# How Do Small Formal and Informal Firms in the Arab Republic of Egypt Compare?

*Caroline Krafft*
*Ragui Assaad*
*Khandker Wahedur Rahman*
*Maakwe Cumanzala*

## Abstract

Formalizing firms can potentially increase the tax base, expand safety and social protections for workers, create good jobs, and grow the economy. However, the costs and processes of formality may be too challenging for firms, particularly the smallest firms, to bear. Thus, informal firms may not be able to survive the transition to formality and attempts to expand formality may be harmful and counterproductive to job creation and growth. This paper investigates the potential for currently informal firms to formalize in the Arab Republic of Egypt. The paper compares the characteristics and dynamics of micro and small non-agricultural firms by formality and identifies the extent of overlap and potential for formalization. The analysis finds that, beyond firm size, the basic and easily observable characteristics of firms are not closely linked to formality. Firm age, productivity, and owner characteristics such as education are strongly predictive of formality. There is some overlap in the predicted probability of formality between formal and informal firms, suggesting some potential for formalization. The paper develops profiles (groups and clusters) of similar firms to identify those with a higher potential for formalization. In terms of dynamics, new firms tend to be informal and informal firms are more likely to exit (close), but conditional on firm survival, employment growth is similar across formal and informal firms.

**How Do Small Formal and Informal Firms in the Arab Republic of Egypt Compare?**

**An Analysis of Firm Characteristics and Implications for Formalization Efforts[1]**

Caroline Krafft,[2] Ragui Assaad,[3] Khandker Wahedur Rahman,[4] Maakwe Cumanzala[5]

---

[2] Corresponding author. Department of Economics and Political Science, St. Catherine University, cgkrafft@stkate.edu

[3] Humphrey School of Public Affairs, University of Minnesota.

[4] Department of Applied Economics, University of Minnesota.

[5] Department of Economics and Political Science, St. Catherine University.

**1. Introduction**

The Egyptian labor market, since structural adjustment in the 1990s, has been characterized by a sharp decline in formal employment and growth of the informal economy. The public sector shrank from 39% of employment in 1998 to 26% in 2018 (Assaad, AlSharawy, & Salemi, 2019). The private sector was unable to replace the shrinking public sector with good, formal jobs (jobs with social insurance or contracts). From 1998 to 2018 the share of formal private wage employment in total employment grew from 8% to just 12%, while the share of informal private employment grew from 53% to 62% (Assaad, AlSharawy, & Salemi, 2019).[6]

Lack of formality and its associated benefits, such as protections afforded through the labor code, paid leave and retirement benefits, are a key impediment to women's participation in the labor force. Despite rising education, women in the Arab Republic of Egypt have increasingly withdrawn from the labor market (Assaad, Hendy, Lassassi, & Yassin, 2018). Women particularly exit at marriage, in the face of informal employment opportunities in the private sector that are irreconcilable with their domestic responsibilities (Assaad, Krafft, & Selwaness, 2017; Selwaness & Krafft, 2018). Access to formal employment and its associated benefits has come to increasingly rely on class and connections for the younger generation (Assaad, Krafft, & Salemi, 2019; Assaad & Krafft, 2020). Frustrations with the lack of good jobs among young people and the middle class were a key component of the Arab Spring protests (World Bank, 2013; Gatti, Angel-Urdinola, Silva, & Bodor, 2014).

On top of the challenges informal employment presents to Egyptian workers, informal firms are a challenge to the macroeconomy. Informal firms do not pay taxes, which creates a difficult fiscal situation for Egypt (AfDB, 2016). The ongoing growth of informal firms in the economy compounds this issue (Elshamy, 2015). Informal firms, facing lower costs, may out-compete formal firms, creating a particular challenge for growing the formal sector (Ali & Najman, 2016). Informal firms may also stay small and not create jobs in order to avoid the attention of authorities (AfDB, 2016). There have therefore been longstanding calls in policy circles in Egypt to formalize micro and small firms to expand the tax base, increase access to formal employment with its various benefits and protections, and facilitate firm access to finance and international trade opportunities (Egyptian Center for Economic Studies, 2005; World Bank, 2014).

Yet formality may create challenges for job creation overall, if the process of formalizing or the requisites of formality include burdensome regulations and substantially higher costs (Sparks & Barnett, 2010). Micro and small enterprises play a critical role in employment creation and dynamics (Li & Rama, 2015). Individuals may start firms more out of necessity or survival (primarily self-employment) than entrepreneurial opportunity (Naudé, 2008, 2010; Brixiova, 2010; Krafft & Rizk, 2018). If the costs of formality are sizeable and fixed in nature, they may prove particularly burdensome for the smallest firms. There are heterogenous results across countries and studies in terms of whether informal firms are

---

[6] This percentage may include a small fraction of formal self-employment, but it is unlikely to exceed a few percentage points.

similarly productive to formal firms (Gelb, Mengistae, Ramachandran, & Shah, 2009; Benjamin & Mbaye, 2012). In some contexts, informal firms were more like formal firms, suggesting potential for formalization.

The reasons firms in Egypt are informal are tied up with costs, although it is not clear if the costs would preclude firm survival and growth. Employers who are informal in Egypt state that their most common reason for remaining informal is that it "saves time and effort" followed closely by "formalization does not involve advantages" (25%-30%). Financial reasons – "no need to pay employees' social insurance" and "no need to pay taxes" are the next most common reasons (around 15%), followed by no need to apply labor laws (around 6%-7%) (AfDB, 2016). Since firms are presumably engaged in a cost-benefit analysis when making their decisions to remain informal, that calculus will likely have to change to induce them to formalize (World Bank, 2014).

There have been substantial recent improvements in the procedures to start a new business (or to formalize an existing business) in Egypt. Although the cost of formally starting a business declined from 69% of income per capita in 2007 to 20% in 2020, it may still be prohibitive for very small firms (World Bank, 2006, 2020). Similarly, the time required to start a business fell from 19 days in 2007 to 12 days in 2020. Egypt's rank in starting a business has accordingly improved from 165th out of 175 economies in 2007 to 90th out of 190 economies in 2020 (World Bank, 2006, 2020). However, other aspects of formality, such as paying taxes, enforcing contracts, and registering property remain extremely burdensome, with Egypt ranking 156th out of 190 economies in paying taxes, 166th in enforcing contracts, and 130th in registering property in 2020 (World Bank, 2020).

The assumption underlying the calls to formalize informal firms is that small informal firms, or at least subsets of them, can be formalized. This implies that informal and formal firms are similar to each other in terms of productivity, employment, types of activities, location, etc. This assumption implies that formalization would be feasible and requiring formality would not result in the firm instead disappearing altogether. This paper sets out to test whether currently informal firms in Egypt are "formalizable" and, if so, which firms, in particular, might be susceptible to formalization.

We use multiple data sources from Egypt to assess which firms are more likely to be formal. We then quantify the amount of overlap between formal and informal firms in terms of their predicted probability of formality, based on observable characteristics. While we find some overlap in the probability to be formal, particularly among firms in fixed establishments, we also find that many informal firms have a low probability of formality and may not be formalizable. We create profiles (clusters, groups) of firms with similar characteristics to explore which informal firms are more, or less, likely to formalize. We also explore the relationship between firm dynamics and formality. We conclude that, in the current landscape in Egypt, while enforcement and encouragement may yield some progress in expanding the formal sector, more substantive reforms to the process and practice of formality are likely needed to substantially expand the formal sector.

**2. Data**

**2.1. Data sources**

Since we are interested in the phenomenon of firm informality, for this study we focus on a universe of private non-agricultural firms with fewer than 25 workers, but situate this universe within employment in Egypt. This focus is because public enterprises and government establishments are not at risk of being informal.[7] Formality is typically not relevant for most agricultural activities, and informality among firms of 25 workers or more is rare (Assaad, AlSharawy, & Salemi, 2019). Given this universe, we use data sets from household surveys with a household enterprise module and establishment surveys or censuses that represent both formal and informal firms of all sizes.

The first data source we use is the Egypt Labor Market Panel Survey (ELMPS) carried out by the Economic Research Forum in collaboration with Egypt's Central Agency for Public Mobilization and Statistics (CAPMAS). Four waves of this survey were carried out in 1998, 2006, 2012, and 2018.[8] We primarily use the 2012 and 2018 waves for contemporaneity with our other data sources, but use the 1998 and 2006 waves when exploring firm dynamics. The second data source is the 2012/13 Economic Census (EcC), which is an in-depth survey of establishments carried out by CAPMAS every five to six years. The third is the Egypt 2017 Establishment Census (EsC), which is a full census of establishments, also carried out by CAPMAS in conjunction with the decennial population and housing census. These data sources have different coverage and depth of detail, as we discuss below.

**2.1.1. Egypt Labor Market Panel Survey**

The ELMPS is a nationally-representative household survey that has tracked a panel of households (continually adding refresher samples) since 1998.[9] The ELMPS has, since its inception, included a non-agricultural household enterprise module. Information on agricultural activities is collected in other modules that are not relevant for our purpose. Any household which reports a member who is self-employed, an employer, or unpaid family worker in either their primary or secondary jobs must respond to either the non-agricultural or agricultural enterprise module.[10] Households can report on multiple enterprises. There were 2,315 enterprises in 2,195 households in the 2012 wave of the ELMPS and 2,228

---

[7] The public sector does not have informal firms, but it does, increasingly, have some informal employment (Selwaness & Ehab, 2019).

[8] For more information on the data, see Krafft, Assaad, and Rahman (2019). The data are publicly available at www.erfdataportal.com.

[9] The ELMPS does not cover the Frontier governorates of Matrouh, New Valley, Red Sea and North and South Sinai, which together comprise about 2% of Egypt's population.

[10] Households that did not initially report members as employed in these statuses may also report non-agricultural enterprises; indeed, allowing this structure identifies substantially more employment among women (Krafft, Keo, & Fedi, 2019).

enterprises in 2,131 households in the 2018 wave.[11] For each enterprise, there are questions about enterprise characteristics, household members' participation, employment of non-household members, net revenues and expenditures.

### 2.1.2. Economic Census and Establishment Census

We use the public release 50% sub-sample of Egypt's 2012/2013 EcC. Despite being called a census, the EcC is a survey of fixed establishments, with the largest establishments having a sampling rate of 100%. The 50% sub-sample we have access to includes 62,108 establishments. Although the 2017 EsC is truly a census of fixed establishments, we only have access to a 20% random sample of these establishments. Thus, our sample is made up of 772,432 establishments. The EcC and EsC cover all non-governmental establishments and therefore exclude any firms that are not in fixed establishments, i.e. ones that involve mobile workers or ones whose activities take place in a field or open worksite, such as the vast majority of firms engaged in construction, agriculture, and transportation activities.[12] The EcC and EsC include public enterprises, but we exclude those from our analyses since they are by definition formal. We also exclude the very small number of agricultural firms that show up in both data sources for comparability with the ELMPS. Since the ELMPS excludes the Frontier governorates, we likewise exclude these governorates from the EsC and EcC data. Since the EcC and EsC only include firms that operate in fixed establishments but the ELMPS includes firms that operate both in and out of establishments, most of our analysis is limited to firms operating within fixed establishments.

### 2.2. Measuring Formality

Firm formality can be defined and operationalized in a variety of ways. We test two increasingly expansive measures of firm formality. First, our basic measure of formality is commercial registration. However, in the EcC, the registration question was not asked of firms in a number of industries.[13] In our extended definition, we expand this measure to include firms that paid social insurance premiums (available in EcC and EsC) or taxes (available in the ELMPS).

### 2.3. Firm characteristics

We have a limited set of harmonizable, common characteristics across all the data sets we use. We compare firms across these characteristics in terms of whether they are describing firms in our universe in the same fashion and whether the different data sources find similar patterns of formality across characteristics. We then conduct some analyses on a broader set of characteristics present only in the EcC or only in the ELMPS.

---

[11] For the dynamic analyses we also use the 1998 wave (1,059 households and 1,134 enterprises) and 2006 wave (1,899 households and 2,044 enterprises).

[12] Firms operating in fixed establishments include firms that operate in private homes.

[13] The EcC uses 11 different survey instruments for firms in different industries. Some of these did not include a question on commercial registration.

### 2.3.1. Common characteristics

The firm's size is one of our common characteristics. We categorize firms into one worker (primarily self-employment), two workers, 3-4 workers, or 5-24 workers. Workers may be paid or unpaid. Industry, at the one-digit level (or aggregated further to present meaningful descriptives), is available in all three data sets. The 2012 and 2018 ELMPS and EcC have the calendar year the firm started, while the EsC has the calendar year if 2010 or later and then switches to decades (e.g. 2000-2009). We therefore create a harmonized firm age variable that is single years of age from 0-7 years old, and then 8-17 years old, 18-27 years old, 28+, or do not know.[14] We further have information on the governorate, which we present in terms of regions: Greater Cairo, Alexandria and Suez Canal, Lower Egypt, and Upper Egypt.[15]

### 2.3.2. Characteristics in the Economic Census

In the EcC, we have particularly rich detail on firm characteristics. In addition to the common characteristics described above, we include indicators for having unpaid workers and wage workers. We also measure the share (percentage) of workers who are female. Since we have detailed information on value added, invested capital, and labor, we analyze the relationship between formality and the log of capital per worker and the log of value added per worker. We also include information on whether the firm is a sole proprietorship or some type of collective ownership arrangement.

### 2.3.3. Characteristics in the ELMPS

In the ELMPS, we can identify additional detail about firm characteristics and owners. We consider the owner to be the person who worked the most on the enterprise.[16] We include the owner's sex, age group (<30, 30-49, 50+), education (illiterate, less than secondary, secondary, higher education), and occupation (professional/technical, other white collar, blue collar craft, blue collar non-craft[17]).

In addition to owner characteristics, we use firm characteristics from the ELMPS enterprise module. Specifically, we have current capital (categorically), the type of the workplace (e.g. office, workshop, etc.), and sole proprietorship (household) vs. collective ownership (partnership). Using the number of workers and the net revenues of the enterprise going to the household, we calculate net revenue/worker as a

---

[14] Since the ELMPS 1998 and 2006 waves have start years categorically, when analyzing dynamics (which we can only do using the ELMPS) we switch to harmonized categorical start years.

[15] Greater Cairo includes the governorates of Cairo, Giza, and Qalyoubia. Alexandria and the Suez Canal includes Alexandria, Port-Said, Suez, and Ismailia. Lower Egypt includes Damietta, Dakahliya, Sharkiya, Kafr El-Sheikh, Gharbiya, Menofiya, and Behira. Upper Egypt includes Beni Sueif, Fayyoum, Menia, Assuit, Souhag, Qena, Aswan, and Luxor.

[16] This generally overlaps with who makes the most decisions, a variable added in 2018 (Krafft, Keo, & Fedi, 2019).

[17] Individuals who did not report being employed in the preceding three months were classified as blue-collar non-craft.

crude measure of productivity. We also use indicators for whether the firm hires relatives from outside the household. A similar measure indicates whether the firm hires any non-relatives.[18]

## 2.4. Firm Dynamics

Since the ELMPS follows households in four waves, 1998, 2006, 2012 and 2018, we can trace the dynamics of firms in our data set. Dynamics can occur over three pairs of waves: 1998-2006, 2006-2012, or 2012-2018. An observation for our dynamic model thus has data from a base wave and a subsequent wave, e.g. for 1998-2006 the base wave is 1998 and the subsequent wave is 2006.

Creating a panel of firms from a household survey presents a number of complexities. We restrict all analyses to cases where the household was observed in both the base and subsequent wave of a pair, so we do not classify a firm as being created simply because a household joined our data set, nor a firm exited because the household attrited. We do track splits, so cases such as where an unmarried son may have had a household enterprise and then took it with him to his new household when he married and moved out (split) are captured as a firm surviving, so long as we successfully located the split household. We define an enterprise that existed in the base wave as having survived over a pair of waves if at least one household member still has an enterprise of the same economic activity (identified on the one-digit level using ISIC 4 coding). We examine the outcome of exit (closure) as the complement to an enterprise surviving. We also consider the formation of new household enterprises, that is, among households that were not engaging in an economic activity in the base wave of a pair, new enterprises in the subsequent wave. We further consider – among enterprises that survived across a pair of waves – the dynamics of (in)formalization, that is, changing formality status between the base and subsequent wave. We also consider, again among surviving enterprises across a pair of waves, employment growth, specifically whether the total number of workers in the enterprise is larger in the subsequent period than it was in the base period.

## 3. Methods

### 3.1. Situating our universe of firms and employment

We begin our analyses by situating our focus – private non-agricultural establishments with 1-24 workers – within the landscape of employment in Egypt. We describe the percentage of employment and the number of workers in this segment of the economy, in comparison to other sectors. We also describe employment formality by segment; note that this is distinct from firm formality, both in terms of the denominator, and in that, while a firm must be formal to have formal employees, not all employees within a formal firm are necessarily formal.

---

[18] The distinction between relatives and non-relatives is not available in 1998 or 2006, so when analyzing dynamics we simply use an indicator for employment outside the household, which is available in all waves.

**3.2. Comparisons of establishment characteristics and formality across data sets**

We then focus on our universe of interest, comparing the formality of private non-agricultural establishments with 1-24 workers across the three data sets, using various definitions. We start by examining the distribution of firms along various characteristics across data sets to ascertain if they represent similar firms. Subsequently, we examine the patterns of formality across the common characteristics and data sets. We present our results descriptively and also test whether there are significant differences using regressions. Specifically, we pool the three data sets and run a logit model for probability of being formal, with all the common characteristics fully interacted with a categorical variable indicating the data source, to understand whether there are significant differences in formality across data sources.

**3.3. Predictors of formality**

A key focus of our research is investigating what factors predict formality and whether formal and informal firms are comparable or different. We run logit models to predict formality. In "Model 1" we include only the few characteristics that are common across all three data sources. "Model 2" includes additional firm characteristics and can only be run on the EcC and ELMPS (with a somewhat different set of variables in each survey; see above). "Model 3," which adds the owner's characteristics, is only possible for the ELMPS.

**3.4. Comparing Distributions of Predicted Probability of Formality by Formality Status**

After running the various models mentioned above, we predict the probability of being formal using each model and data source. We compare the distribution of the predicted probability of formality by whether or not firms are actually formal. We take the size of the overlap in the distribution functions of predicted probability across formal and informal firms as a test and metric of whether informal firms are similar to formal firms and thus might potentially formalize. We acknowledge this overlap is, in part, driven by the number of covariates included in the model. We also look at the proportion of informal firms in different quartiles of the predicted probability to ascertain the proportion of informal firms that are most eligible for formalization.

**3.5. Grouping Analysis**

We undertake grouping of firms, by characteristics that were particularly important in our regression models, as one approach to creating profiles of firms that vary in their potential to formalize. We specifically group together informal firms with common characteristics and explore the proportion in the different quartiles of the predicted probability of formality. We undertake this work only for the ELMPS (model 3) and EcC (model 2), since the EsC has a very limited set of covariates.

For the ELMPS, we pool 2012 and 2018. We group on firm size, firm age, current capital, whether the firm hired a non-relative outside of the household, quartile of the log of revenue per worker, and owner's education level. We use a collapsed set of categories for these covariates to have a manageable number

of resulting groups and reasonable cell size. We further collapse categories as needed.[19] After this exercise we have 43 groups all of which have 10 or more informal firms.

Similarly, we group firms in the EcC. We use firm size, firm age, invested capital, whether the firm hired a wage worker, and quartile of value added per worker. Similar to ELMPS, after obtaining the number of informal firms for each combination of categories from these variables, we impose the constraint that no category can have fewer than 10 informal firms.[20] After this exercise we have 76 groups all of which have 10 or more informal firms.

### 3.6. Cluster Analysis

As an additional approach to identifying profiles of firms with varying potential for formalization, we undertake cluster analysis. Cluster analysis is a data reduction and classification method, designed to create a small number of clusters (groups) where we have maximized similarity within the groups and difference between the groups (Everitt, Landau, Leese, & Stahl, 2011). We use k-medians clustering and Gower's dissimilarity coefficient (since it is amenable to a mix of binary and continuous data) to create 10 clusters in each of the ELMPS (pooling 2012 and 2018) and EcC. For ELMPS, we use the variables in model 3 for the cluster analysis. For EcC, we use the variables in model 2 for the cluster analysis.

After obtaining the clusters, we explore the characteristics of each of them. Based on the characteristics of these clusters, we identify the delineating features for each clusters and name clusters accordingly. We present descriptive statistics on categorical variables showing the ratio of each category within the cluster relative to the category overall in the sample.[21] In addition, we standardize the continuous variables in each of the clusters relative to the mean and standard deviation of that variable in the sample.

### 3.7. Dynamics

We descriptively present the relationship between firm dynamics and formality over pairs of waves in terms of five measures: (1) firm exit, (2) the share of new vs. not new firms that are formal, (3) the share of firms that are new by formality, (4) changing formality status by base wave formality status (conditional on firm survival), and (5) employment growth (as a binary indicator) by base wave formality. We then

---

[19] We impose the constraint that no category can have fewer than 10 informal firms and aggregate groups that have fewer than 10 informal firms with groups of similar characteristics (i.e., all variables have the same value except the one that is used for aggregation) using the following algorithm: i) group quartiles of the log of revenue per worker on the same side of the median together, i.e., quartiles 1 and 2 together, and quartiles 3 and 4 together; ii) aggregate secondary and above secondary education levels of the owner together; iii) aggregate all capital categories together; iv) aggregate all levels of owner's education together; v) aggregate all quartiles of log of revenue per worker; and vi) aggregate all categories of hiring non-relatives from outside the household.

[20] To aggregate groups that have fewer than 10 informal firms with groups of similar characteristics, i) we first group quartiles of the log of revenue per worker on the same side of the median together, i.e., quartiles 1 and 2 together, and quartiles 3 and 4 together; ii) and then aggregate all quartiles of log of revenue per worker.

[21] For ELMPS, we use the sample mean of each wave.

estimate logit models of firm exit, changing formality, and employment growth using base wave characteristics. Since entering firms do not exist in the base wave, we do not estimate models for new firms.

**4. Results**

**4.1. Situating our universe of firms and employment**

We focus on non-agricultural, private sector employment, in firms in establishments with 1-24 workers. This is about a fifth of employment in Egypt's economy (22%, Table 1), a quarter of employment is in the public sector (24%), another quarter outside of establishments (24%) and a further fifth (19%) in agriculture. Lastly, 11% of employment is in non-agricultural private firms in establishments that have 25+ employees. The segment we focus on represents 5.3 million workers, out of the 24.4 million workers in Egypt.

The table also measures employment formality in terms of social insurance. Not even in the public sector do all employees have social insurance (only 82% do). Among non-agricultural private firms in establishments with 25+ workers, half (52%) of workers have social insurance. The share is much lower, only 12%, among those working in non-agricultural private sector firms with 1-24 workers, and lower still outside establishments (8%) and in agriculture (2%). It is important to note that many of the workers who lack social insurance may still be in firms that are registered or otherwise formal, so this is an under-estimate of workers who are working in formal firms. Hereafter we focus on non-agricultural, private sector employment in establishments with 1-24 workers and turn to the firm as the unit of analysis.

**Table 1. Employment and employment formality by firm type**

|  | Percentage of emp. | No. of emp. in millions | Percentage w/ soc. ins. |
|---|---|---|---|
| **Public sector** | 24 | 5.9 | 82 |
| **Private agriculture** | 19 | 4.6 | 2 |
| **Non-ag. private outside est.** | 24 | 5.9 | 8 |
| **Non-ag private in est. 1-24** | 22 | 5.3 | 12 |
| **Non-ag private in est. 25+** | 11 | 2.8 | 52 |
| **Total** | 100 | 24.4 | 31 |

Source: Authors' calculations based on ELMPS 2018

## 4.2. Comparison of firm formality and firm characteristics across data sources

### 4.2.1. Firm formality by data source and definition

As explained above, we have two increasingly expansive measures of firm formality. The basic measure depends exclusively on commercial registration. The extended measure adds to registration paying social insurance premiums or taxes (depending on the data source). We examine the share of firms that are formal according to these two measures for each data source, across firms in establishments and out of establishments (the latter using ELMPS only). Figure 1 shows that firms outside establishments have very low levels of formality across the two measures. From 6-13% of firms outside establishments are formal (depending on the wave and measure). This low rate of formality for firms outside of establishments suggests that firm location – and in particular being in a fixed establishment as compared to outside an establishment – is a key predictor of formality. The low rates of firm formality outside establishments mean that this segment is unlikely to be formalizable. Firms outside establishments – inside one's own home or involving mobile work – may face different relative costs and benefits to formalization, as well as different visibility and enforcement.

**Figure 1. Percentage of firms that are formal by measure and data source**
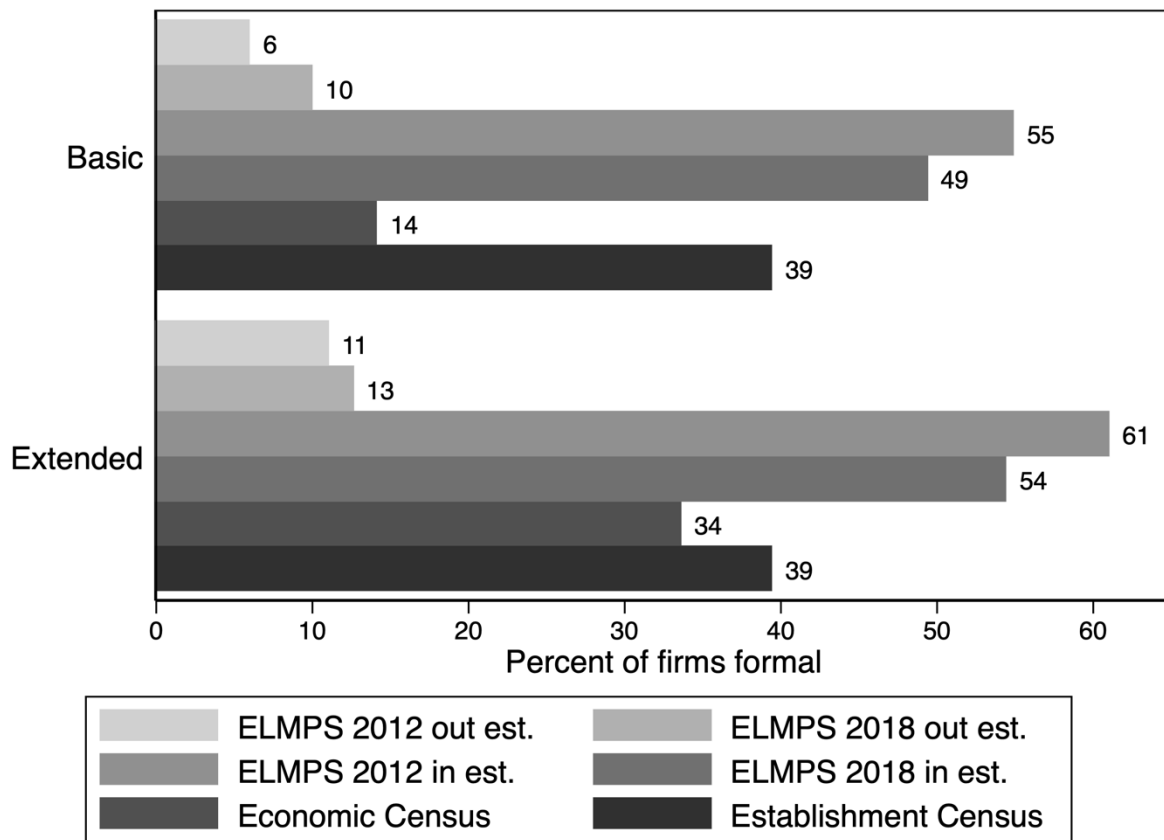


Source: Authors' calculations based on ELMPS 2012, ELMPS 2018, EcC 2012/13, EsC 2017

Focusing now on firms in establishments, the proportion of firms that are formal according to the basic definition is comparable across the two ELMPS waves (49%-55%), somewhat lower in the EsC (39%), and much lower in the EcC (14%), likely because of the way that some EcC survey instruments excluded the question about registration. Moving to the extended definition more than doubles the share of firms that are formal in the EcC (34%), does not change the estimate of 39% for EsC, and raises the estimate of formality by a few percentage points (to 54-61%) in the two ELMPS waves. Thus, the extended estimates are broadly consistent with each other and suggest that somewhere from over a third to half of all private non-agricultural firms of fewer than 25 workers operating in establishments are formal, depending on the data source. Hereafter, we use the extended definition of formality as our primary measure and focus on firms that operate inside fixed establishments in our analysis. We note any exceptions to this approach.

Although Figure 1 shows the rate of formality for firms, since we have the number of employees in each firm, we can also calculate the number of employees in each data source within the universe of non-agricultural firms with 1-24 workers in establishments. Likewise, among these workers, we can calculate the number within formal firms versus informal firms. In the ELMPS 2012, using the extended definition, there are 3.5 million workers overall and 2.4 million within formal firms – although not necessarily formal employment. This compares to 5.8 million workers in the contemporaneous Economic Census, of whom 2.6 million were within formal firms. In the ELMPS 2018, there were 3.0 million workers, of whom 1.9 million would be within formal firms. The Establishment Census generates a much larger number of workers, 9.2 million, of whom 4.5 million were within formal firms.

### 4.2.2. Distribution of firm characteristics across data sources

We now examine firm characteristics across the four data sets to determine whether we are observing roughly the same firms. As shown in Figure 2, the two waves of the ELMPS have a somewhat higher proportion of one-person firms and are conversely somewhat under-represented in the three larger size categories. This may be due to branches of firms, which tend to be larger, being included as separate establishments in the EcC and EsC. Despite this tilt toward smaller firm sizes, the ELMPS data produces higher estimates of formality than the other two sources, as we have seen above. Besides the fact that EcC has a higher representation of two-person firms (35%) and the EsC a higher representation of three- or four-person firms (25%), these two data sources produce fairly comparable estimates in the three other size categories. Overall, firms in our universe skew toward the smaller size categories, with one or two-person firms making up 66% to 86% of all firms in the universe and firms of 5-24 workers making up at most 10%.

**Figure 2. Number of workers by data source (percentage of firms)**
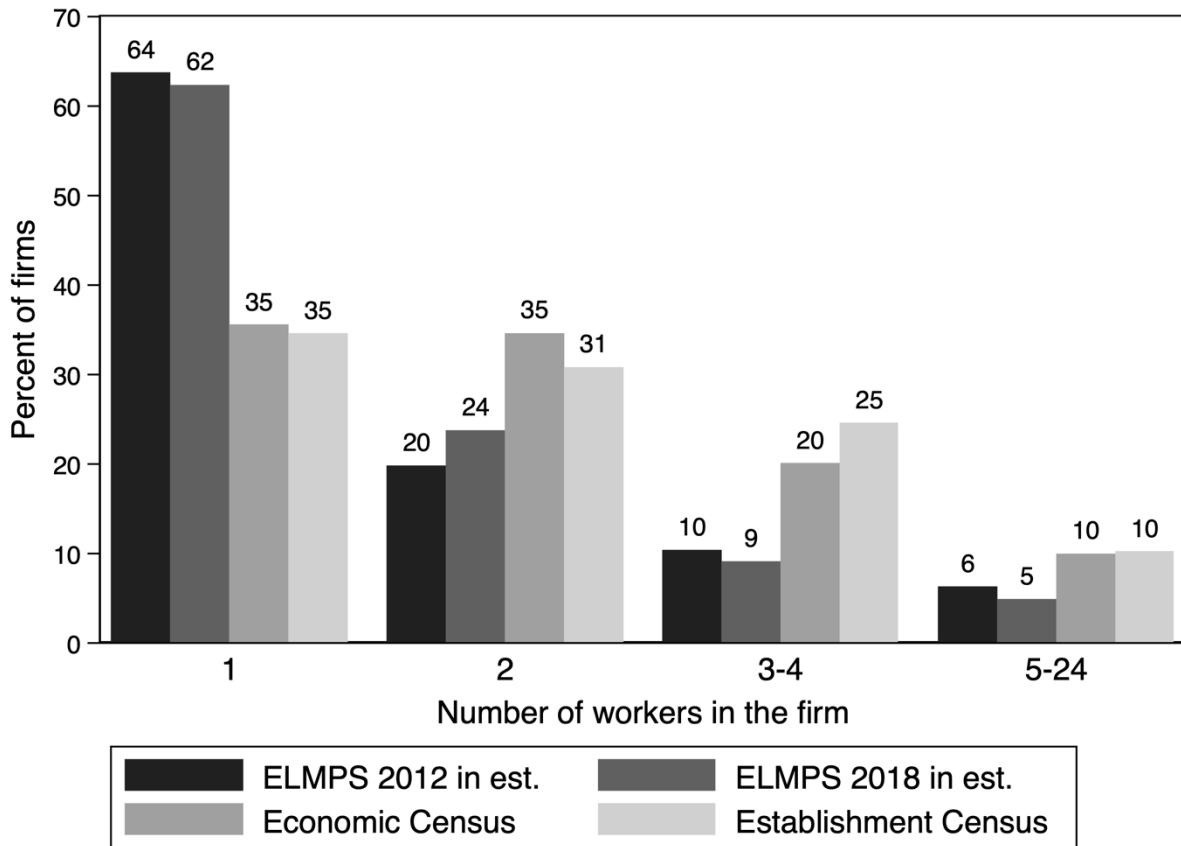


Source: Authors' calculations based on ELMPS 2012, ELMPS 2018, EcC 2012/13, EsC 2017

Figure 3 shows firm age, which was calculated based on start year. Since surveys were fielded at different times in the year, differences in firms aged zero should be interpreted with some caution, as these are likely driven by fielding time more so than differences in the firms captured. Overall, the EcC tends to capture younger firms, with 34% aged 0-2. The three surveys agree that 7-8% of firms are three years old. The EsC then finds more firms that are 4-7 years old, although the sharp spike at 7 is likely driven by heaping, in that this was a report of a start year of 2010. The ELMPS has more older firms, with 28-31% aged 8-17 and 14-16% aged 18-27. The ELMPS may be capturing longstanding but micro family businesses or self-employment more effectively.

**Figure 3. Firm age in years by data source (percentage of firms)**
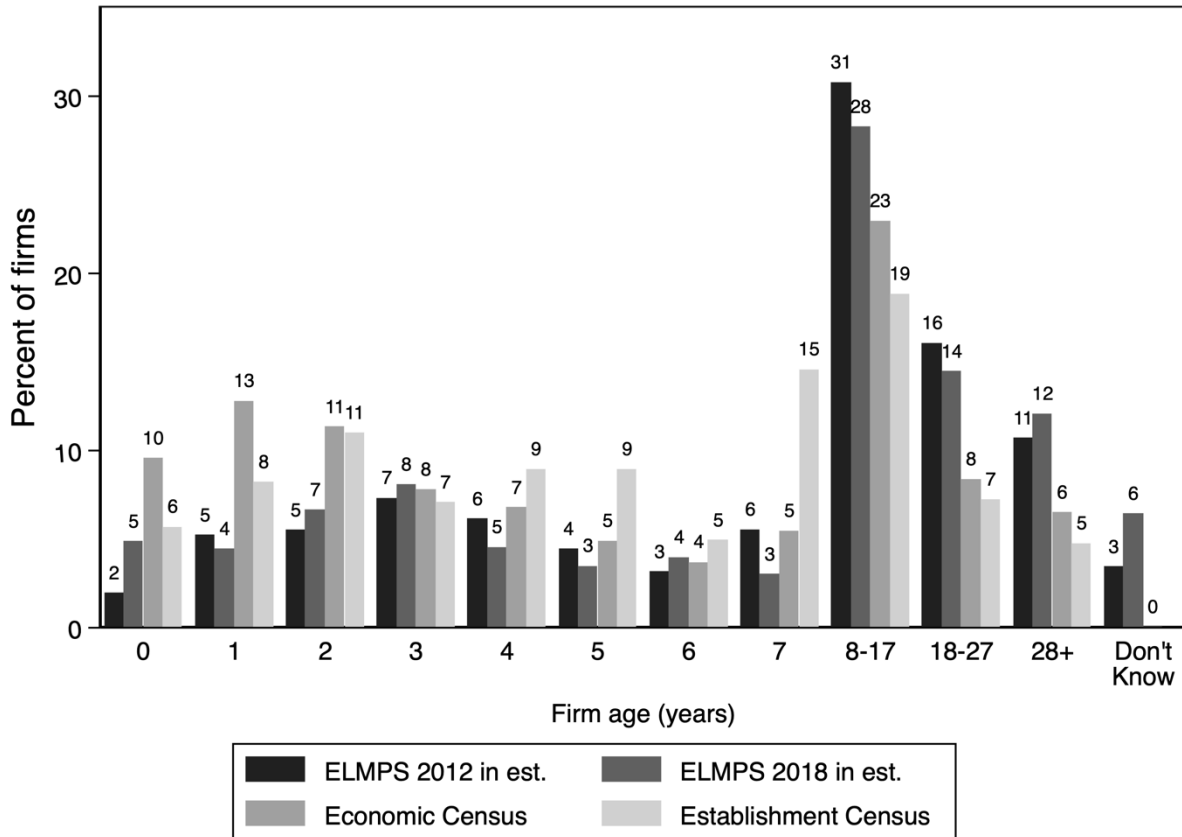


Source: Authors' calculations based on ELMPS 2012, ELMPS 2018, EcC 2012/13, EsC 2017

The distribution of firms across regions is fairly consistent across the four data sets. As shown in Figure 4, the highest share of firms is in Lower Egypt, which accounts for more than a third of firms (with estimates ranging from 37% to 42%). Greater Cairo and Upper Egypt each account for about a quarter of firms, with the ELMPS 2018 somewhat understating Greater Cairo and overstating Upper Egypt compared to the other three data sets. The Alexandria and Suez Canal region accounts for the remaining 9-15%.

**Figure 4. Region by data source (percentage of firms)**
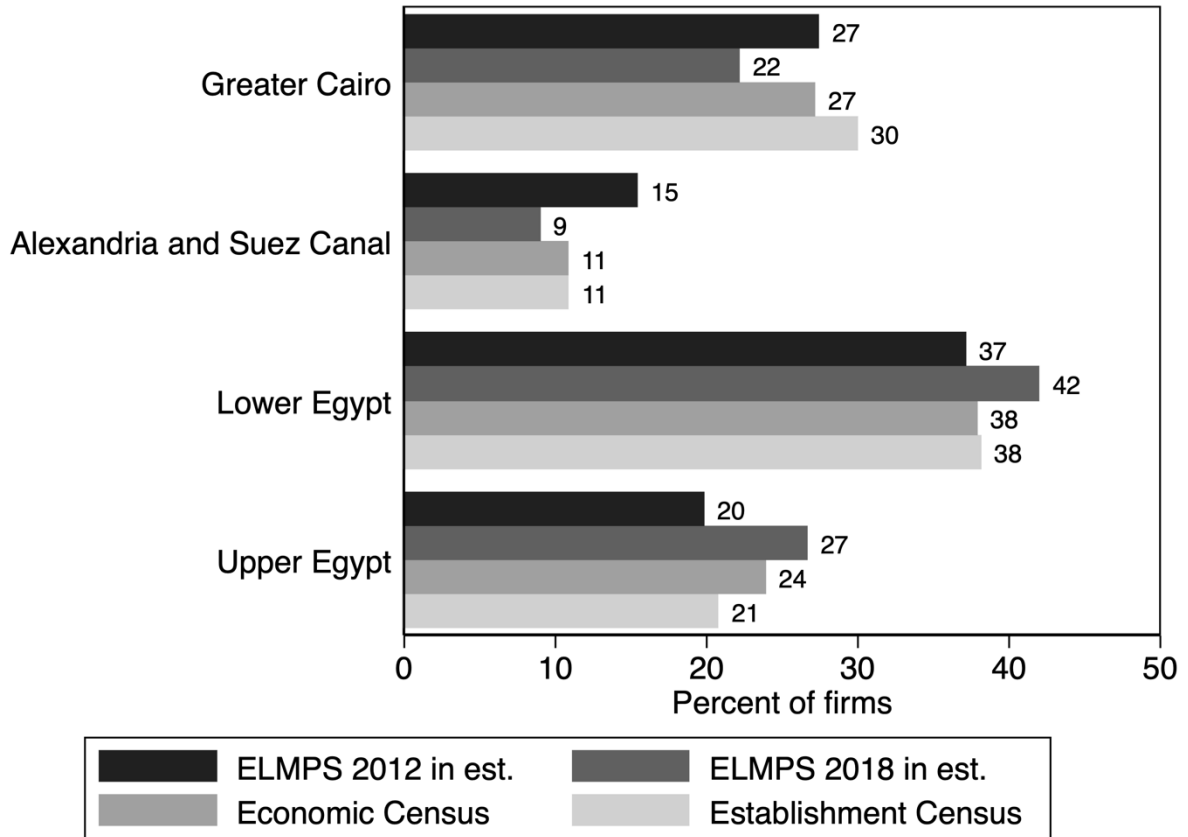


Source: Authors' calculations based on ELMPS 2012, ELMPS 2018, EcC 2012/13, EsC 2017

The distribution of firms across broad industry groups is also quite consistent across data sets. As shown in Figure 5, manufacturing and related trades (13-16%), accommodation and food service (4-6%), and various professional activities (10-12%) are similarly represented across the different sources. The establishment census under-represents retail and wholesale trade (53% vs. 58-60%) and over-represents construction (3%) and transportation and storage firms (6%) compared to the other three sources (less than 1% for transportation and storage; 0-2% for construction). Other services are somewhat under-represented in the ELMPS waves (7%) compared to the EcC and EsC (9-10%).

**Figure 5. Industry by data source (percentage of firms)**
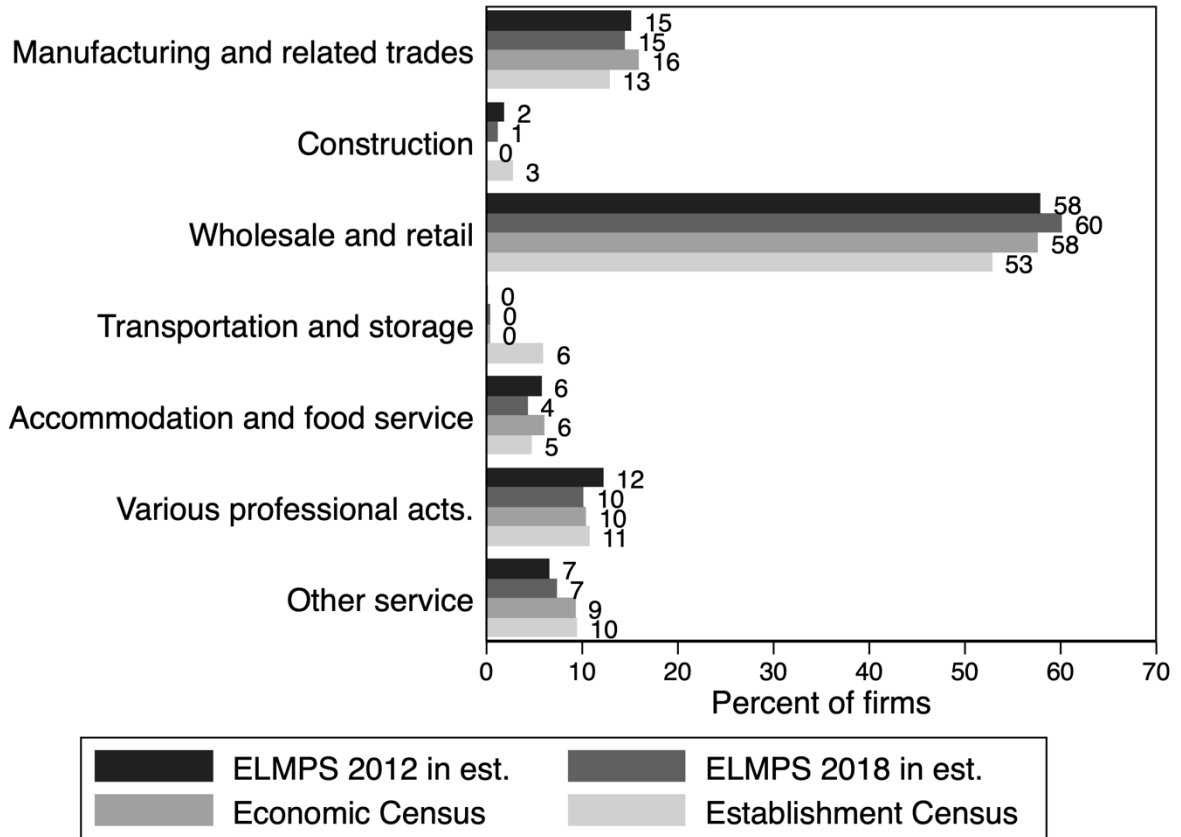


Source: Authors' calculations based on ELMPS 2012, ELMPS 2018, EcC 2012/13, EsC 2017

### 4.2.3. Formality by firm characteristics

We now examine how the percentage of firms that are formal varies across data sets and firm characteristics. We present graphs one characteristic at a time and discuss whether there are significant differences compared to ELMPS 2018 based on a pooled model interacting data and covariates (Table 11 in the Appendix). It is useful to note at the outset that all data sets show an increasing formality rate by firm size (Figure 6), with the exception of a slightly lower rate in 3-4 worker firms (61%) than 2 worker firms (63%) in ELMPS 2018. We also note that the two ELMPS waves indicate a higher rate of formality across all size categories. After accounting for overall higher reporting in the ELMPS, only differences for firm size 3-4 are significant.

16

**Figure 6. Percentage of firms that are formal by number of workers and data source**
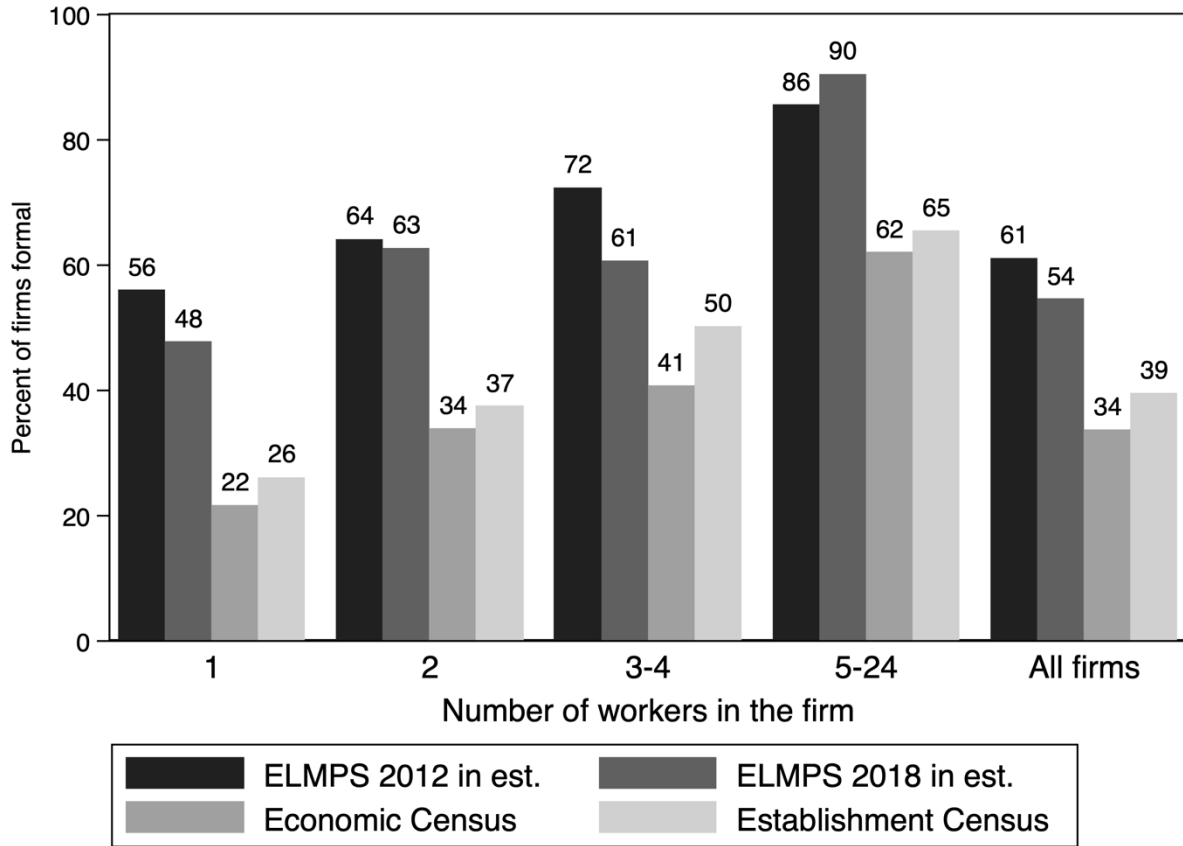


Source: Authors' calculations based on ELMPS 2012, ELMPS 2018, EcC 2012/13, EsC 2017

Formality generally rises with firm age (Figure 7). For example, the EsC reports 23% of firms aged zero are formal, rising steadily each year to 40% by age 7, 49% for ages 8-17, 56% for ages 18-27 and 64% for ages 28+. All the surveys follow the same pattern, albeit at different levels and with some noisiness in the ELMPS given the small sample size for any particular single year of firm age. After accounting for the average level of formality in each data set, only three of the 30 coefficients are significant for interactions between firm age and data set, all for age 18-27, where the ELMPS 2018 has a low rate of formality.

**Figure 7. Percentage of firms that are formal by firm age**
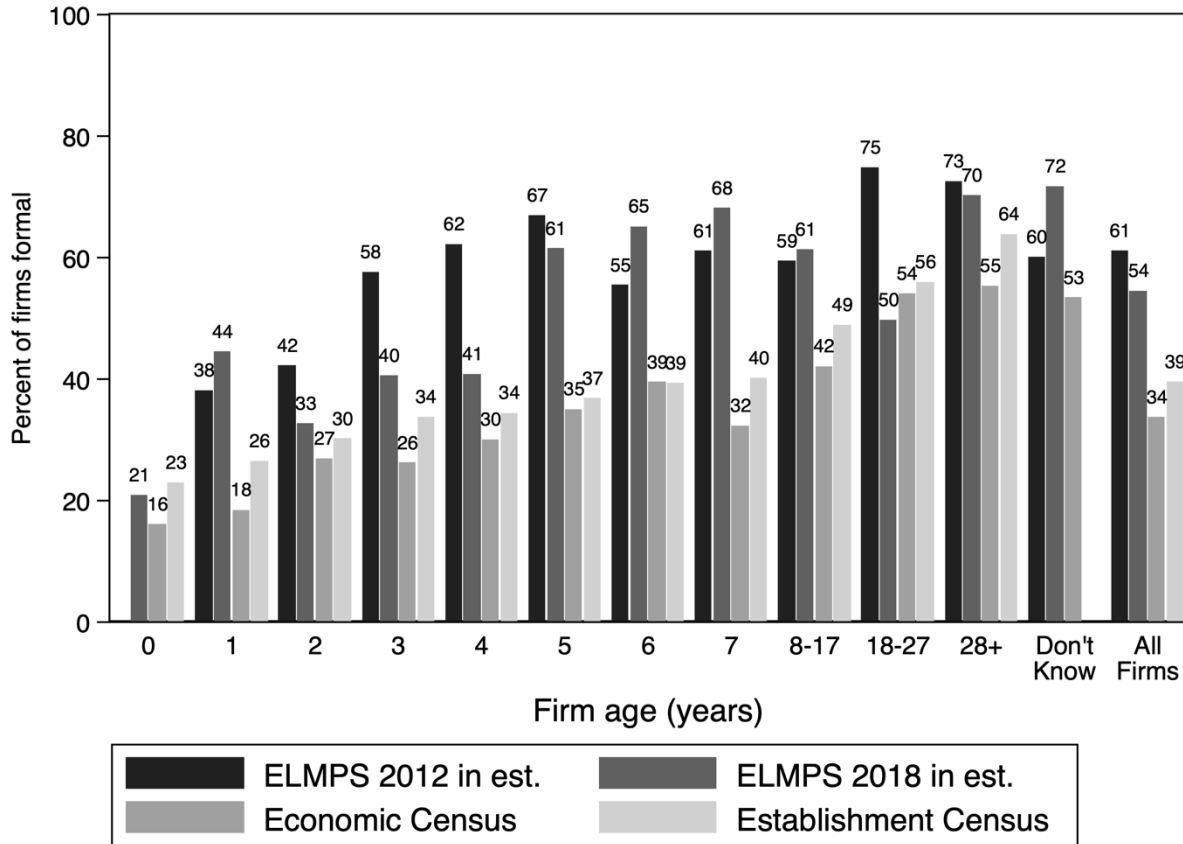


Source: Authors' calculations based on ELMPS 2012, ELMPS 2018, EcC 2012/13, EsC 2017

Note: Age 0 suppressed in ELMPS 2012 and don't know in EsC due to N<30.

Similar patterns to those for firm age and size can be observed if we examine formality by region in Figure 8. The two ELMPS waves report higher rates of formality and the EcC lower rates in all regions except Upper Egypt, where the rates are comparable across EcC and EsC. After accounting for overall higher reporting in the ELMPS, there is a significantly different pattern in the EcC by region, but not the EsC. Notably, the rates of formality are comparable across Greater Cairo, Lower Egypt, and Upper Egypt in both EsC (36-40% formal) and the ELMPS (53-57% formal in 2018). Alexandria and Suez Canal are also similar in the ELMPS (55% formal in 2018).

**Figure 8. Percentage of firms that are formal by region**
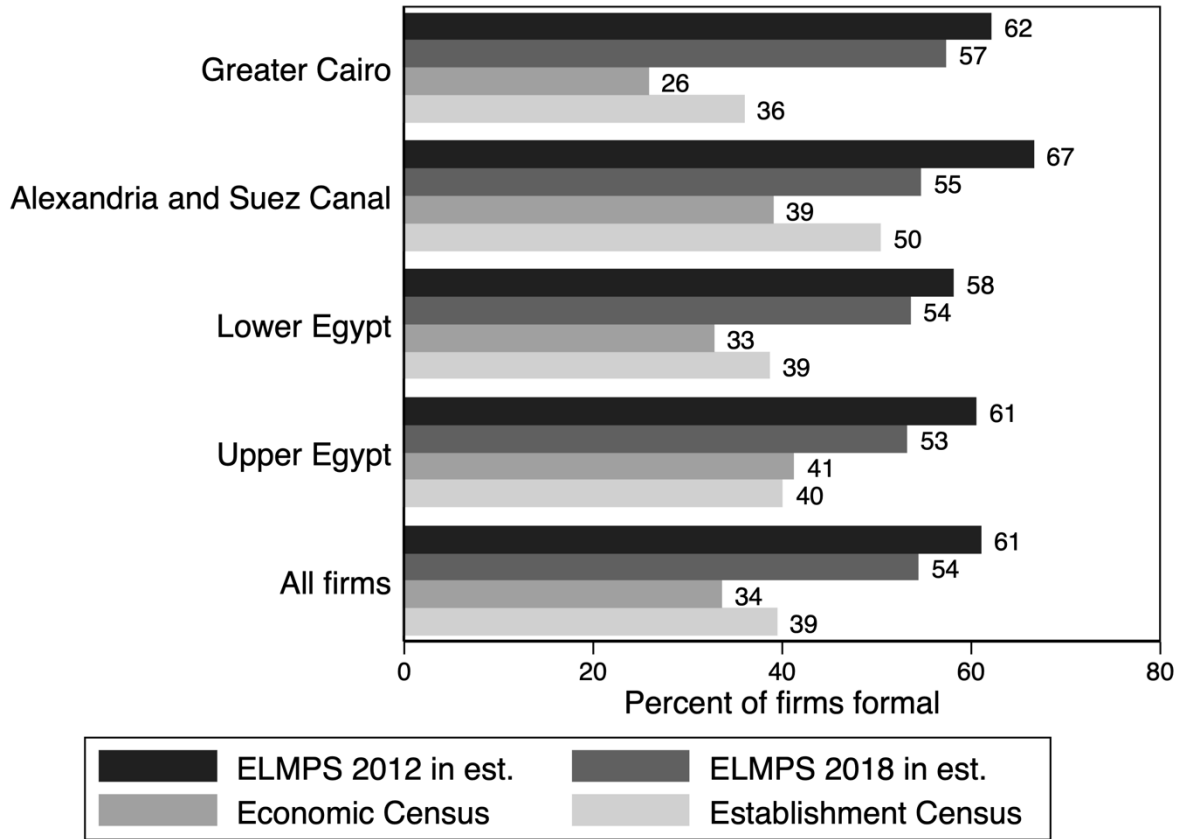


Source: Authors' calculations based on ELMPS 2012, ELMPS 2018, EcC 2012/13, EsC 2017


Figure 9 reveals that there is not a strong relationship between formality and industry. The formality rate across industries varies across a fairly narrow range, with the exception of the EcC which reports an implausibly high formality rate for the construction industry. After accounting for overall higher rates in the ELMPS, there are some significant differences – especially for smaller categories, like construction – in the relationship between industry and formality by data source.

**Figure 9. Percentage of firms that are formal by industry**
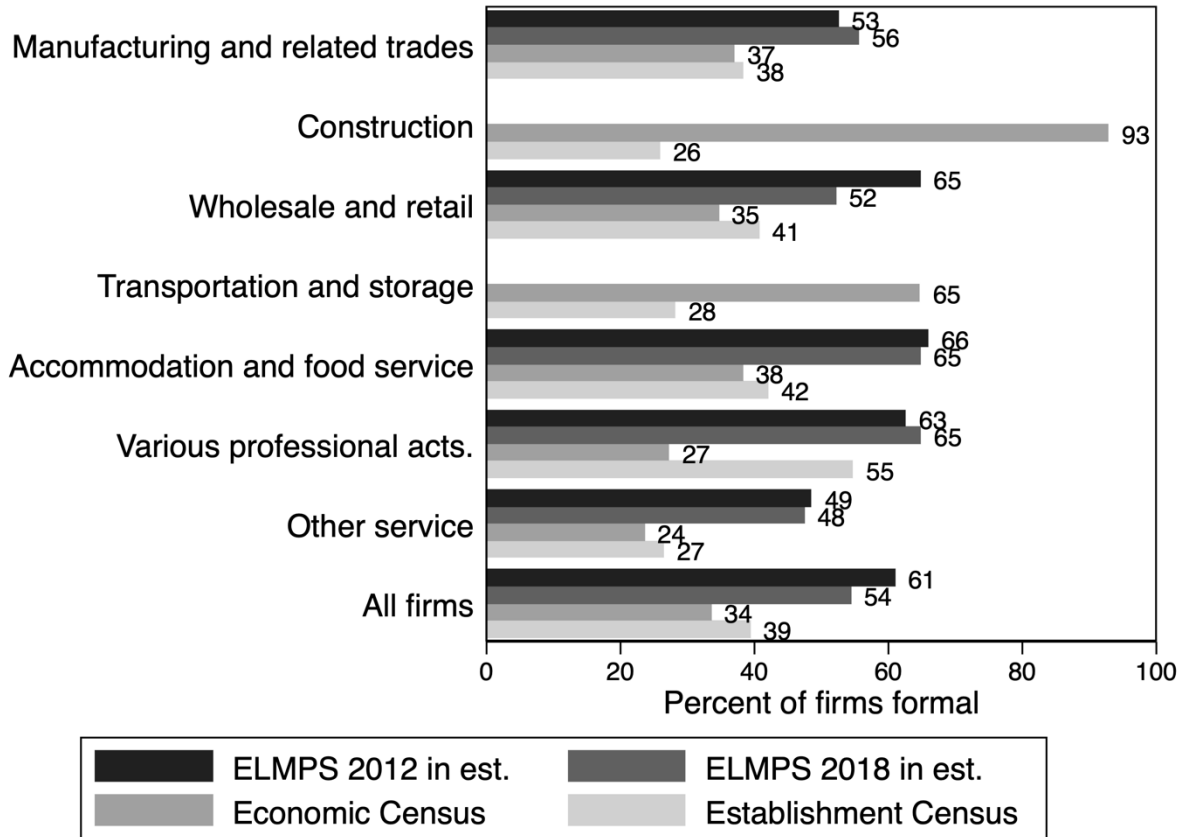


Source: Authors' calculations based on ELMPS 2012, ELMPS 2018, EcC 2012/13, EsC 2017

Notes: The construction and transportation and storage categories for ELMPS suppressed due to N<30

Overall, examining the distribution by data source and formality by data source suggests that there are similar patterns in the data and relationships captured by the different data sources. The ELMPS has higher reported formality rate overall but, after accounting for that, limited differences in the relationship between characteristics and formality.

### 4.3. Modeling formality

### 4.3.1. Predictors of formality

We estimate a series of increasingly rich logit models of predictors of formality. Model 1, shown in Table 2, is a relatively parsimonious model that includes the variables that are common to all four data sets. Model 2a, shown in Table 3, includes additional firm-related characteristics only available in the EcC, such as capital-to-labor ratio, value-added per worker, the presence of unpaid and wage workers in the firm,

and the percentage of workers female. Model 2b, shown in Table 4, includes additional firm-related characteristics only available in the ELMPS data sets, such as the size of the firm's capital (in categories), net revenue per worker, the presence of non-household members among the workers, separately for relatives and non-relatives, and the type of workplace. Model 3, also shown in Table 4, adds to the firm-related characteristics in Model 2b some characteristics related to the firm owner, specifically the owner's sex, educational attainment, broad occupational group, and age group.

We present in Table 2 odds ratios from a logit regression for the probability of being formal using the common set of variables across all four data sets (Model 1). In the first column we show results for the pooled data, with dummy variables indicating the data source. Separate results by data set are shown in the four subsequent columns. As expected, firm size is a strong predictor of formality in all four data sets. According to the pooled model, compared to one-person firms, the odds of being formal are 66% higher for two-worker firms, more than twice as high for 3-4 worker firms, and more than five times as high for 5-24 worker firms.

Older firms are significantly more likely to be formal. Compared to firms aged one year old, those zero are less likely to be formal, usually significantly so. The surveys vary in terms of which ages start to predict significantly higher formality, with the ELMPS 2018 showing a later and weaker relationship. In the pooled model, formality largely rises steadily with age, and the "don't know" firms tend to follow the pattern of older firms (likely don't know means a firm is old enough that the owner cannot remember when, exactly, it started).

All industry groups except "various professional activities" have lower odds of formality compared to wholesale and retail trade, usually significantly so in the pooled model. The industry that is least likely to have formal firms, on average, is construction, followed by transport and storage. However, these results appear to differ across data set. Construction is least likely to be formal in the ELMPS 2012 and highly likely to be formal in EcC, but that effect is measured with a great deal of imprecision. Similarly, transport and storage is less likely to be formal in the two ELMPS waves than in the EcC. It should be kept in mind that few construction and transport firms are actually within establishments, so the group showing up here may be highly and differently selected.

As suggested by the descriptive statistics, firms in the Alexandria and Suez Canal region have higher odds of being formal than those in the Greater Cairo Region. Surprisingly, according to two of the four data sets (EcC and EsC), firms in Lower Egypt and Upper Egypt, two provincial regions, also have higher odds of being formal than those in Greater Cairo. This strongly suggests that being in a metropolitan region does not necessarily raise the odds of formality.

The pooled data results confirm the pattern observed in the descriptives that the two ELMPS waves produce higher measures of formality than the EcC and the EsC. After correcting for size, industry and region, the EcC produces odds of formality that are three-fifths lower and the EsC produces odds that are about half of what they are in ELMPS 2018.

**Table 2. Odds ratios from logit "Model 1" of extended formality by data source**

| | Pooled | ELMPS 2012 in est. | ELMPS 2018 in est. | Economic Census | Establishment Census |
|---|---|---|---|---|---|
| **Data (ELMPS 2018 in est. omit)** | | | | | |
| ELMPS 2012 in est. | 1.349** | | | | |
| | (0.147) | | | | |
| Economic Census | 0.400*** | | | | |
| | (0.035) | | | | |
| Establishment Census | 0.532*** | | | | |
| | (0.045) | | | | |
| **Ent. size (one worker omit.)** | | | | | |
| 2 | 1.655*** | 1.462 | 1.887*** | 1.814*** | 1.628*** |
| | (0.094) | (0.292) | (0.358) | (0.130) | (0.011) |
| 3-4 | 2.414*** | 2.096** | 1.394 | 2.651*** | 2.750*** |
| | (0.141) | (0.506) | (0.368) | (0.275) | (0.019) |
| 5-24 | 5.661*** | 7.228*** | 9.753*** | 7.441*** | 5.108*** |
| | (0.402) | (3.428) | (5.122) | (0.622) | (0.051) |
| **Ent. age (one year old omit.)** | | | | | |
| 0 | 0.728** | 0.685 | 0.339* | 0.825 | 0.831*** |
| | (0.078) | (0.442) | (0.167) | (0.094) | (0.013) |
| 2 | 1.165 | 0.995 | 0.668 | 1.622*** | 1.181*** |
| | (0.108) | (0.420) | (0.308) | (0.162) | (0.015) |
| 3 | 1.461*** | 1.903 | 0.959 | 1.601*** | 1.353*** |
| | (0.168) | (0.730) | (0.456) | (0.172) | (0.019) |
| 4 | 1.619*** | 2.400* | 0.937 | 1.830*** | 1.410*** |
| | (0.150) | (0.958) | (0.439) | (0.199) | (0.018) |
| 5 | 1.946*** | 3.073* | 2.116 | 2.255*** | 1.547*** |
| | (0.190) | (1.451) | (1.127) | (0.280) | (0.020) |
| 6 | 2.063*** | 1.678 | 2.681 | 2.507*** | 1.706*** |

| | Pooled | ELMPS 2012 in est. | ELMPS 2018 in est. | Economic Census | Establishment Census |
|---|---|---|---|---|---|
| | (0.239) | (0.805) | (1.360) | (0.321) | (0.026) |
| 7 | 1.907*** | 2.162 | 2.875 | 1.976*** | 1.725*** |
| | (0.162) | (0.870) | (1.590) | (0.341) | (0.020) |
| 8-17 | 2.426*** | 2.401** | 1.931 | 2.780*** | 2.449*** |
| | (0.196) | (0.734) | (0.712) | (0.314) | (0.028) |
| 18-27 | 3.190*** | 4.597*** | 1.299 | 4.899*** | 3.223*** |
| | (0.326) | (1.601) | (0.514) | (0.490) | (0.044) |
| 28+ | 4.302*** | 4.414*** | 2.869* | 5.045*** | 4.795*** |
| | (0.495) | (1.677) | (1.222) | (0.542) | (0.074) |
| Don't know | 3.313*** | 2.166 | 3.590** | 1.859*** | |
| | (0.842) | (1.021) | (1.685) | (0.287) | |
| **Industry (wholesale/retail omit.)** | | | | | |
| Manufacturing and related trades | 0.661*** | 0.442*** | 0.972 | 0.757*** | 0.639*** |
| | (0.049) | (0.096) | (0.256) | (0.054) | (0.005) |
| Construction | 0.362*** | 0.147*** | 0.491 | 7.319** | 0.523*** |
| | (0.085) | (0.080) | (0.480) | (4.675) | (0.009) |
| Transportation and storage | 0.526*** | 0.464 | 0.563 | 1.903 | 0.514*** |
| | (0.020) | (0.469) | (0.430) | (0.714) | (0.006) |
| Accommodation and food service | 0.842 | 0.816 | 1.574 | 0.794** | 0.759*** |
| | (0.075) | (0.249) | (0.532) | (0.068) | (0.009) |
| Various professional acts. | 1.074 | 0.756 | 1.552 | 0.530*** | 1.562*** |
| | (0.075) | (0.183) | (0.402) | (0.041) | (0.013) |
| Other service | 0.620*** | 0.488** | 0.922 | 0.715*** | 0.537*** |
| | (0.046) | (0.130) | (0.251) | (0.051) | (0.005) |
| **Region (Greater Cairo omit.)** | | | | | |
| Alexandria and Suez Canal | 1.667*** | 1.227 | 1.086 | 2.139*** | 2.048*** |
| | (0.130) | (0.291) | (0.356) | (0.162) | (0.019) |

|  | Pooled | ELMPS 2012 in est. | ELMPS 2018 in est. | Economic Census | Establishment Census |
|---|---|---|---|---|---|
| Lower Egypt | 1.331*** | 0.948 | 0.937 | 1.956*** | 1.438*** |
|  | (0.081) | (0.187) | (0.233) | (0.123) | (0.009) |
| Upper Egypt | 1.431*** | 0.932 | 1.022 | 2.738*** | 1.414*** |
|  | (0.096) | (0.202) | (0.263) | (0.282) | (0.011) |
| N | 766840 | 1200 | 1155 | 52287 | 712198 |
| Pseudo R-squared | .113 | .0821 | .0953 | .119 | .0989 |

Source: Authors' calculations based on ELMPS 2012, ELMPS 2018, EcC 2012/13, EsC 2017

Notes: *p<0.05; **p<0.01; ***p<0.001.

We now move to the results from Model 2a estimated in the EcC data and shown in Table 3. The results on size are in the same direction as before, but are somewhat attenuated because of the inclusion of other regressors associated with firm size. The results on firm age, industry, and region are little changed when compared to the results of Model 1 on the EcC data. With regard to the new variables added, we see that collective ownership as opposed to sole proprietorship slightly and significantly raises the odds of formality. Higher labor productivity is associated with substantially higher odds of formality, as is higher capital intensity. As expected, having unpaid family workers reduces the odds of formality, but having wage workers increases it. The proportion of female workers in the firm's workforce is not associated with the odds of formality.

**Table 3. Odds ratios from logit "Model 2a" of extended formality in EcC 2012/13**

| Ent. size (one worker omit.) | |
| --- | --- |
| 2 | 1.640*** |
| | (0.157) |
| 3-4 | 2.005*** |
| | (0.254) |
| 5-24 | 4.945*** |
| | (0.617) |
| **Industry (wholesale/retail omit.)** | |
| Manufacturing and related trades | 0.697*** |
| | (0.056) |
| Construction | 4.965** |
| | (3.018) |
| Transportation and storage | 1.579 |
| | (0.572) |
| Accommodation and food service | 0.764** |
| | (0.070) |
| Various professional acts. | 0.492*** |
| | (0.038) |
| Other service | 0.748*** |
| | (0.054) |
| **Region (Greater Cairo omit.)** | |
| Alexandria and Suez Canal | 2.146*** |
| | (0.168) |
| Lower Egypt | 2.052*** |
| | (0.139) |
| Upper Egypt | 3.085*** |
| | (0.341) |

**Partnership (sole proprietorship omit.)**

| | |
|---|---|
| Collective Ownership | 1.185* |
| | (0.101) |

**Ent. age (one year old omit.)**

| | |
|---|---|
| 0 | 0.917 |
| | (0.106) |
| 2 | 1.629*** |
| | (0.163) |
| 3 | 1.577*** |
| | (0.172) |
| 4 | 1.773*** |
| | (0.196) |
| 5 | 2.235*** |
| | (0.279) |
| 6 | 2.441*** |
| | (0.316) |
| 7 | 1.853** |
| | (0.358) |
| 8-17 | 2.635*** |
| | (0.298) |
| 18-27 | 4.727*** |
| | (0.480) |
| 28+ | 4.915*** |
| | (0.523) |
| Don't know | 2.274*** |
| | (0.390) |
| **Ln capital to labor ratio** | 1.078*** |
| | (0.023) |

| | |
|---|---|
| **Ln value added per worker** | 1.188*** |
| | (0.044) |
| **Has unpaid workers (no omit.)** | |
| Yes | 0.516*** |
| | (0.083) |
| **Has wage workers (no omit.)** | |
| Yes | 1.460*** |
| | (0.119) |
| **Percentage of workers female** | 0.999 |
| | (0.001) |
| **N** | 51323 |
| **Pseudo R-squared** | .133 |

Source: Authors' calculations based on EcC 2012/13

Notes: *p<0.05; **p<0.01; ***p<0.001.

We now turn to the results of Model 2b and Model 3 estimated on the two waves of the ELMPS data (Table 4). The results on the effect of the firm characteristics do not change appreciably once we add the owner characteristics in Model 3, so for these characteristics we discuss both models for both waves together and then move to the results on owner characteristics in Model 3.

First, we note that firm size plays the same steadily increasing and highly significant role we observed in Model 1, but, as in the case of Model 2a, adding additional regressors attenuates the effect of size. The effects of industry become more mixed and lose significance as more regressors are added to the model. We conclude from this that industry is not a reliable predictor of formality once other characteristics have been taken into account. The effect of region on formality also becomes somewhat mixed when additional regressors are added. Firms in the Alexandria and Suez Canal region are more likely to be formal only in 2012 but not in 2018.  Thus, region is also not a reliable predictor of formality. Firm age is more likely to be significant in 2012 than 2018 and effects are attenuated as more regressors are added.

As we found with the EcC data, firms with collective ownership are more likely to be formal relative to sole proprietorships across years and across the two models, but the effect is larger in 2018 than in 2012. Having more capital increases the odds of being formal, but the effect appears to kick in only when the capital exceeds EGP 1000. As in the case of labor productivity in Model 2a above, higher log net revenue per worker is associated with a higher odds of being formal.

Hiring relatives from outside the household does not seem to be associated with higher formality rates, but hiring non-relatives does, as we found to be the case with wage workers in Model 2a. The type of workplace is also associated with formality. Shops and workshop/factories are most likely to be formal, whereas being located in a flat, room or building are associated with lower odds of formality. Being in a hut or kiosk is associated with very low levels of formality in 2012, but not in 2018.

We now examine results on the characteristics of the firm owner included in Model 3. Having a female as opposed to a male owner is associated with a lower odds of being formal in 2018 but not in 2012 and in both cases is insignificant. Firms with a more educated owner have a substantially higher odds of being formal, with owners with higher education having between 3.5- and 4.0-times higher odds of formalizing their firms than illiterate owners. Beyond education, occupation does not seem to have an additional effect, expect for possibly a lower odds of formality for craft blue collar occupations, and only in 2018. Finally, older owners are more likely to have formal firms than young owners, significantly so in 2012.

**Table 4. Odds ratios from logit "Model 2b" and "Model 3" of extended formality in ELMPS 2012 and 2018**

| | Model 2 | | Model 3 | |
|---|---|---|---|---|
| | 2012 | 2018 | 2012 | 2018 |
| **Ent. size (one worker omit.)** | | | | |
| 2 | 1.224 | 1.510 | 1.215 | 1.572* |
| | (0.271) | (0.324) | (0.270) | (0.332) |
| 3-4 | 1.683 | 0.793 | 1.801 | 0.758 |
| | (0.500) | (0.236) | (0.584) | (0.241) |
| 5-24 | 3.823* | 4.572** | 3.329* | 4.460** |
| | (2.066) | (2.623) | (1.744) | (2.566) |
| **Industry (wholesale/retail omit.)** | | | | |
| Manufacturing and related trades | 0.501* | 1.093 | 0.570 | 1.382 |
| | (0.136) | (0.338) | (0.169) | (0.462) |
| Construction | 0.183** | 0.874 | 0.174* | 1.185 |
| | (0.114) | (0.777) | (0.126) | (1.064) |
| Transportation and storage | 1.533 | 0.428 | 2.666 | 0.415 |
| | (1.489) | (0.362) | (2.600) | (0.372) |
| Accommodation and food service | 1.041 | 1.416 | 1.337 | 1.236 |
| | (0.341) | (0.497) | (0.434) | (0.519) |
| Various professional acts. | 1.002 | 1.991* | 0.796 | 1.381 |
| | (0.324) | (0.626) | (0.265) | (0.493) |
| Other service | 0.756 | 0.996 | 0.891 | 1.124 |
| | (0.213) | (0.279) | (0.282) | (0.336) |
| **Region (Greater Cairo omit.)** | | | | |
| Alexandria and Suez Canal | 1.635 | 0.973 | 1.925* | 1.072 |
| | (0.423) | (0.340) | (0.527) | (0.381) |
| Lower Egypt | 1.257 | 1.049 | 1.327 | 1.078 |
| | (0.278) | (0.268) | (0.303) | (0.297) |

|  | Model 2 | | Model 3 | |
| --- | --- | --- | --- | --- |
|  | **2012** | **2018** | **2012** | **2018** |
| Upper Egypt | 1.170 | 1.171 | 1.374 | 1.269 |
|  | (0.280) | (0.306) | (0.333) | (0.362) |
| **Partnership (sole proprietorship omit.)** | | | | |
| Collective Ownership | 1.439 | 2.087** | 1.591 | 2.408*** |
|  | (0.339) | (0.512) | (0.381) | (0.624) |
| **Capital (none omit.)** | | | | |
| less than LE 100 | 1.166 | 0.580 | 1.529 | 0.511 |
|  | (0.606) | (0.292) | (0.893) | (0.265) |
| LE 100-499 | 1.637 | 0.683 | 2.545 | 0.634 |
|  | (0.810) | (0.342) | (1.381) | (0.346) |
| LE 500-999 | 1.428 | 0.898 | 1.872 | 0.627 |
|  | (0.658) | (0.389) | (0.940) | (0.285) |
| LE 1000-4999 | 2.812* | 1.017 | 4.014** | 0.822 |
|  | (1.296) | (0.441) | (2.039) | (0.379) |
| LE5000-9999 | 7.440*** | 1.469 | 9.265*** | 0.999 |
|  | (3.406) | (0.587) | (4.620) | (0.425) |
| LE10 000 or more | 4.076* | 2.431 | 6.451** | 1.893 |
|  | (2.392) | (1.331) | (4.291) | (1.072) |
| **Ln net revenue per worker** | 1.205* | 1.114** | 1.187* | 1.125** |
|  | (0.092) | (0.040) | (0.092) | (0.043) |
| **Ent. age (one year old omit.)** | | | | |
| 0 | 0.704 | 0.246** | 0.724 | 0.274* |
|  | (0.389) | (0.128) | (0.458) | (0.149) |
| 2 | 0.991 | 0.493 | 0.881 | 0.419 |
|  | (0.449) | (0.237) | (0.406) | (0.211) |
| 3 | 2.129 | 0.706 | 2.095 | 0.666 |

|  | Model 2 | | Model 3 | |
| --- | --- | --- | --- | --- |
|  | **2012** | **2018** | **2012** | **2018** |
|  | (0.847) | (0.365) | (0.889) | (0.361) |
| 4 | 2.872* | 0.702 | 3.094** | 0.629 |
|  | (1.201) | (0.356) | (1.314) | (0.334) |
| 5 | 3.511** | 1.530 | 3.965** | 1.397 |
|  | (1.662) | (0.941) | (1.855) | (0.772) |
| 6 | 1.853 | 2.332 | 2.016 | 2.959 |
|  | (0.882) | (1.301) | (1.012) | (1.916) |
| 7 | 2.016 | 1.817 | 1.967 | 2.338 |
|  | (0.850) | (1.049) | (0.850) | (1.429) |
| 8-17 | 2.388** | 1.404 | 2.117* | 1.414 |
|  | (0.748) | (0.550) | (0.730) | (0.588) |
| 18-27 | 5.658*** | 1.049 | 4.837*** | 1.125 |
|  | (2.078) | (0.437) | (1.979) | (0.499) |
| 28+ | 5.694*** | 2.346 | 3.441** | 2.349 |
|  | (2.307) | (1.060) | (1.507) | (1.166) |
| Don't know | 2.146 | 2.366 | 1.695 | 2.204 |
|  | (1.084) | (1.209) | (0.921) | (1.267) |
| **Relatives outside the HH employed (no omit.)** | | | | |
| Yes | 0.753 | 1.045 | 0.760 | 0.905 |
|  | (0.190) | (0.318) | (0.193) | (0.270) |
| **Non-relatives outside the HH employed (no omit.)** | | | | |
| Yes | 1.344 | 2.293*** | 1.309 | 1.796** |
|  | (0.269) | (0.488) | (0.275) | (0.402) |
| **Location (shop omit.)** | | | | |
| Office/flat/building/rooms | 0.583 | 0.433*** | 0.331*** | 0.447** |
|  | (0.186) | (0.105) | (0.110) | (0.116) |

|  | Model 2 | | Model 3 | |
|---|---|---|---|---|
|  | **2012** | **2018** | **2012** | **2018** |
| Workshop/factory | 0.911 | 0.858 | 1.043 | 1.087 |
|  | (0.279) | (0.247) | (0.328) | (0.331) |
| Kiosk/hut | 0.015*** | 0.757 | 0.012*** | 0.964 |
|  | (0.016) | (0.380) | (0.013) | (0.545) |
| **Sex (male omit.)** |  |  |  |  |
| Female |  |  | 1.141 | 0.594 |
|  |  |  | (0.301) | (0.159) |
| **Education (illit. omit.)** |  |  |  |  |
| Less than sec. |  |  | 1.180 | 1.672 |
|  |  |  | (0.306) | (0.492) |
| Secondary |  |  | 1.730* | 2.201** |
|  |  |  | (0.454) | (0.617) |
| Higher education |  |  | 3.965*** | 3.521*** |
|  |  |  | (1.176) | (1.169) |
| **Occupation (prof./tech. omit.)** |  |  |  |  |
| Other white collar |  |  | 0.660 | 0.864 |
|  |  |  | (0.157) | (0.234) |
| Blue collar craft |  |  | 0.967 | 0.441* |
|  |  |  | (0.279) | (0.148) |
| Blue collar non-craft |  |  | 0.868 | 0.730 |
|  |  |  | (0.305) | (0.264) |
| **Age group (<30 omit.)** |  |  |  |  |
| 30-49 |  |  | 1.892** | 1.105 |
|  |  |  | (0.429) | (0.295) |
| 50+ |  |  | 4.112*** | 1.385 |
|  |  |  | (1.135) | (0.426) |

|  | Model 2 | | Model 3 | |
| --- | --- | --- | --- | --- |
|  | **2012** | **2018** | **2012** | **2018** |
| **N** | 1222 | 1149 | 1221 | 1107 |
| **Pseudo R-squared** | 0.222 | 0.168 | 0.262 | 0.209 |

Source: Authors' calculations based on ELMPS 2012 and ELMPS 2018

Notes: *p<0.05; **p<0.01; ***p<0.001.

### 4.3.2. Are informal firms like formal firms?

Using the preceding models, we now turn to considering whether the predicted probabilities of being formal overlap between formal and informal firms. We take the extent of this overlap to be a metric of the potential to formalize. However, we recognize that this is entirely driven by the predictive power of the models, which is in turn driven by data availability. The pseudo-R-squareds of the models are another metric of this; consistent with our finding that location and industry were poor predictors of formality, Model 1, which includes these, firm size, and firm age, has pseudo-R-squareds that range from 0.082-0.119. Model 2a for EcC is 0.133, and the ELMPS model 2b 0.221 (2012) and 0.168 (2018). Owner characteristics in Model 3 raise this to 0.262 (2012) and 0.209 (2018).

Consistent with the low R-squareds from Model 1, Figure 10 shows high overlap between predicted probabilities, ranging from a 67% overlap in the EcC to an 76% overlap in ELMPS 2012. Since these are the easily observable characteristics, this bodes poorly for the potential of policymakers to target firms that are likely to formalize based on readily observed characteristics, whether for incentives or enforcement. Model 2a (Figure 11) for the EcC with firm characteristics shows 65% overlap. There is a distinct mode at 0.2 probability for informal firms and a mode at 0.4 for formal firms.

**Figure 10. Model 1 predicted probability of formality, by actual formality and data set**
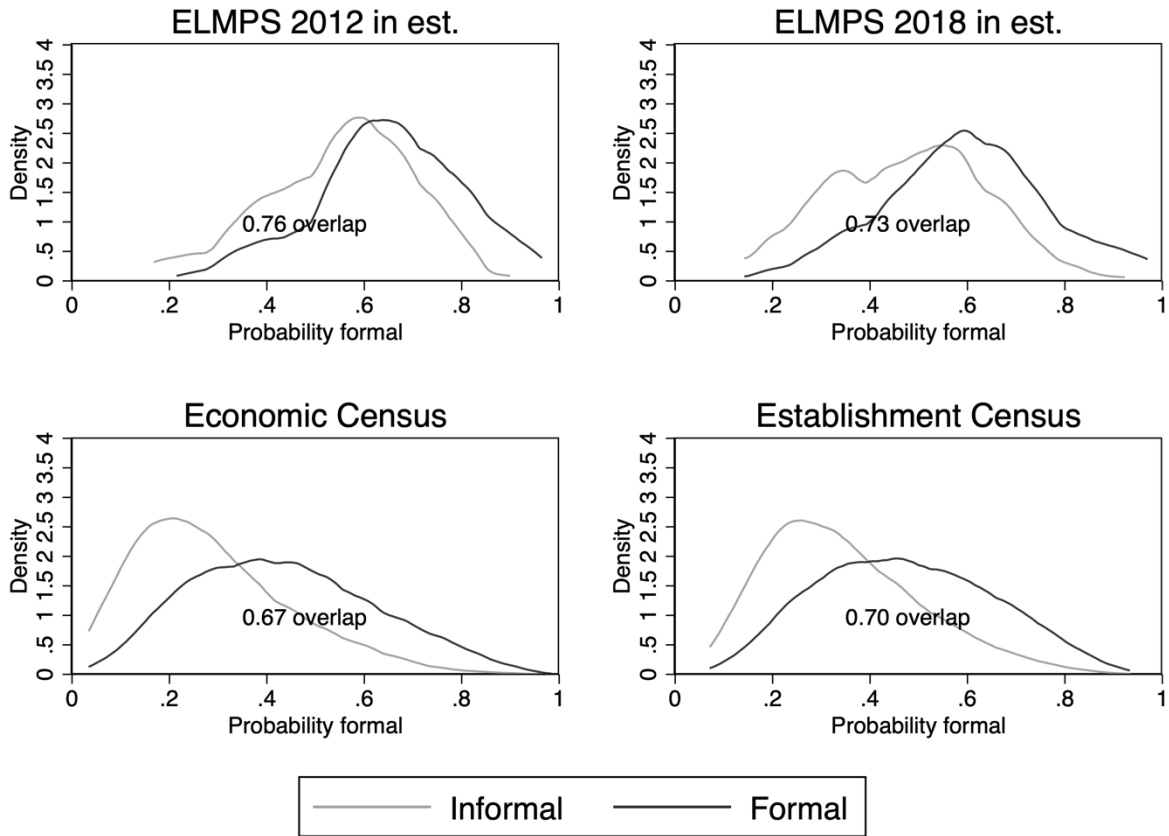


Source: Authors' calculations based on ELMPS 2012, ELMPS 2018, EcC 2012/13, EsC 2017

Notes: Kernel density functions with Epanechnikov kernel and bandwidth 0.05, predictions based on models in Table 2. Overlap denotes the proportion of the area beneath the curves that overlaps.

**Figure 11. Model 2a predicted probability of formality, by actual formality, EcC 2012/13**



## EC
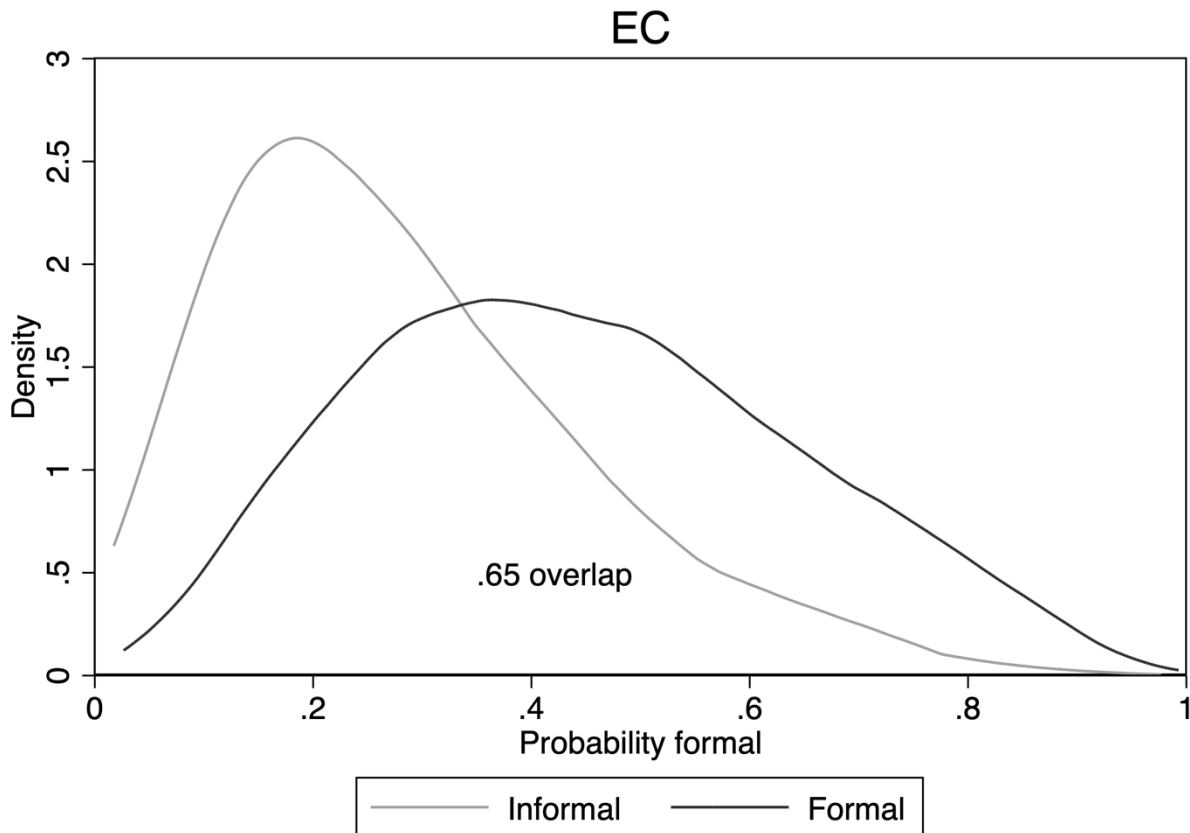
Source: Authors' calculations based on EcC 2012/13

Notes: Kernel density functions with Epanechnikov kernel and bandwidth 0.05, predictions based on models in Table 3. Overlap denotes the proportion of the area beneath the curves that overlaps.

ELMPS Model 2b (Figure 12) and Model 3 (Figure 13) for firms in establishments show diminishing overlap as more characteristics are added; from 64% overlap to 56% overlap in 2012 and 71% to 62% in 2018. In model 3 the modes for informal firm predicted probability remains around 0.2, but the formal firms shift to 0.6-0.8. Given that we still have only a limited number of covariates within the model, and that the overlap includes both the low-probability of formality formal firms and the high-probability of formality informal firms, this overlap is likely a substantial over-estimate of the share of firms that could easily formalize. Since the ELMPS also includes data on firms outside of establishments, we add these firms to Model 2b and Model 3 and re-estimate the results (see Table 12 in the appendix for the full regression). The overlap drops to 38% for model 3 in 2012 and 47% for model 3 in 2018. There is a clear mode among informal firms around an 0.05 probability of formality; these are the firms outside of establishments that are very unlikely to be formalizable.

**Figure 12. Model 2b predicted probability of formality, by actual formality, ELMPS 2012 and 2018**
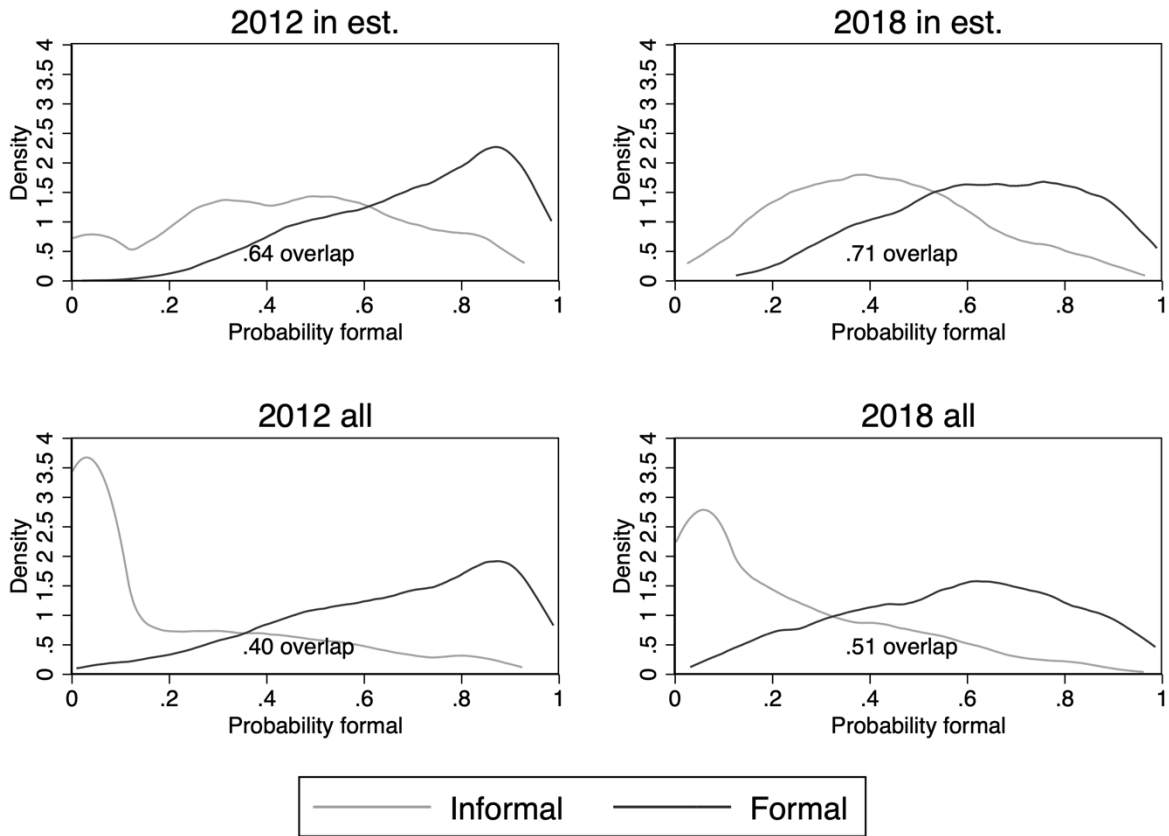


Source: Authors' calculations based on ELMPS 2012 and 2018

Notes: Kernel density functions with Epanechnikov kernel and bandwidth 0.05, predictions based on models in Table 4 and Table 12. Overlap denotes the proportion of the area beneath the curves that overlaps.

**Figure 13. Model 3 predicted probability of formality, by actual formality, ELMPS 2012 and 2018**
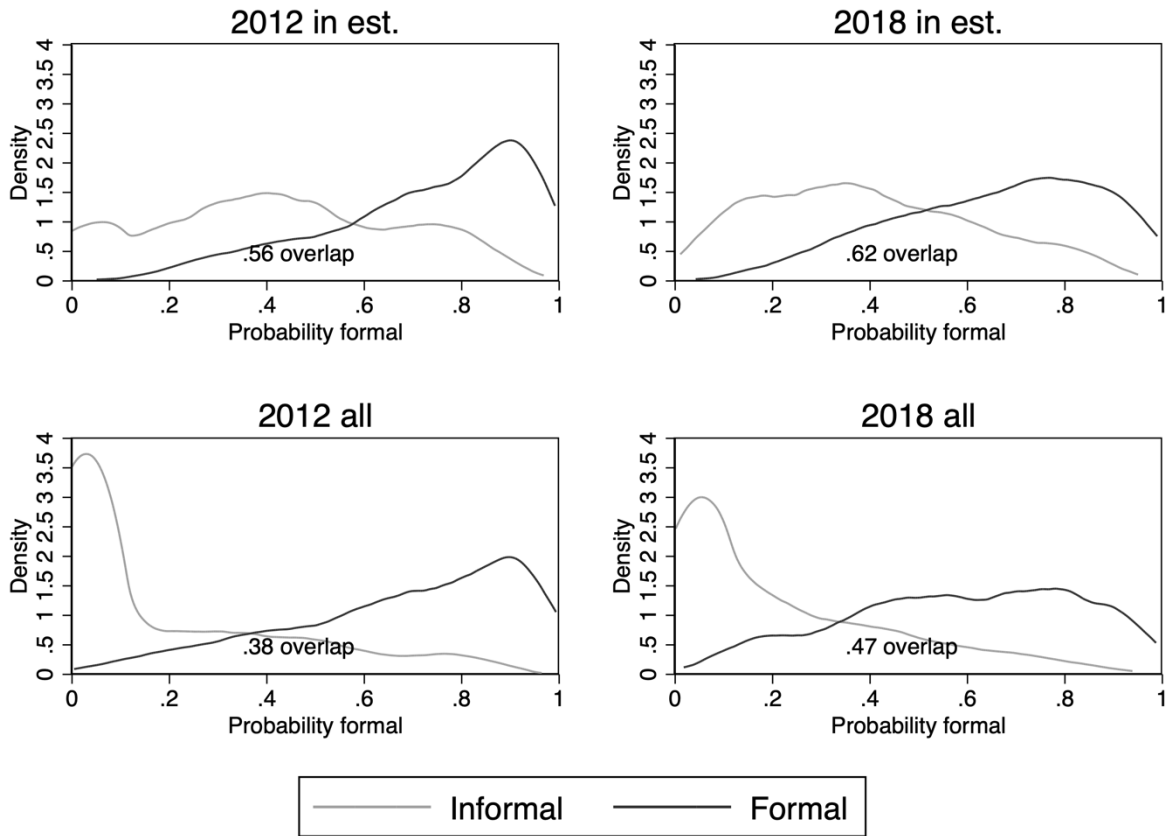


Source: Authors' calculations based on ELMPS 2012 and 2018

Notes: Kernel density functions with Epanechnikov kernel and bandwidth 0.05, predictions based on models in Table 4 and Table 12. Overlap denotes the proportion of the area beneath the curves that overlaps.

We further summarize these distributions in Table 5, where we present the quartiles of predicted probability of formality by model and data source, for informal firms in establishments. These reflect in part the higher probability in the ELMPS than the other data sources. We focus on the most detailed model in the discussion of our results. Probabilities above 50% might be considered a high chance of formalization. In the EsC Model 1, this is 17.5% of firms. In the EcC model 2, this is 10.9% of firms. In Model 3 of the ELMPS, this is 36.2% in 2012 and 32.2% in 2018. Recalling that only 22% of employment is in this segment of the economy, this suggests a relatively limited segment of the economy has the potential to formalize – and not all of that segment can necessarily even be formalized.

**Table 5. Quartiles of predicted probability of formality by model and data source, informal firms**

| Predicted probability of formality: | <25% | 25% to <50% | 50% to <75% | 75%+ |
|---|---|---|---|---|
| **Model 1** | | | | |
| **ELMPS 2012 in est.** | 5.8 | 26.2 | 62.4 | 5.5 |
| **ELMPS 2018 in est.** | 6.8 | 42.6 | 46.7 | 3.8 |
| **Economic Census** | 48.4 | 39.8 | 11.0 | 0.8 |
| **Establishment Census** | 30.9 | 51.7 | 16.1 | 1.4 |
| **Model 2** | | | | |
| **Economic Census** | 50.2 | 39.0 | 9.9 | 1.0 |
| **ELMPS 2012 in est.** | 19.7 | 37.0 | 30.6 | 12.7 |
| **ELMPS 2018 in est.** | 17.3 | 44.1 | 30.1 | 8.5 |
| **Model 3** | | | | |
| **ELMPS 2012 in est.** | 27.2 | 36.5 | 23.2 | 13.0 |
| **ELMPS 2018 in est.** | 30.1 | 37.8 | 23.3 | 8.9 |

Source: predictions based on models in Table 2, Table 3, and Table 4

## 4.4. Profiles: Grouping and Clustering Analysis

In what follows we attempt to characterize which informal firms operating inside establishments are least likely to be formalizable and which have a higher potential for formalization.[22] We do this first by grouping informal firms into different profiles (collections of characteristics that explain formality) based on Model 2 for the EcC data and Model 3 for the pooled ELMPS data. We also use cluster analysis to classify informal firms into 10 clusters based on the variables in EcC Model 2 and ELMPS Model 3. After examining how the mean characteristics of these clusters compare to those of all informal firms within establishments (shown in Table 13 and Table 14 in the appendix), we attempt to assign these clusters names that best describe the firms that fall within them. Finally, we identify, separately for each data set, which groups or clusters of informal firms are least likely and most likely to be formalizable based on the proportion that falls into each quartile of the probability of being formal.

As shown in Table 6, according to the ELMPS, the firms most represented in the bottom quartile of the probability of formality are one or two-person firms that are relatively young (<5 years of age) and that

---

[22] The very high informality rates for firms operating outside fixed establishments precludes this kind of analysis for these firms since few of them are likely to be appropriate candidates for formalization.

do not hire unrelated workers. The owners of these low-probability firms have below secondary or secondary education. These kinds of informal firms have more than a 50% chance of being in the lowest quartile of the probability of formality and almost no chance of being in the top quartile. These relatively young small firms with no hired workers constitute about 19% of all informal firms inside establishments in the ELMPS. Older one-person firms with relatively low revenue per worker come next in terms of representation in the lowest quartile, irrespective of owner's education. These constitute an additional 20% of all informal firms. Adding these two groups together suggests that approximately 40% of all informal firms operating in establishments have a low chance of formalization.

Results based on Model 2 using EcC data, shown in Table 7, confirm this pattern. Here, young one and two-person firms with no wage workers are also the most highly represented category in the lowest quartile of the probability of formality, irrespective of the amount of capital they have and irrespective of their value added per worker. These young small firms represent 34% of informal firms in the EcC. It appears that being young and not hiring wage workers are the primary predictors here, since larger firms with these characteristics are also highly represented in the bottom quartile although they only constitute another 1% of informal firms.

The cluster analysis highlights additional variables that characterize the informal firms that have the lowest probabilities of formality. As shown in Figure 14, the analysis using ELMPS data highlights low probability of formalization clusters than can be characterized as "women-owned retail firms" (8% of informal firms in ELMPS, 62% of which are in the lowest quartile of predicted formality) and "youth-owned young firms" (10% of informal firms, 42% of which are in the lowest quartile of predicted formality). These are followed by a cluster we term "smaller professional firms" (4% of informal firms, 39% of which are in the lowest quartile of predicted formality).[23] All three of these clusters skew toward young firms, with the proportion of zero-year-old firms being twice the overall average in the first and third clusters and the proportion of one and two year-old firms being 2.5 times higher than average in the second cluster.

The analysis using EcC data, shown in Figure 15, highlights clusters that we term "smaller professional firms" (5% of informal firms in EcC), "smaller service firms" (7%), "smaller manufacturing and food service firms" (6%) and "one-person retail firms" (27%). From 65% to 95% of these firms are in the lowest quartile of predicted formality. Firm age does not appear to be as much of a distinguishing feature of a low probability of formality in the EcC cluster results as it was in the grouping analysis and the ELMPS cluster results.

We now move to the other end of spectrum to characterize the profiles of informal firms that have a relatively high probability of formality and can presumably be targeted in formalization efforts. According to the ELMPS grouping analysis, these include older firms with above median revenue per worker, irrespective of size or owner education (21% of informal firms in ELMPS), older firms with more educated

---

[23] We assigned names to the clusters after examining the predominant characteristics that define them. These characteristics are summarized in Table 13 and Table 14 in the Appendix for the 10 clusters we identify in the data from each data set.

owners but below median revenue per worker, irrespective of size (6%), and older firms with three or more workers, irrespective of revenue per worker or owner's education (5%).

The EcC results for informal firms with high predicted probability of formality also highlight older 3-24 person firms that hire wage workers, irrespective of value added per worker (10% of informal firms), and those 3-24 person firms that do not hire wage workers but have high value added per worker (0.3%) as having the highest probability of formality. It further highlights older, two-person firms with value added per worker in the top quartile (3%). Their counterparts in the second and third quartile of value added per worker come next in priority (5%).

As before, cluster analysis provides a different lens to characterize the informal firms that are most amenable to formalization. According to the analysis using ELMPS data, shown in Figure 14, these include two clusters we describe as "larger professional" (8% of informal firms in ELMPS) and "larger non-professional" firms" (7% of informal firms). The latter cluster in particular is also disproportionately made up of firms with collective forms of ownership and relatively high capital. Both of these clusters disproportionately hire workers from outside the household. The first hires mostly non-relatives and the second hires relatives. These are followed by clusters we describe as "shops with older owners" (14%), and "workshops" (11%). These two clusters tend to skew older in terms of firm age and the latter is more likely to hire wage workers than average. The four clusters that have the lowest share of the lowest quartile of predicted probability of formality (5%-25%) still have only 14-26% of their firms in the top quartile of predicted probability.

Cluster analysis using the EcC data, shown in Figure 15, also highlights the larger professional (2% of informal firms in the EcC) and non-professional firm clusters (14%) as well as a clusters we call "larger retail" (8%) and "non-retail with 3+ workers" (10%). Like before, firms in these clusters tend to skew older than average and to be more likely to hire wage workers. Although these clusters have the lowest shares of the lowest quartile of predicted probability of formality (8-31%) between less than 1% and 13% of firms have predicted probabilities of formality in the top quartile, although 13-38% are in quartile three.

**Table 6. Percentage of firms in each quartile of predicted probability of formality by group, ordered by quartile one, pooled ELMPS 2012 and 2018**

| Profile Group | Firm Size | Firm Age | Capital | Non-rel. out of HH hired | Quartile of Ln(Revenue per Worker) | Owner's Education | N (Obs.) | % of Informal Firms in Group | Q1 | Q2 | Q3 | Q4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | Less than 5 years | All capital (Aggregate) | No | 1 & 2 (Aggregate) | Below secondary | 72 | 6.7 | 80.9 | 13.9 | 3.6 | 1.7 |
| 2 | 1 | Less than 5 years | <5K | No | 1 | Secondary | 28 | 2.6 | 67.6 | 32.4 | 0.0 | 0.0 |
| 3 | 2 | Less than 5 years | All capital (Aggregate) | No | 1 & 2 (Aggregate) | Below secondary | 20 | 2.6 | 62.6 | 26.5 | 10.8 | 0.0 |
| 4 | 1 | Less than 5 years | All capital (Aggregate) | No | 3 & 4 (Aggregate) | Below secondary | 34 | 3.4 | 57.4 | 36.2 | 6.4 | 0.0 |
| 5 | 1 | Less than 5 years | <5K | No | 2 | Secondary | 16 | 1.4 | 51.2 | 31.6 | 17.3 | 0.0 |
| 6 | 2 | Less than 5 years | All capital (Aggregate) | No | 1 & 2 (Aggregate) | Secondary+ (Aggregate) | 25 | 2.6 | 50.7 | 40.3 | 9.0 | 0.0 |
| 7 | 1 | 5 years or above | <5K | No | 2 | Below secondary | 53 | 5.5 | 47.4 | 39.8 | 8.7 | 4.2 |
| 8 | 1 | Less than 5 years | <5K | No | 1 & 2 (Aggregate) | Above secondary | 23 | 2.2 | 47.2 | 41.5 | 11.3 | 0.0 |
| 9 | 1 | 5 years or above | <5K | No | 1 | Below secondary | 60 | 6.0 | 44.9 | 38.8 | 16.3 | 0.0 |
| 10 | 1 | 5 years or above | All capital (Aggregate) | Yes | 1 & 2 (Aggregate) | Below secondary | 10 | 1.0 | 43.7 | 32.4 | 18.1 | 5.7 |
| 11 | 1 | 5 years or above | <5K | No | 1 | Secondary | 30 | 2.9 | 42.2 | 42.1 | 12.3 | 3.3 |
| 12 | 1 | 5 years or above | <5K | No | 3 | Below secondary | 22 | 1.9 | 36.6 | 28.3 | 35.1 | 0.0 |

| Profile Group | Firm Size | Firm Age | Capital | Non-rel. out of HH hired | Quartile of Ln(Revenue per Worker) | Owner's Education | N (Obs.) | % of Informal Firms in Group | Q1 | Q2 | Q3 | Q4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 13 | 1 | 5 years or above | 5K+ | No | 1 & 2 (Aggregate) | Below secondary | 11 | 1.3 | 30.1 | 30.4 | 22.9 | 16.6 |
| 14 | 1 | 5 years or above | <5K | No | 1 | Above secondary | 14 | 1.9 | 26.9 | 29.7 | 43.4 | 0.0 |
| 15 | 1 | Less than 5 years | All capital (Aggregate) | Yes | 3 & 4 (Aggregate) | All education (Aggregate) | 17 | 1.7 | 25.6 | 31.9 | 39.3 | 3.2 |
| 16 | 3-24 | Less than 5 years | All capital (Aggregate) | All (Aggregate) | All quartile (Aggregate) | All education (Aggregate) | 26 | 3.1 | 25.4 | 42.9 | 27.8 | 3.8 |
| 17 | 2 | Less than 5 years | All capital (Aggregate) | No | 3 & 4 (Aggregate) | All education (Aggregate) | 13 | 1.4 | 25.1 | 37.1 | 22.4 | 15.4 |
| 18 | 3-24 | 5 years or above | All capital (Aggregate) | No | All quartile (Aggregate) | All education (Aggregate) | 22 | 2.2 | 24.4 | 48.9 | 9.1 | 17.6 |
| 19 | 2 | Less than 5 years | All capital (Aggregate) | Yes | All quartile (Aggregate) | All education (Aggregate) | 11 | 1.3 | 23.9 | 49.8 | 22.9 | 3.4 |
| 20 | 1 | Less than 5 years | All capital (Aggregate) | No | 1 & 2 (Aggregate) | Secondary+ (Aggregate) | 15 | 1.4 | 23.6 | 38.4 | 29.3 | 8.7 |
| 21 | 1 | 5 years or above | <5K | No | 2 | Above secondary | 12 | 1.4 | 22.7 | 60.2 | 7.8 | 9.2 |
| 22 | 1 | Less than 5 years | All capital (Aggregate) | No | 3 & 4 (Aggregate) | Secondary+ (Aggregate) | 53 | 5.4 | 19.4 | 52.1 | 21.4 | 7.1 |
| 23 | 2 | 5 years or above | All capital (Aggregate) | No | 1 & 2 (Aggregate) | Secondary+ (Aggregate) | 12 | 1.0 | 19.0 | 20.8 | 60.1 | 0.0 |
| 24 | 1 | 5 years or above | 5K+ | No | 3 & 4 (Aggregate) | Below secondary | 27 | 2.6 | 18.5 | 33.8 | 27.2 | 20.4 |
| 25 | 2 | 5 years or above | All capital (Aggregate) | No | 1 & 2 (Aggregate) | Below secondary | 29 | 3.1 | 17.9 | 49.2 | 25.0 | 7.9 |
| 26 | 1 | 5 years or above | <5K | No | 3 | Secondary | 19 | 2.0 | 13.7 | 68.5 | 17.8 | 0.0 |

| Profile Group | Firm Size | Firm Age | Capital | Non-rel. out of HH hired | Quartile of Ln(Revenue per Worker) | Owner's Education | N (Obs.) | % of Informal Firms in Group | % of Firms in Each Quartile of Predicted Probability | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | Q1 | Q2 | Q3 | Q4 |
| 27 | 2 | 5 years or above | All capital (Aggregate) | Yes | 1 & 2 (Aggregate) | Secondary+ (Aggregate) | 14 | 1.5 | 13.2 | 24.4 | 24.5 | 37.9 |
| 28 | 1 | Less than 5 years | All capital (Aggregate) | Yes | 1 & 2 (Aggregate) | All education (Aggregate) | 13 | 1.3 | 11.6 | 48.2 | 32.5 | 7.7 |
| 29 | 3-24 | 5 years or above | All capital (Aggregate) | Yes | 1 & 2 (Aggregate) | All education (Aggregate) | 19 | 2.4 | 8.4 | 35.7 | 10.1 | 45.7 |
| 30 | 1 | 5 years or above | <5K | No | 4 | Below secondary | 29 | 2.4 | 7.2 | 56.5 | 35.0 | 1.3 |
| 31 | 2 | 5 years or above | All capital (Aggregate) | Yes | 1 & 2 (Aggregate) | Below secondary | 18 | 2.1 | 6.7 | 41.7 | 26.0 | 25.6 |
| 32 | 2 | 5 years or above | All capital (Aggregate) | No | 3 & 4 (Aggregate) | All education (Aggregate) | 18 | 2.4 | 6.6 | 34.3 | 38.4 | 20.8 |
| 33 | 1 | 5 years or above | All capital (Aggregate) | Yes | 3 & 4 (Aggregate) | Secondary+ (Aggregate) | 20 | 1.9 | 4.9 | 20.3 | 38.1 | 36.7 |
| 34 | 1 | 5 years or above | <5K | No | 2 | Secondary | 33 | 3.3 | 3.4 | 69.4 | 27.2 | 0.0 |
| 35 | 1 | 5 years or above | <5K | No | 3 & 4 (Aggregate) | Above secondary | 18 | 2.0 | 2.3 | 19.6 | 37.1 | 40.9 |
| 36 | 1 | 5 years or above | All capital (Aggregate) | Yes | 3 & 4 (Aggregate) | Below secondary | 10 | 1.1 | 0.0 | 59.8 | 18.9 | 21.3 |
| 37 | 1 | 5 years or above | 5K+ | No | 1 & 2 (Aggregate) | Secondary+ (Aggregate) | 21 | 1.7 | 0.0 | 22.3 | 60.0 | 17.6 |
| 38 | 1 | 5 years or above | <5K | No | 4 | Secondary | 18 | 1.9 | 0.0 | 19.0 | 51.5 | 29.5 |
| 39 | 3-24 | 5 years or above | All capital (Aggregate) | Yes | 3 & 4 (Aggregate) | All education (Aggregate) | 20 | 1.7 | 0.0 | 18.4 | 34.8 | 46.8 |
| 40 | 1 | 5 years or above | 5K+ | No | 3 & 4 (Aggregate) | Secondary | 23 | 2.0 | 0.0 | 17.4 | 58.4 | 24.2 |
| 41 | 1 | 5 years or above | 5K+ | No | 3 & 4 (Aggregate) | Above secondary | 15 | 1.4 | 0.0 | 16.1 | 25.9 | 58.0 |
| 42 | 2 | 5 years or above | All capital (Aggregate) | Yes | 3 & 4 (Aggregate) | All education (Aggregate) | 13 | 1.1 | 0.0 | 7.7 | 30.9 | 61.4 |
| 43 | 1 | 5 years or above | All capital (Aggregate) | Yes | 1 & 2 (Aggregate) | Secondary+ (Aggregate) | 13 | 1.1 | 0.0 | 7.7 | 55.7 | 36.6 |

Source: Authors' calculations based on predicted probability from model 3 for ELMPS 2012 and ELMPS 2018, see Table 4

Notes: Firm age category "don't know" was combined with 5 years or above.

**Table 7. Percentage of firms in each quartile of predicted probability of formality by group, ordered by quartile one, EcC 2012/13**

| Profile Group | Firm Size | Firm Age | Capital | Wage Worker | Quartile of Ln(Labor Value Added) | N (Obs.) | % of Informal Firms in Group | % of Firms in Each Quartile of Predicted Probability | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Q1 | Q2 | Q3 | Q4 |
| 1 | 1 | Less than 5 years | 5K+ | No | 1 | 677 | 3.7 | 96.3 | 3.7 | 0.0 | 0.0 |
| 2 | 1 | Less than 5 years | <5K | No | 1 | 963 | 5.1 | 96.2 | 3.8 | 0.0 | 0.0 |
| 3 | 1 | Less than 5 years | 5K+ | No | 2 | 721 | 3.9 | 94.9 | 5.1 | 0.0 | 0.0 |
| 4 | 1 | Less than 5 years | <5K | No | 2 | 722 | 3.3 | 94.7 | 5.3 | 0.0 | 0.0 |
| 5 | 1 | Less than 5 years | 5K+ | No | 3 | 702 | 3.3 | 93.8 | 6.2 | 0.0 | 0.0 |
| 6 | 1 | Less than 5 years | <5K | No | 3 | 523 | 2.3 | 88.9 | 11.1 | 0.0 | 0.0 |
| 7 | 1 | Less than 5 years | <5K | No | 4 | 231 | 1.0 | 84.1 | 15.9 | 0.0 | 0.0 |
| 8 | 2 | Less than 5 years | 5K+ | No | 1 | 544 | 3.5 | 82.1 | 17.9 | 0.0 | 0.0 |
| 9 | 1 | Less than 5 years | 5K+ | No | 4 | 548 | 2.7 | 81.2 | 18.8 | 0.0 | 0.0 |
| 10 | 2 | Less than 5 years | <5K | No | 1 | 442 | 2.6 | 74.7 | 25.3 | 0.0 | 0.0 |
| 11 | 2 | Less than 5 years | <5K | No | 2 | 135 | 0.8 | 73.6 | 26.4 | 0.0 | 0.0 |
| 12 | 3-24 | Less than 5 years | <5K | No | All quartile (Aggregate) | 82 | 0.5 | 73.5 | 26.5 | 0.0 | 0.0 |
| 13 | 3-24 | Less than 5 years | 5K+ | No | 3 | 29 | 0.1 | 70.4 | 29.6 | 0.0 | 0.0 |
| 14 | 2 | Less than 5 years | 5K+ | No | 2 | 300 | 1.9 | 68.5 | 31.5 | 0.0 | 0.0 |
| 15 | 3-24 | Less than 5 years | <5K | Yes | 1 | 180 | 0.5 | 68.4 | 30.9 | 0.6 | 0.0 |
| 16 | 1 | 5 years or above | 5K+ | No | 1 | 335 | 1.9 | 67.8 | 31.6 | 0.6 | 0.0 |
| 17 | 2 | Less than 5 years | <5K | No | 4 | 15 | 0.1 | 64.1 | 35.9 | 0.0 | 0.0 |
| 18 | 3-24 | Less than 5 years | 5K+ | No | 1 | 126 | 0.6 | 63.5 | 36.5 | 0.0 | 0.0 |
| 19 | 3-24 | Less than 5 years | 5K+ | No | 2 | 40 | 0.2 | 62.8 | 36.9 | 0.3 | 0.0 |
| 20 | 2 | Less than 5 years | 5K+ | Yes | 1 | 451 | 1.6 | 61.7 | 38.1 | 0.2 | 0.0 |
| 21 | 1 | 5 years or above | <5K | No | 1 | 666 | 2.7 | 61.0 | 37.6 | 1.4 | 0.0 |
| 22 | 2 | Less than 5 years | 5K+ | No | 3 | 146 | 0.7 | 58.4 | 41.6 | 0.0 | 0.0 |
| 23 | 2 | Less than 5 years | <5K | Yes | 1 | 244 | 0.9 | 57.4 | 42.5 | 0.1 | 0.0 |
| 24 | 3-24 | Less than 5 years | <5K | Yes | 3 | 64 | 0.1 | 57.3 | 29.5 | 13.2 | 0.0 |

| Profile Group | Firm Size | Firm Age | Capital | Wage Worker | Quartile of Ln(Labor Value Added) | N (Obs.) | % of Informal Firms in Group | Q1 | Q2 | Q3 | Q4 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 25 | 2 | Less than 5 years | <5K | Yes | 2 | 198 | 0.5 | 56.5 | 43.5 | 0.0 | 0.0 |
| 26 | 2 | Less than 5 years | 5K+ | No | 4 | 81 | 0.4 | 54.9 | 44.8 | 0.3 | 0.0 |
| 27 | 2 | Less than 5 years | <5K | No | 3 | 38 | 0.2 | 54.6 | 45.4 | 0.0 | 0.0 |
| 28 | 3-24 | Less than 5 years | <5K | Yes | 2 | 107 | 0.3 | 53.5 | 45.4 | 1.1 | 0.0 |
| 29 | 1 | 5 years or above | 5K+ | No | 2 | 436 | 1.9 | 52.2 | 45.5 | 2.3 | 0.0 |
| 30 | 3-24 | Less than 5 years | 5K+ | Yes | 1 | 733 | 2.0 | 49.9 | 44.6 | 5.4 | 0.1 |
| 31 | 2 | Less than 5 years | 5K+ | Yes | 3 | 539 | 1.8 | 49.0 | 50.1 | 0.9 | 0.0 |
| 32 | 2 | Less than 5 years | 5K+ | Yes | 2 | 609 | 2.1 | 46.3 | 52.9 | 0.8 | 0.0 |
| 33 | 1 | 5 years or above | <5K | No | 2 | 655 | 2.2 | 45.6 | 52.6 | 1.8 | 0.0 |
| 34 | 2 | Less than 5 years | 5K+ | Yes | 4 | 399 | 1.1 | 45.3 | 48.9 | 5.8 | 0.0 |
| 35 | 1 | 5 years or above | 5K+ | No | 3 | 559 | 1.9 | 43.9 | 56.1 | 0.0 | 0.0 |
| 36 | 2 | Less than 5 years | <5K | Yes | 4 | 53 | 0.1 | 41.6 | 58.4 | 0.0 | 0.0 |
| 37 | 1 | 5 years or above | <5K | No | 3 | 605 | 2.0 | 40.0 | 58.7 | 1.3 | 0.0 |
| 38 | 2 | Less than 5 years | <5K | Yes | 3 | 137 | 0.6 | 37.7 | 61.7 | 0.6 | 0.0 |
| 39 | 3-24 | Less than 5 years | 5K+ | Yes | 2 | 786 | 2.3 | 36.9 | 54.3 | 8.4 | 0.4 |
| 40 | 3-24 | Less than 5 years | 5K+ | Yes | 3 | 903 | 3.0 | 33.9 | 60.2 | 5.7 | 0.1 |
| 41 | 3-24 | 5 years or above | 5K+ | No | 1 | 209 | 0.5 | 33.7 | 54.3 | 11.9 | 0.0 |
| 42 | 3-24 | Less than 5 years | <5K | Yes | 4 | 31 | 0.1 | 29.9 | 68.2 | 1.7 | 0.2 |
| 43 | 1 | 5 years or above | <5K | No | 4 | 365 | 1.3 | 25.8 | 60.4 | 13.8 | 0.0 |
| 44 | 2 | 5 years or above | 5K+ | Yes | 1 | 335 | 1.1 | 21.8 | 59.7 | 18.4 | 0.1 |
| 45 | 2 | 5 years or above | <5K | Yes | 1 | 193 | 0.4 | 20.8 | 58.6 | 20.6 | 0.0 |
| 46 | 2 | 5 years or above | 5K+ | No | 1 | 302 | 2.8 | 19.2 | 77.8 | 3.0 | 0.0 |
| 47 | 3-24 | Less than 5 years | 5K+ | Yes | 4 | 725 | 2.1 | 18.0 | 69.4 | 12.2 | 0.4 |
| 48 | 1 | 5 years or above | 5K+ | No | 4 | 437 | 2.8 | 17.7 | 74.6 | 7.7 | 0.0 |
| 49 | 3-24 | Less than 5 years | 5K+ | No | 4 | 17 | 0.2 | 17.1 | 82.9 | 0.0 | 0.0 |
| 50 | 2 | 5 years or above | <5K | No | 1 | 373 | 1.9 | 16.7 | 72.0 | 11.3 | 0.0 |

| Profile Group | Firm Size | Firm Age | Capital | Wage Worker | Quartile of Ln(Labor Value Added) | N (Obs.) | % of Informal Firms in Group | Q1 | Q2 | Q3 | Q4 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | % of Firms in Each Quartile of Predicted Probability | | |
| 51 | 2 | 5 years or above | 5K+ | Yes | 2 | 619 | 1.6 | 14.3 | 62.6 | 23.1 | 0.0 |
| 52 | 3-24 | 5 years or above | <5K | No | 1 | 156 | 0.5 | 14.3 | 56.3 | 29.3 | 0.0 |
| 53 | 2 | 5 years or above | 5K+ | No | 2 | 214 | 0.9 | 14.1 | 67.9 | 18.0 | 0.0 |
| 54 | 2 | 5 years or above | 5K+ | Yes | 3 | 761 | 1.8 | 13.3 | 51.4 | 35.2 | 0.0 |
| 55 | 3-24 | 5 years or above | 5K+ | No | 2 | 64 | 0.2 | 13.0 | 70.7 | 16.3 | 0.0 |
| 56 | 3-24 | 5 years or above | <5K | No | 2 | 49 | 0.2 | 12.1 | 58.7 | 26.4 | 2.7 |
| 57 | 2 | 5 years or above | <5K | Yes | 2 | 359 | 0.9 | 11.8 | 68.6 | 19.4 | 0.2 |
| 58 | 2 | 5 years or above | 5K+ | No | 3 | 120 | 0.4 | 10.2 | 65.1 | 24.7 | 0.0 |
| 59 | 2 | 5 years or above | <5K | No | 3 | 83 | 0.3 | 8.9 | 55.1 | 36.0 | 0.0 |
| 60 | 2 | 5 years or above | <5K | No | 2 | 197 | 0.9 | 8.7 | 63.9 | 27.3 | 0.0 |
| 61 | 3-24 | 5 years or above | 5K+ | Yes | 1 | 885 | 1.3 | 7.1 | 45.9 | 38.7 | 8.3 |
| 62 | 2 | 5 years or above | 5K+ | Yes | 4 | 912 | 1.6 | 5.9 | 43.1 | 50.9 | 0.1 |
| 63 | 2 | 5 years or above | 5K+ | No | 4 | 75 | 0.4 | 5.9 | 61.9 | 24.9 | 7.3 |
| 64 | 2 | 5 years or above | <5K | Yes | 3 | 261 | 0.5 | 5.1 | 69.5 | 25.4 | 0.0 |
| 65 | 3-24 | 5 years or above | 5K+ | Yes | 2 | 915 | 1.9 | 4.9 | 58.1 | 31.9 | 5.1 |
| 66 | 3-24 | 5 years or above | <5K | Yes | 1 | 346 | 0.5 | 4.4 | 54.6 | 38.0 | 3.0 |
| 67 | 3-24 | 5 years or above | 5K+ | No | 3 | 34 | 0.1 | 3.6 | 76.8 | 19.6 | 0.0 |
| 68 | 3-24 | 5 years or above | 5K+ | Yes | 3 | 1101 | 2.3 | 3.3 | 52.1 | 36.8 | 7.8 |
| 69 | 2 | 5 years or above | <5K | No | 4 | 26 | 0.1 | 2.9 | 20.8 | 72.3 | 4.1 |
| 70 | 2 | 5 years or above | <5K | Yes | 4 | 200 | 0.2 | 2.8 | 46.6 | 49.3 | 1.4 |
| 71 | 3-24 | 5 years or above | <5K | Yes | 2 | 218 | 0.4 | 1.7 | 62.0 | 33.1 | 3.1 |
| 72 | 3-24 | 5 years or above | <5K | No | 3 & 4 (Aggregate) | 29 | 0.1 | 1.5 | 37.2 | 52.2 | 9.1 |
| 73 | 3-24 | 5 years or above | 5K+ | Yes | 4 | 1292 | 2.2 | 0.5 | 30.9 | 49.0 | 19.6 |
| 74 | 3-24 | 5 years or above | <5K | Yes | 3 | 178 | 0.4 | 0.5 | 34.0 | 60.6 | 5.0 |
| 75 | 3-24 | 5 years or above | 5K+ | No | 4 | 30 | 0.1 | 0.0 | 40.6 | 52.3 | 7.1 |
| 76 | 3-24 | 5 years or above | <5K | Yes | 4 | 115 | 0.9 | 0.0 | 5.9 | 92.4 | 1.6 |

Source: Authors' calculations based on predicted probability from model 2 using EcC 2012/13, see Table 3.

Notes: Firm age category "don't know" was combined with 5 years or above. Firms that reported firm size of one and that hire wage workers were recoded to firm size two.

**Figure 14. Percentage of informal firms in each quartile of predicted probability of formality by cluster, pooled ELMPS 2012 and 2018**
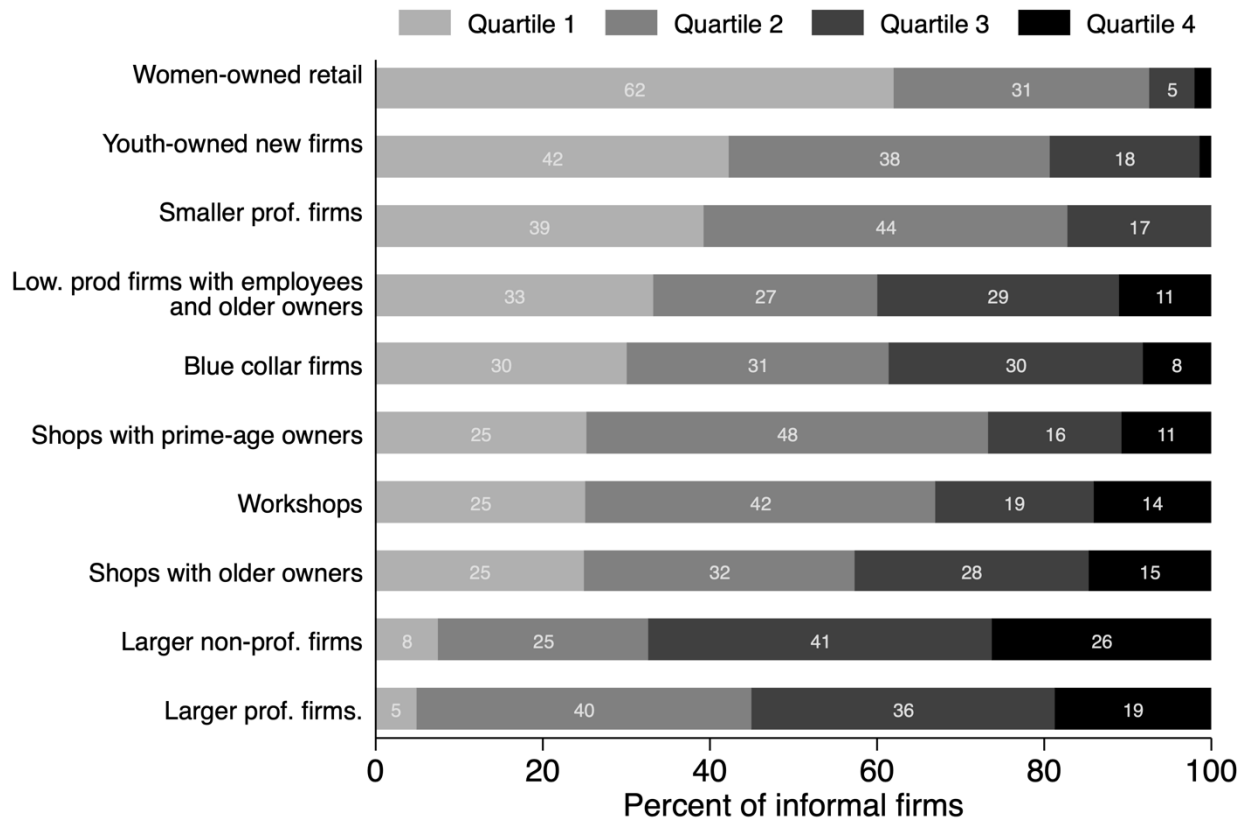


Source: Authors' calculations based on predicted probability from model 3 for ELMPS 2012 and ELMPS 2018, see Table 4

**Figure 15. Percentage of informal firms in each quartile of predicted probability of formality by cluster, EcC 2012/13**
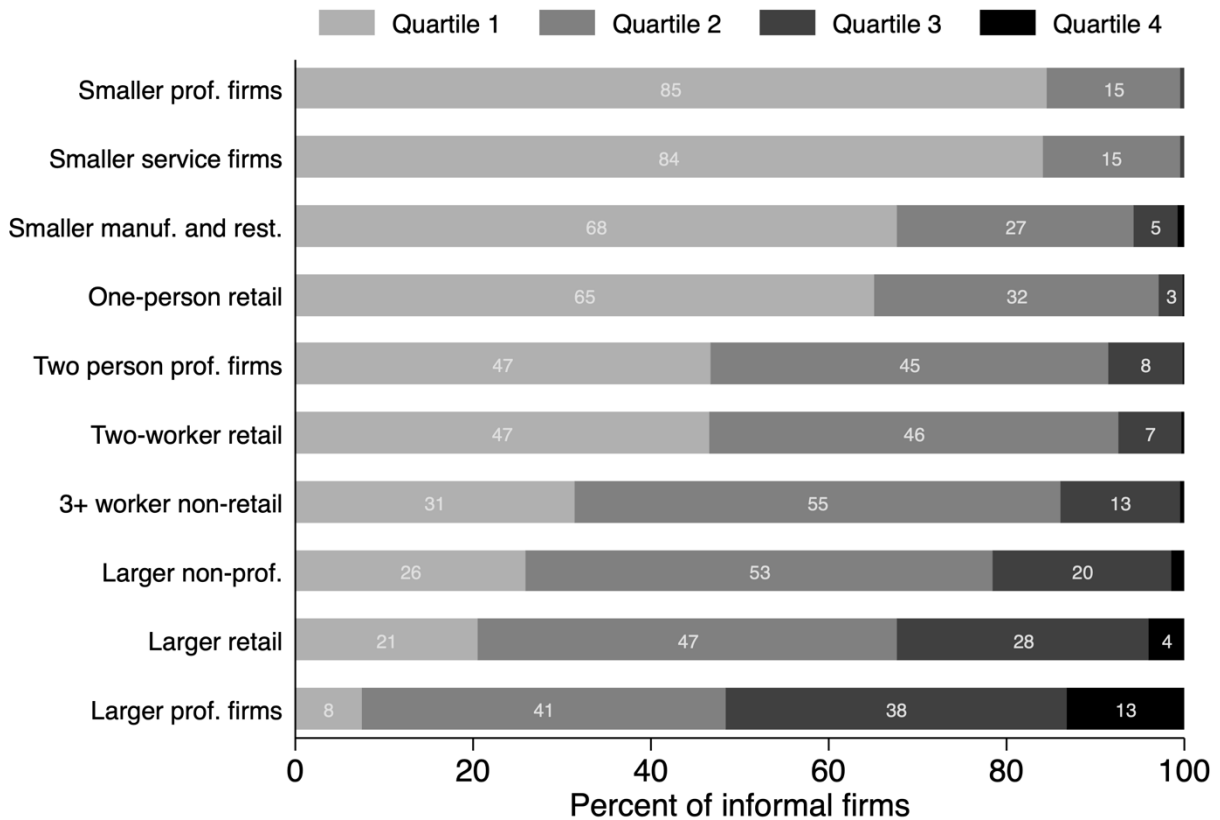


Source: Authors' calculations based on predicted probability from model 2 using EcC 2012/13, see Table 3

## 4.5. Firm Dynamics

In this section we turn to firm dynamics, using the ELMPS, which is the only data source that allows for the construction of a panel of firms. We analyze a variety of outcomes: exit of firms, entry of firms, formalization of existing firms, and whether firms grow. We relate these to firm formality and firm characteristics at the base of a pair of waves, except for new firms, which of course have no base characteristics. We present our results outcome by outcome, first descriptively and then (if possible) with multivariate models. Note that our measure of formality is now basic definition of formality (commercial registration), since the extended definition was not available in 1998 or 2006. We include firms both in and out of establishments in the dynamic analyses, especially since a firm could change locations.

### 4.5.1. Exit of firms

Informal firms are more likely to exit (close) than formal firms (Figure 16). Specifically, 74% of the firms that were informal in 2012 exited by 2018 compared to only 62% of formal firms from 2012. Exit rates have increased substantially over time, from 53% in 1998-2006 (which notably was a longer period) to 62% in 2006-2012 and 71% in 2012-2018. There has been a persistent gap with higher rates of exit among informal firms, with the gap being largest in 2006-2012 (69% of informal vs. 49% of formal firms exiting) and smallest in 2012-2018 (74% of informal and 62% of formal firms exiting).

**Figure 16. Percentage of firms exiting between waves by base wave formality**
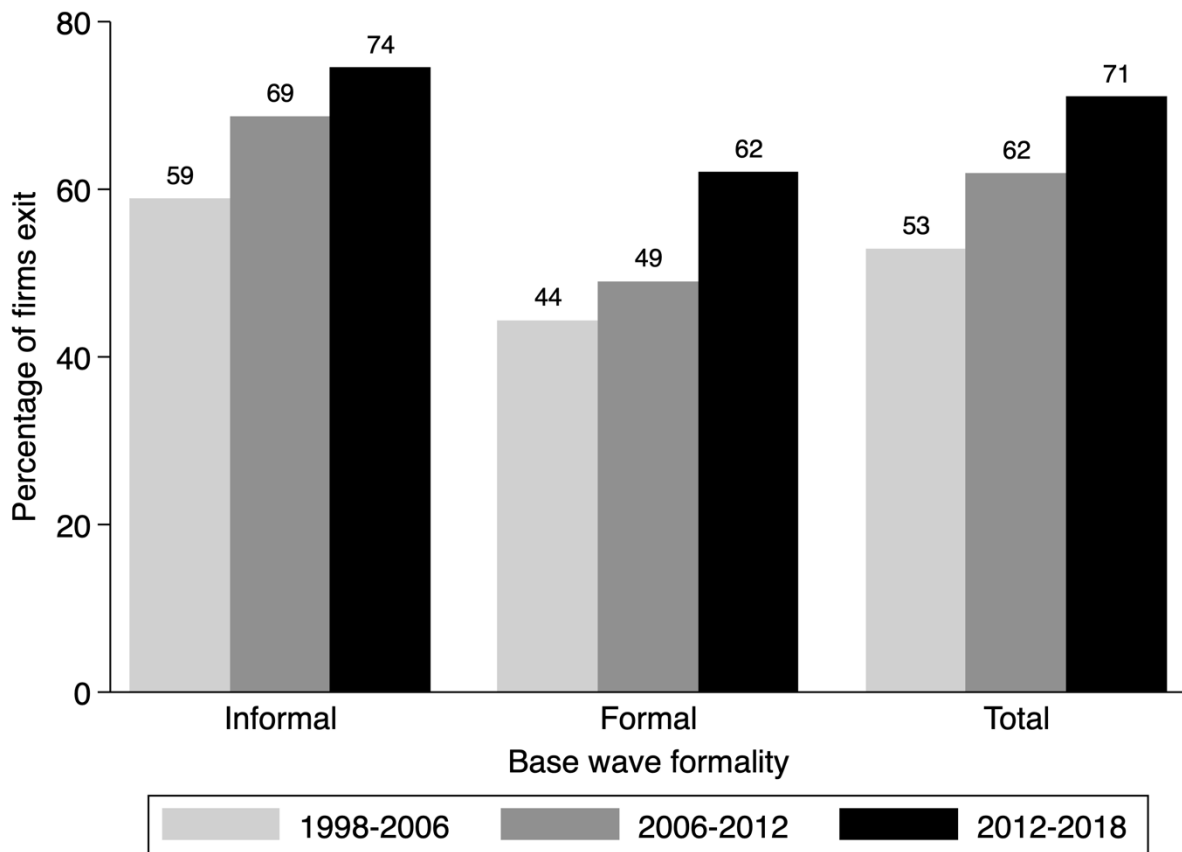


Source: Authors' calculations based on ELMPS 1998, ELMPS 2006, ELMPS 2012, ELMPS 2018

Table 8 presents a multivariate model of firm exit, by pair of waves (2012-2018 is presented without and with log revenue per worker). We primarily focus our discussion on the relationship between formality and exit, since other work has discussed firm dynamics more generally (Krafft, 2016). After accounting for other characteristics, firms that are formal are less likely to exit between waves, but only significantly so

in 2006-2012. There are significant differences over time in the probability of exit, but not significant differences in the relationship between formality and exit.

**Table 8. Odds ratios from logit model of firm exit, by pair of waves, ELMPS 1998-2018**

|  | 1998-2006 | 2006-2012 | 2012-2018 | 2012-2018 | Pool |
|---|---|---|---|---|---|
| **Formality (informal omit.)** | | | | | |
| Formal | 0.811 | 0.625* | 0.919 | 0.915 | 0.957 |
|  | (0.243) | (0.117) | (0.159) | (0.158) | (0.193) |
| **Wave (1998-2006 omit.)** | | | | | |
| 2006-2012 |  |  |  |  | 1.481* |
|  |  |  |  |  | (0.232) |
| 2012-2018 |  |  |  |  | 1.745*** |
|  |  |  |  |  | (0.286) |
| **Interactions between formality and wave** | | | | | |
| Formality*2006-2012 |  |  |  |  | 0.710 |
|  |  |  |  |  | (0.161) |
| Formality*2012-2018 |  |  |  |  | 0.873 |
|  |  |  |  |  | (0.196) |
| **Ent. size (one worker omit.)** | | | | | |
| 2 | 0.950 | 0.885 | 0.714 | 0.722 | 0.863 |
|  | (0.259) | (0.159) | (0.131) | (0.134) | (0.099) |
| 3-4 | 1.090 | 0.830 | 0.677 | 0.687 | 0.853 |
|  | (0.351) | (0.198) | (0.180) | (0.185) | (0.129) |
| 5-24 | 1.417 | 0.743 | 2.043 | 2.096 | 1.275 |
|  | (0.699) | (0.289) | (0.817) | (0.845) | (0.309) |
| **Industry (wholesale/retail omit.)** | | | | | |
| Manufacturing and related trades | 0.477 | 1.313 | 1.661* | 1.668* | 1.202 |
|  | (0.197) | (0.294) | (0.427) | (0.430) | (0.180) |
| Construction | 0.563 | 2.099* | 4.624*** | 4.624*** | 2.107*** |
|  | (0.353) | (0.656) | (1.492) | (1.490) | (0.436) |

| | 1998-2006 | 2006-2012 | 2012-2018 | 2012-2018 | Pool |
|---|---|---|---|---|---|
| Transportation and storage | 2.346 | 1.271 | 3.852* | 3.847* | 2.030 |
| | (1.913) | (0.646) | (2.288) | (2.291) | (0.803) |
| Accommodation and food service | 0.790 | 2.241** | 1.252 | 1.256 | 1.428 |
| | (0.358) | (0.693) | (0.394) | (0.395) | (0.280) |
| Various professional acts. | 1.559 | 0.562 | 2.158** | 2.158** | 1.233 |
| | (0.854) | (0.189) | (0.621) | (0.622) | (0.254) |
| Other service | 2.106 | 1.434 | 1.642 | 1.651 | 1.993*** |
| | (0.919) | (0.531) | (0.532) | (0.534) | (0.413) |
| **Region (Greater Cairo omit.)** | | | | | |
| Alexandria and Suez Canal | 0.999 | 1.113 | 1.035 | 1.035 | 1.018 |
| | (0.299) | (0.265) | (0.255) | (0.255) | (0.151) |
| Lower Egypt | 1.131 | 0.905 | 0.595** | 0.597* | 0.823 |
| | (0.291) | (0.157) | (0.119) | (0.120) | (0.092) |
| Upper Egypt | 0.815 | 0.988 | 0.653* | 0.657* | 0.816 |
| | (0.226) | (0.178) | (0.137) | (0.138) | (0.096) |
| **Partnership (sole proprietorship omit.)** | | | | | |
| Collective Ownership | 1.775 | 0.761 | 1.045 | 1.044 | 1.023 |
| | (0.549) | (0.186) | (0.252) | (0.252) | (0.150) |
| **Capital (none omit.)** | | | | | |
| less than LE 100 | 1.001 | 0.621 | 0.623 | 0.628 | 0.705 |
| | (0.377) | (0.172) | (0.205) | (0.207) | (0.129) |
| LE 100-499 | 0.341* | 0.545* | 0.788 | 0.791 | 0.562** |
| | (0.146) | (0.158) | (0.267) | (0.268) | (0.107) |
| LE 500-999 | 0.837 | 0.647 | 0.707 | 0.708 | 0.712 |
| | (0.332) | (0.189) | (0.220) | (0.221) | (0.134) |
| LE 1000-4999 | 0.518 | 0.490* | 0.880 | 0.875 | 0.645* |

|  | 1998-2006 | 2006-2012 | 2012-2018 | 2012-2018 | Pool |
|---|---|---|---|---|---|
|  | (0.246) | (0.144) | (0.293) | (0.291) | (0.125) |
| LE5000-9999 | 0.736 | 0.575 | 0.630 | 0.624 | 0.609* |
|  | (0.344) | (0.167) | (0.202) | (0.201) | (0.118) |
| LE10 000 or more | 0.654 | 0.780 | 0.660 | 0.655 | 0.792 |
|  | (0.433) | (0.296) | (0.314) | (0.311) | (0.216) |
| **Ln net revenue per worker** |  |  |  | 1.022 |  |
|  |  |  |  | (0.060) |  |
| **Firm start year (1980-1989 omit.)** |  |  |  |  |  |
| Don't know | 1.143 | 0.530 | 1.088 | 1.090 | 1.114 |
|  | (0.812) | (0.436) | (0.451) | (0.451) | (0.350) |
| Before 1950 | 0.512 | 0.809 | 0.803 | 0.797 | 0.677 |
|  | (0.351) | (0.531) | (0.959) | (0.955) | (0.284) |
| 1950-1959 | 1.424 | 0.820 | 0.466 | 0.462 | 0.957 |
|  | (0.721) | (0.429) | (0.407) | (0.407) | (0.317) |
| 1960-1969 | 1.208 | 1.050 | 0.558 | 0.558 | 0.981 |
|  | (0.535) | (0.436) | (0.400) | (0.398) | (0.272) |
| 1970-1979 | 1.364 | 0.864 | 1.076 | 1.080 | 1.041 |
|  | (0.447) | (0.221) | (0.397) | (0.398) | (0.178) |
| 1990-1999 | 1.304 | 1.382 | 1.228 | 1.233 | 1.326* |
|  | (0.310) | (0.241) | (0.284) | (0.285) | (0.158) |
| 2000-2009 |  | 1.789** | 1.352 | 1.359 | 1.621*** |
|  |  | (0.357) | (0.309) | (0.311) | (0.222) |
| 2010-2018 |  |  | 2.124** | 2.150** | 2.321*** |
|  |  |  | (0.605) | (0.612) | (0.521) |
| **People outside the HH employed (no omit.)** |  |  |  |  |  |
| Yes | 0.509** | 1.022 | 0.771 | 0.767 | 0.799* |
|  | (0.128) | (0.195) | (0.133) | (0.133) | (0.089) |

|  | 1998-2006 | 2006-2012 | 2012-2018 | 2012-2018 | Pool |
|---|---|---|---|---|---|
| **Location (shop omit.)** | | | | | |
| Own home | 1.738 | 1.418 | 1.084 | 1.090 | 1.378* |
| | (0.665) | (0.362) | (0.258) | (0.262) | (0.221) |
| Office/flat/building/rooms | 0.685 | 1.487 | 0.589 | 0.586 | 0.901 |
| | (0.352) | (0.499) | (0.198) | (0.197) | (0.188) |
| Workshop/factory | 3.551** | 1.456 | 1.323 | 1.329 | 1.651** |
| | (1.549) | (0.438) | (0.384) | (0.385) | (0.312) |
| Mobile worker | 2.100 | 1.052 | 1.399 | 1.405 | 1.341 |
| | (1.177) | (0.290) | (0.358) | (0.361) | (0.242) |
| Kiosk/hut | 1.741 | 0.812 | 1.693 | 1.708 | 1.376 |
| | (0.928) | (0.457) | (0.770) | (0.779) | (0.387) |
| Transport based | 0.512 | 0.742 | 0.323 | 0.324 | 0.574 |
| | (0.534) | (0.421) | (0.191) | (0.191) | (0.251) |
| Street/field/farm | 0.943 | 0.958 | 1.194 | 1.193 | 1.085 |
| | (0.451) | (0.278) | (0.341) | (0.342) | (0.200) |
| **Sex (male omit.)** | | | | | |
| Female | 1.192 | 1.452 | 1.598* | 1.613* | 1.449** |
| | (0.385) | (0.326) | (0.349) | (0.353) | (0.204) |
| **Education (illit. omit.)** | | | | | |
| Less than sec. | 1.443 | 1.285 | 0.914 | 0.914 | 1.183 |
| | (0.381) | (0.223) | (0.173) | (0.173) | (0.133) |
| Secondary | 3.235*** | 1.521* | 0.960 | 0.956 | 1.460** |
| | (1.114) | (0.290) | (0.185) | (0.185) | (0.184) |
| Higher education | 2.398* | 1.272 | 1.336 | 1.333 | 1.546** |
| | (0.890) | (0.312) | (0.295) | (0.295) | (0.229) |
| **Occupation (prof./tech. omit.)** | | | | | |
| Other white collar | 1.709 | 0.980 | 1.263 | 1.269 | 1.194 |

|  | 1998-2006 | 2006-2012 | 2012-2018 | 2012-2018 | Pool |
|---|---|---|---|---|---|
|  | (0.510) | (0.191) | (0.258) | (0.258) | (0.152) |
| Blue collar craft | 2.229* | 1.000 | 0.960 | 0.955 | 1.184 |
|  | (0.828) | (0.208) | (0.212) | (0.212) | (0.162) |
| Blue collar non-craft | 1.029 | 1.317 | 1.667* | 1.669* | 1.377* |
|  | (0.397) | (0.331) | (0.382) | (0.382) | (0.206) |
| **Age group (<30 omit.)** | | | | | |
| 30-49 | 0.738 | 0.662* | 0.614* | 0.613* | 0.657*** |
|  | (0.246) | (0.116) | (0.119) | (0.118) | (0.078) |
| 50+ | 1.226 | 0.959 | 0.893 | 0.896 | 0.964 |
|  | (0.467) | (0.219) | (0.214) | (0.215) | (0.145) |
| **N** | 806 | 1503 | 1832 | 1832 | 4141 |
| **Pseudo R-squared** | 0.120 | 0.070 | 0.085 | 0.085 | 0.075 |

Source: Author's calculations based on ELMPS 1998-2018

Notes: *p<0.05; **p<0.01; ***p<0.001.


### 4.5.2. Entry of firms

We turn now to entry of firms. We consider new firms to be those that started between waves. Figure 17 shows the percentage of firms that are new by subsequent wave formality. Informal firms are more likely to be new firms. In 2018, 73% of all informal firms were new compared to 55% for formal firms. Figure 18 shows the percentage of firms that are formal by whether they are new. The figure confirms that already established firms are more likely to be formal compared to new ones. Formality has fallen over time across both established and new firms. In 2018, 44% of established firms were formal compared to 26% of new firms. There is suggestive evidence that firms may start as informal (and later formalize). Taken with the higher rates of exit for informal firms, it may also indicate that this sector is more dynamic, has greater churn, or is a fallback form of employment (necessity entrepreneurship).

**Figure 17. Percentage of firms that are new by subsequent wave formality**
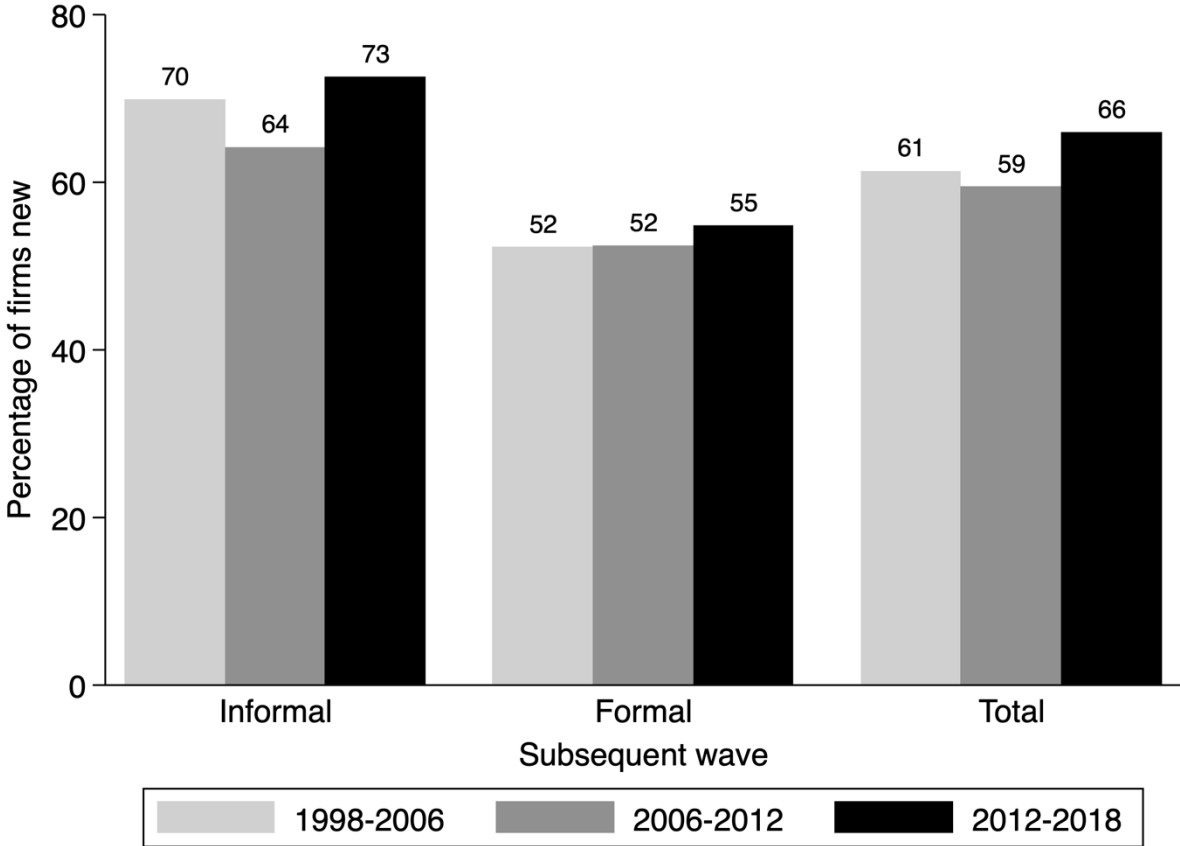


Source: Authors' calculations based on ELMPS 1998, ELMPS 2006, ELMPS 2012, ELMPS 2018

**Figure 18. Percentage of firms formal in subsequent wave by whether firm is new**
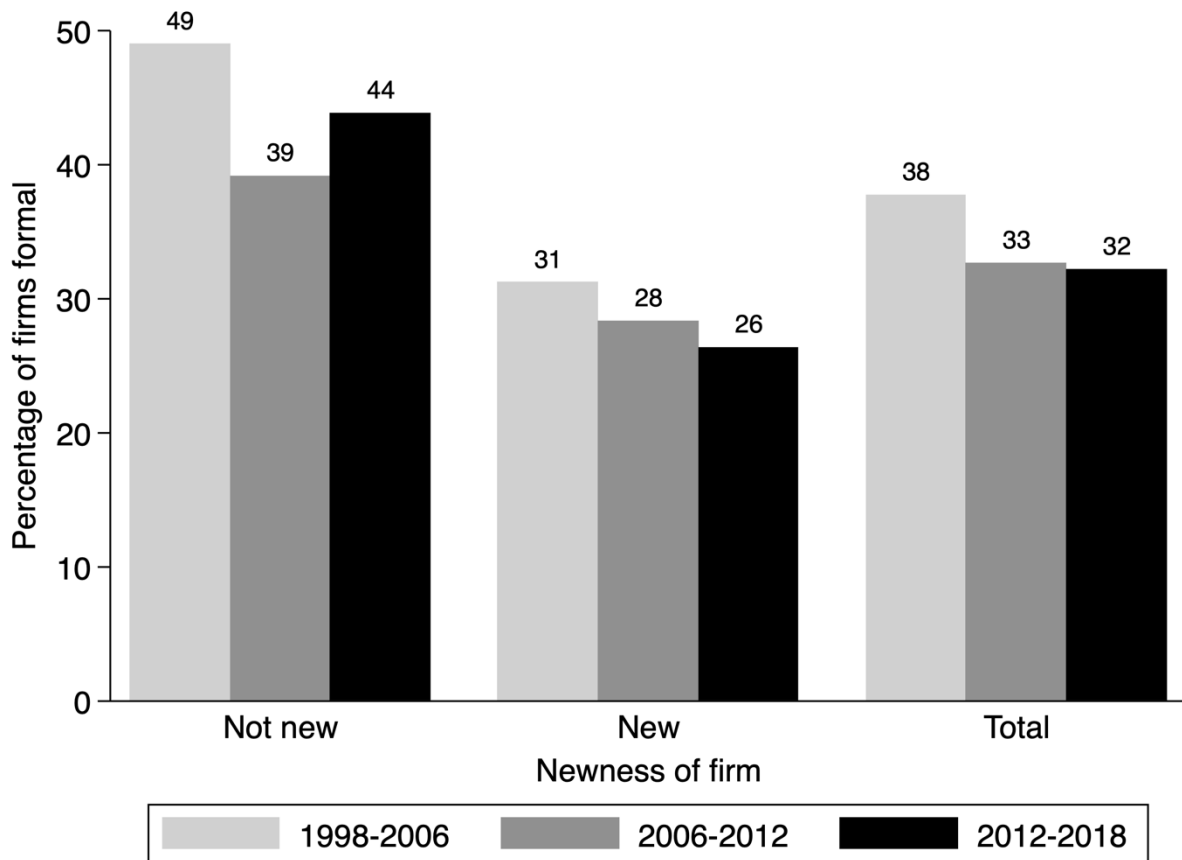


Source: Authors' calculations based on ELMPS 1998, ELMPS 2006, ELMPS 2012, ELMPS 2018

### 4.5.3. (In)formalization of firms

We now turn to examining the dynamics of (in)formalization among firms that survived across a pair of waves. Firms have to renew their commercial registration every five years, including bringing a copy of their original registration, documentation of taxes, Chamber of Commerce membership, and documentation of paying workers' insurance (Almaal, 2020). Thus, between waves, every firm that was formal would have had to renew their commercial registration to remain formal, otherwise they would potentially informalize. Firms that already existed could also have registered and formalized. Figure 19 shows the percentage of firms changing formality in the subsequent wave by their base wave formality. Rates should be interpreted with some caution given potential for measurement error. Only 27% of firms that existed in 2012 formalized by 2018, however, this was an increase from 2006-2012 where only 18% did so and 1998-2006 when only 24% of them did so. Some firms also informalized, more (29-30%) in 2006-2012 and 2012-2018 than in 1998-2006.

**Figure 19. Percentage of firms changing formality in subsequent wave by base wave formality, firms that survived**
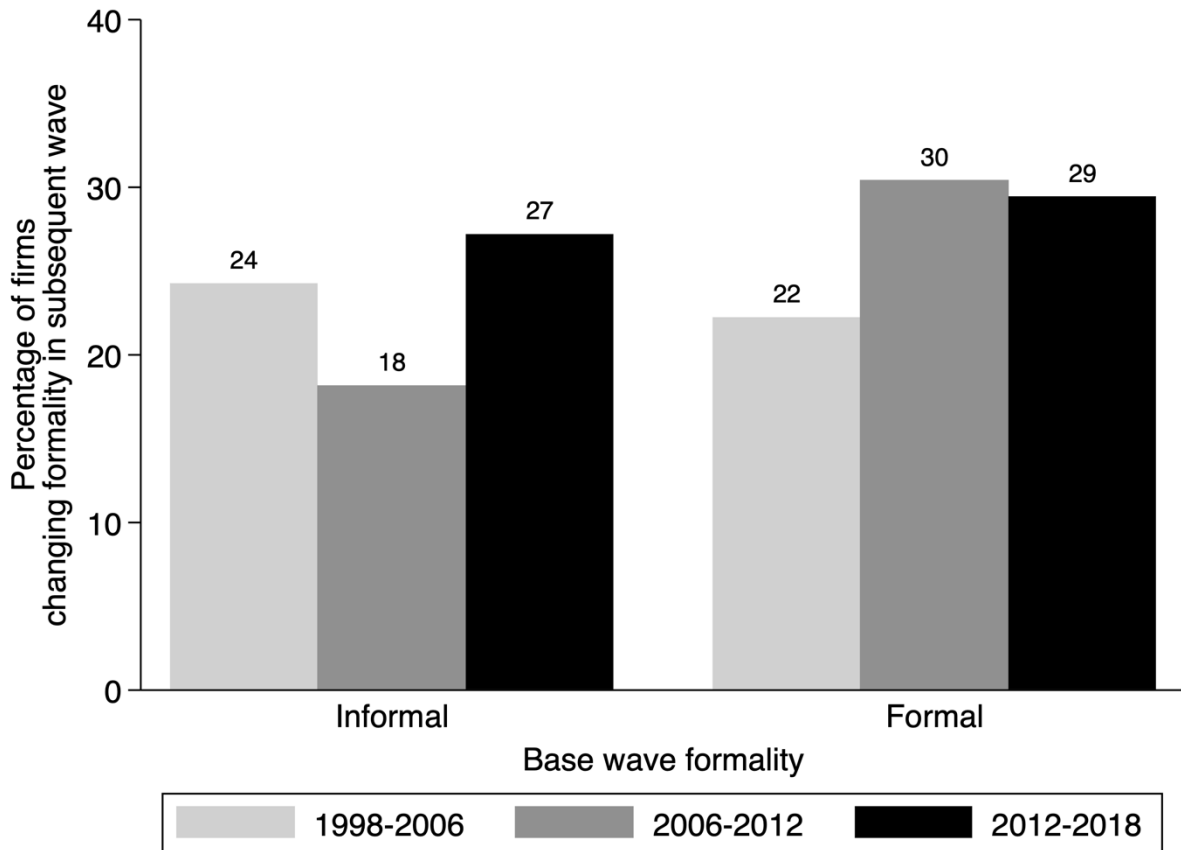


Source: Authors' calculations based on ELMPS 1998, ELMPS 2006, ELMPS 2012, ELMPS 2018

Table 9 explores how changing formality depends on base wave firm characteristics, for firms that survived. The pairs of waves are pooled given the limited sample size, but split into initially informal and initially formal firms. We also show results for 2018, but a number of characteristics drop due to being perfect predictors. Compared to the reference period of 1998-2006, formalization is less common in other periods, significantly so in 2006-2012, and informalization more common, although insignificantly so. There are not significant relationships with firm size in the base wave. Results by economic activity appear driven by sparse cells. Compared to Greater Cairo, other areas are more likely to formalize, but also some are more likely to informalize. There are no significant results by capital, but there is some suggestive evidence that older firms are significantly more likely to formalize. Location appears to be one of the strongest predictors; firms outside establishments are significantly less likely to formalize than those in shops. Woman-owned informal firms are significantly less likely to formalize. Formalization increases with education and informalization decreases, one of the clearest relationships. Base wave log revenue per worker is not a significant predictor of (in)formalization in 2012-2018.

**Table 9. Odds ratios from logit model of changing formality in subsequent wave by base wave formality, ELMPS 1998-2018**

| | Pooled base informal | Pooled base formal | 2012-2018 Base informal | 2012-2018 Base formal |
|---|---|---|---|---|
| **Wave (1998-2006 omit.)** | | | | |
| 06-12 | 0.514* | 1.762 | | |
| | (0.156) | (0.545) | | |
| 12-18 | 0.541 | 1.734 | | |
| | (0.187) | (0.635) | | |
| **Ent. size (one worker omit.)** | | | | |
| 2 | 0.793 | 1.271 | 1.453 | 1.508 |
| | (0.279) | (0.349) | (1.081) | (0.707) |
| 3-4 | 0.650 | 0.948 | 0.178 | 2.696 |
| | (0.316) | (0.308) | (0.181) | (1.804) |
| 5-24 | 2.677 | 0.363 | | |
| | (2.489) | (0.228) | | |
| **Industry (wholesale/retail omit.)** | | | | |
| Manufacturing and related trades | 1.534 | 2.271* | 1.768 | 0.674 |
| | (0.655) | (0.907) | (1.217) | (0.621) |
| Construction | 0.871 | 304.036*** | 0.597 | |
| | (0.685) | (354.038) | (0.832) | |
| Transportation and storage | 1.009 | 551.995*** | 0.864 | |
| | (0.806) | (825.188) | (1.012) | |
| Accommodation and food service | 3.036 | 0.560 | 2.276 | 0.520 |
| | (2.022) | (0.276) | (4.233) | (0.334) |
| Various professional acts. | 0.808 | 27.048*** | 1.118 | 14.497 |
| | (0.570) | (26.003) | (1.600) | (22.405) |
| Other service | 1.520 | 2.460 | 0.935 | 2.534 |

|  | Pooled base informal | Pooled base formal | 2012-2018 Base informal | 2012-2018 Base formal |
|---|---|---|---|---|
|  | (0.793) | (1.476) | (0.845) | (2.456) |
| **Region (Greater Cairo omit.)** |  |  |  |  |
| Alexandria and Suez Canal | 2.676* | 1.046 | 5.429 | 1.324 |
|  | (1.236) | (0.422) | (4.908) | (1.172) |
| Lower Egypt | 2.394* | 2.152* | 3.390 | 6.303* |
|  | (0.846) | (0.688) | (2.227) | (4.584) |
| Upper Egypt | 2.653* | 1.564 | 4.122* | 2.933 |
|  | (1.035) | (0.536) | (2.949) | (2.093) |
| **Partnership (sole proprietorship omit.)** |  |  |  |  |
| Collective Ownership | 2.087 | 0.646 | 2.832 | 1.239 |
|  | (0.805) | (0.204) | (1.669) | (0.728) |
| **Capital (none omit.)** |  |  |  |  |
| less than LE 100 | 0.882 | 1.801 | 1.639 | 1.504 |
|  | (0.404) | (2.017) | (1.333) | (2.146) |
| LE 100-499 | 0.649 | 2.744 | 0.606 | 1.495 |
|  | (0.300) | (2.866) | (0.564) | (2.003) |
| LE 500-999 | 0.953 | 2.423 | 0.393 | 0.867 |
|  | (0.418) | (2.474) | (0.301) | (1.130) |
| LE 1000-4999 | 0.590 | 4.010 | 0.649 | 2.869 |
|  | (0.281) | (4.132) | (0.607) | (3.455) |
| LE5000-9999 | 1.416 | 2.472 | 1.325 | 0.722 |
|  | (0.698) | (2.593) | (1.165) | (0.854) |
| LE10 000 or more | 1.783 | 1.976 | 4.577 |  |
|  | (1.298) | (2.367) | (4.983) |  |
| **Ln net revenue per worker** |  |  | 1.164 | 1.410 |
|  |  |  | (0.195) | (0.290) |

|  | Pooled base informal | Pooled base formal | 2012-2018 Base informal | 2012-2018 Base formal |
|---|---|---|---|---|
| **Firm start year (1980-1989 omit.)** | | | | |
| Don't know | 4.605* | 0.290 | 4.288 | 0.318 |
|  | (3.176) | (0.296) | (4.394) | (0.363) |
| 1950-1959 | 0.664 | 0.695 |  | 1.367 |
|  | (0.614) | (0.557) |  | (2.219) |
| 1960-1969 | 2.126 | 0.410 |  |  |
|  | (1.397) | (0.287) |  |  |
| 1970-1979 | 2.875* | 0.848 | 1.805 | 2.490 |
|  | (1.226) | (0.330) | (2.677) | (2.949) |
| 1990-1999 | 0.882 | 1.003 | 1.618 | 1.859 |
|  | (0.269) | (0.313) | (1.010) | (1.370) |
| 2000-2009 | 1.447 | 1.171 | 1.771 | 1.560 |
|  | (0.510) | (0.443) | (1.002) | (1.085) |
| 2010-2018 | 1.335 | 3.099 | 2.166 | 4.014 |
|  | (0.833) | (2.013) | (1.864) | (3.605) |
| Before 1950 |  | 0.429 |  |  |
|  |  | (0.494) |  |  |
| **People outside the HH employed (no omit.)** | | | | |
| Yes | 1.175 | 0.666 | 1.380 | 0.810 |
|  | (0.384) | (0.182) | (0.733) | (0.434) |
| **Location (shop omit.)** | | | | |
| Own home | 0.321** | 0.223* | 0.205* | 0.249 |
|  | (0.122) | (0.166) | (0.147) | (0.308) |
| Office/flat/building/rooms | 0.571 | 0.056** | 0.555 | 0.131 |
|  | (0.378) | (0.057) | (0.670) | (0.211) |
| Workshop/factory | 1.125 | 0.784 | 0.650 | 0.493 |

63

| | Pooled base informal | Pooled base formal | 2012-2018 Base informal | 2012-2018 Base formal |
|---|---|---|---|---|
| | (0.517) | (0.343) | (0.490) | (0.542) |
| Mobile worker | 0.164*** | 0.031* | 0.132* | |
| | (0.076) | (0.044) | (0.128) | |
| Kiosk/hut | 0.520 | 3.740 | 1.544 | |
| | (0.390) | (3.275) | (1.630) | |
| Transport based | 0.185* | | 0.065** | |
| | (0.148) | | (0.066) | |
| Street/field/farm | 0.276** | 44.965** | 0.503 | |
| | (0.117) | (62.582) | (0.287) | |
| **Sex (male omit.)** | | | | |
| Female | 0.195*** | 0.801 | 0.183* | 0.454 |
| | (0.095) | (0.403) | (0.132) | (0.423) |
| **Education (illit. omit.)** | | | | |
| Less than sec. | 1.817* | 0.489* | 2.039 | 0.247 |
| | (0.523) | (0.161) | (1.194) | (0.196) |
| Secondary | 1.556 | 0.286** | 1.191 | 0.202* |
| | (0.526) | (0.113) | (0.732) | (0.164) |
| Higher education | 2.399* | 0.268** | 2.829 | 0.207 |
| | (1.064) | (0.115) | (2.562) | (0.180) |
| **Occupation (prof./tech. omit.)** | | | | |
| Other white collar | 1.021 | 1.719 | 0.785 | 2.114 |
| | (0.371) | (0.707) | (0.472) | (1.401) |
| Blue collar craft | 0.488 | 0.762 | 0.162** | 0.554 |
| | (0.213) | (0.286) | (0.102) | (0.458) |
| Blue collar non-craft | 0.974 | 0.778 | 2.094 | 0.254 |
| | (0.324) | (0.552) | (1.315) | (0.284) |

|  | Pooled base informal | Pooled base formal | 2012-2018 Base informal | 2012-2018 Base formal |
|---|---|---|---|---|
| **Age group (<30 omit.)** | | | | |
| 30-49 | 1.120 | 1.140 | 0.985 | 1.963 |
|  | (0.368) | (0.446) | (0.505) | (1.756) |
| 50+ | 1.106 | 0.775 | 2.063 | 2.646 |
|  | (0.434) | (0.360) | (1.480) | (2.703) |
| **N** | 788 | 675 | 295 | 190 |
| **Pseudo R-squared** | 0.193 | 0.153 | 0.279 | 0.192 |

Source: Author's calculations based on ELMPS 1998-2018

Notes: *p<0.05; **p<0.01; ***p<0.001. Blank cells indicate dropped perfect predictors

### 4.5.4. Formality and growth

We turn now to our last dynamic outcome, growth (an increase in the number of workers) among surviving firms (Figure 20). Formal firms are slightly more likely to grow than informal firms, although the gap in growth has narrowed over time. In 1998-2006, while 29% of informal firms grew, 35% of formal firms did so. However, in 2012-2018, while 19% of informal firms grew 21% of formal firms did so. Growth among surviving firms was clearly highest in 1998-2006 and lowest in 2006-2012; the increase in growth must be tempered by rising rates of exit (Figure 16).

**Figure 20. Percentage of firms that grew (increased workers) by base wave formality, firms that survived**
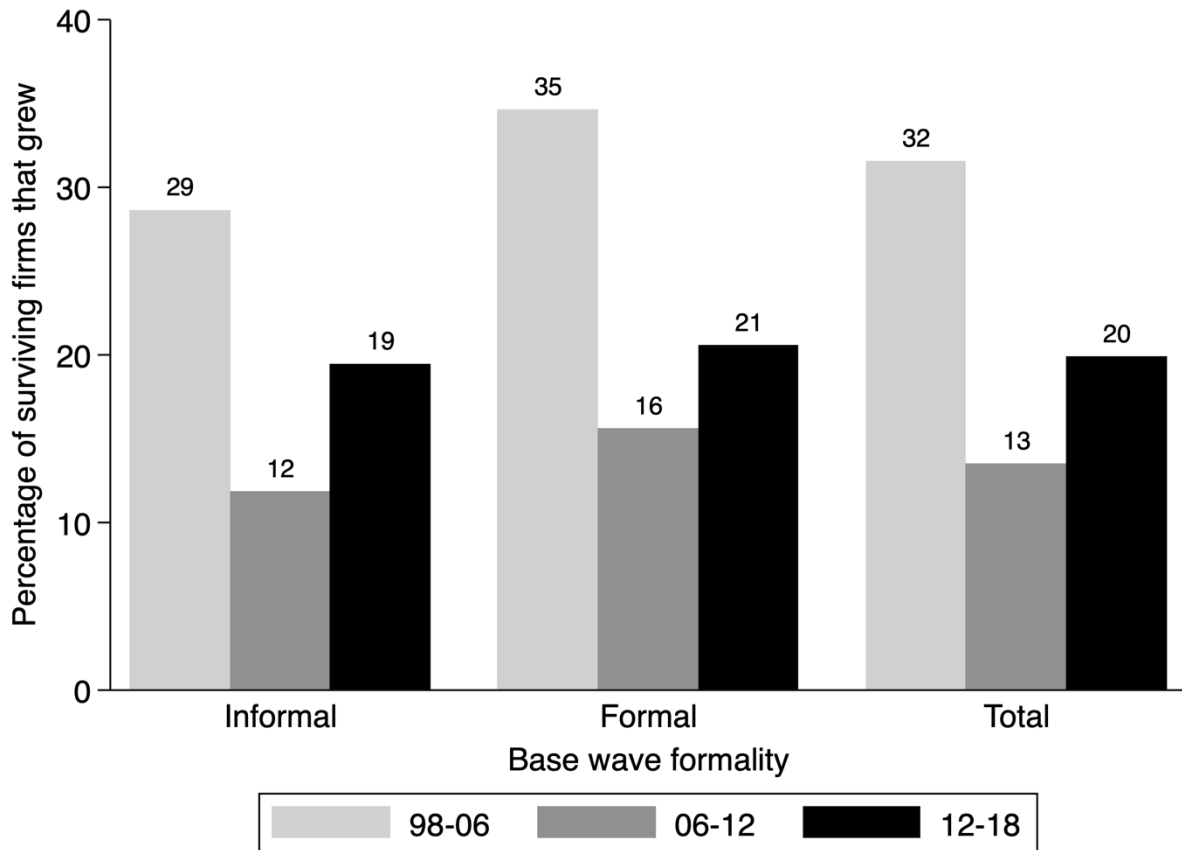


Source: Authors' calculations based on ELMPS 1998, ELMPS 2006, ELMPS 2012, ELMPS 2018

Table 10 presents the results of a model predicting, for firms that survived, whether they grew. There is a pooled model with base wave formality and interactions with round as well as a pooled model that adds subsequent wave formality. There is also a model with just 2012-18 to include log revenue per worker in the base wave. We focus our discussions on the results about formality, since models of growth have been presented elsewhere (Krafft, 2016). Base wave formality is not associated with growth. Nor are there significant interactions with wave. Although subsequent wave formality for the reference, 1998-2006 period, is not significantly associated with growth, when the interactions in 2006-2012 are added to the main effect, they are significant. That is, the firms that were formal in 2012 after accounting for their formality in 2006 (i.e. those who formalized) also grew. It may be that the firms that grew were better able to afford to formalize, but we cannot determine the exact sequencing of these changes.

**Table 10. Odds ratios from logit model of growth, by wave, ELMPS 1998-2018**

| | Pool | Pool with subsequent formality | 2012-2018 |
|---|---|---|---|
| **Base wave formality (informal omit.)** | | | |
| Formal | 1.096 | 1.233 | 0.729 |
| | (0.339) | (0.414) | (0.243) |
| **Subsequent wave formality (informal omit.)** | | | |
| Formal | | 1.287 | |
| | | (0.412) | |
| **Wave (1998-2006 omit.)** | | | |
| 06-12 | 0.371*** | 0.341*** | |
| | (0.105) | (0.106) | |
| 12-18 | 0.560* | 0.503* | |
| | (0.162) | (0.166) | |
| **Interactions between base formality and wave** | | | |
| 06-12 | 1.056 | 0.713 | |
| | (0.407) | (0.329) | |
| 12-18 | 0.897 | 0.779 | |
| | (0.337) | (0.332) | |
| **Interactions between subsequent formality and wave** | | | |
| 06-12 | | 1.660 | |
| | | (0.760) | |
| 12-18 | | 1.350 | |
| | | (0.580) | |
| **Ent. size (one worker omit.)** | | | |
| 2 | 0.308*** | 0.319*** | 0.332** |
| | (0.073) | (0.076) | (0.142) |
| 3-4 | 0.188*** | 0.174*** | 0.236* |

|  | Pool | Pool with subsequent formality | 2012-2018 |
|---|---|---|---|
|  | (0.063) | (0.059) | (0.152) |
| 5-24 | 0.191*** | 0.170*** | 0.557 |
|  | (0.082) | (0.073) | (0.510) |
| **Industry (wholesale/retail omit.)** |  |  |  |
| Manufacturing and related trades | 1.955* | 2.200* | 1.577 |
|  | (0.602) | (0.694) | (0.712) |
| Construction | 2.879* | 3.355** | 0.913 |
|  | (1.316) | (1.564) | (1.022) |
| Transportation and storage | 0.184*** | 0.189** | 0.102* |
|  | (0.091) | (0.098) | (0.099) |
| Accommodation and food service | 2.111* | 2.207* | 4.235* |
|  | (0.774) | (0.816) | (2.596) |
| Various professional acts. | 0.950 | 0.756 | 0.184 |
|  | (0.465) | (0.368) | (0.176) |
| Other service | 0.561 | 0.580 | 1.856 |
|  | (0.256) | (0.274) | (1.237) |
| **Region (Greater Cairo omit.)** |  |  |  |
| Alexandria and Suez Canal | 0.855 | 0.875 | 0.816 |
|  | (0.250) | (0.262) | (0.476) |
| Lower Egypt | 0.729 | 0.747 | 0.928 |
|  | (0.154) | (0.164) | (0.398) |
| Upper Egypt | 1.014 | 1.066 | 0.594 |
|  | (0.233) | (0.250) | (0.276) |
| **Partnership (sole proprietorship omit.)** |  |  |  |
| Collective Ownership | 1.136 | 1.079 | 1.041 |
|  | (0.292) | (0.276) | (0.415) |

|  | Pool | Pool with subsequent formality | 2012-2018 |
|---|---|---|---|
| **Capital (none omit.)** | | | |
| less than LE 100 | 0.895 | 0.878 | 0.986 |
| | (0.340) | (0.354) | (0.869) |
| LE 100-499 | 1.489 | 1.261 | 2.343 |
| | (0.591) | (0.525) | (2.039) |
| LE 500-999 | 1.557 | 1.521 | 1.267 |
| | (0.569) | (0.583) | (1.051) |
| LE 1000-4999 | 1.993 | 1.874 | 1.623 |
| | (0.775) | (0.753) | (1.340) |
| LE5000-9999 | 2.316* | 1.936 | 3.102 |
| | (0.934) | (0.806) | (2.616) |
| LE10 000 or more | 3.956** | 3.349* | 2.552 |
| | (2.038) | (1.790) | (2.958) |
| **Ln net revenue per worker** | | | 1.136 |
| | | | (0.137) |
| **Firm start year (1980-1989 omit.)** | | | |
| Don't know | 0.440 | 0.367 | 0.483 |
| | (0.297) | (0.246) | (0.432) |
| Before 1950 | 1.937 | 2.017 | |
| | (1.162) | (1.194) | |
| 1950-1959 | 1.873 | 1.837 | |
| | (1.211) | (1.209) | |
| 1960-1969 | 1.138 | 1.016 | |
| | (0.515) | (0.475) | |
| 1970-1979 | 1.975* | 1.821* | 2.668 |
| | (0.564) | (0.513) | (1.936) |

|  | Pool | Pool with subsequent formality | 2012-2018 |
|---|---|---|---|
| 1990-1999 | 1.211 | 1.208 | 1.641 |
|  | (0.273) | (0.272) | (0.802) |
| 2000-2009 | 1.130 | 1.126 | 1.177 |
|  | (0.309) | (0.314) | (0.568) |
| 2010-2018 | 2.327* | 2.382* | 2.428 |
|  | (0.991) | (1.050) | (1.459) |
| **People outside the HH employed (no omit.)** |  |  |  |
| Yes | 1.799** | 1.738* | 0.947 |
|  | (0.404) | (0.394) | (0.367) |
| **Location (shop omit.)** |  |  |  |
| Own home | 1.287 | 1.189 | 1.786 |
|  | (0.420) | (0.398) | (0.901) |
| Office/flat/building/rooms | 1.106 | 1.332 | 3.412 |
|  | (0.524) | (0.632) | (3.262) |
| Workshop/factory | 1.340 | 1.388 | 1.351 |
|  | (0.480) | (0.516) | (0.768) |
| Mobile worker | 1.399 | 1.608 | 0.871 |
|  | (0.504) | (0.614) | (0.586) |
| Kiosk/hut | 0.195 | 0.219 | 0.389 |
|  | (0.181) | (0.209) | (0.757) |
| Transport based | 1.271 | 1.683 | 0.152* |
|  | (0.630) | (0.907) | (0.129) |
| Street/field/farm | 1.069 | 1.220 | 1.202 |
|  | (0.381) | (0.449) | (0.702) |
| **Sex (male omit.)** |  |  |  |
| Female | 0.693 | 0.625 | 0.685 |

|  | Pool | Pool with subsequent formality | 2012-2018 |
|---|---|---|---|
|  | (0.208) | (0.201) | (0.390) |
| **Education (illit. omit.)** |  |  |  |
| Less than sec. | 1.075 | 0.909 | 1.344 |
|  | (0.252) | (0.219) | (0.566) |
| Secondary | 0.888 | 0.801 | 0.895 |
|  | (0.242) | (0.225) | (0.424) |
| Higher education | 0.723 | 0.648 | 1.201 |
|  | (0.219) | (0.202) | (0.642) |
| **Occupation (prof./tech. omit.)** |  |  |  |
| Other white collar | 0.967 | 1.039 | 0.528 |
|  | (0.257) | (0.288) | (0.249) |
| Blue collar craft | 0.605 | 0.614 | 0.405 |
|  | (0.184) | (0.191) | (0.205) |
| Blue collar non-craft | 1.106 | 1.158 | 2.129 |
|  | (0.373) | (0.393) | (1.137) |
| **Age group (<30 omit.)** |  |  |  |
| 30-49 | 0.821 | 0.783 | 0.602 |
|  | (0.208) | (0.205) | (0.212) |
| 50+ | 0.689 | 0.662 | 0.417 |
|  | (0.210) | (0.209) | (0.204) |
| **N** | 1533 | 1478 | 529 |
| **Pseudo R-squared** | 0.117 | 0.129 | 0.161 |

Source: Author's calculations based on ELMPS 1998-2018

Notes: *p<0.05; **p<0.01; ***p<0.001. Blank cells indicate dropped perfect predictors

## 5. Discussion and Conclusions

Egypt has struggled to create good, formal jobs for its young and growing labor force (World Bank, 2013; Gatti, Angel-Urdinola, Silva, & Bodor, 2014; Assaad, AlSharawy, & Salemi, 2019). Informal firms are a challenge not only for Egypt's workers, but also the macroeconomy and tax base (AfDB, 2016; Ali & Najman, 2016). For some time, there have been recommendations to formalize the largely informal sector of micro and small firms in Egypt in order to expand the tax base and increase the availability of good jobs (Egyptian Center for Economic Studies, 2005; World Bank, 2014). Yet such calls assume that firms that are currently informal are formalizable, that they could afford and survive formalization, without negatively impacting job creation. This assumption has not been tested or substantiated in Egypt.

In an attempt to answer whether informal firms may be formalizable, in this paper we compared formal and informal non-agricultural firms with 1-24 workers. We used multiple data sources on micro and small firms in Egypt and analyzed which firm and owner characteristics predict formality. We used the predicted probability of formality to characterize informal firms that are like formal firms and may be amenable to formalization efforts, in contrast to those firms that are quite different and are thus highly unlikely to formalize. For the most part we used an expanded definition of formality that counts a firm as formal if it has a commercial registration or if it pays social insurance premiums or taxes. In the dynamic analyses, we used a more basic definition that just depends on the commercial registration status of the firm.

Although the different data sources – the 2017 Establishment Census (EsC), the 2012-13 Economic Census (EcC), and the various waves of the Egypt Labor Market Panel Survey (ELMPS) – produce different rates of formality for our universe, they demonstrated similar relationships between formality and observable characteristics. Results on the relationship between formality and characteristics were comparable across firm size (number of workers), firm age, economic activity and region. In a subset of the data sets, measures of firm productivity and whether or not they hired outside workers also showed comparable results. Only the ELMPS data include firms that operate outside fixed establishments, which are predominantly informal. After establishing the highly informal nature of firms outside establishments, we restricted most of our analyses to firms operating within fixed establishments, which can be found across all our data sources.

We generally found that firm size and firm age are strong predictors of formality, with larger and older firms being more likely to be formal. Industry and region were, at best, inconsistent if not unreliable predictors of formality. Firms that hired wage workers, especially non-relatives, were more likely to be formal, as were the ones that had higher labor productivity or net revenue per worker. When owner characteristics were available, we found that owners with secondary education, but especially higher education were more likely to have formal firms and so were prime age and older owners compared to owners under the age of 30.

When we compared the distributions of formal and informal firms in terms of their predicted probabilities of being formal, we found that there was some overlap. This result suggests that some informal firms are similar to formal firms and thus may be amenable to formalization. The overlap was larger when only firms that operate in fixed establishments were compared and when there was a limited set of

observables to predict the probability of formality. As the set of predictors expanded, the overlap between the two distributions declined. This reduced overlap means that, while we are better able to identify informal firms that look like some of their formal counterparts, there are fewer of them. With more detailed characteristics, we can see that fewer informal firms are similar to formal firms and fewer will be potentially formalizable. Informal firms operating outside establishments do not seem to have many counterparts that are formal and therefore are particularly unlikely to be formalized.

We used two different methods to group or cluster informal firms based on their observable characteristics and then classify these groupings according to their predicted probability of being formal. The first method was to group informal firms based on a combination of the most important predictors of formality, broken down into discrete categories. We made sure to aggregate these groups in such a way that no groups contained fewer than 10 informal firms. The second method was to use cluster analysis to create 10 clusters with maximum similarity between the observables in each cluster and dissimilarity across clusters. Using these methods separately on the EcC data and pooled ELMPS data, we identified the profiles of firms that were least likely to be amenable to formalization and those of firms that were most likely to have formal counterparts and be potentially formalizable.

The grouping analysis suggests that firms not amenable to formalization include young one- or two-person firms with no hired labor, irrespective of productivity. These firms make up 22% of informal firms in the ELMPS and 34% of informal firms in the EcC. Groups with low probability of formality also include older one-person firms with relatively low revenue per worker, which constitute another 20% of informal firms in the ELMPS. Cluster analysis on ELMPS data identified three clusters as least amenable to formalization, which we named "women-owned retail firms," "youth-owned young firms," and "smaller professional firms." These clusters together make up 22% of informal firms and tend to skew young. Cluster analysis on the EcC data identified small professional, service and manufacturing firms and one-person retail firms as least amenable to formalization, which together made up 45% of informal firms in the EcC. On the other end of the spectrum, we identified the profiles of firms that have a relatively high probability of formality. The ELMPS grouping analysis identified older, more productive firms, older firms with more educated owners, and older larger firms. The grouping analysis on EcC identifies older larger firms that either hire wage workers or that are more productive as having a higher probability of formality.

What do our results imply about the determinants of formality and why they might matter? The cluster analysis on both data sets puts the emphasis on older and larger firms that hire wage workers as more amenable to formalization. These patterns are consistent with the fact that formalization imposes certain fixed costs and administrative hurdles (AfDB, 2016; World Bank, 2020) that younger and smaller firms are not able to afford.

Higher productivity is associated with greater formality, but the direction of causality is not clear. It seems more likely that more productive firms are better able to afford the costs of formality than it is that formality leads to higher productivity. Moreover, the grouping and cluster analyses demonstrated that firms' productivity level was subordinate to firm size and age in terms of predicting formality. While productivity is a factor in decisions to formalize, it does not appear to be the driving factor in Egypt.

Globally, overlap in productivity between formal and informal firms varies substantially by country context (Gelb, Mengistae, Ramachandran, & Shah, 2009; Benjamin & Mbaye, 2012).

Having a more educated owner is more conducive to formalization presumably because it makes it easier to negotiate the complex bureaucracy of formality (Almaal, 2020; World Bank, 2020). The strong association of formality with firm age also has two possible explanations: firms formalize later in the life cycle or informal firms are less likely to survive to old age. Our analysis of firm dynamics and informality, which we turn to next, suggests that the second explanation has some truth to it, but it is not clear whether the difference in survival between formal and informal firms is sufficient to account for the strong association between firm age and formality.

Our analysis of firms' dynamics used panel data from different waves of the ELMPS to examine the relationship between formality and new firm formation, firm survival, and firm growth and formalization among firms that survive. Our findings point to the fact that informal firms experience more churn, with higher rates of firm formation and destruction among informal firms. Informal firms also seem to be more susceptible to macroeconomic conditions, exhibiting a clearer pro-cyclical pattern than formal firms.

Our data show that formalization does occur during the life cycle of firms, but that informalization also occurs, which we interpret as a possible failure to renew firm registration when it expires. In any case, registration of new businesses cannot be interpreted as new firm formation, since formalization could come fairly late in the life cycle of existing informal firms. Moreover, requiring firms to renew their registration every five years and the costs that this imposes may be a factor that contributes to higher rates of informality. Conditional on survival across waves, employment growth at the firm level does not seem to be strongly associated with the formality status in the base year, but is associated with subsequent formality, suggesting that firms that grow are also more able to afford formality.

Our results have important implications for current debates about the need to extend the scope of formality to a broader range of firms. First, it is clear that such efforts must be accompanied by attempts to streamline the process and dramatically reduce its costs, while enhancing the immediate benefits of formality, such has access to credit and input and product markets. While the costs of formality declined dramatically in Egypt from 69% to 20% of per capita income from 2007 to 2020, they may still be prohibitive for newly formed small firms (World Bank, 2006, 2020). Other aspects of formality, such as paying taxes, enforcing contracts and registering property remain prohibitively onerous and expensive. Firms presumably compare the costs and benefits of formalization when making decisions and employers who are informal in Egypt identify high costs (including time and effort) and a lack of advantages as their reasons for remaining informal (World Bank, 2014; AfDB, 2016).

Any push toward formalization should be targeted to firms that may be able to afford it, namely somewhat larger and older informal firms that hire wage workers. Efforts should, at least initially and in the current policy environment, avoid firms for whom there are few formal counterparts, such as firms operating outside fixed establishments or young/small firms. This and other work suggests that such firms, along with those with younger or female owners, no hired labor, or older low-productivity firms may be more survival self-employment than growth-oriented entrepreneurship (Krafft, 2016; Krafft & Rizk, 2018).

It is important to keep in mind that employment in 1-24 worker non-agricultural firms in establishments is only 22% of Egypt's employment. Informality of employment, even among formal firms, remains an issue in other economic sectors. Within the micro and small establishment segment of the economy, our results suggest that there are some currently informal firms that are potentially amenable to formalization. However, only some of those potentially formalizable firms are likely to actually formalize even if renewed efforts to increase formality are rolled out. There is also a large fraction of informal firms that are very unlikely to formalize given the current economic and policy environment. It is an important and unanswered question whether further reducing the costs and raising the benefits of formality would shift this calculus.

Reforms designed to lower the cost and complexity of formality for micro-enterprises in Brazil did increase formality there (Fajnzylber, Maloney, & Montes-Rojas, 2011). Particularly notably, reform-induced increases in formality also led to higher profitability and increased employment. A key channel for this effect was shifting production to a permanent location. Given the rapid growth of employment outside establishments in Egypt, employment with the worst working conditions (Assaad, AlSharawy, & Salemi, 2019), reforms that lowered costs in ways that shifted employment into establishments could be positive. Conversely, increased enforcement of existing regulations, without changing the costs and benefits of formality, might drive employment out of establishments and worsen working conditions as well as other economic outcomes. Efforts to expand formality must be carefully designed to maximize the benefits for the economy and workers and minimize the costs.

**References**

AfDB. (2016). *Addressing Informality in Egypt. AfDB Working Paper North Africa Policy Series*.

Ali, N., & Najman, B. (2016). *Informal Competition, Firms' Productivity and Policy Reforms in Egypt. Economic Research Forum Working Paper Series No. 1025*. Cairo, Egypt.

Almaal. (2020). The Validity Period of the Commercial Registration in Egypt and the Conditions Necessary for Its Extraction (Arabic). Retrieved April 11, 2020 from https://www.almaal.org/commercial-register#i-4

Assaad, R., AlSharawy, A., & Salemi, C. (2019). *Is The Egyptian Economy Creating Good Jobs? Job Creation and Economic Vulnerability from 1998 to 2018. Economic Research Forum Working Paper Series No. 1354*. Cairo, Egypt.

Assaad, R., Hendy, R., Lassassi, M., & Yassin, S. (2018). *Explaining the MENA Paradox: Rising Educational Attainment, Yet Stagnant Female Labor Force Participation. IZA Discussion Paper Series No. 11385*. Bonn, Germany.

Assaad, R., & Krafft, C. (2020). Excluded Generation: The Growing Challenges of Labor Market Insertion for Egyptian Youth. *Journal of Youth Studies*.

Assaad, R., Krafft, C., & Salemi, C. (2019). *Socioeconomic Status and the Changing Nature of the School-to-Work Transition in Egypt, Jordan, and Tunisia. Economic Research Forum Working Paper Series No. 1287*. Cairo, Egypt.

Assaad, R., Krafft, C., & Selwaness, I. (2017). *The Impact of Marriage on Women's Employment in the Middle East and North Africa. Economic Research Forum Working Paper Series No. 1086*. Cairo, Egypt.

Benjamin, N. C., & Mbaye, A. A. (2012). The Informal Sector, Productivity, and Enforcement in West Africa: A Firm-Level Analysis. *Review of Development Economics*, *16*(4), 664–680.

Brixiova, Z. (2010). Unlocking Productive Entrepreneurship in Africa's Least Developed Countries. *African Development Review*, *22*(3), 440–451.

Egyptian Center for Economic Studies. (2005). *The Case for Formalization of Business in Egypt. ECES Policy Viewpoint No. 17*.

Elshamy, H. M. (2015). Measuring the Informal Economy in Egypt. *International Journal of Business Management and Economic Research*, *6*(2), 137–142.

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis*. Chicester, UK: John Wiley & Sons.

Fajnzylber, P., Maloney, W. F., & Montes-Rojas, G. V. (2011). Does Formality Improve Micro-Firm Performance? Evidence from the Brazilian SIMPLES Program. *Journal of Development Economics*, *94*(2), 262–276.

Gatti, R., Angel-Urdinola, D. F., Silva, J., & Bodor, A. (2014). *Striving for Better Jobs: The Challenge of Informality in the Middle East and North Africa*. Washington, DC: World Bank.

Gelb, A., Mengistae, T., Ramachandran, V., & Shah, M. K. (2009). *To Formalize or Not to Formalize? Comparisons of Microenterprise Data from Southern and East Africa*. *Center for Global Development Working Paper Series No. 175*.

Krafft, C. (2016). *Understanding the Dynamics of Household Enterprises in Egypt: Birth, Death, Growth, and Transformation*. *Economic Research Forum Working Paper Series No. 983*. Cairo, Egypt.

Krafft, C., Assaad, R., & Rahman, K. W. (2019). *Introducing the Egypt Labor Market Panel Survey 2018*. *Economic Research Forum Working Paper Series No. 1360*. Cairo, Egypt.

Krafft, C., Keo, C., & Fedi, L. (2019). *Rural Women in Egypt: Opportunities and Vulnerabilities*. *Economic Research Forum Working Paper Series No. 1359*. Cairo, Egypt.

Krafft, C., & Rizk, R. (2018). *The Promise and Peril of Youth Entrepreneurship in MENA*. *Economic Research Forum Working Paper Series No. 1257*. Cairo, Egypt.

Li, Y., & Rama, M. (2015). Firm Dynamics, Productivity Growth, and Job Creation in Developing Countries: The Role of Micro- and Small Enterprises. *The World Bank Research Observer*, *30*(1), 3–38.

Naudé, W. (2008). *Entrepreneurship in Economic Development*. *UNU-WIDER Research Paper No. 2008/20*. Helsinki, Finland.

Naudé, W. (2010). *Promoting Entrepreneurship in Developing Countries: Policy Challenges*. *UNU Policy Brief No. 4*.

Selwaness, I., & Ehab, M. (2019). *Social Protection and Vulnerability in Egypt: A Gendered Analysis*. *Economic Research Forum Working Paper Series No. 1363*. Cairo, Egypt.

Selwaness, I., & Krafft, C. (2018). *The Dynamics of Family Formation and Women's Work: What Facilitates and Hinders Female Employment in the Middle East and North Africa? Economic Research Forum Working Paper Series No. 1192*. Cairo, Egypt.

Sparks, D. L., & Barnett, S. T. (2010). The Informal Sector In Sub-Saharan Africa: Out Of The Shadows To Foster Sustainable Employment and Equity? *International Business & Economics Research Journal*, *9*(5), 1–12.

World Bank. (2006). *Doing Business 2007: How to Reform*. Washington, DC: World Bank.

World Bank. (2013). *Jobs for Shared Prosperity: Time for Action in the Middle East and North Africa*. Washington, DC: World Bank.

World Bank. (2014). *Arab Republic of Egypt: More Jobs, Better Jobs: A Priority for Egypt*. Washington, DC: World Bank.

World Bank. (2020). *Doing Businnss 2020: Economy Profile: Egypt, Arab. Rep. Doing Business*. Washington, DC: World Bank.

**Table 11. Odds ratios from logit model of extended formality with firm size, region, and industry interactions with data source**

| | |
|---|---|
| **Data (ELMPS 2018 omit.)** | |
| ELMPS 2012 in est. | 1.181 |
| | (0.613) |
| Economic Census | 0.143*** |
| | (0.060) |
| Establishment Census | 0.316** |
| | (0.129) |
| **Ent. size (one worker omit.)** | |
| 2 | 1.887*** |
| | (0.358) |
| 3-4 | 1.394 |
| | (0.368) |
| 5-24 | 9.753*** |
| | (5.119) |
| **Ent. size and data int.** | |
| ELMPS 2012 in est.*2 | 0.775 |
| | (0.213) |
| ELMPS 2012 in est.*3-4 | 1.504 |
| | (0.538) |
| ELMPS 2012 in est.*5-24 | 0.741 |
| | (0.524) |
| Economic Census*2 | 0.961 |
| | (0.195) |
| Economic Census*3-4 | 1.902* |
| | (0.539) |

| | |
|---|---|
| Economic Census*5-24 | 0.763 |
| | (0.405) |
| Establishment Census*2 | 0.863 |
| | (0.164) |
| Establishment Census*3-4 | 1.973* |
| | (0.521) |
| Establishment Census*5-24 | 0.524 |
| | (0.275) |
| **Ent. age (one year old omit.)** | |
| 0 | 0.339* |
| | (0.167) |
| 2 | 0.668 |
| | (0.307) |
| 3 | 0.959 |
| | (0.456) |
| 4 | 0.937 |
| | (0.438) |
| 5 | 2.116 |
| | (1.126) |
| 6 | 2.681 |
| | (1.360) |
| 7 | 2.875 |
| | (1.589) |
| 8-17 | 1.931 |
| | (0.712) |
| 18-27 | 1.299 |
| | (0.514) |
| 28+ | 2.869* |

|  |  |
|---|---|
|  | (1.221) |
| Don't know | 3.590** |
|  | (1.684) |
| **Ent. age and data  int.** |  |
| ELMPS 2012 in est.*0 | 2.021 |
|  | (1.639) |
| ELMPS 2012 in est.*2 | 1.489 |
|  | (0.930) |
| ELMPS 2012 in est.*3 | 1.984 |
|  | (1.211) |
| ELMPS 2012 in est.*4 | 2.561 |
|  | (1.575) |
| ELMPS 2012 in est.*5 | 1.453 |
|  | (1.033) |
| ELMPS 2012 in est.*6 | 0.626 |
|  | (0.437) |
| ELMPS 2012 in est.*7 | 0.752 |
|  | (0.514) |
| ELMPS 2012 in est.*8-17 | 1.243 |
|  | (0.596) |
| ELMPS 2012 in est.*18-27 | 3.540* |
|  | (1.866) |
| ELMPS 2012 in est.*28+ | 1.539 |
|  | (0.878) |
| ELMPS 2012 in est.*Don't know | 0.603 |
|  | (0.401) |
| Economic Census*0 | 2.435 |
|  | (1.229) |

| | |
|---|---|
| Economic Census*2 | 2.428 |
| | (1.144) |
| Economic Census*3 | 1.668 |
| | (0.813) |
| Economic Census*4 | 1.952 |
| | (0.938) |
| Economic Census*5 | 1.066 |
| | (0.583) |
| Economic Census*6 | 0.935 |
| | (0.489) |
| Economic Census*7 | 0.688 |
| | (0.398) |
| Economic Census*8-17 | 1.440 |
| | (0.555) |
| Economic Census*18-27 | 3.772** |
| | (1.540) |
| Economic Census*28+ | 1.759 |
| | (0.772) |
| Economic Census*Don't know | 0.518 |
| | (0.256) |
| Establishment Census*0 | 2.451 |
| | (1.206) |
| Establishment Census*2 | 1.768 |
| | (0.814) |
| Establishment Census*3 | 1.411 |
| | (0.670) |
| Establishment Census*4 | 1.504 |
| | (0.704) |

| | |
|---|---|
| Establishment Census*5 | 0.731 |
| | (0.389) |
| Establishment Census*6 | 0.636 |
| | (0.323) |
| Establishment Census*7 | 0.600 |
| | (0.332) |
| Establishment Census*8-17 | 1.269 |
| | (0.468) |
| Establishment Census*18-27 | 2.482* |
| | (0.983) |
| Establishment Census*28+ | 1.672 |
| | (0.712) |
| Establishment Census*Don't know | 1.000 |
| | (.) |
| **Industry (wholesale/retail omit.)** | |
| Manufacturing and related trades | 0.972 |
| | (0.256) |
| Construction | 0.491 |
| | (0.479) |
| Transportation and storage | 0.563 |
| | (0.430) |
| Accommodation and food service | 1.574 |
| | (0.532) |
| Various professional acts. | 1.552 |
| | (0.402) |
| Other service | 0.922 |
| | (0.251) |
| **Industry and data int.** | |

| | |
|---|---|
| ELMPS 2012 in est.*Manufacturing and related trades | 0.455* |
| | (0.156) |
| ELMPS 2012 in est.*Construction | 0.300 |
| | (0.336) |
| ELMPS 2012 in est.*Transportation and storage | 0.824 |
| | (1.045) |
| ELMPS 2012 in est.*Accommodation and food service | 0.519 |
| | (0.236) |
| ELMPS 2012 in est.*Various professional acts. | 0.487* |
| | (0.172) |
| ELMPS 2012 in est.*Other service | 0.529 |
| | (0.202) |
| Economic Census*Manufacturing and related trades | 0.779 |
| | (0.213) |
| Economic Census*Construction | 14.906* |
| | (17.389) |
| Economic Census*Transportation and storage | 3.381 |
| | (2.879) |
| Economic Census*Accommodation and food service | 0.504* |
| | (0.176) |
| Economic Census*Various professional acts. | 0.342*** |
| | (0.092) |
| Economic Census*Other service | 0.775 |
| | (0.218) |
| Establishment Census*Manufacturing and related trades | 0.658 |
| | (0.173) |
| Establishment Census*Construction | 1.065 |
| | (1.040) |

| | |
|---|---|
| Establishment Census*Transportation and storage | 0.914 |
| | (0.698) |
| Establishment Census*Accommodation and food service | 0.482* |
| | (0.163) |
| Establishment Census*Various professional acts. | 1.006 |
| | (0.261) |
| Establishment Census*Other service | 0.582* |
| | (0.158) |
| **Region (Greater Cairo omit.)** | |
| Alexandria and Suez Canal | 1.086 |
| | (0.356) |
| Lower Egypt | 0.937 |
| | (0.233) |
| Upper Egypt | 1.022 |
| | (0.263) |
| **Region and data int.** | |
| ELMPS 2012 in est.*Alexandria and Suez Canal | 1.130 |
| | (0.457) |
| ELMPS 2012 in est.*Lower Egypt | 1.012 |
| | (0.321) |
| ELMPS 2012 in est.*Upper Egypt | 0.911 |
| | (0.307) |
| Economic Census*Alexandria and Suez Canal | 1.969* |
| | (0.662) |
| Economic Census*Lower Egypt | 2.088** |
| | (0.536) |
| Economic Census*Upper Egypt | 2.679*** |
| | (0.743) |

| | |
|---|---|
| Establishment Census*Alexandria and Suez Canal | 1.885 |
| | (0.618) |
| Establishment Census*Lower Egypt | 1.535 |
| | (0.383) |
| Establishment Census*Upper Egypt | 1.384 |
| | (0.357) |
| **N** | 766840 |
| **Pseudo R-squared** | .129 |

Source: Authors' calculations based on ELMPS 2012, ELMPS 2018, EcC 2012/13, EsC 2017

Notes: *p<0.05; **p<0.01; ***p<0.001.

**Table 12. Odds ratios from logit "Model 2b" and "Model 3" of extended formality in ELMPS 2012 and 2018 including firms outside establishments**

| | Model 2 | | Model 3 | |
|---|---|---|---|---|
| | 2012 | 2018 | 2012 | 2018 |
| **Ent. size (one worker omit.)** | | | | |
| 2 | 1.262 | 1.417 | 1.268 | 1.454 |
| | (0.246) | (0.276) | (0.249) | (0.285) |
| 3-4 | 1.639 | 0.935 | 1.768 | 0.862 |
| | (0.460) | (0.282) | (0.534) | (0.254) |
| 5-24 | 5.705** | 6.067*** | 5.424** | 5.118** |
| | (3.596) | (3.221) | (3.456) | (2.661) |
| **Industry (wholesale/retail omit.)** | | | | |
| Manufacturing and related trades | 0.534* | 1.062 | 0.595 | 1.419 |
| | (0.138) | (0.309) | (0.166) | (0.425) |
| Construction | 0.202** | 1.198 | 0.198** | 1.538 |
| | (0.108) | (0.572) | (0.117) | (0.780) |
| Transportation and storage | 0.916 | 8.525*** | 0.879 | 7.282*** |
| | (0.691) | (4.208) | (0.666) | (4.122) |
| Accommodation and food service | 1.134 | 1.187 | 1.378 | 1.184 |
| | (0.370) | (0.393) | (0.432) | (0.450) |
| Various professional acts. | 1.378 | 1.910* | 1.038 | 1.469 |
| | (0.441) | (0.530) | (0.332) | (0.434) |
| Other service | 0.766 | 1.056 | 0.858 | 1.216 |
| | (0.205) | (0.271) | (0.249) | (0.328) |
| **Region (Greater Cairo omit.)** | | | | |
| Alexandria and Suez Canal | 1.251 | 1.390 | 1.412 | 1.522 |
| | (0.285) | (0.449) | (0.335) | (0.492) |
| Lower Egypt | 1.204 | 1.691* | 1.272 | 1.734* |
| | (0.238) | (0.411) | (0.257) | (0.449) |

| | Model 2 | | Model 3 | |
|---|---|---|---|---|
| | **2012** | **2018** | **2012** | **2018** |
| Upper Egypt | 0.901 | 1.325 | 1.053 | 1.400 |
| | (0.189) | (0.333) | (0.219) | (0.371) |
| **Partnership (sole proprietorship omit.)** | | | | |
| Collective Ownership | 1.377 | 1.745** | 1.529 | 1.942** |
| | (0.300) | (0.360) | (0.338) | (0.410) |
| **Capital (none omit.)** | | | | |
| less than LE 100 | 1.220 | 1.035 | 1.543 | 1.115 |
| | (0.507) | (0.378) | (0.701) | (0.413) |
| LE 100-499 | 1.892 | 0.846 | 2.682* | 0.929 |
| | (0.766) | (0.343) | (1.156) | (0.408) |
| LE 500-999 | 2.004 | 1.392 | 2.483* | 1.198 |
| | (0.766) | (0.464) | (1.010) | (0.419) |
| LE 1000-4999 | 3.782*** | 1.930 | 4.866*** | 1.780 |
| | (1.481) | (0.744) | (2.015) | (0.712) |
| LE5000-9999 | 8.939*** | 2.762*** | 10.450*** | 2.296* |
| | (3.372) | (0.846) | (4.174) | (0.743) |
| LE10 000 or more | 6.346*** | 4.203*** | 9.311*** | 3.456** |
| | (3.303) | (1.701) | (5.377) | (1.417) |
| **Ln net revenue per worker** | 1.212** | 1.071* | 1.212** | 1.073* |
| | (0.076) | (0.030) | (0.078) | (0.032) |
| **Ent. age (one year old omit.)** | | | | |
| 0 | 0.740 | 0.354* | 0.752 | 0.374* |
| | (0.372) | (0.151) | (0.434) | (0.172) |
| 2 | 1.222 | 0.538 | 1.115 | 0.465 |
| | (0.517) | (0.211) | (0.489) | (0.193) |
| 3 | 1.942 | 1.118 | 2.016 | 1.018 |

|  | Model 2 | | Model 3 | |
|---|---|---|---|---|
|  | **2012** | **2018** | **2012** | **2018** |
|  | (0.673) | (0.479) | (0.746) | (0.462) |
| 4 | 2.778** | 0.702 | 2.905** | 0.646 |
|  | (1.002) | (0.297) | (1.028) | (0.288) |
| 5 | 3.248** | 1.019 | 3.421** | 0.935 |
|  | (1.303) | (0.442) | (1.334) | (0.389) |
| 6 | 2.221* | 2.095 | 2.420* | 2.293 |
|  | (0.899) | (0.866) | (1.003) | (1.034) |
| 7 | 2.030 | 1.476 | 2.070* | 1.521 |
|  | (0.741) | (0.619) | (0.754) | (0.670) |
| 8-17 | 2.801*** | 1.466 | 2.544** | 1.406 |
|  | (0.761) | (0.467) | (0.748) | (0.488) |
| 18-27 | 5.449*** | 1.247 | 4.868*** | 1.268 |
|  | (1.689) | (0.424) | (1.652) | (0.460) |
| 28+ | 6.237*** | 2.581** | 4.360*** | 2.564* |
|  | (2.216) | (0.934) | (1.657) | (1.036) |
| Don't know | 2.798* | 2.043 | 2.400 | 1.891 |
|  | (1.268) | (0.780) | (1.156) | (0.783) |
| **Relatives outside the HH employed (no omit.)** | | | | |
| Yes | 0.995 | 1.022 | 1.026 | 0.928 |
|  | (0.250) | (0.277) | (0.259) | (0.242) |
| **Non-relatives outside the HH employed (no omit.)** | | | | |
| Yes | 1.458* | 2.179*** | 1.457 | 1.840** |
|  | (0.269) | (0.435) | (0.281) | (0.376) |
| **Location (shop omit.)** | | | | |
| Own home | 0.120*** | 0.255*** | 0.111*** | 0.311*** |
|  | (0.037) | (0.089) | (0.032) | (0.108) |

|  | Model 2 | | Model 3 | |
| --- | --- | --- | --- | --- |
|  | **2012** | **2018** | **2012** | **2018** |
| Office/flat/building/rooms | 0.467* | 0.467** | 0.288*** | 0.494** |
|  | (0.150) | (0.109) | (0.095) | (0.121) |
| Workshop/factory | 0.799 | 0.813 | 0.907 | 1.074 |
|  | (0.244) | (0.236) | (0.286) | (0.328) |
| Mobile worker | 0.031*** | 0.018*** | 0.037*** | 0.020*** |
|  | (0.014) | (0.007) | (0.017) | (0.008) |
| Kiosk/hut | 0.017*** | 0.849 | 0.016*** | 1.093 |
|  | (0.018) | (0.452) | (0.016) | (0.655) |
| Transport based | 0.181* | 0.034*** | 0.209* | 0.045*** |
|  | (0.138) | (0.017) | (0.162) | (0.024) |
| Street/field/farm | 0.020*** | 0.014*** | 0.026*** | 0.019*** |
|  | (0.012) | (0.010) | (0.015) | (0.012) |
| **Sex (male omit.)** | | | | |
| Female | | | 1.181 | 0.746 |
|  | | | (0.293) | (0.212) |
| **Education (illit. omit.)** | | | | |
| Less than sec. | | | 1.247 | 1.538 |
|  | | | (0.278) | (0.375) |
| Secondary | | | 1.803* | 2.273** |
|  | | | (0.413) | (0.587) |
| Higher education | | | 3.930*** | 3.471*** |
|  | | | (1.030) | (0.947) |
| **Occupation (prof./tech. omit.)** | | | | |
| Other white collar | | | 0.808 | 1.101 |
|  | | | (0.181) | (0.246) |
| Blue collar craft | | | 1.070 | 0.571* |

|  | Model 2 | | Model 3 | |
| --- | --- | --- | --- | --- |
|  | **2012** | **2018** | **2012** | **2018** |
|  |  |  | (0.286) | (0.159) |
| Blue collar non-craft |  |  | 1.118 | 1.100 |
|  |  |  | (0.288) | (0.287) |
| **Age group (<30 omit.)** |  |  |  |  |
| 30-49 |  |  | 1.849** | 1.140 |
|  |  |  | (0.370) | (0.246) |
| 50+ |  |  | 3.468*** | 1.408 |
|  |  |  | (0.826) | (0.354) |
| **N** | 2243 | 2140 | 2241 | 2067 |
| **Pseudo R-squared** | 0.410 | 0.296 | 0.434 | 0.321 |

Source: Authors' calculations based on ELMPS 2012, ELMPS 2018

Notes: *p<0.05; **p<0.01; ***p<0.001.

**Table 13. Ratio of characteristic to average (or standardized) characteristic, by cluster, pooled ELMPS 2012 and 2018**

| | Women-owned retail | Youth-owned new firms | Smaller prof. firms | Low-prod. firms with employees and older owners | Blue collar firms | Shops with prime-age owners | Workshops | Shops with older owners | Larger non-prof. firms | Larger prof. firms. |
|---|---|---|---|---|---|---|---|---|---|---|
| **Number of employees** | | | | | | | | | | |
| 1 | 1.17 | 1.01 | 1.34 | 0.09 | 1.23 | 1.07 | 1.04 | 1.25 | 1.02 | 0.53 |
| 2 | 0.91 | 1.11 | 0.18 | 4.19 | 0.54 | 1.15 | 0.68 | 0.33 | 0.52 | 1.61 |
| 3-4 | 0.97 | 0.97 | 0.00 | 2.01 | 0.26 | 0.63 | 1.25 | 0.87 | 2.28 | 3.82 |
| 5-24 | 0.00 | 0.00 | 1.25 | 1.75 | 0.00 | 0.00 | 2.25 | 0.00 | 4.87 | 6.52 |
| **Industry** | | | | | | | | | | |
| Manufacturing and related trades | 0.71 | 0.23 | 0.30 | 1.28 | 0.45 | 0.16 | 5.25 | 0.44 | 0.64 | 1.01 |
| Construction | 0.00 | 0.30 | 0.00 | 2.04 | 1.89 | 0.00 | 0.00 | 4.08 | 2.56 | 1.30 |
| Wholesale and retail | 1.54 | 1.19 | 0.41 | 0.86 | 1.32 | 1.37 | 0.22 | 1.19 | 1.12 | 0.12 |
| Transportation and storage | 0.00 | 5.23 | 0.00 | 0.00 | 0.00 | 2.47 | 2.59 | 0.00 | 0.00 | 0.00 |
| Accommodation and food service | 0.70 | 2.54 | 0.97 | 3.94 | 0.23 | 1.11 | 0.00 | 0.84 | 0.81 | 1.18 |
| Various professional acts. | 0.00 | 0.36 | 4.57 | 0.95 | 0.63 | 0.25 | 0.35 | 0.16 | 0.53 | 6.75 |
| Other service | 0.24 | 1.30 | 3.94 | 0.84 | 0.96 | 1.07 | 0.48 | 1.22 | 1.34 | 0.41 |
| **Region** | | | | | | | | | | |
| Greater Cairo | 0.71 | 0.85 | 1.34 | 1.21 | 0.96 | 1.05 | 0.64 | 1.65 | 0.75 | 1.30 |

|  | Women-owned retail | Youth-owned new firms | Smaller prof. firms | Low-prod. firms with employees and older owners | Blue collar firms | Shops with prime-age owners | Workshops | Shops with older owners | Larger non-prof. firms | Larger prof. firms. |
|---|---|---|---|---|---|---|---|---|---|---|
| Alexandria and Suez Canal | 1.54 | 1.69 | 1.50 | 1.26 | 1.12 | 0.77 | 0.75 | 0.90 | 0.76 | 1.64 |
| Lower Egypt | 0.79 | 0.82 | 1.08 | 1.00 | 0.91 | 0.84 | 1.87 | 0.79 | 1.60 | 0.78 |
| Upper Egypt | 1.56 | 1.22 | 0.72 | 0.92 | 1.21 | 1.36 | 0.26 | 0.77 | 0.88 | 1.19 |
| **Legal status of firm/establishment** | | | | | | | | | | |
| Collective Ownership | 0.26 | 1.05 | 0.98 | 0.69 | 0.11 | 0.43 | 1.31 | 0.30 | 7.51 | 0.82 |
| Sole Proprietorship | 1.09 | 1.00 | 1.03 | 1.06 | 1.09 | 1.06 | 1.01 | 1.07 | 0.33 | 1.04 |
| **Capital amount (in EGP)** | | | | | | | | | | |
| None | 1.24 | 0.10 | 5.50 | 0.82 | 1.26 | 0.70 | 1.18 | 0.56 | 0.75 | 2.78 |
| less than LE 100 | 2.65 | 1.06 | 0.57 | 1.07 | 2.03 | 0.78 | 0.60 | 1.33 | 0.60 | 0.53 |
| LE 100-499 | 1.60 | 1.10 | 0.50 | 0.71 | 0.93 | 0.95 | 1.05 | 1.51 | 0.74 | 1.62 |
| LE 500-999 | 0.98 | 0.80 | 1.88 | 1.46 | 1.21 | 0.97 | 0.94 | 1.22 | 0.35 | 0.55 |
| LE 1000-4999 | 1.12 | 1.26 | 0.16 | 1.23 | 0.90 | 1.47 | 1.23 | 0.37 | 1.17 | 1.06 |
| LE5000-9999 | 0.06 | 1.10 | 0.95 | 0.93 | 0.88 | 0.96 | 1.17 | 0.98 | 2.45 | 1.40 |
| LE10 000 or more | 2.99 | 1.87 | 0.00 | 1.45 | 0.38 | 0.41 | 0.74 | 1.97 | 1.86 | 1.49 |
| **Age of firm (years)** | | | | | | | | | | |
| 0 | 2.21 | 1.31 | 2.52 | 1.86 | 0.72 | 1.03 | 0.97 | 0.37 | 1.76 | 0.48 |
| 1 | 1.63 | 2.42 | 1.13 | 0.68 | 0.19 | 0.99 | 0.60 | 0.63 | 1.28 | 2.01 |

| | Women-owned retail | Youth-owned new firms | Smaller prof. firms | Low-prod. firms with employees and older owners | Blue collar firms | Shops with prime-age owners | Workshops | Shops with older owners | Larger non-prof. firms | Larger prof. firms. |
|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 1.18 | 2.53 | 0.82 | 0.60 | 0.78 | 1.00 | 0.84 | 0.68 | 2.63 | 0.32 |
| 3 | 0.57 | 1.17 | 1.78 | 0.85 | 0.32 | 1.65 | 1.41 | 1.05 | 0.97 | 1.36 |
| 4 | 0.83 | 1.32 | 1.43 | 1.74 | 0.57 | 1.38 | 1.43 | 0.79 | 1.58 | 0.34 |
| 5 | 0.43 | 2.17 | 2.16 | 1.43 | 0.57 | 0.97 | 0.99 | 1.55 | 2.06 | 0.64 |
| 6 | 2.33 | 1.54 | 0.96 | 1.80 | 0.00 | 1.76 | 0.51 | 0.00 | 2.79 | 0.84 |
| 7 | 1.23 | 1.46 | 0.18 | 1.71 | 0.00 | 1.40 | 2.19 | 0.44 | 0.38 | 1.67 |
| 8-17 | 0.90 | 0.75 | 0.96 | 0.80 | 2.79 | 0.72 | 1.01 | 0.75 | 0.49 | 1.19 |
| 18-27 | 1.07 | 0.21 | 1.25 | 1.43 | 0.46 | 1.15 | 1.59 | 1.62 | 0.95 | 1.35 |
| 28+ | 0.50 | 0.00 | 1.67 | 2.24 | 0.25 | 0.93 | 1.29 | 3.40 | 0.46 | 1.06 |
| Don't know | 1.89 | 0.62 | 0.00 | 1.25 | 0.21 | 1.17 | 0.40 | 1.30 | 1.57 | 1.97 |
| **Relatives outside the HH employed** | | | | | | | | | | |
| No | 1.08 | 0.99 | 1.02 | 0.96 | 1.04 | 1.04 | 0.93 | 1.04 | 0.83 | 0.97 |
| Yes | 0.00 | 1.17 | 1.02 | 1.41 | 0.97 | 0.53 | 1.78 | 0.78 | 2.96 | 1.46 |
| **Non-relatives outside the HH employed** | | | | | | | | | | |
| No | 1.23 | 1.14 | 1.19 | 0.87 | 1.11 | 1.12 | 0.79 | 1.12 | 1.03 | 0.21 |
| Yes | 0.07 | 0.59 | 0.41 | 1.46 | 0.56 | 0.51 | 1.82 | 0.52 | 1.02 | 4.10 |
| **Location** | | | | | | | | | | |

| | Women-owned retail | Youth-owned new firms | Smaller prof. firms | Low-prod. firms with employees and older owners | Blue collar firms | Shops with prime-age owners | Workshops | Shops with older owners | Larger non-prof. firms | Larger prof. firms. |
|---|---|---|---|---|---|---|---|---|---|---|
| Shop | 0.78 | 1.29 | 0.34 | 1.05 | 0.81 | 1.43 | 0.54 | 1.30 | 1.34 | 0.21 |
| Office/flat/building/rooms | 2.18 | 0.18 | 4.80 | 1.04 | 0.88 | 0.23 | 0.74 | 0.28 | 1.17 | 4.84 |
| Workshop/factory | 0.25 | 0.28 | 0.00 | 1.68 | 1.78 | 0.05 | 4.93 | 0.39 | 0.61 | 0.87 |
| Kiosk/hut | 3.23 | 2.11 | 0.79 | 1.32 | 2.05 | 0.75 | 0.00 | 1.28 | 1.17 | 0.53 |
| **Owner's sex** | | | | | | | | | | |
| Male | 0.05 | 0.95 | 0.99 | 1.14 | 1.13 | 1.04 | 1.14 | 1.18 | 1.03 | 1.09 |
| Female | 5.87 | 1.38 | 1.25 | 0.32 | 0.39 | 0.80 | 0.44 | 0.20 | 0.89 | 0.63 |
| **Owner's education** | | | | | | | | | | |
| Illiterate | 3.82 | 0.31 | 0.26 | 2.51 | 0.35 | 0.83 | 0.73 | 1.26 | 0.37 | 0.09 |
| Less than sec. | 0.66 | 1.17 | 0.76 | 0.98 | 0.45 | 1.18 | 1.88 | 1.41 | 0.70 | 0.41 |
| Secondary | 0.07 | 1.51 | 0.22 | 0.53 | 2.26 | 1.01 | 0.87 | 0.59 | 1.75 | 0.83 |
| Higher education | 0.09 | 0.82 | 4.08 | 0.64 | 0.27 | 0.95 | 0.33 | 1.04 | 0.89 | 3.40 |
| **Owner's occupation** | | | | | | | | | | |
| Professional/technical | 0.85 | 0.67 | 2.64 | 1.81 | 0.88 | 0.49 | 0.46 | 0.88 | 1.68 | 2.72 |
| Other white collar | 1.72 | 1.66 | 0.00 | 0.31 | 0.73 | 1.94 | 0.09 | 1.05 | 0.56 | 0.08 |
| Blue collar craft | 0.00 | 0.53 | 0.18 | 1.04 | 1.58 | 0.19 | 3.92 | 1.29 | 0.67 | 0.26 |
| Blue collar non-craft | 1.72 | 1.23 | 0.62 | 1.02 | 1.90 | 1.12 | 0.77 | 0.72 | 1.19 | 1.19 |

| | Women-owned retail | Youth-owned new firms | Smaller prof. firms | Low-prod. firms with employees and older owners | Blue collar firms | Shops with prime-age owners | Workshops | Shops with older owners | Larger non-prof. firms | Larger prof. firms. |
|---|---|---|---|---|---|---|---|---|---|---|
| **Owner's age group** | | | | | | | | | | |
| <30 | 0.41 | 7.10 | 2.70 | 0.83 | 0.21 | 0.00 | 0.61 | 0.00 | 0.66 | 0.47 |
| 30-49 | 0.48 | 0.00 | 0.43 | 0.52 | 1.66 | 1.72 | 1.39 | 0.00 | 1.48 | 1.57 |
| 50+ | 2.59 | 0.00 | 1.18 | 2.17 | 0.01 | 0.00 | 0.44 | 3.74 | 0.44 | 0.28 |
| | | | | | | | | | | |
| **Ln net revenue per worker** | -0.44 | 0.06 | 0.01 | -0.48 | 0.40 | -0.05 | 0.07 | -0.03 | 0.29 | 0.27 |
| **N** | 81 | 119 | 46 | 69 | 103 | 193 | 112 | 120 | 72 | 74 |
| **Percentage of Informal Firms in the Cluster** | 8.0 | 9.8 | 4.3 | 7.4 | 10.1 | 20.3 | 11.0 | 13.6 | 7.2 | 8.2 |

Source: Authors' calculations based on ELMPS 2012 and ELMPS 2018

Notes: Categorical variables are relative to the overall mean percentage for that category in the wave, continuous variables are standardized to that wave's mean.

**Table 14. Ratio of characteristic to average (or standardized) characteristic, by cluster, EcC 2012/13**

| | Smaller prof. firms | Smaller service firms | Smaller manuf. and rest. | One-person retail | Two person prof. firms | Two-worker retail | 3+ worker non-retail | Larger non-prof. | Larger retail | Larger prof. firms |
|---|---|---|---|---|---|---|---|---|---|---|
| **Number of employees** | | | | | | | | | | |
| 1 | 1.91 | 2.21 | 1.85 | 2.33 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 |
| 2 | 0.43 | 0.14 | 0.38 | 0.00 | 2.35 | 2.76 | 0.00 | 2.75 | 0.30 | 0.03 |
| 3-4 | 0.22 | 0.12 | 0.29 | 0.13 | 0.64 | 0.25 | 4.68 | 0.00 | 4.57 | 0.59 |
| 5-24 | 0.14 | 0.06 | 0.76 | 0.02 | 1.31 | 0.06 | 2.87 | 0.88 | 1.34 | 15.41 |
| **Industry** | | | | | | | | | | |
| Manufacturing and related trades | 0.00 | 0.00 | 5.90 | 0.00 | 0.00 | 0.34 | 3.43 | 1.54 | 0.00 | 1.34 |
| Construction | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 6.67 | 0.00 | 4.79 |
| Wholesale and retail | 0.00 | 0.00 | 0.00 | 1.75 | 0.00 | 1.45 | 0.00 | 0.93 | 1.76 | 0.35 |
| Transportation and storage | 0.58 | 0.13 | 1.24 | 0.00 | 0.00 | 0.12 | 5.65 | 2.04 | 0.00 | 1.89 |
| Accommodation and food service | 0.99 | 0.78 | 1.80 | 0.07 | 0.00 | 0.97 | 3.50 | 1.61 | 0.00 | 1.92 |
| Various professional acts. | 8.24 | 0.00 | 0.00 | 0.00 | 8.74 | 0.21 | 1.42 | 0.00 | 0.00 | 4.06 |
| Other service | 0.00 | 8.89 | 0.00 | 0.00 | 0.00 | 0.43 | 1.00 | 1.33 | 0.00 | 0.23 |
| **Region** | | | | | | | | | | |
| Greater Cairo | 1.00 | 1.13 | 0.57 | 0.66 | 1.19 | 0.72 | 1.98 | 0.80 | 1.64 | 1.78 |
| Alexandria and Suez Canal | 1.31 | 1.29 | 0.76 | 0.89 | 1.85 | 0.82 | 0.67 | 1.35 | 0.75 | 1.43 |
| Lower Egypt | 0.67 | 0.98 | 1.66 | 1.20 | 0.68 | 1.09 | 0.65 | 1.06 | 0.54 | 0.51 |

| | Smaller prof. firms | Smaller service firms | Smaller manuf. and rest. | One-person retail | Two person prof. firms | Two-worker retail | 3+ worker non-retail | Larger non-prof. | Larger retail | Larger prof. firms |
|---|---|---|---|---|---|---|---|---|---|---|
| Upper Egypt | 1.46 | 0.72 | 0.53 | 1.18 | 0.91 | 1.31 | 0.38 | 1.00 | 1.04 | 0.58 |
| **Legal status of firm/establishment** | | | | | | | | | | |
| Collective Ownership | 0.60 | 0.45 | 0.41 | 0.10 | 0.19 | 1.87 | 1.01 | 0.11 | 1.89 | 12.41 |
| Sole Proprietorship | 1.03 | 1.04 | 1.05 | 1.07 | 1.06 | 0.93 | 1.00 | 1.07 | 0.93 | 0.11 |
| **Age of firm (years)** | | | | | | | | | | |
| 0 | 0.86 | 0.85 | 0.79 | 1.14 | 0.70 | 1.36 | 0.73 | 0.88 | 0.75 | 0.86 |
| 1 | 0.95 | 1.07 | 1.06 | 1.09 | 0.67 | 1.12 | 0.95 | 0.85 | 0.92 | 0.84 |
| 2 | 0.94 | 1.03 | 1.08 | 1.06 | 0.88 | 0.87 | 0.91 | 1.13 | 1.03 | 0.72 |
| 3 | 1.15 | 1.02 | 0.77 | 1.23 | 0.68 | 0.81 | 1.14 | 1.02 | 0.76 | 0.56 |
| 4 | 0.87 | 1.27 | 0.82 | 0.99 | 0.87 | 0.84 | 1.28 | 1.03 | 1.13 | 0.87 |
| 5 | 0.94 | 0.86 | 0.71 | 1.16 | 0.91 | 1.08 | 1.00 | 0.90 | 0.93 | 0.92 |
| 6 | 0.84 | 0.97 | 1.01 | 0.89 | 1.00 | 1.05 | 1.04 | 1.30 | 0.89 | 0.77 |
| 7 | 1.33 | 0.79 | 0.85 | 1.37 | 0.84 | 0.77 | 0.93 | 0.97 | 0.64 | 1.11 |
| 8-17 | 1.21 | 0.82 | 1.02 | 0.78 | 1.58 | 0.95 | 1.00 | 1.08 | 1.33 | 1.31 |
| 18-27 | 0.83 | 1.22 | 1.47 | 0.56 | 1.72 | 0.98 | 1.24 | 1.10 | 1.23 | 1.21 |
| 28+ | 0.69 | 1.57 | 1.73 | 0.72 | 0.85 | 0.99 | 1.18 | 0.81 | 0.96 | 2.06 |
| Don't know | 0.00 | 0.00 | 0.00 | 0.00 | 1.15 | 0.00 | 0.26 | 0.29 | 0.00 | 35.66 |
| **Has unpaid workers** | | | | | | | | | | |

|  | Smaller prof. firms | Smaller service firms | Smaller manuf. and rest. | One-person retail | Two person prof. firms | Two-worker retail | 3+ worker non-retail | Larger non-prof. | Larger retail | Larger prof. firms |
|---|---|---|---|---|---|---|---|---|---|---|
| No | 0.00 | 0.17 | 0.23 | 0.00 | 1.10 | 0.00 | 0.85 | 0.66 | 1.82 | 25.56 |
| Yes | 1.01 | 1.00 | 1.00 | 1.01 | 1.00 | 1.01 | 1.00 | 1.00 | 1.00 | 0.87 |
| **Has wage workers** | | | | | | | | | | |
| No | 1.59 | 1.59 | 1.53 | 1.59 | 0.07 | 1.59 | 0.04 | 0.00 | 0.10 | 0.11 |
| Yes | 0.00 | 0.00 | 0.10 | 0.00 | 2.57 | 0.00 | 2.62 | 2.69 | 2.51 | 2.50 |
| **Ln capital to labor ratio** | 0.48 | -0.12 | 0.27 | 0.19 | 0.18 | -0.15 | -0.05 | -0.18 | -0.35 | -0.12 |
| **Ln value added per worker** | -0.04 | -0.03 | 0.16 | 0.01 | 0.16 | -0.57 | 0.36 | 0.12 | 0.38 | -0.08 |
| **Percentage of workers female** | -0.35 | -0.29 | -0.32 | 0.00 | 0.78 | 0.41 | -0.25 | -0.08 | -0.11 | 0.11 |
| **N** | 2325 | 2773 | 2230 | 2946 | 3693 | 2527 | 4055 | 3686 | 1548 | 2167 |
| **Percentage of Informal Firms in the Cluster** | 4.8 | 7.3 | 6.3 | 27.2 | 3.9 | 16.6 | 9.9 | 13.6 | 8.3 | 2.3 |

Source: Authors' calculations based on EcC 2012/13

Notes: Categorical variables are relative to the overall mean percentage for that category continuous variables are standardized