

# What Determines Firms' Decisions to Formalize?

## Empirical Evidence from Rural Indonesia\*

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

Preliminary version, 16. February 2009

### 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 in an IV approach the effect of formalization on tax payments, corruption, access to credit and revenue. We take firm heterogeneity into account and show how the effect of formalization depends on firm characteristics such as firm size, education and ethnicity of the owner. We identify firm and owner characteristics that make formalization more likely.

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, but also developed countries have sizeable informal sectors (Enste and Schneider 2000, Schneider 2005). These large informal sectors are potentially detrimental to economic development as they escape government taxation and regulation. The government is lacking necessary resources for financing development and cannot ensure minimum safety and labor standards in this sector. Moreover, informal firms may not have equal access to formal (credit) institutions, government contracts and other official or private institutions and lesser visibility and thus may have lesser chances to grow. It would thus 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 high entry barriers as they find the resulting few large tax payer easier to tax. Loayza (1997) finds that large informal sectors are a consequence of excessive taxation, overregulation and weak government institutions resulting in low growth rates.

Not all firms are affected by government policies in the same way or to the same extent. Which are the firms that decide to go formal, who decides to stay away from being registered and licensed? This is the concern of the paper. 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. Then we seek to identify the characteristics of formal firms as opposed to informal firms. To identify the characteristics of the informal firms would help to design policies that increased the formal sector and allowed for a less distorting and more efficient regulation.

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

National Micro Enterprise Survey 1992. Jäckle and Lee (2003) confirm Levenson and Maloney's hypothesis in a panel approach that allows capturing firm dynamics by using the Peruvian Living Standard Measurement Survey for 1994, 1997, and 2000. Insightful as they are, they focus on two firm characteristics only, yet other characteristics such as gender, ethnicity and religion of the manager may matter if formalization is not entirely voluntary and certain groups may be more scrutinized 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 a formal status that the authors do not address. Older and more successful firms may be more likely to go formal but formal firms may 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 (2007) use an instrumental variable *approach* to address the endogeneity. 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 the 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 potential source for it. Moreover it remains unclear whether distance to tax office is a good instrument as it could capture also 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>1</sup> which covers 2500 small and micro non-farm enterprises and their communities in rural Indonesia. This survey has been conducted by the World Bank and is among the first to be followed by a number of RICS in other countries. Indonesia is particularly suited for such an analysis because its diversity allows studying systematically the role of ethnicity, religion and community characteristics. We adopt the 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 the status (formal versus informal) and we interact the formalization variable with a number of household characteristics to account for the systematically different effect of formalization on

---

<sup>1</sup> [www.worldbank.org/id/rica](http://www.worldbank.org/id/rica)

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

The remainder of the paper is structured as follows. Section 2 presents the RICS dataset (Sect. 2.1), provides a definition of formality and reports on current licensing practices in Indonesia (Sect. 2.2), and introduces the potential costs and benefits that are analyzed in the empirical section. After introducing the empirical setup, Section 3 presents and discusses the estimation results. Ordinary Least Squares (OLS) and the IV estimates will be compared, before assessing the heterogeneous treatment effect of licensing through different firm characteristics. The last subsection of the empirical part presents the results of estimating a reduced form Probit regression on firms' decision to formalize. Section 4 concludes.

## **2. Data and Definitions**

### **2.1. The Indonesian Rural Investment Climate Survey**

The Indonesian Rural Investment Climate Survey (RICS) is an in-depth quantitative survey of households, non-farm enterprises (NFE) and communities that intends to capture the rural investment climate. It has been conducted in six districts (*Kabupaten*) in different parts of Indonesia that all provide distinct geographical environments to local enterprises (World Bank 2006: 10):

- 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
- Sumbawa, NTB – a dryland area

The sampling within these districts was done in three steps<sup>2</sup>. Based on the 2006 PODES data 30 villages (*desa*) 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

---

<sup>2</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).

residence), ‘standalone enterprises’ (not located in the place of residence) and ‘pure households’ was drawn. Additionally a number of enterprises registered with the district statistical office<sup>3</sup> were sampled to ensure the inclusion of enterprises of different size, the assumption being that those ‘listed firms’ are likely to be larger and balance a possible bias of smaller household enterprises. This stratified sampling 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 (cf. WB 2008: 11f).

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. Table 1 summarizes all data collected.

**Table 1: Data in RICS**

<b>subject</b>	<b>obs</b>	<b>questionnaire answered</b>
pure households	1057	(1) household
household enterprises	1757	(1) household & (2) enterprise
standalone enterprises	618	(2) enterprise
listed enterprises	146	(2) enterprise
villages	149	(3) community

This paper will utilize a merged enterprise dataset, which means that pure households not operating an enterprise have been excluded. However, information from the household questionnaire answered by household enterprises has been retained. Additionally, information from the community questionnaire exists for each village in which the enterprises surveyed are located. Sorted by employment groups, the merged dataset contains 2137 micro enterprises (1-4 employees), 263 small enterprises (5-19 employees), 51 medium enterprises (20-99 employees), and nine large enterprises (> 100 employees)<sup>4</sup>. The analysis in this paper will focus on small and micro firms. Thus the 60 medium and large firms have been excluded from the sample for further analysis. Of the remaining 2461 firms almost 90 percent of firms are ‘micro’ firms and the mean number of employees (2.6) is rather low.

<sup>3</sup> Registration with the statistical office does not necessarily imply that those firms are operating in the formal sector (Schulze and Quadros 2007).

<sup>4</sup> The formation of employment groups follows the BPS classification of micro (1-4 employees), small (5-19), medium (20-99), and large enterprises (>100). The number of employees also includes the owner or manager of the enterprise himself as well as unpaid laborer.

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 villages are still classified as either 'urban' or 'rural' The term urban hereby refers to a set of village characteristics (share of agriculture, population density, and several services and facilities associated with urban living) that also applies to peri-urban or smaller urban centers (WB 2006: 4). According to this definition about half of the firms in the sample are located in rural (49 %) areas. 54 % operate in the trading sector, 35 % in the service sector and the remainder is in the manufacturing sector.<sup>5</sup> 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.

For about 90 percent of firms the owner is also the primary manager of the business. 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. The majority of firms surveyed are managed by men (63 %). The mean age of the managers is 41 years. Education is at a very low level for almost half of all firm managers. 13 percent state that they did not enjoy any schooling and about 30 percent have only completed elementary school. 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. In Malang and Barru more than 90 percent of managers belong to the local ethnic group (Javanese and Bugis), about 75 percent in Sumbawa (Sumbawa/ Sasak), 70 percent in Kutai (Kutai/ Dayak), 65 percent in Badung (Balinese) and 58 percent Labuhan Batu the share of indigenous managers accounts for 65 percent and respective. Another ethnic group specified here are the Chinese, accounting for 1.7% of enterprise managers throughout the six Kabupaten. Compared to ethnicity, religion is quite homogeneous. Except for Badung in Bali, the share of Islamic managers is more than 90 percent in all Kabupaten. In Badung the majority are Hindus (64%), with only 28 percent of managers being Moslems.<sup>6</sup>

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

Business licensing in Indonesia is characterized by complex, time-consuming, and costly procedures. According to the 'Doing Business Report' by the World Bank and the

---

<sup>5</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.

<sup>6</sup> Descriptive statistics for all variables used for the analysis are summarized in the Appendix (Table A2).

International Financial Corporation (2008), Indonesia ranks among the worst-performing countries in the world in terms of 'starting a business' and 'dealing with licenses'.<sup>7</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 30 provinces, subdivided into 440 districts (Kabupaten and Kota). Those 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). One central feature of the licensing system in Indonesia is the fact that it is overregulated while at the same marked by overlapping functions and redundancies (ibid: 30, TAF 2007: 7, 10). Local governments appreciate licensing services more as a means of generating local revenues than for regulating markets or to efficiently collect information: since decentralization about 1600 new regulations and local government licenses have been introduced, at least 30% of which can be 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. These processes can be roughly described as follows (cf. 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 register for a tax identification number (NPWP) with the central government. Second, the firm needs physical permits such as a construction, building location (IMB) or nuisance permit. 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 in a fifth step.<sup>8</sup>

The RICS data contains information on the following components of this 'package': building permit (IMB), industrial permit (TDI), trade permit (SIUP), enterprise registration (TDP), and

---

<sup>7</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. (Doing Business 2008)

<sup>8</sup> For a limited liability company these steps are even more numerous. For a detailed description of all processes involved see Doing Business (2008: 53ff).

tax identification number (NPWP)<sup>9</sup>. In terms of the formalization procedure explained above and taking into account the data available, a firm would be 'fully registered' if it has a tax number, the building permit (if it operates in a separate building), 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 reality, it is hard to tell which licenses a business exactly needs. The trade permit for example, although officially designed for companies engaged in trading activities, is generally the most common license even among industrial firms (KPPOD 2008: 32). In RICS the trade permit is almost as common among manufacturing firms as the industrial permit, with 13 percent holding the former and 14 the latter. Legally, the building license is a prerequisite for firms operating in a separate building in order to obtain the industrial or trade permit and subsequent enterprise registration<sup>10</sup>. In practice, this permit is also often issued to firms that operate their business inside the dwelling. In RICS more than eleven percent of those firms that run their business inside their homes and almost four percent of firms that do not even operate from a fixed location do have an IMB. Hence one main problems of why such a small number of firms are 'fully formalized' in the legal sense seems to be one of information.<sup>11</sup>

This paper will therefore rely on a more realistic identification strategy: 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 23.44% or 541 enterprises. Since the policies on local licenses are being determined at 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%) and Badung, Sumbawa and Kutai with shares between 23 and 29 percent. In Barru as much as 42 percent of firms are holding at least one of the four main local licenses.

---

<sup>9</sup> As opposed to the local licenses, the data does not contain any additional information on the tax registration process, but simply the information on whether or not a firm has the NPWP.

<sup>10</sup> Its purpose is to provide evidence for the fact that an enterprise is really located in a certain building and also includes neighborhood safety guarantees.

<sup>11</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 to obtain a license is because they think it is not required for them. This is the case for 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 TDP, as the only license that is needed by all firms no matter from where or in which sector they operate, reveals the same picture. Only 7.5% of firms hold an enterprise registration at all. From the remaining 2011 firms more than 51 % state that TDP it is not required. Yet it is not entirely clear whether this is the truth or just an excuse as we cannot exclude an 'experimenter effect'.



## 2.3. Costs and Benefits of Formality

Similar to the studies summarized in the introduction, the analysis in this paper will follow a cost-benefit approach. The decision to formalize will depend on “firms perceiving that it is in their self-interest” (Kenyon 2007: 5), i.e. when the net benefits of doing so outweigh the net benefits of operating informally. Those potential costs and benefits can be categorized into three subcomponents: the one-time costs of entering the formal sector, the costs and the benefits of operating formal.

### Entry costs

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.<sup>12</sup> Entry costs in this paper will be captured 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). On average, an enterprise included in the RICS data has to pay IDR 548,000 and wait for 10.8 days to obtain a license.

### Operating formal costs

A formal enterprise has to abide by government regulations that are often very complex. Besides the direct monetary costs in terms of taxes (or bribes) to government officials, the firm has to bear the indirect costs of time spent for submitting government documents or fulfilling requirements such as product or working standards (cf. Djankov et al. 2003: 66f; Ishengoma and Kappel 2006: 16f). These costs of operating formally can also be considered as the forgone benefits of operating informally. An informal firm would save time and money by avoiding government regulations (but not necessarily taxes). At the same time, it is more dependent on the ‘goodwill’ of government officials and the police, which makes bribe payments more unpredictable or even higher for informal firms (cf. *ibid*: 18, Djankov et al. 2003: 71). Anecdotal evidence from a qualitative evaluation of One Stop Shops for Business Licensing in Indonesia suggests that one main reason for firms to get a license is to reduce unofficial payments to business inspectors (LabSosio 2008: 41).

Operating formal costs will be analyzed by looking at the levies a firm had to pay in 2005, subdivided into total taxes and ‘other levies’. Total taxes include the amount of taxes a firm has paid to the central, provincial and local governments. The ‘other levies’ comprise payments to security officials, thugs, and sub-district or village officials and will be used as a proxy for unofficial payments and bribes.

---

<sup>12</sup> The registration process differs tremendously from country to country. While in 2008 it took only two procedures and two business days at no cost to start a new business in Australia, opening a model business in Indonesia required 12 procedures in 105 business days and costs 80% of per capita GDP (Doing Business 2008). In their analysis of entry regulation, Djankov et al. (2002) find that all three measures of entry costs - the number of procedures, cost-to-per-capita-GDP ratio and time to meet government requirements – decrease with a country's GDP per capita. This means that the emergence of new (formal) firms is more constrained in countries with lower GDP per capita.

## Operating formal benefits

Similar to the costs of operating as a formal enterprise, the benefits can easily be understood as the opportunity costs of informality (Ishengoma and Kappel 2006: 18). An informal firm needs to stay small or might even need to change location more often in order to escape detection and harassment by the police or government officials (cf. *ibid*; Djankov et al. 2003: 70). Hence it might not consider business expansion, even if this would be profitable. For the same reason it may avoid investment in fixed assets or technology. Investments are also constrained by informal firms' limited access to financial resources or risk-sharing mechanisms (cf. *ibid*, Leveson and Maloney 1998: 4 Ishengoma and Kappel 2006). Thus, an informal firm operates under larger uncertainty and additionally does not enjoy legal protection. It can easily be shut down by the police or local governments and is not able to use the legal system in solving disputes. Furthermore, informal firms are excluded from markets such as public procurement and cooperation with the formal sector is limited for informal firms (cf. Ishengoma and Kappel 2006: 18f).

The benefits of operating formally are analyzed here by looking at total firm revenue, access to government contracts and access to credit. Access to government contracts is measured by the share of total firm sales made to governments. Only 3 percent of firms are doing business with governments at all, of which the mean share of sales to governments is 38 percent.

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.<sup>13</sup>

While the initial costs of entering the formal sector are a self-explanatory direct cost of formalizing, payment of levies and performance or expansion measures will only qualify as a cost or benefit if they are actually altered by the fact that a firm is formal. Comparing the means of several measures of firm performance confirms the basic assumption that had already been underlying the analysis of Levenson and Maloney (1997): compared to informal

---

<sup>13</sup> One potential benefit of formality that is missing in this list is the opportunity to use the legal system. The question of whether licensed firms have access or better access to different modes of conflict resolution, enforcement of contracts or property rights cannot be empirically addressed within this paper due to limited data on this topic in RICS. From the 2.8% of firms reporting any payment dispute at all and the 4.47% of firms stating that they had problems with competitors, suppliers or buyers in the last three years, none of them had ever tried using the court system to solve those problems. Most firms would rather address these kinds of conflict directly or informally by involving a respected individual, neighbor or friend.

firms formal enterprises are generally found to be older and larger, both in terms of total sales and number of employees (Table 2).

**Table 2: Mean Size and Enterprise 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

The same expected differences can be found for the potential operating formal costs and benefits discussed in this Section. The comparison of mean values in Table 3 shows that formal firms pay more taxes, more ‘other levies’, sell a larger share to government institutions, and are more likely to have been granted a loan.

**Table 3: 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

However, these differences could simply be due to other firm or manager characteristics rather than formality itself – the most straightforward reason being that formal firms are generally found to be larger (see Table 2). The identification of such a causal effect of licensing on the potential costs and benefits will be the central aim of the empirical analysis in the next Section.

### 3. Empirical Analysis

### 3.1. Empirical Setup

The estimation of the effect of formality on potential costs and benefits needs to address to issues. First, formal status is not exogenous to costs and benefits and thus needs to be instrumented for. Second, formality affects heterogeneous firms differently. Thus the overall coefficient of the formality variable in the respective cost or benefit regression gives little insight into the main question at hand: which kind of firms will actually benefit from licensing and find it worthwhile to register in the end?

In an environment marked by regulatory inefficiency and deficient enforcement, formality is a choice made by firms. The fact that informal firms are generally found to be smaller and less productive might simply reflect smaller firms' choice to stay informal rather than being a consequence of informality (cf. McKenzie and Sakho 2007: 2). That is, if formality is a beneficial option for only some firms; those who do not find it beneficial will simply stay informal. Hence we face a problem of simultaneous equations<sup>14</sup>: the firm's decision to formalize depends on the costs and benefits of operating formally, while at the same time we assume that those costs and benefits are themselves a function of being formal. One possible solution to this problem that will be applied in this paper is an instrumental variables (IV) approach estimated by Two Stage Least Squares (2SLS). In the first stage, the instrument ( $I$ ) together with a set of other exogenous variables ( $X$ ) will be used to predict fitted values of formality ( $L$ ).

$$L = \alpha_0 + \alpha_1 X + \alpha_2 I + \nu \quad (1)$$

In the second stage this predicted value ( $\hat{L}$ ) will replace the observed value of formality in the regressions on costs and benefits ( $Y$ ):

$$Y = \beta_0 + \beta_1 \hat{L} + \beta_2 X + \varepsilon \quad (2)$$

A valid instrument must fulfill two conditions: first, it has to be exogenous to both formality and the potential costs and benefits. Second, it has to be correlated with licensing, but not the dependent variables estimated in the second stage. Any correlation of the instrument with the costs and benefits should be explainable through the effect on the instrumented variable. This paper will use community averages of licensing as an instrument. Section 3.2 will discuss the reasoning for choosing this instrument and provide the first stage estimation results.

---

<sup>14</sup> For a detailed discussion of endogeneity and the instrumental variable approach see for example Wooldridge (2002: 50ff, 83ff).

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

$$Y = \beta_0 + \beta_1 L + \beta_2 M + \beta_3 F + \beta_4 C + \varepsilon \quad (3)$$

where  $Y$  denotes the respective cost or benefit, i.e. taxes, 'other levies', firm revenue, sales made to the government, and access to credit.  $M$ ,  $F$ , and  $C$  are controls for manager, firm and location (community) characteristics. The results of estimating an OLS regression that does not take into account endogeneity and employs the original formality measure ( $L$ ) will be compared to the results of the 2SLS approach, where  $L$  is replaced by the first-stage IV estimates ( $\hat{L}$ ).

$\hat{L}$  enters the equation separately in order to estimate the overall effect of licensing, but also in the form of interaction terms with the respective firm, manager and location characteristics:

$$Y = \beta_0 + \beta_1 \hat{L} + \beta_2 \hat{L} * M + \beta_3 \hat{L} * F + \beta_4 \hat{L} * C + \beta_5 M + \beta_6 F + \beta_7 C + \varepsilon \quad (4)$$

These interaction terms mirror the underlying assumption of a heterogeneous treatment effect of licensing. McKenzie and Sakho (2007) show that the effect of formality on firm profits varies with the number of employees. This paper will further disentangle the overall effect of licensing and assume treatment effects to vary with further manager, firm and location characteristics in order to analyze whether licensing effects on prospective costs and benefits are dependent on specific characteristics, explaining why only some firms choose to formalize and others do not.

Finally, the probability of being formal will be estimated in a reduced form Probit regression (section 3.4).

$$\Pr(L=1) = \delta_0 + \delta_1 M^S + \delta_2 F^S + \delta_3 C^S + \delta_4 EC + \mu \quad (5)$$

This equation does not contain the costs and benefits as such, but those manager, firm and location characteristics ( $M^S$ ,  $F^S$  and  $C^S$ ) that in interaction with formality cause a variation of  $Y$  in equation (4). The regression also includes the one-time costs of entering the formal sector ( $EC$ ), proxied by registration costs and time. Again, the problem of endogeneity persists in this equation: all firm characteristics that are dependent on firm performance, i.e. measures of firm size and enterprise age, will be endogenous, since they are also affected by the firm's formality status. To mitigate this endogeneity problem at least to some degree, those variables will enter the equation as quintiles dummies instead of using the actual variable values.

The present study will follow a similar approach to McKenzie and Sakho (2007), but differ from their paper in important ways. McKenzie and Sakho only consider heterogeneous effects of formality on firm profits according to firm size. This paper will look at a more comprehensive list of owner, firm and location characteristics that might cause heterogeneous treatment effects on a number of potential costs and benefits such as gender and ethnicity. Instead of analyzing urban firms, we investigate rural non-farm enterprises, thus giving some more insight into the different dynamics of rural areas as compared to firms in metropolitan areas. Lastly our analysis is the first to use data from Asia.

### **3.2. IV Estimate: The First Stage**

The most difficult task in applying the IV approach is to find an instrumental variable that fulfils the two conditions named above: (1) it has to be exogenous to both, licensing and the potential costs and benefits and (2) the instrument should only affect the second stage through its impact on licensing.

In this paper we will use community averages of licensing as an instrument.<sup>15</sup> As an aggregate measure, village licensing averages are exogenous to the question of whether an individual firm is formal as well as all other firm-specific potential costs and benefits. Thus condition (1) stated above is fulfilled. The second condition requires average licensing to be correlated with the second stage dependent variables only through its impact on licensing. Here a potential shortcoming of the instrument arises: a village with a high share of licensed firms can also be expected to perform better in other fields of economic policy and to provide a more favourable business climate. Failing to control for these confounding effects would again result in biased estimates, since the instrument in the second stage (equations (3) and (4)) might also capture other village-level influences on the dependent variables. To address this problem each regression will control for village averages of the respective cost or benefit analyzed.<sup>16</sup>

The results of the first stage estimates are summarized in Table A4. The Probit regressions to predict the fitted value of licensing controls for the exogenous firm, manager and location

---

<sup>15</sup> The approach of using averages as an instrument is quite common, cf. for example Wößmann and West 2006. We tried to use community characteristics (frequency of village meetings, different measures for conflict, local business organizations and education of the village head) as instruments

<sup>16</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.

characteristics that will also be used to estimate the costs and benefits in the second stage<sup>17</sup>. Performance of the instrument was tested for overidentification and weakness. The Sargan test of overidentification requires the application of at least one more instrument than endogenous variables (cf. Wooldridge 2002: 122f). To perform this test, some of the weaker instruments discussed 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 will therefore be 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.

### 3.3. Quantifying the Costs and Benefits

This section will summarize the results from estimating the impact of formality (*com\_form*) on official and unofficial levies as well as business performance and expansion. For each dependent variable the results of OLS estimates will be compared to those of 2SLS.<sup>18</sup> For testing heterogeneity of treatment effects interaction terms of the fitted values for licensing and the various control variables were employed. Only those interaction terms significantly entering the cost-benefit regressions are reported in Tables A5-A11.

Each cost and benefit regression will control for the same firm, manager and location characteristics. The manager characteristics are highest level of education (*edu*), gender (*female*), age (*age*), ethnicity (*indethn*), a dummy for whether or not the manager is Muslim (*islam*) and a dummy that denotes whether the manager lives in the same village that the enterprise operates in (*resdum*). The ethnicity dummy (*indethn*) equals one if the manager belongs to the local indigenous group. The religion dummy indicates whether a manager belongs to the majority religion Islam. Balinese managers are excluded from this variable,

---

<sup>17</sup> In order to obtain consistent estimates the same control variables should be included in both stages (cf. Wooldridge 2002: 51). To reduce 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>18</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 (cf. section 2.1). 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 (cf. Greene 2002: 341ff). 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.

since the majority religion here is Hinduism (cf. section 3.1)<sup>19</sup>. Furthermore a dummy for managers of Chinese descent will be included separately in the regression.

The regressions control for firm size<sup>20</sup> in terms of the number of employees (*employee*), log total sales (*lnsales*), and log total value of fixed assets (*lnfasset*)<sup>21</sup>. Further firm-level variables included are the number of years the firm has been in operation (*ent\_age*), and sector (*manufac* and *service*).

Controls for firm location are the respective village averages of the costs and benefits (*villtax*, *villcorrupt*, *villsales*, *villgovs*, and *villcred*), a dummy classifying villages as either rural or urban (*rural*), and district dummies. The estimates on access to credit will also include a dummy for whether or not there is a bank available in the village area (*bank*).

### **3.3.1. Taxes and Levies**

Micro enterprises with less than five employees are mainly subsistence firms, for which the prospect of getting involved with any kind of regular payments will come as a major threat (LabSosio 2008: 53). Thus, analyzing the additional impact of licenses on levies in the form of taxes and unofficial payments ('other levies') will give important information about whether this concern is justified.

Only 47 percent of firms in the sample are paying any taxes at all, and only 27 percent report positive values on paying 'other levies'. The following estimations will thus take into account the possibility of sample selection bias. If the amount of levies paid by firms is dependent on the underlying decision of firms on whether or not to pay any levies at all, OLS estimations of the model will contain a bias that can be considered as a specification error (Heckman 1979). A Heckman selection model takes into account the correlation between selection and outcome stage by including the "inverse Mill's ratio" into the latter, a function representing the inverse probability that an observation is selected into the sample.

### **Taxes**

A potential problem arises when estimating a selection model with two stages that have a lot of explanatory variables in common. The quality of the estimations depends on whether an

---

<sup>19</sup> Hindu managers in Bali are not included as a dummy separately. 95 percent of Balinese managers are Hindu. Thus this majority religion indicator corresponds with the ethnicity variable.

<sup>20</sup> 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>21</sup> Total fixed asset value includes buildings, land, equipment and machinery, furniture, storage facilities, and vehicles.



'exclusion restriction' can be found, i.e. one or more variables explaining the selection but not the outcome stage. If no good instrument is available, the inverse Mill's ratio could be collinear with the other explanatory variables and the results of a selection model will not be robust (Puhani 2000: 57, 65). Different exclusion restrictions were tried for the case of tax payments. However, with the set of potential instruments available in the RICS data, the null hypothesis of no selection bias, i.e. no correlation between the two equations, could not be rejected. In this case, a separate estimation of the outcome stage is more advisable than the Heckman selection model.

The results of estimating the impact of formality on the log of total tax payments by OLS and 2SLS are displayed in Table A5. The OLS estimation (Column 1) that does not control for endogeneity and uses the actual value of licensing 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.

A heterogeneous effect of licensing on taxes can be stated with regard to firm sales. If actual values of log total sales are interacted with licensing a multicollinearity problem arises: the coefficients of formality and the interaction term are jointly significant, but not individually. Using sales quintiles instead alleviates multicollinearity. The results in Column 3 indicate 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. A second heterogeneous licensing effect occurs with gender. Column 4 shows that licensing effect is significantly smaller for firms with female managers. For rural firms the benefits from licensing through tax reduction more than double to those of urban firms (Column 5).

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. Particularly small and rural enterprises substantially benefit from a reduction in tax payments if they are licensed..

A possible explanation for the result that the smallest formal firms benefit most from licensing in terms of taxes could be found in the tax collection process in Indonesia. In his case study on local governance in six Kabupaten, von Luebke (2006: 13) found that tax collection practices are inefficient and to a large extent based on rough estimates or personal negotiations. In this sense, the results above could suggest that licensing increases the 'bargaining power' of very small enterprises in those tax negotiations and thus helps to reduce the amount of taxes paid by these firms.

## **'Other Levies'**

Analogous to tax payments, a Heckman selection model has been estimated for unofficial payments in order to account for a possible selection bias. In this case the null hypothesis of independent equations could be rejected: selection and outcome stage are significantly correlated, suggesting that estimating both stages separately would yield biased results.<sup>22</sup>

However, the difficulty of finding an appropriate excluding instrument prevails also for the case of other levies. Tables A6 and A7 present the results of using two different variables as exclusion restrictions. In the former the firm's number of employees is excluded from the outcome stage. Number of employees is the most 'visible' of all firm size measures while at the same time it does not give indications on firm performance in terms of revenue or profit. 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 flexible in changing location to escape this harassment. The actual amount of other levies paid will, on the other hand, rather depend on firm revenue or profit than the number of employees. Table A7 reports the results of applying managers' gender as an restriction. The predominantly male officers or thugs might be more reserved in approaching female managers and explain the fact that gender enters the selection stage significantly and negative, but not the outcome stage. The results of estimating the Heckman selection model are robust to both specifications and provide almost the same point estimates for all variables.

Columns 1 and 2 (Tables 6 and 7) display the results of Heckman selection with licensing not being instrumented. No significant impact of licensing on payments of 'other levies' can be determined on both stages. In contrast to this, both specifications using the instrumented value of licensing (Columns 3 and 4, Tables 6 and 7) show a significant negative impact of licensing on bribe payments in the selection as well as the outcome stage. A formal firm will be less likely to pay any other levies at all and, for firms that do pay a positive amount of other levies, licensing reduces the sum paid.

This bias in OLS estimates has the expected direction. If the underlying effect of licensing is to reduce corruption payments especially those firms will incur the costs of licensing that would have to pay large bribes otherwise. This will be the large firms that will pay larger bribes also after becoming formal. OLS does not take into account this endogenous positive selection effect of 'other levies' on licensing. As a result, the estimate is upward biased – instead of a negative effect no significant impact can be stated.

---

<sup>22</sup> For reasons of comparison, Table A8 reports the results from estimating 'other levies' with OLS and 2SLS. The overall effect of licensing on unofficial payments is significantly negative. The significant interaction effects in the 2SLS model are female (reducing the benefit from licensing) and the residence dummy (for managers living in the same village, licensing further reduces bribe payment). However, the discussion in the following part will be based on the preferred outcome of the Heckman selection model.

This assumption is supported by the heterogeneous treatment effects of licensing depending on size. Columns 5 and 6 suggest that bribe payments for formal firms decrease significantly with log sales. While for the smallest firms licensing will result in an increase of unofficial payments of approximately 5%, a formal firm with mean log sales will already benefit from formality by a reduction in other levies of about 1%. The same is true for the interaction terms with log fixed assets. Payments of other levies significantly decrease with a firm's log value of fixed assets (Columns 7 and 8).

For rural firms, the effect of licensing on selection and outcome stage is pointing in different directions (Columns 9 and 10). Licensing reduces the likelihood of rural firms to pay any 'other levies' at all even more than it does for urban firms. On the outcome stage however, the results show that once a rural firm pays a positive amount of unofficial levies, licensing will further increase this sum, while urban firms benefit from licensing through reduced payments. Significant interactions also occur with manager religion and ethnicity (Columns 11-14). For managers who belong to the local indigenous group, holding a license decreases the likelihood of paying any 'other levies' at all as compared to managers of other ethnicity. The reduction in bribe payments for formal Muslim managers is much less in the outcome stage than it is for non-Muslim managers. For female managers, formality will reduce the likelihood of paying bribes significantly less. In the outcome stage no significant interaction effect can be stated: licensing reduces bribe payments for female and male managers alike (Columns 15 and 16).

Similar to taxes, the payment of 'other levies' can be considered a benefit of formality: on average, licensing reduces not only the amount of unofficial payments made (outcome stage), but also the likelihood of paying those levies at all (selection stage). However, while in the case of tax payments smaller firms (measured in terms of sales) benefit more from licensing than larger ones, this effect is reversed for firms' payments of 'other levies': here the larger firms, both in terms of sales and fixed asset value, experience a reduction in bribe payments as compared to smaller firms.

### **3.3.2. Firm Performance and Business Expansion**

In section 2.3 we argued that firms can benefit from formality through access to different resources, markets, and the ability to grow. Therefore we analyze whether formality affects positively firm performance and expansion. OLS estimates, however, can be expected to overestimate this impact of formality by failing to take into account the endogeneity problem: Larger firms are more likely to formalize and, conversely, formalization will contribute to firm growth.

Size measures and enterprise age are endogenous to all of the three variables – total sales, share of sales to governments, and access to credit – analyzed in the following. To mitigate this endogeneity problem size and enterprise age will enter the equations as dummies instead of the real variable values.

### **Total revenue**

Total firm revenue is measured as log total sales in 2005. As expected, OLS estimates (Table A9, Column 1) indeed show a significant gain to total firm sales from licensing of 43%. Comparing these results with 2SLS (Column 2) confirms the assumption of an upward bias in OLS estimates stated above. After controlling for endogeneity by including the instrumented value of licensing the results do not show a significant average effect of licensing on total sales anymore.

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 by an increase in total sales of 47%. This finding corresponds with the fact that micro firms are usually subsistence enterprises that would not be able to take advantage of the opportunities opened up by formality to increase revenue, such as the facilitation of trading across village borders or new business relations or establishment of new business relations based on licenses as a sign of legality and trustworthiness (LabSosio 2008: 43ff).

### **Access to government contracts**

For the case of access to government contracts one could again expect an upward bias in OLS estimates caused by endogeneity. 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. Formality will reversely influence the access to government contracts. However, no significant average impact of licensing on the share of sales to governments can be established in either OLS or 2SLS estimates (Table A10, Columns 1 and 2).

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>23</sup> are even selling a lower share of

---

<sup>23</sup> 87 percent of firms in the lowest quintile do not own any fixed assets at all.

output to governments than their informal counterparts. Firms in the highest fixed asset quintile instead do benefit from licensing with an increase in the share of sales made to governments of 1.8 percent. A positive interaction effect of size on government contracts can also be established with 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 6). However only 74 firms (3%) in the RICS sample report positive sales to governments – small and micro firms are not a typical suppliers of the government. Thus a potential effect of licensing on access to government contracts can be expected to be small.

### **Credit**

Access to credit has widely been considered as a central argument in favor of licensing (cf. section 2.3). In fact the regulations of the Bank of Indonesia require a firm to have legal business status in the form of TDP or SIUP in order to apply for a 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 on grants 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 demand lower requirements.

The results of a Probit regression on the access to credit are presented in Table A11. In accordance with the literature licenses turn out not to be decisive for access to credit, neither in the OLS nor 2SLS regression. A heterogeneous impact of licensing on access to credit appears through the firm's number of employees (Column 3). 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 the status of a formal firm, while the opposite holds for the largest firms. A second significant interaction effect of licensing can be observed with firm sector (Column 4). While firms in the service and even more so 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.

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. Merely eight percent of those firms report that their loan requests have been

rejected. This low rejection rate suggests that a main restriction in access to credit seems to be limited information or discouraged borrowers: most firms do not even apply for a loan. In fact 33 percent of firms in the RICS would not apply for a loan from a formal financial institution, because they do not know where or how to apply.

### 3.4. The Determinants of Formality

The empirical analysis presented in the previous section addresses two questions. First, it identified the average effect of licensing on taxes, bribe payments, firm revenue, access to government contracts and access to credit. It turned out that formal status was beneficial for tax and corruption payments, whereas no significant overall effect of licensing on any of the three indicators for firm performance and expansion could be established.

Second, we established that the effect of licensing depend on a number of firm and manager characteristics. These characteristics can in fact be considered as the determinants of firms' decisions to formalize. Table 4 summarizes those characteristics that have shown significant interaction effects with licensing in Tables A5-A11. The signs in parentheses denote whether potential costs or benefits are increasing or decreasing function of the respective interacting characteristic.

**Table 4: The costs and benefits of formality**

<b>Interacting characteristic</b>	<b>cost or benefit</b>
log total sales	taxes (+) 'other levies': outcome stage (-) sales to government (+)
log value of fixed assets	'other levies': outcome stage (-) sales to government (+)
number of employees	credit (-) revenue (+)
female	taxes (+) 'other levies': selection stage (+)
rural	taxes (-) 'other levies': selection stage (-) outcome stage (+)
manufacturing sector	credit (-) sales to government (+)
indethn	'other levies': selection stage (-)
chinese	sales to government (-)
islam	revenue (-) other levies: outcome stage (+)

No variation in licensing effects could be found by enterprise age, manager education, manager age, and managers' place of residence.

The decision to formalize can now be expressed as a reduced form function of these characteristics (cf. equation (5), Section 3.1). The regression also includes time and costs of licensing as a proxy for entry costs in the formal sector. The expected average registration time and cost for each firm was calculated as a weighted mean over employment quartiles, sector and sub-districts<sup>24</sup>, based on the existing data on registration processes provided by the licensed firms in RICS.

The estimation results are presented in Table A11. Again, measures of firm size and age enter the equation as quintile dummies in order to mitigate endogeneity. While the estimates in Column (1) only use the significant interaction characteristics as independent variables, manager education has been added in Column (2). Although no significant interaction effect of education and licensing could be found in Section 3.3, education serves as a proxy for information as a component of entry costs. Higher educated managers will have easier access to lower costs of information, which will lower the costs of entry in the formal sector. Information is of special importance, keeping in mind that most firms are unaware of the package of licenses they are supposed to have (Section 2.2). The coefficient of education shows the expected positive impact on a firm's probability of being formal. Although slightly different in point estimates, the results of the other coefficients are robust to the inclusion of the 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>25</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. This result corresponds to the findings above. 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 governments. Those benefits seem to outweigh the finding that tax payments for large formal firms are higher than for informal firms.

---

<sup>24</sup> For those sub-districts that had no data entries on time and costs, district averages have been used.

<sup>25</sup> A more meaningful indicator would be a variable that captures (un)certainly 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.

Female managers have a lower incentive to go formal: they have a lower gain from formalization in reduced taxes or bribes. However, Table 12 shows that gender does affect the formalization decision significantly. .

For managers who belong to the local indigenous ethnic group, the probability of being formal decreases by about 10 percent. Chinese managers are 35 % more likely to register their firm. These findings seem to be at odds with those of the preceding section For managers of the local ethnicity, licensing resulted in a differential reduction of the probability to pay 'other levies', which actually provides an incentive to formalize. 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. Such an effect could not have been detected in the previous regressions. A similar case could be made for managers that do not live in the place of business. In fact, the coefficient of the residence dummy (*resdum*) shows a significant negative impact on the probability of being formal (Column 3).

For rural firms, licensing reduces the amount of taxes paid and the likelihood of paying other levies at all. On the other hand it might increase the actual amount of 'other levies' paid by rural firms. In the reduced form Probit regression '*rural*' significantly reduces the probability of being formal by more than ten percent. To explain this result, one should take into account that the costs of obtaining licenses will be higher in rural areas. Administrative services and facilities will be more costly to reach and information provision more limited. 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.

#### **4. Conclusion**

In this paper we have analyzed the determinants of going formal for small and micro non-farm enterprises in rural Indonesia. In an instrumental variable approach that accounts for the endogeneity of the legal status we have established that firms have an incentive to get licensed as this reduces tax and corruption payment. This result is contrary to the notion that formalization will increase tax payments as firms appear on the tax authorities' radar screen.

We have 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 and the sector the firm operates in, and whether they are located in rural or semi-rural areas. This applies not only to tax and corruption payments, but also to firm revenue. Access to credit and to government contracts seem largely unaffected by the formal or informal status of a firm. That



is due to the fact that most firms in our sample of small and microfirms are discouraged borrowers and do not sell to the government quite independent of their status.

We have analyzed the probability of going formal. Larger firms are more likely to get licensed as are firms that are owned/ managed by Chinese or more highly educated people. Firms in rural areas or those run by managers who belong to the local indigenous group and those who live in the village their companies is located are less likely to become formal. That points to a different treatment by local authorities who favor locals and members of their own ethnicity and put Chinese managers under closer scrutiny. In the end the assumption of completely voluntary decision to become formalized may not hold in a strict sense.<sup>26</sup>

---

<sup>26</sup> It would have been desirable to account explicitly for differences in enforcement (apart from using district dummies); unfortunately this information is unavailable.

## Literature

- Angeini, 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. In: *The Quarterly Journal of Economics*, Vol. CXVII (1): 1-37.
- 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.
- Hussmanns, 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 Ka: el (2006): Economic Growth and Poverty: Does Formalization of Informal Enterprises Matter? GIGA Working Paper No. 20.
- Jäckle, Annette E. and Carmen A. Lee (2003): Firm Dynamics and Institutional Participation: A Case Study on Informality of Micro-Enterprises in Peru. No 552 Economics Discussion Papers from University of Essex.
- Kenyon, Thomas (2007): A Framework for Thinking about Enterprise Formalization Policies in Developing Countries. World Bank Policy Research Working Paper 4235.
- K: OD (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): 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 Sumatra, Bali and NTB. Paper presented for the Rural Investment Climate Assessment, World Bank, Jakarta, processed.

- McKenzie, David and Seynabou Yaye Sakho (2007): Does it pay firms to register for taxes? The impact of formality on firm profitability. IZA Discussion Paper Series Np. 3179.
- 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.
- Enste, Dominik and Friedrich Schneider (2000) Shadow Economies: Sizes, Causes, and Consequences, *Journal of Economic Literature*, Vol. 38: 77-114.
- 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.
- The World Bank (2008): The Rural Investment Climate. Analysis and Findings. Report No.44009-GLB.
- The World Bank and the International Financial Corporation (2007): Doing Business 2008: Country Report Indonesia.
- The World Bank Office Jakarta (2006): Revitalizing the rural economy: An assessment of the investment climate faced by non-farm enterprises at the district level. Consultative Draft.  
<http://siteresources.worldbank.org/INTINDONESIA/Resources/Publication/280016-1152870963030/RICAFINALE.pdf> (09/04/2008)
- Wooldridge, Jeffrey M. (2002): *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, London.

## 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
com_form	dummy =1 for firms that hold at least one of the four most common local licenses (TDI, SIUP, IMB or TDP)
corrptdum	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)
edu	manager education (ordinal scale)
empldum2	dummy for firms with >5 employees
employee	total number of employees
ent_age	enterprise age
entagedum2	dummy = 1 for firms longer in operation than median
fassdum2	dummy = 1 for firms with fixed assets above median
female	dummy =1 for female manager
ful_form	dummy =1 for firms being fully formalized
govsales	share of sales made to government
indethn	dummy =1 for manager who belongs to a local indigenous group
int_*	prefix indicating interaction term with fitted value of com_form
islam	dummy = 1 for muslim manager
Incorrpt2	log of total 'other levies' at central, provincial and district level
Infasset	log of total fixed assets 2005 (includes land, buildings, equipment, furniture, vehicles)
Insales	log of total sales 2005
Inttax2	log of total taxes at central, provincial and district level
loc_form	dummy = 1 for firms formalized at the local level
mannufac	dummy for firm in manufacturing sector
meancost	average expected registration costs averaged over employment, location and sector
meantime	expected registration time, averaged over employment, location and sector
qemployee1-5	dummies for employee quintiles
qent_age	dummies for ent_age quintiles
qInfasset1-5	dummies for log fixed asset quintiles
qInsales	quintiles for Insales
qInsales1-5	dummies for log sales quintiles
resdum	dummy =1 for manager who lives in the same village as the business operate
rural	dummy for rural =1 and urban =0
saledum2	dummy = 1for firms with sales above median

service	dummy for firm in service sector
taxdum	dummy for firms with positive tax payments
villcom	village average of com_form
villcorrpt2	village average of Incorrpt2
villcred	village average of credit
villgovs	village average of govsales
villsales	village average of Insales
villtax2	village average of Inttax2

**Table A2.1 Descriptive Statistics**

Variable Name	Obs.	Mean	St.Dev.	Min	Max
ful_form	2461	0.02	0.138	0	1
loc_form	2461	0.029	0.167	0	1
com_form	2308	0.234	0.424	0	1
villcom	2183	0.230	0.231	0	1
taxdum	1155	5.251	1.959	0	12.11
Inttax2	2461	2.479	2.942	0	12.11
villtax2	2461	2.479	1.599	0	11.98
villtax	2326	5.159	1.228	0.23	11.98
corrptdum	673	4.805	1.756	0	12.26
Incorrpt2	2461	1.322	2.334	0	12.26
villcorrpt2	2461	1.322	1.431	0	8.01
villcorrpt	2013	4.584	1.398	0.69	9.16
age	2371	41.64	11.79	16	85
female	2380	0.3744	0.484	0	1
indethn	2370	0.659	0.474	0	1
islam	2380	0.858	0.349	0	1
resdum	2173	0.862	0.345	0	1
chinese	2370	0.017	0.130	0	1
employee	2400	2.58	2.235	1	19
ent_age	2461	9.063	8.783	1	66
Insales	2382	10.177	1.624	3.76	17.74
govsales	2461	1.156	8.687	0	100
villgovs	2461	1.156	4.387	0	100
Infasset	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
villcred	2427	0.196	0.193	0	1
meantime	2303	10.264	8.815	1	60
meancost	2281	550.444	467.385	5	2000

**Table A2.2 Education**

Owner/Manager Last Level of Education	Frequency	Percent
no schooling	317	13.32
elementary school	729	30.64
junior high school	436	18.33
high school	606	25.47
vocational school	103	4.33
diploma I/II	32	1.35
diploma III	29	1.22
diploma IV/ B.A.	121	5.09
Masters/PhD	6	0.25
Total	2,379	100.00

**Table A3. Correlation Matrix**

	com_form	villcom	lnttax2	villtax2	lncorr~2	villco~2	lnsales
com_form	1.0000						
villcom	0.5556*	1.0000					
lnttax2	0.3677*	0.3199*	1.0000				
villtax2	0.3194*	0.5826*	0.5435*	1.0000			
lncorrpt2	0.1640*	0.1210*	0.3173*	0.2774*	1.0000		
villcorrpt2	0.1016*	0.2058*	0.2459*	0.4525*	0.6131*	1.0000	
lnsales	0.2922*	0.1878*	0.3392*	0.2137*	0.3025*	0.2557*	1.0000
villsales	0.1939*	0.3859*	0.2388*	0.4410*	0.3209*	0.5252*	0.4834*
govsales	0.1494*	0.1032*	0.1471*	0.1133*	0.0346	0.0403*	0.0904*
villgovs	0.1305*	0.1966*	0.1219*	0.2243*	0.0489*	0.0798*	0.0368
credit	0.0858*	0.0390	0.1280*	0.1109*	0.0783*	0.1153*	0.0519*
villcred	0.0046	0.0188	0.1200*	0.2224*	0.1806*	0.2913*	0.0451*
edu	0.3328*	0.2957*	0.2924*	0.3065*	0.1411*	0.1923*	0.2620*
age	0.0642*	0.0265	0.0198	-0.0087	-0.0626*	-0.0743*	-0.0566*
female	-0.1105*	-0.0632*	-0.1290*	-0.0876*	-0.0593*	-0.0025	-0.0854*
resdum	-0.1340*	-0.0943*	-0.1410*	-0.1508*	-0.2327*	-0.2727*	-0.2105*
indethn	-0.0207	-0.0483*	-0.0294	-0.0678*	-0.0536*	-0.1107*	-0.1470*
chinese	0.1868*	0.2187*	0.1459*	0.1539*	0.0812*	0.0678*	0.1438*
islam	-0.1370*	-0.1770*	-0.1855*	-0.2733*	-0.3399*	-0.4899*	-0.1683*
lnfasset	0.2410*	0.1963*	0.2649*	0.1828*	0.1181*	0.0392	0.1444*
employee	0.3199*	0.1557*	0.3102*	0.1932*	0.2405*	0.1580*	0.3859*
rural	-0.1900*	-0.3539*	-0.2228*	-0.4100*	-0.1395*	-0.2275*	-0.1486*
manufac	0.0142	-0.0706*	-0.0176	-0.0847*	-0.0460*	-0.0676*	-0.0235
service	0.0070	0.0325	0.0287	0.0483*	-0.0094	-0.0330	-0.2000*
bank	0.1481*	0.2693*	0.1142*	0.2201*	0.0146	0.0242	-0.0322

	villsa~s	govsales	villgovs	credit	villcred	edu	age
villsales	1.0000						
govsales	0.0315	1.0000					
villgovs	0.0681*	0.5051*	1.0000				
credit	0.0252	-0.0089	0.0246	1.0000			
villcred	0.0952*	0.0188	0.0370	0.4463*	1.0000		
edu	0.2293*	0.1713*	0.1209*	0.0004	0.0695*	1.0000	
age	-0.0582*	-0.0017	-0.0033	0.0057	-0.0219	-0.2968*	1.0000
female	-0.0642*	-0.0492*	-0.0246	0.0327	0.0069	-0.1220*	-0.0018
resdum	-0.1804*	-0.0181	-0.0558*	-0.0545*	-0.1237*	-0.1965*	0.0853*
indethn	-0.2827*	0.0265	0.0466*	0.0192	-0.0010	0.0127	0.0179
chinese	0.1576*	0.0208	0.0579*	-0.0132	-0.0029	0.0665*	0.0654*
islam	-0.2701*	-0.0632*	-0.1443*	-0.1061*	-0.2123*	-0.2397*	0.0568*
lnfasset	0.0908*	0.0793*	0.0696*	0.0598*	-0.0855*	0.1793*	0.0178
employee	0.1822*	0.1957*	0.1406*	0.0725*	0.0748*	0.2298*	0.0468*
rural	-0.3144*	-0.0391	-0.0775*	-0.1075*	-0.2819*	-0.2175*	-0.0101
manufac	-0.0989*	0.2046*	0.0806*	-0.0605*	-0.0348	-0.0191	0.0436*
service	-0.0657*	-0.0064	0.0061	-0.0186	0.0380	0.0677*	-0.0067
bank	-0.0707*	0.0485*	0.1435*	0.1152*	0.2261*	0.0976*	0.0233
	female	resdum	indethn	chinese	islam	lnfasset	employee
female	1.0000						
resdum	0.0301	1.0000					
indethn	0.0495*	0.0610*	1.0000				
chinese	-0.0294	-0.0946*	-0.1847*	1.0000			
islam	-0.0521*	0.2623*	-0.0731*	-0.2908*	1.0000		
lnfasset	-0.1021*	0.0102	-0.0134	0.0459*	-0.0489*	1.0000	
employee	-0.1215*	-0.1838*	-0.0141	0.1440*	-0.1530*	0.2049*	1.0000
rural	-0.0041	0.1424*	-0.0651*	-0.0799*	0.1879*	0.0565*	-0.0986*
manufac	-0.1342*	-0.0361	0.0525*	-0.0011	0.0234	0.0815*	0.1330*
service	-0.2056*	0.0316	0.0187	-0.0242	0.0319	0.0791*	0.0077
bank	-0.0160	0.0269	0.1745*	0.0470*	-0.0725*	-0.0063	0.0435*
	rural	manufac	service	bank			
rural	1.0000						
manufac	0.0785*	1.0000					
service	-0.0398*	-0.2487*	1.0000				
bank	-0.3111*	-0.0041	0.0516*	1.0000			

**Table A4. First Stage IV Estimates**

Marginal effects from Probit regression on com\_form

	(1)	(2)	(4)	(5)	(6)
villcom	0.562 (0.086)***	0.547 (0.086)***	0.538 (0.081)***	0.582 (0.084)***	0.531 (0.094)***
edu	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)
resdum	-0.246 (0.076)***	-0.244 (0.077)***	-0.247 (0.074)***	-0.242 (0.074)***	-0.253 (0.074)***
indethn	-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)
empldum2	0.085 (0.046)*	0.086 (0.046)*	0.108 (0.041)***	0.087 (0.046)*	0.081 (0.044)*
ivsaledum2	0.044 (0.032)	0.041 (0.033)		0.039 (0.033)	0.044 (0.033)
ivfassdum2	0.069 (0.018)***	0.070 (0.018)***	0.069 (0.016)***	0.070 (0.017)***	0.068 (0.018)***
entagedum2	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)
manufac	-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)
villtax2	-0.002 (0.007)				
villcorrpt2		0.007 (0.012)			
villsales			0.018 (0.018)		
villgovs				-0.008 (0.004)**	
villcred					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 A5. Total Taxes**

	(1) OLS	(2) 2SLS	(3) 2SLS	(4) 2SLS	(5) 2SLS
com_form	1.045 (0.381)***	-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)**		
edu	0.342 (0.060)***	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	-0.404 (0.228)*	-0.359 (0.213)*	-0.389 (0.193)**	-0.667 (0.242)***	-0.384 (0.220)*
resdum	0.271 (0.373)	-0.439 (0.544)	-0.375 (0.585)	-0.424 (0.506)	-0.459 (0.546)
indethn	-0.189 (0.270)	-0.374 (0.251)	-0.470 (0.257)*	-0.378 (0.249)	-0.387 (0.251)
chinese	0.826 (0.636)	0.705 (0.880)	0.327 (0.910)	0.135 (0.817)	0.550 (0.946)
islam	-0.469 (0.317)	-0.587 (0.345)*	-0.676 (0.334)**	-0.566 (0.348)	-0.628 (0.353)*
employee	0.105 (0.036)***	0.099 (0.045)**	0.082 (0.044)*	0.112 (0.048)**	0.098 (0.045)**
lnsales	0.251 (0.083)***	0.347 (0.059)***		0.330 (0.058)***	0.355 (0.061)***
lnfasset	0.087 (0.037)**	0.089 (0.031)***	0.092 (0.028)***	0.086 (0.032)***	0.091 (0.031)***
ent_age	-0.021 (0.015)	-0.010 (0.015)	-0.009 (0.014)	-0.008 (0.015)	-0.010 (0.014)
rural	-0.298 (0.297)	0.099 (0.202)	-0.014 (0.205)	0.128 (0.203)	0.255 (0.241)
manufac	0.247 (0.465)	0.579 (0.323)*	0.656 (0.289)**	0.569 (0.337)*	0.601 (0.330)*
service	0.251 (0.215)	0.225 (0.201)	0.149 (0.175)	0.209 (0.205)	0.227 (0.197)
villtax2		0.804 (0.062)***	0.819 (0.061)***	0.802 (0.063)***	0.797 (0.061)***
Constant	-3.056	-3.950	-0.572	-3.702	-4.088

	(1.196)**	(1.467)***	(1.357)	(1.481)**	(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>0.33</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 A6.1. 'Other Levies' – Heckman Selection (1)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Incorrpt2	corrptdum	Incorrpt2	corrptdum	Incorrpt2	corrptdum	Incorrpt2	corrptdum
com_form	-0.190 (0.307)	0.109 (0.155)	-1.968 (0.958)**	-1.457 (0.464)***	8.225 (2.697)***	-2.150 (1.445)	2.613 (1.632)	-1.368 (0.808)*
int_Insales					-0.902 (0.234)***	0.053 (0.123)		
int_Infasset							-0.402 (0.112)***	-0.007 (0.061)
edu	0.014 (0.083)	-0.097 (0.042)**	0.146 (0.091)	0.008 (0.055)	0.150 (0.087)*	0.013 (0.055)	0.104 (0.085)	0.005 (0.055)
age	0.009 (0.012)	-0.014 (0.006)**	0.013 (0.013)	-0.010 (0.007)	0.007 (0.013)	-0.010 (0.007)	0.019 (0.012)	-0.010 (0.007)
female	-0.275 (0.307)	-0.234 (0.147)	-0.396 (0.307)	-0.221 (0.150)	-0.383 (0.277)	-0.229 (0.150)	-0.259 (0.290)	-0.221 (0.149)
resdum	-0.551 (0.308)*	-0.088 (0.160)	-1.165 (0.447)***	-0.664 (0.208)***	-1.417 (0.415)***	-0.662 (0.211)***	-0.869 (0.446)*	-0.645 (0.207)***
indethn	-0.042 (0.338)	0.011 (0.139)	-0.182 (0.308)	0.015 (0.170)	-0.303 (0.282)	0.023 (0.171)	-0.116 (0.286)	0.048 (0.169)
chinese	0.755 (0.621)	0.471 (0.411)	1.244 (0.705)*	0.512 (0.529)	1.399 (0.565)**	0.635 (0.500)	1.242 (0.635)*	0.484 (0.466)
islam	0.291 (0.397)	0.250 (0.286)	0.069 (0.402)	-0.235 (0.319)	-0.123 (0.372)	-0.196 (0.313)	-0.155 (0.349)	-0.198 (0.299)
Insales	0.070 (0.089)	0.174 (0.043)***	0.170 (0.095)*	0.130 (0.043)***	0.394 (0.118)***	0.111 (0.060)*	0.159 (0.088)*	0.127 (0.043)***
Infasset	0.045 (0.034)	-0.017 (0.017)	0.079 (0.033)**	0.011 (0.017)	0.080 (0.031)***	0.010 (0.017)	0.150 (0.034)***	0.009 (0.020)
ent_age	0.005 (0.015)	-0.007 (0.008)	0.006 (0.019)	-0.010 (0.009)	0.005 (0.019)	-0.010 (0.009)	0.008 (0.019)	-0.010 (0.009)
rural	-0.537 (0.240)**	0.329 (0.133)**	-0.525 (0.269)*	0.122 (0.139)	-0.371 (0.268)	0.099 (0.142)	-0.415 (0.277)	0.125 (0.139)
manufac	0.750 (0.371)**	-0.459 (0.206)**	0.713 (0.406)*	-0.367 (0.218)*	0.734 (0.462)	-0.361 (0.220)	0.996 (0.473)**	-0.403 (0.223)*
service	0.413	0.120	0.474 (0.283)*	-0.161 (0.148)	0.556 (0.261)**	-0.168 (0.149)	0.623 (0.257)**	-0.175 (0.148)
villcorrpt2			0.271 (0.227)	0.721 (0.070)***	0.337 (0.212)	0.724 (0.070)***	0.264 (0.201)	0.719 (0.070)***
employee		0.065 (0.028)**		0.073 (0.033)**		0.071 (0.036)**		0.077 (0.031)**
Constant	3.579 (1.390)**	-1.474 (0.722)**	2.056 (1.492)	-0.747 (0.783)	0.502 (1.463)	-1.818 (0.801)**	2.450 (1.223)**	-0.792 (0.766)
<b>Observations</b>	<b>1901</b>	<b>1901</b>	<b>1782</b>	<b>1782</b>	<b>1782</b>	<b>1782</b>	<b>1782</b>	<b>1782</b>

Incorrpt2: outcome stage  
corrptdum: selection stage

Robust standard errors in parentheses

Regression contains district dummies

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

Table A6.2. 'Other Levies' – Heckman Selection (1 contd.)

	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	Incorrpt2	corrptdum	Incorrpt2	corrptdum	Incorrpt2	corrptdum	Incorrpt2	corrptdum
com_form	-2.159 (0.943)**	-1.300 (0.472)***	-2.279 (1.166)*	-0.596 (0.516)	-3.244 (1.045)***	-1.719 (0.631)***	-1.894 (0.967)*	-1.663 (0.469)***
int_rural	2.805 (0.952)***	-1.242 (0.641)*						
int_indethn			0.494 (0.959)	-1.178 (0.494)**				
int_islam					2.592 (0.889)***	0.236 (0.604)		
int_female							0.296 (1.030)	0.938 (0.489)*
edu	0.133 (0.092)	0.010 (0.055)	0.145 (0.091)	0.012 (0.054)	0.121 (0.090)	0.012 (0.053)	0.141 (0.092)	0.010 (0.053)
age	0.011 (0.013)	-0.009 (0.007)	0.012 (0.013)	-0.009 (0.007)	0.014 (0.014)	-0.010 (0.007)	0.013 (0.013)	-0.010 (0.007)
female	-0.365 (0.302)	-0.237 (0.150)	-0.404 (0.300)	-0.205 (0.148)	-0.376 (0.288)	-0.233 (0.148)	-0.454 (0.368)	-0.364 (0.197)*
resdum	-1.178 (0.442)***	-0.692 (0.206)***	-1.200 (0.434)***	-0.592 (0.206)***	-1.248 (0.435)***	-0.665 (0.208)***	-1.105 (0.465)**	-0.637 (0.208)***
indethn	-0.126 (0.304)	-0.012 (0.167)	-0.307 (0.402)	0.235 (0.210)	-0.200 (0.292)	0.037 (0.168)	-0.187 (0.305)	0.018 (0.170)
chinese	1.359 (0.684)**	0.335 (0.515)	1.451 (0.806)*	-0.142 (0.558)	1.908 (0.703)***	0.700 (0.455)	1.048 (0.829)	0.242 (0.617)
islam	0.121 (0.396)	-0.281 (0.316)	0.090 (0.399)	-0.294 (0.320)	-0.916 (0.465)**	-0.225 (0.371)	0.072 (0.395)	-0.243 (0.321)
insales	0.162 (0.093)*	0.137 (0.043)***	0.173 (0.096)*	0.130 (0.043)***	0.160 (0.095)*	0.130 (0.042)***	0.160 (0.094)*	0.120 (0.044)***
infasset	0.078 (0.033)**	0.013 (0.017)	0.079 (0.033)**	0.014 (0.017)	0.084 (0.031)***	0.011 (0.017)	0.078 (0.033)**	0.010 (0.017)
ent_age	0.003 (0.019)	-0.010 (0.009)	0.005 (0.019)	-0.010 (0.009)	0.002 (0.019)	-0.010 (0.009)	0.006 (0.019)	-0.010 (0.009)
rural	-0.827 (0.310)***	0.261 (0.170)	-0.503 (0.272)*	0.094 (0.140)	-0.497 (0.268)*	0.125 (0.140)	-0.525 (0.274)*	0.141 (0.140)
manufac	0.653 (0.432)	-0.365 (0.216)*	0.680 (0.397)*	-0.324 (0.220)	0.867 (0.460)*	-0.371 (0.221)*	0.723 (0.407)*	-0.371 (0.215)*
service	0.468 (0.281)*	-0.165 (0.149)	0.481 (0.282)*	-0.171 (0.148)	0.514 (0.272)*	-0.162 (0.148)	0.474 (0.282)*	-0.171 (0.149)
villcorrpt2	0.261 (0.211)	0.724 (0.070)***	0.262 (0.208)	0.725 (0.070)***	0.337 (0.193)*	0.722 (0.071)***	0.218 (0.228)	0.722 (0.070)***
employee		0.077 (0.033)**		0.067 (0.032)**		0.075 (0.031)**		0.080 (0.034)**
Constant	2.366 (1.452)	-0.855 (0.773)	2.205 (1.411)	-0.960 (0.761)	2.417 (1.414)*	-0.799 (0.793)	2.357 (1.450)	-0.689 (0.794)
<b>Observations</b>	<b>1782</b>	<b>1782</b>	<b>1782</b>	<b>1782</b>	<b>1782</b>	<b>1782</b>	<b>1782</b>	<b>1782</b>

Incorrpt2: outcome stage  
corrptdum: selection stage

Robust standard errors in parentheses  
Regression contains district dummies

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

**Table A7.1. 'Other Levies' – Heckman Selection (2)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	lnccrpt1	corrptdum	lnccrpt2	corrptdum	lnccrpt2	corrptdum	lnccrpt2	corrptdum
com_form	-0.155 (0.309)	0.100 (0.157)	-2.095 (1.035)**	-1.427 (0.455)***	9.113 (2.667)***	-2.270 (1.466)	2.650 (1.614)	-1.429 (0.801)*
int_Insales					-1.000 (0.230)***	0.066 (0.125)		
int_Infasset							-0.432 (0.104)***	0.000 (0.061)
edu	-0.026 (0.075)	-0.089 (0.041)**	0.128 (0.091)	0.010 (0.053)	0.124 (0.086)	0.016 (0.054)	0.079 (0.086)	0.010 (0.054)
age	0.008 (0.012)	-0.014 (0.006)**	0.015 (0.013)	-0.011 (0.007)	0.008 (0.013)	-0.010 (0.007)	0.021 (0.012)*	-0.011 (0.007)
resdum	-0.486 (0.310)	-0.092 (0.162)	-1.089 (0.448)**	-0.664 (0.208)***	-1.375 (0.411)***	-0.664 (0.212)***	-0.824 (0.440)*	-0.658 (0.206)***
indethn	-0.064 (0.337)	0.019 (0.139)	-0.250 (0.313)	0.026 (0.169)	-0.390 (0.289)	0.033 (0.171)	-0.167 (0.294)	0.052 (0.169)
Chinese	0.599 (0.581)	0.543 (0.393)	1.093 (0.660)*	0.532 (0.497)	1.303 (0.510)**	0.670 (0.482)	1.229 (0.594)**	0.556 (0.456)
islam	0.228 (0.402)	0.280 (0.279)	0.002 (0.398)	-0.207 (0.311)	-0.233 (0.369)	-0.161 (0.308)	-0.257 (0.350)	-0.151 (0.299)
employee	0.075 (0.054)	0.048 (0.030)	0.070 (0.064)	0.062 (0.033)*	0.090 (0.064)	0.059 (0.033)*	0.083 (0.060)	0.061 (0.032)*
Insales	0.034 (0.096)	0.182 (0.044)***	0.141 (0.098)	0.137 (0.043)***	0.378 (0.118)***	0.114 (0.060)*	0.123 (0.088)	0.135 (0.043)***
Infasset	0.045 (0.034)	-0.016 (0.017)	0.085 (0.034)**	0.012 (0.017)	0.086 (0.031)***	0.011 (0.017)	0.160 (0.034)***	0.008 (0.020)
ent_age	0.002 (0.015)	-0.006 (0.008)	0.004 (0.017)	-0.010 (0.009)	0.003 (0.016)	-0.010 (0.009)	0.006 (0.017)	-0.009 (0.009)
rural	-0.524 (0.248)**	0.339 (0.132)**	-0.530 (0.274)*	0.130 (0.138)	-0.365 (0.265)	0.104 (0.141)	-0.415 (0.276)	0.136 (0.139)
manufac	0.686 (0.368)*	-0.456 (0.203)**	0.695 (0.406)*	-0.379 (0.216)*	0.663 (0.443)	-0.368 (0.220)*	0.917 (0.458)**	-0.405 (0.222)*
service	0.434 (0.283)	0.112 (0.137)	0.515 (0.275)*	-0.170 (0.145)	0.584 (0.251)**	-0.174 (0.147)	0.629 (0.245)**	-0.177 (0.146)
villcorrpt2			0.245 (0.204)	0.719 (0.069)***	0.326 (0.198)*	0.723 (0.070)***	0.286 (0.187)	0.718 (0.070)***
female		-0.294 (0.123)**		-0.287 (0.134)**		-0.283 (0.141)**		-0.275 (0.134)**
Constant	3.579 (1.390)**	-1.474 (0.722)**	3.585 (1.344)***	-0.802 (0.785)	0.525 (1.402)	-1.800 (0.797)**	0.888 (1.211)	-0.845 (0.776)
<b>Observations</b>	<b>1901</b>	<b>1901</b>	<b>1782</b>	<b>1782</b>	<b>1782</b>	<b>1782</b>	<b>1782</b>	<b>1782</b>

lnccrpt2: outcome stage  
corrptdum: selection stage

Robust standard errors in parentheses  
Regression contains district dummies

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

**Table A7.2. ‘Other Levies’ – Heckman Selection (2 contd.)**

	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	lncorrpt2	corrptdum	lncorrpt2	corrptdum	lncorrpt2	corrptdum	lncorrpt2	corrptdum
com_form	-2.415 (1.024)**	-1.270 (0.466)***	-2.485 (1.171)**	-0.571 (0.505)	-3.325 (1.167)***	-1.705 (0.635)***	-1.924 (1.015)*	-1.660 (0.460)***
int_rural	3.391 (0.950)***	-1.309 (0.644)**						
int_indethn			0.584 (0.932)	-1.177 (0.494)**				
int_islam					2.539 (0.882)***	0.249 (0.604)		
int_female							-0.337 (0.799)	1.083 (0.462)**
edu	0.108 (0.091)	0.013 (0.054)	0.127 (0.091)	0.014 (0.053)	0.106 (0.087)	0.014 (0.052)	0.124 (0.090)	0.012 (0.052)
age	0.013 (0.013)	-0.010 (0.007)	0.014 (0.013)	-0.009 (0.007)	0.015 (0.014)	-0.010 (0.007)	0.015 (0.013)	-0.010 (0.007)
resdum	-1.121 (0.440)**	-0.694 (0.206)***	-1.133 (0.432)***	-0.591 (0.205)***	-1.177 (0.451)***	-0.663 (0.208)***	-1.064 (0.463)**	-0.640 (0.206)***
indethn	-0.179 (0.305)	-0.003 (0.166)	-0.402 (0.404)	0.248 (0.209)	-0.260 (0.302)	0.049 (0.167)	-0.239 (0.311)	0.027 (0.170)
chinese	1.281 (0.645)**	0.369 (0.491)	1.342 (0.753)*	-0.107 (0.533)	1.753 (0.661)***	0.738 (0.451)	1.193 (0.841)	0.210 (0.568)
islam	0.052 (0.388)	-0.247 (0.309)	0.024 (0.395)	-0.264 (0.312)	-0.952 (0.484)**	-0.197 (0.372)	0.004 (0.397)	-0.212 (0.313)
employee	0.082 (0.065)	0.063 (0.033)*	0.072 (0.062)	0.055 (0.032)*	0.061 (0.067)	0.064 (0.032)**	0.066 (0.061)	0.068 (0.034)**
lnsales	0.125 (0.096)	0.145 (0.043)***	0.144 (0.099)	0.137 (0.043)***	0.136 (0.096)	0.136 (0.042)***	0.137 (0.095)	0.126 (0.044)***
lnfasset	0.084 (0.033)**	0.014 (0.017)	0.084 (0.034)**	0.014 (0.017)	0.089 (0.032)***	0.011 (0.016)	0.086 (0.034)**	0.010 (0.017)
ent_age	0.000 (0.017)	-0.010 (0.009)	0.004 (0.017)	-0.010 (0.009)	0.001 (0.018)	-0.010 (0.009)	0.004 (0.017)	-0.010 (0.009)
rural	-0.899 (0.315)***	0.276 (0.170)	-0.503 (0.277)*	0.101 (0.139)	-0.497 (0.272)*	0.133 (0.140)	-0.559 (0.280)**	0.154 (0.139)
manufac	0.578 (0.432)	-0.373 (0.214)*	0.653 (0.401)	-0.335 (0.220)	0.859 (0.450)*	-0.384 (0.220)*	0.666 (0.413)	-0.377 (0.213)*
service	0.488 (0.272)*	-0.173 (0.146)	0.522 (0.275)*	-0.180 (0.145)	0.557 (0.258)**	-0.173 (0.145)	0.497 (0.275)*	-0.177 (0.146)
villcorrpt2	0.249 (0.194)	0.723 (0.070)***	0.238 (0.189)	0.723 (0.069)***	0.313 (0.178)*	0.719 (0.071)***	0.202 (0.205)	0.720 (0.070)***
female		-0.303 (0.134)**		-0.275 (0.132)**		-0.305 (0.131)**		-0.453 (0.171)***
Constant	3.867 (1.309)***	-0.925 (0.777)	3.725 (1.280)***	-1.017 (0.764)	4.451 (1.352)***	-0.854 (0.798)	3.773 (1.291)***	-0.738 (0.796)
<b>Observations</b>	<b>1782</b>	<b>1782</b>	<b>1782</b>	<b>1782</b>	<b>1782</b>	<b>1782</b>	<b>1782</b>	<b>1782</b>

lncorrpt2: outcome stage

corrptdum: selection stage

Robust standard errors in parentheses

Regression contains district dummies

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

**Table A8. 'Other Levies' – No Selection**

	(1)	(2)	(3)	(4)
	OLS	2SLS	2SLS	2SLS
com_form	0.264 (0.300)	-1.806 (0.448)***	-2.053 (0.465)***	-0.938 (0.794)
int_female			0.894 (0.504)*	
int_resdum				-1.069 (0.639)*
edu	-0.128 (0.049)**	0.013 (0.063)	0.015 (0.062)	0.014 (0.063)
age	-0.016 (0.008)**	-0.009 (0.006)	-0.009 (0.007)	-0.008 (0.006)
female	-0.422 (0.203)**	-0.388 (0.184)**	-0.544 (0.234)**	-0.385 (0.182)**
resdum	-0.350 (0.196)*	-1.194 (0.300)***	-1.188 (0.290)***	-0.673 (0.526)
indethn	-0.110 (0.221)	-0.085 (0.139)	-0.089 (0.142)	-0.089 (0.137)
chinese	1.695 (0.818)**	1.674 (0.794)**	1.391 (0.765)*	1.546 (0.745)**
islam	0.845 (0.403)**	0.170 (0.507)	0.183 (0.511)	0.174 (0.530)
employee	0.108 (0.047)**	0.107 (0.045)**	0.114 (0.042)***	0.101 (0.047)**
lnsales	0.203 (0.061)***	0.127 (0.040)***	0.119 (0.039)***	0.125 (0.041)***
lnfasset	-0.004 (0.026)	0.030 (0.018)	0.028 (0.018)	0.030 (0.018)*
ent_age	-0.003 (0.010)	-0.006 (0.011)	-0.005 (0.011)	-0.006 (0.011)
rural	0.287 (0.240)	0.036 (0.122)	0.052 (0.122)	0.005 (0.117)
manufac	-0.347 (0.269)	-0.173 (0.192)	-0.180 (0.194)	-0.175 (0.191)
service	0.241 (0.205)	-0.005 (0.152)	-0.012 (0.149)	-0.009 (0.154)
villcorrupt2		1.032 (0.084)***	1.026 (0.085)***	1.034 (0.083)***
Constant	-1.198 (0.948)	-0.011 (1.037)	0.072 (1.063)	-0.493 (1.175)
<b>Observations</b>	<b>1901</b>	<b>1782</b>	<b>1782</b>	<b>1782</b>
<b>R-squared</b>	<b>0.41</b>	<b>0.54</b>	<b>0.54</b>	<b>0.54</b>

Robust standard errors in parentheses

Regression contains district dummies

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

**Table A9. Total Firm Revenue 2005**

	(1)	(2)	(3)
	OLS	2SLS	2SLS
com_form	0.424 (0.233)*	-0.379 (0.291)	-1.226 (0.349)***
int_qemployee_2			0.701 (0.594)
int_qemployee_4			0.516 (0.502)
int_qemployee_5			1.699 (0.465)***
int_islam			
edu	0.084 (0.030)***	0.129 (0.028)***	0.123 (0.027)***
age	-0.009 (0.003)**	-0.005 (0.005)	-0.005 (0.005)
female	-0.193 (0.146)	-0.108 (0.139)	-0.110 (0.141)
indethn	-0.294 (0.164)*	-0.089 (0.152)	-0.108 (0.161)
chinese	-0.057 (0.600)	0.164 (0.434)	-0.197 (0.461)
islam	0.089 (0.200)	-0.019 (0.155)	-0.075 (0.162)
resdum	-0.260 (0.176)	-0.225 (0.247)	-0.163 (0.226)
employee_2	0.120 (0.178)	0.221 (0.148)	0.111 (0.128)
employee_4	0.391 (0.176)**	0.575 (0.158)***	0.502 (0.194)**
qemployee_5	1.364 (0.204)***	1.503 (0.207)***	1.004 (0.287)***
qInfasset_2	-0.179 (0.176)	-0.191 (0.156)	-0.156 (0.155)
qInfasset_3	-0.267 (0.150)*	-0.189 (0.137)	-0.149 (0.135)
qInfasset_4	0.064 (0.148)	-0.056 (0.153)	-0.013 (0.160)
qInfasset_5	0.097 (0.139)	0.113 (0.130)	0.137 (0.130)
entagedum_2	0.445 (0.107)***	0.436 (0.107)***	0.438 (0.107)***
entagedum_3	0.434 (0.141)***	0.475 (0.172)***	0.482 (0.166)***
rural	-0.242 (0.160)	0.175 (0.139)	0.135 (0.138)
service	-0.723 (0.115)***	-0.677 (0.114)***	-0.705 (0.110)***
manufac	-0.363 (0.187)*	-0.316 (0.168)*	-0.338 (0.173)*
villsales		0.836 (0.095)***	0.863 (0.091)***
Constant	10.630	1.421	1.278



	(0.441)***	(1.180)	(1.111)
<b>Observations</b>	<b>1901</b>	<b>1782</b>	<b>1782</b>
<b>R-squared</b>	<b>0.39</b>	<b>0.45</b>	<b>0.46</b>

Robust standard errors in parentheses

Regression contains district dummies

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

**Table A10. Share of Total Sales to Governments**

	(1)	(2)	(3)	(4)	(5)
	OLS	2SLS	2SLS	2SLS	2SLS
com_form	1.606 (1.018)	0.766 (1.671)	-2.828 (2.724)	-4.520 (2.207)**	-0.726 (2.313)
int_qlnsale2			5.558 (4.608)		
int_qlnsale3			15.354 (9.187)*		
int_qlnsale4			0.919 (2.802)		
int_qlnsale5			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_manufac					37.179 (8.161)***
int_service					1.192 (2.661)
edu	0.625 (0.276)**	0.688 (0.389)*	0.734 (0.360)**	0.725 (0.393)*	0.413 (0.404)
age	0.027 (0.035)	0.034 (0.035)	0.032 (0.034)	0.035 (0.034)	0.005 (0.029)
female	0.373 (0.527)	0.361 (0.632)	0.499 (0.623)	0.228 (0.678)	0.344 (0.598)
resdum	1.680 (1.004)*	2.231 (1.725)	1.923 (1.506)	1.973 (1.804)	2.728 (1.948)
indethn	-0.844 (0.615)	-1.166 (0.735)	-1.085 (0.675)	-1.225 (0.680)*	-1.233 (0.709)*
Chinese	-2.148 (1.059)**	-1.223 (1.160)	-0.068 (1.550)	-1.201 (1.275)	1.088 (1.351)
islam	1.452 (1.281)	2.709 (1.655)	2.854 (1.681)*	2.712 (1.433)*	3.258 (1.695)*
qemployee_2	-0.398	-0.370	-0.389	-0.447	-0.389

	(0.649)	(0.530)	(0.502)	(0.534)	(0.511)
qemployee_4	-0.302	-0.140	-0.221	-0.098	0.105
	(0.637)	(0.778)	(0.771)	(0.770)	(0.779)
qemployee_5	1.607	1.241	1.327	1.193	0.353
	(1.354)	(1.750)	(1.836)	(1.793)	(1.629)
qlnsales_2	0.472	0.525	-0.140	0.459	0.872
	(0.607)	(0.598)	(0.603)	(0.617)	(0.509)*
qlnsales_3	2.450	2.232	-0.096	2.216	2.120
	(1.242)*	(1.328)*	(0.723)	(1.244)*	(1.266)*
qlnsales_4	-0.748	-0.584	-0.446	-0.567	-0.223
	(0.979)	(1.033)	(0.847)	(1.042)	(0.680)
qlnsales_5	-0.071	0.201	0.732	0.296	0.286
	(0.879)	(1.353)	(1.243)	(1.355)	(1.393)
qlnfasset_2	-0.638	-0.286	-0.285	-0.276	-0.242
	(0.693)	(0.563)	(0.563)	(0.603)	(0.399)
qlnfasset_3	0.650	0.528	0.498	-0.445	0.353
	(0.795)	(0.956)	(0.905)	(0.862)	(0.675)
qlnfasset_4	0.567	-0.495	-0.390	-0.684	-0.261
	(0.630)	(0.876)	(0.844)	(0.849)	(0.704)
qlnfasset_5	0.343	0.583	0.588	-0.379	0.157
	(0.933)	(1.103)	(1.111)	(1.002)	(1.111)
entagedum_2	-0.565	-0.656	-0.599	-0.645	-0.884
	(0.678)	(0.737)	(0.668)	(0.669)	(0.751)
entagedum_3	-0.434	-0.607	-0.488	-0.633	-0.242
	(0.942)	(1.071)	(1.013)	(1.073)	(0.971)
rural	0.257	-0.148	-0.058	-0.188	-0.277
	(0.922)	(0.670)	(0.581)	(0.590)	(0.437)
manufac	8.492	9.157	9.011	8.899	2.951
	(4.429)*	(4.683)*	(4.455)**	(4.522)*	(2.442)
service	1.151	0.656	0.654	0.516	0.400
	(0.463)**	(0.645)	(0.634)	(0.723)	(0.382)
villgovs		0.702	0.730	0.698	0.509
		(0.096)***	(0.100)***	(0.103)***	(0.154)***
Constant	-6.576	-8.159	-8.025	-7.259	-6.914
	(2.660)**	(2.127)***	(2.191)***	(2.067)***	(2.294)***
Observations	1901	1782	1782	1782	1782
R-squared	0.14	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 A11. Access to Credit**  
Marginal effects from Probit regression on *credit*

	(1)	(2)	(3)	(4)
	OLS	IV	IV Interaction	
com_form	-0.001 (0.063)	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_manufac				-0.366 (0.134)***
int_service				-0.115 (0.069)*
edu	-0.022 (0.012)*	-0.013 (0.012)	-0.015 (0.013)	-0.013 (0.013)
age	-0.001 (0.002)	-0.001 (0.002)	-0.001 (0.002)	-0.001 (0.002)
female	0.049 (0.047)	0.044 (0.040)	0.044 (0.034)	0.040 (0.041)
resdum	0.042 (0.054)	0.079 (0.023)***	0.055 (0.024)**	0.075 (0.024)***
indethn	0.015 (0.038)	0.029 (0.029)	0.028 (0.027)	0.027 (0.030)
chinese	0.072 (0.204)	-0.011 (0.136)	0.076 (0.216)	-0.034 (0.098)
islam	0.066 (0.059)	-0.011 (0.037)	-0.007 (0.035)	-0.011 (0.039)
qemployee_2	-0.011 (0.038)	0.014 (0.034)	0.064 (0.043)	0.017 (0.033)
qemployee_4	-0.023 (0.042)	-0.017 (0.049)	-0.035 (0.054)	-0.015 (0.049)
qemployee_5	0.142 (0.085)*	0.202 (0.101)**	0.387 (0.159)**	0.212 (0.104)**
qlnsales_2	-0.032 (0.043)	-0.049 (0.030)*	-0.047 (0.024)*	-0.049 (0.029)*
qlnsales_3	0.067 (0.052)	-0.005 (0.038)	-0.008 (0.032)	-0.001 (0.037)
qlnsales_4	0.011 (0.053)	-0.004 (0.040)	-0.022 (0.036)	-0.012 (0.040)
qlnsales_5	0.020 (0.070)	-0.035 (0.056)	-0.027 (0.052)	-0.036 (0.055)
qlnfasset_2	0.082 (0.069)	0.027 (0.041)	0.030 (0.040)	0.018 (0.039)
qlnfasset_3	0.135 (0.061)**	0.063 (0.050)	0.068 (0.044)	0.050 (0.047)
qlnfasset_4	0.044 (0.064)	0.055 (0.044)	0.067 (0.044)	0.044 (0.044)
qlnfasset_5	0.237 (0.099)**	0.176 (0.072)**	0.188 (0.071)***	0.161 (0.073)**

entagedum_2	-0.011 (0.035)	-0.034 (0.028)	-0.024 (0.028)	-0.034 (0.028)
entagedum_3	0.013 (0.039)	0.011 (0.028)	0.016 (0.033)	0.010 (0.027)
bank	0.063 (0.044)	0.000 (0.024)	-0.006 (0.022)	-0.006 (0.024)
rural	-0.026 (0.044)	0.001 (0.025)	0.006 (0.024)	0.004 (0.025)
manufac	-0.081 (0.052)	-0.038 (0.045)	-0.030 (0.043)	0.046 (0.084)
service	0.022 (0.041)	-0.009 (0.029)	-0.003 (0.025)	0.015 (0.039)
villcred		0.732 (0.079)***	0.723 (0.077)***	0.735 (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 A11. Firms' Decision to Formalize**  
Marginal effects from probit regression on *com\_form*

	(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	0.149	0.149

	(0.070)**	(0.068)**	(0.068)**
qlnfasset_5	0.206	0.183	0.183
	(0.069)***	(0.067)***	(0.067)***
female	-0.022	-0.016	-0.016
	(0.036)	(0.036)	(0.036)
indethn	-0.102	-0.091	-0.091
	(0.041)**	(0.042)**	(0.042)**
islam	-0.083	-0.059	-0.059
	(0.084)	(0.077)	(0.077)
Chinese	0.353	0.379	0.379
	(0.158)**	(0.156)**	(0.156)**
rural	-0.195	-0.163	-0.163
	(0.031)***	(0.032)***	(0.032)***
manufac	0.035	0.032	0.032
	(0.086)	(0.074)	(0.074)
service	0.025	0.012	0.012
	(0.037)	(0.034)	(0.034)
meancost	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
meantime	0.002	0.002	0.002
	(0.002)	(0.002)	(0.002)
edu		0.045	0.045
		(0.010)***	(0.010)***
resdum		-0.155	-0.155
		(0.064)**	(0.064)**
provdum1		0.051	0.051
		(0.090)	(0.090)
Observations	1851	1682	1682

Robust standard errors in parentheses

Regression contains district dummies

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