

# Household Enterprises: The Impact of Formality on Productivity and Profits

Nesma Ali and Mohamed Ali Marouani



# **HOUSEHOLD ENTERPRISES: THE IMPACT OF FORMALITY ON PRODUCTIVITY AND PROFITS**

Nesma Ali<sup>1</sup> and Mohamed Ali Marouani<sup>2</sup>

**Working Paper No. 1452**

**December 2020**

This research was supported by the Carnegie Corporation of New York through a fellowship on Economic Vulnerability in the Arab World to the Humphrey School of Public Affairs at University of Minnesota. Useful comments were received at the Economic Research Forum annual conference, the RIEME conference in Edinburgh, the research seminar at Swansea University; and from fellow colleagues at Düsseldorf University. We would also like to thank Yosra El-Chinnawy and Markus Funke for their excellent research assistance.

Send correspondence to:

Nesma Ali

Heinrich-Heine University Düsseldorf

[ali@dice.hhu.de](mailto:ali@dice.hhu.de)

---

<sup>1</sup> Düsseldorf Institute for Competition Economics (DICE), Heinrich-Heine University Düsseldorf.

<sup>2</sup> UMR Développement et sociétés, IRD and Paris 1 Pantheon-Sorbonne University, email: marouani@univ-paris1.fr

First published in 2020 by  
The Economic Research Forum (ERF)  
21 Al-Sad Al-Aaly Street  
Dokki, Giza  
Egypt  
[www.erf.org.eg](http://www.erf.org.eg)

Copyright © The Economic Research Forum, 2020

All rights reserved. No part of this publication may be reproduced in any form or by any electronic or mechanical means, including information storage and retrieval systems, without permission in writing from the publisher.

The findings, interpretations and conclusions expressed in this publication are entirely those of the author(s) and should not be attributed to the Economic Research Forum, members of its Board of Trustees, or its donors.

### **Abstract**

In this paper, we assess the impact of commercial registration of household enterprises on their labor productivity and profits. Based on the 2012 and 2018 rounds of the Egyptian Labor Market Panel Surveys, we employ an instrumental variable strategy and find a positive effect of formal registration on profit and a much higher positive effect on labor productivity. The main channels are higher capital intensity and assets. In addition to the owner's gender and education level, labor productivity varies mainly with firms' assets and shared ownership, while profits are mostly determined by the age of the firm and that of its owner. We also find that the positive effect of formality on performance holds for a select category of owners and firms. Finally, our analysis allows us to identify different policy intervention tools for household firms according to their productivity levels and their distance to a threshold of formality fixed costs.

**Keywords:** Informal sector, Entrepreneurship, Productivity, Egypt.

**JEL Classifications:** O17, O12, D22, L26.

# 1 Introduction

One of the main features of developing economies is a sizable informal sector, which comprises small businesses or self-employed entrepreneurs that legally produce goods and services but evade legal registration, financial reporting, and taxation. Informal employment is also part of the informal sector. Large crises often provide an opportunity to relaunch the debate on the informal sector. On the one hand, governments need resources and the informal sector contributes poorly. On the other hand, informal firms are difficult to support in tough times (such as in lockdowns imposed by the coronavirus disease of 2019) because they are not registered. A rich and nationally representative labor market and household longitudinal survey of Egypt (Assaad and Krafft, 2013), the most populated country in the Middle-East and the third in Africa (100 million), allows us to contribute new ideas to the debate on the causes and consequences of the informal sector. Our two main contributions are to explain the persistence of the informal sector by the gap between private and social benefits of formality and to identify the establishments and entrepreneurs for which the costs of formality matter. We also study the channels through which formality affects firms' performance indicators.

Our paper is related to the literature on the existence and persistence of the informal sector. While Maloney (2004) argues for a new micro-entrepreneurial view of the informal sector, La Porta and Shleifer (2014) focus on the dualism between the informal and formal sectors where the latter is considered the main source of growth and the former is considered unproductive and inefficient. Allen et al. (2018) challenge this strict duality by finding overlapping productivity distributions between formal and informal firms. In their review of competing theories about the informal sector's existence, Rothenberg et al. (2016) add to the dual economy explanation what they call "the exclusion model" and the "rational exit model." The former is based on De Soto et al. (1989) model of burdensome regulations and the second on the rationality of informal firms' response to high entry costs and uncertain benefits. The evidence in Indonesia is in favor of the two latter explanations. Other studies investigate the safety net or last resort role of the informal sector and demonstrate that both voluntary and last resort segments co-exist (Loayza and Rigolini, 2011; Günther and Launov, 2012). Finally, Ulyssea (2018) presents a unifying framework for the main competing explanations on informality (regulatory costs, "parasite" behavior<sup>1</sup>, and survival strategies).

Our paper also belongs to the literature showing formality's positive effects on firms' various outcomes in different countries, which emerged in the past decade (i.e., on profitability in Bolivia: McKenzie and Sakho, 2010; on firms' productivity in West Africa: Benjamin and Mbaye, 2012; on firms' value added in Vietnam: Demenet et al., 2016). McKenzie and Sakho (2010) argue that

---

<sup>1</sup>Firms that are productive enough to be formal but prefer informality to avoid taxes.

the effect of informality is heterogeneous and mostly depends on owners' and firms' characteristics. Amin and Islam (2015) find that small informal firms in Africa outperform the larger ones in terms of productivity. Using data from the same region, Grimm et al. (2012) empirically identify a new segment of informal entities, "constrained gazelles," which are characterized by low levels of capital (similar to firms in the lower tier of firm size distribution) but high levels of returns. They argue that "constrained gazelles" are potentially the best targets for informal sector interventions aimed at boosting growth.

Our paper is also related to the literature on the impact of formal registration on informal firms' performance. In one of the earlier studies, McKenzie and Woodruff (2006) argue that formalization secures better access to credit for firms, and therefore, translates into higher profitability. Fajnzylber et al. (2011) find that formalization increases the revenues and profits for a sample of Brazilian firms. They attribute improvements in firms' performance to a permanent physical location, which facilitates the accumulation of capital and labor hiring. Similarly, Rand and Torm (2012) find that the increase in formal labor contracts is one of the main channels of better performance of firms in Vietnam. Finally, Benhassine et al. (2018) use experimental evidence from Benin to illustrate that formalization costs exceed their benefits, and instead, suggest targeting informal firms that have characteristics closest to formal firms.

The ongoing debate on the causes and consequences of formality <sup>2</sup> brings our attention to how challenging it is to measure and empirically test formality efficiency gains. Most existing studies limit their analysis to a single performance indicator and lack explanation on the main transmission channels. This is mostly due to data availability and the sensitivity of available datasets, which might suffer from measurement errors. Further, very few studies account for the endogeneity of formality (or informality), which arises because of either the causality effects or the fact that the decision to formally operate a firm (or not) mostly depends on firms' and owners' characteristics, which also determine firms' performance.

In this paper, we attempt to address these challenges using a rich panel survey of Egyptian household firms. We investigate the impact of formality on both labor productivity and profit of firms to highlight the trade-off between social and private gains. While increasing labor productivity can be considered as a social benefit, higher profit primarily benefits the owner. We hypothesize that the social benefits of formality are much higher than private benefits, which can explain why informality is so widespread and persistent in developing countries. Our second objective is to identify the characteristics of owners and firms that impact each performance variable the most. The use of an instrumental variable setting allows us to determine the group of complier firms

---

<sup>2</sup>See Ulyssea (2020) for a recent survey.

for which the cost of formality matters the most. Finally, we investigate the channels through which formality affects firms' performance. Moreover, our identification strategy considers formality endogeneity bias by relying on the distance to the registration office as an instrument for formality (following McKenzie and Sakho, 2010). It assumes that the closer is the location of the firm to its tax office where registration occurs, the easier and cheaper are the taxation procedures, the better is the access to information on the registration process and its benefits, and the higher is the probability of the firm getting caught because of tax evasion. Thus, firms located closer to tax offices are more likely to register and operate formally.

We find that formality affects both labor productivity and profit positively, although it has a much larger effect on the former. Labor productivity and profits are not affected by the same set of variables characterizing the firm and its owner. In addition to the owner's gender and education level, labor productivity is mostly determined by the firm's assets and shared ownership, whereas profits are mostly determined by the age of the firm and that of its owner. We also find that the private benefit of formality is limited to low productive firms. As for heterogeneous effects, we find that the positive effect of formality on firms' labor productivity holds for self-employed and mid-sized firms, higher educated male entrepreneurs, and higher capital intensity firms. Furthermore, our findings allow us to classify household firms according to their performance and a hypothetical threshold of formality fixed costs. Finally, we demonstrate that firms' capital intensity and assets are the main channels through which formality increases firms' performance.

The remainder of the paper is structured as follows. Section 2 explores the dataset and the main variables included in the model. Section 3 presents the identification strategy. Section 4 discusses the results and section 5 provides concluding remarks.

## 2 Data and variables

### 2.1 Data

In this paper, we use surveys on non-farm household enterprises extracted from the Egyptian Labor Market Panel Surveys (ELMPS) for 2012 and 2018. The ELMPS is nationally representative and designed by the Central Agency for Public Mobilization and Statistics and the Economic Research Forum.<sup>3</sup> These surveys are considered as the main source of data for the labor market and human development research in Egypt and have served as a model for similar data in Jordan and Tunisia. The 2012 and 2018 rounds focus on over-sampling the poorest villages in Egypt using a stratified

---

<sup>3</sup>The 2018 wave represents the fourth wave of ELMPS. Previous waves (1998 and 2006) are not used in this paper as they do not include information on the formality and performance of non-farm household firms. The dataset is available and downloadable through the Economic Research Forum data portal (OAMDI, 2018): <http://erf.org.eg/data-portal/>.

sample defined at the governorate level (first administrative level), urban\rural areas, and poor versus non-poor villages within rural areas. These surveys provide information at the individual-level, household-level, and household-firm-level (if any).<sup>4</sup>

The data cover a representative sample of 3,765 unique household firms surveyed in 2012 and 2018 (1,975 and 1,790 firms respectively), among which 707 household firms were surveyed in both rounds. This sample accounts for 16% of the total households included in both ELMPS rounds. On average, households report owning one enterprise managed by one member of the household (hereafter “the owner,” who is also the survey respondent). We find that 63.7% of the firms included in the sample are one-person firms (self-employed), 30.7% are micro firms (with 2–4 workers), and 5.6% are small firms (with at least 5 workers).<sup>5</sup> Firms operate in 18 different sectors of activity covering wholesale and retail trade, transportation, and manufacturing. They are distributed over 6 regions, 22 governorates, 242 qism (hereafter, cities), and 709 shyakha (hereafter, town).<sup>6</sup> This dataset provides unique and accurate information on firms’ location in each administrative division in Egypt, which allows us to create an instrument based on the distance to the local tax office where registration occurs. It also provides information on firms’ total number of workers, output, and costs, which we use to compute firms’ productivity and profit.

## 2.2 Main variables

### 2.2.1 Measuring firm-level outcomes

Our main outcome variables are firm-level labor productivity and profit.<sup>7</sup> Consider a household firm  $i$  at time  $t$ ; then, a firm’s labor productivity and profit is measured as follows:

$$Profit_{it} = NE_{it} \tag{1}$$

$$Labor\ productivity_{it} = \frac{VA_{it}}{L_{it}} \tag{2}$$

where  $NE_{it}$  denotes firm monthly net-earnings, which is extracted directly from the survey based on the response to the following question: “What is the average net-earnings of your enterprise per month during the past 12 months?”  $Labor\ productivity_{it}$  denotes a firm’s monthly labor productivity, which is the ratio of the firm’s value added ( $VA_{it}$ ) and its total workers ( $L_{it}$ ), including the

---

<sup>4</sup>See Krafft et al. (2019) for more details on the ELMPS surveys.

<sup>5</sup>The number of workers includes the owner. Workers might be members of the household and/or hired from outside.

<sup>6</sup>In Egypt, qism is the second administrative level after governorate. It refers to a city and can be rural and urban. The third administrative level is shyakha and refers to a town or village. Shyakha can be either rural or urban. For simplification, we will refer to qism and shyakha as city and town, respectively.

<sup>7</sup>The dataset do not allow us to estimate productivity using more advanced methods (e.g., Olley and Pakes) because of the limited number of repeated firms across years and incomplete information on material costs.



owner, regular workers (from inside and outside the household), and unpaid family workers.  $VA_{it}$  is calculated using the sum of the firm’s net-earnings ( $NE_{it}$ ), labor costs ( $L_{it}$ ), and operating capital costs (including taxes) ( $K_{it}$ ).<sup>8</sup>

Information on firms’ labor cost ( $L_{it}$ ) are provided if the firm hires workers from inside or outside the household. To have a more precise measure of a firm’s labor costs, we impute some of these costs to include owners’ and unpaid family workers’ labor costs. To do so, we use unique identifiers to match firms’ owners and unpaid family workers to the individual ELMPS survey, which contains detailed information on their labor market characteristics. We then estimate their hourly wage using data on their age, gender, education level, and year of the survey.<sup>9</sup> We measure a firm’s operating capital costs ( $K_{it}$ ) using the information on the firm’s current capital, which is a categorical variable. Thus, we assign each firm’s capital cost to the mid-point of each category and use the average book value of capital to assign a value for the top-coded category.<sup>10</sup>

### 2.2.2 Measuring a firm’s formality status

A firm’s formality status is our independent variable of interest. According to the Egyptian commercial registry law (no. 34/1976), investment law (no. 72/2017), and tax law (no. 91/2005), starting a formal business in Egypt requires the registration of a firm in the commercial registry. This registration can only be approved after obtaining a tax identity card (hereafter, tax ID), which must be requested at the public tax office in the city where the firm is located. This step is compulsory for all firm types and activities and even in the case of tax exemption. A firm’s commercial registration expires every five years, and its renewal is ensured if the firm regularly declares and pays its annual taxes (if any). Hence, a registered firm (what we observe in the data) is a firm that has already acquired a tax ID from its corresponding tax office, respects the annual tax declaration, and regularly pays its taxes. Besides, business licensing depends on the firm’s type of activity and is a third step after the firm’s acquisition of the tax ID and official registration.

Accordingly, a firm’s commercial registration is the best proxy measure of its formality status and can be directly extracted from the response to the following survey question: “does the enterprise have commercial registration?” Based on this question, we generate a “formality” dummy variable, which is considered as our main independent variable throughout the estimation process. To overcome potential endogeneity biases, in the next section, we explain how we instrument this

---

<sup>8</sup>All monetary values are in Egyptian pound and deflated using corresponding year’s consumer price index (CPI).

<sup>9</sup>In case of missing values, we assign a total of 160 hours per month, which corresponds to an average of 8 working hours per day.

<sup>10</sup>For example, we assign a capital cost of EGP 250 if the current capital of the firm is coded between EGP 1–499 and a value of EGP 750 if it is between EGP 500–999. For the top-coded category (EGP 50,000 and more), we assign the average book value of capital, which corresponds to EGP 83,600.

variable using the distance of the firm to its corresponding tax office.

Furthermore, the survey provides two other questions that can be used for more restrictive definitions of firms' formality status: "do you have a business license?" and "do you keep regular accounting books?" Only 18% of the firms in the sample report keep accounting books in addition to having a commercial registration and a business license.<sup>11</sup> Thus, we decide to use these "more restrictive definitions" in our robustness check estimations to test the validity of our main results. Moreover, the dataset provides information on the owner's formality status by investigating the incidence of his or her work social insurance. This information allows us to partly investigate the intensive margin of informality (informal employment) as recently illustrated by Ulyssea (2018), who demonstrate the importance of including this intensive margin when tackling firm informality effects (the extensive margin).

### 2.3 Stylized facts on the Egyptian formal/informal sector

As in many other developing countries, informality is becoming the new normal in Egypt, accounting, on average, for 34% of gross domestic product between 2004 and 2015 and 49.6% of total non-agriculture activities in 2012 (Medina and Schneider, 2018; Charmes, 2012), thereby ranking Egypt among the highest in the contribution of the informal economy in North Africa. Formal and informal firms in Egypt do not operate in strictly separated industries. They rather co-exist in the same industry showing a great overlap in productivity. Figure 1 shows that in aggregate over all industries in our sample, formal firms are usually more productive than informal ones. Nevertheless, some informal firms have the same level or even higher productivity than formal firms. Figures 2 and 3 indicate that the productivity overlap remains significant within industries and tend to be greater in more advanced industries with higher fixed costs (manufacturing vs. wholesale and retail industries). This evidence is in line with the theoretical and empirical models of Ulyssea (2018) and Allen et al. (2018), supporting the significance of this sector and its persistence in the economy.

Various factors, such as the lack of enforcement of many stringent economic reforms, the failure of main institutions supporting the formal market (i.e., competition authority, labor law, and tax system), and the embeddedness of corruption and cronyism, help nourish the prevalence of informality in Egypt. In addition, economic crisis and political instability (i.e., 2008 economic crisis and 2011 Egyptian revolution) helped shape the path toward a more prominent informal sector, which became the only resort for many new entrants. Barriers to formal jobs encouraged many people to start small activities. According to the 2013 Egyptian economic census, micro firms (less than 10

---

<sup>11</sup>Keeping accounting books is a burdensome and bidding constraint for small firms in Egypt, especially informal firms

employees) represent 96.9% of total non-agricultural establishments and 55.9% of total employment. Yet, most of the jobs are being created in informal micro firms rather than formal ones. We can notice a similar pattern in our dataset—63.7% of the firms included in our sample are self-employed, among which, only 25% are formal firms, while 36.3% are micro and small-sized firms (2–4), among which, 48% are formal.

Tables 1 and A1 indicate that formal firms are significantly better off than informal firms in terms of many economic indicators. They are, on average, bigger in size, more productive, and profitable. They are also older in age and tend to favor shared ownership over sole proprietorship and accumulate assets. Owners of formal firms are mostly male, older in age, with a work social insurance and a significantly higher education level<sup>12</sup>. Despite these positive indicators, firms’ transition into formality between 2012 and 2018 is very limited, as shown in Table 2. Only 16% of firms became formal, against 15% that became informal.<sup>13</sup> Although most firms remain informal (40% vs. 28% that remain formal), they are the most disadvantaged group, especially in terms of the distance to the tax office, which is double that of the firms that remain formal (Table A2). This evidence indicates that the distance to the tax office is considered as one of the main determinants of the formalization cost. The longer is the distance, the harder is the access to information on the registration process, and the higher are the costs of entering the formal sector as well as the ongoing costs of staying formal. As shown in Ulyssea (2020), the latter costs are considered as the main factors that push firms out of formality in developing countries.

These descriptive statistics also underline the existing heterogeneity within the informal sector, which includes not only low-skilled unproductive firms that will never be able to formalize but also productive firms that are kept out of formality because of high formalization costs (De Soto’s firms) and firms that are productive enough to become formal but voluntarily choose to remain informal to earn higher profits from the cost advantage of not complying with taxes and regulation (parasite firms)<sup>14</sup>. Hence, the decision to operate as an informal entity is no longer made exclusively out of necessity but is becoming an attractive opportunity, which explains why the effect of informality can be very heterogeneous and largely depends on owners’ and firms’ characteristics.

---

<sup>12</sup>In case of shared ownership, the characteristics of the owner refers to the main owner, who is also the manager of the firm and the respondent to the questionnaire.

<sup>13</sup>These percentages are based on a total of 707 firms that were interviewed in both rounds of the survey.

<sup>14</sup>The terms “survival firms,” “De Soto’s firms,” and “parasite firms” have been used in many studies to characterize the different groups of firms composing the informal sector (see, for example, La Porta and Shleifer, 2014 and Ulyssea, 2020).

### 3 Empirical strategy

#### 3.1 Baseline specification

Our empirical strategy aims to identify the effect of firms’ formal registration on their performance. We start by analyzing the following regression using ordinary least squares (OLS) estimation.

$$y_{it} = \alpha_0 + \alpha_1 \text{formality}_{it} + \alpha_2 X_{it} + d_t + \gamma_j + \theta_g + u_{it} \quad (3)$$

where  $y_{it}$  is the logarithm of firm-level labor productivity or profits measured as explained in section 2.2.1;  $\text{formality}_{it}$  is a dummy variable that refers to a firm’s formality status and equals one if the firm is formal; and  $X_{it}$  is a vector of firm and owner characteristics such as a firm’s size, age, ownership status, and assets, and owner’s age, education level, and gender.<sup>15</sup> We include dummies for the year of the survey ( $d_t$ ), the firm’s sector of activity ( $\gamma_j$ ), and governorate ( $\theta_g$ ) to control for year, sector, and governorate specific effects.<sup>16</sup> We also cluster our standard errors at the town-level.

Our baseline specification assumes that firms’ formality status is exogenous to firms’ performance and conditional on firm-level controls and fixed effects. Nevertheless, potential endogeneity bias might arise because of the reverse causality between a firm’s formality and its performance level. Formality may lead to higher productivity and profits. Further, better performance may induce firms to formalize. Moreover, potential omitted variable and self-selection biases could alter our results because the firm’s decision to operate formally and to perform better depends on the characteristics of the firm and its owner (i.e., the firm’s sector of activity and size, and the owner’s gender and education level).

To overcome these biases, we instrument the formality status of the firms and control for various relevant observable characteristics. We use a distance instrument in the spirit of McKenzie and Sakho (2010), who instrumented formality using the distance of the firm to the tax office where registration occurs.

#### 3.2 Instrumental variable approach

Egypt is divided into three administrative levels. The first level divides the country into 27 major areas called “governorates.” The second level divides each governorate into multiple cities called “qism.” Finally, the third level divides each city into multiple towns called “shyakha.” Firms’

---

<sup>15</sup>The owner of the firm should be at least 21 years old to register a firm. Thus, we restrict our sample to this age limit, resulting in dropping 95 observations from our analysis. This sample restriction does not bias our final results.

<sup>16</sup>Only 18% of surveyed firms have been interviewed over the 2 rounds of the survey, firms’ transition in formality status is very limited and is subject to measurement error. Therefore, the panel structure of the sample cannot be used. The data are treated as a pooled cross-section with year fixed effects.

registration and tax collection occur at public tax offices that are spread across the country at the city-level. The dataset provides information on the location of the firm at the narrowest administrative level—the town-level. Firms located in a given town are assigned to a given tax office in the city and are not allowed to register at any other tax office located in another city. With this information in hand, the distance between the centroid of the town where each firm is located and the exact address of its corresponding tax office in the city is used as an instrument for a firm’s formality status, as shown in figure 4.<sup>17</sup> This distance instrument variable is measured in kilometers and based on the longitudinal and latitudinal coordinates of each location. Based on this instrument, we implement a two-stage least squares (2SLS) estimation assuming the following: 1) firms’ distance to their corresponding tax offices in the city determine their formality status; and 2) firms’ distance to their corresponding tax authorities has no direct effect on their performance indicators (labor productivity and profits).

According to our first assumption, and as previously mentioned, firms’ registration (formality) is applicable after obtaining a tax ID, which is requested at the tax office located in the city.<sup>18</sup> Therefore, the closer is the firm to its corresponding tax office, the higher is its probability to acquire a tax ID and register the firm. One can also argue that the closer is the firm to its corresponding tax office, the lower are the fixed costs of registration, the easier is the access to information on registration and taxation procedures, and the higher is the probability of getting detained due to tax inspection. Nevertheless, one might think that the second assumption is violated because of urban agglomeration in Egypt. Most firms are located in the capital city and big cities surrounding the capital, where most of the demand originates due to higher population density (The World Bank, 2012). This means that more productive firms would be located in the core (residential and industrial zones with sound infrastructures), while less productive firms would be located near the periphery. The level of urban agglomeration in Egypt has led to strong diseconomies of scale (i.e., city congestion, poor infrastructure, informal settlements), which has pushed firms to diversify their location choice.<sup>19</sup> Our data support the evidence on firms’ diversification. High or low productive firms are not exclusively concentrated in some specific governorates, cities, or towns. Rather, they are distributed across different locations regardless of their productivity levels. Moreover, some of the most productive firms are located exactly in the same governorate/city/town levels as the least

---

<sup>17</sup>The exact addresses of public tax offices in each city are extracted from the Egyptian tax authority website (<http://www.incometax.gov.eg/map.asp>). In 10 cases, the corresponding coordinates were not found. For these cases, we use the coordinates corresponding to the centroid of the city.

<sup>18</sup>In 2017, the government created a special unit for the development of micro, small, and medium enterprises. This unit facilitates a firm’s creation through a one-stop-shop that allows a firm to obtain the tax ID and the commercial registration from a single place. Yet, this unit mostly concerns projects receiving social funds (which corresponds to four enterprises in our sample) and is not as well spread geographically as tax authorities (almost one unit per governorate).

<sup>19</sup>See Krugman (1991) and Ellison et al. (2010) for a full review on urban agglomeration and firms’ performance.

productive firms.

To fully ensure that our second estimation assumption is validated, we test the correlation between firms’ distance to their corresponding tax offices and market characteristics defined at the sector and year levels. Specifically, we regress the firms’ distance to their corresponding tax offices in the current year on the current and lagged levels of the market size (captured by the log of firms’ profits), capital intensity (captured by the log ratio of capital costs to labor costs), and population density. Using an OLS estimation, Table 3 shows that none of these measures are significantly correlated with firms’ distance to their tax offices. Nevertheless, in our 2SLS estimations, we control for the intensity of economic activity in each town using night-lights data extracted from the Earth Observatory group.<sup>20</sup> The latter is normalized with respect to the brightest town per city to obtain a variable between 0 (towns with the lowest level of economic activity) and 1 (towns with the highest level of economic activity). We also weight this variable using population density in each town.<sup>21</sup> Therefore, for two different towns with an equal level of economic activity and population density, firms located in the town closest to the tax office are more likely to operate formally.

Hence, the effect of firms’ formality on their labor productivity or profits is estimated using the following equation:

$$y_{it} = \beta_0 + \beta_1 \text{formality}_{it} + \beta_2 \text{EcoActivityPop}_s + \beta_4 X_{it} + d_t + \gamma_j + \theta_g + u_{it} \quad (4)$$

where  $y_{it}$  is the logarithm of firm-level labor productivity or profits measured as explained in section 2.2.1;  $\text{formality}_{it}$  is the independent variable of interest and refers to the formality status of firm  $i$ , which is estimated in a first step using  $\text{distance}_{s,k}$  as an instrument; and  $\text{distance}_{s,k}$  is the logarithm of the distance in kilometer between the centroid of the town  $s$  where the firm is located and its corresponding tax office address in the city  $k$ . The variable  $\text{EcoActivityPop}_s$  proxies for the town-level economic activity weighted by population density. The vector  $X_{it}$  represents firms’ and owners’ characteristics, including firms’ age, size, assets, ownership status, and owner’s age.<sup>22</sup> We also include dummies for the year of the survey ( $d_t$ ), the firm’s sector of activity ( $\gamma_j$ ), and governorate ( $\theta_g$ ) to control for year, sector, and governorate specific effects. Further, we cluster our standard errors at the town-level.

---

<sup>20</sup>Several papers have used the night-lights data to measure economic activity or gross domestic product (e.g., Henderson et al., 2011 and Storeygard, 2016). These data are derived from a digital number ranging between 0 and 63, where 0 represents the least bright point (i.e., the lowest level of economic activity) and 63 represents the brightest point (i.e., the highest level of economic activity). The data are downloadable through the following link: <https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html>

<sup>21</sup>Data on population density in each town are extracted from the Central Agency for Public Mobilization and Statistics.

<sup>22</sup>We remove owners’ gender and education levels because of the sensitivity of our 2SLS estimation to the inclusion of these variables, as explained in section 4.2.1.

## 4 Results

This section presents the effect of firms' formality on their labor productivity and profits. We start by analyzing the results of our baseline OLS specification. Next, we present the 2SLS specification results. In this subsection, we, first, emphasize the relevance of our distant instrument in estimating a firm's formality status. We then highlight the trade-off in formality efficiency gains and formality heterogeneous effects that explain the differences between the OLS and 2SLS specification results. We end this section by determining the potential channels of transmission explaining the effect of formality on firms' labor productivity and profits.

### 4.1 Baseline specification

Table 4 presents the effect of firms' formality on their labor productivity (columns 1–3) and profits (column 4) using OLS estimation. In column (1), we exclude all explanatory variables and fixed effects. In column (2), we add fixed effects for years, sectors of activity, and governorates, and account for firms' characteristics. We then add the owner's characteristics in columns (3) and (4). The results indicate a significant positive relationship between firms' formality and the outcomes of interest in all specifications. Operating formally induces a significant increase in firms' labor productivity by more than 70% and in profits by more than 30%. These positive effects are expected and are in line with most of the literature studying formality effects on firms' performance indicators in different countries (e.g., on firms' profitability in Latin America: McKenzie and Sakho, 2010; on firms' productivity in West Africa: Benjamin and Mbaye, 2012; on firms' value added in Vietnam: Demenet et al., 2016).

However, jointly estimating the effect of formality on both labor productivity and profits of firms offers some interesting new insights. On the one hand, labor productivity and profits are not affected by the same set of variables characterizing the firm and its owner. In addition to the owner's gender and education level, labor productivity is mostly determined by the firm's assets and shared ownership, whereas profits are mostly determined by the age of the firm and that of its owner. On the other hand, the resultant increase in labor productivity due to formality is more than twice the resultant increase in profits. This suggests the existence of a large gap between the social and private benefits of formality, which can be explained by the other determinants of firms' labor productivity and profits.

Column (4) illustrates that the 30% increase in profits due to formality can be attained by an informal firm that is operated by a younger male educated entrepreneur. This is quite plausible as 63% of firms with these underlined characteristics in our sample are informal. Hence, for a certain

group of entrepreneurs, operating informally can still be beneficial if the firm’s main target is profit gains. These are the entrepreneurs who have the prerequisites to operate formally but choose not to do so to benefit from tax reductions.

However, this is not the case for the 70% increase in labor productivity, which cannot be attained unless the firm has assets and better access to finance through shared ownership (column 3). These conditions are almost impossible to achieve for an informal firm that has no collateral and seeks to remain hidden from authority. Therefore, for the same group of entrepreneurs, operating informally is no longer beneficial if their firms’ main target is labor productivity gains. Even though informality can bring some private benefits (increase in profit), these benefits are not sustainable as informality prevents firms from accessing formal sources of finance and realizing economies of scale and scope. By contrast, formality will induce higher productivity levels, which will translate into higher investment capacity (through better access to finance and accumulation of assets) and higher profits in the long-run.

## **4.2 Using exogenous variation from a firm’s distance to the tax office**

As already explained in section 3, the results of our baseline specification should be interpreted with caution because of potential endogeneity biases. In the next subsection, we control for these biases by implementing an instrumental variable approach that uses firms’ distance to its corresponding tax office as an instrument for firms’ formality status. This exercise also allows us to shed light on the trade-off in formality efficiency gains and on some of the main formality heterogeneous effects.

### **4.2.1 Firm’s formality status and distance to the tax office**

In our empirical strategy, we first start by demonstrating the extent to which a firm’s distance to its corresponding tax office would determine its probability of operating formally. Table 5 reports the average marginal effects from probit regression in which the firm’s formality status is the dependent variable. We show variants of this first stage specification by adding some characteristics at the firm and owner levels. Column (1) indicates a significant negative relationship between the firm’s distance to the tax office and its probability of operating formally. The firm’s willingness to acquire a commercial registration and operate formally drops by 0.7 percentage point if the firm’s distance to the tax office increases by 1 km from the mean (i.e., 7 km away). This negative effect remains valid when controlling for the characteristics of the firm and its owner in columns (2) and (3).<sup>23</sup> In addition to the firm’s distance to the tax office, column (2) indicates that firm and owner characteristics matter

---

<sup>23</sup>This negative effect remains also valid when setting the distance to the tax office to different cut-offs, as shown in Table A3.



significantly in determining the firm’s willingness to formalize. The latter increases by almost 20 percentage points if the firm hires more labor, by 16 percentage points if the firm favors shared ownership over sole proprietorship, and by 4 percentage points if the firm has assets. Regarding the owners’ characteristics, the older they are and the higher their education levels, the higher is their willingness to operate formally. Further, male owners are relatively more susceptible to register their firms compared to their female counterparts.

These results confirm our identification assumptions and are in line with most of the studies. The meta-analysis of Ulyssea (2020) showed that better access to information on the registration process and its benefits, as well as the reduction of the costs of entering the formal sector and the ongoing costs of remaining formal, are the most effective methods to encourage entrepreneurs to operate formally. According to our results, this could be well achieved by locating the firm closer to the tax office, which will reduce the fixed costs of formalization and ease access to information. We also show that increasing the size of the firm as well as its funding capacity through shared ownership (or through other means such as access to loans and credits) are among the main factors explaining a firm’s decision to operate formally in developing countries (as shown in La Porta and Shleifer, 2014). This decision is also subject to male dominance in the market and his ability to build a strong reputation over the years and increase his understanding capacity of formalization procedures and benefits through education. Yet, our specification might be sensitive to some owner characteristics (e.g., education level and owner’s gender), which might also be highly collinear and endogenous to our second stage regression. Therefore, we choose to not include these variables in our second stage regressions.

#### **4.2.2 Second stage results: Trade-off in formality efficiency gains**

Table 6 reports the estimated effects of firms’ formality on their labor productivity (columns 1–2) and profits (column 3), using the 2SLS specification with fixed effects for the year, sector of activity, and governorates. The results indicate a significant positive effect of formality on firms’ labor productivity in both specifications (with or without firm and owner characteristics) and on firms’ profits. The results remain very similar to the OLS baseline specification, except for the effect of firms’ ownership on their labor productivity, which loses its significance.

We can once again notice the gap between the social and private benefits of formality. The estimated effect of formality on firms’ productivity is almost twice its effect on firms’ profits. To pin down this gap, we re-estimate these regressions by disentangling the labor productivity gains according to the level of profit, and the profit gains according to the level of productivity. Table 7 shows that the effect of formality on a firm’s labor productivity remains positive and significant

regardless of the firm’s profit level. In other words, the social benefit of formality (i.e., increase in labor productivity) is not limited to a certain level of profit.

By contrast, Table 8 shows that the private benefit of formality (i.e., increase in firms’ profits) is actually limited to low productive firms. This means that for high productive firms, operating formally or informally is inconsequential in terms of profit gain because the percentage of profit loss due to informality shrinks with higher productivity levels (as shown in Allen et al., 2018). Therefore, informality might be beneficial in terms of profit gains only if the firm is productive enough, which is quite implausible in the first place as informality tends to lower firms’ productivity. Nevertheless, informal firms are common in developing countries because of the high levels of corruption and cronyism that allow them to remain informal despite being productive enough to operate formally (i.e., parasite firms according to (Ulyssea, 2020)).

The reported first-stage-F-test confirms the validity of our distance instrument. Thus, we can be confident about the interpretation of the underlined results. Nevertheless, the large gap in magnitude between the OLS and 2SLS coefficients might raise some concerns. We explain that these coefficients show different effects. The OLS reports the average treatment effect, while the 2SLS reports the local average treatment effect, which is the causal effect for the group of compliers (i.e., firms that are formal just because they are located closer to the tax office). According to Carneiro et al. (2011), this difference mostly arises when the treatment effect is heterogeneous (i.e., the impact of formality differs across firms) or due to measurement errors. One might think that systematic measurement error is the reason as people usually tend to hide the informality status of their firms. However, the data collection process along with the first estimated results allows us to believe that this difference is instead due to heterogeneous effects.

On the one hand, the data used in the paper is primarily from a labor market survey administered for individuals and households. Among the different sections of this survey, there is a form dedicated to households owning a firm, and the different questions dealing with “Household Firms” emerge only at that point, implying that at this point, the respondent will be more confident that investigating about taxation or formality for formal authority purposes is not the aim of the survey. Further, data collectors are well trained in addressing these types of sensitive questions. Questions on the formality status of the firm are rather indirect questions on the acquisition of commercial registration, licensing, or accounting books. Our results also show that the effect of formality remains the same when using a more restricted definition of formality, which rules out the possibility of measurement error. Table A4 demonstrates that our 2SLS results remain valid to the use of the following measures: the firm is registered at the commercial registry (column 1), the firm is registered at the commercial registry and has a business license (column 2), and the firm is registered

at the commercial registry, has a business license, and keeps accounting books (column 3). Column (4) adds the intensive margin of formality to our baseline definition. Here, formality refers to formal firms (acquiring commercial registration and business license) or to formal firms owned by formal entrepreneurs (acquiring work insurance).<sup>24</sup> The results remain positive and significant once again.

On the other hand, and as explained in section 4.2.1, our specification is sensitive to the introduction of some observable characteristics such as the gender and education level of the owner. These pieces of evidence support the argument in favor of the heterogeneous effect because of observable characteristics, which is emphasized in the next section.

### 4.2.3 Formality heterogeneous effects: Determining the group of complier firms

In this section, we aim to determine the group of complier firms—firms for which the distance to the tax office can shape their decision to formalize. The idea is to find the main observable characteristics that define this group of firms. To do so, we re-estimate our 2SLS regression (eq.4) that estimates the impact of formality on firms’ labor productivity and profits, by splitting it according to various variables that have shown significant impacts on firms outcomes in our previous results (see Tables 4 and 6). We repeat this exercise according to the following variable splits: the size of the firm (self-employed, mid-sized, and larger firms), the owner’s gender and intermediate level of education, and the median-level of firms’ age and capital intensity.

The results of this exercise on firms’ labor productivity are presented in Table 9. According to the reported first-stage-F-test, we observe that the positive effect that formality has on firms’ labor productivity holds for self-employed and mid-sized firms, for higher educated male entrepreneurs, and higher capital intensity firms. It also holds for both younger and older firms, although formal younger firms are three times more productive than older formal firms. These results remain very similar when repeating this exercise on firms’ profits (Table 10), except for owners’ gender and education levels, which hold and remain positive but are no longer significant.<sup>25</sup>

Therefore, we can confirm that the effect of formality on firms’ labor productivity or profits is not the same for all types of firms. Instead, it largely depends on firm and owner observable characteristics. Our 2SLS results seem to represent only the treatment effect for our group of complier firms. This includes male entrepreneurs who acquired an adequate level of education and experience, allowing them to better understand the market and outweigh the advantages of formality

---

<sup>24</sup>Combining the intensive and extensive margins of formality create different categories (only the firm is formal, only the owner is formal, both are formal, and both are informal). According to Table A5 in the Appendix, the categories that matter the most are “both formal” and “firm is formal.” Therefore, we decide to define formality in column (4) of Table A4 as explained.

<sup>25</sup>Panel A of Table 10 replicates the findings of McKenzie and Sakho (2010), who showed that the positive effect that formality has on firms’ profits only holds for mid-sized firms. Here, we add that this effect also holds for self-employed entrepreneurs and other observable firm characteristics (i.e., firms’ age and capital intensity).

over its costs. These entrepreneurs realize that starting a business formally generates sustainable levels of productivity and profits that would eventually cover the cost of formality. They have the skills to efficiently reallocate these gains to better investment opportunities to increase the capital intensity of the firm. Having all these strong characteristics in hand, these entrepreneurs can compete and survive in the market despite their firm being small. More importantly, what makes this group of entrepreneurs different is their strong willingness and ability to formalize if the fixed costs of formality get reduced, which in turn would increase their labor productivity and profits.<sup>26</sup>

Another interesting insight from these results is that we can now clearly categorize these household firms according to their productivity (or profit) levels and a hypothetical threshold of formality fixed costs. Firms below this threshold are informal. These firms are either too far away from this threshold because they are unproductive firms with low capital intensity and are managed by uneducated entrepreneurs who are trapped into informality (survival firms) or they are very close to this threshold and have all the prerequisite to formalize (part of the complier group) but voluntarily chose to stay informal to benefit from informality cost advantages. This parasite group of firms has probably gained all necessary means to stay informal and to hide from authority despite being productive and large enough. By contrast, firms above this threshold are formal. These firms either started formally because they could bear the cost of formality—which allowed them to generate sustainable levels of productivity and profits to stay formal—or are part of the complier group that decides to formalize because of a reduction in formality fixed costs (De Soto’s firms). Hence, to encourage formalization in Egypt, policies should not only reduce formality fixed costs but also find ways to induce parasite firms to join the formal sector. The policies should also prioritize the transition of more survivalist firms into the complier group through the provision of better access to finance and capital stock, as highlighted in the next section.

### 4.3 Transmission channels

In this section, we attempt to provide evidence on how formality increases firms’ labor productivity and profits. We also investigate why this positive effect is more applicable to the group of complier firms. Our findings, shed light on some interesting and important features characterizing this group. The complier firms have higher capital intensity despite being small. Further, their labor productivity can increase by 15% on average by acquiring assets (Table 6). These pieces of evidence draw

---

<sup>26</sup>Our findings that the group of compliers includes male entrepreneurs rather than female should not be considered as a result *per se* but is rather related to the reality of the Egyptian labor market. The ratio of female to male entrepreneurs in Egypt is very small because of social and cultural barriers (only 14% of entrepreneurs in our sample are females). They are either too small and unproductive to formalize, or they chose to stay informal because they can easily hide from authority. They are usually not taken seriously in the market or they might face unfair competition from their male counterparts.

attention to the importance of firms' capital stock in determining formality effects.

In Table 11, we test the extent to which formality could differently affect firms' capital and labor costs. Our findings show that for the same level of assets and labor, formal firms spend four times more on capital than on labor. For the complier group, a reduction in formality fixed costs would encourage them to formalize and to use saving costs to accumulate more capital, which leads to higher investment capacity, and in turn, translates into higher productivity and profit that is more sustainable over time.

Nevertheless, firms need a strong financial base to be able to accumulate capital and to invest. In Table 12, we test firms' ability to have better access to sources of finance through formality. We do so by using a firm's probability of having shared ownership versus sole proprietorship, as a proxy for the firm's financial capacity.<sup>27</sup> The results indicate that access to finance through shared ownership is 7 percentage points higher for formal firms compared to informal firms.<sup>28</sup>

Taken together, a reduction in formality cost is susceptible to encouraging a certain group of firms to join the formal sector. Formality will, in turn, enable these firms to have a stronger financial capacity, allowing them to efficiently allocate resources into capital, which will translate into a better investment, and significantly higher productivity and profits. This is more likely to hold for the group of complier firms (self-employed and mid-sized firms that have a high capital intensity and are owned by educated male entrepreneurs), for which, a reduction in formality fixed costs would actually affect their behavior toward formality.

## 5 Conclusion

In this paper we investigated the impact of commercial registration on labor productivity and profitability of household enterprises in Egypt. Using the 2012 and 2018 rounds of ELMPS, we also explored the heterogeneous effects of formality on performance. To overcome the selection into formal registration that unobservable firm characteristics may drive, we instrumented formality using the distance to the local tax office where registration occurs.

The results of the OLS estimation indicated a positive relationship between firms' formality and the outcomes of interest in all specifications. Operating formally induces a significant increase in firms' profit and a much higher increase in labor productivity (more than double that for profit). Moreover, labor productivity and profits are not affected by the same variables of firm and owner characteristics. In addition to the owner's gender and education level, labor productivity is mostly

---

<sup>27</sup>We use firms' ownership status as a proxy for financial capacity because our dataset lacks information on firms' access to sources of finance or their financial capacity.

<sup>28</sup>Results in Table 12 are based on a probit model. A Wald test was performed using an instrumental variable probit model to ensure the absence of endogeneity.

determined by the firm's assets and shared ownership, whereas profits are mostly determined by the age of the firm and that of its owner.

Moreover, for a certain group of entrepreneurs, operating informally can still be beneficial in terms of profit gains. These are the entrepreneurs who have the prerequisites to operate formally but choose not to do so to benefit from tax reductions. This is not the case for the increase in labor productivity, which cannot be attained unless the firm has assets and better access to finance, through shared ownership.

The first stage of the 2SLS estimation showed that firm and owner characteristics significantly impact the willingness to formalize. The latter increases if the firm favors shared ownership over sole proprietorship, if it has assets, and when the owner is older and more educated. The results of the second stage of the 2SLS confirmed the results of the OLS specification. They also showed that the effect of formality on firms' labor productivity remains positive and significant regardless of the firm's profit level. By contrast, the increase in firms' profits is actually limited to low productive firms as the percentage of profit loss due to informality shrinks with higher productivity levels.

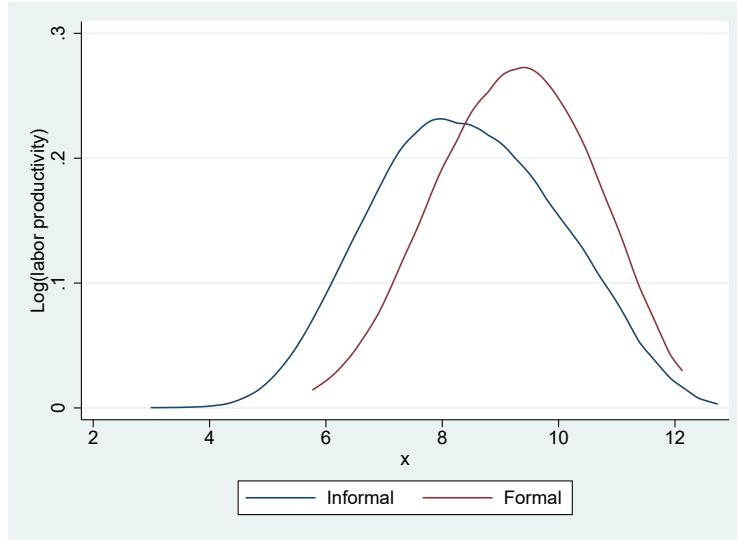
Our 2SLS estimated results represent the treatment effect for our group of compliers. This includes male entrepreneurs who acquired an adequate level of education and experience, allowing them to better understand the market and outweigh the advantages of formality over its costs. What makes this group of entrepreneurs different is their strong willingness and ability to formalize if the fixed costs of formality get reduced, which, in turn, would increase their labor productivity and profits.

Another interesting insight from these results is that we can now categorize these household firms according to their performance level and a hypothetical threshold of formality fixed costs. Firms below this threshold are informal. They are either too far from this threshold because they are unproductive or very close to this threshold but chose voluntarily to stay informal to benefit from informality cost advantages. Firms above this threshold are formal.

Finally, we provided evidence on how formality increases firms' labor productivity and profits and why this is more likely to happen for the group of compliers, which includes firms with higher capital intensity despite being relatively small in size and firms with higher assets. Hence, to encourage formalization, governments should not only think about reducing formality fixed costs but also prioritize the transition of more firms into the complier group through the provision of better access to finance and capital.

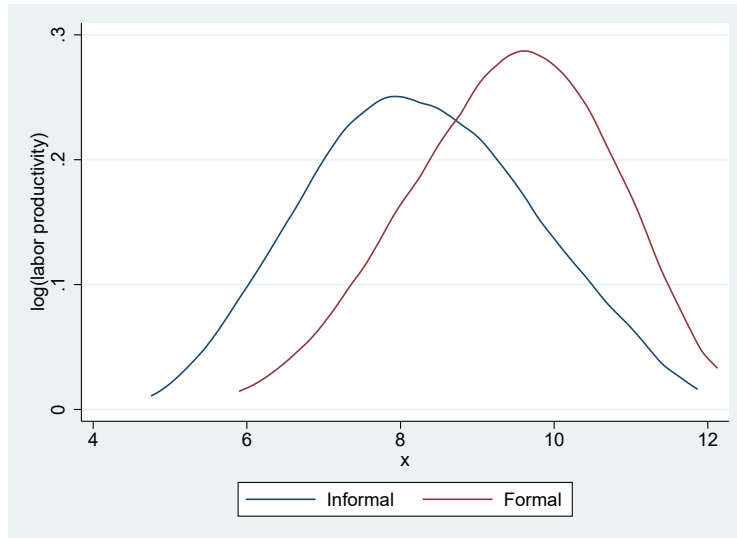
# Figures

Figure 1: Productivity overlap in aggregate



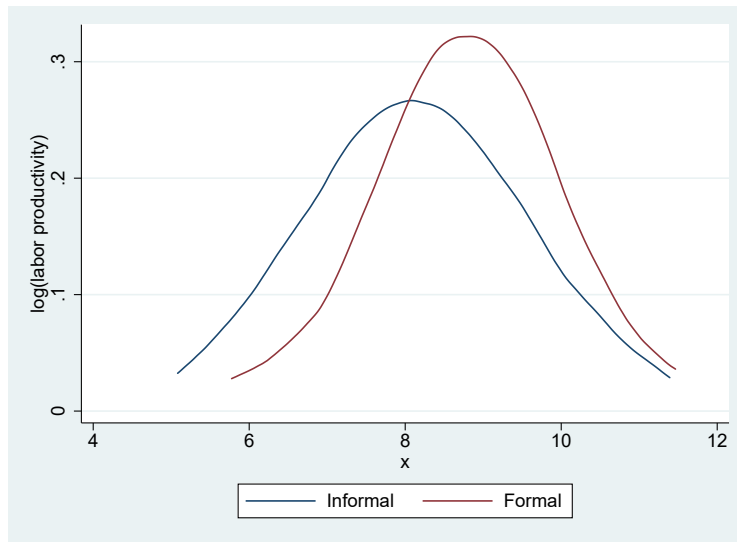
*Notes.* The figure shows overlap in productivity distribution in aggregate over all sectors of activity. Productivity is measured as explained in section 2.2.1.

Figure 2: Productivity overlap in wholesale and retail



*Notes.* The figure shows overlap in productivity distribution in the wholesale and retail sector. Productivity is measured as explained in section 2.2.1.

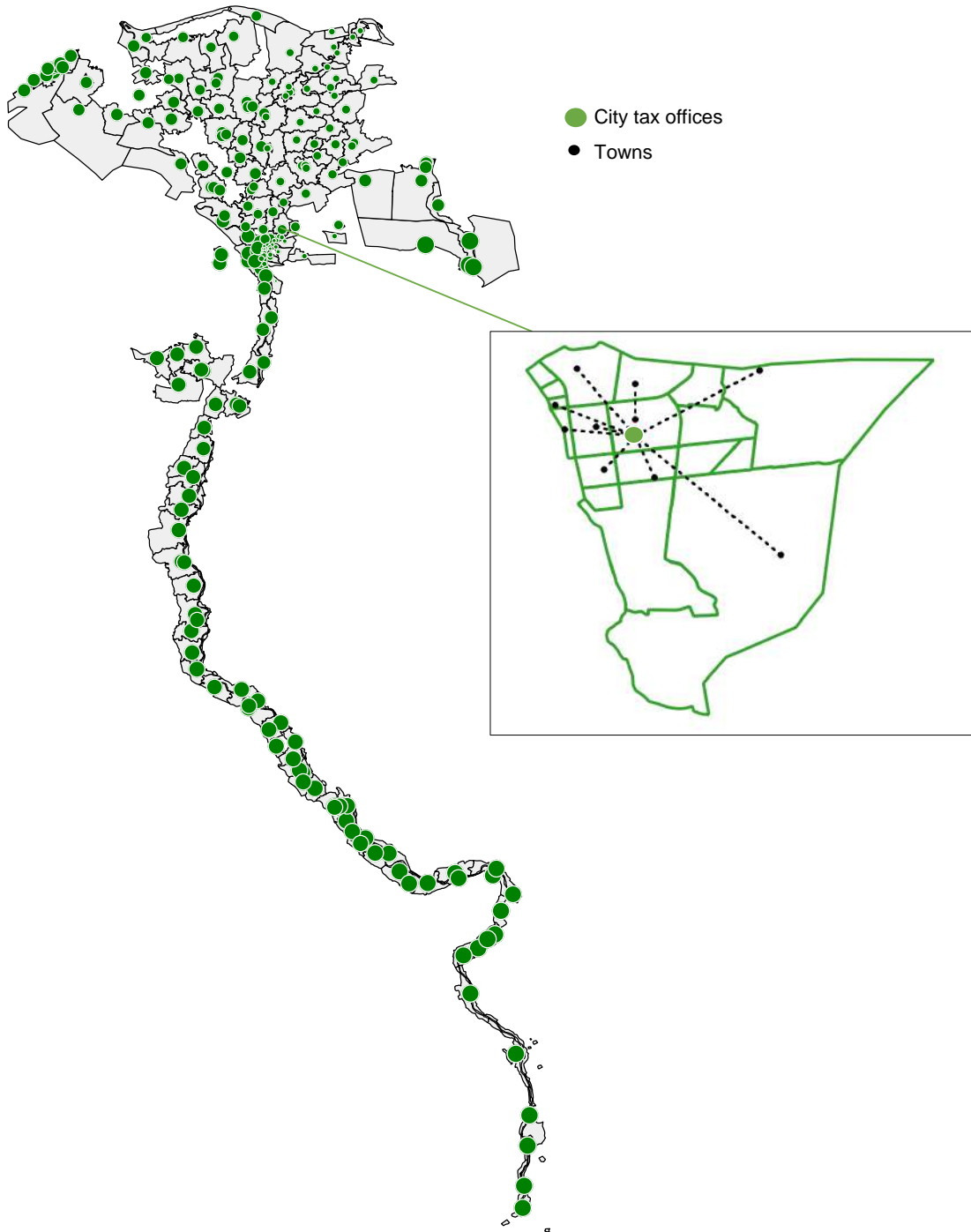
Figure 3: Productivity overlap in manufacturing



*Notes.* The figure shows overlap in productivity distribution in the manufacturing sector. Productivity is measured as explained in section 2.2.1.



Figure 4: Distribution of tax offices and towns across Egypt's governorates



## Tables

Table 1: Formal vs. informal firms

	(1)	(2)	(3)	(4)	(5)
	Log(labor)	Log(labor productivity)	Log(profit)	Shared ownership	Firm age
<i>Formal firm</i>	0.351*** (0.026)	0.821*** (0.049)	0.648*** (0.052)	0.102*** (0.015)	2.351*** (0.468)
<i>N</i>	3113	3113	3113	3113	3113
	(6)	(7)	(8)	(9)	(10)
	Assets	Owner age	Informal owner	Higher education	Female owner
<i>Formal firm</i>	0.080*** (0.020)	2.001*** (0.478)	-0.280*** (0.025)	0.191*** (0.020)	-0.117*** (0.013)
<i>N</i>	3113	3113	1710	3113	3113

*Notes.* This table reports coefficients from the OLS estimation with year, governorate, and sector fixed effects. Labor productivity and profit are measured as described in section 2.2.1. See Table A1 for definition of variables. Standard errors are clustered at the town-level and reported in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table 2: Registration matrix

	2018 Informal	2018 Formal	Total
2012 Informal	283	117	400
2012 Formal	109	198	307
Total	392	315	707

*Notes.* This table reports firm registration matrix over the years 2012 and 2018. Only firms included in both rounds are considered.

Table 3: Instrument validity

	(1)	(2)	(3)	(4)	(5)	(6)
	Log(distance to tax office)					
<i>Log(market size)</i>	-0.003 (0.023)					
<i>Ln(market size)<sub>t-1</sub></i>		-0.018 (0.022)				
<i>Log(capital intensity)</i>			0.024 (0.021)			
<i>Ln(capital intensity)<sub>t-1</sub></i>				0.018 (0.022)		
<i>Log(market population)</i>					-0.003 (0.032)	
<i>Ln(market population)<sub>t-1</sub></i>						-0.024 (0.023)
<i>N</i>	4401	790	4401	790	4401	790
Cluster se-level	sector	sector	sector	sector	sector	sector

*Notes.* This table reports coefficients from the OLS estimation. Standard errors are clustered at the sector-level and reported in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table 4: OLS estimation

	Log(labor productivity)			Log(Profit)
	(1)	(2)	(3)	(4)
<i>Formal firm</i>	0.714*** (0.049)	0.849*** (0.051)	0.717*** (0.050)	0.325*** (0.052)
<i>Log(labor)</i>		-0.239*** (0.039)	-0.301*** (0.038)	0.519*** (0.039)
<i>Firmage</i>		0.001 (0.002)	0.003 (0.002)	0.010*** (0.002)
<i>Shared ownership(ref.sole)</i>		0.344*** (0.068)	0.271*** (0.066)	0.118 (0.075)
<i>Firm has assets</i>		0.224*** (0.047)	0.188*** (0.045)	0.021 (0.050)
<i>Ownerage</i>			-0.003 (0.002)	-0.009*** (0.002)
<i>High education level</i>			0.383*** (0.048)	0.192*** (0.046)
<i>Female owner</i>			-0.791*** (0.069)	-0.718*** (0.069)
<i>N</i>	3149	3113	3113	3113
Year FE	No	Yes	Yes	Yes
Sector FE	No	Yes	Yes	Yes
Governorate FE	No	Yes	Yes	Yes
Cluster se-level	-	Town-level	Town-level	Town-level

*Notes.* This table reports coefficients from the OLS estimation. Standard errors are clustered at the town-level and reported in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table 5: First stage results: Firms' formality and distance to the tax office

	(Pr(formal firm))		
	(1)	(2)	(3)
<i>Distance to tax of fice</i>	-0.007*** (0.002)	-0.006*** (0.002)	-0.004** (0.002)
<i>Log(labor)</i>		0.233*** (0.018)	0.210*** (0.018)
<i>Firm age</i>		0.004*** (0.001)	0.003** (0.001)
<i>Shared ownership(ref. sole)</i>		0.159*** (0.031)	0.155*** (0.030)
<i>Firm has assets</i>		0.042* (0.022)	0.041* (0.022)
<i>Night light (w. pop)</i>		-0.183 (0.122)	-0.150 (0.116)
<i>Owner age</i>			0.004*** (0.001)
<i>High education level</i>			0.188*** (0.020)
<i>Female owner</i>			-0.134*** (0.033)
<i>N</i>	3007	2993	2993
<i>Year FE</i>	Yes	Yes	Yes
<i>Sector FE</i>	Yes	Yes	Yes
<i>Governorate FE</i>	Yes	Yes	Yes
<i>Cluster se-level</i>	Town-level	Town-level	Town-level

*Notes.* This table reports average marginal effects from the probit estimation. Standard errors are clustered at the town-level and reported in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table 6: Second stage results: Firms' formality, labor productivity, and profits

	Log(labor productivity)		Log(profit)
	(1)	(2)	(3)
<i>Formal firm</i>	2.308*** (0.592)	2.599*** (0.709)	1.400** (0.587)
<i>Log(labor)</i>		-0.604*** (0.154)	0.357*** (0.126)
<i>Firm age</i>		0.002 (0.003)	0.011*** (0.003)
<i>Shared ownership(ref.sole)</i>		0.027 (0.147)	0.012 (0.121)
<i>Firm has assets</i>		0.151** (0.065)	0.016 (0.061)
<i>Owner age</i>		-0.013*** (0.003)	-0.014*** (0.002)
<i>Night light (w.pop)</i>		0.355* (0.207)	0.432* (0.261)
<i>N</i>	3007	2993	2993
Year FE	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes
Governorate FE	Yes	Yes	Yes
Cluster se-level	Town-level	Town-level	Town-level
First-stage-F-test	29.693	22.035	22.035

*Notes.* This table reports coefficients from the 2SLS estimation. Standard errors are clustered at the town-level and reported in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table 7: The extent of formality productivity gains

	(1)	(2)
	Log(labor productivity)	
	High profit firms	Low profit firms
<i>Formal firm</i>	1.307* (0.692)	1.583** (0.761)
<i>N</i>	1564	1428
First-stage-F-test	11.428	15.670

*Notes.* This table reports coefficients from the 2SLS estimation with year, governorate, and sector fixed effects. Log(productivity) is the dependent variable. Control for night light intensity (weighted by population ratio) are included in all regressions. Standard errors are clustered at the town-level and reported in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table 8: The extent of formality profit gains

	(1)	(2)
	Log(profit)	
	High productivity firms	Low productivity firms
<i>Formal firm</i>	0.373 (0.680)	1.718** (0.788)
<i>N</i>	1575	1417
First-stage-F-test	12.313	15.507

*Notes.* This table reports coefficients from the 2SLS estimation with year, governorate, and sector fixed effects. Log(profit) is the dependent variable. Control for night light intensity (weighted by population ratio) are included in all regressions. Standard errors are clustered at the town-level and reported in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table 9: Treatment heterogeneous effects: Labor productivity

	Log(labor productivity)		
	(1)	(2)	(3)
Panel A: by size	Self-employed	Mid-size	Larger
<i>Formal firm</i>	3.117*** (0.957)	1.350** (0.647)	8.415 (16.869)
<i>N</i>	1846	958	181
First-stage-F-test	12.417	16.350	0.204
Panel B: by education	Lower education level	Higher education level	
<i>Formal firm</i>	1.711 (1.246)	2.096*** (0.629)	
<i>N</i>	1370	1621	
First-stage-F-test	5.833	16.465	
Panel C: by owner gender	Female	Male	
<i>Formal firm</i>	2.875* (1.529)	1.412** (0.603)	
<i>N</i>	403	2587	
First-stage-F-test	7.187	18.798	
Panel D: by firm age	Younger firm	Older firms	
<i>Formal firm</i>	3.515*** (1.156)	1.185* (0.609)	
<i>N</i>	1484	1509	
First-stage-F-test	11.260	17.016	
Panel E: by capital intensity	Lower capital intensity	Higher capital intensity	
<i>Formal firm</i>	2.022* (1.071)	1.225*** (0.400)	
<i>N</i>	1411	1582	
First-stage-F-test	7.848	21.952	

*Notes.* This table reports the coefficients from the 2SLS estimation with year, governorate, and sector fixed effects. The dependent variable is the firms' labor productivity. Self-employed refers to one-person firms (i.e., the owner). Mid-size firms refer to firms with 2 to 4 workers (including owners). Larger firms refer to firms with more than 4 workers (including owners). Lower education refers to owners with education below the intermediate level (categories: illiterate, reads and writes, or less than intermediate). Higher education refers to owners with at least an intermediate level of education (categories: intermediate, above intermediate, university or post-graduate). Younger and older firms are defined according to the median firm age. Capital intensity is the ratio of capital expenditures to labor expenditures. Sample split is based on the median value of capital intensity. Control for night light intensity (weighted by population ratio) is included in all regressions. Standard errors are clustered at the town-level and reported in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table 10: Treatment heterogeneous effects: Profit

		Log(profit)		
		(1)	(2)	(3)
Panel A: by size		Self-employed	Mid-size	Larger
<i>Formal firm</i>		1.659** (0.818)	1.153* (0.609)	-0.183 (5.128)
<i>N</i>		1846	958	181
First-stage-F-test		12.417	16.350	0.204
Panel B: by education		Lower education level	Higher education level	
<i>Formal firm</i>		2.090 (1.309)	0.987 (0.599)	
<i>N</i>		1370	1621	
First-stage-F-test		5.833	16.465	
Panel C: by owner gender		Female	Male	
<i>Formal firm</i>		2.042 (1.551)	0.726 (0.545)	
<i>N</i>		403	2587	
First-stage-F-test		7.187	18.798	
Panel D: by firm age		Younger firm	Older firms	
<i>Formal firm</i>		1.753* (0.902)	1.226** (0.607)	
<i>N</i>		1484	1509	
First-stage-F-test		11.260	17.016	
Panel E: by capital intensity		Lower capital intensity	Higher capital intensity	
<i>Formal firm</i>		2.020* (1.105)	1.077** (0.514)	
<i>N</i>		1411	1582	
First-stage-F-test		7.848	21.952	

*Notes.* This table reports the coefficients from the 2SLS estimation with year, governorate, and sector fixed effects. The dependent variable is the firms' labor productivity. Self-employed refers to one-person firms (i.e., the owner). Mid-size firms refer to firms with 2 to 4 workers (including owners). Larger firms refer to firms with more than 4 workers (including owners). Lower education refers to owners with education below the intermediate level (categories: illiterate, reads and writes, or less than intermediate). Higher education refers to owners with at least an intermediate level of education (categories: intermediate, above intermediate, university or post-graduate). Younger and older firms are defined according to the median firm age. Capital intensity is the ratio of capital expenditures to labor expenditures. Sample split is based on the median value of capital intensity. Control for night light intensity (weighted by population ratio) is included in all regressions. Standard errors are clustered at the town-level and reported in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .



Table 11: Labor cost vs. capital cost

	(1)	(2)
	Log(capital cost)	Log(labor cost)
<i>Formal firm</i>	4.212*** (1.386)	1.019*** (0.310)
<i>Firm has assets</i>	0.496*** (0.125)	-0.019 (0.029)
<i>Log(labor)</i>	0.505* (0.295)	0.394*** (0.071)
<i>Firm age</i>	-0.013* (0.007)	0.001 (0.002)
<i>Shared ownership</i>	0.027 (0.283)	-0.264*** (0.065)
<i>Owner age</i>	-0.014*** (0.005)	0.000 (0.001)
<i>Night light (w.pop)</i>	-0.991 (0.638)	0.026 (0.099)
<i>N</i>	2993	2993
First-stage-F-test	22.035	22.035

*Notes.* This table reports coefficients from the 2SLS estimation with year, governorate, and sector fixed effects. Capital and labor costs are measured as explained in section 2.2.1. Standard errors are clustered at the town-level and reported in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table 12: Access to finance through shared ownership

	Pr(shared ownership)
<i>Formal firm</i>	0.073*** (0.012)
<i>Firm has assets</i>	0.009 (0.012)
<i>Log(labor)</i>	0.044*** (0.009)
<i>Firm age</i>	0.001 (0.001)
<i>Owner age</i>	-0.003*** (0.001)
<i>N</i>	3103

*Notes.* This table reports average marginal effects from the probit estimation with year, governorate, and sector fixed effects. Standard errors are clustered at the town-level and reported in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

## References

- Allen, Jeffrey, Shanthi Nataraj, and Tyler C Schipper**, “Strict duality and overlapping productivity distributions between formal and informal firms,” *Journal of Development Economics*, 2018, *135*, 534–554.
- Amin, Mohammad and Asif Islam**, “Are large informal firms more productive than the small informal firms? Evidence from firm-level surveys in Africa,” *World Development*, 2015, *74*, 374–385.
- Assaad, Ragui and Caroline Krafft**, “The Egypt labor market panel survey: introducing the 2012 round,” *IZA Journal of Labor & Development*, 2013, *2* (1), 8.
- Benhassine, Najy, David McKenzie, Victor Pouliquen, and Massimiliano Santini**, “Does inducing informal firms to formalize make sense? Experimental evidence from Benin,” *Journal of Public Economics*, 2018, *157*, 1–14.
- Benjamin, Nancy C and Ahmadou Aly Mbaye**, “The Informal Sector, Productivity, and Enforcement in West Africa: A Firm-level Analysis,” *Review of Development Economics*, 2012, *16* (4), 664–680.
- Carneiro, Pedro, James J Heckman, and Edward J Vytlacil**, “Estimating marginal returns to education,” *American Economic Review*, 2011, *101* (6), 2754–81.
- Charmes, Jacques**, “The informal economy worldwide: Trends and characteristics,” *Margin: The Journal of Applied Economic Research*, 2012, *6* (2), 103–132.
- Demenet, Axel, Mireille Razafindrakoto, and François Roubaud**, “Do informal businesses gain from registration and how? Panel data evidence from Vietnam,” *World Development*, 2016, *84*, 326–341.
- Ellison, Glenn, Edward L Glaeser, and William R Kerr**, “What causes industry agglomeration? Evidence from coagglomeration patterns,” *American Economic Review*, 2010, *100* (3), 1195–1213.
- Fajnzylber, Pablo, William F Maloney, and Gabriel V Montes-Rojas**, “Does formality improve micro-firm performance? Evidence from the Brazilian SIMPLES program,” *Journal of Development Economics*, 2011, *94* (2), 262–276.
- Grimm, Michael, Peter Knorringa, and Jann Lay**, “Constrained Gazelles: High Potentials in West Africa’s Informal Economy,” *World Development*, 2012, *7* (40), 1352–1368.

- Günther, Isabel and Andrey Launov**, “Informal employment in developing countries: Opportunity or last resort?,” *Journal of development economics*, 2012, *97* (1), 88–98.
- Henderson, Vernon, Adam Storeygard, and David N Weil**, “A bright idea for measuring economic growth,” *American Economic Review*, 2011, *101* (3), 194–99.
- Krafft, Caroline, Ragui Assaad, Khandker Wahedur Rahman et al.**, “Introducing the Egypt labor market panel survey 2018,” in “Economic Research Forum Working Paper Series (Forthcoming)” 2019.
- Krugman, Paul**, “Increasing returns and economic geography,” *Journal of political economy*, 1991, *99* (3), 483–499.
- Loayza, Norman V and Jamele Rigolini**, “Informal employment: safety net or growth engine?,” *World Development*, 2011, *39* (9), 1503–1515.
- Maloney, William F**, “Informality revisited,” *World development*, 2004, *32* (7), 1159–1178.
- McKenzie, David and Yaye Seynabou Sakho**, “Does it pay firms to register for taxes? The impact of formality on firm profitability,” *Journal of Development Economics*, 2010, *1* (91), 15–24.
- McKenzie, David J and Christopher Woodruff**, “Do entry costs provide an empirical basis for poverty traps? Evidence from Mexican microenterprises,” *Economic development and cultural change*, 2006, *55* (1), 3–42.
- Medina, Leandro and Friedrich Schneider**, “Shadow economies around the world: what did we learn over the last 20 years?,” 2018.
- OAMDI**, “Labor Market Panel Surveys (LMPS), <http://erf.org.eg/data-portal/>. Version (3.0) of Licensed Data Files; ELMPS 2018. Egypt: Economic Research Forum (ERF),” 2018.
- Porta, Rafael La and Andrei Shleifer**, “Informality and development,” *Journal of Economic Perspectives*, 2014, *28* (3), 109–26.
- Rand, John and Nina Torm**, “The benefits of formalization: Evidence from Vietnamese manufacturing SMEs,” *World development*, 2012, *40* (5), 983–998.
- Rothenberg, Alexander D, Arya Gaduh, Nicholas E Burger, Charina Chazali, Indrasari Tjandraningsih, Rini Radikun, Cole Sutera, and Sarah Weiland**, “Rethinking Indonesias informal sector,” *World Development*, 2016, *80*, 96–113.
- Soto, Hernando De et al.**, “The other path,” 1989.

**Storeygard, Adam**, “Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa,” *The Review of economic studies*, 2016, *83* (3), 1263–1295.

**The World Bank**, “Arab Republic of Egypt. Reshaping Egypt’s Economic Geography: Domestic Integration as a Development Platform,” Technical Report june 2012.

**Ulysea, Gabriel**, “Firms, informality, and development: Theory and evidence from Brazil,” *American Economic Review*, 2018, *108* (8), 2015–47.

– , “Informality: Causes and Consequences for Development,” *Annual Review of Economics*, 2020, *12*.

## 6 Appendix

Table A1: Firms' characteristics: Mean (standard deviation).

	Definition	All firms	Informal firms	Formal firms
<i>Formal firms</i>	=1 if firm is registered	0.336 (0.472)	0	1
<i>Labor</i>	Total labor (monthly)	1.837 (1.922)	1.601 (1.533)	2.498 (2.653)
<i>Labor productivity</i>	Value added over total workers (monthly)	12745.3 (17272.0)	11126.5 (16225.8)	17000.3 (18270.3)
<i>Profit</i>	Total net-earning (monthly)	3730.1 (6956.1)	3153.8 (5866.1)	5565.8 (9318.2)
<i>Shared ownership</i>	=1 if shared ownership and 0 if sole ownership	0.101 (0.302)	0.0755 (0.264)	0.175 (0.380)
<i>Assets</i>	=1 if firm has assets (e.g., building, land, machinery)	0.504 (0.500)	0.473 (0.499)	0.535 (0.499)
<i>Firm age</i>	Firm age in years	16.50 (10.90)	15.10 (10.61)	18.55 (11.47)
<i>Owner age</i>	Owner age in years	41.48 (12.09)	40.78 (11.93)	43.52 (12.33)
<i>Informal owner</i>	=1 if owner has no work insurance	0.749 (0.434)	0.832 (0.374)	0.556 (0.497)
<i>Owner education</i>	=1 if owner's education is above intermediate	0.529 (0.499)	0.479 (0.500)	0.668 (0.471)
<i>Female owner</i>	=1 for female owners	0.142 (0.349)	0.169 (0.375)	0.0719 (0.258)
<i>Distance to tax office</i>	Distance to tax office in km	7.237 (6.127)	7.746 (6.132)	6.214 (5.989)
<b>Total</b>		<b>3149</b>	<b>2092</b>	<b>1057</b>

*Notes.* All monetary variables are in Egyptian pounds and deflated using corresponding year's CPI. Labor productivity and profit are measured as described in section 2.2.1. Distance to the tax office is measured as described in section 3.2.

Table A2: Panel firms characteristics: Mean (standard deviation).

	Remained informal	Became formal	Became informal	Remained formal
<i>Formal firms</i>	0	0.425	0.550	1
		(0.496)	(0.499)	
<i>Labor</i>	1.548	1.954	2.206	2.808
	(1.104)	(1.426)	(2.013)	(3.035)
<i>Labor productivity</i>	9199.0	14679.6	14903.9	18830.8
	(14583.4)	(17989.7)	(17254.3)	(19767.2)
<i>Profit</i>	2419.6	4269.3	4693.7	6409.3
	(4367.2)	(6668.1)	(6864.8)	(11282.0)
<i>Shared ownership</i>	0.0645	0.115	0.139	0.196
	(0.246)	(0.320)	(0.347)	(0.397)
<i>Firm has assets</i>	0.525	0.506	0.572	0.491
	(0.500)	(0.501)	(0.496)	(0.501)
<i>Firm age</i>	16.99	16.72	19.54	21.34
	(10.23)	(10.04)	(11.36)	(11.53)
<i>Owner age</i>	41.78	39.99	45.93	45.54
	(10.68)	(10.64)	(12.17)	(12.52)
<i>Informal owner</i>	0.857	0.736	0.694	0.530
	(0.350)	(0.442)	(0.462)	(0.500)
<i>Owner education</i>	0.382	0.592	0.561	0.690
	(0.487)	(0.493)	(0.498)	(0.463)
<i>Female owner</i>	0.180	0.0747	0.0833	0.0427
	(0.384)	(0.264)	(0.277)	(0.203)
<i>Distance to tax office</i>	8.039	9.261	6.460	4.502
	(6.498)	(6.852)	(5.589)	(4.977)

*Notes.* All monetary variables are in Egyptian pounds and deflated using corresponding year's CPI. Labor productivity and profit are measured as described in section 2.2.1. Distance to the tax office is measured as described in section 3.2. See Table A1 for definition of variables.

Table A3: Robustness check: Firm formality and distance to the tax office

	Pr(formal firm)	
	(1)	(2)
<i>Distance to tax office</i>	-0.037*** (0.013)	-0.017*** (0.005)
<i>Log(labor)</i>	0.258*** (0.027)	0.224*** (0.021)
<i>Firm age</i>	0.002 (0.002)	0.002 (0.001)
<i>Shared ownership (ref.sole)</i>	0.177*** (0.052)	0.160*** (0.038)
<i>Firm has assets</i>	0.038 (0.033)	0.036 (0.028)
<i>Owner age</i>	0.005*** (0.001)	0.005*** (0.001)
<i>High education level</i>	0.217*** (0.033)	0.205*** (0.026)
<i>Female owner</i>	-0.153*** (0.055)	-0.141*** (0.039)
<i>Night light (w.pop)</i>	-0.404** (0.200)	-0.257** (0.119)
<i>N</i>	1431	2116

*Notes.* This table reports average marginal effects from probit estimation. The distance to the tax office is set to a maximum of 5 km in column (1) and to a maximum of 10 km in column (2). These values correspond to the distance median value and the 75th percentile value, respectively. Standard errors are clustered at the town-level and reported in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table A4: Robustness check: Formality definition

	Log(firm labor productivity)			
	(1)	(2)	(3)	(4)
<i>Registration</i>	2.567*** (0.711)			
<i>Registration + license</i>		2.789*** (0.776)		
<i>Registration + license + accounting</i>			2.449*** (0.661)	
<i>Formality firm and owner</i>				2.994*** (0.881)
<i>N</i>	2,993	3,111	3,117	3,424
<i>First-stage-F-test</i>	21.65	19.74	35.14	21.87

*Notes.* This table reports coefficients from the 2SLS estimation with year, governorate, and sector fixed effects. Control for night light intensity (weighted by population ratio) are included in all regressions. Standard errors are clustered at the town-level and reported in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table A5: Combining formality intensive/extensive margins

	Log(firm labor productivity)
<i>Firm – owner formality (ref. informal firm owner)</i>	
<i>Formal firm owner</i>	0.666*** (0.105)
<i>Only firm is formal</i>	0.837*** (0.104)
<i>Only owner is formal</i>	-0.356*** (0.0874)
<i>Log(labor)</i>	-0.245*** (0.0499)
<i>Firm age</i>	0.00280 (0.00316)
<i>Shared ownership (ref. sole)</i>	0.262*** (0.0909)
<i>Owner age</i>	-0.00845*** (0.00281)
<i>N</i>	3,404

*Notes.* This table reports coefficients from the 2SLS estimation with year, governorate, and sector fixed effects. Control for night light intensity (weighted by population ratio) are included in all regressions. Standard errors are clustered at the town-level and reported in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$