

# What Determines Firms' Decisions to Formalize?

## Evidence from Rural Indonesia\*

Neil McCulloch,<sup>\*</sup> Günther G. Schulze<sup>^</sup> and Janina Voss<sup>▼</sup>

Version, 20 April 2010

### Abstract

In this paper we analyze the decision of small and micro firms to formalize, i.e. to obtain business and other licenses in rural Indonesia. We use the rural investment climate survey (RICS) that comprises around 2500 non-farm rural enterprises in six Indonesian districts and analyze the effect of formalization on tax payments, corruption, access to credit and revenue, taking into account the endogeneity of the formalisation decision to such benefits and costs. We show that, contrary to some of the literature, formalisation reduces tax and corruption payments. However, the benefits of formalisation, and therefore the likelihood of being formal, also depend on characteristics such as firm size, as well as the education and ethnicity of the owner.

JEL classification: O17, O18

Keywords: Formalization, rural development, rural investment climate

---

\* We are grateful to Krisztina Kis-Katos for very helpful comments.

<sup>\*</sup> Institute of Development Studies at the University of Sussex, Brighton BN1 9RE, U.K.

<sup>^</sup> Corresponding author: University of Freiburg, Dept. of International Economic Policy, 79085 Freiburg, Germany, [guenther.schulze@vwl.uni-freiburg.de](mailto:guenther.schulze@vwl.uni-freiburg.de).

<sup>▼</sup> University of Freiburg, Dept. of International Economic Policy, 79085 Freiburg, Germany.

# 1. Introduction

Most developing countries are characterized by large informal sectors.<sup>1</sup> These large informal sectors are potentially detrimental to economic development as they escape government taxation and regulation. The former means that governments have fewer resources for financing development, whilst the latter can lead to welfare losses due to the lack of minimum safety and labor standards in the informal sector. Moreover, a larger share of informal firms in an economy could reduce growth because of their unequal access to formal credit institutions, government contracts and other official or private institutions. It would thus seem to be in the interest of a benevolent government to reduce the size of the informal economy. Yet, a large informal economy may be created by misguided government policies. Auriol and Walter (2004) argue that myopic governments create large informal sectors through deliberately creating high entry barriers, as they find the resulting small numbers of large tax payers easier to tax. Similarly, Loayza (1997) finds that large informal sectors are a consequence of excessive taxation, overregulation and weak government institutions resulting in low growth rates.

In order for policymakers to be able to design appropriate policies to reduce the size of the informal sector, it is necessary to understand why some firms decide to go formal, whilst others decide to avoid being registered and licensed. This paper examines the determinants of formalization in Indonesia. In the presence of weak government institutions firms decide to become formal if the benefits from formalization exceed the costs. Therefore we analyze the effects of formalization on tax payments, corruption, access to credit and to government contracts as well as on revenues. We allow for firm heterogeneity and identify firm characteristics that affect the gain from formalization.

Even though the issue at hand is of substantial policy relevance (Perry et al. 2007), surprisingly little attention has been devoted to it so far. Levenson and Maloney (1998) regard formality as an input into a firm's production function. In exchange for the benefits of formalization – enforceability of contracts, credibility, access to capital and access to public risk-pooling mechanisms – a firm has to pay initial entry costs and periodic 'taxes', such as reporting requirements or insurance payments. They assume that emerging firms face uncertainty over their managerial ability and production costs, which resolves over time. Given that entry costs are assumed to be identical, the firms that become successful over time will find it profitable to become formal and thus the older and larger firms will have a higher probability of being formal. They confirm this hypothesis by using the Mexican

---

<sup>1</sup> See Enste and Schneider (2000) and Schneider (2005) for reviews of the size, causes and consequences of shadow economies around the world.

National Micro Enterprise Survey 1992. Jäckle and Lee (2006) confirm Levenson and Maloney's hypothesis in a panel approach that captures firm dynamics using the Peruvian Living Standard Measurement Survey for 1994, 1997, and 2000. Insightful as they are, they focus on only two firm characteristics, yet other characteristics such as gender, ethnicity and religion of the manager may matter if formalization is not entirely voluntary and certain groups are more subject to scrutiny than others. Moreover, if formality is considered an input in the firm's production function, it clearly affects firm revenue and longevity (considering the high failure rates of newly emerging firms in developing countries). Thus there is an endogeneity problem with respect to formal status that the authors do not address. Older and more successful firms may be more likely to go formal but formal firms may also become more successful and long-lived.

Fajnzylber et al. (2006) solve the endogeneity problem in a quasi-experimental regression discontinuity approach. The introduction of the Brazilian SIMPLES program, which aimed at reducing registration costs and tax rates for small firms, lead to higher formalization rates. Comparing the firms established before and after the introduction of the program they show that formalized firms do better in terms of revenue, investment or employment. McKenzie and Sakho (2009), by contrast, use an instrumental variable *approach* to address the endogeneity problem. They analyze Bolivian micro- and small enterprises in urban areas that are within 10 kilometers of the city's tax office and use the distance to the tax office as an instrument for formalization. They show that formalization increases profits, but that this effect is limited to medium sized firms (2-5 workers) while very small and larger firms lose. Formality reduces tax payments but does neither improve access to credit nor reduce perceived corruption. While the authors take firm heterogeneity into account, they only look at firm size as a potential source for it. Moreover it remains unclear whether distance to the tax office is a good instrument, as it could also capture distance to market or other institutions if tax offices are centrally located.

In this paper we use the Indonesian Rural Investment Climate Survey (RICS),<sup>2</sup> which covers 2500 small and micro non-farm enterprises and their communities in rural Indonesia. Indonesia is particularly suited for such analysis because its diversity allows studying systematically the role of ethnicity, religion and community characteristics. We adopt a cost-benefit approach and investigate the effect of formalization on a number of profit components using a wide array of firm and community characteristics. We employ an IV approach to account for the endogeneity of firm status (formal versus informal) and we interact the formalization variable with a number of household characteristics to explore the different effects of formalization on heterogeneous firms. We find that formalization reduces tax

---

<sup>2</sup> The survey was conducted by the World Bank and one of a series of RICS in other countries. See [www.worldbank.org/id/rica](http://www.worldbank.org/id/rica) for details.

payments and the incidence of corruption as well as their magnitude, but that this effect is non-uniform and depends, inter alia, on firm size and the gender of the manager. We find that larger firms are more likely to be formal, as are firms with non-local managers or those that are Chinese or of a non-dominant ethnic group.

The remainder of the paper is structured as follows. Section 2 presents the RICS dataset, explains licensing practices in Indonesia which affect our definition of formality. Section 3 introduces the costs and benefits of formalization. Section 4 lays out the estimation strategy, section 5 lays out our results after which the final section concludes.

## 2. Data and Definitions

### The Indonesian Rural Investment Climate Survey

The Indonesian Rural Investment Climate Survey (RICS) is a quantitative survey of households, non-farm enterprises (NFE) and communities that attempts to measure the business environment faced by rural firms. It was conducted in early 2006 in six districts (*Kabupaten*) purposively chosen to reflect the distinct geographical environments faced by local enterprises in different parts of Indonesia (World Bank, 2006).<sup>3 4</sup>

The sampling within these districts was done in three stages.<sup>5</sup> Based on a 2006 village census, 30 villages were selected with probabilities of selection proportional to their number of businesses. In a second step census blocks were constructed in each village from which in the third stage a sample of 'household enterprises' (located in the place of residence), 'standalone enterprises' (not located in the place of residence) and 'pure households' (i.e. with no household enterprise) was drawn. This sampling strategy runs the risk of failing to sample the small number of large firms at the district level who might not be located in the sampled census blocks. To try and redress this potential bias, an additional sample was taken from the list of enterprises registered with the district statistical office<sup>6</sup>, on the assumption that these 'listed firms' tend to be somewhat larger. This stratified sampling

---

<sup>3</sup> The survey locations were: Labuhan Batu, North Sumatra – a plantation area; Kutai, East Kalimantan – an area rich in mineral resources; Barru, South Sulawesi – a forest fringe area; Malang, East Java – a rich agricultural area; Badung, Bali – a semi-urban agglomeration area and; Sumbawa, NTB – a dryland area.

<sup>4</sup> There are over 400 districts in Indonesia. Because the focus of the survey was the exploration of the investment climate at the district level, the RICS takes a large representative sample in each district to facilitate cross-district comparisons. The survey is not, therefore, nationally representative (which would have required extensive but sparse sampling), but the choice of regions attempts to encompass the main types of economic geography in the country.

<sup>5</sup> A more detailed review of the sampling strategy, the data collection and data processing is provided by LPEM-FEUI in the Data User Guide (2006a) and the Field Survey Report (2006b). A summary of the sampling procedure and the main results of the RICS can be found in Schulze and Quadros (2007).

<sup>6</sup> Registration with the statistical office does not necessarily imply that those firms are operating in the formal sector (Schulze and Quadros 2007).

design requires the use of sampling weights in the empirical analysis in order to obtain unbiased estimation results that are representative of the population of NFEs at the district level (World Bank, 2008).

The survey consisted of three linked questionnaires: (1) the household questionnaire, collecting information on household economic activity and consumption patterns, (2) the enterprise questionnaire, that aims at investigating the rural investment climate for NFEs by looking at their costs, revenues and constraints, and (3) the community questionnaire, that provides information on local infrastructure and governance. From these we constructed an enterprise dataset which contains 2137 micro enterprises (1-4 employees), 263 small enterprises (5-19 employees), 51 medium-sized enterprises (20-99 employees), and nine large enterprises (> 100 employees)<sup>7</sup>. Since our analysis focuses on formalization, we exclude the 60 medium and large firms from our sample.

Table 2 shows the characteristics of the firms and their managers in our sample. The number of firms is distributed fairly equally over the six *Kabupatens* with 400 to 420 firms in each district. Although the sample focuses on rural areas and does not include any major metropolitan centers, the statistical bureau still classifies villages as either 'urban' or 'rural' based on a set of village characteristics (share of agriculture, population density, and several services and facilities associated with urban living – see World Bank (2006)). According to this definition about half of the firms in the sample are located in rural areas, although this share is much higher in Labuhan Batu and much lower in the peri-urban area of Badung. Typically more than half of the sampled firms in each area operate in the trading sector, around 35 % in the service sector, with the remainder in the manufacturing sector.<sup>8</sup> Of the 2461 firms in the sample, almost 90 percent are 'micro' firms; as a result the mean number of employees is only 2.6. The mean enterprise age is nine years. However, more than 50 percent of firms in the sample have been operating for six years or less and 12 percent have existed for only one year.

---

<sup>7</sup> These employment groups follow the BPS size classification of enterprises. The number of employees also includes the owner or manager of the enterprise as well as unpaid laborers.

<sup>8</sup> The three sectors comprise the following activities: trade, including wholesale and retail trade; the service sector, comprising of repair shops, hotels, food and beverages, transportation, finance, real estate, health and public services; manufacturing activities, defined as mining and excavation, manufacturing including the processing of agricultural goods, electricity, gas and water provision and construction.

For about 90 percent of firms the owner is also the primary manager of the business.<sup>9</sup> The majority of firms surveyed are managed by men (63 %) with a mean age of 41 years. Education is at a very low level for almost half of firm managers: 13 percent state that they did not complete primary school, and a further 30 percent have primary school as their highest level of schooling. Only eight percent of firm managers report that they have a university level education.

Ethnicity is highly diversified in the sample as is true for Indonesia as a whole. However, in each Kabupaten one local ethnicity dominates among firm managers. Only in Kutai and Labuan Batu are there more managers from non-indigenous ethnicities than indigenous managers. Chinese Indonesians play a particularly important role in the Indonesian economy and so are specified separately in Table 2. Although they only account for 1.7% of enterprise managers throughout the six Kabupaten, they represent a much higher share of small firm managers than micro firm managers. Compared to ethnicity, the religion of the manager is quite homogeneous. The share of Islamic managers is more than 90 percent in all Kabupaten, except Badung in Bali, where the majority of managers are Hindus (64%), with only 28 percent being Moslems.

### **Business Licensing in Indonesia: Defining Formality**

Business licensing in Indonesia is characterized by complex, time-consuming, and costly procedures. According to the World Bank's 'Doing Business Report', Indonesia ranks among the worst-performing countries in the world in terms of 'starting a business' and 'dealing with licenses' (World Bank, 2008).<sup>10</sup> The far-reaching decentralization undertaken in 2001 made processes even less transparent. Indonesia has rapidly transformed from decades of authoritarian rule to one of the most decentralized countries in the world with 33 provinces, subdivided into 459 districts (Kabupaten and Kota). These districts have become responsible for a large share of economic policies, among them business licensing (von Luebke 2006: 2). Hence licensing practices differ throughout Indonesia. Additionally there are still licenses being issued at the provincial or central government level, which further complicates the process by requiring cooperation between the different administrative levels (KPPOD 2008: 31). Local governments tend to see licensing services more as a means of generating local revenues than for regulating markets or collecting information: since decentralization about

---

<sup>9</sup> In the subsequent analysis, all variables referring to manager characteristics are taking the value of the owner if he is managing the business himself and the primary manager otherwise.

<sup>10</sup> In 2006, when RICS was conducted, it took 12 procedures, 151 days at a cost of 101.7% of GNI per capita to officially start a new business in Indonesia. For dealing with all required licenses to build a warehouse in Jakarta, a SME company with 20 employees has to take into account 10 procedures, 224 days and 370.5% of income per capita (World Bank, 2008).

1600 new regulations and local government licenses have been introduced, at least 30% of which are considered to distort economic activity (ibid: 14).

This complexity also makes the identification of formal and informal firms more difficult. In theory business licensing comes as a 'package' in Indonesia. To be fully registered an enterprise must complete a number of administrative processes at the national as well as the local level (TAF 2007: 8f). First, to make sure a new enterprise fulfils all requirements for a formal business or company, it needs a deed of establishment from a notary and a tax identification number (NPWP) from the central government. Second, the firm needs physical permits such as a building permit (IMB) and a nuisance permit (HO). Third, a sectoral license has to be obtained to allow operation in one of the major sectors. The main sectoral licenses are the trade permit (SIUP) and the industrial registration/permit (TDI). Only after those requirements have been fulfilled can a firm process the business registration (TDP) at the local level. Depending on which products or activities a firm deals with, it may additionally need to obtain product- and activity-specific licenses.<sup>11</sup>

The RICS data contains information on the main components of this 'package': building permit (IMB), industrial permit (TDI), trade permit (SIUP), enterprise registration (TDP), and tax identification number (NPWP). In terms of the formalization procedure explained above, a firm would be 'fully registered' if it has a tax number, a building permit (if it operates in a building separate from a household residence), one of the sectoral licenses (trade or industrial permit), and the enterprise registration. According to this definition, only 2% or 48 firms in the sample are fully formalized. Leaving out the tax registration number from the definition – as it is the only license not issued at the district level – results in 2.9% of firms being 'fully registered with the local government'.

In practice, there is considerable confusion about which licenses a business really needs. The trade permit for example, although officially designed for companies engaged in trading activities, is the most common license held even among industrial firms (KPPOD 2008: 32). In the RICS, the trade permit is almost as common among manufacturing firms (13 percent hold it) as the industrial permit (14 percent). Similarly, the building license is legally required only for firms operating in a separate building, but is often issued to firms that operate their business from their dwelling (11 percent) and even firms that do not operate from a fixed location (4 percent). Indeed the belief that certain licenses are not needed may be one reason why such a small number of firms are 'fully formalized'.<sup>12</sup>

---

<sup>11</sup> For a limited liability company these steps are even more numerous. For a detailed description of all processes involved see World Bank (2008: 53ff).

<sup>12</sup> When asked about the reason for not holding a certain permit, only three to five percent of firms consider the respective license to be too expensive and merely one percent criticizes complicated procedures. The main stated reason for firms not obtaining a license is because they think it is not required for them. This is the case for

This paper will therefore rely on a more practical definition of formality: a firm is classified as being formal if it has at least one of the local licenses (IMB, SIUP, TDI or TDP), which is the case for 23.4% or 541 enterprises. Since the policies on local licenses are being determined at the Kabupaten level, there are significant differences across the six districts in the share of formal firms. The lowest share can be observed in Malang (7.4%), followed by Labuhan Batu (17.9%); Badung, Sumbawa and Kutai have shares between 23 and 29 percent, whilst in Barru 42 percent of firms hold at least one of the four main local licenses.

### **3. Costs and Benefits of Formality**

A firm's decision to formalize will depend on "firms perceiving that it is in their self-interest"

(Kenyon 2007: 5). We therefore anticipate that firms will choose to formalize when the expected profit of operating formally is larger than the expected profit of operating informally.<sup>13</sup> However, the formality status of the firm will also influence the revenue and cost functions that it faces.

#### **The costs of operating formally**

Each firm that wants to 'legally' open a business or register an already existing one faces an initial entry cost in terms of time and fees (Djankov, 2002). We define entry costs by the mean cost and time it takes to register, where registration is defined as holding at least one of the four local licenses (SIUP, TDI, TDP or IMB).

After registering, a formal firm has to abide by government regulations that are often complex. Besides the direct monetary costs in terms of taxes (or bribes) to government officials, the firm also has to bear the indirect costs of time spent submitting government documents or fulfilling product or labor standards (cf. Djankov et al. 2003: 66f; Ishengoma and Kappel 2006: 16f).

An informal firm, by contrast, saves on these costs. At the same time, it is more dependent on the 'goodwill' of government officials and the police. This may make bribe payments more unpredictable and potentially higher for informal firms (cf. *ibid*: 18, Djankov et al. 2003: 71).

---

52% of firms that operate from a fixed location but do not have an IMB, 49 % of trading firms without SIUP, and 48% of manufacturing firms without TDI. Even the enterprise registration certificate, TDP, which is the only license that is needed by all firms no matter where they operate or in which sector, reveals the same picture. Only 7.5% of firms hold an enterprise registration at all – of the remaining 2011 firms, more than 51 % state that TDP it is not required.

<sup>13</sup> If firms are not risk neutral, their decision will also take into account the relative uncertainty of the profit streams under formality and informality, but we do not explore this here.



Anecdotal evidence from a qualitative evaluation of One Stop Shops for Business Licensing in Indonesia suggests that an important reason firms get licenses is to reduce unofficial payments to business inspectors (LabSosio 2008: 41). Informal firms may respond to such costs by staying small or changing location in order to escape detection and harassment by the police or government officials (cf. *ibid*; Djankov et al. 2003: 70). For the same reason such firms may avoid investment in conspicuous fixed assets or technology.

In this paper, we calculate the costs of operating formally as the value of levies a firm has to pay, subdivided into total taxes and 'other levies'. Total taxes include the taxes a firm pays to the central, provincial and local governments; 'other levies' comprises payments to security officials, thugs, and sub-district or village officials which we use as a proxy for unofficial payments and bribes.

### **The benefits of operating formally**

There are also several potential benefits of operating formally. These can include: the ability to access formal credit markets; access to contract governments; access to legal dispute resolution mechanisms, such as courts; and the ability to enter into collaborative agreements with large firms. None of these benefits are typically available to informal firms.

We calculate the benefits of formality using total firm revenue, access to government contracts and access to credit.<sup>14</sup> We did not examine the benefits associated with using the courts, because, of the 2.8% of firms in our sample that had any kind of payment dispute, none had used the courts to resolve the problem.

### **Costs, benefits and firm characteristics**

Table 2 shows the average value of these costs and benefits for formal and informal firms separately. At face value it would appear that the costs, in terms of taxes and levies, are much higher for formal rather than informal firms, but that the benefits – sales to government and better access to credit – are also commensurately larger.

However, these differences could be due to other firm or manager characteristics rather than formality itself. In particular, larger average costs and benefits may simply be because formal firms tend to be larger and older (Levenson and Maloney, 1998). Table 3 shows that this is indeed the case.

---

<sup>14</sup> Access to credit is measured by a dummy variable that equals one for all firms that have had a loan approved in the last twelve months. The counterfactual group only contains those firms who state that they currently need additional funding for their enterprise. According to this definition, 82 percent of firms have constrained access to credit in the sense that they need additional financial support, but did not receive credit.

It is therefore not possible to draw conclusions about the impact of formality on firm revenues and costs from the stylized facts above. Rather we need to estimate the causal effect of licensing on firm revenues and costs taking into account the endogeneity of the formality decision to the net benefits of formalization.

#### 4. Approach to Estimation

Our basic equation to estimate the impact of formality on potential costs and benefits is the following:

$$Y = \beta_0 + \beta_1 L + \beta_2 M + \beta_3 F + \beta_4 C + \varepsilon \quad (1)$$

where  $Y$  denotes the respective cost or benefit, i.e. taxes, ‘other levies’, firm revenue, sales made to the government, and access to credit.  $L$  is the formality dummy, and  $M$ ,  $F$ , and  $C$  are controls for manager, firm and location (community) characteristics.<sup>15</sup>

Because the formality decision is endogenous, we instrument it and estimate using 2SLS. The instrument we use is the community average level of licensing.<sup>16</sup> As an aggregate measure, village licensing averages are correlated with the formality status of individual firms, but should have little influence on the costs and benefits obtained by individual firms, other than through its influence upon licensing. However, it could be argued that a village with a high share of licensed firms can also be expected to perform better in other fields of economic policy. If village averages of licensing are correlated with a more favorable business climate and this reduces firm costs (or increases their benefits), then our second stage estimates could be biased. To address this problem each regression will control for village averages of the respective cost or benefit analyzed.<sup>17</sup>

In addition to estimating the overall impact of formality on costs and benefits, we are interested in how the impact of formality differs across firms with different characteristics. We therefore enter our first-stage estimate of licensing,  $\hat{L}$ , both directly, in order to estimate the overall effect of licensing, as well as in the form of interaction terms with the firm, manager and location characteristics i.e.

---

<sup>15</sup> Appendix Table A1 provides the list of variables used in the analysis; Table A2 provides descriptive statistics.

<sup>16</sup> The approach of using village or city averages as an instrument is quite common – see Wößmann and West (2006) or Dollar et al (2005). We also explored a range of other community characteristics as instruments (frequency of village meetings, different measures for conflict, local business organizations and education of the village head) but these instruments did not pass our tests.

<sup>17</sup> If there are only few firms in each village, endogeneity of village averages with the firm-specific data could be a potential problem. If too highly correlated with the individual values, the village averages would take explanatory power from the other independent variables and again provide inconsistent results. A comparison of results with and without the averages showed that estimates of the other covariates were robust to the inclusion of village averages.

$$Y = \beta_0 + \beta_1 \hat{L} + \beta_2 \hat{L} * M + \beta_3 \hat{L} * F + \beta_4 \hat{L} * C + \beta_5 M + \beta_6 F + \beta_7 C + \varepsilon \quad (2)$$

Our approach is therefore very similar to McKenzie and Sakho (2009) who show that the effect of formality on firm profits varies with the number of employees. We use a more comprehensive list of owner, firm and location characteristics that might cause heterogeneous treatment effects on a broader range of potential costs and benefits.

## 5. Results

### Constructing the instrument

The results of the first stage estimates are summarized in Table 4. The firm, manager and location variables used to predict the fitted value of licensing will also be used to estimate the costs and benefits in the second stage.<sup>18</sup> The Sargan test of overidentification requires the application of at least one more instrument than endogenous variables. To perform this test, some of the weaker instruments mentioned above (frequency of village meetings, registration time and cost) were additionally included in the regression. The null hypothesis of instrument validity – denoting that all instruments are uncorrelated with the error term and thus have correctly been excluded – could not be rejected for either combination. However, a redundancy test shows that the weaker instruments do not improve the efficiency of estimates. Village averages of licensing are therefore applied as the only instrument in the following. The high F-statistic of the excluded instrument (between 71 and 126, depending on the dependent variable in the second stage) indicates that there is no problem of weak identification.

### The impact of formality on costs and benefits

We now summarize the results from estimating the impact of formality on official and unofficial levies as well as on revenues, access to government contracts, and access to credit. In each case we compare our 2SLS results with OLS.<sup>19</sup> For testing heterogeneity of

---

<sup>18</sup> In order to obtain consistent estimates the same control variables should be included in both stages (Wooldridge 2002). To reduce the endogeneity with licensing, actual values of size and enterprise age were replaced by dummies. For the number of employees the dummy equals one for firms with more than five employees. The dummies for total fixed assets, sales, and enterprise age equal one if the firm belongs to the group above the median.

<sup>19</sup> Since costs and benefits are being determined simultaneously at the firm-level, one could also estimate a system of equations rather than each equation individually. A seemingly unrelated regression model (SUR) that takes into account possible correlations of the error term might produce more efficient results. However, a SUR setup does not allow for the use of sampling weights that are applied here to balance the stratified sampling design. Even though they might be less efficient, individual OLS estimates of the same cost and benefit equations outside SUR will still be unbiased and consistent (Greene, 2002). Sampling weights on the other hand will correct for sample stratification bias. Applying sampling weights will thus be preferred over the SUR setup in the following analysis.

treatment effects, interaction terms of the fitted values for licensing and the various control variables were employed.

Each cost and benefit regression will control for the same firm, manager and location characteristics. The manager characteristics are highest level of education, gender, age, ethnicity, a dummy for whether or not the manager is Muslim and a dummy that denotes whether the manager lives in the same village as the enterprise. The ethnicity dummy equals one if the manager belongs to the local indigenous group. The religion dummy indicates whether a manager belongs to the majority religion, Islam.<sup>20</sup> A dummy for managers of Chinese descent is included separately in the regression.

---

<sup>20</sup> Balinese managers are excluded from this variable, since the majority religion in Bali is Hinduism. Hindu managers in Bali are not included as a dummy separately, because, although overall only 64 percent of managers in Bali are Hindu, almost all (95 percent) of the Balinese managers are Hindu. Thus this majority religion indicator corresponds with the ethnicity variable.

The regressions control for firm size in terms of the number of employees, log total sales, and log total value of fixed assets.<sup>21</sup> Further firm-level variables included are the number of years the firm has been in operation, and the sector of operation (manufacturing, service or trade<sup>22</sup>). As noted above, we also control for the how the general investment climate might influence the respective cost or benefit, by including the village averages of the relevant costs and benefits. We also include a dummy for whether villages are classified as rural or urban as well as district dummies. The estimates on access to credit also include a dummy for whether or not there is a bank in the village.

### **Taxes**

The results of estimating the impact of formality on the log of total tax payments are shown in Table 5. OLS estimation that does not control for the endogeneity of licensing (Column 1) reveals a positive and significant impact of licensing on tax payments. The 2SLS results (Column 2), however, suggest that OLS significantly overestimates the impact of licensing. The IV estimates show that formality is on average associated with a significant decrease in firms' total tax payments. The bias in the OLS result has exactly the sign we expect. If the underlying effect of licensing is to reduce tax payments, then firms with characteristics that make them likely to pay high taxes will want to get a license. OLS does not take into account this endogenous positive selection effect and therefore significantly underestimates the tax reducing effect of obtaining a license.

The effect of licensing on taxes varies with the level of firm sales. The results in Column 3 show that licensing reduces tax payments for firms in the lowest sales quintile the most, while the effect is smaller for the third, fourth and fifth sales quintiles.<sup>23</sup> Column 4 also shows that the licensing effect is significantly smaller for firms with female managers. Similarly, for rural firms, the benefits from licensing through tax reduction are more than double those of urban firms (Column 5).

Aside from licensing, several other characteristics influence the amount of taxes paid. Better educated and older owners tend to pay more tax (probably because these are also

---

<sup>21</sup> Total fixed asset value includes buildings, land, equipment and machinery, furniture, storage facilities, and vehicles. Although number of employees, log total sales, and log of total fixed asset value are different measures for firm size, no additional multicollinearity could be detected when entering all three variables at the same time. Since the information given by each of them is slightly different, including them all was preferred to using a compound index of firm size.

<sup>22</sup> Trade is the excluded category.

<sup>23</sup> Interaction with sales quintiles rather than simply with the value of firm sales was used to alleviate multicollinearity.

correlated with firm performance). Female owners, by contrast, pay less, as do Muslim owners. Again this may be due to correlation between these characteristics and firm performance, but could also be evidence for discrimination in favor of these groups. Unsurprisingly, large firms pay significantly more tax, as do firms in the manufacturing sector. And the general investment climate also appears to matter, with firms that are situated in villages with high average tax payments paying more.

These other factors aside, however, it is interesting to note that, on average, formalization reduces tax payments rather than increasing them. Our results thus contradict conventional wisdom that informality keeps firms under the tax administration's radar screen. Small and rural enterprises benefit from a substantial reduction in tax payments if they are licensed. A possible explanation for this result may be found in the tax collection process in Indonesia. In his case study on local governance in six Kabupatens, von Luebke (2006) found that tax collection practices are inefficient and to a large extent based on rough estimates or personal negotiations. The results above suggest that licensing may increase the 'bargaining power' of very small enterprises in such negotiations and thus help to reduce the amount of taxes paid by these firms.

### **Other Levies**

Our questionnaire also asked firms about the payment of 'other levies' aside from taxes. These include payments to security officials (i.e. the police), *preman* (organized thugs that extort money), as well as other 'informal' payments requested by sub-district or village officials. In short, 'other levies' are (generally small scale) bribes and extortion payments.

In estimating the impact of formality on other levies, we need to take account of the low incidence of such payments in our dataset. Only 27 percent of firms report positive values for 'other levies', so there is a possibility that estimation may be subject to sample selection bias (Heckman, 1979). To account for this possibility, we tested and, to the extent possible, corrected for sample selection bias using a Heckman selection model.<sup>24</sup>

Table 6 shows the results using the different approaches. Column 1 displays the simple OLS results, with no instrumentation of licensing and no correction for sample selection. As before, we see positive but insignificant association between licensing and other levies. Column 2 shows the 2SLS results in which licensing has been instrumented, but still with no correction for sample selection. The sign of the licensing variable is now negative and

---

<sup>24</sup> Note that the same problem potentially arises with tax payments, since only 47 percent of firms in the sample pay any taxes at all. A variety of exclusion restrictions were tried for the case of tax payments. However, with the set of potential instruments available in our data, the null hypothesis of no selection bias, could not be rejected.

strongly significant, suggesting that, other things equal, having a license reduces the amount of other levies paid. Several other variables also appear to significantly influence the amount of other levies: female owners appear to pay less, as do owners that live in the same village as the enterprise. By contrast Chinese owners pay more, as do larger firms (whether in terms of employees or sales) and firms in villages with generally high payments.

Columns 3 and 4 of Table 6 show the results when we correct for sample selection bias with a Heckman selection model using the firm's number of employees as the exclusion restriction. The number of employees is the most 'visible' of all firm size measures. An enterprise with a larger number of employees will be more 'exposed' to security officers, thugs or other officials asking for unofficial payments and less able to change location to escape such harassment. However, the actual amount of other levies paid may depend more on firm revenue or profit than on the number of employees. Similarly, Columns 5 and 6 report the results of applying the manager's gender as the exclusion restriction. Predominantly male officers or thugs might be more reserved in approaching female managers for other levies, but this may not affect the amount paid if they are asked.

One of the advantages of the Heckman model is that it allows us to assess the influence of various characteristics on the probability of paying 'other levies' at all. Columns 4 and 6 show that female owners and owners that live in the same village as the enterprise are less likely to pay 'other levies', as are firms in the manufacturing sector. Larger firms are much more likely to pay such levies as are firms in villages with high levels of 'other levies'. Having a license also substantially reduces the probability of paying these levies, in addition to reducing the amount that is paid.

The estimates from the outcome stage of both Heckman selection models are very similar to those from the 2SLS with no correction for selection bias. Almost exactly the same variables are statistically significant and in most cases the values of the coefficients are similar. However, this may reflect the weakness of our exclusion restrictions, since there is significant collinearity between the inverse Mills ratio and the explanatory variables of the outcome stage regressions. Evidence from Monte Carlo simulations suggests that, where such collinearity exists, it may be more robust to simply estimate using OLS on the censored dataset (i.e. in our case the sample of firms that pay positive levies) (Puhani, 2000). This is shown in Table 7 (Column 1). Although this dramatically reduces the sample size, almost exactly the same variables as before are statistically significant. However, the strength of the impact of licensing on other levies increases when firms that do not pay are excluded, suggesting that selection bias may indeed be a problem.

Table 7 also explores the heterogeneous effects of licensing on other levies, using the censored sample estimation.<sup>25</sup> The benefit of being licensed (due to reduced payments of other levies) increases with firm size, whether measured by sales (Column 2) or fixed assets (Column 3).<sup>26</sup> Firms in the smallest size decile still benefit from licensing (although the absolute size of the benefit is small) but the absolute size of the benefit increases significantly with size. Other levies are different therefore from taxes, where the benefits of licensing accrued more strongly to smaller firms. In addition, the benefit of licensing is greater for urban rather than rural firms (Column 4).<sup>27</sup> The interactions with ethnicity and gender are not statistically significant (Columns 5 and 7) (although our Heckman selection models with interaction effects suggest that both characteristics significantly strengthen the impact of licensing on not paying 'other levies' at all). Licensing also appears to reduce 'other levies' for Muslim managers much less than for non-Muslim managers (Column 6)

---

<sup>25</sup> We also estimated Heckman selection models with interaction effects with broadly similar results which are available on request.

<sup>26</sup> Mean log sales of those that pay other levies is 10.8; mean log fixed assets is 7.8. Thus licensing significantly reduces the level of other levies paid.

<sup>27</sup> Note that this result, like all the results in Table 7, is conditional on paying some other levies. Our Heckman selection model with interaction effects suggests that being rural makes it much more likely that the firm pays no levies at all.



## **Firm Performance and Business Expansion**

As noted above, being licensed may also provide firms with access to new markets and customers, as well as potentially better access to finance. We therefore explore the impact of licensing on three variables, total sales, the share of sales to government, and access to credit. OLS estimates can be expected to overestimate the impact of formality, because formalization will contribute to firm growth and larger firms are more likely to formalize. Indeed measures of size and enterprise age are endogenous to all three dependent variables. To mitigate this endogeneity problem, size and enterprise age will enter the equations as dummies.

### **Total sales**

As expected, OLS estimates (Table 8, Column 1) show a gain to total firm sales from licensing of 24%, although this is not statistically significant. Comparing these results with 2SLS (Column 2) confirms the assumption of an upward bias in OLS estimates stated above, although again the result is not precisely estimated.

Interacting licensing with employment quartiles, however, shows that the gain to total sales from licensing depends on firm size. While firms in the lowest employment quartile experience a decrease in total sales from formality, firms located in the highest quartile will gain from formality with total sales more than doubling. This finding corresponds with the evidence from a qualitative survey of firms using licensing services in Indonesia that found that smaller enterprises feel that they are not able to take advantage of the opportunities opened up by formality to increase revenue, such as the facilitation of trading across a wider region or the use of licenses as a sign of legality and trustworthiness to establish new business relations (LabSosio 2008: 43ff).

Table 8 also suggests that businesses with Muslim owners are less likely to experience a sales benefit from licensing (column 4). This could reflect higher enforcement of licensing among non-Muslims. Similarly, the benefit of higher sales from having a license is much larger for older enterprises than for recently formed enterprises.

### **Access to government contracts**

For the case of access to government contracts one might again expect an upward bias in OLS estimates. Government institutions in Indonesia usually require licenses for businesses who want to interact with them (LabSosio 2008: 43). If this is the case, a higher share of sales made to governments will denote a higher probability that a firm is formal. However, on average, find no impact of licensing on the share of sales to governments either from our

OLS or 2SLS estimates (Table 9, Columns 1 and 2). However, a heterogeneous effect of the instrumented licensing variable occurs in interaction with firm size. The share of sales made to governments increases with fixed asset value (Column 4): licensed firms that belong to the lowest fixed asset quintile<sup>28</sup> sell an even lower share of output to governments than their informal counterparts. Conversely, firms in the highest fixed asset quintile have a share of sales to government which is 1.8 percentage points higher than their unlicensed counterparts. A similar effect is found with respect to total sales (Column 3) - compared to firms in the lowest sales quintile, enterprises in the third sales quintile benefit from licensing by an increase of 15.4 percent in sales made to governments. While licensing does not make any difference for firms operating in the service or trade sector, formal manufacturing enterprises experience a large gain in the share of government contracts of 37 percent (Column 5). Chinese owners are somewhat less likely to sell to government (Column 6). However, we treat these results with some caution since only 74 firms (3%) in the RICS sample report positive sales to governments – small and micro firms are typically not suppliers to local government.

## **Credit**

Access to credit has widely been considered as a central argument in favor of licensing. The regulations of the Bank of Indonesia require a firm to have a legal business status in the form of a TDP or a SIUP in order to apply for credit from a commercial bank (WB 2006: 51). In practice, banks in Indonesia do not seem to consider licensing as a central determinant for the approval of a loan request. Decisions are rather based on a survey of business feasibility that banks conduct themselves and the collateral or other securities a firm has to offer (ibid: 50, LabSosio 2008: 44ff). Alternative loan sources such as cooperatives or private lending are widely used by small firms, since they are more flexible and have lower requirements.

The results of a Probit regression on the access to credit are presented in Table 10. In accordance with the literature, licenses turn out not to be decisive for access to credit. However, the results show that for firms belonging to the smallest employment quartile (i.e. enterprises with only one employee) access to credit is improved by having a license, while the opposite holds for the largest firms (Column 3). The impact of licensing also appears to depend on the firm's sector (Column 4). While firms in the service and the manufacturing sector benefit from increased access to credit through licensing, no additional effect on access to credit can be found for formal firms operating in trade.

---

<sup>28</sup> 87 percent of firms in the lowest quintile do not report any fixed assets at all.

An important issue that is not captured sufficiently in the presented regression is the problem of information on access to finance. From the 57 percent of firms in the sample who state that they need additional funding, only 15 percent have applied for a loan in the last twelve months, but of these, only eight percent report that their loan requests have been rejected. One reason for the fact that most firms do not even apply for a loan, despite needing additional funding, may be a lack of information. A third of firms in the RICS stated that they would not apply for a loan from a formal financial institution, because they do not know where or how to apply.

### **The Determinants of Formality**

The empirical analysis presented in the previous section has shown that the impact of licensing on both costs and benefits depends on a range of firm and manager characteristics. We therefore conclude by estimating the probability of being formal in reduced form, in order to identify whether the characteristics which reduce costs (and enhance benefits) are indeed associated with being formal.

$$\Pr(L=1) = \delta_0 + \delta_1 M^S + \delta_2 F^S + \delta_3 C^S + \delta_4 EC + \mu \quad (3)$$

Note that equation (3) is quite different from the first stage of our 2SLS procedure, since it contains neither the village level licensing instrument nor the general investment climate variables. Moreover, it does not contain the costs and benefits directly, but rather those manager, firm and location characteristics ( $M^S$ ,  $F^S$  and  $C^S$ ) that, in interaction with formality, were shown to influence the costs and benefits. The regression also includes the one-time costs of entering the formal sector ( $EC$ ), proxied by registration costs and time.<sup>29</sup> Again, we have an endogeneity problem since some firm characteristics that are dependent on firm performance (e.g. firm size and enterprise age) will be endogenous, since they are also affected by the firm's formality status. To mitigate this, we enter potentially endogenous variables as quintile dummies.

The results are presented in Table 11. The estimates in Column (1) only use the significant interaction characteristics as independent variables; in column (2), manager education has been added as a proxy for the information available to the firm about licensing procedures. The coefficient on education shows the expected positive impact on a firm's probability of

---

<sup>29</sup> The expected average registration time and cost for each firm was calculated as a weighted mean over employment quartiles, sector and sub-districts, based on the existing data on registration processes provided by the licensed firms in RICS. For those sub-districts that had no data entries on time and costs, district averages have been used.

being formal. Although slightly different in point estimates, the results of the other coefficients are robust to the inclusion of education.

Both indicators for entry costs – registration cost and time – have insignificant coefficients close to zero. This outcome could be explained by the fact that uncertainty over costs and licensing procedures, as well as incomplete information, are more important in determining firms' decisions to formalize than the actual amount paid or time waited (cf. Section 3.2; LabSosio 2008: 36).<sup>30</sup>

For all measures of firm size – number of employees, log fixed assets and log total sales – the probability of being formal is significantly larger for enterprises located in the highest quintiles, while no significant impact can be found for firms of medium size. The largest firms in the sample benefit from licensing through a reduction in bribes, a gain in revenues and an increased share of sales made to local governments. Those benefits seem to outweigh the finding that tax payments for large formal firms are higher than for similar informal firms.

Female managers are no less likely than male manager to be licensed, despite gaining less in reduced taxes. For managers who belong to the local indigenous ethnic group, the probability of being formal decreases by about 10 percent. By contrast, Chinese managers are 35 % more likely to register their firm. This is surprising since Chinese managers of formal firms were found to make a lower share of their sales to governments. However if managers of a different ethnicity, and especially Chinese managers, are under closer scrutiny of local authorities (or are less trusted by other businesspeople) they may have a stronger incentive to get licensed. A similar case could be made for managers that do not live in the place of business, who are significantly more likely to be formal than those living in the village.

For rural firms, licensing reduces the amount of taxes paid and the likelihood of paying other levies at all. One might therefore expect rural firms to be more likely to be licensed. In fact being rural reduces the probability of being formal by more than ten percent. This may reflect higher information, time and monetary costs of obtaining licenses for those in rural areas (Mackenzie and Sakho, 2007). Thus, the benefits from licensing will have to be high enough for rural firms in order to outweigh the additional transaction costs of entering the formal sector.

---

<sup>30</sup> A more meaningful indicator would be a variable that captures uncertainty over licensing processes by calculating the deviation of the observed time and cost in RICS from officially announced values. However, no data was available on these official figures.

## 6. Conclusion

In this paper we have analyzed the determinants of going formal for small and micro non-farm enterprises in rural Indonesia. After taking into account the endogeneity of licensing we find that licensing does provide advantages for some firms in terms of reduced tax and corruption payments. This result is contrary to the conventional wisdom that formalization will increase tax payments as firms appear on the tax authorities' 'radar screen'.

We have also shown that the benefits of formalization depend on the size of firm and other firm characteristics such as the ethnicity, religion and gender of the manager, the sector the firm operates in, and whether they are located in rural or peri-urban areas. Access to credit and to government contracts seem on average to be unaffected by the formality status of a firm. This may be due to the fact that most firms in our sample are discouraged borrowers and do not sell to the government quite independent of their formal status.

We have also analyzed the probability of going formal. Larger firms are more likely to get licensed, as are firms that are owned or managed by Chinese Indonesians, and by those with more education. Firms in rural areas, those run by managers who belong to the local indigenous group, and those who live in the village where their companies are located, are less likely to become formal. This suggests that local authorities may favor locals and members of their own ethnicity and put Chinese managers under closer scrutiny.

## References

- Angelini, John and Kenichi Hirose (2004): Extension of Social Security Coverage for the Informal Economy in Indonesia. ILO Working Paper 11, December 2004.
- Auriol, Emmanuelle and Michael Walters (2004): Taxation base in developing countries. In: *Journal of Public Economics* 89: 625-646.
- Djankov, Simeon, Ira Lieberman, Joyita Mukherjee and Tatiana Nenova (2003): Going Informal: Benefits and Costs. In: Belev, Boyan (ed.) *The Informal Economy in the EU Accession Countries* Center for the Studies of Democracy, 63-80.
- Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silanes and Andrei Shleifer (2002): The Regulation of Entry. *The Quarterly Journal of Economics*, Vol. CXVII (1): 1-37.
- Dollar, D, M. Hallward-Driemeier and T. Mengistae (2005). 'Investment Climate and Firm Performance in Developing Economies', *Economic Development and Cultural Change*, 54(1):1-31.
- Enste, Dominik and Friedrich Schneider (2000) Shadow Economies: Sizes, Causes, and Consequences, *Journal of Economic Literature*, Vol. 38: 77-114.
- Fajnzylber, Pablo, William F. Maloney and Gabriel V. Montes Rojas (2006): Does Formality Improve Micro-Firm Performance? Quasi-Experimental Evidence from the Brazilian SIMPLES Program. [http://www.nber.org/books\\_in\\_progress/IASE06/iasef06/fajn.pdf](http://www.nber.org/books_in_progress/IASE06/iasef06/fajn.pdf) (12/05/2008)
- Greene, William H. (2003): *Econometric Analysis*. Fifth Edition. Pearson Education Inc, New Jersey.
- Heckman, James J. (1979): Sample Selection Bias as a Specification Error. In: *Econometrica*, Vol.47 (1): 153-161.
- Husmanns, Ralf and Farhad Mehran (1999) Statistical definition of the informal sector – International standards and national practices. <http://www.gdrc.org/informal/huss0772.pdf> (12/05/2008)
- Ishengoma, Ester and Robert Kappel (2006): Economic Growth and Poverty: Does Formalization of Informal Enterprises Matter? GIGA Working Paper No. 20.
- Jäckle, Annette E. and Carmen A. Lee (2006): Firm Dynamics and Institutional Participation: A Case Study on Informality of Micro-Enterprises in Peru, *Economic Development and Cultural Change*, 54:557–578.
- Kenyon, Thomas (2007): A Framework for Thinking about Enterprise Formalization Policies in Developing Countries. World Bank Policy Research Working Paper 4235.
- KPPOD (Regional Autonomy Watch) and The Asia Foundation (2008): Local Economic Governance in Indonesia. <http://k:od.org/ind/datapdf/rating2007/rating07.pdf> (23/07/2008)
- LabSosio and The Asia Foundation (2008): A Qualitative Evaluation of the Impact of One Stop Shop Business Licensing Service on Local Economies and Governance. Mimeo.
- Levenson, Alec R. and William F. Maloney (1998): The informal sector, firm dynamics and institutional participation. World Bank Working Paper 1988.
- Loayza, Normal A. (1997): The economics of the informal sector. A simple model and some empirical evidence from Latin America. World Bank Policy Research Working Paper 1727.
- LPEM-FEUI (2006a): RICS Data User Guide and Documentation. Mimeo.

- LPEM-FEUI (2006b): Field Survey Report. Indonesian Rural Investment Climate Survey January 22-February 6, 2006. Mimeo.
- Luebke, Christian Von (2006): Leadership and Voice in Local Governance: Political Economy of District Business Regulations in Central Java, West Sumatrea, Bali and NTB. Paper presented for the Rural Investment Climate Assessment, World Bank, Jakarta, processed.
- McCulloch, Neil (ed.) (2009) *Rural Investment Climate in Indonesia*, Institute of Southeast Asian Studies: Singapore.
- McKenzie, David and Seynabou Yaye Sakho (2009): Does it pay firms to register for taxes? The impact of formality on firm profitability, *Journal of Development Economics*, forthcoming.
- Perry, Guillermo, William Maloney and Omar Arias (eds.) (2007): Informality: Exit and Exclusion. World Bank Publications, 1-20.
- Puhani, Patrick A. (2000): The Heckman Correction for Sample Selection and its Critique. In: *Journal of Economic Surveys*, Vol.14(1): 53-68.
- Schneider, Friedrich (2005) Shadow Economies around the World: What do we really know?, *European Journal of Political Economy* 21/3: : .598-642.
- Schulze, Günther G. and Susanne Quadros (2007): Report on the Results of the Rural Investment Climate Assessment (RICA) in Indonesia 2006. Mimeo.
- The Asia Foundation (2007): Making Sense of Business Licensing in Indonesia: A Review of Business Licensing Policy and Survey of One Stop Shop Service Centers.
- World Bank (2006) Revitalizing the rural economy: An assessment of the investment climate faced by non-farm enterprises at the district level. Consultative Draft. World Bank Office Jakarta. <http://siteresources.worldbank.org/INTINDONESIA/Resources/Publication/280016-1152870963030/RICAFINALe.pdf> (09/04/2008)
- World Bank (2007) Doing Business 2008: Country Report Indonesia. International Financial Corporation, Jakarta.
- World Bank (2008) The Rural Investment Climate. Analysis and Findings. Report No.44009-GLB.
- Wooldridge, Jeffrey M. (2002): *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, London.

## Tables

**Table 1: Firm and manager characteristics**

	District					
	Labuan B.	Malang	Badung	Sumbawa	Kutai K.	Barru
<b>Firm characteristics</b>						
# of firms	418	410	399	414	409	411
% rural	80	45	16	47	61	45
% manufacturing	9	18	9	11	6	8
% trade	55	43	62	58	54	54
% services	36	39	29	31	40	38
Mean no. of employees	2.3	2.6	3.3	2.5	2.2	2.6
Mean age of firm (years)	7.2	11.8	6.8	8.9	7.6	12
Median age of firm (years)	4	8	5	6	5	10
<b>Manager characteristics</b>						
Age	41	44	39	42	41	43
% female	32	34	42	51	37	29
% indigenous ethnicity	42	93	65	75	30	92
% chinese Indonesian	4	1	2	3	0	0
<i>Highest level of education</i>						
% did not complete primary	5	26	7	16	13	13
% primary	34	39	19	35	32	23
% junior secondary	26	15	15	16	20	18
% senior secondary	27	12	35	27	24	29
% vocational/university	8	8	24	6	11	16

**Table 2: Mean Costs and Benefits**

**Table 2: Mean Costs and Benefits**

	Formal	Informal
Mean of log total taxes	4.47	1.90
Mean of log total 'other levies'	2.06	1.15
Mean % of sales to government	3.65	0.49
Access to credit	0.24	0.15

**Table 3: Mean Enterprise Size and Age**

**Table 3: Mean Enterprise Size and Age**

	Formal	Informal
Mean number of employees	3.93	2.23



Mean of log total sales	11.03	9.92
Mean enterprise age	10.77	8.54

---

**Table 4: First Stage IV Estimates**

Marginal effects from Probit regression on Licensed

	(1)	(2)	(4)	(5)	(6)
Village Licensed	0.562 (0.086)***	0.547 (0.086)***	0.538 (0.081)***	0.582 (0.084)***	0.531 (0.094)***
Education	0.039 (0.007)***	0.039 (0.007)***	0.039 (0.007)***	0.039 (0.007)***	0.037 (0.008)***
Age	0.003 (0.001)***	0.003 (0.001)***	0.003 (0.001)***	0.003 (0.001)***	0.003 (0.001)***
Female	-0.017 (0.024)	-0.017 (0.024)	-0.017 (0.023)	-0.015 (0.024)	-0.016 (0.024)
Resides in village	-0.246 (0.076)***	-0.244 (0.077)***	-0.247 (0.074)***	-0.242 (0.074)***	-0.253 (0.074)***
Indigenous	-0.085 (0.058)	-0.084 (0.056)	-0.081 (0.054)	-0.085 (0.055)	-0.081 (0.053)
Chinese	0.056 (0.105)	0.048 (0.105)	0.046 (0.100)	0.047 (0.100)	0.049 (0.101)
Islam	-0.043 (0.046)	-0.047 (0.047)	-0.044 (0.046)	-0.049 (0.047)	-0.054 (0.049)
>5 employees	0.085 (0.046)*	0.086 (0.046)*	0.108 (0.041)***	0.087 (0.046)*	0.081 (0.044)*
>median sales	0.044 (0.032)	0.041 (0.033)		0.039 (0.033)	0.044 (0.033)
>median fixed assets	0.069 (0.018)***	0.070 (0.018)***	0.069 (0.016)***	0.070 (0.017)***	0.068 (0.018)***
Older enterprise	0.014 (0.029)	0.013 (0.029)	0.017 (0.028)	0.014 (0.028)	0.015 (0.028)
Rural	-0.015 (0.029)	-0.015 (0.028)	-0.011 (0.027)	-0.005 (0.026)	0.001 (0.027)
Manufacturing	-0.032 (0.028)	-0.033 (0.028)	-0.035 (0.028)	-0.029 (0.029)	-0.028 (0.028)
Service	-0.028 (0.019)	-0.029 (0.019)	-0.037 (0.019)*	-0.025 (0.018)	-0.028 (0.018)
Bank					0.032 (0.024)
Village Ln taxes	-0.002 (0.007)				
Village Other levies		0.007 (0.012)			
Village Ln sales			0.018 (0.018)		
Village Sales to gov				-0.008 (0.004)**	
Village Credit					0.104 (0.056)*
<b>Observations</b>	<b>1676</b>	<b>1676</b>	<b>1715</b>	<b>1676</b>	<b>1636</b>

Robust standard errors in parentheses

Regression contains district dummies

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

**Table 5: Total Taxes**

	(1) OLS	(2) 2SLS	(3) 2SLS	(4) 2SLS	(5) 2SLS
Licensed	0.755 (0.402)*	-1.423 (0.581)**	-3.760 (1.116)***	-1.910 (0.592)***	-1.219 (0.588)**
Int_Female				1.775 (0.695)**	
Int_Rural					-1.432 (0.838)*
Int_Qlnsales_2			1.147 (0.944)		
Int_Qlnsales_3			2.293 (1.184)*		
Int_Qlnsales_4			2.684 (0.995)***		
Int_Qlnsales_5			2.997 (1.445)**		
Education	0.253 (0.063)***	0.388 (0.096)***	0.411 (0.089)***	0.392 (0.095)***	0.394 (0.095)***
Age	0.013 -0.011	0.021 (0.012)*	0.022 (0.011)**	0.021 (0.012)	0.023 (0.011)**
Female	-0.255 -0.209	-0.359 (0.213)*	-0.389 (0.193)**	-0.667 (0.242)***	-0.384 (0.220)*
Resides in village	0.116 -0.391	-0.439 (0.544)	-0.375 (0.585)	-0.424 (0.506)	-0.459 (0.546)
Indigenous	-0.16 -0.235	-0.374 (0.251)	-0.470 (0.257)*	-0.378 (0.249)	-0.387 (0.251)
Chinese	0.1 -0.738	0.705 (0.880)	0.327 (0.910)	0.135 (0.817)	0.550 (0.946)
Islam	-0.499 (0.293)*	-0.587 (0.345)*	-0.676 (0.334)**	-0.566 (0.348)	-0.628 (0.353)*
# Employees	0.122 (0.033)***	0.099 (0.045)**	0.082 (0.044)*	0.112 (0.048)**	0.098 (0.045)**
Ln sales	0.272 (0.058)***	0.347 (0.059)***		0.330 (0.058)***	0.355 (0.061)***
Ln fixed assets	0.066 (0.034)*	0.089 (0.031)***	0.092 (0.028)***	0.086 (0.032)***	0.091 (0.031)***
Enterprise age	-0.016 -0.015	-0.010 (0.015)	-0.009 (0.014)	-0.008 (0.015)	-0.010 (0.014)
Rural	0.277 -0.174	0.099 (0.202)	-0.014 (0.205)	0.128 (0.203)	0.255 (0.241)
Manufacturing	0.256 -0.336	0.579 (0.323)*	0.656 (0.289)**	0.569 (0.337)*	0.601 (0.330)*
Service	0.252 -0.188	0.225 (0.201)	0.149 (0.175)	0.209 (0.205)	0.227 (0.197)
Village Ln taxes	0.75 (0.067)***	0.804 (0.062)***	0.819 (0.061)***	0.802 (0.063)***	0.797 (0.061)***
Constant	-4.368 (1.082)***	-3.950 (1.467)***	-0.572 (1.357)	-3.702 (1.481)**	-4.088 (1.413)***
<b>Observations</b>	<b>1901</b>	<b>1782</b>	<b>1782</b>	<b>1782</b>	<b>1782</b>
<b>R-squared</b>	<b>0.42</b>	<b>0.42</b>	<b>0.43</b>	<b>0.42</b>	<b>0.42</b>

Robust standard errors in parentheses

Regression contains district dummies

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

**Table 6: Other levies and selection bias**

	OLS	2SLS	Heckman outcome (employee)	Heckman selection (employee)	Heckman outcome (female)	Heckman selection (female)
	-1	-2	-3	-4	-5	-6
	Ln other levies	Ln other levies	Ln other levies	Pay other levies	Ln other levies	Pay other levies
Licensed	0.174	-1.806	-1.968	-1.457	-2.095	-1.427
	-0.288	(0.448)***	(0.958)**	(0.464)***	(1.035)**	(0.455)***
Education	-0.108	0.013	0.146	0.008	0.128	0.01
	(0.049)**	-0.063	-0.091	-0.055	-0.091	-0.053
Age	-0.018	-0.009	0.013	-0.01	0.015	-0.011
	(0.007)**	-0.006	-0.013	-0.007	-0.013	-0.007
Female	-0.328	-0.388	-0.396	-0.221		-0.287
	(0.185)*	(0.184)**	-0.307	-0.15		(0.134)**
Resides in village	-0.332	-1.194	-1.165	-0.664	-1.089	-0.664
	(0.185)*	(0.300)***	(0.447)***	(0.208)***	(0.448)**	(0.208)***
Indigenous	0.023	-0.085	-0.182	0.015	-0.25	0.026
	-0.163	-0.139	-0.308	-0.17	-0.313	-0.169
Chinese	1.035	1.674	1.244	0.512	1.093	0.532
	-0.825	(0.794)**	(0.705)*	-0.529	(0.660)*	-0.497
Islam	0.317	0.17	0.069	-0.235	0.002	-0.207
	-0.487	-0.507	-0.402	-0.319	-0.398	-0.311
# Employees	0.111	0.107		0.073	0.07	0.062
	(0.037)***	(0.045)**		(0.033)**	-0.064	(0.033)*
Ln sales	0.1	0.127	0.17	0.13	0.141	0.137
	(0.054)*	(0.040)***	(0.095)*	(0.043)***	-0.098	(0.043)***
Ln fixed assets	0.007	0.03	0.079	0.011	0.085	0.012
	-0.017	-0.018	(0.033)**	-0.017	(0.034)**	-0.017
Enterprise age	-0.003	-0.006	0.006	-0.01	0.004	-0.01
	-0.01	-0.011	-0.019	-0.009	-0.017	-0.009
Rural	0.252	0.036	-0.525	0.122	-0.53	0.13
	(0.147)*	-0.122	(0.269)*	-0.139	(0.274)*	-0.138
Manufacturing	-0.201	-0.173	0.713	-0.367	0.695	-0.379
	-0.242	-0.192	(0.406)*	(0.218)*	(0.406)*	(0.216)*
Service	0.095	-0.005	0.474	-0.161	0.515	-0.17
	-0.182	-0.152	(0.283)*	-0.148	(0.275)*	-0.145
Village Other levies	-0.471	-0.115	0.318	-1.092	0.271	-1.086
	-0.516	-0.193	-0.512	(0.287)***	-0.508	(0.285)***
Constant	-0.273	0	-0.845	-1.002	-0.922	-0.993
	-0.507	0	(0.428)**	(0.236)***	(0.435)**	(0.235)***
Observations	0	-0.094	-1.396	-1.265	-1.53	-1.221
R-squared	0.53	0.54				

Robust standard errors in parentheses

Regression contains district dummies

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

**Table 7: Other levies with interaction effects**

	OLS with instrument but no zeros					
	No interaction -1	sales -2	fixed assets -3	rural -4	ethnicity -5	islam -6
	Ln other levies	Ln other levies	Ln other levies	Ln other levies	Ln other levies	Ln other levies
Licensed	-2.864 (1.206)**	8.345 (2.796)***	1.527 -2.237	-3.105 (1.185)**	-3.127 (1.292)**	-4.034 (0.975)***
Education	0.133 (0.060)**	0.126 (0.058)**	0.092 -0.067	0.116 (0.060)*	0.133 (0.059)**	0.115 (0.054)**
Age	0.011 -0.012	0.006 -0.01	0.018 (0.011)*	0.01 -0.012	0.01 -0.012	0.012 -0.011
Female	-0.489 (0.291)*	-0.442 -0.277	-0.361 -0.236	-0.469 -0.294	-0.498 (0.280)*	-0.475 -0.295
Resides in village	-1.392 (0.235)***	-1.599 (0.273)***	-1.133 (0.258)***	-1.426 (0.242)***	-1.425 (0.214)***	-1.479 (0.243)***
Indigenous	-0.171 -0.216	-0.327 -0.251	-0.131 -0.228	-0.121 -0.222	-0.281 -0.327	-0.2 -0.234
Chinese	1.678 -1.234	1.728 (0.911)*	1.715 (0.960)*	1.801 -1.206	1.834 -1.267	2.278 (0.898)**
Islam	-0.027 -0.417	-0.255 -0.316	-0.286 -0.395	0.008 -0.429	-0.016 -0.423	-0.948 (0.462)**
# Employees	0.08 -0.074	0.096 -0.064	0.093 -0.076	0.091 -0.075	0.082 -0.074	0.071 -0.081
Ln sales	0.172 (0.088)*	0.397 (0.115)***	0.162 (0.072)**	0.162 (0.087)*	0.174 (0.090)*	0.172 (0.089)*
Ln fixed assets	0.08 (0.047)*	0.083 (0.043)*	0.152 (0.041)***	0.079 (0.047)*	0.079 (0.046)*	0.083 (0.041)**
Enterprise age	0.001 -0.012	0.001 -0.012	0.002 -0.012	-0.002 -0.012	0.001 -0.012	-0.002 -0.012
Rural	-0.434 -0.42	-0.309 -0.395	-0.325 -0.4	-0.71 -0.445	-0.42 -0.409	-0.402 -0.416
Manufacturing	0.469 -0.426	0.463 -0.471	0.736 -0.469	0.377 -0.45	0.43 -0.438	0.631 -0.484
Service	0.38 (0.156)**	0.456 (0.160)***	0.515 (0.207)**	0.362 (0.156)**	0.375 (0.158)**	0.42 (0.150)***
Village Other levies	0.613 (0.204)***	0.597 (0.206)***	0.652 (0.190)***	0.626 (0.201)***	0.608 (0.207)***	0.699 (0.189)***
Int_Lnsales		-0.981 (0.226)***				
Int_Ln fixedassets			-0.398 (0.124)***			
Int_Female						
Int_Islam						2.392 (0.826)***
Int_Indigenous					0.372 -0.728	
Int_Rural				2.62 (1.063)**		
Constant	0.795 -1.698	-1.155 -1.728	-0.478 -1.458	1.03 -1.666	0.936 -1.59	0.949 -1.622
Observations	484	484	484	484	484	484
R-squared	0.44	0.48	0.49	0.45	0.44	0.46

Table 8: Total Sales

Dependent variable	OLS	2SLS	2SLS	2SLS
	-1 Ln sales	-2 Ln sales	-3 Ln sales	-4 Ln sales
Licensed	0.241	-0.379	-1.226	0.135
	-0.226	-0.291	(0.349)***	-0.381
Education	0.081	0.129	0.123	0.126
	(0.023)***	(0.028)***	(0.027)***	(0.027)***
Age	-0.008	-0.005	-0.005	-0.006
	(0.004)**	-0.005	-0.005	-0.004
Female	-0.13	-0.108	-0.11	-0.106
	-0.138	-0.139	-0.141	-0.138
Indigenous	-0.14	-0.089	-0.108	-0.092
	-0.133	-0.152	-0.161	-0.152
Chinese	-0.307	0.164	-0.197	-0.107
	-0.455	-0.434	-0.461	-0.393
Islam	-0.137	-0.019	-0.075	0.19
	-0.15	-0.155	-0.162	-0.222
Resides in village	-0.259	-0.225	-0.163	-0.178
	-0.185	-0.247	-0.226	-0.235
Int_qemployee_2	0.205	0.221	0.111	0.22
	-0.159	-0.148	-0.128	-0.149
Int_qemployee_4	0.55	0.575	0.502	0.6
	(0.146)***	(0.158)***	(0.194)**	(0.159)***
Int_qemployee_5	1.4	1.503	1.004	1.476
	(0.199)***	(0.207)***	(0.287)***	(0.202)***
Int_qlnfasset_2	-0.237	-0.191	-0.156	-0.191
	-0.151	-0.156	-0.155	-0.152
Int_qlnfasset_3	-0.202	-0.189	-0.149	-0.186
	-0.136	-0.137	-0.135	-0.136
Int_qlnfasset_4	0.052	-0.056	-0.013	-0.06
	-0.147	-0.153	-0.16	-0.149
Int_qlnfasset_5	0.111	0.113	0.137	0.112
	-0.135	-0.13	-0.13	-0.129
Int_entagedum_2	0.399	0.436	0.438	0.435
	(0.106)***	(0.107)***	(0.107)***	(0.105)***
Int_entagedum_3	0.369	0.475	0.482	0.482
	(0.146)**	(0.172)***	(0.166)***	(0.172)***
Rural	0.241	0.175	0.135	0.169
	(0.119)**	-0.139	-0.138	-0.139
Service	-0.67	-0.677	-0.705	-0.682
	(0.113)***	(0.114)***	(0.110)***	(0.112)***
Manufacturing	-0.398	-0.316	-0.338	-0.337
	(0.158)**	(0.168)*	(0.173)*	(0.167)**
Village Ln sales	0.813	0.836	0.863	0.835
	(0.100)***	(0.095)***	(0.091)***	(0.096)***
Int_Qemployee_2			0.701	
			-0.594	
Int_Qemployee_4			0.516	
			-0.502	
Int_Qemployee_5			1.699	
			(0.465)***	
Int_Islam				-0.656
				(0.278)**
Constant	1.992	1.442	1.362	1.242
	(1.159)*	-1.142	-1.06	-1.129
Observations	1901	1782	1782	1782
R-squared	0.46	0.45	0.46	0.45

**Table 9: Share of Total Sales to Governments**

	(1) OLS	(2) 2SLS	(3) 2SLS	(4) 2SLS	(5) 2SLS
Licensed	1.269 (0.965)	0.766 (1.671)	-2.828 (2.724)	-4.520 (2.207)**	-0.726 (2.313)
Int_Qlnsales2			5.558 (4.608)		
Int_Qlnsales3			15.354 (9.187)*		
Int_Qlnsales4			0.919 (2.802)		
Int_Qlnsales			0.833 (2.499)		
Int_Qlnfasset2				1.646 (1.842)	
Int_Qlnfasset3				8.427 (7.808)	
Int_Qlnfasset4				3.822 (2.059)*	
Int_Qlnfasset5				6.316 (2.541)**	
Int_Chinese					
Int_Manufacturing					37.179 (8.161)** *
Int_Service					1.192 (2.661)
Education	0.602 (0.258)**	0.688 (0.389)*	0.734 (0.360)**	0.725 (0.393)*	0.413 (0.404)
Age	0.019 (0.031)	0.034 (0.035)	0.032 (0.034)	0.035 (0.034)	0.005 (0.029)
Female	0.381 (0.51)	0.361 (0.632)	0.499 (0.623)	0.228 (0.678)	0.344 (0.598)
Resides in village	1.623 (0.962)*	2.231 (1.725)	1.923 (1.506)	1.973 (1.804)	2.728 (1.948)
Indigenous	-0.767 (0.525)	-1.166 (0.735)	-1.085 (0.675)	-1.225 (0.680)*	-1.233 (0.709)*
Chinese	-2.027 (0.954)**	-1.223 (1.160)	-0.068 (1.550)	-1.201 (1.275)	1.088 (1.351)
Islam	2.195 (1.028)**	2.709 (1.655)	2.854 (1.681)*	2.712 (1.433)*	3.258 (1.695)*
Qemployee_2	-0.607 (0.605)	-0.370 (0.530)	-0.389 (0.502)	-0.447 (0.534)	-0.389 (0.511)
Qemployee_4	-0.525 (0.632)	-0.140 (0.778)	-0.221 (0.771)	-0.098 (0.770)	0.105 (0.779)
Qemployee_5	1.003 (1.32)	1.241 (1.750)	1.327 (1.836)	1.193 (1.793)	0.353 (1.629)
Qlnsales_2	0.551 (0.664)	0.525 (0.598)	-0.140 (0.603)	0.459 (0.617)	0.872 (0.509)*
Qlnsales_3	2.547	2.232	-0.096	2.216	2.120

	(1.243)**	(1.328)*	(0.723)	(1.244)*	(1.266)*
Qlnsales_4	-0.416	-0.584	-0.446	-0.567	-0.223
	(1.02)	(1.033)	(0.847)	(1.042)	(0.680)
Qlnsales_5	0.562	0.201	0.732	0.296	0.286
	(0.884)	(1.353)	(1.243)	(1.355)	(1.393)
Qlnfasset_2	-0.564	-0.286	-0.285	-0.276	-0.242
	(0.73)	(0.563)	(0.563)	(0.603)	(0.399)
Qlnfasset_3	0.481	0.528	0.498	-0.445	0.353
	(0.888)	(0.956)	(0.905)	(0.862)	(0.675)
Qlnfasset_4	0.156	-0.495	-0.390	-0.684	-0.261
	(0.607)	(0.876)	(0.844)	(0.849)	(0.704)
Qlnfasset_5	0.364	0.583	0.588	-0.379	0.157
	(1.011)	(1.103)	(1.111)	(1.002)	(1.111)
entagedum_2	-0.438	-0.656	-0.599	-0.645	-0.884
	(0.724)	(0.737)	(0.668)	(0.669)	(0.751)
entagedum_3	-0.263	-0.607	-0.488	-0.633	-0.242
	(0.974)	(1.071)	(1.013)	(1.073)	(0.971)
Rural	-0.006	-0.148	-0.058	-0.188	-0.277
	(0.731)	(0.670)	(0.581)	(0.590)	(0.437)
Manufacturing	8.247	9.157	9.011	8.899	2.951
	(4.403)*	(4.683)*	(4.455)**	(4.522)*	(2.442)
Service	0.899	0.656	0.654	0.516	0.400
	(0.420)**	(0.645)	(0.634)	(0.723)	(0.382)
Village Sales to gov	0.781	0.702	0.730	0.698	0.509
	(0.077)**	(0.096)**	(0.100)**	(0.103)**	(0.154)**
	*	*	*	*	*
Constant	-6.804	-8.159	-8.025	-7.259	-6.914
	(2.615)**	(2.127)**	(2.191)**	(2.067)**	(2.294)**
	*	*	*	*	*
Observations	1901	1782	1782	1782	1782
R-squared	0.18	0.19	0.21	0.19	0.28

Robust standard errors in parentheses

Regression contains district dummies

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%



**Table 10. Access to Credit**

Marginal effects from Probit regression on access to credit

	(1) OLS	(2) IV	(3) IV Interaction	(4)
Licensed	0.012 (0.066)	0.021 (0.124)	0.175 (0.088)**	0.093 (0.123)
Int_Qemployee_2			-0.280 (0.094)***	
Int_Qemployee_4			0.136 (0.186)	
Int_Qemployee_5			-0.353 (0.137)***	
Int_Manufacturing				-0.366 (0.134)***
Int_Service				-0.115 (0.069)*
Education	-0.022 (0.012)*	-0.013 (0.012)	-0.015 (0.013)	-0.013 (0.013)
Age	-0.002 (0.002)	-0.001 (0.002)	-0.001 (0.002)	-0.001 (0.002)
Female	0.038 (0.042)	0.044 (0.040)	0.044 (0.034)	0.040 (0.041)
Resides in village	0.033 (0.055)	0.079 (0.023)***	0.055 (0.024)**	0.075 (0.024)***
Indigenous	0.019 (0.029)	0.029 (0.029)	0.028 (0.027)	0.027 (0.030)
Chinese	0.021 (0.164)	-0.011 (0.136)	0.076 (0.216)	-0.034 (0.098)
Islam	0.025 (0.046)	-0.011 (0.037)	-0.007 (0.035)	-0.011 (0.039)
Qemployee_2	-0.01 (0.033)	0.014 (0.034)	0.064 (0.043)	0.017 (0.033)
Qemployee_4	-0.013 (0.043)	-0.017 (0.049)	-0.035 (0.054)	-0.015 (0.049)
Qemployee_5	0.162 (0.086)*	0.202 (0.101)**	0.387 (0.159)**	0.212 (0.104)**
QInsales_2	-0.046 (0.036)	-0.049 (0.030)*	-0.047 (0.024)*	-0.049 (0.029)*
QInsales_3	0.046 (0.048)	-0.005 (0.038)	-0.008 (0.032)	-0.001 (0.037)
QInsales_4	0 (0.046)	-0.004 (0.040)	-0.022 (0.036)	-0.012 (0.040)
QInsales_5	-0.012 (0.064)	-0.035 (0.056)	-0.027 (0.052)	-0.036 (0.055)
QInfasset_2	0.021 (0.05)	0.027 (0.041)	0.030 (0.040)	0.018 (0.039)
QInfasset_3	0.095 (0.054)*	0.063 (0.050)	0.068 (0.044)	0.050 (0.047)
QInfasset_4	0.05 (0.061)	0.055 (0.044)	0.067 (0.044)	0.044 (0.044)
qInfasset_5	0.202	0.176	0.188	0.161

	(0.089)**	(0.072)**	(0.071)***	(0.073)**
Enterprise age_2	0.004	-0.034	-0.024	-0.034
	(0.033)	(0.028)	(0.028)	(0.028)
Enterprise age_3	0	0.011	0.016	0.010
	(0.034)	(0.028)	(0.033)	(0.027)
Bank	-0.001	0.000	-0.006	-0.006
	(0.023)	(0.024)	(0.022)	(0.024)
Rural	0.001	0.001	0.006	0.004
	(0.025)	(0.025)	(0.024)	(0.025)
Manufacturing	-0.046	-0.038	-0.030	0.046
	(0.055)	(0.045)	(0.043)	(0.084)
Service	0.009	-0.009	-0.003	0.015
	(0.037)	(0.029)	(0.025)	(0.039)
Village Credit	0.858	0.732	0.723	0.735
	(0.078)***	(0.079)***	(0.077)***	(0.080)***
<b>Observations</b>	<b>1162</b>	<b>1076</b>	<b>1076</b>	<b>1076</b>
<b>Pseudo R-squared</b>	<b>0.1812</b>	<b>0.3084</b>	<b>0.3336</b>	<b>0.3158</b>

Robust standard errors in parentheses

Regression contains district dummies

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

**Table 11. Firms' Decision to Formalize**  
Marginal effects from probit regression on *Licensed*

	(1)	(2)	(3)
Qlnsales_2	0.023 (0.052)	-0.026 (0.045)	-0.026 (0.045)
Qlnsales_3	-0.010 (0.051)	-0.015 (0.047)	-0.015 (0.047)
Qlnsales_4	0.117 (0.067)*	0.103 (0.063)	0.103 (0.063)
Qlnsales_5	0.221 (0.079)***	0.161 (0.074)**	0.161 (0.074)**
Qemployee_2	0.048 (0.044)	0.050 (0.043)	0.050 (0.043)
Qemployee_4	0.069 (0.054)	0.055 (0.051)	0.055 (0.051)
Qemployee_5	0.175 (0.072)**	0.087 (0.063)	0.087 (0.063)
Qlnfasset_2	0.106 (0.068)	0.121 (0.069)*	0.121 (0.069)*
Qlnfasset_3	-0.005 (0.058)	0.002 (0.061)	0.002 (0.061)
Qlnfasset_4	0.149 (0.070)**	0.149 (0.068)**	0.149 (0.068)**
Qlnfasset_5	0.206 (0.069)***	0.183 (0.067)***	0.183 (0.067)***
Female	-0.022 (0.036)	-0.016 (0.036)	-0.016 (0.036)
Indigenous	-0.102 (0.041)**	-0.091 (0.042)**	-0.091 (0.042)**
Islam	-0.083 (0.084)	-0.059 (0.077)	-0.059 (0.077)
Chinese	0.353 (0.158)**	0.379 (0.156)**	0.379 (0.156)**
Rural	-0.195 (0.031)***	-0.163 (0.032)***	-0.163 (0.032)***
Manufacturing	0.035 (0.086)	0.032 (0.074)	0.032 (0.074)
Service	0.025 (0.037)	0.012 (0.034)	0.012 (0.034)
Mean Cost	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Mean Time	0.002 (0.002)	0.002 (0.002)	0.002 (0.002)
Education		0.045 (0.010)***	0.045 (0.010)***
Resides in village		-0.155 (0.064)**	-0.155 (0.064)**
Observations	1851	1682	1682

Robust standard errors in parentheses  
Regression contains district dummies

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

## Appendix

**Table A1. List of Variables**

Variable Name	Description
Age	manager age in years
Bank	dummy for whether a financial institution is available in the village area
Chinese	dummy = 1 for Chinese manager
Licensed	dummy =1 for firms that hold at least one of the four most common local licenses (TDI, SIUP, IMB or TDP)
Pay other levies	dummy for firms with positive payments of 'other levies'
Credit	access to credit (dummy =1 for firms that have received a loan or credit in the last 12 months; dummy=0 for all firms that need additional funding)
Education	manager education (ordinal scale)
>5 employees	dummy for firms with >5 employees
# Employees	total number of employees
Enterprise age	Age of the enterprise in years
Older enterprise	dummy = 1 for firms longer in operation than median
Fixed assets dummy	dummy = 1 for firms with fixed assets above median
Female	dummy =1 for female manager
Fully formalized	dummy =1 for firms being fully formalized
Sales to gov	share of sales made to government
Indigenous	dummy =1 for manager who belongs to a local indigenous group
Int_*	prefix indicating interaction term with fitted value of Licensed
Islam	dummy = 1 for muslim manager
Ln other levies	log of total 'other levies' at central, provincial and district level
Ln fixed assets	log of total fixed assets 2005 (includes land, buildings, equipment, furniture, vehicles)
Ln sales	log of total sales 2005
Ln taxes	log of total taxes at central, provincial and district level
Manufacturing	dummy for firm in manufacturing sector
Mean cost	average expected registration costs averaged over employment, location and sector
Mean time	expected registration time, averaged over employment, location and sector
Qemployee1-5	dummies for employee quintiles
Qent_age	dummies for enterprise age quintiles
Qlnfasset1-5	dummies for ln fixed asset quintiles
Qlnsales	quintiles for ln sales
Qlnsales1-5	dummies for ln sales quintiles
Resides in village	dummy =1 for manager who lives in the same village as the business operate
Rural	dummy for rural =1 and urban =0
High sales	dummy = 1 for firms with sales above median
Service	dummy for firm in service sector

Pays tax	dummy for firms with positive tax payments
Village Licensed	village average of Licensed
Village Other levies	village average of Ln other levies
Village Credit	village average of Credit
Village Sales to gov	village average of Sales to gov
Village Ln sales	village average of Ln sales
Village Ln taxes	village average of Ln taxes

**Table A2 Descriptive Statistics**

Variable Name	Obs.	Mean	St.Dev.	Min	Max
Fully formalized	2461	0.02	0.138	0	1
Licensed	2308	0.234	0.424	0	1
Village Licensed	2183	0.230	0.231	0	1
Pays tax	1155	5.251	1.959	0	12.11
Ln taxes	2461	2.479	2.942	0	12.11
Village Ln taxes	2461	2.479	1.599	0	11.98
Pay other levies	673	4.805	1.756	0	12.26
Ln other levies	2461	1.322	2.334	0	12.26
Village Other levies	2461	1.322	1.431	0	8.01
Age	2371	41.64	11.79	16	85
Female	2380	0.3744	0.484	0	1
Indigenous	2370	0.659	0.474	0	1
Islam	2380	0.858	0.349	0	1
Resides in village	2173	0.862	0.345	0	1
Chinese	2370	0.017	0.130	0	1
# Employees	2400	2.58	2.235	1	19
Enterprise age	2461	9.063	8.783	1	66
Ln sales	2382	10.177	1.624	3.76	17.74
Sales to gov	2461	1.156	8.687	0	100
Village Sales to gov	2461	1.156	4.387	0	100
Ln fixed assets	2413	7.842	4.235	0	18.36
Rural	2461	0.491	0.500	0	1
Bank	2423	0.592	0.492	0	1
Credit	1518	0.177	0.382	0	1
Village Credit	2427	0.196	0.193	0	1
Mean Time	2303	10.264	8.815	1	60
Mean Cost	2281	550.444	467.385	5	2000