behavioral response it generates is the movement between different sector options. In addition to the extensive margin behavioral response, the income cap also triggers some income distortion due to bunchers that under-report their true income in order to be eligible for the program.

I use a federal registry of all firms in Brazil as my primary data source to assess the effects of the IME program on the number of formal firms and the transitions of firms that would otherwise be formal into the IME program. This data, recently made available by the *Receita Federal* (RFB), the Brazilian tax authority, contains all firms with a tax identifier, including those registered as Individual Micro-Entrepreneurs. It provides crucial information on the dates of opening and closing, as well as the dates of entry into and exit from the IME program for each firm, enabling the construction of a panel of active firms. I aggregate this firm-level information to create a monthly panel detailing the total number of firms at the industry-by-state level from January 2006 to December 2017.

[1]The fixed fee system has been implemented in several countries, including Brazil, Ecuador, and a non-exhaustive list of seventeen countries in Sub-Saharan Africa (see Hoy et al. (2024)).

I estimate the main results using a difference-in-difference strategy that compares the growth in the number of formal firms across industries with different eligibility statuses. Analyzing the RFB data, I find that the implementation of the lower tax IME program increases the number of formal firms by an average of 43% over the sample period. This effect intensifies over time and is primarily observed among micro and sole proprietorship firms, as it should be by the program rules. Furthermore, although some firms strategically transition from other formal options to the IME program, the number of firms exiting the program more than doubles this figure, reaching approximately 80,000 in 2017. This helps explain that, in aggregate, my results indicate no evidence of strategic sorting into the IME program of firms that would be formal otherwise. I show that implementing the IME program does not significantly decrease the number of formal firms in the alternative tax system.

Because the administrative data does not capture non-registered businesses, i.e., informal firms, I use the Brazilian Monthly Labor Force Survey—*Pesquisa Mensal de Emprego (PME)*—to measure the flow into the program from the informal sector. I first aggregate the total number of formal and informal entrepreneurs for each industry-by-metropolitan region cell from January 2006 to December 2015 to estimate the overall effect of the program[2]. The results reveal contrasting effects for the formal and informal sectors: the number of formal self-employed individuals and employers increases by 14%, aligning with the administrative data pattern, while the number of informal self-employed individuals and employers decreases by 16.5% on average over the sample period. In a second analysis, I estimate the effects of the IME program on the share of formality using the survey microdata. The IME program significantly increases the share of formal entrepreneurs by 3.2 percentage points, representing an effect of 10% relative to the baseline formality share in the control group.

After analyzing the evidence indicating that the program effectively facilitated the transition of entrepreneurs from the informal to the formal sector without significantly switching them from other higher-tax formal options, the empirical findings reveal an increase in tax revenue following the program's implementation. This aligns with descriptive evidence examining tax revenue growth across different two-digit industries between 2009 and 2016. I find a positive correlation between tax revenue growth and the relevance of the IME sector within each industry in 2016, as measured by the share of firms that are IME participants. Specifically, for each one percentage point increase in the IME firm share within an industry, there is a corresponding 0.54 percent increase in total tax revenue. Notably, when excluding IME participants, the point estimate for this correlation remains positive but non-significant, indicating no evidence of strategic switching into the IME program from other formal tax regimes.

[2]Entrepreneurs are defined as self-employed individuals and employers with one employee, in accordance with the criteria established by the IME program.

I explore other potential margins that the program's introduction could influence and find no evidence that formal employees change their occupational status to be part of the program and also find no evidence that formal firms with more than 1 employee downsize to be eligible for the program. However, there is some evidence that the program reduces the number of informal employees and informal firms, suggesting that part of these groups may have transitioned into the program. I also test two additional heterogeneities. First, I find that the formalization results are stronger for middle-income self-employed and employers, consistent with them having higher tax incentives. Second, the formalization effects are stronger in industries where entrepreneurs sell more to the final consumer in the pre-period, meaning that the program reached small businesses not constrained by the production chain's self-enforcing mechanisms.

Interpreting the empirical results through the framework of my model leads to two key conclusions. First, the introduction of the program enhances welfare. Both the extensive margin findings and the descriptive tax revenue data indicate that the IME program boosted tax revenue. This serves as a sufficient, though not necessary, condition for the program's implementation to be welfare-improving, as entrepreneurs' private utility cannot decrease when they have access to more options. Second, I use the extensive margin estimates to examine the conditions that would justify varying IME tax rates and income eligibility thresholds. The sufficient statistics for the fee suggest that, under reasonable assumptions, increasing the program's fee toward the end of the sample period would lead to welfare improvements. Meanwhile, the sufficient statistics for the income cap indicate that, to justify the current cap, the average size distortion created by it should be at least three times higher than the distortion caused by the current output tax.

I also make parametric assumptions to fully characterize and estimate the model using the Simulated Method of Moments (SMM). With the estimated parameters, I determine the optimal IME tax and income cap, analyzing the roles of simplification and tax reduction in shaping the program's impact on formalization. The findings indicate that the optimal policy entails increasing the IME fee from R$36 to R$109 and lowering the IME income cap from R$5,000 to R$3243. This adjustment would reduce enrollment to approximately 44% compared to the levels observed in 2014, but would increase tax revenue by 10%. To assess the relative importance of simplification versus tax reduction, I simulate two counterfactual scenarios: one in which the IME incurs the same compliance cost as the alternative formal system and another in which it remits the same tax as that system. The results demonstrate that simplification has a more significant impact than tax reduction; specifically, eliminating simplification would lead to a 3 percentage point greater decrease in the IME share than removing the tax benefits in comparison to the alternative formal system.

This paper contributes to several strands of the literature. First, my findings expand the limited evidence on size-based taxation for small firms and self-employed individuals. Dual tax regimes are prevalent in many tax systems, as profit taxes typically impose higher compliance costs on firms and create greater administrative burdens for governments. As a result, countries often establish special tax treatments for small businesses operating below a revenue cutoff. Existing studies have explored size-based tax designs that provide value-added tax exemptions for small firms (Harju, Matikka, and Rauhanen, 2019), as well as the use of output taxes as an alternative to profit taxes for the self-employed (Akcigit et al., 2022) in developed countries, and found that individuals highly value and respond to the simplicity of these arrangements. I provide evidence on a new type of dual tax regime that establishes a simpler alternative to the turnover tax for self-employed. I show evidence of how the simplification brought by taxing small business with a set fee can increase business creation, formalization and tax revenue.

Second, this paper contributes to the theoretical literature on public finance that examines the implications of limited tax capacity for optimal taxation (Best et al., 2015; Kanbur and Keen, 2014; Keen and Slemrod, 2017; Dharmapala, Slemrod, and Wilson, 2011; Gordon and Li, 2009; Basri et al., 2021). Some countries with low state capacity have implemented fixed fee tax options, which are also discussed in the literature as presumptive tax measures to enhance compliance (Engelschalk, 2007). However, there has been no formal analysis addressing the trade-offs associated with such a system or the sufficient statistics for welfare. This paper addresses this gap by presenting a model that analyzes the trade-offs of implementing a fixed fee system alongside an output taxation system in contexts with high levels of informality. I provide three theoretical results. First, I demonstrate that as long as output taxation entails higher compliance costs than the fixed fee taxation, it is always possible to find a fee and income cap that improve welfare compared to a pure output system. Second, I show that an increase in tax revenue is a sufficient, but not necessary, condition to conclude the welfare impact of this type of intervention. Third, the sufficient statistics for defining the optimal parameters of the fixed fee tax regime depend crucially on the flow of individuals transitioning from both informality and the formal output system into this option.

Third, this paper adds to the developing literature that studies formalization programs a non-experimental evidence suggesting that reducing the ongoing costs associated with formality can be an effective policy for enhancing formalization and boosting tax revenue. Much of the existing research has focused on interventions that lower registration costs, which have shown only limited effectiveness (Ulyssea, 2020). The few studies that look at the effects of reducing tax liabilities find a more promising increase in formalization (De Mel, McKenzie, and Woodruff, 2013; Benhassine et al., 2018; Rocha, Ulyssea, and Rachter, 2017; Rocha and Farias, 2021; Hoy et al., 2024). Closer to this paper are Rocha, Ulyssea, and Rachter (2017) and

Rocha and Farias (2021), which empirically analyze the effects of the IME program. Rocha, Ulyssea, and Rachter (2017), the first to examine the introduction of the IME, relies on administrative data that does not require Individual Micro-Entrepreneurs to report unless they have employees, which constitutes only about 2% of IME participants[3]. My results address this gap by using novel data that covers all IME participants and align with those of a contemporaneous study by Rocha and Farias (2021), which also employs the same data. I further differentiate my empirical analysis in two key ways. First, I estimate the aggregate transition to the IME for firms that would otherwise fall under the alternative formal tax system, which is a crucial margin for fully characterizing the welfare impact of the IME program according to my model. Second, I support my findings by presenting evidence of the program's implementation on tax revenue, which is essential for inferring the welfare implications of its introduction.

Finally, this paper contributes to the broader empirical literature on how high levels of non-compliance constrain taxation capacity in developing countries. One empirical fact highlighted by the literature and often attributed to the lack of state capacity is that developing countries usually collect less tax revenue than developed countries. Low-income countries typically collect taxes of between 10 to 20 percent of GDP, while the average for high-income countries is around 40 percent (Besley and Persson, 2014). The weak state capacity increases the elasticity of the tax base to the tax rate, since non-compliance becomes the first-order behavioral response. This paper adds to the literature another example where non-compliance drives a negative correlation between tax revenue and tax liabilities. (Waseem, 2018; Bergeron, Tourek, and Weigel, 2024).

2 Institutional Background:

The Individual Micro-Entrepreneur (IME) was introduced in July 2009 as part of an agenda to improve the business environment for micro and small firms. According to Rocha, Ulyssea, and Rachter (2017), the program has 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. To accomplish these targets, the program reduces the formalization cost in different ways. First, it reduces the registration cost in monetary and nonmonetary terms. Entrepreneurs enroll online in the program and do not pay any fee. Second, the micro-entrepreneurs have to pay only a single monthly contribution consisting of a fixed-based minimum wage rate. Third, the pro-

[3]They use the universe set of formal contracts collected annually by the Brazilian Ministry of Labor in the *Relação Anual de Informações Sociais* (RAIS) dataset. This matched employer-employee data is only mandatory for firms with employees to report. At the time, this represented the most comprehensive administrative data available, as the RFB data was released only later.

gram eliminates all nonmonetary costs related to remaining formal. Besides complying with the monthly contribution, the only obligation is to inform the annual revenue at the end of the year.

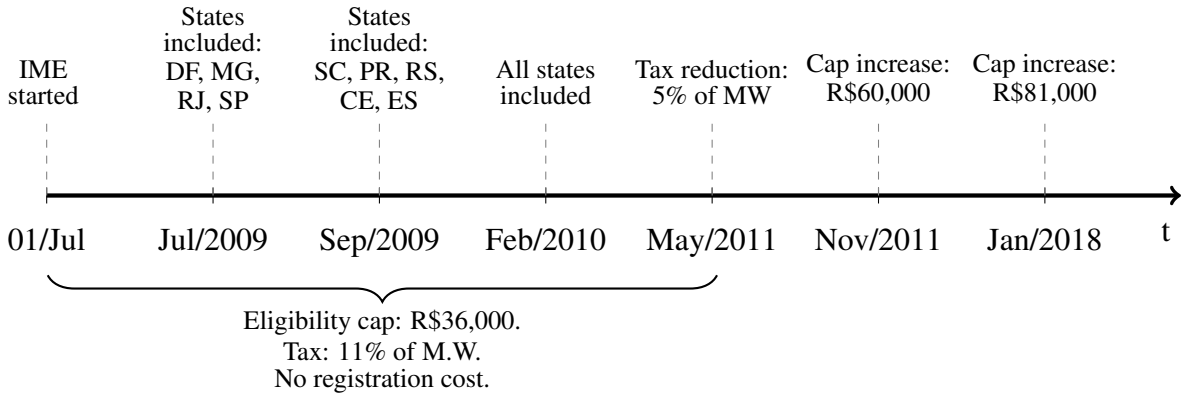
The new legislation improves on a previous tax intervention designed to reduce and simplify the taxes obligations of micro and small firms. The Simples system, created in 1996, unified many different federal taxes and mandatory contributions into a single and monthly based-rate payment and significantly decreased the tax liabilities. For example, micro firms have to pay a total rate from 3% to 5% (depending on the sector) under the Simples instead of paying from 5% to 11% of gross revenue and 20% of the payroll to social security. The eligibility for this program is established according to the firm's activity sector defined by the 7-digit industry level and included manufacturing, transportation, and other services not subject to specific regulation (Monteiro and Assunção, 2012). Even though the Simples system reduces and simplifies the tax costs, the firms are still required to comply with many bureaucratic procedures. More specifically, firms are required to have an accountant expert, as well as to pay some taxes that were not unified by the system. Additionally, this tax regime did not change the registration costs.

The eligibility to the IME program follows three different criteria: entrepreneurs can only register if they work in specific activity sectors, have at most one employee, and have annual revenue below a cap. To illustrate the program's extension, from the total of 1332 7-digit activity sectors, 417 different activities were allowed by the program at some point. In 2008, one year before the implementation, the eligible industries comprised around 70% of all self-employed workers in Brazil. Although 16 different law changes regulated the sectors allowed by the program over time, the initial set of eligible sectors relied on the ones defined by the Simples system in 1996. More specifically, some sectors were included and excluded, but 303 out of 417 were allowed by all legislation pieces.

Besides the activity sectors, other important program features also changed over time. As displayed in the timeline (figure 1), the annual revenue cap was initially set at R$36,000 per year (around US$18,000) and later expanded to R$60,000 in November 2011 and to R$81,000 in January 2018. The mandatory monthly contribution was first fixed at 11% of the minimum wage from July 2009 to April 2011 and then set at 5% from April 2011 onward. Another important variation source relies on the fact that the program's implementation time varied by the state due to technical problems in setting the program's online platform. The program started in July/2009, September/2009, and February/2010 for 4, 5, and 18 states, respectively.

Before the Individual Micro-Entrepreneur Program, the formal options of self-employment were: to open a micro firm in the Simples System or work as a formal self-employed in case the occupation was regulated (dentist, physicians, auditors, and others.).

Figure 1: IME timeline

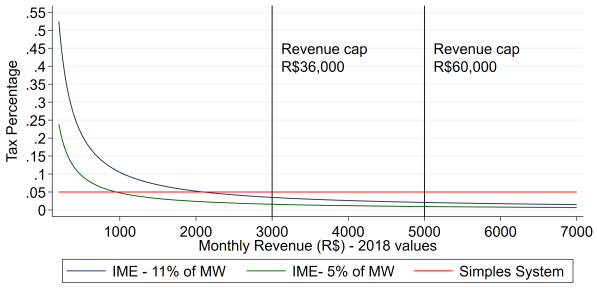


Strictly analyzing in terms of taxes, self-employment under the IME program is a more attractive option for high earning self-employed individuals as presented in figure 2a. Since the taxes under the program are defined as a fixed amount, entrepreneurs pay a lower tax percentage as their revenue increases, as long as it does not exceed the revenue cap. For all the different phases, the IME program makes the taxes even lower when compared to the Simples System for an extensive range of average monthly revenues. In contrast, low-earning individuals pay more under the IME program than under the Simples. However, they can still have incentives to enroll as an IME, considering that the complexity of tasks required for a self-employed to remain formal under Simples is not trivial. The Micro-Entrepreneur Program changed this by basically eliminating the compliance cost.

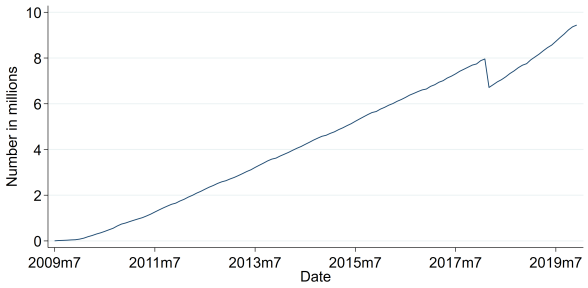
Since the implementation of the program, the number of Individual Micro-entrepreneurs has increased from 53,040 in Dec/2009 to 9,442,251 in Dec/2019, as presented in figure 2b. It is important to highlight that the IME program coexists with the Simples tax system so that entrepreneurs can choose which one is the best option for them. If the IME program's revenue cap is exceeded, the tax authority enforces the eligibility in two different ways. If the additional revenue is lower than 20% of the program's cap, the entrepreneur must switch to the Simples system in the following year and pay the Simples tax over the excess revenue. If the revenue exceeds more than 20% the cap, the entrepreneurs are immediately dis-enrolled from the IME program and have to retroactively pay Simples tax over the earned revenue of the entire year. In any case, the Simples tax system is the natural option for entrepreneurs that grow out of the IME program.

Figure 2: Size of the program and tax incentives

(a) Tax comparison between IME and Simples



(b) Number of IME



3 Data:

I use five different datasets in this paper. The primary dataset is the federal registry of the universe of all firms in Brazil that was recently made available by the *Receita Federal* (RFB), the Brazilian tax authority. All formal firms need to register with the tax authority to get a tax identifier, including the self-employed enrolled as an Individual Micro-Entrepreneurs. The new data provide information on the firm's legal structure, legal and fictitious name of the firm, tax identifier of the firm, date of opening, date of closing, when the firm entered and left the Simples tax system, and the IME program. This dataset does not provide direct revenue information, however it brings information if the firm is micro or small, defined based on some revenues cutoffs.

Although the RFB dataset is a cross-section of all firms in 2020, it is possible to retroactively build a panel of active firms relying on the date of opening and closing for each firm. I also use the Simples and IME entering and leaving date information to track each period the firms were part of the Simples system and the IME program. The final data present two important limitations. First, the cross-section structure do not allow me to retroactively create some specific variables, such as firms' size, since the recorded information refers to the moment of data collection (or last information if the firm is closed). Second, it is not possible to check if the firms were actually active in some specific period besides being registered. This concern could be somewhat mitigated considering that registered but inactive Individual Micro-Entrepreneurs would still need to pay their monthly tax contribution, having incentives to close their firms whenever they stop operating.

I collapse this firm-level data to create a monthly panel containing the total number of firms at the industry-by-state level from January 2006 to December 2017. The main outcome variable is the logarithm of the number of formal firms that will allow me to measure how the IME program increased the overall number of formal firms using the activity sector variation.

My final data contains 2,464,992 industry-state-month observations. I use the 5-digit activity sectors to define the treatment and, from the total number of 672 activity sectors, I keep 469 after excluding the industries that went into/out of the IME program to avoid selection bias and also excluding the sectors that were included in the Simples later. For more details, figures 11 and 12 present how the number of eligible sectors changed over the sample period in the Appendix.

I use a second administrative dataset as a baseline to assess how the RFB data change the measured results of the IME program. *Relação Anual de Informações Sociais* (RAIS) is the main source of formal labor market data in Brazil and the only option available to measure the IME program before the RFB data was released. This is an employer-employee administrative dataset assembled yearly by the Brazilian Ministry of Labor. It is a high-quality census of the Brazilian formal labor market that comprises detailed contractual information on 76.1 million contracts of registered workers and 3.9 million in registered firms. Each observation represents an employment contract between a firm and a worker and contains information on monthly wage, age, gender, race, education level, sector, occupation, the month of admission and separation, establishment size, and location.

Considering that RAIS is a matched employer-employee data, Individual Micro-Entrepreneurs are not required to report their information to the Ministry of Labor unless they employ someone else. Therefore, it is likely that RAIS underreport the true number IME individuals. I check it by using the contract level information presented in RAIS to create a monthly panel containing the total number of firms with at least one active contract at the industry-by-state level from January 2006 to December 2017. I use the same main outcome, logarithm of the number of formal firms, and follow the same procedure used in the RFB final data construction. My final data comprises 2,464,992 industry-state-month observations after excluding the industries that went into/out of the program

More than measuring the aggregate impact of the IME program on the total number of formal firms, it is also important to provide evidence on where the new firms are coming from. One important channel potentially consists on firms that were already formal switching to a lower tax regime provided by the new program. Besides precisely measuring the number of IME participants, another contribution of using the RFB data is the possibility of checking if there was any substitution between the IME program and the firms part of the Simples tax system. I use the collapsed industry-state-month RFB data to measure if the introduction of the IME program is also related with less firms being part of the alternative formal options.

Another important flow into the program is generated by firms coming from the informal sector. To measure this channel, I rely heavily on the strategy used by Rocha, Ulyssea, and Rachter (2017). Since by definition the administrative data does not allow me to observe infor-

mal firms, I use the Brazilian monthly labor force survey - Pesquisa Mensal de Emprego (PME) - as my third dataset. PME is conducted by the Brazilian Institute of Geography and Statistics and is designed as a rotating panel of individuals in the six largest metropolitan regions in Brazil. Households are interviewed in two waves of four consecutively months, with an one year distance between the beginning of the first and second waves. I first use the individual level observation to classify self-employed and employers into the formal and informal sectors. I use the same definition used in the literature and define as part of the formal sector everyone that contributes to the social security[4]. After classifying individuals into the formal and informal sectors, I create a panel with the total number of formal and informal self-employed and employers for each industry-by-metropolitan region cells from January 2006 to December 2015.

Focusing on the aggregate measure instead of on the individual flows allows me to capture a long-term measurement of the effects of the IME program. More precisely, it is possible that, in the long run, individuals would primarily sort themselves as an Individual Micro-Entrepreneur if the program succeeds. In this case, an individual measure of the flow from the informal sector to the IME program would not capture the entire effect of the intervention considering that, after the program, individuals would have the option to sort into their preferred option without the need to make any transition. In contrast, this will be captured by an aggregate measure of the total number of formal individuals. To get a sense that the increase in the number of IME participants is mainly driven by formalization, I check how the program's implementation affects the total number of formal and informal self-employed and employers. If this channel is relevant, the program should increase the total number of formal individuals and decrease the number of informal individuals in the sectors allowed by the program. I also work with the micro-level data to measure what happens with the probability of being formal conditional on being an entrepreuner for industries allowed and not allowed by the program.

One limitation with the PME data is that the information about the individuals' activity sectors is recorded at the 2-digit level. At the same time, the eligibility for the program is defined at the 7-digit level, a much more disaggregated measure. I follow a similar strategy used by Rocha, Ulyssea, and Rachter (2017) to determine the eligibility in the PME sectors' measure using a fourth dataset. The Brazil National Household Sample Survey (PNAD) is an annual household cross-section representative of the entire country that contains a more disaggregated activity sector measure, a 5-digit level. I use the sample from 2002 to 2008 (before implementing the IME legislation) to measure the share of self-employed and employers in the eligible 5-digit industries within each 2-digit sector. I define a 2-digit sector as treated if this share is greater than 0. The list of treated and control industries is presented in tables 22 and

[4]By legislation, a formal self-employed or employer are subject to mandatory contributions to the social security system.

23.

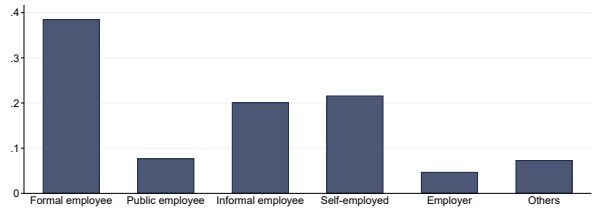
To validate the results estimated on the extensive margin long-term flows, I use an additional RFB dataset that measures the aggregated tax remittances for the 85 distinct 2-digit industries separately in 2009 and 2016. This dataset provides information regarding the total number of firms and the total tax revenues collected for various formal tax regimes, including the IME program. The information disaggregated by the different formal options allows me to measure the importance of the IME program both for the growth of the number of firms and of the tax revenue for each industry. It also allows me to assess what happens to the growth of the number of firms and of the tax revenue in the alternative formal tax categories, which works as an additional test if firms are strategically transitioning to the lower tax regime offered by the new program.

4 Descriptive Statistics:

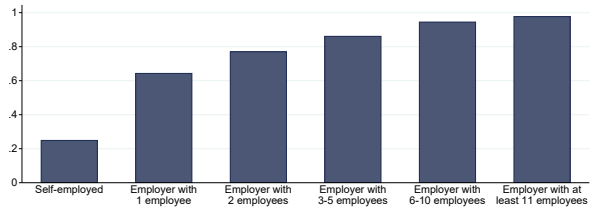
Before the introduction of the Individual Micro-Entrepreneur (IME) Program, individuals had a few formal options: becoming an employee, starting a business, or working as a formally self-employed person. In addition to these formal choices, individuals could also work informally as an employee, self-employed person, or employer. As illustrated in Figure 3a, by 2009, self-employed individuals made up more than 20% of the workforce, with self-employed and employers together constituting about 25% of the total workforce. Figure 3b shows that informality was particularly prevalent among entrepreneurs targeted by the IME program, with around 80% of self-employed individuals and about 40% of employers with one employee operating informally. This high level of informality is also seen in other groups of small employers, and it tends to decrease with the size of the firm.

Figure 3: Share of the workforce by employment type:

(a) Employment type:



(b) Share of formal self-employed and employers by size:



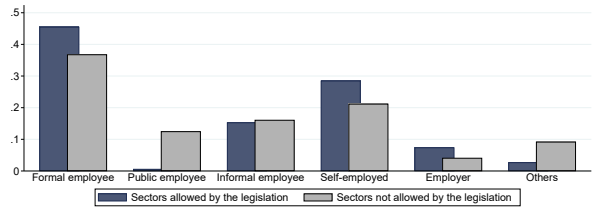
Source: PNAD 2009. Household survey.

Figure 4 compares employment types between industries eligible for the IME program and those that are not, at the 5-digit level in the PNAD. The workforce composition varies between these two groups: sectors not covered by the program tend to have a higher proportion

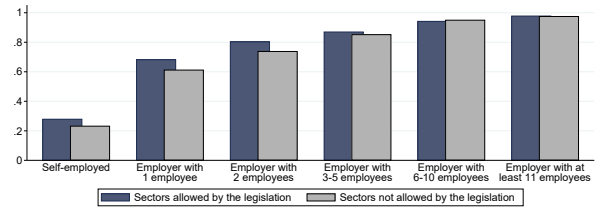
of public employees and a lower proportion of self-employed individuals. Conversely, the level of formality among self-employed individuals and small employers is similar across different activity sectors, though there is a slight increase in informality in sectors excluded from the IME legislation.

Figure 4: Share of the workforce by employment type - IME and non-IME sectors:

(a) Employment type:



(b) Share of formal self-employed and employers by size:



Source: PNAD 2009. Household survey.

5 Model:

Consider that entrepreneurs choose how much output Y_s to produce, how much of earnings to evade, e_s, and the sector s in which they want to operate — formal (Simples or IME) and informal. The agents face a perfectly elastic demand and sell at a price normalized to one. Entrepreneurs have a quasilinear utility function that depends on their profits, $\Pi_{i,s}$, the government's total expenditure on public goods, g, and some idiosyncratic preference for the sector, $\xi_{i,s}$. Each individual has a type a_i that defines her productivity. The cost of increasing total income is strictly convex due to the disutility generated by labor effort - $\psi_Y(Y_s, a)$ and $\psi_{YY}(Y_s, a)$ are positive - but high-ability individuals incur a lower total and marginal cost of generating income, such that $\psi_a(Y_s, a)$ and $\psi_{Ya}(Y_s, a)$ are negative. Evading taxes is also costly, and the evasion cost function, $c(e, \alpha)$, is a strictly convex and increasing function of the amount being evaded, e, and of a government enforcement level, α. Government enforcement also increases the marginal cost of evasion, such that $c_{ea}(e, \alpha) > 0$. I assume that both cost functions are the same for all the sector options, but both the tax function, T_s, and the compliance cost, θ_s, are sector-specific. Entrepreneurs also receive utility from the public good, which is an increasing and strictly concave function, $v_g(g) > 0$ and $v_{gg}(g) < 0$, and only depends on the amount of government expenditure g.

$$U_{i,s} = \underbrace{Y_s - T_s - \psi(Y_s, a_i) - c(e_s, \alpha) - \theta_s}_{\Pi_{i,s}} + v(g) + \xi_{i,s}$$

Simples option: Self-employed individuals remit output taxes at a rate t_o on their reported output, such that $T_{Sim} = t_o(Y_{Sim} - e_{Sim})$, and have a positive compliance cost, considering that they have to comply with many bureaucratic procedures, such as hiring an accountant expert or spending time and mental effort in complying with all the rules[5]. I assume formality as a non-binary dimension and, as such, individuals registered in the Simples option choose their income and evasion level separately and are allowed to evade some taxes by concealing part of their actual income[6]. The first-order conditions are characterized by:

$$Y: \quad \psi_Y(Y^*_{Sim}, a_i) = 1 - t_o$$
$$e: \quad c_e(e^*_{Sim}, \alpha) = t_o$$

Entrepreneurs in the Simples regime keep evading until their marginal cost of concealing equals the benefit of not paying the output tax; and they keep working until their marginal disutility of working equals their marginal after-tax income[7].

IME option: Self-employed individuals' only responsibilities in this option are to report their annual income at the end of every fiscal year and to remit a monthly fee, τ, such that $T_{Ime} = \tau$. Considering that the fee value is always the same irrespective of the output, I assume the compliance cost to be zero, $\theta_{Ime} = 0$, considering that there is no need to hire an accountant, and the time effort to comply is insignificant[8]. To be eligible for this option, entrepreneurs must report their income below a cap Y_C. In case their actual income is above the cap, individuals can evade the excess income, $e_{Ime} = Y_{Ime} - Y_C$, to become eligible. The optimal conditions are characterized by:

$$\psi_Y(Y^*_{Ime}, a_i) = 1 \qquad \text{if } Y^*_{Ime} \leq Y_C$$
$$\psi_Y(Y^*_{Ime}, a_i) = 1 - c_e(Y^*_{Ime} - Y_C, \alpha) \qquad \text{if } Y^*_{Ime} > Y_C$$

Note that if $Y^*_{Ime} > Y_C$, the evasion cost function incorporates the cost of the evasion necessary to be eligible for the program, capturing the idea that individuals could under-report their true income to be eligible for this option.

[5] I follow Kanbur and Keen (2014) in the structure of the compliance cost.

[6] I follow Keen and Slemrod (2017) and consider formality as a non-binary dimension using the same properties of the evasion cost used in their paper. I also follow the literature and assume that the output tax influences the amount being evaded. This result is subject to the penalty functional form.

[7] Note that in case of a corner solution for a low enough ability level, such that $e^*_{Sim} \geq Y^*_{Sim}$, the entrepreneur does not remit any taxes, but still have to pay the compliance cost. Also, for simplicity, as in Keen and Slemrod (2017), labor supply is independent of the concealment costs. This assumption is relaxed for the IME and informal options.

[8] The assumption of zero compliance cost is with no loss of generality. The main results would still be valid assuming a compliance cost lower than the one in the Simples option.

Informal option: Informal self-employed individuals are "invisible" to the tax authority. Considering they do not formally register their firms in the tax system, they do not comply with any tax obligations by definition, such that $T_{Inf} = \theta_{Inf} = 0$. As such, they do not remit any taxes, do not face any compliance costs, and evade their entire income, $e_{Inf} = Y_{Inf}$. The income first-order condition in this option is defined as:

$$\psi_Y(Y^*_{Inf}, a) = 1 - c_e(Y^*_{Inf}, \alpha)$$

Although informal entrepreneurs remit no taxes, they incur a higher marginal evasion cost for every additional income they generate. Therefore, they keep working until their marginal disutility of working equals their marginal after-evasion cost income.

5.1 Sector choice:

I assume that entrepreneurs compute their utilities in each different option and choose option s if

$$U_{i,s}(Y^*_s, e^*_s) \geq U_{i,s'}(Y^*_{s'}, e^*_{s'}) \quad \forall s' \neq s$$

The model incorporates two different sources of sector choice variation. First, entrepreneurs have different ability levels that come from an exogenous distribution $F(a)$ with support on (0,∞). Higher-ability individuals tend to sort into the formal options, especially in the Simples tax system, because it is cheaper to remit the tax than to incur the evasion cost when they are high-income. Second, the model also assumes that individuals have different sector idiosyncratic preferences that come from a smooth exogenous distribution F(ξ). This additional source of heterogeneity makes the model more flexible by allowing individuals with the same ability to have different sector choices, which permits the model to replicate the empirical overlapping in the income distribution of the different sectors. The idiosyncratic preferences can represent different mechanisms that affect the sector choices but are not explicitly modeled, such as different risk tolerance, different individual levels of attention, and different preferences for the type of business organization.[9] [10]

To show the sector choice patterns, I start by keeping the idiosyncratic preferences fixed and considering the incentives for entrepreneurs with the same set of sector-specific preferences.

[9] The IME sector does not allow entrepreneurs to have a business partner.

[10] This is a non-exhaustive list of potential reasons, and identifying each of them is beyond the scope of this paper.

Proposition 1: *Without the IME option, there is an ability level, $a_{I,S}(t_0, \alpha) \geq 0$, such that every entrepreneur with ability equal or above it sorts into the Simples option.*

Proof: Appendix.

Intuitively, even though higher-ability individuals have a lower marginal cost of generating income in both sectors, this is dominated by a higher marginal cost of evasion in case individuals sort into the informal sector because, by definition, they evade all their income. The informal sector arises endogenously as a consequence of the taxes and compliance costs.

Proposition 2: *Including the IME option implies:*

a) Regardless of the other government instruments, whenever the IME income cap is positive, $Y_C > 0$, there is an ability cutoff level, $a_{I,Ime}(\tau, \alpha, Y_C)$, such that every entrepreneur with ability equal or above it prefers the IME over the informal option.

b) Suppose there is an ability level, a_l, such that it is the lowest ability in which the IME is preferred over the Simples, then there is also an ability cutoff, $a_{Ime,S}(t_o, \tau, \alpha, Y_C) \geq a_l$, such that every entrepreneur with ability between a_l and $a_{Ime,S}(t_o, \tau, \alpha, Y_C)$ prefers the IME over the Simples option, and every entrepreneur with ability above $a_{Ime,S}(t_o, \tau, \alpha, Y_C)$ prefers the Simples over the IME. Note that if $\tau \leq \theta$, then $a_l = 0$.[11]

Proof: Appendix.

Intuitively, the utility as an IME is lower than the utility in the informal sector for low levels of ability due to the fee format of the tax, but it increases faster with ability given that the marginal cost of evasion is lower even when the entrepreneurs evade to be eligible for the program since it is only incurred in the income that exceeds the IME income cap. Regarding the two formal options, for sufficiently low values of the lump-sum comparatively to the compliance cost, individuals would prefer the IME to the Simples option. However, as ability grows, the cost of the evasion needed to still be eligible for the program would get so pronounced that the Simples would be a preferred option.

Proposition 3: *In case all sectors appear in equilibrium for some IME combination of parameters, there are always two cutoffs $a_{I,Ime}$ and $a_{Ime,S}$, such that entrepreneurs prefer:*

i) The informal option if $a \in (0, a_{I,Ime})$,

ii) The Simples option if $a \in [a_{Ime,S}, ...)$.

Proof: Appendix.

The model estimation section illustrates Proposition 3. Note that the propositions were

[11] Note that the existence of the ability cutoffs between the two options of the formal sector depends on the parameters of the IME option. Without parameters restrictions, it is possible that the Simples option will always be preferred over the IME (Example: low Y_C and high τ). The opposite, however, is not possible.

proven while the idiosyncratic preferences were fixed. Ultimately, each set of sector-specific preferences would potentially generate a different set of ability cutoffs that keeps the sorting presented in Proposition 3. The implication of this Proposition is that the informal, IME and Simples participants are loosely concentrated in the low, middle, and high ability positions, respectively.

5.2 Welfare Analysis:

Government tax revenue provides the public good and covers the administrative cost involved in enforcing tax legislation. Because this is the empirically relevant formulation, I assume that all the different types of self-employed are observed in equilibrium, as in Proposition 3, after the IME program is implemented.

Define the ability and idiosyncratic draws as $\boldsymbol{\gamma}$ and the set of best options as $B_s = \{\boldsymbol{\gamma} : U_{i,s}(Y_s^*(\boldsymbol{\gamma}), e_s^*(\boldsymbol{\gamma})) \geq U_{i,s'}(Y_{s'}^*(\boldsymbol{\gamma}), e_{s'}^*(\boldsymbol{\gamma})) \quad \forall s' \neq s\}$, where $s \in \{Inf, Ime, Sim\}$.

The government budget constraint is given by:

$$g + A(\alpha) = \int_{\boldsymbol{\gamma} \in B_{Ime}} \tau f(\boldsymbol{\gamma}) d\boldsymbol{\gamma} + \int_{\boldsymbol{\gamma} \in B_{Sim}} t_o(Y_{Sim}^* - e_{Sim}^*) f(\boldsymbol{\gamma}) d\boldsymbol{\gamma}$$

where $A(\alpha)$ is the administration cost that the government incurs to monitor individuals irrespective of their sectors[12] and g is the expenditure in the public good.

Social welfare consists of the sum of the utilities of entrepreneurs, sorted into the different sectors, and the utility generated by the public good:

$$W(t_o, \alpha, \tau, Y_C) = \underbrace{\int_{\boldsymbol{\gamma} \in B_{Inf}} U(Y_{Inf}^*; \alpha, \boldsymbol{\gamma}) f(\boldsymbol{\gamma}) d\boldsymbol{\gamma}}_{\text{Informal entrepreneurs}} + \underbrace{\int_{\boldsymbol{\gamma} \in B_{Ime}} U(Y_{Ime}^*; \tau, Y_C, \alpha, \boldsymbol{\gamma}) f(\boldsymbol{\gamma}) d\boldsymbol{\gamma}}_{\text{IME participants}}$$
$$+ \underbrace{\int_{\boldsymbol{\gamma} \in B_{Sim}} U(Y_{Sim}^*; t_o, \alpha, \boldsymbol{\gamma}) f(\boldsymbol{\gamma}) d\boldsymbol{\gamma}}_{\text{Entrepreneurs part of the Simples tax system}} + \underbrace{v(g)}_{\text{Value of public funds}}$$

Proposition 4: *Increasing the tax revenue is a sufficient, but not necessary, condition for the implementation of the IME program to be welfare-increasing.*

By revealed preference, entrepreneurs enroll in the IME program only if it is a utility-maximizing decision. Therefore, including an additional option can never decrease the sum of their private utility. The final effect on welfare depends on the provision of the public good,

[12] This assumption can be relaxed to make enforcement a complement of the program's instruments

which is a direct function of tax revenue. If the tax revenue net of administrative costs increases, this is sufficient to conclude that the program introduction increases welfare. In case the tax revenue decreases, the final welfare change will depend on which effect dominates.

The tax revenue change can be translated into the extensive margin effects of the program. In case all new program participants come from the informal sector, the new option certainly increases welfare because it can only increase the tax revenue[13]. When the number of participants that come from the Simples option increases, the effect on welfare is indeterminate. Even though IME participants save on the compliance cost and have their optimal income less distorted by the tax format, they usually remit less taxes than they would under the Simples option[14].

Proposition 5: *Given that output taxation involves a positive compliance cost, there will always be an IME implementation that improves welfare*[15].

Proof: Appendix.

As stated in Proposition 1, each combination of sector-specific preferences includes an ability cutoff that distinguishes individuals in the informal sector from those in the output taxation system prior to the IME implementation. The income cap and fee can be utilized to ensure that the IME fee surpasses the output tax remittances of the marginal switcher, while also positioning this individual as close as possible to the ability cutoff. Once the cap is set, this can be accomplished by establishing the IME fee slightly below the total of the output taxes and compliance costs. This approach guarantees that any inflow into the IME program will boost tax revenue and, consequently, enhance welfare, as discussed in Proposition 4.

I hold the other parameters constant and focus on the optimal conditions of the IME policy instruments - IME tax (τ) and income eligibility cap (Y_C). By deriving the welfare function by the program's parameters and using the envelope theorem, the following sufficient statistics measure the program's welfare[16]:

$$\frac{\partial W(\tau,\alpha,t_o,Y_C)}{\partial \tau} = \underbrace{[v'(g)-1]P(Ime)}_{\substack{\text{Net welfare of a higher}\\ \text{tax on IME participants}\\ >0}} + v'(g)[\underbrace{\int_a (\frac{\partial P(Sim|a)}{\partial \tau}}_{\substack{\text{Flow from IME}\\ \text{into Simples}\\ \text{per ability}\\ >0}} \underbrace{(t_o(Y^*_{Sim}-e^*_{Sim})-\tau)}_{\substack{\text{Net revenue of switchers}\\ Undetermined}} - \underbrace{\frac{\partial P(Inf|a)}{\partial \tau}}_{\substack{\text{Flow from IME}\\ \text{into Informal}\\ \text{per ability}\\ >0}} \tau) f(a)da] \quad (1)$$

[13] Assuming that the administrative cost remains the same.

[14] This intuition only applies when considering the government decisions about the program's instruments Y_C and τ given. The first-order conditions will discuss what are the parameters' optimal choice.

[15] This paper assumes a fixed output system and does not address scenarios where the fee taxation completely replaces the output taxation system.

[16] Note that the income eligibility cap cannot be equal to 0 for the first-order condition with respect to the IME tax to have some meaning

$$\frac{\partial W(\tau,\alpha,t_o,Y_C)}{\partial Y_C} = \underbrace{\int_a P(Ime, Y^*_{Ime} > Y_C|a) c_e(Y^*_{Ime} - Y_C, \alpha) f(a) da}_{\substack{\text{Marginal cost of evasion for} \\ \text{IME under reporting participants} \\ >0}} + v'(g)[\underbrace{\int_a (\frac{\partial P(Sim|a)}{\partial Y_C}}_{\substack{\text{Flow from IME} \\ \text{into Simples} \\ \text{per ability} \\ <0}} \underbrace{(t_o(Y^*_{Sim} - e^*_{Sim}) - \tau)}_{\substack{\text{Net revenue of switchers} \\ Undetermined}} - \underbrace{\frac{\partial P(Inf|a)}{\partial Y_C}}_{\substack{\text{Flow from IME} \\ \text{into Informal} \\ \text{per ability} \\ <0}} \tau) f(a) da] \tag{2}$$

The first-order condition regarding the IME tax is outlined in Equation 1. The first term demonstrates that increasing the IME fee mechanically results in a higher amount collected from IME participants to fund public goods. Given that the marginal value of the public good exceeds 1, this implies a positive net welfare effect of a higher IME fee for the participants. While a higher IME fee does not affect entrepreneurs' optimal decisions on the intensive margin, it does lead them to exit the program, shifting to the informal sector or the Simples tax system, as indicated by $\frac{\partial P(Inf|a)}{\partial \tau}$ and $\frac{\partial P(Sim|a)}{\partial \tau}$ being positive. The second term accounts for the tax implications of these extensive margin behavioral responses. An optimal fee would balance the positive net welfare gain for IME participants and the increased tax revenue from those transitioning to the Simples program against the revenue loss from individuals moving to the informal sector.

Proposition 6: *Exemption is dominated by an option with a positive IME fee.*

Exemption is a commonly studied policy alternative in the literature. The current model can simulate an exemption by considering an IME option with a positive income cap, $Y_C > 0$, and zero IME fee, $\tau = 0$. In this context, the government has strong incentives to raise the IME tax for two reasons. First, with a zero fee, the IME program achieves maximum enrollment at each income cap, leading to the highest possible mechanical tax collection from increasing the IME tax. Second, the behavioral response, reflected in the extensive margin movement out of the program, can only increase tax revenue. Specifically, directing individuals to the Simples option increases tax revenue, while pushing individuals into informality has no impact.[17].

The first-order condition regarding the IME income cap is outlined in Equation 2. The first term indicates that raising the income eligibility cap reduces the marginal cost of evasion for any IME participant who underreports income to qualify for the program. This results in reduced evasion and increased real income for these participants[18]. The second term addresses the tax implications of the behavioral responses at the extensive margin. A higher income cap leads to an increase in enrollment in the IME program, as $\frac{\partial P(Inf|a)}{\partial Y_C}$ and $\frac{\partial P(Sim|a)}{\partial Y_C}$ are negative. Since increasing the cap typically attracts individuals with incomes higher than the cap in their

[17] If $\tau = 0$, $\frac{\partial W(\tau,\alpha,t_o,Y_C)}{\partial \tau} = [v'(g) - 1]P(Ime) + v'(g)[\int_a (\frac{\partial P(Sim|a)}{\partial \tau}(t_o(Y^*_{Sim} - e^*_{Sim}))f(a)da] > 0$.

[18] The actual responses would be even more pronounced if the model did not permit evasion to achieve eligibility.

alternative system, the net tax revenue effect for those switching from Simples to IME tends to be negative as the income cap rises. Additionally, because high-income individuals are generally more concentrated in the Simples system than in the informal sector, raising the cap will draw more entrepreneurs from Simples into the IME program, relative to those moving from the informal sector.

Even in the absence of an IME fee, $\tau = 0$, it might still be optimal to implement a positive income eligibility cap. This policy effectively functions as a tax exemption for self-employed individuals who report income below the specified cap. In this scenario, the net tax revenue from switchers will always be negative, as these individuals do not remit taxes if their reported income falls below the eligibility threshold. Without an IME tax, the share of individuals participating in the IME program is maximized for a given income cap. A lower income cap increases the benefit of reducing the marginal cost of evasion. Thus, even if the net tax revenue from switchers is negative, the government may find it advantageous to set a lower income cap to reduce distortions in real income caused by evasion costs.

The model formulation highlights the crucial importance of evaluating how changes in the program's parameters impact decisions at the extensive margin. This is the goal of the next section.

6 Empirical Estimation of the Aggregate Effect of the IME Program:

6.1 Identification Strategy

I leverage the implementation of the IME program and the exogeneity of the industries covered by the new legislation to estimate a difference-in-differences model. The results are analyzed by industry, geographic region, and month/year level. I use the administrative data to measure the effect of the IME program on the total number of formal firms, and analyze the PME survey to measure the effect on the total number of formal and informal entrepreneurs[19]. I also examine the PME individual microdata to assess the probability of entrepreneurs operating formally. More precisely, I assume that:

$$y_{g,j,t} = \delta_1 \cdot Post1_t \cdot Treated_j + \delta_2 \cdot Post2_t \cdot Treated_j + \lambda_3 \cdot Treated_j \cdot \mu_g + \mu_g + \mu_j + \mu_t + \varepsilon_{g,j,t} \tag{3}$$

where $y_{g,j,t}$ is the logarithm of the number of entrepreneurs or a formal dummy (when an-

[19] I define entrepreneurs as self-employed individuals and employers with one employee, in accordance with the eligibility criteria of the IME program.

alyzing the survey microdata) within the geographic region g, industry j, and at month t. $Treated_j$ is a dummy variable that equals 1 if the industry was eligible by the IME program. $Post1_t \cdot Treated_j$ and $Post2_t \cdot Treated_j$ are the difference in difference coefficients estimated separately for the periods when the program had an 11% and 5% tax rate over the minimum wage, respectively. μ_g, μ_j and μ_t are geographic region, industry, and month/year fixed effects. The model also allows for different geographic region fixed effects for treated and control industries. For the results using the administrative data, I de-trend the outcome variables by estimating different linear trends in the pre-period by groups and subtracting the prediction value from the real outcome[20].

6.2 Estimation of the Effects on the Total Number of Formal Firms:

The main results of the regression are estimated using the logarithm of the total number of firms as the dependent variable and all Brazilian states as the geographic region. By focusing on the total number of firms, I aim to address concerns about potential biases from firms artificially adjusting their revenue size to qualify for the program, as such reclassifications would not impact the overall count of firms. As presented in table 1, I find that the introduction of the IME program is associated with an increase of 43% in the total number of formal firms. By decomposing the effect into the two different phases of the program, I show that, as expected, a lower tax of 5% is associated with a higher increase in the total number of firms. More specifically, the 11% tax phase of the program is related to an increase of 11.2% in the total number of firms, and the 5% tax phase is related to a 51.7% increase in the total number of firms. The decrease in the IME tax rate from 11% to 5% of the minimum wage leads to a 40.5% increase in the number of formal firms.

I estimate the same regression model using the RAIS data and present the results in columns 3 and 4 of table 1. The same pattern is found, although the estimated effects of the program are much lower. More precisely, the IME program is estimated to cause the total number of formal firms to increase by 6%. The first and the second phases of the program are associated with an increase of 2.9% and 6.9%, respectively. Recall that the results reflect one important difference between the two datasets: a firm is only counted in RAIS as long as it has some contract with at least one employee. Therefore, ignoring the firms that the owner is self-employed would bias down the true effect of the program. This indicates that the majority of formal business creation occurs among purely self-employed individuals, which aligns with the fact that around 98% of the program's participants do not have any employees.

To be more precise about the effects of the IME program over time, instead of estimating

[20]The results without controlling for pre-trends are presented in the appendix.

Table 1: IME effects on the total number of firms

	Dep. Var: log(Total number of firms)			
	IRS data		RAIS data	
Treatment	0.4309***		0.0603*	
	(0.0505)		(0.0352)	
Treatment1		0.1121***		0.0295
		(0.0288)		(0.0189)
Treatment2		0.5177***		0.0687*
		(0.0571)		(0.0400)
Observations	1484707	1484707	1309612	1309612

Note: Significance levels: * 10%, ** 5%, ***1%. The results are clustered by industries. All regressions include state, industry, and month/year fixed effects. It also allows for different state fixed effects for treated and control industries by including their interactions. $Treatment1$ is the coefficient that captures the effect of the first phase of the program when the IME tax was 11% of the minimum wage. $Treatment2$ captures the effect for the period when the tax was set as 5% of the minimum wage. All results are estimated using the same time span: from Jan/2006 to Dec/2017.

the results for the two different tax rate phases, I estimate the month-by-month effects of the program since its implementation. Figure 5a shows that the results estimated using the IRF data are quite large. As an example, the total number of firms increased 68% more in the IME-eligibile industries compared to the industries not permitted by the program in the last period in my analysis, December 2017. Notably, the effects of the program are smoothly increasing over time; no sharp effect is detected after the implementation of a lower tax and the revenue cap expansion, which implies that measuring the program's impact using a narrow time window could lead to underestimation.

The results estimated using RAIS are plotted in figure 5b. Even though most of the point estimates are not significant, it is possible to see an increase in the number of firms after implementing the program. The estimated effect smoothly increases and stays constant at around 4% from 2015 onward. The comparison between both graphs shows that ignoring the self-employed owners would underestimate the impact of the program and its dynamics after 2015.

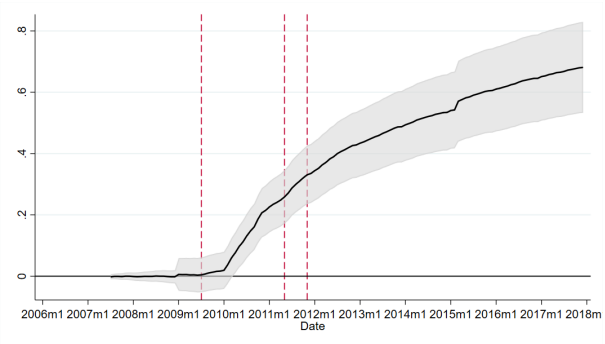
6.3 Robustness Estimation of the Effects on the Total Number of Formal Firms:

To make sure that the effect comes from the program, I rely on the fact that all firms that are part of the IME program are required by the legislation to be micro firms, have sole proprietorship as their business legal structure, and be part of the Simples[21]. If the program

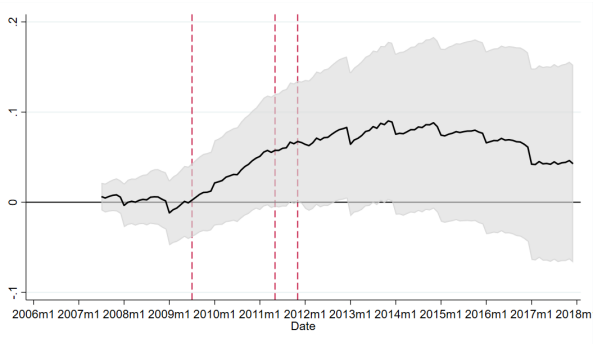
[21]Micro and small firms are defined by legislation as those with gross revenues below R$360,000 and between R$360,000 and R$4.8 million per year, respectively. Considering the IME income cap of R$60,000, all the IME are classified as micro firms. Additionally, the legislation stipulates that IME participants cannot have any business partners, automatically designating them as sole proprietorships. The size and business structure criteria enable effective targeting of IME effects, while also providing definitions applicable to the industries in the control group.

Figure 5: Effects of the program in the total number of firms over the sample period

(a) IRF data:



(b) RAIS data:



Note: The results are clustered by industries. All regressions include state, industry, and month/year fixed effects. It also allows for different state fixed effects for treated and control industries by including their interactions. All results are estimated using the same time span: from Jan/2006 to Dec/2017. The confidence interval is generated using a 5% significance level. The first vertical line indicates the program's introduction in 07/2009. All states were included only in 02/2010. The second vertical line represents the tax reduction from 11% to 5% of the minimum wage. The third line shows the period when the income cap was increased from R$36,000 to R$60,000 a year.

is driving the increase in the total number of firms, the effects should come only from the firms framed as micro and with sole proprietorship as their business legal structure, which is precisely what is found in table 2. More specifically, the program increased the total number of formal firms by 70.5% and 51.7% when I focus on firms with sole proprietorship and firms classified as micro, respectively. Reassuringly, no effects are found when analyzing the total number of firms with other legal structures, and with small size, which alleviates concerns about industry-specific shocks coinciding with the introduction of the IME program.

Table 2: IME effects on the total number of firms by legal structure and size

Dep. Var	Legal structure		Revenue size	
	log(Sole proprietorship)	log(Other Legal structures)	log(Micro)	log(Small)
Treatment	0.7058***	0.0012	0.5172***	0.0194
	(0.0622)	(0.0323)	(0.0465)	(0.0215)
Observations	1287155	1443821	1410611	1048457

Note: All regressions include state, industry, and month/year fixed effects. It also allows for different state fixed effects for treated and control industries by including their interactions. $Treatment1$ is the coefficient that captures the effect of the first phase of the program when the IME tax was 11% of the minimum wage. $Treatment2$ captures the effect for the period when the tax was set as 5% of the minimum wage. All results are estimated using the same time span: from Jan/2006 to Dec/2017.

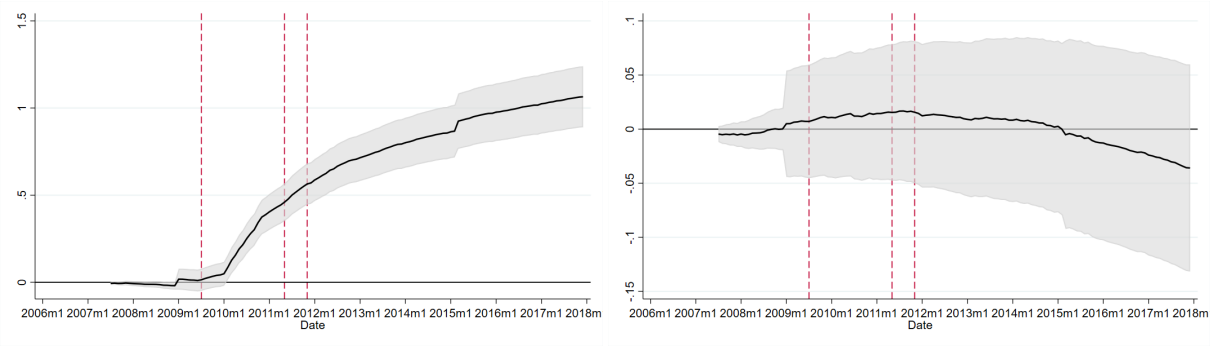
The month-by-month estimates for the different classifications of firms are presented in figure 6. As observed in the main result, the program's effects increase smoothly and are only observed for firms that meet the program's criteria: micro and sole proprietorships. To illustrate, in the last period analyzed, December 2017, the estimated effects are the highest, showing increases of approximately 106% for sole proprietorships and 79% for firms classified as micro. Consistent with aggregated results, no significant effect is found by analyzing the total number of firms with other legal structures over the sample period. When focusing on small firms, if anything, there is an insignificant increase in the point estimates shortly after the

change to the second phase of the program that goes to zero towards the final period[22].

Figure 6: Over time effects of the program in the number of firms by legal structure and size

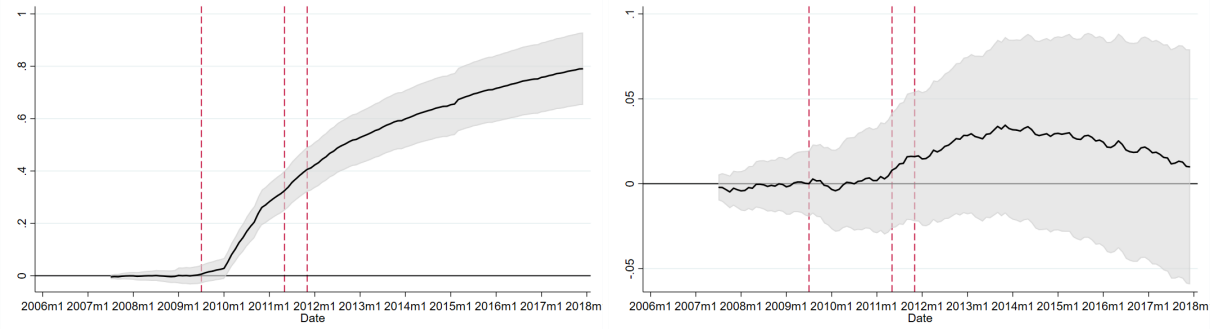
(a) Sole proprietorship:

(b) Other legal structures:



(c) Micro firms:

(d) Small:



Note: The results are clustered by industries. All regressions include state, industry, and month/year fixed effects. It also allows for different state fixed effects for treated and control industries by including their interactions. All results are estimated using the same time span: from Jan/2006 to Dec/2017. The confidence interval is generated using a 5% significance level. The first vertical line indicates the program's introduction in 07/2009. All states were included only in 02/2010. The second vertical line represents the tax reduction from 11% to 5% of the minimum wage. The third line shows the period when the income cap was increased from R$36,000 to R$60,000 a year.

6.4 Estimation of the Effects on the Informal Sector:

The analysis up to this point has focused exclusively on the creation of firms in the formal sector. However, an essential part of the economic activity comes from the informal sector in developing countries. Considering that the administrative data, by definition, does not capture this part of the economy, I use the Brazilian monthly labor force survey - *Pesquisa Mensal de Emprego (PME)* - to measure the effects of IME the program on both formal and informal sectors. The goal is to measure some aggregated switches from the informal sector towards the formal sector generated by the program and to exclude the possibility of industry-specific

[22]As explained in the data section, it is important to note that the IRF data is cross-sectional, meaning that the revenue size definitions reflect the most recent information. Thus, size classifications should be interpreted with caution, as some selection bias may influence the size of the firms that survive. Consequently, an increasing effect on the number of small firms could indicate spillover effects from firms that outgrow the program over the long term.

economic shocks coinciding with its implementation, which could potentially increase the total number of entrepreneurs in both sectors.

I estimate a slightly modified version of regression equation 3 separated by the formal and the informal sectors. Two modifications are necessary. First, since PME data covers only the six largest metropolitan regions in Brazil, I replace the state-fixed effects with metropolitan region fixed effects. Second, as explained in the data section, I use the 2-digit industry classification available in PME to control for industry fixed effects and to define the treated sectors. Considering that the program targets entrepreneurs with at most one employee, this intervention is expected to affect both self-employed individuals and small employers. Additionally, if formalization is an important channel, the introduction of the program should be associated with an increase in the number of formal entrepreneurs and a decrease in the number of informal entrepreneurs in eligible industries.

The results presented in table 3 show that the introduction of the IME program generates considerably different effects for the size of the formal and informal sectors. When comparing eligible and non-eligible industries, the program is associated with a higher number of individuals in the formal sector and a lower number in the informal sector. More specifically, the IME program increases the number of formal self-employed by 16.2% and decreases the number of informal self-employed by 16.3%, as displayed in columns 1 and 2 of table 3. The point estimates go in the same direction when focusing on the number of employers, as presented in columns 3 and 4. The program increases the number of formal employers by 10.6% and decreases the number of informal employers by 16.6%. The results jointly estimated for self-employed and employers indicate a 14.8% increase in the total number of formal and a 16.5% decrease in the number of informal individuals. Note that even though the aggregated effects are not significant for self-employed and both groups together, the effects are significant when estimated separately for the lower tax phase of the program (table 15 in Appendix).

Table 3: IME effects on the number of formal and informal entrepreneurs

Dep. Var	log(N. Self-Emp.)		log(N. 1 Employee)		log(N. Self-Emp.+1 Employee)	
	Formal	Informal	Formal	Informal	Formal	Informal
Treatment	0.1626	-0.1637**	0.1064**	-0.1663***	0.1486	-0.1656***
	(0.0982)	(0.0622)	(0.0399)	(0.0328)	(0.0920)	(0.0612)
Observations	19310	24852	9434	9271	19950	25050

Note: Significance levels: * 10%, ** 5%, ***1%. The results are clustered by industries. All regressions include metropolitan region, industry, and month/year fixed effects. It also allows for different metropolitan region fixed effects for treated and control industries by including their interactions. All results are estimated using the same time span: from Jan/2006 to Dec/2015.

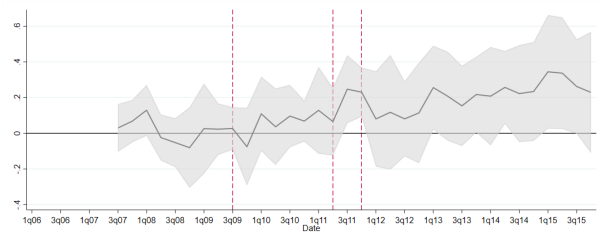
I present the estimates of the quarter-by-quarter effects in figure 7[23]. The results for all three categories - self-employed, employers, and self-employed+employers with one employee - are naturally more imprecise, considering there are fewer geographical units (only six

[23]I estimate quarter results to have more stable coefficients because the sample size is small for each month

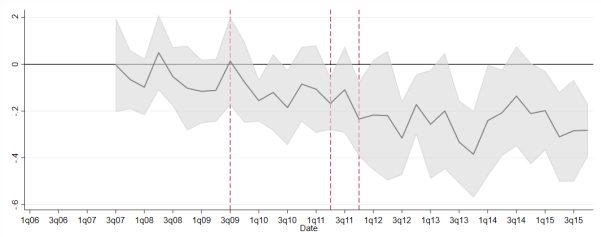
metropolitan regions) and fewer industries. Nonetheless, figure 7a presents some significant and increasing effects in the total number of formally self-employed respondents, a pattern similar to the one observed using the administrative data, and figure 7b shows that the number of respondents that are informally self-employed is significantly decreasing over time. In contrast, not much is observed when the focus is only on the employers. One potential reason is insufficient power to estimate the month-by-month effect for this category. They consist of only a small share of interviewed individuals. Finally, the results become more evident when I aggregate employers and self-employed respondents. The jointly estimated effects for formal self-employed and employers increase and stabilize from 20% to 30% and the effects for informal self-employed and employers decrease to less than -20%.

Figure 7: Over time effects of the program in the number of firms in the formal and informal sector

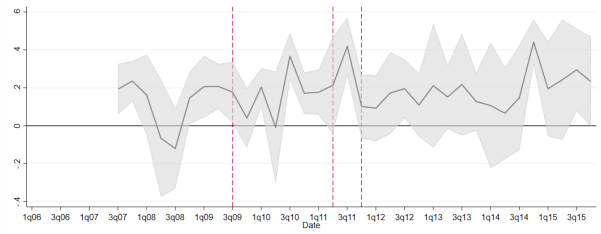
(a) log(Formal Self-employed):



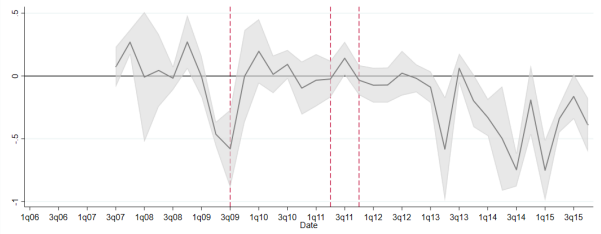
(b) log(Informal Self-employed):



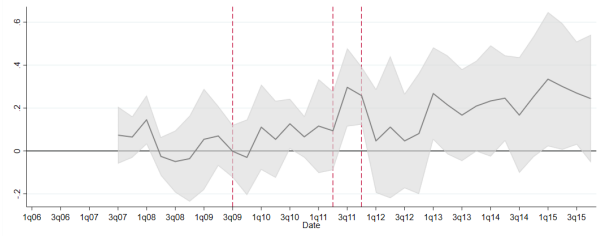
(c) log(Formal Employer):



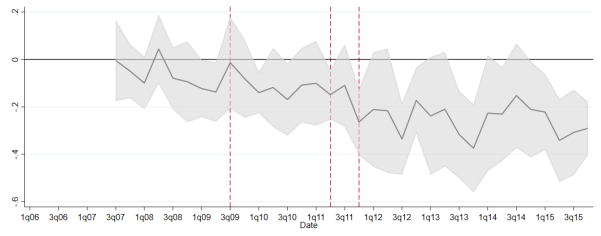
(d) log(Informal Employer):



(e) log(Formal Self-employed and Employer):



(f) log(Informal Self-employed and Employer):



Note: The results are clustered by industries. All regressions include state, industry, and month/year fixed effects. It also allows for different state fixed effects for treated and control industries by including their interactions. All results are estimated using the same time span: from Jan/2006 to Dec/2017. The confidence interval is generated using a 5% significance level. The first vertical line indicates the program's introduction in 07/2009. All states were included only in 02/2010. The second vertical line represents the tax reduction from 11% to 5% of the minimum wage. The third line shows the period when the income cap was increased from R$36,000 to R$60,000 a year.

More important to the welfare analysis is the estimation of how the share of formal workers

increased with the program's implementation. The variation in the share would proxy for a flow of entrepreneurs from the informal sector to the IME program in the long run. It is worth noting that in the long run, with the availability of the program, some individuals who would otherwise sort themselves into the informal sector may enter the IME sector directly. Table 4 presents individual-level regressions on the probability of being formal, conditional on being self-employed, an employer with one employee, or both groups combined. The results show that introducing the program increases the share of formal individuals by 2.8 p.p., 9.3 p.p., and 3.3 p.p., for self-employed, employers with one employee, or both groups, respectively. The results are mainly driven by the second phase of the program when the taxes went down to 5% of the minimum wage. The quarter-by-quarter estimates are presented in the Appendix in figure 16.

Table 4: IME effects on the probability of being formal

Dep. Var: 1{Formal=1}	Self-Emp.		1 Employee		Self-Emp.+1 Employee	
Treatment1		-0.0066		0.0311		-0.0017
		(0.0252)		(0.0405)		(0.0240)
Treatment2		0.0400***		0.1133***		0.0447***
		(0.0104)		(0.0407)		(0.0105)
Treatment	0.0279***		0.0935***		0.0328***	
	(0.0083)		(0.0265)		(0.0086)	
Average in the control	0.3239	0.3239	0.5788	0.5788	0.3381	0.3381
Observations	1018047	1018047	69440	69440	1087487	1087487

Note: Significance levels: * 10%, ** 5%, ***1%. The results are clustered by industries. All regressions include state, industry, and month/year fixed effects. It also allows for different state fixed effects for treated and control industries by including their interactions. All results are estimated using the same time span: from Jan/2006 to Dec/2017.

6.5 Estimation of the Effects of Within Formal Sector Switches:

The lower tax regime implemented by the IME program may also lead to unintended consequences. Firms that would otherwise fall under the alternative formal tax system might choose to enroll in the program to reduce their tax liabilities. As indicated in the model, this flow can have significant implications for the welfare analysis of the IME program. To assess the relevance of this effect, I examine the impact of the IME program's introduction on the total number of firms that are not part of the program. I rely on the same identification strategy to estimate the change in the total number of firms outside the IME program for both eligible and non-eligible industries.

The results presented in table 5 show no evidence of strategic transitions from the Simples tax system to the IME program. If this flow were important, the total number of firms not part of the IME program would be expected to drop in the eligible industries after introducing the IME program, generating a negative difference-in-difference coefficient. I find negative

point estimates for the number of all firms and for the ones classified as sole proprietorships but positive point estimates for the micro firms related to the lower tax phase of the program. Notably, no coefficient is statistically significant. Note that part of the zero effects could be explained by spillovers of firms growing out of the program in the long run. Figure 17 in the Appendix shows that over time the number of firms growing out of the program increased and reached more than 80,000 transitions in 2017. This number is more than twice the number of firms that strategically move from other formal options to the IME program. Figure 17 also shows that the share of IME firms that grow out of the program stabilizes around 5%.

Table 5: IME program effects in the total number of firms not part of the IME program

	Dep. Var: log(Total number of firms not part of IME)					
	All		Sole proprietorship		Micro	
Treatment	-0.0089		-0.0196		0.0086	
	(0.0347)		(0.0394)		(0.0261)	
Treatment1		0.0105		0.0267		0.0203
		(0.0252)		(0.0286)		(0.0157)
Treatment2		-0.0142		-0.0322		0.0055
		(0.0375)		(0.0427)		(0.0291)
Observations	1480308	1480308	1273848	1273848	1403580	1403580

Note: Significance levels: * 10%, ** 5%, ***1%. The results are clustered by industries. All regressions include state, industry, and month/year fixed effects. It also allows for different state fixed effects for treated and control industries by including their interactions. $Treatment1$ is the coefficient that captures the effect of the first phase of the program when the IME tax was 11% of the minimum wage. $Treatment2$ captures the effect for the period when the tax was set as 5% of the minimum wage. All results are estimated using the same time span: from Jul/2007 to Dec/2017.

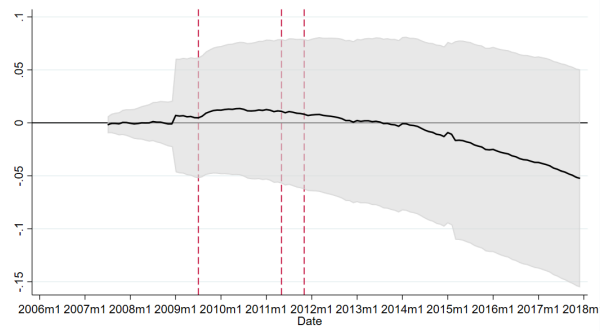
To further investigate the effects of the IME program on the firms that otherwise would be part of the formal system, I present the estimates of the month-by-month effects in figure 8. The results provide no evidence of strategic transitions from formal firms to the IME program. Although the point estimates are not statistically significant, they become increasingly negative toward the end of the analyzed period. While these estimates do not achieve significance in any of the examined periods, they may suggest a higher level of formal-to-IME transitions as participation in the program increases.

6.6 Estimation of the Tax Revenue Effects:

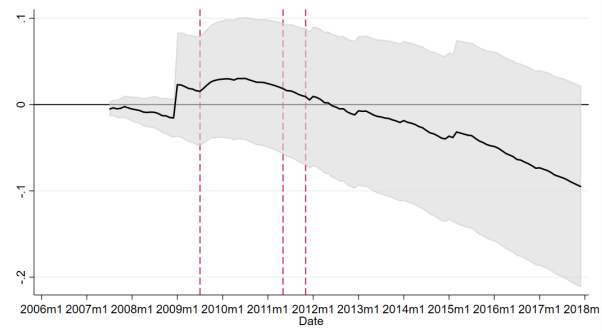
The extensive margin effects estimated in the previous sections provide evidence that enrollment in the IME program is determined by entrepreneurs who otherwise would be part of the informal sector, with no evidence of a substantial displacement from other formal options. To confirm that the IME program participants are not displacing other forms of formal firms, figure 9 illustrates the growth from 2009 to 2016 in the number of formal firms and the tax revenue collected across 87 distinct 2-digit industries, ranked by the significance of IME firms

Figure 8: Over time effects of the program in the total number of firms not part of the IME

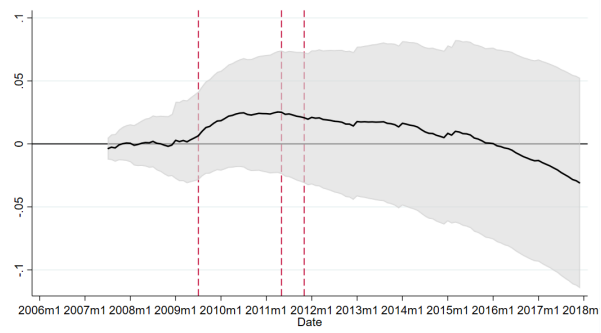
(a) log(Total number of firms not IME):



(b) log(Total number of sole proprietorship firms not IME):



(c) log(Total number of micro firms not IME):



Note: The results are clustered by industries. All regressions include state, industry, and month/year fixed effects. It also allows for different state fixed effects for treated and control industries by including their interactions. All results are estimated using the same time span: from July/2007 to Dec/2017. The confidence interval is generated using a 5% significance level. The first vertical line indicates the program's introduction in 07/2009. All states were included only in 02/2010. The second vertical line represents the tax reduction from 11% to 5% of the minimum wage. The third line shows the period when the income cap was increased from R$36,000 to R$60,000 a year.

as a share of the total number of formal firms in 2016. The growth figures for the total number of firms and tax revenue are calculated both including and excluding the firms enrolled in the IME program in 2016. If displacement is associated with the IME program, we would expect the growth in both the number of firms and tax revenue after excluding IME participants to be negatively correlated with the share of IME firms in 2016.

The descriptive evidence supports the idea that the inflow from the informal to the formal sector is the primary driver of enrollment in the IME program. First, as shown in Figure 9a, the growth in the total number of formal firms is positively correlated with the relevance of the IME sector in each industry in 2016. This aligns with the main results indicating a higher increase in formal firms in industries eligible for the IME. Specifically, an increase of 0.01 in the share of IME firms within an industry corresponds to a 2.2% rise in the total number of firms in that industry. Second, the growth in the total number of firms after excluding IME participants (formal non-IME firms) is also positively correlated with the IME share in 2016. An increase of 0.01 in the share of IME firms is associated with an insignificant 0.14% increase in non-IME formal firms, suggesting that there is no material displacement from other formal tax regimes.

Figure 9b presents the difference between the growth in the total number of firms, including and excluding the IME firms. Mechanically, the higher the share of IME firms in 2016, the higher the growth in total number of firms explained by the IME program. To illustrate, for the two sectors with the highest share of IME participants in 2016, Personal Care (87% in 2016) and Housekeeping services (99% in 2016), the IME firms alone would represent a growth of 2.04 and 4.86 times the number of firms in 2009.

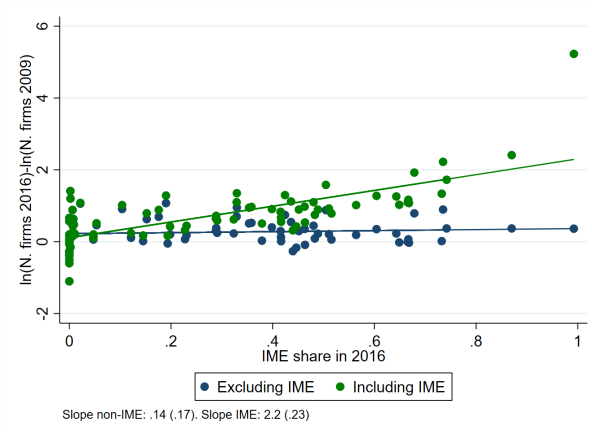
The same pattern is observed for the industry-level tax revenue growth presented in figure 9c. The change in tax revenue is positively correlated with the significance of the IME sector in each industry in 2016. Specifically, an increase of 0.01 in the share of IME firms is associated with a 0.54% rise in total tax revenue for that industry. Additionally, the growth in tax revenue, after excluding revenue from IME participants, also shows a positive correlation, with a non-significant 0.31% increase for each 0.01 rise in the share of IME firms. Together, these correlations suggest evidence that the IME program successfully increased tax revenue, with no significant displacement from other formal tax systems occurring.

Figure 9d presents the difference between the growth in the tax revenue, including and excluding the amount remitted by the IME firms. While the IME participants are important in explaining the growth in the number of firms, this is not reflected in tax revenue. Because the IME firms are very small compared to other firms, the IME tax revenue alone would account for at most a 5% growth in the tax revenue collected in most industries. In contrast, for the two industries with the highest share of IME participants in 2016, Personal Care (87% in 2016)

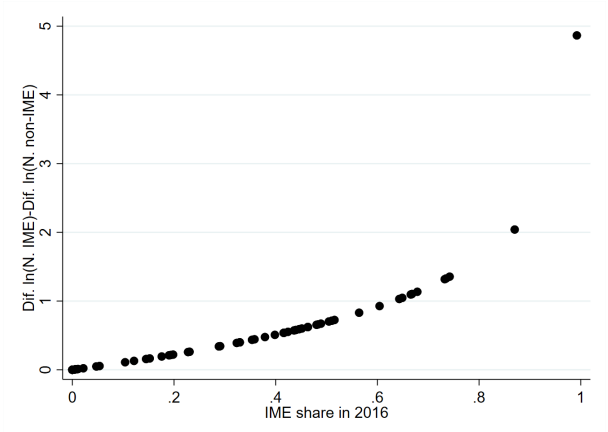
and Housekeeping services (99% in 2016), the IME firms alone would represent a tax revenue growth of 11% and 170%[24].

Figure 9: Growth in the number of firms and tax revenue from 2009 to 2016 by the share of IME firms in each sector

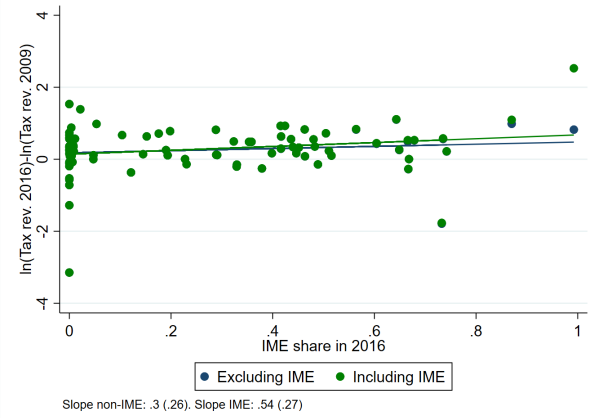
(a) Growth in the number of firms excluding and including IME firms:



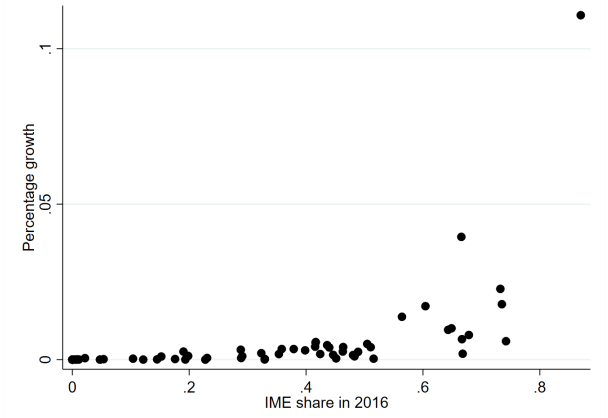
(b) Growth in the number of firms explained by the IME:



(c) Growth in the tax revenue excluding and including IME firms:



(d) Growth in the tax revenue explained by the IME:



Note: The graphs are based on a RFB data that measure the aggregated tax remittances and total number of firms for the 85 different 2-digit industries separately in 2009 and 2016. For each sector, I plot in figures (a) and (c) the difference in logarithms of the outcome in the Y-axis from 2009 to 2016 and correlate it with the share of the total firms that were part of the IME program in the same industry in 2016. The results are displayed both including and excluding the IME firms from the growth calculations. Figures (b) and (d) take the difference in the growth when I include and do not include the IME firms.

7 Robustness Estimation of the Main Mechanisms:

In this section, I check other potential margins that the program introduction could influence. Table 6 presents results testing whether the program affected employees' decisions

[24]The results for housekeeping services are not shown in figure 9d due to the scale of the tax revenue growth, 170%

to transition to self-employment and whether firms with more than two employees downsized to qualify for the program. This analysis is particularly important, as the theoretical model does not account for the behavioral responses of non-self-employed individuals. I re-estimate the difference-in-difference specification comparing the change in the number of employees, firms with 2 employees, and firms with 3 to 5 employees in the control and treated industries. Columns 1 and 2 show no significant difference in the change in the number of formal employees; however, treated industries experienced a 14.3% reduction in the number of informal employees compared to control industries. Columns 3 and 4 indicate no significant change in the number of formal and informal firms with two employees. Similarly, Columns 5 and 6 show no significant change in the number of formal firms with three to five employees, but a 6.3% decrease in the number of informal firms within this group is observed. In summary, there is no evidence of occupational changes or downsizing among formal firms. However, the program appears to be associated with a decrease in both informal employees and informal firms, suggesting that some individuals from these groups may have transitioned into the IME program. Given that these potentially affected groups are part of the informal sector, the tax implications of moving into the IME program would be similar to those seen with increased formalization. Therefore, I do not model these specific channels, as this would complicate the analysis without providing additional gains in intuition.

Table 6: IME effects on the probability of being formal

	log(Employee)		log(2 employees)		log(3 to 5 employees)	
	Formal	Informal	Formal	Informal	Formal	Informal
Treatment	-0.0075	-0.1439***	-0.0175	-0.0465	0.0486	-0.0633**
	(0.0356)	(0.0420)	(0.0325)	(0.0414)	(0.0548)	(0.0280)
Observations	35232	28089	8953	6980	11559	7315

Note: Significance levels: * 10%, ** 5%, ***1%. The results are clustered by industries. All regressions include state, industry, and month/year fixed effects. It also allows for different state fixed effects for treated and control industries by including their interactions. All results are estimated using the same time span: from Jan/2006 to Dec/2017.

Apart from testing alternative margins, I conduct several additional robustness tests. First, I estimate the program's effects on formalization among self-employed individuals in different income brackets. Table 16 shows that greater formalization is found for the middle-income groups that are right below the eligibility cap. Second, I also re-estimate all the results with the PNAD. The advantage of using PNAD is that this survey contains the 5-digit industry of everyone working. On the other hand, this survey is conducted annually, has fewer observations, and does not cover the period of the first phase of the IME program[25]. The results are presented in the Appendix and are similar to the ones estimated using the PME. Third, I estimate different effects according to the share in an industry that sells to the government or a firm and

[25]The PNAD survey is not conducted in Census years (every ten years). Therefore, there is no data for 2010.

present the results in table 21 of the appendix. The results show that the formalization effects are stronger in industries where self-employed and employers with one employee sell more to the final consumer in the pre-period. Fourth, I present the results for different specifications of the regression model in table 20 of the appendix. The results are similar after controlling for composition effects and different unemployment effects by industry. It is also robust to the inclusion of more years, taking some cities out and changing the definition of the treatment.

8 Evaluating the IME Optimal Policy Parameters:

As stated in Proposition 4, increased tax revenue is a sufficient condition for the introduction of the program to be welfare-improving, given entrepreneurs' private utility cannot decrease when they have more options. The empirical evidence presented in Section 7 indicates that the primary channel for enrollment in the program is through transitions from the informal sector, with no significant movement from firms in other formal tax regimes to the IME. Therefore, taken at face value, the empirical results can only imply a higher tax revenue generated by the program, which leads to the conclusion that the program introduction was welfare-increasing. It is important to note that this is a partial equilibrium result and, as such, assumes away any general equilibrium effects related to increased enforcement.

Another significant policy-related exercise involves evaluating the first-order conditions to assess whether the income cap and IME tax were optimally set following the program's introduction. To evaluate the first-order conditions, I first adjust them to reflect the variations in the shares of informal and formal non-IME entrepreneurs, as estimated in my empirical results. My analysis concentrates on the estimates from the second phase of the IME program, during which the tax was reduced from 11% to 5% of the minimum wage, and the income cap was raised from R$36,000 to R$60,000 per year[26]. Two assumptions are necessary. First, because the implementation of the lower tax and the higher income cap occurred within a short time frame, I am unable to estimate their separate effects on the shares. Consequently, I assume that both changes were equally important[27]. Second, because my empirical estimates capture the impact on both shares when transitioning from the first phase of the program to the second phase (with a lower tax and a higher income cap), I assume the effects of the cap and the tax to be linear. This allows me to account for the impact of changing each by R$1.

[26] I work with the monthly values of the cap even though the restriction are valid only for the total income in a year

[27] In the empirical analysis, I estimate the impact of transitioning the program to its second phase on the number of formal firms not participating in the IME program. Notably, since this transition involves both raising the income cap and lowering the tax simultaneously, it yields a single estimate for the variation. I use the same estimate ($\Delta(1 - F(a_{Ime,S}))$) to build the extensive margin effect for the tax and for the income cap, such that $\frac{\partial 1-F(a_{Ime,S})}{\partial \tau} = \frac{\Delta(1-F(a_{Ime,S}))}{\tau^2-\tau^1}$, and $\frac{\partial 1-F(a_{Ime,S})}{\partial Y_C} = \frac{\Delta(1-F(a_{Ime,S}))}{Y_C^2-Y_C^1}$.

Below are the modified versions of the first-order conditions. The blue terms are measured using the estimated results in my empirical exercise. The brown terms are calibrated, and the red term is bounded after the first-order conditions are evaluated.

$$\frac{\partial W(\tau,\alpha,t_o,Y_C)}{\partial \tau} = \underbrace{[v'(g)-1][P(Ime)]}_{\substack{\text{Net welfare of a higher}\\ \text{tax on IME participants}\\ >0}} + v'(g)[\underbrace{\frac{\partial P(Sim)}{\partial \tau}}_{\substack{\text{Change in the}\\ \text{share of formal}\\ \text{non-IME}\\ >0}} \underbrace{(-\tau + t_o E[(Y^*_{Sim} - e^*_{Sim})|\text{Switch after}\Delta\tau = 1])}_{\substack{\text{Net revenue of switchers}\\ Undetermined}} - \underbrace{\frac{\partial P(Inf)}{\partial \tau}}_{\substack{\text{Change in the}\\ \text{share of informal}\\ >0}} \tau]$$

$$\frac{\partial W(\tau,\alpha,t_o,Y_C)}{\partial Y_C} = \underbrace{E[c_e|IME=1, Y^*_{Ime} > Y_C]P(Ime|Y^*_{Ime} > Y_C)}_{\substack{\text{Marginal cost of evasion for}\\ \text{IME under reporting participants}\\ >0}} + v'(g)[\underbrace{\frac{\partial P(Sim)}{\partial Y_C}}_{\substack{\text{Change in the}\\ \text{share of formal}\\ \text{non-IME}\\ >0}} \underbrace{(-\tau + t_o E[(Y^*_{Sim} - e^*_{Sim})|\text{Switch after}\Delta Y_C = 1])}_{\substack{\text{Net revenue of switchers}\\ Undetermined}} - \underbrace{\frac{\partial P(Inf)}{\partial Y_C}}_{\substack{\text{Change in the}\\ \text{share of informal}\\ >0}} \tau]$$

The parameter values used to evaluate the first-order conditions are presented in Table 7. Note that one of the parameters needed is the income that would be reported in the Simples system by the marginal switcher from Simples to IME. Unfortunately, since I cannot access reported income from the administrative data, estimating it with the available data is not feasible. Instead, I use the average income from the 2014 PNAD survey, along with the average income above the cap for formal entrepreneurs not participating in the IME, to calibrate the reported income in the Simples system for the marginal switcher attracted by the fee and the cap changes, respectively[28]. I also choose the Simples to IME flow parameters to be the highest response point estimate estimated in Figure 8a for the end of the second phase of the program, which gives a conservative measure of the welfare effects of the program[29]. This estimate reflects the percentage decrease in formal non-IME firms during the last month of analysis. I convert this percentage decrease into percentage point values by multiplying it by the share of formal entrepreneurs prior to the program's implementation.

Table 8 presents the welfare evaluation of the first-order conditions and includes values for key parameters that would make the IME income cap and fee of the second phase optimal. In line 1, I assess welfare in a scenario where there is no inflow from the Simples system to the IME program. In this case, increasing the IME tax by R$1 would decrease welfare by R$0.022. In line 2, with the chosen parameters described in Table 7, an increase of R$1 in the IME tax

[28]Note that changes in fee and cap potentially attracts entrepreneurs with different ability. The fee makes the IME more attractive for every entrepreneur. The cap only changes the decision for individuals who would report their income above the IME cap in case they were enrolled in the IME.

[29]The greater the flow from other formal sectors to IME, the more substantial the tax increase related to expelling individuals from the program. In this context, assuming a higher formal-to-formal flow implies that the optimal size of the IME program should be smaller, providing a conservative estimate of the program's welfare impact.

Table 7: First-order condition parameters

Parameter	Description	Source	Value
Chosen parameters:			
τ^2	IME tax in phase 2	Statutory	R$36.2 per month
Y_C^2	IME income cap in phase 2	Statutory	R$5,000 per month
t_o	Simples output tax	Statutory	0.05
$v'(g)$	Marginal value of public funds	Keen and Slemrod (2017)	1.2
$\eta_{Fnon-IME}$	Share of formal before the IME program	PME from 2006m1 to 2009m6	0.192
$[F(a_{Ime,S}) - F(a_{I,Ime})]$	IME share	PNAD in 2014	0.1137
$E[(Y^*_{Sim} - e^*_{Sim})\|\text{Switch after}\Delta\tau = 1]$	Average reported income in the Simples system of the marginal switcher due to the fee change	PNAD 2014 - Average income for the formal entrepreneur not in the IME	R$2,711
$E[(Y^*_{Sim} - e^*_{Sim})\|\text{Switch after}\Delta Y_C = 1]$	Average reported income in the Simples system of the marginal switcher due to the cap change	PNAD 2014 - Average income for the formal entrepreneur not in the IME with income above the cap	R$10,046
Estimated parameters:			
$\Delta(F(a_{I,Ime}))$	Variation in the informal share from phase 1 to phase 2	Table 4	-4.47p.p.
$\frac{\partial F(a_{I,Ime})}{\partial \tau}$	$\frac{\Delta(F(a_{I,Ime}))}{\tau^2-\tau^1}$		0.001367
$\Delta(\text{Log(Formal non-IME)})$	Percentage variation in the number of formal firms non-IME	Last point estimate from Figure 8a	-5.25%
$\Delta(1 - F(a_{Ime,S}))$	$[(1 - \Delta(\text{Log(Formal non-IME)}))\eta_{Fnon-IME}]$ - $\eta_{Fnon-IME}$		-1p.p.
$\frac{\partial 1-F(a_{Ime,S})}{\partial \tau}$	$\frac{\Delta(1-F(a_{Ime,S}))}{\tau^2-\tau^1}$		0.000232
$\frac{\partial 1-F(a_{Ime,S})}{\partial Y_C}$	$\frac{\Delta(1-F(a_{Ime,S}))}{Y_C^2-Y_C^1}$		-0.000005

would improve welfare by R$0.0056, assuming a formal-to-formal flow of 1 percentage point. Generally, any Simples to IME flow exceeding 0.8 percentage points suggests that a higher IME fee would enhance welfare, holding other parameters constant. Although the transitions from Simples to IME are not significant in the empirical results, the point estimates indicate an increasing transition rate over time, surpassing the 0.8 percentage point threshold toward the end of the sample period. This implies that the optimal fee should be higher than the one in place during the second phase as the end of the sample period approaches.

I conduct a similar analysis for the marginal switcher's reported income in the Simples system, which was selected based on the average income of the formal entrepreneurs not participating in the IME, as well as for the average marginal cost of evasion faced by IME entrepreneurs who underreport their income to qualify, which remains unknown. First, for any reported income above R$2,302, increasing the IME fee becomes welfare-improving, as expelling individuals from the program enhances tax revenue at this threshold. This value is lower than the average income in the Simples system (R$2,711) and less than half of the eligibility cap (R$5,000)[30]. Second, I demonstrate that the average distortion created by the cap must be at least 0.18 for an increase in the cap to be welfare-improving, which is more than three times the distortion created by the output tax (0.05). Raising the cap results in higher real income for IME participants who were underreporting to qualify, but it also has tax revenue implications, as it attracts more individuals to the program.

[30]Note that the reported income of the marginal switcher can only be lower than the IME cap in theory due to the assumption that individuals have idiosyncratic preferences about their sectors. If the reported income is assumed to be more than the cap, the cutoff for the formal-to-formal flow would go down to at least 0.17 p.p..

Table 8: Welfare evaluation and bounds

Evaluated term	Phase 2 (lower tax, higher cap)
Panel A):	
$\frac{\partial W(\tau,\alpha,t_o,Y_C)}{\partial \tau}$ if Formal to Formal flow equals 0	-R$0.0220
$\frac{\partial W(\tau,\alpha,t_o,Y_C)}{\partial \tau}$	R$0.0056
Keeping everything else constant:	
Calibrated Formal to Formal flow value	1 p.p.
Formal to Formal flow value for which $\frac{\partial W(\tau,\alpha,t_o,Y_C)}{\partial \tau} = 0$	0.8p.p.
Optimal τ change if	
F to F flow < 0.8 p.p.	Decrease
0.8 p.p. $<$ F to F flow	Increase
Calibrated income of marginal switcher value	R$2711
Reported income of marginal switcher value for which $\frac{\partial W(\tau,\alpha,t_o,Y_C)}{\partial \tau} = 0$	R$2302
Optimal τ change if	
0 $<$ Rep. Income of mg. switcher $< R\$2302$	Decrease
$R\$2302.$ $<$ Rep. Income of mg. switcher	Increase
Panel B):	
$\frac{\partial W(\tau,\alpha,t_o,Y_C)}{\partial Y_C}$ depends on the marginal cost of evasion	
Keeping everything else constant:	
Average Mg. evasion cost for which $\frac{\partial W(\tau,\alpha,t_o,Y_C)}{\partial Y_C} = 0$	0.18
Optimal Y_C change if	
Avg. Mg. evasion cost < 0.18	Decrease
$0.18 <$ Avg. Mg. evasion cost	Increase

9 A Structural Model of the IME Program:

The last section indicates that increasing the IME fee towards the end of my sample period would improve welfare. It also provides the relevant range for the IME cap distortion so that policymakers can decide in which direction the income cap should move to increase welfare. To provide a more precise answer about the optimal fee and cap, additional assumptions about the parametric forms of the cost functions and the distribution of the model's stochastic components are required. First, I assume the same functional forms for the production and evasion costs that are used by Akcigit et al. (2022):

$$\psi(Y,a) = \frac{a}{1+\frac{1}{\epsilon}}\left(\frac{Y}{a}\right)^{(1+\frac{1}{\epsilon})} \qquad c(e,\alpha) = \frac{1}{\alpha}\frac{1}{1+\frac{1}{\eta}}(\alpha e)^{(1+\frac{1}{\eta})}$$

Second, I assume the ability distribution follows a log-normal with mean μ and variance σ, and that the sector idiosyncratic preference follows a type 1 extreme value distribution with zero location and a σ_{ext} scale parameters[31].

I introduce two modifications to the model presented in Section 6 that do not alter the conclusions drawn from the first-order condition analysis in the previous section. First, to simplify, I consider a version of the model that does not allow evasion in the *Simples* option. Second, I reformulate the model to span two periods, allowing for greater flexibility and the incorporation of transition costs and inertia. Entrepreneurs can choose between the *Simples* option and the informal sector in the first period. In the second period, their IME utility is contingent upon their choice in the first period, reflecting the assumption that transition costs depend on that choice. More precisely, $U^{t_2}_{Ime}(s^{t_1}) = U_{Ime} - E_{s^{t_1}}$, where $E_{s^{t_1}}$ is the transition cost associated with the choice s in t_1.

Given the assumption that the preference shock, ξ_s, follows a type 1 extreme value distribution, a convenient logistic equation defines the probability of (or share) choosing each option. In the first period, the probability of selecting sector s conditional on the ability and unconditional are respectively:

$$P(s^{t_1} = s|a) = \frac{e^{\frac{U_s}{\sigma_{ex}}}}{e^{\frac{U_{Sim}}{\sigma_{ex}}} + e^{\frac{U_{Inf}}{\sigma_{ex}}}}$$

$$P(s^{t_1} = s) = \int \frac{e^{\frac{U_s}{\sigma_{ex}}}}{e^{\frac{U_{Sim}}{\sigma_{ex}}} + e^{\frac{U_{Inf}}{\sigma_{ex}}}} f(a)da$$

[31]The Type I extreme value distribution is widely utilized in economics, particularly in the field of industrial organization. It offers a closed-form equation for sector choice conditional on each ability level, which helps reduce the dimensionality of the estimation procedure.

In the second period, since the utility of the IME is contingent upon the choice made in the first period, the probabilities of choosing each sector are influenced by these prior decisions. Consequently, the probability of staying in sector s or moving to the IME is:

Staying in sector s in period 2:

$$P(s^{t_2} = s|a) = P(s^{t_1} = s|a) \cdot \frac{e^{\frac{U_s}{\sigma_{ex}}}}{e^{\frac{U_s}{\sigma_{ex}}} + e^{\frac{U^{t_2}_{Ime}(s^{t_1})}{\sigma_{ex}}}}$$

$$P(s^{t_2} = s) = \int P(s^{t_1} = s|a) \cdot \frac{e^{\frac{U_s}{\sigma_{ex}}}}{e^{\frac{U_s}{\sigma_{ex}}} + e^{\frac{U^{t_2}_{Ime}(s^{t_1})}{\sigma_{ex}}}} f(a)da$$

Moving to IME in period 2:

$$P(s^{t_2} = Ime|a) = P(s^{t_1} = Sim|a) \cdot \frac{e^{\frac{U^{t_2}_{Ime}(Sim^{t_1})}{\sigma_{ex}}}}{e^{\frac{U_{Sim}}{\sigma_{ex}}} + e^{\frac{U^{t_2}_{Ime}(Sim^{t_1})}{\sigma_{ex}}}} + P(s^{t_1} = Inf|a) \cdot \frac{e^{\frac{U^{t_2}_{Ime}(Inf^{t_1})}{\sigma_{ex}}}}{e^{\frac{U_{Inf}}{\sigma_{ex}}} + e^{\frac{U^{t_2}_{Ime}(Inf^{t_1})}{\sigma_{ex}}}}$$

$$P(s^{t_2} = s) = \int P(s^{t_1} = Sim|a) \cdot \frac{e^{\frac{U^{t_2}_{Ime}(Sim^{t_1})}{\sigma_{ex}}}}{e^{\frac{U_{Sim}}{\sigma_{ex}}} + e^{\frac{U^{t_2}_{Ime}(Sim^{t_1})}{\sigma_{ex}}}} + P(s^{t_1} = Inf|a) \cdot \frac{e^{\frac{U^{t_2}_{Ime}(Inf^{t_1})}{\sigma_{ex}}}}{e^{\frac{U_{Inf}}{\sigma_{ex}}} + e^{\frac{U^{t_2}_{Ime}(Inf^{t_1})}{\sigma_{ex}}}} f(a)da$$

Because the transition cost may vary depending on the sector choice in period 1, it is essential to calculate the probability of transitioning from each option in period 1 to the IME in period 2. Notably, σ_{ex} dictates the degree of sector variation for individuals with the same ability. It can be estimated in this case since the utility functions are normalized to express the utility in dollar amounts and are not linear[32].

It is important to characterize the optimal income decisions alongside the sector shares defined by the logistic equations. These outputs of the model have empirical counterparts that will be used to inform the model parameters. As mentioned, I assume there is no evasion in the *Simples* option. The first-order conditions for each sector are as follows:

Simples: $Y^*_{Sim} = a(1 - t_o)^{\epsilon}$

Informal: $1 - (\alpha Y^*_{Inf})^{\frac{1}{\eta}} - (\frac{Y^*_{Inf}}{a})^{\frac{1}{\epsilon}} = 0$

IME: $\begin{cases} Y^*_{Ime} = a & \text{if } Y^*_{Ime} \leq Y_C \\ 1 - (\alpha(Y^*_{Ime} - Y_C))^{\frac{1}{\eta}} - (\frac{Y^*_{Ime}}{a})^{\frac{1}{\epsilon}} = 0 & \text{if } Y^*_{Ime} > Y_C \end{cases}$

I use the Simulated Minimum Distance method to find the optimal set of parameters that fully characterize self-employed behavior. I simulate 200,000 ability draws, which remain constant throughout the estimation process. Given that $\phi^{Stat} := \{t_o, \tau, Y_C\}$ are defined statutorily,

[32] In the industrial organization literature, it is not possible to separately estimate the standard deviation of the idiosyncratic preference from the coefficients of the model.

I have 9 parameters to be estimated - $\phi^{Est} := \{\alpha, \eta, \epsilon, \theta, \mu, \sigma, E_{Sim^{t_1}}, E_{Inf^{t_1}}, \sigma_{ex}\}$. The estimation process focuses on finding the parameters that best align the moments derived from the simulated data with the corresponding real empirical moments. This approach ensures that the model accurately reflects observed self-employed behavior.

Define $\hat{m}_N = \frac{1}{N}\sum_{i=1}^{N} m_i$ as the moments calculated from the empirical data, and $\hat{m}_S(\phi^{Est}; \phi^{Stat}) = \frac{1}{S}\sum_{i=1}^{S} m_s(\phi^{Est}; \phi^{Stat})$ the simulated moments. Let $g_{NS}(\phi^{Est}; \phi^{Stat}) = \frac{\hat{m}_N - \hat{m}_S(\phi^{Est}; \phi^{Stat})}{\hat{m}_N}$ be the percentage difference from the simulated and empirical moments. The estimator is given by minimizing the distance between the empirical moments and the simulated moments:

$$\hat{\phi}^{Est} = \arg\min \mathrm{Q}(\phi^{Est}; \phi^{Stat}) = g_{NS}(\phi^{Est}; \phi^{Stat})'\hat{W} g_{NS}(\phi^{Est}; \phi^{Stat})$$

where $\hat{W}$ is a weight matrix with the standard properties: positive, semi-definite m x m matrix, where m is the size of the moments vector. To minimize the asymptotic variance, I choose $\hat{W} = \hat{\Sigma}^{-1}$, where $\hat{\Sigma}$ is the two-step variance-covariance matrix of the percentage error vector, $g_{NS}(\phi^{Est}; \phi^{Stat})$[33]. This choice of weighting matrix ensures that the estimator is efficient and provides reliable parameter estimates by appropriately accounting for the variability in the moments.

I rely on 25 empirical moments based on each sector's share and income distribution. These moments include: the share of individuals in the Simples tax regime prior to the program's introduction, and the share of individuals in the IME during both phase 1 and phase 2 of the program's implementation[34]. The extensive margin estimates of the IME share in phase 2 that transitioned from both the informal sector and other formal tax regimes[35]. The income distribution in each sector option (informal, IME and *Simples*) given by the share of people after the program introduction with income between R$0 and R$1000, R$1000 and R$2000, and R$2000 and R$5000; and the average income in each sector option. The share of people in the IME and *Simples* conditional on the individual earning between R$0 and R$1000, R$1000 and R$2000, R$2000 and R$5000, and above R$5000[36].

[33] In the first stage, I use an identity matrix as the weighting matrix. Once the model is estimated, I utilize the resulting parameter estimates to compute the variance-covariance matrix. This estimated variance-covariance matrix is then employed to construct the optimal weight matrix. With the optimal weight in place, I re-estimate the model to improve efficiency and obtain more precise parameter estimates.

[34] I only observe the share of IME individuals in the second phase of the program. This data is only available in the 2014 PNAD survey. To calculate the share of IME in the first phase, I use the administrative data to get the ratio of the number of IME individuals in 2011 and 2014, $\frac{\text{N. of IME in 2011}}{\text{N. of IME in 2014}}$. I multiply this ratio by the share observed in 2014.

[35] The estimated extensive margins are 4.47p.p. and 0.618p.p. for the informal to IME flow and the other formal to IME flow, respectively. The sum of both margins is less than the total IME share in the survey in 2014 - 11.37%. I normalize the extensive margin estimates to the share that comes from each option, $\frac{\text{Ext. margin from sector S}}{4.47+0.618}$ and multiply it to 11.37%.

[36] The number of income brackets is limited by data availability. Ideally, a finer division into more income

Conditional on the parameters, as well as the ability and idiosyncratic preference draws, I can fully identify the model and generate simulated moments that will match the empirical ones. While I cannot precisely determine what uniquely identifies each parameter, I can provide some intuition regarding the data patterns that will discipline them. First, the income distribution of IME entrepreneurs earning below the cap is crucial for identifying the ability distribution parameters, as ability is the sole determinant of income in the first-order conditions. The income distribution within the *Simples* option further helps identify the tax elasticity of income, ϵ. Meanwhile, the informal income distribution and the share of IME participants earning above the cap inform the identification of the tax elasticity of evasion and the propensity to evade, represented by η and α.

The share of formal firms in the pre-period serves as an identifier for compliance costs. In the post-period, both the overall share of firms in the IME and the share coming from each sector are critical for identifying the transition costs associated with moving from the informal sector or other formal tax regimes into the IME program. Finally, the share of IME and *Simples* participants conditional on income brackets also contributes to identifying the parameters of the evasion cost function. Specifically, the absence of informal firms in the upper-income levels constrains the convexity of the evasion cost function, thereby influencing the estimation of η and α. The income overlap among the different options helps identify the variance of the idiosyncratic shock, while the share of IME participants in phase 1 disciplines this parameter to ensure that the model remains sensitive to tax changes.

The estimated parameters are presented in table 9, and all are statistically significant at the 1% level. The results indicate a higher sensitivity of tax evasion to changes in taxation within the Brazilian context compared to estimates for developed countries. By focusing on self-employed workers in France, Akcigit et al. (2022) estimate a 1.1 tax elasticity of evasion, η, and calibrate the tax elasticity of income, ϵ, to be 0. In contrast, I estimate a tax elasticity of evasion of 2.3, highlighting greater responsiveness to tax changes in Brazil, likely due to the prevalence of the informal sector. Additionally, my estimate of the tax elasticity of income is 0.2, suggesting that non-compliance, rather than changes in income, is the primary behavioral response to tax alterations.

Moreover, Akcigit et al. (2022) estimate the parameter equivalent to the enforcement level, α, in my model to be between 0.0063 and 0.0167, while my estimates are three orders of magnitude lower, with a point estimate of 0.000003. Beyond regulating the total cost of evasion, this parameter is also linked to the maximum evasion level permitted by the model, as indicated by

brackets would be preferable, provided it does not introduce noise into the estimation due to small sample sizes in each group. I have chosen to use only four income brackets to ensure a sufficient number of IME participants in each category. For instance, there are only 181 IME participants in the highest bracket, which illustrates the challenge of maintaining adequate observations across all groups.

the first-order conditions[37]. The stark difference is likely due to the fact that they focus on the evasion of firms bunching around the eligibility cap, whereas my model also accounts for evasion by firms operating in the informal sector. For instance, the implied evaded amount ranges from 500 to 1800 euros per year in their context, while in my context, informal entrepreneurs evade more than 14,000 *reais* annually[38].

The compliance cost of the *Simples* system, θ, is estimated to be R$ 109, which is 4% of the average income of people in the *Simples* system and 9% of the average income of people in the informal sector. The estimated transition costs for both sectors to the IME are higher than the compliance cost of the *Simples* system. These parameters may indicate high inertia in sector choice as well as subjective benefits of the non-IME sectors that are not accounted for by the model[39]. The estimated transition cost from the *Simples* sector, $E_{Sim^{t_1}}$, is R$ 416, which is 15% of the average income of entrepreneurs in this sector. The transition cost from the informal sector, $E_{Inf^{t_1}}$, is R$ 213, which is 18% of the average income of entrepreneurs in this sector. The general ability distribution is concentrated at lower ability levels, with a median of 1094, reflecting the typically low income levels for self-employed individuals in developing countries. The model also estimates significant heterogeneity in idiosyncratic preferences, with a standard deviation, σ_{ex}, of 75. This implies that a sector offering R$ 170 more in profits would attract over 90% of entrepreneurs, ensuring income overlap among the various sector choices.

Table 9: Estimated parameters

Parameters	Description	Estimate	SE
η	Tax elasticity of evasion	2.3098	0.0578
ϵ	Tax elasticity of income	0.2398	0.0267
θ	Compliance cost	109.1102	3.1387
α	Enforcement level	3.8509e-06	4.8397e-07
$E_{Inf^{t_1}}$	Transition cost from informal	213.8476	5.3555
$E_{Sim^{t_1}}$	Transition cost from formal	416.7451	5.2864
μ	Mean of log(ability)	7.0075	0.0023
σ	Standard deviation of log(ability)	0.7406	0.0011
σ_{ex}	Standard deviation of the idiosyncratic shock preference	75.7172	1.7056

All the simulated and empirical moments are presented in table 24 in the Appendix. Figure 10 illustrates the share of entrepreneurs in each option and their respective income distribution.

[37] For each individual, the evasion level can never be higher than the inverse of α.

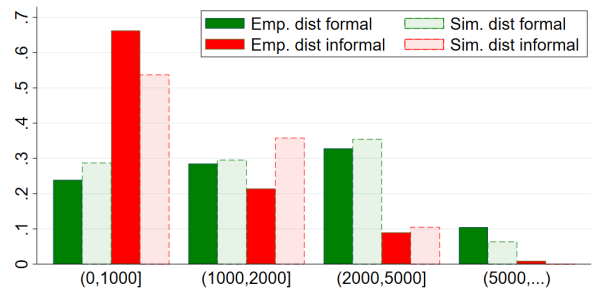
[38] This parameter is sensitive to the measurement unit, and it is the difference in the absolute amount of evasion that primarily explains the variation between the estimates.

[39] The transition costs estimates work as a way to give more flexibility to model and center the idiosyncratic preference in some specific levels. As the idiosyncratic preference, they can rationalize different mechanisms that affect the sector choices but are not explicitly modeled, such as different risk tolerance, different individual levels of attention, and different preferences for the type of business organization. For example, the IME sector does not allow entrepreneurs to have a business partner. This is a non-exhaustive list of potential reasons, and identifying each of them is beyond the scope of this paper.

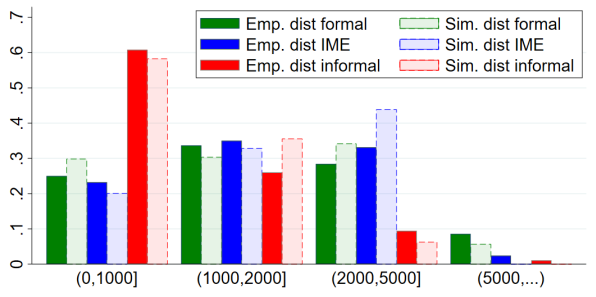
As presented in figure 10c, the model matches the targeted share in each sector well, especially after the program's introduction, presenting only one decimal point of difference from the actual shares observed in the data. The model also captures the income distribution patterns in each sector but with imprecision for some brackets. Figure 10a and 10b respectively present the simulated and empirical moments for the untargeted pre-program income distribution and the targeted post-program income distribution. In both the pre- and post-program periods, the model effectively captures the declining share of informal workers in higher-income groups, although it tends to overestimate the share in the second income bracket. The model also successfully reflects the stability of shares in the first three income groups and the sharp decline in the fourth bracket for both types of formal sector participants, before and after the program's implementation. However, it overestimates the share of IME participants just below the program's income cap (R$5000). The model further estimates the average income with some degree of imprecision, but it correctly represents the higher average income for both formal options compared to the informal option.

Figure 10: Empirical and simulated moments of the income distribution and sector shares:

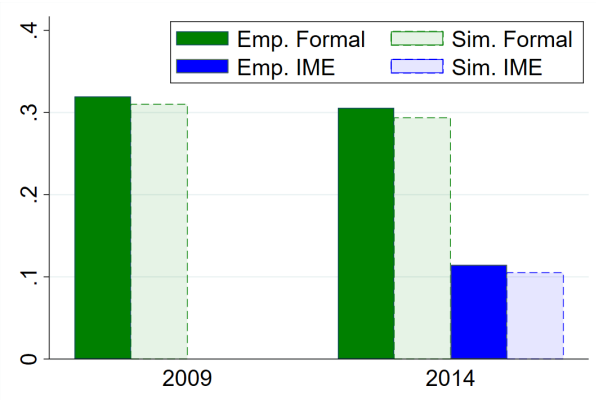
(a) Untargeted pre-income distribution:



(b) Targeted pre-income distribution:



(c) Targeted sector shares:



Source: PNAD 2009 and 2014. Household survey.

I use the estimated parameters to solve for the combination of IME fee and cap that maximizes welfare and to separately check the importance of the simplification and tax reduction

brought by the IME program in increasing formalization. The results are presented in table 10. The total welfare is calculated considering a constant marginal utility of the public good of 1.2, implying that each individual gets their private utility plus 1.2 times the total tax revenue. Column 1 sets the baseline with no IME program. Notice that private welfare can only go up by offering an additional option so that the baseline presents the lowest private welfare. Consistent with the empirical evidence, at the implemented version in phase 2 (low fee, high cap), tax revenue goes up by 3%, which also implies higher total welfare compared to the baseline as presented in column 2. Starting from phase 2, column 3 shows that the optimal policy is achieved by increasing the IME fee from R$36 to R$109 and decreasing the IME cap from R$5000 to R$3243. This leads to an increase of 13% in the tax revenue and total welfare compared to the scenario without the program. The program's size would shrink to approximately 44% of the size observed in the second phase in 2014.

Table 10: Empirical and simulated moments

Simulation	No Program	Implemented	Optimal
Share IME	0	0.105	0.047
Share from informal	0	0.088	0.042
Share from formal	0	0.016	0.004
Tax revenue	6649074.6	6850040.4	7568643.1
Avg. tax revenue	33.2	34.2	37.8
Avg. private welfare	1131.9	1146.3	1136.8
Avg. total welfare	7980021.7	8221194.9	9083508.5
IME fee	0	R$36.2	R$109.5
IME cap	∞.	R$5000	R$3243

I use three counterfactual scenarios to check the importance of the simplification and tax reduction brought by the IME program in increasing formalization and present the results in table 11. First, to check the importance of the compliance cost, I simulate a scenario in which the IME incurs in the *Simples* compliance cost in addition to the IME fee in place in the second phase. The results in column 2 show that the IME share drops from 10.5% to 3.4%. In column 3, I assume a different scenario in which the IME participants remain with no compliance cost but remit a fee of the size of their output tax in case they were in the *Simples*. The share of participants shrinks to 6.4%. In Column 4, I assume that both changes occur simultaneously: IME participants face the same compliance costs and remit a fee equivalent to the tax amount they would remit under the *Simples* regime. Under these conditions, the size of the IME program further decreases 1.7% [40]. By comparing the impact of each change individually, it becomes

[40]Notice that it does not shrink to 0 because the fee does not distort the optimal income level and the idiosyncratic sector preferences.

evident that the reduction in compliance costs under the IME is the primary driver of program enrollment. Eliminating the compliance cost reduction leads to a 3 percentage point larger decrease in program size compared to eliminating the tax benefits alone.

Table 11: Empirical and simulated moments

Simulation	Implemented	Same Compliance	Same tax	Same compliance and tax
Share IME	0.105	0.034	0.064	0.017
Share from informal	0.088	0.029	0.059	0.016
Share from formal	0.016	0.004	0.005	0.001
IME fee	R$36.2	R$36.2 + θ (R$109.5)	t_o Rep. Income Simples	t_o Rep. Income Simples + θ (R$109.5)

10 Conclusion:

Size-based regulation is a policy instrument embedded in many countries' tax systems. It aims to reduce compliance costs and foster business creation while it can also unintentionally create incentives for firms not to grow past the threshold. In developing countries, the potential effect of this type of policy on formal business creation might be enhanced. The lack of enforcement makes small businesses weakly attached to the formal sector and very sensitive to tax liabilities in their non-compliance decision. While policies offering special tax treatment to small businesses have received considerable attention as a means to promote formality, the existing literature lacks sufficient evidence on the impact of size-based taxation on formal business creation, size distortion, and tax revenue in contexts characterized by high informality.

I explore these questions by examining the introduction of a large-scale size-based taxation program implemented in Brazil, in which the key feature was the replacement of all business taxes with a low, fixed monthly tax for micro-businesses below a revenue threshold. This fixed fee relieved participants from the burden of calculating taxes monthly and eliminated the requirement to hire an accountant, which was mandatory under the alternative output tax system. My identification strategy takes advantage of the fact that only small businesses in a subset of industries were eligible to enroll in the new tax program. By leveraging the variation in industry eligibility and the timing of the program's implementation, I conduct a difference-in-difference analysis.

My results show that the new tax option led to a 43% increase in the number of formal firms, primarily by attracting entrepreneurs from the informal sector without significantly impacting transitions from entrepreneurs in other higher-tax formal options. These findings align with an increase in tax revenue, which I further support by showing evidence using tax data. I find that increasing the share of IME firms by 0.01 in the industry is associated with a 0.54% increase in the total tax revenue in the same industry. The correlation remains positive but non-significant after excluding the IME participants, which corroborates that there is no evidence of tax base erosion due to transitions from other formal tax regimes.

I develop a model that formalizes the trade-offs of implementing a fixed fee system and derives sufficient statistics for welfare, which can also be used to interpret my empirical findings. I provide three important results. First, I demonstrate that, as long as the output taxation involves higher compliance costs, finding a fee and an income cap that improve welfare is always possible compared to a pure output system. Second, I show that the intervention in Brazil increased welfare, as a tax increase is sufficient to conclude the program's introduction was welfare-improving in the model. Third, the optimal parameters of the fixed fee tax regime depend critically on the flow of individuals transitioning from informality and the formal output

system into this option. By incorporating the estimated flows into the sufficient statistics, the model suggests that increasing the program's fee toward the end of the sample period would further improve welfare, and that to justify the current income cap, the average size distortion it creates should be at least three times greater than the distortion caused by the current output tax.

I fully characterize and estimate the model using the Simulated Method of Moments (SMM) to determine the optimal IME fee and income cap, and separately assess the impact of simplification and tax reduction on formalization. The optimal policy suggests raising the IME fee from R$36 to R$109 and lowering the income cap from R$5000 to R$3243, resulting in a reduction of more than half in IME enrollment and an increase in the tax revenue by 10% compared to 2014 levels. Moreover, counterfactual simulations reveal that simplification plays a greater role than tax reduction, with the removal of simplification reducing the IME share by 3 percentage points more than eliminating tax benefits.

References

Akcigit, Ufuk, Aghion, Philippe, Gravoueille, Maxime, Lequien, Mathieu, and Stantcheva, Stefanie (2022). "Tax Simplicity or Simplicity of Evasion? Evidence from Self-Employment Taxes in France". *NBER Working Paper* 24049.

Basri, M. Chatib, Felix, Mayara, Hanna, Rema, and Olken, Benjamin A. (2021). "Tax Administration versus Tax Rates: Evidence from Corporate Taxation in Indonesia". *American Economic Review* 111.12, pp. 3827–3871.

Benhassine, Najy, McKenzie, David, Pouliquen, Victor, and Santini, Massimiliano (2018). "Does inducing informal firms to formalize make sense? Experimental evidence from Benin". *Journal of Public Economics* 157, pp. 1–14.

Bergeron, Augustin, Tourek, Gabriel, and Weigel, Jonathan (2024). "The State Capacity Ceiling on Tax Rates: Evidence from Randomized Tax Abatements in the DRC". *Econometrica* 92.

Besley, Timothy and Persson, Torsten (2014). "Why Do Developing Countries Tax So Little?" *Journal of Economic Perspectives* 4.

Best, Michael Carlos, Brockmeyer, Anne, Kleven, Henrik Jacobsen, Spinnewijn, Johannes, and Waseem, Mazhar (2015). "Production versus Revenue Efficiency with Limited Tax Capacity : Theory and Evidence from Pakistan Production versus Revenue Efficiency with Limited Tax Capacity : Theory and Evidence from Pakistan". 123.6, pp. 1311–1355.

De Mel, By Suresh, McKenzie, David, and Woodruff, Christopher (2013). "The Demand for , and Consequences of , Formalization among Informal Firms in Sri Lanka". *American Economic Journal: Applied Economics* 5.2, pp. 122–150.

Dharmapala, Dhammika, Slemrod, Joel, and Wilson, John Douglas (2011). "Tax policy and the missing middle: Optimal tax remittance with firm-level administrative costs". *Journal of Public Economics* 95.9-10, pp. 1036–1047. URL: http://dx.doi.org/10.1016/j.jpubeco.2010.10.013.

Engelschalk, Michael (2007). "Designing a tax system for micro and small businesses: guide for practitioners."

Gordon, Roger and Li, Wei (2009). "Tax structures in developing countries: Many puzzles and a possible explanation". *Journal of Public Economics* 93.7-8, pp. 855–866. URL: http://dx.doi.org/10.1016/j.jpubeco.2009.04.001.

Harju, Jarkko, Matikka, Tuomas, and Rauhanen, Timo (2019). "Compliance costs vs. tax incentives: Why do entrepreneurs respond to size-based regulations?" *Journal of Public Economics* 173.

Hoy, Christopher, Scot, Thiago de, Oguso, Alex, Custers, Anna, Zalo, Daniel, Doino, Ruggero, Karver, Jonathan, and Pillai, Nicolas (2024). "Trade-offs in the Design of Simplified Tax Regimes Evidence from Sub-Saharan Africa ".

Kanbur, Ravi and Keen, Michael (2014). "Thresholds, informality, and partitions of compliance". *International Tax and Public Finance* 21.4, pp. 536–559.

Keen, Michael and Slemrod, Joel (2017). "Optimal tax administration". *Journal of Public Economics* 152, pp. 133–142.

Monteiro, Joana and Assunção, Juliano (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, pp. 105–115.

Rocha, Roberto Hsu and Farias, Alison de (2021). "Formality cost, registration and development of microentreprenuers: Evidence from brazil. Registration and Development of Microentreprenuers: Evidence From Brazil".

Rocha, Rudi, Ulyssea, Gabriel, and Rachter, Laísa (2017). "Do lower taxes reduce informality? Evidence from Brazil". *Journal of Development Economics* 134, pp. 28–49.

Ulyssea, Gabriel (2020). "Informality: Causes and consequences for development". *Annual Review of Economics* 12, pp. 525–546.

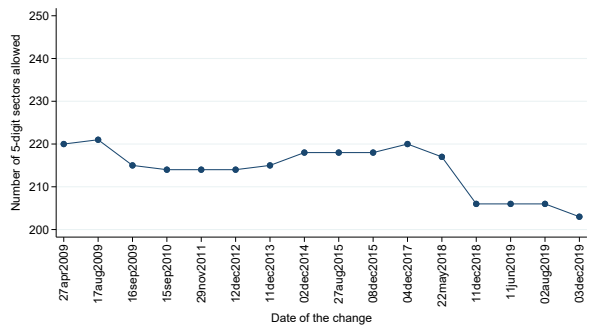
Waseem, Mazhar (2018). "Taxes, informality and income shifting: Evidence from a recent Pakistani tax reform". *Journal of Public Economics* 157.

Appendix:

Appendix - Number of sectors in and out:

Figure 11: Number of sectors allowed by the legislation

(a) 5-digit sectors



(b) 7-digit sectors

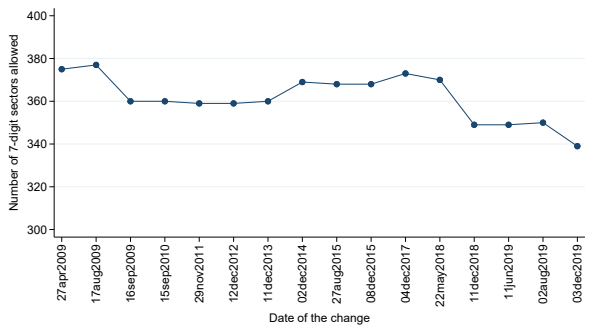
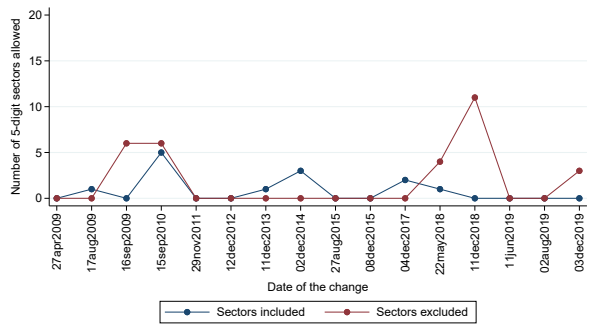
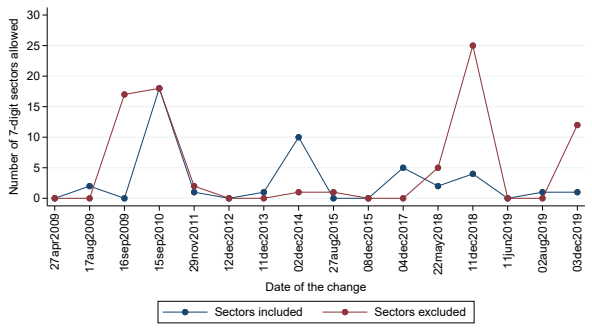


Figure 12: Number of sectors included and excluded by the legislation

(a) 5-digit sectors



(b) 7-digit sectors



Over the years a total of 16 different laws regulated the activity sectors that were eligible to participate in the program. Some sectors went in and out, but the majority of the sectors were allowed by all different laws.

Appendix - Model proofs:

Proposition 1: *Without the IME option, there is an ability level, $a_{I,S}(t_0, \alpha) \geq 0$, such that every entrepreneur with ability equal or above it sorts into the Simples sector.*

Proof:

From the first order conditions, we have i) $c_e(e^*_{Sim}, \alpha) = t_o$, ii) $\psi_Y(Y^*_{Sim}, a) = 1 - t_o$ and iii) $\psi_Y(Y^*_{Inf}, a) = 1 - c_e(Y^*_{Inf}, \alpha)$. Note that a higher ability does not increase the evasion in the Simples, but it increases both the gross income in the Simples and in the informal sector, since $\frac{\partial Y_{Sim}}{\partial a} = \frac{-\psi_{Ya}(Y^*_{Sim},a)}{\psi_{YY}(Y^*_{Sim},a)} > 0$ (from ii) and $\frac{\partial Y_{Inf}}{\partial a} = \frac{-\psi_{Ya}(Y^*_{Inf},a)}{\psi_{YY}(Y^*_{Inf},a)+c_{ee}(Y^*_{Inf},\alpha)} > 0$. Therefore, making the ability low enough to a level a_l eventually implies $Y^*_{Sim}(a_l) = e^*_{Sim}(a_l)$, which is the same income that would be optimal in the informal sector $Y^*_{Sim}(a_l) = e^*_{Sim}(a_l) = Y^*_{Inf}(a_l)$. Using the envelope theorem, it is also possible to show that $\frac{\partial U_{Sim}(a)}{\partial a} = -\psi_a(Y^*_{Sim}, a) > 0$ and $\frac{\partial U_{Inf}(a)}{\partial a} = -\psi_a(Y^*_{Inf}, a) > 0$. Considering the assumption that $\psi_{aY} < 0$, $\frac{\partial U_{Sim}(a)}{\partial a} > (<) \frac{\partial U_{Inf}(a)}{\partial a}$ if $a > (<)\ a_l$. This means that the difference between the utility in the informal sector and the formal sector reaches the maximum at a_l, such that $U_{Inf}(a_l) - U_{Sim}(a_l) = \theta + (\xi_{i,Inf} - \xi_{i,Sim})$. Considering that the formal sector is subject to the restriction that evasion cannot be higher than the income, $Y_{Sim} \geq e_{Sim}$, as long as $\theta + (\xi_{i,Inf} - \xi_{i,Sim}) \geq 0$, there is some $a_{I,S}(t_0, \alpha) \geq a_l$, such that $U_{Inf}(Y^*_{Inf}(a_{I,S}(t_0, \alpha))) = U_{Sim}(Y^*_{Sim}(a_{I,S}(t_0, \alpha)))$. If $\theta + (\xi_{i,Inf} - \xi_{i,Sim}) < 0$, then all entrepreneurs choose the formal sector.

Proposition 2: *Including the IME option implies:*

a) Regardless of the other government instruments, whenever the IME income cap is positive, $Y_C > 0$, there is an ability cutoff level, $a_{I,Ime}(\tau, \alpha, Y_C)$, such that every entrepreneur with ability equal or above it prefers the IME over the informal option.

b) Suppose there is an ability level, a_l, such that it is the lowest ability in which the IME is preferred over the Simples, then there is also an ability cutoff, $a_{Ime,S}(t_o, \tau, \alpha, Y_C) \geq a_l$, such that every entrepreneur with ability between a_l and $a_{Ime,S}(t_o, \tau, \alpha, Y_C)$ prefers the IME over the Simples option, and every entrepreneur with ability above $a_{Ime,S}(t_o, \tau, \alpha, Y_C)$ prefers the Simples over the IME. Note that if $\tau \leq \theta$, then $a_l = 0$.[41]

Proof:

a) Using the envelope theorem, it is also possible to show that $\frac{\partial U_{Ime}(a)}{\partial a} = -\psi_a(Y^*_{Ime}, a) > 0$ and $\frac{\partial U_{Inf}(a)}{\partial a} = -\psi_a(Y^*_{Inf}, a) > 0$. Additionally, irrespective of the features of the market, whenever $Y_C > 0$, $Y^*_{Ime}(a) > Y^*_{Inf}(a)\ \forall\ a$. Both conditions together imply $\frac{\partial U_{Ime}(a)}{\partial a} > \frac{\partial U_{Inf}(a)}{\partial a}\ \forall\ a$. Therefore, exists a ability level $a_{I,Ime}(\tau, \alpha, Y_C)$, such that $U_{Ime}(a_{I,Ime}) = U_{Inf}(a_{I,Ime})$. If

[41] Note that the existence of the ability cutoffs between the two options of the formal sector depends on the parameters of the IME option. Without parameters restrictions, it is possible that the Simples option will always be preferred over the IME (Example: low Y_C and high τ). The opposite, however, is not possible.

$\tau > \xi_{i,Ime} - \xi_{i,Inf}$, $a_{I,Ime} > 0$.

b) Since the income in the IME sector is increasing with ability, $\frac{\partial Y_{Ime}}{\partial a} = \frac{-\psi_{Ya}(Y^*_{Ime},a)}{\psi_{YY}(Y^*_{Ime},a)} > 0$ or $\frac{\partial Y_{Ime}}{\partial a} = \frac{-\psi_{Ya}(Y^*_{Ime},a)}{\psi_{YY}(Y^*_{Ime},a)+c_{ee}(Y^*_{Ime}-Y_C,\alpha)} > 0$, there is an ability level a_l such that $Y^*_{Ime}(a_l) - Y_C = e^*_{Sim}$, implying that $Y^*_{Ime}(a_l) = Y^*_{Sim}(a_l)$. If $a < a_l$ ($a > a_l$), then $Y^*_{Ime}(a) > Y^*_{Sim}(a)$ ($Y^*_{Ime}(a) < Y^*_{Sim}(a)$) and $\frac{\partial U_{Ime}(a)}{\partial a} > \frac{\partial U_{Sim}(a)}{\partial a}$ ($\frac{\partial U_{Ime}(a)}{\partial a} < \frac{\partial U_{Sim}(a)}{\partial a}$), such that $a_l = \underset{a}{\text{argmax}} \{U_{Ime}(a) - U_{Sim}(a)\}$. If $U_{Ime}(Y^*_{Ime}(a_l)) > U_{Sim}(Y^*_{Sim}(a_l))$, since utility in the Simples grows faster with the ability for $a > a_l$, there exists some ability level $a_{Ime,S}(t_o,\tau,\alpha,Y_C) \geq a_l$, such that every entrepreneur with ability between a_l and $a_{Ime,S}(t_o,\tau,\alpha,Y_C)$ prefers the IME over the Simples option, and every entrepreneur with ability above $a_{Ime,S}(t_o,\tau,\alpha,Y_C)$ prefers the Simples over the IME.

Proposition 3: *In case all sectors appear in equilibrium for some IME combination of parameters, there are always two cutoffs $a_{I,Ime}$ and $a_{Ime,S}$, such that entrepreneurs prefer:*

i) The informal option if $a \in (0, a_{I,Ime})$,

ii) The Simples option if $a \in [a_{Ime,S}, ...)$.

Proof:

If all sectors exist, then consider an ability level a^{aux} such that $U_{Ime}(a^{aux}) > U_s(a^{aux})$ for $s = \{Inf, Sim\}$. Considering that $\tau > 0$ and $Y_C > 0$, then $\frac{\partial U_{Ime}(a)}{\partial a} > \frac{\partial U_{Inf}(a)}{\partial a}$ $\forall\, a$ as shown in the proof of 2.a). Therefore, $U_{Ime}(a) > U_{Inf}(a)$ $\forall\, a > a^{aux}$. We know by proposition 2.b that there is a $a_{Ime,s}$ such that $a_{Ime,s} > a^{aux}$, $U_{Ime}(a_{Ime,s}) = U_{Sim}(a_{Ime,s})$, and $U_{Ime}(a) < U_{Sim}(a)$ $\forall\, a > a_{Ime,s}$. Considering that the informal sector exists, then $\theta + (\xi_{i,Inf} - \xi_{i,Sim}) > 0$ and $\tau > \xi_{i,Ime} - \xi_{i,Inf}$. Basically, the informal sector utility should start at a higher level when the ability is 0. Therefore, by propositions 1 and 2.a), there are the two ability cutoffs that define if the utility in the informal sector is higher or lower than in the two formal options, IME, $a_{Inf,Ime}$, and Simples, $a_{Inf,Sim}$. There are possibilities: a) $a_{Inf,Ime} \leq a_{Inf,Sim}$, b) $a_{Inf,Ime} > a_{Inf,Sim}$.

Possibility a) implies:

i) The informal option if $a \in (0, a_{I,Ime})$,

ii) The IME option if $a \in [a_{I,Ime}, a_{Ime,S})$.

iii) The Simples option if $a \in [a_{Ime,S}, ...)$.

Possibility b) implies:

i) The informal option if $a \in (0, a_{Inf,Sim})$,

ii) The Simples option if $a \in [a_{Inf,Sim}, \underline{a_{Ime,S}})$.

iii) The IME option if $a \in [\underline{a_{Ime,S}}, a_{Ime,S})$.

iii) The Simples option if $a \in [a_{Ime,S}, ...)$.

Where $\underline{a_{Ime,S}}$ and $a_{Ime,S}$ are the lower and upper bound ability cutoff between the IME and the Simples.

Proposition 4: *Increasing the tax revenue is a sufficient but not necessary condition for the implementation of the IME program to be welfare increasing.*

No proof is needed.

Proposition 5: *Given that output taxation involves a positive compliance cost, there will always be an IME implementation that improves welfare.*

Proof:

Let $\{a^s_{I,S}\}_s$ represent the set of all ability cutoffs that equalize utilities $U_{Sim} = U_{Inf}$ for different combinations of the sector preferences vector $\boldsymbol{\xi_i}$. Pick the highest ability cutoff, $a^{Max}_{I,S}$, and set the income cap such that $Y_C = Y^*_{Ime}(a^{Max}_{I,S}) - e^*_{Sim}$. This implies that $Y^*_{Ime}(a^{Max}_{I,S}) = Y^*_{Sim}(a^{Max}_{I,S})$, leading to $Y_C = Y^*_{Sim}(a^{Max}_{I,S}) - e^*_{Sim}$. Given that utility in the *Simples* option increases at a faster rate than in the IME for any ability $a > a^{Max}_{I,S}$, the maximum fee corresponds to the point where the individual with ability $a_{I,S}$ is indifferent between the IME and the *Simples* option. This condition is expressed as $\tau^{Max} = t_o Y_C + \theta$, based on the equation $U_{Sim}(a_{I,S}) - U_{Ime}(a_{I,S}) = \tau - (t_o Y_C + \theta)$. Define $a^*(\tau)$ as the ability cutoff satisfying $U_{Sim}(a^*(\tau)) = U_{Ime}(a^*(\tau))$. As the fee is reduced from τ^{Max}, both $a^*(\tau)$ and $Y^*_{Sim}(a^*(\tau))$ increase continuously from $a_{I,S}$ and Y_C, respectively. Therefore, by keeping τ sufficiently close to τ^{Max}, it is possible to ensure $\tau > t_o(Y^*_{Sim}(a^*(\tau)) - e^*_{Sim})$. This guarantees that the inflow into the IME program will enhance tax revenue, thereby improving overall welfare.

Appendix - Results without detrending:

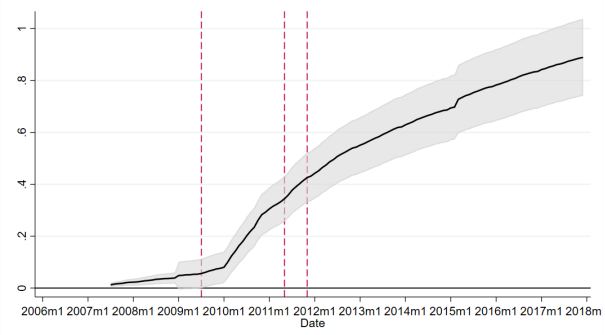
Table 12: IME effects on the total number of firms

	Dep. Var: log(Total number of firms)			
	IRS data		RAIS data	
Treatment	0.5423***		0.1761***	
	(0.0505)		(0.0352)	
Treatment1		0.1615***		0.0810***
		(0.0288)		(0.0189)
Treatment2		0.6459***		0.2021***
		(0.0571)		(0.0400)
Observations	1484707	1484707	1309612	1309612

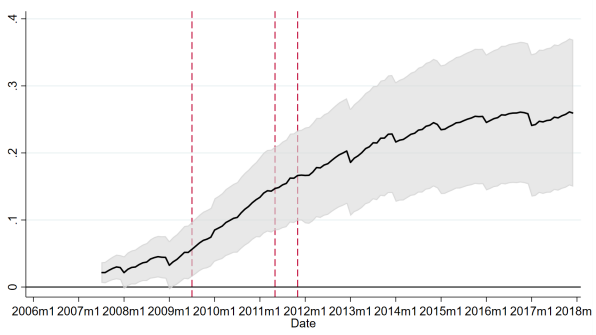
Note: Significance levels: * 10%, ** 5%, ***1%. The results are clustered by industries. All regressions include state, industry, and month/year fixed effects. It also allows for different state fixed effects for treated and control industries by including their interactions. $Treatment1$ is the coefficient that captures the effect of the first phase of the program when the IME tax was 11% of the minimum wage. $Treatment2$ captures the effect for the period when the tax was set as 5% of the minimum wage. All results are estimated using the same time span: from Jan/2006 to Dec/2017.

Figure 13: Effects of the program in the total number of firms over the sample period

(a) IRF data:



(b) RAIS data:



Note: The results are clustered by industries. All regressions include state, industry, and month/year fixed effects. It also allows for different state fixed effects for treated and control industries by including their interactions. All results are estimated using the same time span: from Jan/2006 to Dec/2017. The confidence interval is generated using a 5% significance level. The first vertical line indicates the introduction of the program in 07/2009. All states were included only in 02/2010. The second vertical line represents the tax reduction from 11% to 5% of the minimum wage. The third line shows the period when the income cap was increased from R$36,000 to R$60,000 a year.

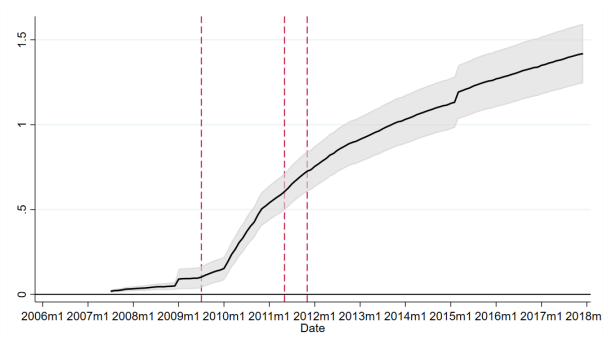
Table 13: IME effects on the total number of firms by legal structure and size

Dep. Var	Legal structure		Revenue size	
	log(Sole proprietorship)	log(Other Legal structures)	log(Micro)	log(Small)
Treatment	0.8958***	0.1004***	0.6502***	0.0900***
	(0.0622)	(0.0323)	(0.0465)	(0.0215)
Average in the control	2.3715	3.0433	2.7766	1.8274
Observations	1287155	1443821	1410611	1048457

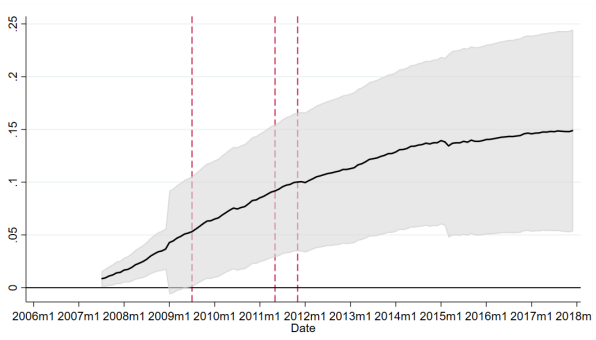
Note: Significance levels: * 10%, ** 5%, ***1%. The results are clustered by industries. All regressions include state, industry, and month/year fixed effects. It also allows for different state fixed effects for treated and control industries by including their interactions. All results are estimated using the same time span: from Jan/2006 to Dec/2017.

Figure 14: Over time effects of the program in the number of firms by legal structure and size

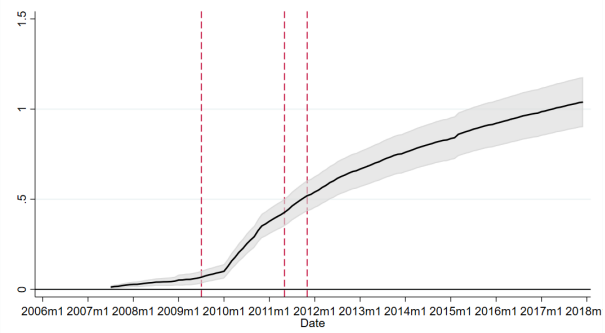
(a) Sole proprietorship:



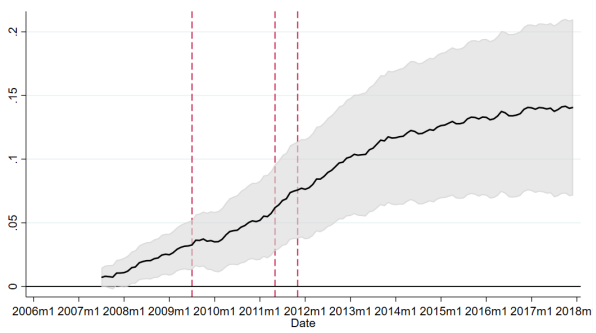
(b) Other legal structures:



(c) Micro firms:



(d) Small:



Note: The results are clustered by activity sectors. All regressions include state, activity sector, and month/year fixed effects. It also allows for different state fixed effects for treated and control activity sectors by including their interactions. All results are estimated using the same time span: from Jan/2006 to Dec/2017. The confidence interval is generated using a 5% significance level. The first vertical line indicates the introduction of the program in 07/2009. All states were included only in 02/2010. The second vertical line represents the tax reduction from 11% to 5% of the minimum wage. The third line shows the period when the income cap was increased from R$36,000 to R$60,000 a year.

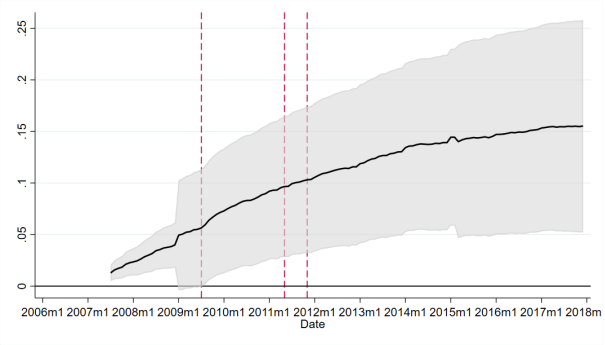
Table 14: IME program effects in the total number of firms not part of the IME program

	Dep. Var: log(Total number of firms not part of IME)					
	All		Sole proprietorship		Micro	
Treatment	0.1024***		0.1701***		0.1415***	
	(0.0347)		(0.0394)		(0.0261)	
Treatment1		0.0599**		0.1109***		0.0792***
		(0.0252)		(0.0286)		(0.0157)
Treatment2		0.1140***		0.1863***		0.1586***
		(0.0375)		(0.0427)		(0.0291)
Average in the control	3.2448	3.2448	2.3715	2.3715	2.7766	2.7766
Observations	1480308	1480308	1273848	1273848	1403580	1403580

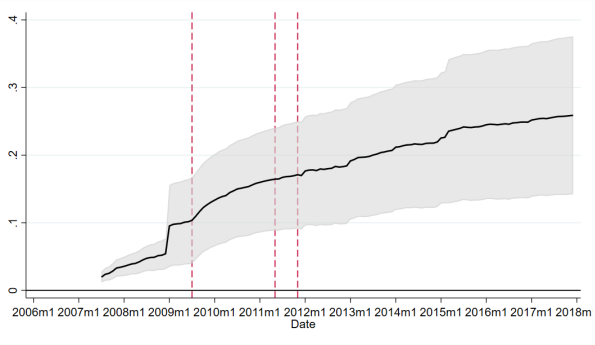
Note: Significance levels: * 10%, ** 5%, ***1%. The results are clustered by activity sectors. All regressions include state, activity sector and month/year fixed effects. It also allows for different state fixed effects for treated and control activity sectors by include their interactions. $Treatment1$ is the coefficient that captures the effect of the first phase of the program, when the IME tax was 11% of the minimum wage. $Treatment2$ captures the effect for the period when the tax was set as 5% of the minimum wage. All results are estimated using the same time span: from Jul/2007 to Dec/2017.

Figure 15: Over time effects of the program in the total number of firms not part of the IME

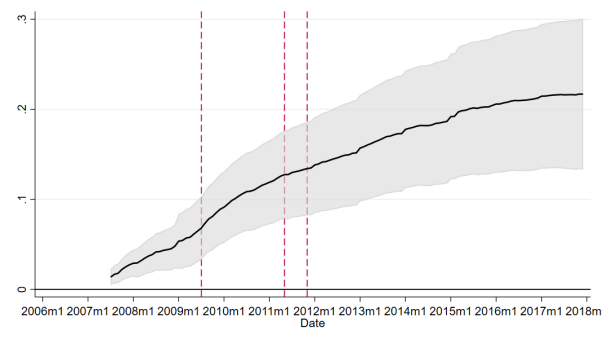
(a) log(Total number of firms not IME):



(b) log(Total number of sole proprietorship firms not IME):



(c) log(Total number of micro firms not IME):



Note: The results are clustered by industries. All regressions include state, industry, and month/year fixed effects. It also allows for different state fixed effects for treated and control industries by including their interactions. All results are estimated using the same time span: from July/2007 to Dec/2017. The confidence interval is generated using a 5% significance level. The first vertical line indicates the introduction of the program in 07/2009. All states were included only in 02/2010. The second vertical line represents the tax reduction from 11% to 5% of the minimum wage. The third line shows the period when the income cap was increased from R$36,000 to R$60,000 a year.

Appendix - Additional results for the informal sector:

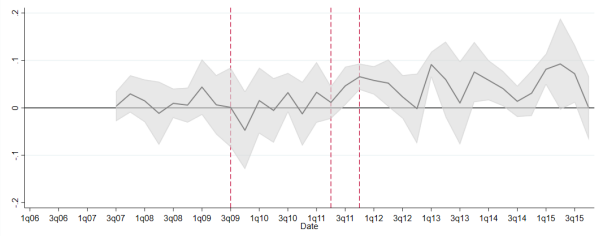
Table 15: IME effects on the number of formal and informal entrepreneurs

Dep. Var	log(N. Self-Emp.)		log(N. 1 Employee)		log(N. Self-Emp.+1 Employee)	
	Formal	Informal	Formal	Informal	Formal	Informal
Treatment1	0.0399	-0.0695	0.0900*	-0.0475	0.0338	-0.0688
	(0.0547)	(0.0526)	(0.0454)	(0.0553)	(0.0472)	(0.0491)
Treatment2	0.2040*	-0.2010***	0.1125**	-0.2069***	0.1871*	-0.2040***
	(0.1152)	(0.0700)	(0.0489)	(0.0349)	(0.1080)	(0.0696)
Observations	19310	24852	9434	9271	19950	25050

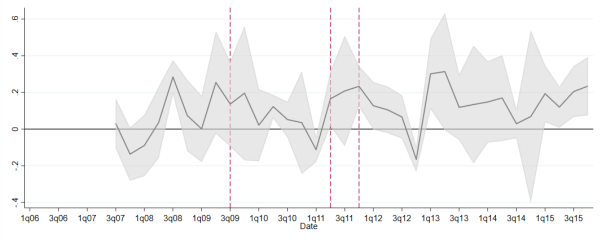
Note: Significance levels: * 10%, ** 5%, ***1%. The results are clustered by industries. All regressions include state, industry, and month/year fixed effects. It also allows for different state fixed effects for treated and control industries by including their interactions. All results are estimated using the same time span: from Jan/2006 to Dec/2015.

Figure 16: Over time effects of the program in the share of formal entrepreneurs

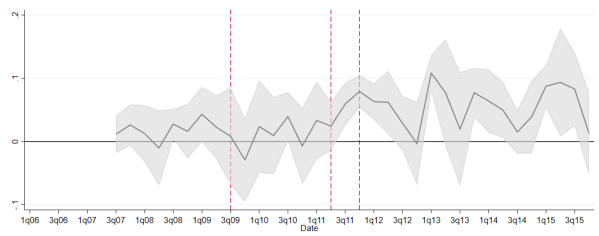
(a) Self-employed:



(b) 1 Employee:



(c) Self-employed and 1 Employee:



Note: The results are clustered by industries. All regressions include state, industry, and month/year fixed effects. It also allows for different state fixed effects for treated and control industries by including their interactions. All results are estimated using the same time span: from Jan/2006 to Dec/2015. The confidence interval is generated using a 5% significance level. The first vertical line indicates the introduction of the program in 07/2009. All states were included only in 02/2010. The second vertical line represents the tax reduction from 11% to 5% of the minimum wage. The third line shows the period when the income cap was increased from R$36,000 to R$60,000 a year.

Table 16: IME effects on the probability of being formal by different income ranges in real terms.

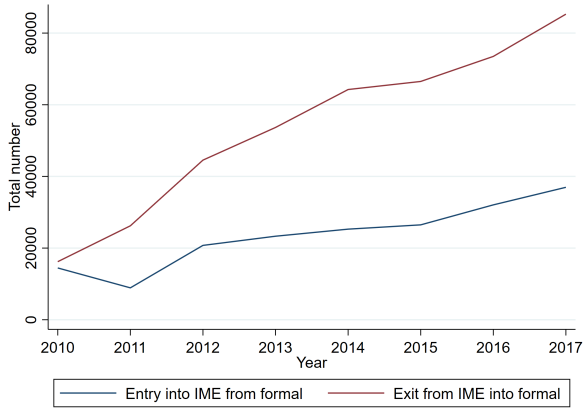
	(1) (0, R$1000]	(2) (R$1000, R$2000]	(3) (R$2000, R$5000]	(4) (R$5000, R$8000]	(5) (R$8000, ...)
Treatment1	0.0251	-0.0209	0.0285	-0.0290	-0.0434
	(0.0243)	(0.0311)	(0.0364)	(0.0278)	(0.0404)
Treatment2	0.0251	0.0217*	0.0482**	-0.0075	-0.0235
	(0.0212)	(0.0128)	(0.0191)	(0.0541)	(0.0420)
Average in the control	0.1051	0.3022	0.4752	0.5415	0.5903
Observations	493057	322594	211506	38883	21447

Note: Significance levels: * 10%, ** 5%, ***1%. The results are clustered by industries. All regressions include state, industry, and month/year fixed effects. It also allows for different state fixed effects for treated and control industries by including their interactions. All results are estimated using the same time span: from Jan/2006 to Dec/2015.

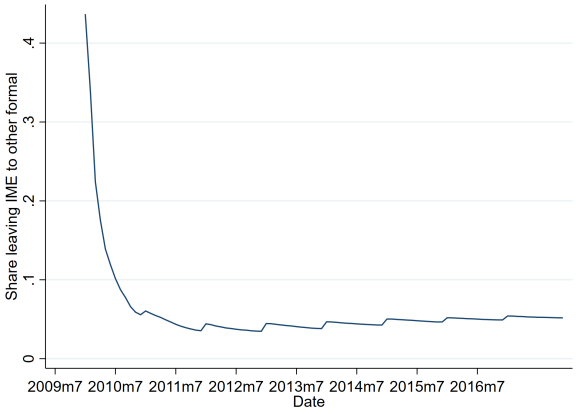
Appendix - Additional results for the formal to formal transition:

Figure 17: Formal to formal transitions

(a) Total number of formal to formal flow:



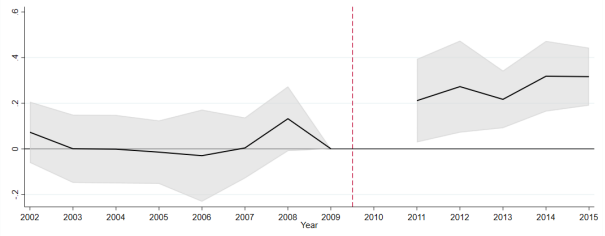
(b) Share leaving IME to formal:



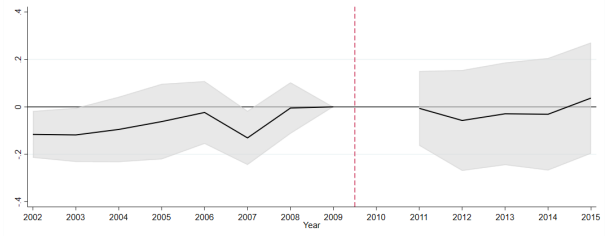
Appendix - results with PNAD at the 5-digit level:

Figure 18: Over time effects of the program in the number of firms in the formal and informal sector - PNAD 5 digits

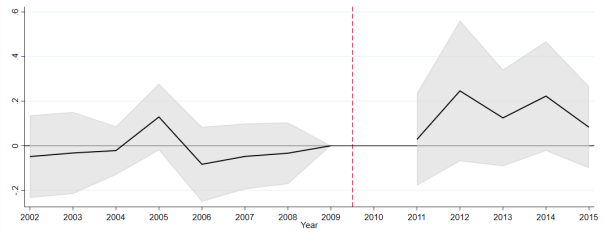
(a) log(Formal Self-employed):



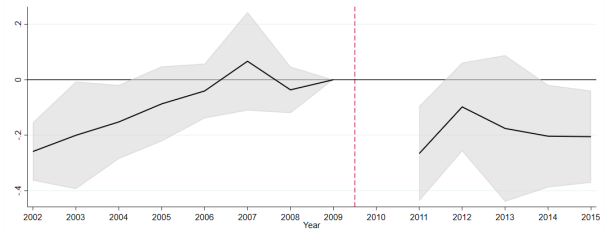
(b) log(Informal Self-employed):



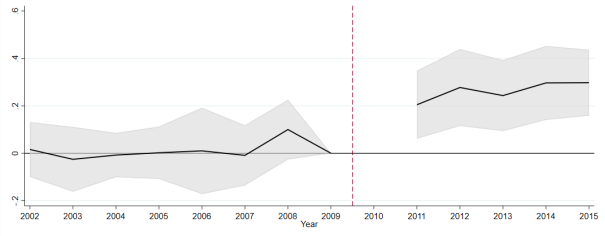
(c) log(Formal Employer):



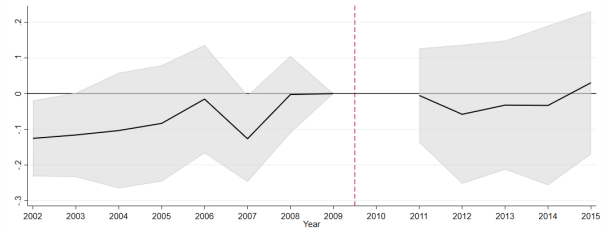
(d) log(Informal Employer):



(e) log(Formal Self-employed and Employer):



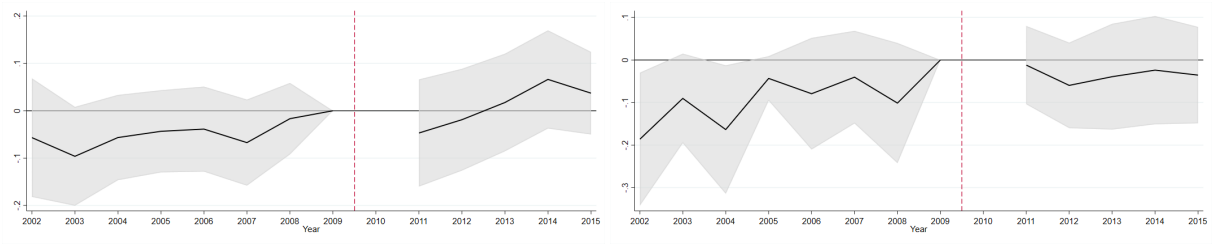
(f) log(Informal Self-employed and Employer):



Note: The results are clustered by industries. All regressions include state, industry, and month/year fixed effects. It also allows for different state fixed effects for treated and control industries by including their interactions. All results are estimated using the same time span: from 2006 to 2015. The confidence interval is generated using a 5% significance level. I consider 2011 the first treated year because the data was not collected in 2010 due to the Census collection. Additionally, the program was introduced in 07/2009 and the 2009 PNAD was collected in 09/2009, only two months after the program was in place.

Figure 19: Over time effects of the program in the number of firms in the formal and informal sector - PNAD 5 digits

(a) log(Formal Employees): (b) log(Informal Employees):



Note: The results are clustered by industries. All regressions include state, industry, and month/year fixed effects. It also allows for different state fixed effects for treated and control industries by including their interactions. All results are estimated using the same time span: from 2006 to 2015. The confidence interval is generated using a 5% significance level. I consider 2011 the first treated year because the data was not collected in 2010 due to the Census collection. Additionally, the program was introduced in 07/2009 and the 2009 PNAD was collected in 09/2009, only two months after the program was in place.

Table 17: IME effects on the number of formal and informal entrepreneurs - PNAD 5 digits

Dep. Var	log(N. Self-Emp.)		log(N. 1 Employee)		log(N. Self-Emp.+1 Employee)	
	Formal	Informal	Formal	Informal	Formal	Informal
Treatment	0.2496***	0.0509	0.1580**	-0.0968	0.2543***	0.0514
	(0.0806)	(0.0905)	(0.0780)	(0.0694)	(0.0783)	(0.0888)
Observations	12143	20071	4862	4815	13284	20587

Note: Significance levels: * 10%, ** 5%, ***1%. The results are clustered by industries. All regressions include state, industry, and month/year fixed effects. It also allows for different state fixed effects for treated and control industries by including their interactions. All results are estimated using the same time span: from 2006 to 2015.

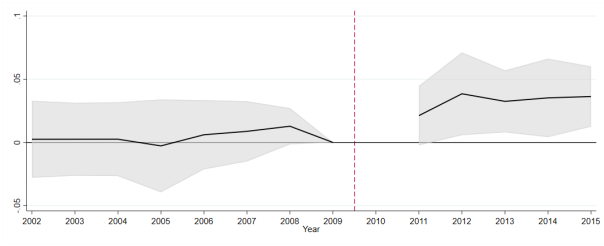
Table 18: IME effects on the number of formal and informal entrepreneurs - PNAD 5 digits Placebo

Dep. Var	log(N. Employees.)		log(N. 2 Employees)		log(N. 3 to 5 Employees)	
	Formal	Informal	Formal	Informal	Formal	Informal
Treatment	0.0577	0.0539	0.0577	-0.0574	0.0374	-0.0523
	(0.0610)	(0.0572)	(0.0623)	(0.0747)	(0.0557)	(0.0591)
Observations	33762	22902	4164	3354	6057	3731

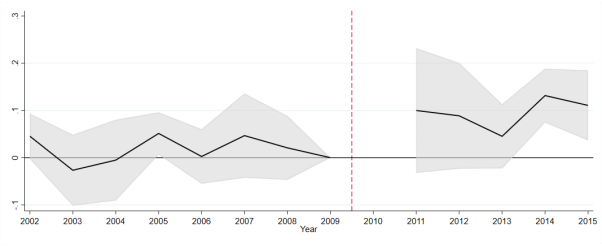
Note: Significance levels: * 10%, ** 5%, ***1%. The results are clustered by industries. All regressions include state, industry, and month/year fixed effects. It also allows for different state fixed effects for treated and control industries by including their interactions. All results are estimated using the same time span: from 2006 to 2015.

Figure 20: Over time effects of the program in the share of formal entrepreneurs - PNAD 5 digits

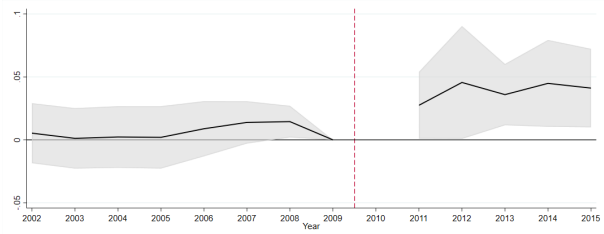
(a) Self-employed:



(b) 1 Employee:



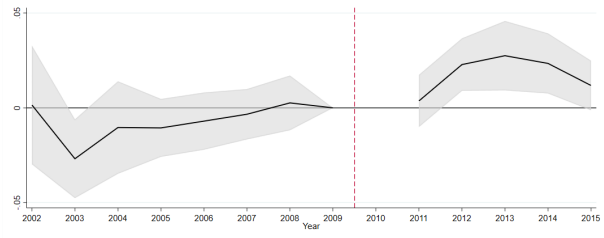
(c) Self-employed and 1 Employee:



Note: The results are clustered by industries. All regressions include state, industry, and month/year fixed effects. It also allows for different state fixed effects for treated and control industries by including their interactions. All results are estimated using the same time span: from 2006 to 2015. The confidence interval is generated using a 5% significance level. I consider 2011 the first treated year because the data was not collected in 2010 due to the Census collection. Additionally, the program was introduced in 07/2009 and the 2009 PNAD was collected in 09/2009, only two months after the program was in place.

Figure 21: Over time effects of the program in the share of formal entrepreneurs - PNAD 5 digits placebo

(a) Employee:



Note: The results are clustered by industries. All regressions include state, industry, and month/year fixed effects. It also allows for different state fixed effects for treated and control industries by including their interactions. All results are estimated using the same time span: from 2006 to 2015. The confidence interval is generated using a 5% significance level. I consider 2011 the first treated year because the data was not collected in 2010 due to the Census collection. Additionally, the program was introduced in 07/2009 and the 2009 PNAD was collected in 09/2009, only two months after the program was in place.

Table 19: IME effects on the probability of being formal - PNAD 5 digits

Dep. Var: 1{Formal=1}	Self-Emp.	1 Employee	Self-Emp.+1 Employee
Treatment	0.0287**	0.0805**	0.0330*
	(0.0140)	(0.0343)	(0.0167)
Average in the control	0.1953	0.6347	0.2202
Observations	302519	19883	322402

Note: Significance levels: * 10%, ** 5%, ***1%. The results are clustered by industries. All regressions include state, industry, and month/year fixed effects. It also allows for different state fixed effects for treated and control industries by including their interactions. All results are estimated using the same time span: from 2006 to 2015.

Appendix - Robustness and heterogeneity:

Table 20: IME effects on the probability of being formal - robustness PME and PNAD

Dep. Var: 1{Formal=1}	Self-Emp.+1 Employee							
Treatment1	-0.0017	0.0013	0.0059	0.0182***	0.0281*	0.0009		
	(0.0240)	(0.0249)	(0.0236)	(0.0037)	(0.0158)	(0.0237)		
Treatment2	0.0447***	0.0474***	0.0334***	0.0394	0.0822***	0.0490***	0.0706**	0.0330*
	(0.0105)	(0.0121)	(0.0090)	(0.0276)	(0.0176)	(0.0119)	(0.0271)	(0.0179)
Time period	2006,2015	2002,2015	2006,2015	2006,2015	2006,2015	2006,2015	2002,2015	2002,2015
Type	Baseline	More years	Control for demographics	No RJ and SP	Control industries with less than 20% eligible for IME	Unemployment rate interacted with the treatment group	PNAD 2-digits	PNAD 5-digits
Observations	1087487	1500236	1087487	648910	1087487	1087487	322402	322402

Note: Significance levels: * 10%, ** 5%, ***1%. The results are clustered by industries. All regressions include state, industry, and month/year fixed effects. It also allows for different state fixed effects for treated and control industries by including their interactions. The first 6 columns are estimated using the PME and the last 2 are estimated using PNAD

Table 21: IME effects on the probability of being formal according to who they sell to

Dep. Var: 1{Formal=1}	Self-Emp.	1 Employee	Self-Emp.+1 Employee
Treatment * Above Median	-0.0423**	-0.0876**	-0.0447**
	(0.0162)	(0.0424)	(0.0175)
Treatment	0.0414***	0.1347***	0.0476***
	(0.0104)	(0.0247)	(0.0101)
Observations	1017235	69412	1086647

Note: Significance levels: * 10%, ** 5%, ***1%. The results are clustered by industries. All regressions include state, industry, and month/year fixed effects. It also allows for different state fixed effects for treated and control industries by including their interactions. All results are estimated using the same time span: from 2006 to 2015. The Above Median is a dummy variable that indicates if the 2-digit industry has an above-the-median share of entrepreneurs that usually sell to the government or a firm. This share is calculated using the 2003 ECINF data.

Table 22: 2-digit sector definitions - part 1

2-digit industry	Total Self-Emp + 1 Employee	Number of 5-digit	Number of treated 5-digit	Share treated by IME	Share formal
Real estate activities	1189	2	0	0	0.301934
Auxiliary financial intermediation activities	500	2	0	0	0.474
Union activities	269	3	0	0	0.223048
Insurance and private pension funds	86	1	0	0	0.453488
Mining or quarrying of metallic minerals	80	2	0	0	0.0125
Basic metallurgy	38	3	0	0	0.31579
Air transport	16	1	0	0	0.375
Electricity, gas, and water	2	1	0	0	0
Manufacture of machinery and equipment for electronic data processing systems	2	1	0	0	0
Fuel production	2	2	0	0	1
Research and development	1	1	0	0	0
Mining or quarrying of non-metallic minerals	202	3	1	0.059406	0.079208
Health and social services	4084	5	1	0.185847	0.536974
Retail and wholesale trade and repair of personal and household goods	52129	21	17	0.522147	0.18506
Manufacture and assembly of motor vehicles	38	3	1	0.657895	0.473684
Education	1924	3	1	0.741684	0.16684
Services primarily to businesses	8920	9	5	0.79787	0.430269
Manufacture of other transport equipment	46	4	3	0.826087	0.26087
Repair and retail of motor vehicles and motorcycles, and retail of fuels	7549	5	4	0.870049	0.232614
Manufacture of chemical products	219	4	2	0.940639	0.141553
Recreational, cultural, and sports activities	3267	8	6	0.99296	0.171105
Manufacture of food and beverages	3368	8	6	0.995546	0.100059
Manufacture of machinery and equipment	304	2	1	0.996711	0.351974
Land transport	12553	4	3	0.99992	0.274994
Construction	30807	2	2	1	0.102736
Personal services	12675	5	5	1	0.133254
Accommodation and food service	12120	3	3	1	0.15363
Manufacture of wearing apparel and accessories	8732	2	2	1	0.136624
Manufacture of furniture and other manufacturing activities	3856	2	2	1	0.121888
Manufacture of textile products	3147	2	2	1	0.047982
Manufacture of metal products - excluding machinery and equipment	1870	2	2	1	0.218717
Information technology activities and related services	1503	2	2	1	0.33666
Auxiliary transport activities and travel agencies	1123	4	4	1	0.182547
Manufacture of wood products	954	1	1	1	0.093291
Publishing, printing, and reproduction of recorded media	612	1	1	1	0.261438

Table 23: 2-digit sector definitions - part 2

2-digit industry	Total Self-Emp + 1 Employee	Number of 5-digit	Number of treated 5-digit	Share treated by IME	Share formal
Rental of vehicles, machinery, and of personal and household goods	584	3	3	1	0.263699
Manufacture of leather goods, travel accessories, and footwear	572	3	3	1	0.152098
Manufacture of non-metallic mineral products	454	3	3	1	0.132159
Manufacture of medical equipments, optical instruments, and industrial automation equipments	333	4	4	1	0.393393
Postal and telecommunication activities	294	2	2	1	0.234694
Manufacture of machinery, appliances, and electrical materials	94	2	2	1	0.234043
Water transport	75	1	1	1	0.093333
Financial intermediation, excluding insurance and private pension funds	68	1	1	1	0.352941
Manufacture of pulp, paper, and paper products	58	2	2	1	0.224138
Manufacture of rubber and plastic products	50	2	2	1	0.46
Manufacture of electronic materials and communication equipment	14	1	1	1	0.357143
Recycling	13	1	1	1	0.153846
Water collection, treatment, and distribution	6	1	1	1	0.333333
Manufacture of tobacco products	5	1	1	1	0
Urban cleaning and sewage; and related activities	3	1	1	1	0.666667

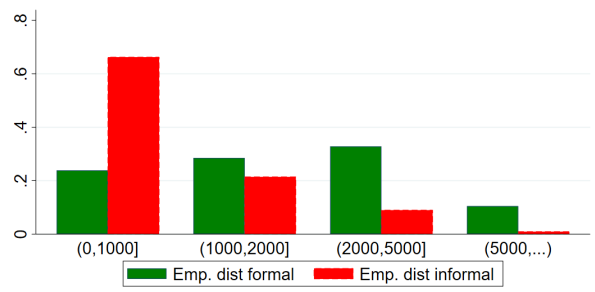
Appendix - Model fit:

Table 24: Empirical and simulated moments

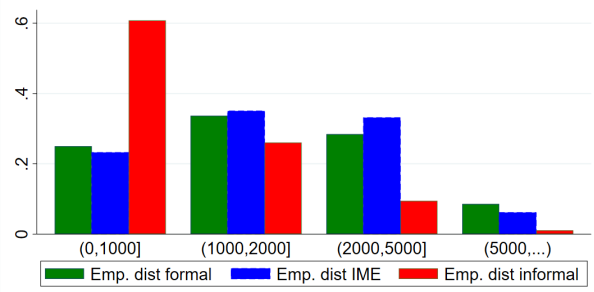
Moment	Simulated	True
Shares in each sector:		
Share of Simples - Pre	.310	.319
Share of IME - phase 1	.063	.026
Share of IME - phase 2	.105	.114
Share of IME from informal - phase 2	.088	.099
Share of IME from formal - phase 2	.016	.013
Average income in each sector:		
Avg. income Simples - Post	2073.65	2680.25
Avg. income Informal - Post	997.31	1184.99
Avg. income IME - Post	2099.53	2468.49
Income distribution in each sector:		
Income dist in Simples - post		
(R$0,R$1000]	.298	.250
(R$1000,R$2000]	.303	.336
(R$2000,R$5000]	.341	.284
Income dist in Informal - post		
(R$0,R$1000]	.582	.608
(R$1000,R$2000]	.355	.260
(R$2000,R$5000]	.062	.094
Income dist in IME - post		
(R$0,R$1000]	.200	.232
(R$1000,R$2000]	.328	.350
(R$2000,R$5000]	.438	.331
Share in each sector by income:		
Share of Simples in each income bracket - post		
(R$0,R$1000]	.190	.164
(R$1000,R$2000]	.264	.344
(R$2000,R$5000]	.545	.477
(R$5000,...)	.830	.659
Share of IME in each income bracket - post		
(R$0,R$1000]	.045	.058
(R$1000,R$2000]	.102	.136
(R$2000,R$5000]	.250	.212
(R$5000,...)	.169	.180

Figure 22: Empirical moments:

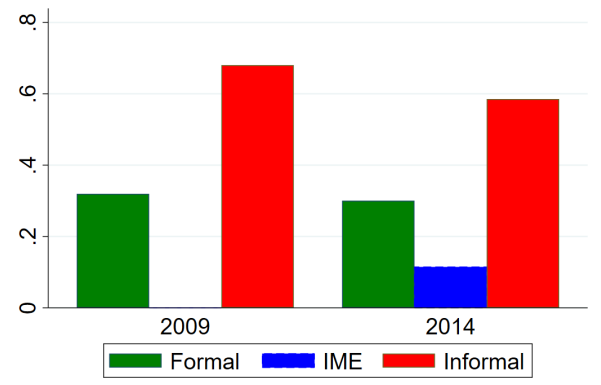
(a) Pre-period empirical size distribution:



(b) Post-period empirical size distribution:



(c) Share in each sector:



Source: PNAD 2009 and 2014. Household survey.