



ERC Working Papers in Economics 09/01
March 2009

Informality and Productivity
Productivity Differentials between Formal and Informal Firms in Turkey

Erol Taymaz
Department of Economics
Middle East Technical University
Ankara 06531 Turkey
etaymaz@metu.edu.tr

Economic Research Center
Middle East Technical University
Ankara 06531 Turkey
www.erc.metu.edu.tr

Informality and Productivity

Productivity Differentials between Formal and Informal Firms in Turkey

Erol Taymaz

Department of Economics
Middle East Technical University
Ankara 06531 Turkey
etaymaz@metu.edu.tr

Abstract

The informal sector constitutes a large share of employment and output in all developing countries. Although the informal sector is regarded by many researchers and policy makers as a source of employment developing countries desperately need, there is ample evidence that documents that informal firms are less productive, employ unskilled labor, and pay lower wages. This study analyzes the sources of productivity difference between informal and formal firms in Turkey. Since the data on the informal sector is likely to be noisy, we use two different approaches to analyze productivity differentials: firm-level analysis and individual-level analysis. In the case of firm-level analysis, we estimate and compare productivity levels of informal and formal firms by using matching propensity score and switching regression methods. In the case of individual-level data, we compare wage differentials between informal and formal wage workers by estimating a multinomial selection model.

Our findings indicate that there is a significant productivity gap between informal and formal firms, and a wage gap between informal and formal workers. Moreover, the hypothesis that more educated entrepreneurs and workers move to the formal sector is supported by the data. This process of self-selection contributes to widen the productivity gap between informal and formal firms.

The theories of life-cycle and learning are also supported by our findings. Older (i.e., more experienced) firms tend to operate in the formal sector. However, the relationship between informality and age is U-shaped for entrepreneurs and workers. Even after controlling for all these factors (self-selection, differences in endowments, and learning), the productivity gap does not disappear.

The findings suggest that there is a substantial but untapped potential to increase productivity in Turkey. The productivity effect of operating formally is higher for services, but we may expect that a large number of informal service firms could not survive if they operate formally.

1. Introduction

The informal sector constitutes a large share of employment and output in all developing countries. Although the informal sector is regarded by many researchers and policy makers as a source of employment developing countries desperately need, there is ample evidence that documents that informal firms are less productive, employ unskilled labor, and pay lower wages. Therefore, there are calls to adopt policies to reduce the size of the informal economy so as to increase productivity and to achieve sustainable growth. Since the informal economy is a source of substantial employment, the benefits and costs of reducing informality should be carefully analyzed.

This study analyzes the sources of productivity difference between informal and formal firms in Turkey. Since the data on the informal sector is likely to be noisy, we use two different approaches to analyze productivity differentials: firm-level analysis and individual-level analysis. In the case of firm-level analysis, we use two datasets collected through specific surveys on formal and informal firms, and estimate and compare productivity levels of informal and formal firms by using matching propensity score and switching regression methods. In the case of individual-level data, collected through the Labor Force Surveys, we compare wage differentials between informal and formal wage workers by estimating a multinomial selection model.

The paper is organized as follows. In the second section of the paper, we define the concept of “informality”, and discuss the (likely) sources of productivity differentials between informal and formal firms. In the third and fourth sections, we present the data, model, and estimation results of the firm-level analyses. The findings of the individual-level analysis are discussed in the fifth section. The sixth section presents the findings of a simulation analysis on the effects of enforcing formality. The last section summarizes the main findings of our analyses.

2. Conceptual framework

2.1. Definition of “informality”

There are a large number of terms and definitions offered for informal-type economic activities. The terms “informal”, “shadow”, “underground”, “uncovered” and “unrecorded” are usually used synonymously, but the definitions of these terms refer to different conceptualizations (for a discussion, see Perry et al., 2007, Chapter 1). However, different definitions could be categorized in two groups, which are also precisely defined by the International Labor Organization (ILO). The first category emphasizes the dualistic and segmented nature of the labor market, and defines “informality” in terms of the characteristics of enterprises and working conditions. The second category refers to the legal status of the economic activity. An economic activity is defined as “informal” if it is legal but not legally recorded/registered.

In the case of employment, the first category is defined by ILO as “employment in the informal sector” and the second one as “informal employment”. “Employment in the informal sector” covers all jobs in informal sector enterprises which are defined as “private unincorporated enterprises (excluding quasi-corporations), i.e. enterprises owned by individuals or households that are not constituted as separate legal entities independently of their owners, and for which no complete accounts are available that would permit a financial separation of the production activities of the enterprise from the other activities of its owner(s)” (for details, see Hussmanns, 2004). Informal employment, however, includes “1) Own-account workers and employers employed in their own informal sector enterprises, 2) Contributing family workers, irrespective of whether they work in formal or informal sector enterprises, 3) Employees holding informal jobs, 4) Members of informal producers’ cooperatives, 5) Own-account workers engaged in the production of goods exclusively for own final use by their household”.

In this study, we prefer to use the term “informal employment” to refer to those employees (wage workers, self-employed and entrepreneurs) who are not registered in any social security organization. The “informal firm” is defined accordingly, as a firm that employs informal employees, and the “informal sector” as the set of informal firms. There are three reasons to choose this restricted and strict definition. First, as Henley et al. (2006) show,

definitions of informality based on occupation and firm size seem to be “arbitrary in practice even if conceptually well-founded”. Second, the availability of data on social security status makes empirical study feasible. Third, as a comprehensive study on Turkey shows (see McKinsey Global Institute, 2003, p.50), this is the most widespread form of evading regulatory obligations in Turkey, because “most of the business [in Turkey] are registered, but they partially report business revenue and employment” due to low cost of registering and strong enforcement.

2.2. Informality and determinants of productivity

There are a substantial number of theoretical and empirical studies that analyze and document the characteristics of informal employment and informal firms. The literature shows almost unequivocally that costly and burdensome labor and product regulations, administrative complexity of taxation, and legal quality are important determinants of informality.¹ In other words, informal firms could avoid a number of costs involved in operating formally (avoiding taxes and social security payments, benefiting from flexible employment and production relations, etc.), but informal firms face with a number of disadvantages and costs. The most obvious cost of informality is the (potential) cost of punishment if the firm is detected, and the probability of detection is likely to increase by firm size so that this fact explains why small firms are more likely to operate informally. Moreover, there are additional disadvantages of informality, like the lack of, or the restricted access to, public services (training, fairs, etc.), infrastructure, and public support schemes, limited access to formal credit, lack of legal protection, high transaction costs, etc. These factors may have a detrimental effect on informal firms’ performance.

One of the most cited stylized fact associated with informality is the productivity differential between informal and formal firms. Informal firms are less productive than formal firms (Dabla-Norris et al., 2005). A recent comprehensive study on informality shows that the difference in labor productivity between those firms that operate informally and formally is about 30, on average, for seven Latin American and Caribbean countries (Perry et al., 2007: 173). A study on Turkey finds a similar productivity gap between

¹ Dabla-Norris et al. (2005) suggests that “the elasticity of informality with respect to the regulation burden is smaller, the better the quality of the legal system is”. Therefore the effect of regulatory burden in encouraging informal activity will be weaker in countries with a strong rule of law.

formal and informal businesses, around 30-40 percent (McKinsey Global Institute, 2003: 65),

Low productivity of informal firms is a policy concern for governments who are frequently advised to reduce the size of the informal economy to sustain economic growth. For example, an OECD study suggests that “overcoming the duality between the formal and informal sectors should be the central point” of any strategy in Turkey, and the government should encourage small and medium-sized firms, by reducing regulatory burden, to move into the formal sector, and to raise productivity through economies of scale (Gönenç et al., 2007: 20).

Any policy towards informal sector that aims to raise productivity should be based on a careful analysis on the determinants of productivity among informal and formal firms,² because the outcome of the policy is not independent of the factors that generate productivity differential. For example, Amaral and Quintin (2006) suggest that when labor markets are segmented, policies aimed at increasing the share of the formal sector can raise productivity because the value of the marginal product of formal workers is higher than that of informal workers. Thus, subsidizing formal employment can increase national income. However, in the case of competitive labor markets, workers in the informal sector are less productive because they have less education. In such a case, policies that aim at reducing the size of the informal sector are “a poor substitute for investments in education, or investments in the quality of formal institutions (e.g. improving enforcement)”. In other words, the policy should attempt to solve the binding constraint that lowers productivity of informal firms. Otherwise, the policy would be ineffective, or even be counterproductive. The literature on the effects of public policies towards informality is extensive. But, unfortunately, there is no consensus on the best policy options, partly because of the fact that the effects of policies are highly context-dependent.³ Thus, there is a need to identify

² A government could aim at reducing the extent of informality because of tax considerations (to enlarge the tax base), to eliminate the cost disadvantages of formal firms, and to improve working conditions. Although these issues are also very important, we focus our attention on the effects of informality on productivity.

³ Numerous studies yield unconventional findings on the effects of public policies towards informality. For example, strict law enforcement may reduce informal employment, but, it may also decrease average wages, productivity and investment by reducing the firm’s access to unregulated labor (Almeida and Carneiro, 2006); an increase in labor tax rate may reduce the share of the informal sector (Badaoui et al., 2007); an increase in the labor tax may reduce job creation in the informal sector, but may increase the average quality of the workforce in the formal sector with a positive effect on job creation (Boeri and Garibaldi, 2005);

the reasons behind the productivity differentials to design policies for reducing informality and alleviating urban poverty.

One of the main factors that may lead to productivity gap between formal and informal firms is the lack of access to markets and services by informal firms (Djankov et al., 2003; Straub, 2005; Perry et al., 2007: 157-158). Informal firms may not benefit from key public goods, enforcement of property rights and contracts. This could increase their transaction costs due to inefficient contractual relations, i.e., a part of informal firms' resources will be wasted due to inefficient institutional mechanisms in which informal firms are forced to operate. Moreover, they will not be able to benefit from various public support schemes (training of employees and managers, technology diffusion services, etc.) that may improve productivity.

The lack of access to credit provided by state-owned or private banks may have a detrimental impact on productivity because of two reasons. First, capital constrained informal firms will scale down their capacity, and operate below the efficient scale of production. Second, high cost of capital or limited outside financing will force informal firms to substitute (low-skill) labor for physical capital (Amaral and Quintin, 2006). Hence, informal firms are likely to have lower capital intensity and lower labor productivity.⁴

A usual suspect for productivity differentials is the existence of economies of scale. The negative correlation between the extent of informality and firm size is one of the robust stylized facts on informality: informal firms are usually small firms. If economies of scale are relevant, at least among very small firms, and if informality is widespread among small firms, then a productivity gap will arise between average informal and formal firms (Perry et al, 1007: 157). In such a case, if informal firms shy away from growth because of fears of detection, eliminating regulatory burden will make it easier for informal small and medium-sized firms to grow, and the average productivity will raise through economies of reducing the cost of formalization may increase the size of the informal sector (Dessy and Pallage, 2003); policies improving the education of the labor force may decrease incentives to formalize (Masatlioglu and Rigolini, 2005).

⁴ If informal firms substitute labor for capital due to relatively higher cost of capital, they would have lower labor productivity as the formal firms do, but the same level of total factor productivity if they operate on the same production function.

scale (Gönenç et al., 2007:20). Although the economies of scale argument is frequently adopted by policy makers, there is no robust empirical evidence on the degree of economies of scale. Moreover, even if the production function exhibit economies of scale, any productivity difference between informal and formal firms will disappear once it is conditioned on firm size.

The discussion on the role of economies of scale implicitly assumes that informal and formal firms adopt the same technology and operate on the same production function. However, because of the differences in the knowledge set available to informal and formal entrepreneurs, and the restrictions they face with, informal and formal firms may indeed use different technologies, i.e., the production function for informal and formal firms could be different, and the differences in underlying production technologies may lead to differences in observed productivity levels.

There are a number of studies that show that new firms in developing countries tend to start their life as informal, and if they perform well, they tend to grow and become formal (Levenson et al., 1998). In this framework, that resembles Jovanovic's (1982) model of learning, firms, if they survive, should move from informal to formal as part of their natural evolution. Since young firms at the early stages of their life-cycle have lower productivity on average, there would be a productivity difference between informal and formal firms. However, this difference will disappear if the life-cycle (for example, the age of the firm) is controlled for.

Productivity differentials could arise not because of intrinsic characteristics of informal and formal firms, but because of self-selection of more productive (more educated) workers and entrepreneurs into the formal sector. There are a large number of theoretical studies, mainly based on heterogeneous workers and/or firms and matching models, that show that more productive workers go to formal sector jobs, whereas less productive workers select into the informal sector (see, for example, Boeri and Garibaldi, 2005; Albrecht et al., 2006, Galiani and Weinschelbaum, 2007). The most talented managers self-select into the formal sector, and that formal managers operate with more physical capital than informal managers in the model developed by Amaral and Quintin (2006). In

all models, there is a (probabilistic) penalty of detection for informal firms operating informally. The model by Rogers and Swinnerton (2004) is different than others by its assumption on voluntary compliance, i.e., there is no cost for informality. In this model, firms differ in productivity, and more productive firms tend to voluntarily comply with the costly “labor standards” because doing so relieves the labor shortage they face. Once firms have been identified as formal firms, workers will look for jobs with them first. This will increase the average productivity of the workforce, as a greater number of workers will go to the more productive firms. Note that in all those models, there is a productivity differential between informal and formal firms, but it arises because of self-selection of more productive/more educated/more talented workers and/or entrepreneurs into the formal sector.

There is some empirical support for the underlying assumptions of these models. Galiani and Weinschelbaum (2007) and Dimova et al. (2008) find that workers with high levels of education allocate to the formal sector and receive high compensation for their education and experience, whereas less educated workers allocate to the informal sector. However, there are some notable exceptions as well. For example, the study by McKenzie and Sakho (2007) reveal that “owners of large firms who have managed to stay informal are of higher entrepreneurial ability than formal firm owners, in contrast to the standard view (correct among smaller firms) that informal firm owners are low ability”.

The discussion (and the papers analyzed) so far assumes that endowments, and most importantly, educational level as a proxy for human capital is fixed or given. However, the agents (firms and workers) will respond to incentives in investing in human capital, and the extent of informality may have a significant impact on incentives. Perry et al. (2007: 161, 171) suggest that, because of working conditions, informal firms may not attract educated workers, and have less incentive to invest in training and innovation. Thus, the returns to education will be lower in an economy dominated by the informal sector. In a similar framework, Masatlioglu and Rigolini (2005) present a model in which educated and unskilled workers are employed in the formal and informal sectors, respectively. They “show that high costs of education make labor migration and firm’s profits in the formal sector an increasing function of its size. Therefore, incentives to reduce informal economic

activity increase with the size of the formal economy, and unless the formal sector has reached a ‘critical mass’ countries remain in a highly informal equilibrium”. The model by Amaral and Quintin (2006) assumes endogenous education decision by workers, and the model generates well the stylized facts that formally employed workers tend to be older, have more education, and earn more than informal workers, and employers that comply with regulations tend to be much larger than informal employers.

The literature on skill traps caused by skill-investment or skill-R&D complementarities (see, for example, Snower, 1994; Redding, 1996; Acemoglu, 2001; Burdett and Smith, 2002) emphasizes the importance of high skill-good jobs that pay high wages for economic performance. This literature can be extended to analyze informality where (high skill-high wage) “good jobs” are created by the formal sector, and (low skill-low wage) “bad jobs” by the informal sector.

To summarize, our review suggests that there are four factors that may lead to a productivity gap between informal and formal firms: 1) the lack of access to public services and markets by informal firms, 2) choice of technology and economies of scale, 3) self-selection of more educated/more productive workers/entrepreneurs to the formal sector, and 4) learning and life-cycle effects. In the subsequent sections, we will analyze the role of these factors in explaining productivity differentials by using firm-level and individual-level data.

3. Informality and productivity: Firm-level evidence for small firms

3.1. The data

The lack of data is the main obstacle in comparing productivity differentials between informal and formal firms, because, almost by definition, informal firms do not keep their accounts, and are not willing to provide information on their activities. There is, however, an exceptional study made on Turkey by Semza Ozar of Bogazici University (2006). As a part of the Economic Research Forum (ERF) project on “Promoting Competitiveness in Micro and Small Enterprises”, Dr. Semsa Ozar and her team conducted an extensive study

on small and medium-sized firms in Turkey, by collecting the data from about 5000 firms employing less than 50 people through field survey.

The study used a national, stratified, multi-stage systematic sampling method to identify firms. In the first stage, 19 provinces were selected from 5 strata defined in terms of socio-economic level. The selection of provinces from each stratum was carried out by weighted probability regarding the number of enterprises in each province. In the second stage, 432 Primary Sampling Units (PSUs) with a minimum of 45 enterprises in urban areas were selected.⁵ The PSUs were defined on the basis of the *2000 Census of Buildings* conducted by Turkstat.

The research team identified 9,280 eligible firms for the survey in elected PSUs, and 7,335 of these firms were selected randomly with respect to the proportions by sub-categories of gender, size and location. A total of 5,000 interviews were carried out. The study was conducted from June to September 2001 (for details, see Ozar, 2006).⁶

A very detailed questionnaire with 322 questions was applied in the survey, and information about the entrepreneur, legal status of the enterprise, types of informality, production, employment, financing, etc. were collected. There are three questions about different types of informality: i) if the enterprise registered with the industrial or commerce register, ii) if the enterprise registered with the tax department (acquired a tax card or a card number), and iii) if the enterprise joined any social insurance scheme.⁷ Table 1 presents the data on the proportion of formal firms by the type of formality at the time of survey.⁸ Firms are classified into two sectors, manufacturing and services.⁹ The data show that most of small firms in Turkey are registered. The proportion of registered firms (both commerce register and tax office) is slightly lower in manufacturing than in services (89

⁵ During this stage, 100 villages in rural areas were also selected, but the rural enterprises from the sample were excluded before the weighting and extrapolating process.

⁶ The database and related information can be downloaded from the ERF web site: www.erf.org.eg

⁷ The specific questions are Q84 (“Enterprise registered (industrial or commercial)?”), Q90 (“Registered with tax department (acquired a tax card or a card number)?”), and Q93 (“Joined social insurance scheme?”). If the response to Q93 is “yes”, then a follow-up question is asked (Q96, “Who was covered? (1) entrepreneur, (2) entrepreneur + some workers, (3) entrepreneur + all workers”).

⁸ Unless otherwise stated, all data presented in this section are derived from the ERF database.

⁹ Manufacturing includes mining as well (only 6 firms). Services includes construction (65), wholesale and retail trade (2887), restaurants and hotels (406), transportation, communication (41), finance (21), real estate, business services (89), and other services (203).

percent vs 96-97 percent). There is not much difference in terms of registering employees with a social security scheme: about one fifth of small firms in Turkey did not register any employee.

Figure 1 shows the Genn diagram for the number of informal firms by three types of informality to visualize overlaps between different types of informality. It is seen that most of the firms who do not register at the commerce register do not register with the tax office and social security as well. This is an expected outcome, because registering the firm with the commerce register is the first step in establishing a business in Turkey. Firms apply to the tax office, and social security scheme after registration. However, a great majority of firms who did not cover their employees with a social security scheme are registered with the commerce register *and* the tax department (626 firm in the sample). These findings show that almost all firms, even very small ones, do register with the commerce register and tax office, and uncovered employment is most common type of informality. This is why we prefer to define “informality” and “informal sector” in terms of “informal employment”.

The information on the factors behind informality indicates that those firms that did not register with the commerce register or the tax office did so because they consider registration as “useless”, but the majority of firms that did not join social security schemes did so because they consider it “expensive”. This finding supports the anecdotal evidence presented in McKinsey Global Institute (2003: 50) that the cost of registering the business in Turkey is low, but small firms partially report employment.

As it is observed in almost all countries, there is a positive correlation between the level of formal employment (defined as covering employees with an official social security scheme) and firm size both in manufacturing and services (see Table 3). Although the share of formal firms is less than 80 percent among firms that employ less than 5 people, it exceeds 95 percent among firms than employ at least 20 people. Entrepreneur’s gender seems to matter as well. Female entrepreneurs, especially those in manufacturing, tend to have much higher proportion of informality. Finally, as suggested by self-selection

theories, more educated entrepreneurs are more likely to establish formal firms. The share of formality is especially very low among those entrepreneurs without any diploma.

3.2. *The models*

A simple comparison between informal and formal firms in Turkey shows that the stylized facts are valid for the Turkish case. Table 4 presents the data on the mean values of a number of variables for informal and formal firms operating in manufacturing and services. Even if our sample includes only small firms employing less than 50 people, as expected, formal firms produce more, achieve higher labor productivity (as measured by output per employee), use more capital intensive techniques, and pay higher wages. Moreover, formal firms are older (because they survive longer), and employ fewer young people (they have longer tenure). Entrepreneurs of formal firms are more educated, older, and have more experience.

Although most of these differences are statistically significant at the 5 percent level, a simple comparison between a group of informal and formal firms would not provide any information why these differences emerge. If, for example, the factors that influence the choice of entry into formal/informal sectors and the characteristics under investigation are correlated, then the observed differences would simply reflect the effect of self-selection process. In order to mitigate the effects of self-selection and to test if there is a productivity gap between informal and formal firms, we use two methods, propensity matching and endogenous switching regression.

The principal question to be answered in assessing the effect of informality is what an informal firm would do had it registered (see Fajnzylber et al., 2006; Arias and Khamis, 2008). Matching method is based on a comparison of the outcomes that would have been observed for informal (“treated”) firms had they registered and become formal. In this framework, let C^1 be the outcome conditional on informality (“treatment”) and C^0 the outcome that would have been observed if the same firm joined the formal sector. The impact of the informality is:

$$(1) \quad \Delta = C^1 - C^0$$

Since only C^1 or C^0 is observed for each firm, Δ is not observable. Although it is not possible to identify the individual effect of informality, the average effect can be estimated under certain assumptions. The main parameter estimated in empirical studies is the mean “impact of treatment on the treated”, which is defined as:

$$(2) \quad TT = E(\Delta | \text{Inf} = 1) = E(C^1 | \text{Inf} = 1) - E(C^0 | \text{Inf} = 1)$$

where $E(\cdot)$ is the expectation operator. “Inf = 1” and “Inf=0” denote the groups of informal and formal firms, respectively.

The counterfactual $E(C^0 | \text{Inf} = 1)$ must be estimated because it is not observed. There are three estimators available: cross-section, before-after, and difference-in-differences (DID). Since we have only a cross-sectional data, we use cross-section estimator which is defined as:

$$(3) \quad \Delta_{CS} = E(C^1 | \text{Inf} = 1) - E(C^0 | \text{Inf} = 0)$$

The cross-section estimator (Δ_{CS}) compares the average values of the outcome variable, C , for the informal firms and a comparable (“control”) group of formal firms. In this paper the “nearest neighbor matching” is used to construct a control group of formal firms. Each informal firm in the sample is matched with its closest formal neighbor on the propensity score (the probability to become informal), which is estimated by a logit model. This method ensures that the constructed groups of informal and formal firms are statistically not different in terms of the variables used in the estimation of propensity scores.¹⁰

A switching regression model for production function is also estimated to check if informal and formal firms use the same technology, to test if there are productivity differences, and to understand the factors behind productivity differences, including the role of economies of scale, if any.

¹⁰ “psmatch2” program is used for this purpose (see Leuven and Sianesi, 2003).

The model is defined as

$$(4) \quad \begin{aligned} I_i &= 1 \text{ if } \gamma Z_i + u_i > 0 \\ I_i &= 0 \text{ if } \gamma Z_i + u_i \leq 0 \\ \text{Informal: } q_{1i} &= \beta_1 X_{1i} + \varepsilon_{1i} \text{ if } I_i = 1 \\ \text{Formal: } q_{0i} &= \beta_0 X_{0i} + \varepsilon_{0i} \text{ if } I_i = 0 \end{aligned}$$

where I_i denotes informality status of the i^{th} firm (1 for informal, 0 for formal), Z a vector of variables that determine the choice of informality, q_{ji} (log) output of firm i under j (informal/formal), and X a vector of inputs (capital, labor, materials, all in log form). β_1 , β_0 , and γ are vectors of parameters to be estimated. It is assumed that u_i , ε_{1i} and ε_{0i} have a trivariate normal distribution with mean vector zero and covariance matrix

$$\Omega = \begin{bmatrix} \sigma_u^2 & \sigma_{1u} & \sigma_{0u} \\ \sigma_{1u} & \sigma_1^2 & \cdot \\ \sigma_{0u} & \cdot & \sigma_0^2 \end{bmatrix}$$

The covariance between ε_{1i} and ε_{0i} is not defined, because q_{1i} and q_{0i} are never observed simultaneously. Further, it is assumed that $\sigma_{2u} = 1$ (γ is estimable only up to a scalar factor). The model is identified by non-linearities because the latent variable in the selection equation is a continuous variable but the selection is a discrete event. Moreover, to improve identification, the selection model may include some exogenous variables that do not have any impact on output so that these variables are excluded from the production function.

The switching regression model assumes that an entrepreneur, given a set of her characteristics represented by Z , makes a decision to enter into informal or formal sectors. The entrepreneur adopts a specific production technology to produce output, and the production technology, due to the differences in operating conditions in the informal and formal sectors. The parameters of the production function will provide information about the sources of productivity difference.

We assume a Cobb-Douglas production function in estimating the switching regression model:

$$(5) \quad q_{ji} = \beta_{j0} + \beta_{j1}k_i + \beta_{j2}l_i + \beta_{j3}m_i + \varepsilon_{ji}$$

where subscripts denote informality status (1/0), and the firm. q , k , l , and m are (log) value of output, value of fixed capital, number of employees, and value of input (raw materials, energy, and others). The production function exhibits economies of scale if $\beta_{j1} + \beta_{j2} + \beta_{j3} > 1$. The intercept term, β_{j0} is the productivity parameter, and it measures the level of total factor productivity. If $\beta_{10} < \beta_{00}$, then we will conclude that formal firms are more productive.

The productivity of the firm is likely to be determined by a number of variables, such as the composition of the workforce, educational level of the entrepreneur, etc. In other words,

$$(6) \quad \beta_{j0} = \alpha_j X_i^*$$

where X^* is a vector of variables that determine productivity. By substituting Equation 6 into Equation 5, an augmented production function can be estimated to shed light on the sources of productivity differences.

3.3. *Determinants of informality*

Propensity matching and endogenous switching regression methods are based on an adjustment for the selection process. Therefore, we will first analyze the determinants of selection by estimating a logit model in which the dependent variable is a binary variable that takes the value 1 if the firm is informal (no employee registered with social security), and 0 otherwise. The following explanatory variables are included in the model:

It is well documented in the literature that there is a negative correlation between firm size and the probability of informality because large firms can easily be identified by the public

officials if they do not register their employees. We therefore use the log number of employees, *ll*, as an explanatory variable in the model. This is the only firm-specific variable in the model because we would like to avoid possible simultaneity of firm-specific variables in the informality status model. Thus, we estimated the same model by omitting the size (*ll*) variable, but the results did not change much.

The model includes a rich set of variables about the characteristics of the entrepreneur: first, we include a variable, *gender* (1 for female, 0 for male), to test if entrepreneur's gender influences his/her decision to operate informally. As also observed in the descriptive analysis, female entrepreneurs could be more likely to enter into the informal sector.

The model includes age and its square to allow for non-linear relationship between entrepreneur's age and his/her tendency to operate informally (*age* and *age2*). We expect a U-type relationship between informality and age, because young entrepreneurs could be more likely to start with informal activities, then to switch to the formal sector. After controlling for the experience of the entrepreneur, we may expect that older entrepreneurs are also more likely to work informally, because the business they are involved in is likely to be their secondary activity. A variable about the experience of the entrepreneur (*lexper*, the log value of experience of the entrepreneur in the last job) is also added to test if experience makes transition to formal activities. We expect a negative coefficient for the *lexper* variable because more experience entrepreneurs are more likely to move to the formal sector.

The theories of self-selection emphasized the importance of education (human capital). We use a set of dummy variable for the educational level, *educ0-educ6* for no diploma, primary school, secondary school, high school, and university, respectively.¹¹ Since the proportion of people with no diploma was very low in our sample, we merged *educ0* and *educ1*, and the merged category (less than secondary school) is used as the base category (the omitted dummy variable). We expect that the probability of informality will decline by

¹¹ The questionnaire includes a question about the year of schooling. We generated the educational level dummies by assuming "normal" time for schooling. Therefore, strictly speaking, *educ0-educ4* refer less than 5 years, 5-7, 8-10, 11-14, 15 and more years of schooling, respectively.

the level of schooling. Although the questionnaire does not differentiate between different types of schooling (for example, high school vs vocational high school), there is a specific question on the vocational training the entrepreneur has ever received. Thus, we include a dummy variable for vocational training, *vocation*, to control for the effects of vocational training on informality.

The region where the entrepreneur was born could be a factor in informality decision because the businesses are established in the urban areas, and those who were born in rural areas may have limited access to social networks in urban areas. We use a dummy variable, *rural*, that takes the value 1 if the entrepreneur was born in rural areas, and 0 if born in urban areas. We expect a positive coefficient for the rural variable: those entrepreneurs with rural background are more likely to operate in the informal sector.

The questionnaire has a specific question about the entrepreneur's main aim in establishing the business, and the options are "suits qualifications", "has experience in business", "capital requirements reasonable", "family business", "desire to set up new enterprise", "to improve living conditions", and "only option available". The last option refers to the fact that the entrepreneur has established the business because of desperation, and these entrepreneurs are more likely to face with credit and other constraints, and, thus, are forced to operate informally. To test if this is the case, we generated a dummy variable, *aim_tina*, that takes the value 1 if the entrepreneur established the business if this was his/her "only option available", and 0 otherwise.¹²

Entrepreneurs could establish more than one business or could be involved in many economic activities. We use a dummy variable, *sole_act*, to check if there is a difference in the tendency towards informality between those entrepreneurs who are involved in only one economic activity, and those who are involved in more activities.

The location of the firm could be an important factor for informality. If the firm is located in a cluster (the *cluster* variable) in which there are neighboring enterprises engaged in

¹² We experimented with a number of dummy variables for other options as well, but they did not have a statistically significant coefficient.

related activities, the firm may tend to operate formally. Hence, we expect a negative coefficient for the *cluster* variable.

Finally, the (log) age of the firm (*lface*) is included to check if the life-cycle of the firm matters for informality. If, as suggested by learning and life-cycle theories, young firms tend to start their life in the informal sector and gradually move to the formal sector, we expect a negative coefficient for the *lface* variable.

The estimation results for the logit model of informality decision in manufacturing and services are presented in Table 5. The findings are consistent with our *a priori* expectations. Large firms are less likely to operate informally, and more educated entrepreneurs tend to move into the formal sector. Vocational training has a strong negative impact on informality. Those entrepreneurs who have received vocational training are more likely to operate in the formal sector. There is a U-type relationship between informality and entrepreneur's age. The probability to operate informally reaches its minimum level around the late-40s in both sector (47 in manufacturing, 50 in services).

If the entrepreneur has established the business because he/she had no other option, then, he/she is more likely to be in the informal sector (the impact is not statistically significant in manufacturing). If the business is the entrepreneur's only economic activity, then the probability of informality declines. Apparently, secondary activities are more likely to be performed informally.

There is a support for life-cycle theories: the tendency to operate informally diminished by entrepreneur's experience and firm's age, but the coefficients of these variables are statistically significant only in services. Moreover, clustering has a positive impact on the extent of formality.

Our findings on the determinants of informality suggest that self-selection of more educated entrepreneurs into the formal sector, life-cycle and learning theories, and the detection and punishment factors (proxied by the firm size) are all supported by the data.

Note that all these factors are likely to increase the productivity of formal firms (more human capital, more experience, and larger size).

3.4. Performance differences between small informal and formal firms

We compare informal and formal firms to test if performance differences between these groups are statistically significant or not. In order to eliminate the effects of self-selection into informality, we need to construct a control group of formal firms. The control group is formed by using the estimated probabilities of being informal from the logit model so that the group of informal firms (the “treated” group) and formal firms in the control group have similar characteristics with respect to the variables that determine the choice of informality (the explanatory variables in Table 5).¹³

Table 6 summarizes the average effects of informality by using matching propensity score estimation. The mean values of variables under investigation are presented for the “treated” group (all informal firms), the group of all formal firms, and the control group (the selected group of formal firms). The test statistics for the control group are based on biased-corrected bootstrap estimator calculated for 250 replications.

The findings indicate that the differences in almost all variables between informal and formal firms are statistically significant at the 10 percent level, both in manufacturing and services. The differences for male and female wage variables are usually not significant, although the difference for average wage rate is significant. The *credit* variable (if the firm received any credit in the last year) has almost the same mean value. About one quarter of all informal and formal firms received credit in the last year.

The effects of selection are controlled for in estimating the differences between the informal group and the control formal group. It seems that even after controlling for selection, formal firms are more productive than informal firms. The (labor) productivity gap in manufacturing is 107 percent and in service 60 percent. Formal firms are more capital intensive, and they pay, on average, higher wages. The differences for a number of variables (*partners*, *youngsh*, *blink*, *rlink*, and *cluster*) become insignificant when the

¹³ Because of the lack of panel data, we cannot control for unobserved effects that could be eliminated by difference-in-difference estimator.

effects of the selection process are controlled for. These findings show that the productivity gap between informal and formal firms cannot be explained only by self-selection of “better” (more educated, more experienced, etc.) entrepreneurs into the formal sector.

3.5. Productivity differences between small informal and formal firms

We estimate production functions for informal and formal firms by using the switching regression method to test the determinants of productivity differences (Equations 4 and 5).¹⁴ The first model estimated includes only input variables (capital, labor and raw materials) as explanatory variables. Then the variables about entrepreneurial characteristics (educational level, vocational training, gender and partnership) are included. Finally, we included a number of variables on the composition of the workforce and types of customers (technology, female employee share, young employee share, type of main consumer and geographical market). These variables, especially the last two variables, are likely to be endogenous.¹⁵ Nevertheless, we estimated the model to check the possible effect of firm-specific variables.

Estimation results are presented in Table 7. The coefficients of production functions for informal and formal firms are significantly different, i.e., there is no homogenous technology used by both informal and formal firms. Moreover, contrary to our *a priori* expectations, there are strong decreasing returns to scale for informal firms. For formal firms, constant returns to scale seem to prevail. The labor elasticity of output is much higher for formal firms than informal firms, i.e., the marginal product of labor is higher in formal firms. This may explain why formal firms pay higher wages.

The coefficient of the intercept term in a model with only inputs measures the level of total factor productivity (TFP). A comparison of that coefficient for informal and formal firms shows that there is a substantial TFP gap: formal firms are almost 150 percent more productive than informal firms in manufacturing as well as in services. The productivity

¹⁴ The selection model includes some variables (age, age2, lexper, lfage, rural, aintina, sole_act and cluster) that are not used in the production function to improve the identification the the switching regression model. F-tests indicate that these variables have statistically significant impact on selection.

¹⁵ We estimated the model by dropping these two variables (blink and rlink) but the estimation results for other variables did not change qualitatively.

gap is observed even after controlling for self-selection of more educated and experienced entrepreneurs into the formal sector.

We then include into the model a number of variables that may explain TFP (Equation 6), first a set of variables about entrepreneurial characteristics, then the second set of firm-specific variables that may improve TFP. The addition of the second set does not change the coefficients of other variables significantly.

Estimation findings show that entrepreneur's educational level matters for productivity. More educated entrepreneurs' firms are more productive both in the informal and formal sectors. However, entrepreneur's gender does not matter: although female entrepreneurs are more likely to enter into the informal sector, conditional on the sector, entrepreneur's gender does not have any impact on productivity.

Vocational training has an unexpected negative impact on productivity in services. It seems that partnership stimulates productivity in services but not in manufacturing.

The composition of workforce is another relevant factor for productivity. Those firms with higher share of young employees are less productive presumably because of the fact that young employees are less experienced. Female employment share has an ambiguous effect: a formal manufacturing or service firm becomes more productive by employing more female workers, but the share of female employees on productivity is negative for informal manufacturing firms.

The type of main consumer and geographical market has a significant impact on productivity. Those firms whose main consumers are more institutionalized users (the government, medium-sized and large firms) are more productive than those firms who sell their products mostly to the households. In a similar way, those firms that sell their products in regional/national markets are more productive than those firms that produce for the local market. Finally, the technology variable has a positive and significant coefficient for informal firms.

When the production function included only input variables, the TFP differential between formal and informal firms is found to be about 150 percent in manufacturing and 140 percent in services (the TFP difference in this case is equal to the difference in the intercept terms). When the set of variables on entrepreneurial characteristics is included, the TFP differential unexplained by entrepreneurial characteristics is found to be 44 percent in manufacturing and 115 percent in services. These results confirm that the differences in entrepreneurs educational levels can explain a non-significant part of productivity differentials between informal and formal firms, especially in manufacturing. When the firm-specific characteristics are also controlled for, the unexplained TFP differential is reduced to 19 percent in manufacturing and 62 percent in services. However,

4. Informality and productivity: Firm-level evidence

4.1. The data

The analysis in the preceding section is based on a database of small and medium-sized firms. Since the share of informality declines sharply by firm size, the exclusion of firms employing 50 or more people is not likely to create a serious bias in estimation results. The World Bank conducted a special, comprehensive survey on informality in Turkey in 2008. We will analyze the World Bank dataset to provide additional evidence on the sources of productivity gap between informal and formal firms.

The World Bank survey was conducted by Frekans Research at the end of 2008, and it includes questions about the characteristics of the firms (sector, products, output, number of employees, etc.), registration status, and a large number of questions about the perceptions of the respondent on informality, the role of the state, trust in various institutions, etc. There are about 1000 firms surveyed, but the item response rate is as low as 60 percent for some questions. Half of the firms surveyed are sampled from the western part of the country (İstanbul, Kocaeli, İzmir, Bursa, Ankara and Denizli), and the rest from the “eastern” provinces (İçel, Malatya, Konya, Trabzon, Gaziantep and Adıyaman). In about 57 percent of the firms, respondents are “owner of the firm” or “partner”, and in the 43 percent of cases, “manager”, “manager of finance/accounting”, or other managers.

The survey has a specific question about informal employment (F7, “Do you have workers whom you could not insured? If yes, how many?”). We define as “informal” all firms that have a positive response to this question. The data on the extent of informality is presented in Table 8. As shown in this table, the proportion of formal firms in the dataset exceeds 90 percent both in manufacturing and services.¹⁶ There is a positive correlation between the extent of formality and firm size: the share of informality among micro- and very small firms (that employ less than 20 people) is about 9.5 percent, and it declines to 5.6 percent for those that employ 20 or more people.¹⁷

There are a few female entrepreneurs (“owners” and “partners”) in the World Bank database (only 7 in manufacturing, and 17 in services).¹⁸ The rate of informality is lower for female entrepreneurs than males, but the low number of observations on female entrepreneurs makes the comparison not meaningful.

The self-selection of hypothesis receives further support from the World Bank survey. There is a strong, positive correlation between the extent of formality and the level of education. More educated entrepreneurs are more likely to operate in the formal sector.

Descriptive statistics for informal/formal firms operating in manufacturing and services are provided in Table 9. Since a significant proportion of respondents are not entrepreneurs, the data on individual characteristics are calculated for all respondents (“respondents’ characteristics”). The data on manufacturing are consistent with the data provided in the preceding section: formal manufacturing firms are larger, older, and more productive, use more capital intensive technologies, pay higher wages, earn more profit and achieve higher levels of efficiency¹⁹ than their informal counterparts. Moreover, formal firms are more likely to be “incorporated” companies, to deposit wages in a bank (instead of direct, cash-

¹⁶ The survey excludes agriculture and electricity/gas/water sectors. “Services” includes construction (84 firms), trade (329 firms) and transportation and communication (189 firms).

¹⁷ The informality rates in the World Bank database are quite lower than the rates calculated for the ERF database. For those firms that employ less than 20 people (50 people), the rates of informality for the World Bank and ERF databases are 9.5 percent and 17.5 percent (9.1 percent and 17.2 percent), respectively. The difference is likely to be due to differences in definitions and sampling methodologies.

¹⁸ “Owners” and “partners” are defined as “entrepreneur”. Therefore, the data on entrepreneur’s gender and education is based on only those respondents that designate themselves as “owner”/“partner”.

¹⁹ “Profitability” and “efficiency” are self-designated, subjective evaluations.

payments). The owners/managers of formal firms have are more educated but not older than the owners/managers of informal firms.

The data on firms operating in services do not reveal much difference between informal and formal firms. There is no significant difference between informal and formal service firms in terms of size, productivity and capital intensity, but formal firms pay higher wages than informal firms do (the wage differential is, on average, about 20 percent). In spite of these similarities, there seems to be self-selection of owners/managers in services as well. The owners/managers of formal service firms are more educated than the owners/managers of informal service firms.

The survey asks a question on how much trust the respondents have in a number of people and institutions (“most people living in this country”, “police”, “municipal officials”, “TÜSİAD”, “TOBB”, “judges”, “workers’ unions”, “finance/tax officials”, “political parties”, “government”, “members of the Armed Forces”, “most people he/she do business with”). Since there is a strong correlation between these variables, we calculated the average value for each respondent (the “trust” variable in Table 9). There seems to be a significant difference between formal and informal firms. Owners/managers of informal firms have a much lower trust in institutions their formal counterparts, both in manufacturing and services. This finding is consistent with the stylized fact that those people who do not have any strong faith in legal and business institutions are more likely to operate in the informal sector.

4.2. Determinants of informality

We estimated a logit model of informality by using the World Bank dataset to identify the variables that have a significant impact on the choice of informality. The size and its quadratic term (log number of employees, ll , and its square, ll^2) are used to check the relationship between firm size and informality, whereas the (log) age of the firm ($lfage$) is included to test the life-cycle hypothesis. A number of individual-level variables, on gender, age (age and its square, age^2), experience (log number of years in the same job, $lexper$), educational level ($educ2-educ5$), and trust in legal and business institutions ($trust$), are included.

Estimation results are presented in Table 10. Compared to the ERF database (see Table 5), there are fewer number of variables with statistically significant coefficients. Interestingly, there seems to be an inverted-U type relationship between firm size and informality. The probability of informality first increases by size (up to 7 employees in manufacturing and 24 employees in services), then tend to decline. Contrary to our *a priori* expectations, older firms are more likely to operate informally in services. In both sectors, those firms whose owners/managers have university diploma are less likely to operate in the informal sector. Finally, trust in institutions has a significant impact on the decision to be informal.

Before discussing the findings on the performance differences between informal and formal firms, a major caveat for the current analysis needs to be emphasized. The models for informality have low explanatory power, and the model for services as whole is not statistically significant at the 10 percent level (see the Chi-square statistic in Table 10). This is possibly because of the lack of sufficient number of observations on informality (40 out of 393 firms in manufacturing and 46 out of 514 firms in services). Since the informality models have low explanatory power, the analyses that control the process of self-selection by using these models are likely not to be robust.

4.3. Performance and productivity differences between informal and formal firms

We estimate Equation 3 for a number of variables to test if the differences between the average values of these variables for informal and formal firms are statistically significant or not. We first compare the differences for all informal and formal firms, and, then, by forming a control group of formal firms by using the propensity score matching method, between the groups of informal firms and a similar set of formal firms (the control group).

The average effects of informality are presented in Table 11. Although formal manufacturing firms are more productivity, use more capital intensive technologies, and pay higher wages than their informal counterparts, none of these differences are statistically significant at the 10 percent level. But, the owners/managers of formal firms strongly believe that they are more profitable than an average firm in the market.

Formal service firms are more productive than their informal counterparts, but they neither use more capital intensive technology, nor pay higher wages than their informal counterparts. There is also no significant difference in relative profitability between informal and formal service firms. Interestingly, the self-assessed level of technology in formal service firms is lower than the level in informal firms.

Formal firms are more likely to be incorporated companies, and pay wages through a bank. These findings reveal that formal firms are more likely to cooperate with formal financial institutions, but these relationships do not lead to a significant advantage in terms of using credit: the difference in the proportions of informal and formal firms that use credit in the last year is not statistically different from zero.

Finally, we estimated production functions to understand the sources of productivity differences between informal and formal firms. Since the number of informal firms is rather small, we are not able to estimate separate production functions for informal and formal firms. We assume that formal and informal firms operate on the same production function, but their productivity levels (the “total factor productivity” term, i.e., the intercept term) differ.

We estimated treatment effects model in which a variable on the status of informality, *informal*, is added into the production function. The coefficient of the *informal* variable is a measure of productivity gap between informal and formal firms. Since the informal variable is endogenous to the model, it is assumed to be a linear function of a set of variables (for explanatory variables, see Table 10), and the production function and the informality decision model in which the *informal* variable is the dependent variable are estimated simultaneously by using the maximum likelihood method.

The estimation results for production functions of manufacturing and services are presented in Table 11. The results should be interpreted with extra caution because, as we have seen in the previous subsection, the low explanatory power of the informality models are likely to cause weak instrument problems.

There are two models estimated for both manufacturing and services. The first model includes only the input variables (capital, material inputs, and labor), and the informality variables, whereas the second model includes, in addition to these variables, a number of variables on the educational level and gender of the owner/manager, and dummy variables on technological level, large suppliers, and large customers.

The coefficients of all input variables are statistically significant. The estimation results suggest that there are mild increasing returns to scale in manufacturing and services. Educational level of owners/managers has a weak, positive impact on productivity. Dummy variables for technology, large suppliers, and large buyers do not have statistically significant impact on productivity. The main variable under investigation, the *informal* variable, has a statistically significant negative coefficient in both sectors, i.e., the average level of total factor productivity is substantially lower in informal firms than in formal firms.

5. Informality and productivity: Individual-level evidence

5.1. *The data*

We use individual-level wage data to double check the degree of productivity differentials between informal and formal firms/employees, and to identify the sources of these differences. The wage and labor market participation data come from the Labor Force Surveys (LFS) conducted by Turkstat. The LFS questionnaire includes questions about a large number of variables for each household and individual surveyed on labor market status. The survey includes questions regarding the wage income for wage workers, and social security protection. We use the data for 2006 to estimate Mincer-type wage equations. The same analysis was performed by using the 2005 data. Since the results were qualitatively same, we will report here only the findings from the 2006 dataset.

The wage rate at the individual level is expected to be positively correlated with labor productivity. If the labor market is competitive, the wage rate will be equal to the marginal product of labor, and any difference in the wage rates in the informal and formal sectors can be used as an indication of productivity differences. In this framework, the marginal

product of labor, and, consequently, the wage rate will differ by the level of human capital. Thus, there will be no difference in wages and productivity once human capital and other factors are accounted for.

If the labor market is not competitive, and, for example, the wage rate is determined by a bargaining between workers and the firm, the wage rate will be positively correlated with productivity because the workers will share a part of productivity. In such a case, we observe a positive correlation between wages and productivity, but a wage differential between informal and formal firms may arise without any productivity differential if the workers in the formal sector have a stronger bargaining power than those workers employed in the informal sector. Thus, the findings on wage differentials could be interpreted with caution.

As in the previous section, we here define “informal employment” as those employees who are not covered by any social security scheme. Since employment dynamics are very different in urban and rural areas and most of the employees are informal in rural areas, we analyze only the urban regions.

Table 13 presents the data on the extent of formal employment for male and female employees by age and employment status.²⁰ As in the case of entrepreneurs, there seems to be a U-shape relationship between workers’ age and informality. The share of formal employees is much lower among young and old people. Females are more likely than males to work in the informal sector. The share of informality is especially very high among casual workers.

There is a positive correlation between the extent of formality and firm size in both manufacturing and services (Table 14). Although more than half of all employees working in firms employing less than 10 people are informal, the share of formal employees exceeds 90 percent in firms that employ 50 or more people.

²⁰ Unless otherwise stated, all data used in this section are from the 2006 LFS.

There are significant wage differentials between informal and formal workers even after controlling for gender and firm size. The wage differential is much higher in services than in manufacturing (about 55 percent in services and 35 percent in manufacturing), and somewhat higher for female workers than male workers. The wage differentials does not change much by firm size, i.e., irrespective of firm size, formal firms pay higher than their informal counterparts in the same size category.

5.2. The model

We estimate a Mincer-type reduced form wage equation to analyze wage differences between the informal and formal sectors. Since the wage rate is observed only for those people who are employed as wage workers, the selection process needs to be taken into account in estimating the wage equation. Thus, we use a multinomial logit selection model to take care of the selection process. The model is defined as follows:

$$(7) \quad U_{ji} = h(Z_i) \\ w_{ki} = \beta_k X_{ki} \text{ for } U_{ki} \geq U_{ji}, j \in \{h, fm, fs, im, is, e, se\}$$

where U_{ji} is the utility of being at labor market state j for individual i , w is the (log) wage rate for wage workers.

We assume that there are six types of jobs available for an individual: formal jobs in manufacturing (*fm*) and services (*fs*), informal jobs in manufacturing (*im*) and services (*is*), entrepreneur (employer, *e*) and self-employed (*se*). If the person stays at home, she is considered to be at the “non-employment” status (*h*). The person will chose the state that maximizes her utility. If the chosen labor market state is any one of wage employment (*fm*, *fs*, *im*, or *is*), the wage rate is observed for that individual.

We estimate a multinomial logit model to understand the determinants of labor market decision for men and women aged 15 or more living in urban areas. There are seven labor market outcomes and the base outcome is non-employment.²¹ The next four outcomes are

²¹ This outcome includes non-participation in the labor market, and unemployment. We experimented with separate “agriculture” and “unpaid family worker” outcomes as well, but the estimation results for these outcomes were not significantly different from non-employment in most of the cases. Therefore, we added “agricultural employment” and “unpaid family workers” to the “non-employment” category.

about wage employment in formal/informal manufacturing and services. The last two outcomes, the “employer” and “self-employed” categories are not classified into sectors because of the lack of sufficient number of observations. The following variables are used as explanatory variables in the multinomial logit model.

The age of a person has a significant impact on labor market decisions. We added a number of age dummies to allow for non-monotonic effects of the age variable. “15-19 years old” is the omitted age dummy. We expect that there would be a U-shaped relationship between age and the probability of non-employment outcomes.

The status in the household is captured by the *Child* dummy that takes the value 1 if the person is “daughter/son”, “daughter-/son-in-law”, “grand daughter/son”, or “other relative/non relative” aged less than 30, and 0 otherwise. The omitted variable is the “Parent” category that includes all other people not included in the *Child* category.

The effects of educational level are captured by five dummy variables: *Primary* for literates and primary school graduates²², *Secondary* for secondary school graduates, *High school* for high school graduates, *Vocational* for vocational high school graduates, and *College* for 2- and 4- year higher education graduates. The omitted variable is the *Illiterate* category, i.e., the educational level dummy variables measure the effects of relevant levels of education relative to illiteracy.

There are two dummy variables for marital status: *Single* for never-married singles, and *Divorced* for the divorced and widowed. The omitted variable is the *Married* category.

In order to test the effects of household size, we include to the model the *Parent*household size* and *Child*household size* interactions. The household size is measured by the (log) number of people in the household. It is interacted with the *Parent* and *Child* dummy variables because the effects of household size on parents and children are likely to differ. We expect that the *Parent*household size* variable may have a negative effect on the probability of non-employment for men, but it may have a positive effect on

²² The “literate without any diploma” and “primary school diploma” categories are merged together because the number of people in the former category was too low.

the probability of non-employment for women, because women are more likely to be involved in home production, and the need for home production will increase by the household size. In other words, the value of household production will increase for women by household size, whereas the need for workplace employment will increase for men.

We use a dummy variable, *Any registered*, to test if the availability of social security benefits provided by another formally employed person in the household increases the non-employment probability and decreases the formal employment probability of other persons in the household. We expect that if there is a formally employed person in the household, other members of the household are likely to benefit from social security coverage (health insurance, etc.) so that if they get a formal job, the value of additional non-wage benefits will be low. This may discourage other household members to get a (formal) job.²³

Finally, we use a dummy variable for persons in a household whose “head” is unemployed, *Unemployed HH*. This variable takes the value 1 for a person whose household head is not employed, and 0 otherwise.²⁴ The household head is expected to earn the bread and butter for the household. If the household head is unemployed, the incentives for other household members may change.

5.3. Estimation results

The estimated coefficients from the multinomial logit model are difficult to interpret quantitatively. Thus, we calculated the marginal effects of each variable on labor market outcomes. For continuous explanatory variables, the marginal effect is the change in the probability of the relevant outcome’s realization in response to a slight change in the dependent variable i.e., the marginal effect is defined as $\partial \text{Pr}(j) / \partial x_k$ where $\text{Pr}(j)$ is the probability that the labor market outcome j will be chosen, and x_k is the k th explanatory variable. For dummy variables, the marginal effect shows the change in the probability $\text{Pr}(j)$ induced when the dummy variable changes from 0 to 1.

²³ Although the employment decision, as discussed here, is likely to be made at the household level, we model it at the individual level because of the lack of panel dimension in our data.

²⁴ By definition, the value of this variable is zero for household heads because we test the effect for other household members.

The marginal effects on labor market participation decision are presented in Table 16. It is apparent that education is one of the most important factors that determine the labor market participation decision. The probability of unemployment decreases significantly by education level for both men and women. Moreover, more educated people are less likely to be employed in the informal sector, i.e., self-selection is evident in entering informal/formal sectors.

We estimated Mincerian wage equations to understand wage differentials between informal and formal sectors. Since the wage rate is observed for only those who work as wage workers, we used multinomial logit sample selection model in estimating the wage equations to eliminate the selection bias. The wage equation is estimated, by taking into account the selection process, for four categories of wage employment (formal/informal and manufacturing/services) for men and women separately. The following variables, in addition to the sample selection correction terms, are included into the model.

The dummy variables for age categories are used to estimate the age-wage profile and to estimate the effects of age/experience on wages. All educational level dummies are included to estimate the effects of education on wages. Dummy variables for firm size (small, 10-24 employees; medium-sized, 25-49 employees; large, 50 and more employees) are added to test the effect of firm size on the wage rate because it is known that large firms tend to pay higher wages. The micro-firm (less than 10 employees) dummy is the omitted firm size variable. The (log) working time per week and a dummy variable for full time employees are used to control for working time and employment status on wages.

The estimation results for the wage equation are presented in Table 17. Our findings indicate that, with the exception of female workers employed in manufacturing, there are significant wage gaps between informal and formal workers even after controlling for selection and individual characteristics. These differences in wages are likely to be caused by productivity differentials between informal and formal firms. Moreover, educational level and firm size have also positive effect on wages. These two factors also contribute the observed wage differentials because, as we have seen before, the share of formal employees is much higher among more educated and large firms.

6. Enforcing formality: A simulation analysis

Our findings in the previous sections suggest that there are substantial productivity differences between informal and formal firms. The policy of enforcing formality could be effective in increasing productivity in the economy. However, our analysis has also documented that entrepreneurs self-select into informality by taking into account the profits they could make under these two conditions, and there are significant differences between the characteristics of entrepreneurs and firms operating in the informal and formal sectors.

Figure 2 depicts the kernel density (frequency) functions for informal and formal firms by size. As expected, informal firms have a higher concentration in small size in both manufacturing and services, whereas the distribution of formal firms is skewed to the right, i.e., larger size. The observed size differences between informal and formal firms can be explained by the probability of enforcements (large firms are more likely to be identified and inspected by the authorities), and productivity differentials because small firms are, on average, less productivity and thus has a stronger incentive to operate informality to reduce the cost of compliance.

There is also a strong correlation between entrepreneur's educational level and the extent of informality (see Figure 3). More educated entrepreneurs are more likely to operate formally, due to cultural factors, types of activities conducted, etc.

Finally, there are significant differences between informal and formal firms in terms of entrepreneurs' and firms' ages (see Figures 4 and 5). Young entrepreneurs and young firms are more likely to operate informally. We can observe similar differences in other variables that determine the choice of informality (see Table 5).

Our analysis indicates that some entrepreneurs prefer to establish informal firms because it is more profitable to do so for their characteristics that cannot be changed easily. Thus, if those entrepreneurs who would prefer to work informally are forced to operate formally

through strict enforcement, they could not achieve the productivity level achieved by those entrepreneurs who operate formally. We need to take into account the fact that entrepreneurs' characteristics (like gender and educational level) will not change at least in the short- and medium-term after enforcing formality.

We experiment with three scenarios to understand the employment effects of strict enforcement of formality. In the first scenario, we assume that all informal firms start to operate formally. In the second scenario, we assume that all informal firms exit from the market because they lose their competitiveness as a result of high costs of formality (social security contributions, taxes, etc.), and the formal firms expand, or new formal firms are established to satisfy the demand that was satisfied by informal firms such that total output does not change. It is obvious that these two scenarios represent two extreme cases: all informal firms survive in the first one, and all informal firms exit in the second one. The third scenario is the same as the second one, but we assume that total output increases by 10 percent due to lower prices achieved thanks to higher productivity or lower taxes.

Scenario 1: All informal firms become formal

In this scenario, we assume that all informal firms start to operate formally due to strict enforcement, and they switch to the production function of formal firms without any change in input levels (same amounts of capital, material inputs, and labor), and entrepreneurs' characteristics. However, we assume that other firm characteristics (technology, composition of labor, customer relations, etc.) change as a result switching to formality. Thus we use the second estimated models of production functions in Table 7.

In order to assess the productivity effect of enforcing formality, we need to calculate the conditional output level for informal firms. The conditional expected output level of an informal firm is defined as follows:

$$(8a) \quad E(q_{li} | I_i = 1, X_{li}) = \beta_1 X_{li} + \sigma_1 \rho_1 f(\gamma Z_i) / F(\gamma Z_i)$$

$$(8b) \quad E(q_{li} | I_i = 0, X_{li}) = \beta_0 X_{li} + \sigma_1 \rho_1 f(\gamma Z_i) / (1 - F(\gamma Z_i))$$

where $f(\cdot)$ and $F(\cdot)$ are probability density and cumulative distribution functions, respectively, and ρ_1 is the coefficient of correlation between ε_{1i} and u_i (see Equation 4). Equation 8a defines the expected output level of an informal firm, i , conditional on operating informally, and Equation 8b defines the expected output level for the same firm conditional on operating formally. Note that the equations take into account the differences in production functions (the β parameters), and the effects of the selection process.

The average (log) expected conditional output levels for informal and formal firms are presented in Table 18. The average (log) output of informal manufacturing firms operating informally in our sample of firms was 12.62. The output of these firms would be 12.67 had they operated formally, i.e., formality would increase their output by only 5 percent.²⁵ The output (and productivity) effect of formality on existing informal firms would be relatively higher in services: the average increase in output due to formality is about 25 percent in services.

Figure 6 depicts the kernel density function of (log) output for informal and formal firms in manufacturing and services. The dotted line (“Informal mod”) represents the expected output of informal firms conditional on operating formally. As seen in the figure, formality has a stronger positive effect on small informal manufacturing firms, but larger informal manufacturing firms would produce less if they operate informally, so that the average value would not change much. In the case of services, we observe almost a parallel shift in the kernel density function for informal firms.

To summarize, our estimation results indicate that if all informal firms start to operate formally, the output and productivity effect would be negligible in manufacturing, whereas output and productivity would increase considerably in services.

Scenario 2. All informal firms exit from the market

In the second scenario, we assume that the productivity improvement due to formality would not be sufficient to compensate for the costs of formality (higher taxes and social security contributions) so that all informal firms will exit from the market. The loss of

²⁵ Since we assume constant input levels, the increase in output reflects productivity enhancement.

output due to the exit of informal firms will be balanced by formal firms. Since formal firms are more productive, we expect a decrease in total employment.

In order to calculate the employment effects in this scenario, we need to estimate productivity differentials between informal and formal firms. We define “productivity” here as output/labor ratio because we assume that there would be no change in total output.

The productivity differential between informal and formal firms in the ERF sample data is 131 percent in manufacturing and 24 percent in services. The differential in manufacturing seems to be substantial but it is in the order of magnitude of the productivity differential between very small firms that employ 10-24 people and others that employ 25 or more people. According to the Turkstat *Annual Survey of Manufacturing Industries* in 2000, the productivity differential between these two groups of manufacturing firms was about 191 percent. Thus, we use the ratios calculated from the ERF database in our scenario.

We use the *Labor Force Survey 2006* data to analyze changes in the composition of employment as a result of enforcing formality. Table 19 presents the number of informal (unregistered) and formal (registered) employees in urban regions in 2006. There were about 11 million formal and 5.5 million informal employees in manufacturing and services in urban regions (agriculture is excluded from our analysis). We assume that formal manufacturing (service) employees produce 131 (24) percent more than informal employees do. If all informal jobs are replaced with more productive formal jobs, total employment (all formal) would drop to 15 million, i.e., there would be 9.7 decline in employment.²⁶ The most significant decline in employment would be observed in manufacturing (18.2 percent decline) because of substantial productivity differential and widespread use of informal labor in manufacturing. Average wages would increase by 18 percent because formal employees are paid much higher than informal employees.

Simulated changes in the composition of employment reveal that the burden of transition from informality to formality is not shared equally among different employment categories. Uneducated (less than high school level), young (aged 15-24), old (aged 45 and

²⁶ If the productivity differential between informal and formal manufacturing firms was 50 percent, instead of 131 percent, the decline in total employment would be 7.7 percent.

above) and female employees would experience substantial employment loss. People who would benefit the most from formality are those with vocational school training or university degree holders, and mid-aged (35-44 years old) people.

Scenario 3. All informal firms exit from the market and the market grows 10 percent

In the previous scenario, we assumed that there is no change in total output. However, when all informal firms exit from the market, there would be an increase in average productivity as a result of the increase in the share of more productive sector (the formal sector). Thus, the supply function would shift to the right, and cause a decline in the price and an increase in the quantity demanded. Moreover, the increase in the share of formality could lead to a substantial increase in tax revenue and social security contributions (we expect about 37 percent increase in tax revenue and social security contribution in the second scenario). If the government reduces tax and social security rates, the prices could decline further, and total output would receive another impetus. In this scenario, we assume that total output increases by 10 percent because of these two factors.

A 10 percent increase in output would compensate almost completely the loss of employment (see Table 20). Although the aggregate employment would remain almost at the same level, its composition would change significantly. As in the case of scenario 2, the main disadvantaged groups are less educated, young, old, and female employees.

Our simulations on hypothetical changes in enforcement reveal that transition from informality to formality would be beneficial in raising productivity, wages, and social security revenue. However, such a transition involves a serious and painful adjustment. First, a large number of informal firms are likely to exit from the market because the existing informal entrepreneurs would not be able to achieve a sizeable productivity increase when they operate formally. Second, even if formal firms fill in the vacuum created by the exit of informal firms and generate a compensating growth, the composition of employment changes radically. Disadvantaged groups (less educated, young and old, and female) would experience a decline in the demand for their labor, whereas more educated, mid-aged and male workers would have a much better employment prospects. Therefore, the public policy towards eliminating the informal sector should be

complemented by a policy supporting the disadvantaged groups during the transition process, and a comprehensive training and education policy to satisfy increasing demand for more educated labor.

The transition toward formality would increase the demand for more educated workers. The shift in the demand function would increase the returns to education, and would create (private) incentives to invest in education. This could be especially important for female workers who have traditionally low labor market participation rates.

7. Conclusions

Our main findings can be summarized as follows:

1. There is a significant productivity gap between informal and formal firms, and a wage gap between informal and formal workers. The findings are robust with respect to sectors (manufacturing and services), firm size and gender.
2. The hypothesis that more educated entrepreneurs and workers move to the formal sector is supported by the data. This process of self-selection contributes to widen the productivity gap between informal and formal firms.
3. The theories of life-cycle and learning are also supported by our findings. Older (i.e., more experienced) firms tend to operate in the formal sector. However, the relationship between informality and age is U-shaped for entrepreneurs and workers. New firms tend to start their operations in the informal sector and the point of entry to employment is the informal sector for (male) workers. Entrepreneurs and workers tend to move to the formal sector over time, but after a certain age (the late 40s-early 50s), they tend to have informal jobs/businesses possible for their secondary activities.
4. The hypothesis on the importance of economies of scale for productivity growth is not supported by the data. Informal and formal firms are likely to operate on different production functions, and decreasing returns to scale prevail for informal firms, whereas

there are constant returns to scale for formal firms. If informal firms could move to the formal sector adopt a new production technology, the net effect would be similar to the effect of economies of scale.

5. Even after controlling for all these factors (self-selection, differences in endowments, and learning), the productivity gap does not disappear. The remaining gap is due to access to public services and infrastructure, access to markets, and unobserved factors.

6. The findings suggest that there is a substantial but untapped potential to increase productivity in Turkey. The analysis on the sources of productivity differentials suggests that policies towards reducing the size of the informal economy would be beneficial in increasing productivity and boosting growth, and both “stick” (enforcement) and “carrot” (training, clustering, market access, etc) policies would be adopted to reduce informality. If informal firms are forced to operate formally through strict enforcement, many informal firms in manufacturing are likely to exit from the market because they could not increase their productivity sufficiently when they operate formally. The productivity effect of operating formally is higher for services, but we may expect that a large number of informal service firms could not survive if they operate formally.

7. The transition to formality would be a painful process especially for disadvantaged groups (less educated, young, old, and female). Therefore, policies aimed at reducing informality should be complemented by social policies that help the disadvantaged groups during transition, and education policies that help to provide skills needed by formal firms.

References

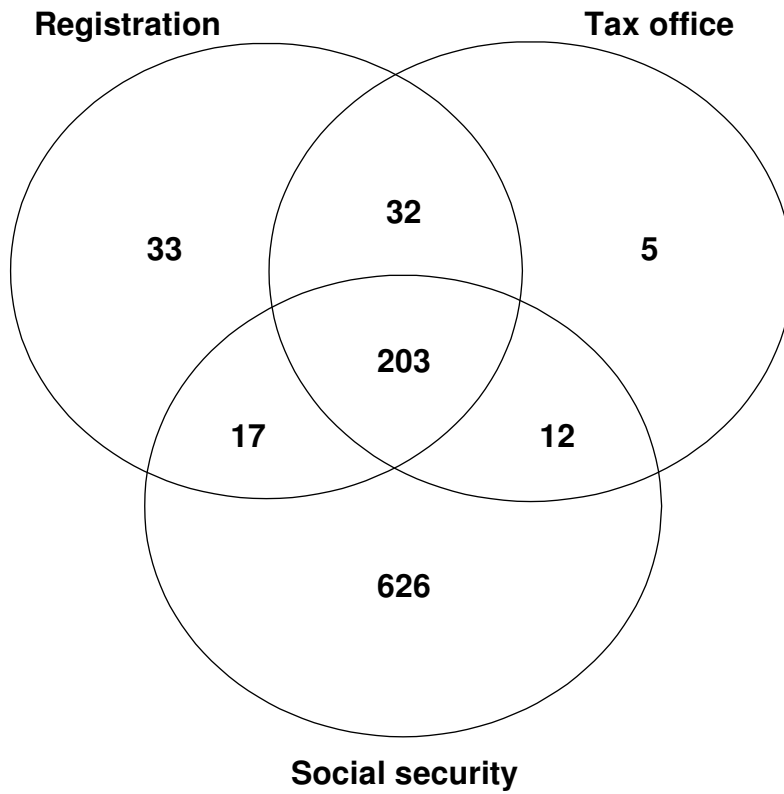
- Acemoglu, D. (2001), "Good Jobs versus Bad Jobs", *Journal of Labor Economics* (19): 1-21.
- Agénor, P. Jensen, H.T., Verghis, M. and Yeldan, E. (2006), *Disinflation, Fiscal Sustainability, and Labor Market Adjustment in Turkey*, World Bank Policy Research Working Paper 3804.
- Albrecht, J., Navarro, L. and Vroman, S. (2005), *The Effects of Labor Market Policies in an Economy with an Informal Sector*, IZA Discussion Paper No. 2141, IZA, Bonn.
- Almeida, R. and Carneiro, P. (2006), *Enforcement of Labor Regulation, Informal Labor, and Firm Performance*, CEPR Discussion Papers No. 5976, CEPR, London.
- Amaral, P.S. and Quintin, E. (2006), "A competitive model of the informal sector", *Journal of Monetary Economics* (53): 1541-1553.
- Arias, O. and Khamis, M. (2008), *Comparative Advantage, Segmentation and Informal Earnings: A Marginal Treatment Effects Approach*, IZA Discussion Paper No. 3916, IZA, Bonn.
- Badaoui, E., Strobl, E. and Walsh, F. (2007), *The Formal Sector Wage Premium and Firm Size*, IZA Discussion Paper No. 3145, IZA, Bonn.
- Boeri, T. and Garibaldi, P. (2005), *Shadow Sorting*, paper presented at the NBER Macroeconomic Conference, Budapest, June 2005.
- Bourguignon, F., Fournier, M. and Gurgand, M. (2007), "Selection Bias Corrections Based on the Multinomial Logit Model: Monte Carlo Comparisons", *Journal of Economic Surveys* (21): 174-205.
- Brown, J.D., Hotchkiss, J.L. and Quispe-Agnoli, M. (2009), *Undocumented Worker Employment and Firm Survival*, IZA Discussion Paper No. 3936, IZA, Bonn.
- Burdett, K. and Smith, E. (2002), "The Low Skill Trap", *European Economic Review* (46): 1439- 1451.
- Dabla-Norris, E., Gradstein, M. and Inchauste, G. (2005), *What Causes Firms to Hide Output? The Determinants of Informality*, IMF Working Paper No. 05/160, IMF, Washington DC.
- Dessy, S. and Pallage, S. (2003), "Taxes, Inequality and the Size of the Informal Sector", *Journal of Development Economics* (70): 225-233.
- Dimova, R., Nordman, C.J. and Roubaud, F. (2008), *Allocation of Labour in Urban West Africa: Implication for Development Policies*, IZA Discussion Paper No. 3558, IZA, Bonn.
- Djankov, S., Lieberman, I., Mukherjee, J. and Nenova, T. (2003), "Going Informal: Benefits and Costs", in B. Belev (ed.), *The Informal Economy in the EU Accession Countries: Size, Scope, Trends and Challenges to the Process of EU Enlargement*, Sofia: CSD, pp. 63-80.
- Fajnzylber, P., Maloney, W.F. and Rojas, G.V.M. (2006), *Releasing Constraints to Growth or Pushing on a String? The Impact of Credit, Training, Business*

Associations and Taxes on the Performance of Mexican Micro-Firms, World Bank Policy Research Working Paper 3807, Washington, DC: World Bank.

- Galiani, S. and Weinschelbaum, F. (2007), *Modeling Informality Formally: Households and Firms*, CEDLAS Working Papers No. 0047, Universidad Nacional de La Plata .
- Gönenç, R., Leibfritz, W. and Yilmaz, G. (2007), *Enhancing Turkey's Growth Prospects by Improving Formal Sector Business Conditions*, OECD Economics Department Working Papers, No. 542, OECD Publishing, doi:10.1787/316551532607.
- Henley, A., Arabsheibani, G.R. and Carneiro, F.G. (2006), *On Defining and Measuring the Informal Sector*, IZA Discussion Paper No. 2473, IZA, Bonn.
- Husmanns, R. (2004), *Measuring the Informal Economy: From employment in the Informal Sector to Informal Employment*, ILO Working Paper No. 53, ILO, Geneva.
- Jovanovic, B. (1982), "Selection and The Evolution of Industry", *Econometrica*, 50: 649-670.
- Heckman, J. J., and Hotz, V. (1986), "An Investigation of Labor Market Earnings of Panamanian Males", *Journal of Human Resources* (21): 507-542.
- E. Leuven and B. Sianesi. (2003). *PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing*, <http://ideas.repec.org/c/boc/bocode/s432001.html>. This version 3.0.0.
- Levenson, A.R., Maloney, W.F. and Ventura, G. (1998), *The Informal Sector, Institutional Participation, and Micro-firm Dynamics*, unpublished paper.
- Masatlioglu, Y. and Rigolini, J. (2005), *Informality Traps*, unpublished paper.
- McKenzie, D. and Sakho, Y.S. (2007), *Does It Pay Firms to Register for Taxes? The Impact of Formality on Firm Profitability*, IZA Discussion Paper No. 3179, IZA, Bonn.
- McKinsey Global Institute (2003), *Turkey: Making the Productivity and Growth Breakthrough*, New York: McKinsey.
- Ozar, S. (2006), *Micro and Small Enterprises in Turkey: Uneasy Development*, Project on "Promoting Competitiveness in Micro and Small Enterprises in the MENA Region", ERF Research Report Series No.: 0420, Cairo.
- Pagés, C. and Stampini, M. (2007), *No Education, No Good Jobs? Evidence on the Relationship between Education and Labor Market Segmentation*, IZA Discussion Paper No. 3187, IZA, Bonn.
- de Paula, Á. and Scheinkman, F.A. (2008), *The Informal Sector*, NBER Working Paper No: 13486, Cambridge, Mass: NBER.
- Perry, G.E., Maloney, W.F., Arias, O.S., Fajnzylber, R., Mason, A.D. and Saavedra-Chanduvi, J. (2007), *Informality: Exit and Exclusion*, Washington, DC: World Bank.
- Pratapa, S. and Quintin, E. (2006), "Are Labor Markets Segmented in Developing Countries? A Semiparametric Approach", *European Economic Review* (50): 1817-1841.

- Redding, S. (1996), “The Low-Skill, Low-Quality Trap: Strategic Complementarities between Human Capital and R&D”, *Economic Journal* (106): 458-470.
- Rogers, C.A. and Swinnerton, K.A. (2004), *A Model of Informal Sector Labor Markets*, unpublished paper.
- Saracoğlu, D.Ş. (2008), “The Informal Sector and Tax on Employment: A Dynamic General Equilibrium Investigation”, *Journal of Economic Dynamics and Control* (32) 529–549.
- Snowder, D.J. (1994), *The Low-Skill, Bad-Job Trap*, IMF Working Paper WP/94/83.
- Straub, S. (2005), “Informal Sector: The Credit Market Channel”, *Journal of Development Economics* (78): 299-321.
- Tannuri-Pianto, M., Pianto, D.M. and Arias, O. (2004), *Informal Employment in Bolivia: A Lost Proposition?* 2004 Econometric Society Latin American Meetings, No. 149.

Figure 1. Number of informal firms, by type of informality, 2001



Source: Calculated from the ERF dataset on small firms

Figure 2. Firm size and informality

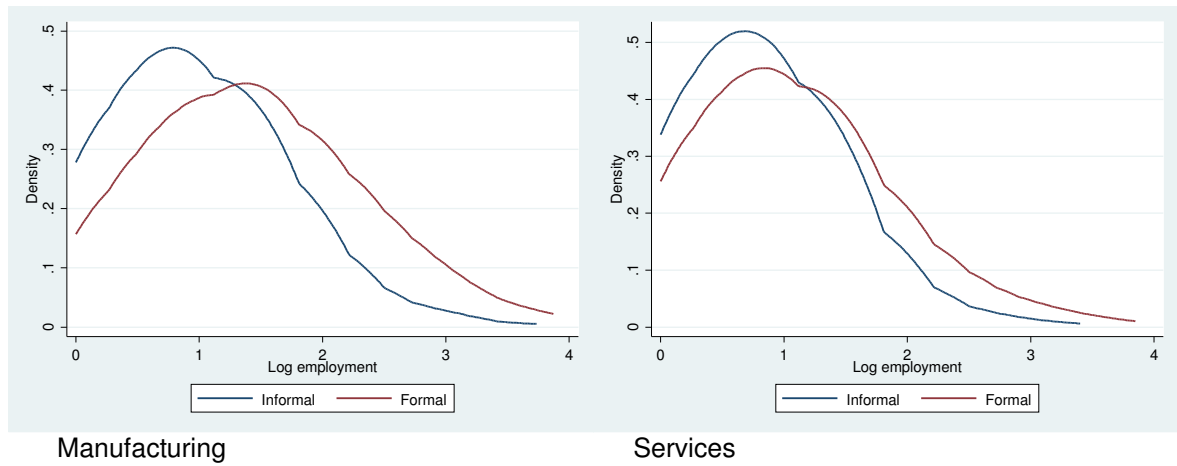


Figure 3. Entrepreneur's education level and informality

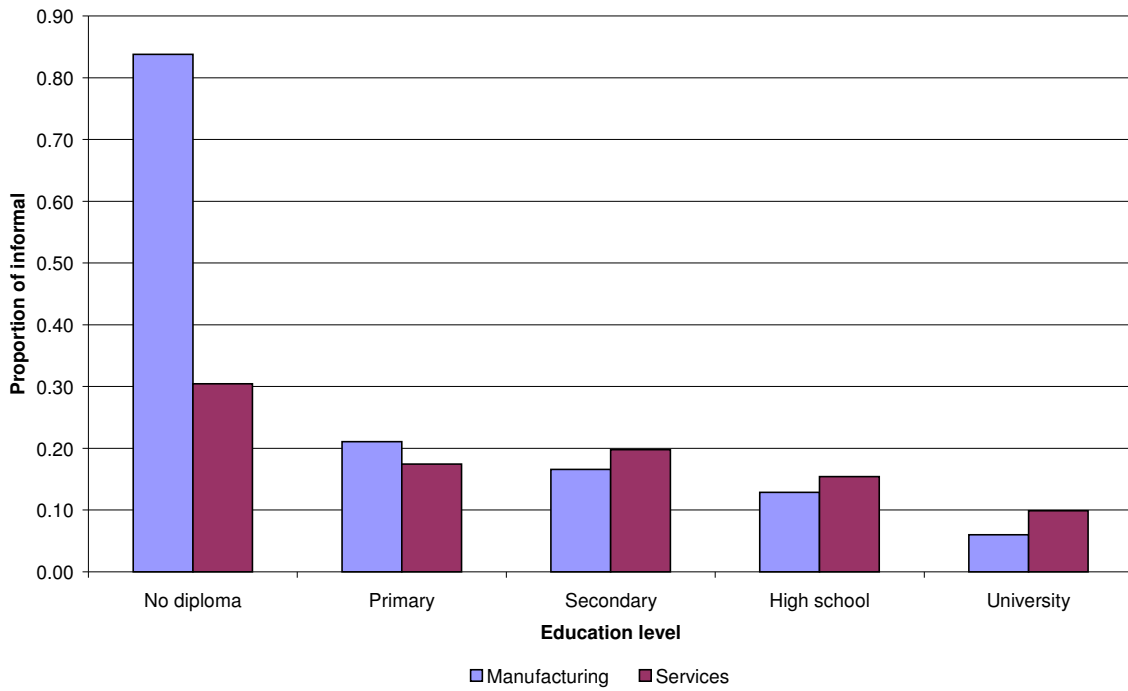


Figure 4. Entrepreneur's age and informality

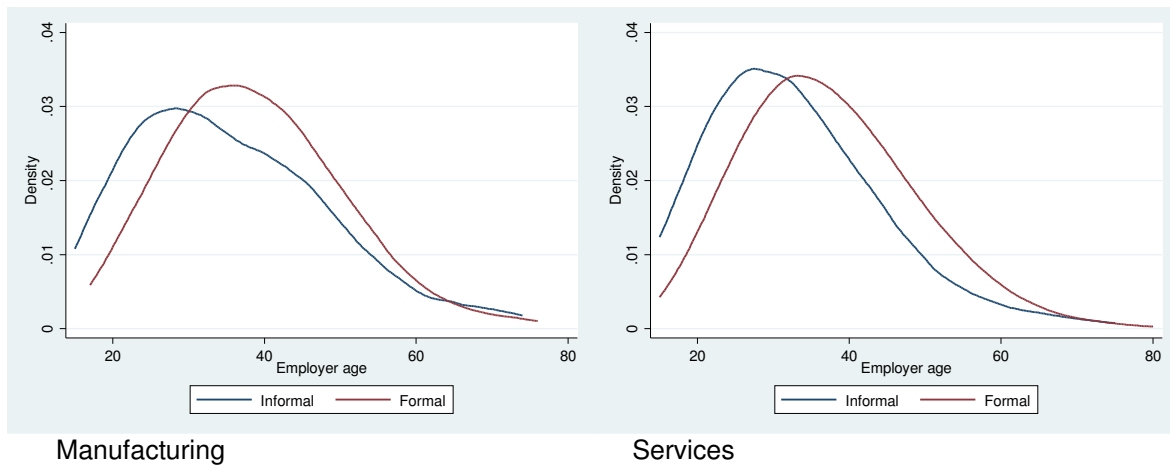


Figure 5. Firm's age and informality

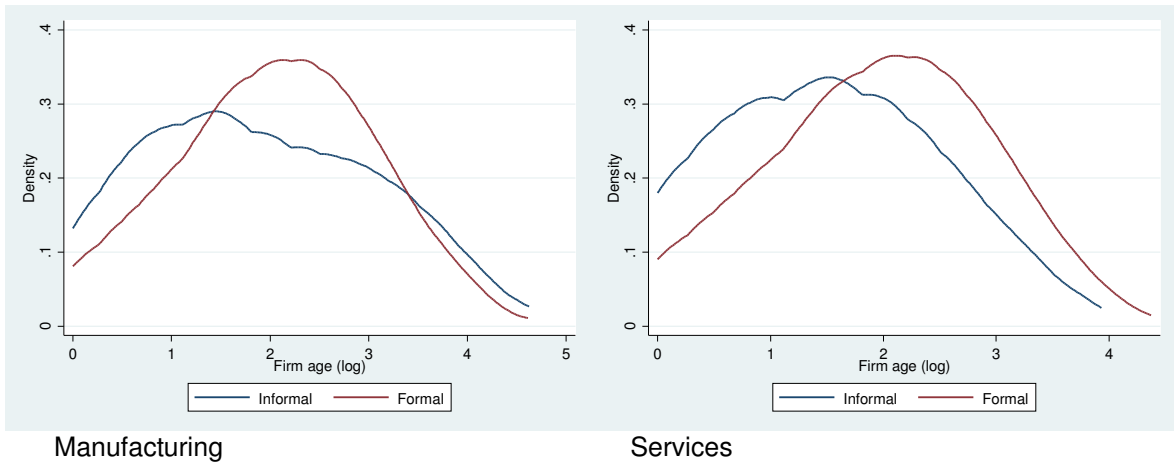


Figure 6. Output effects of formality

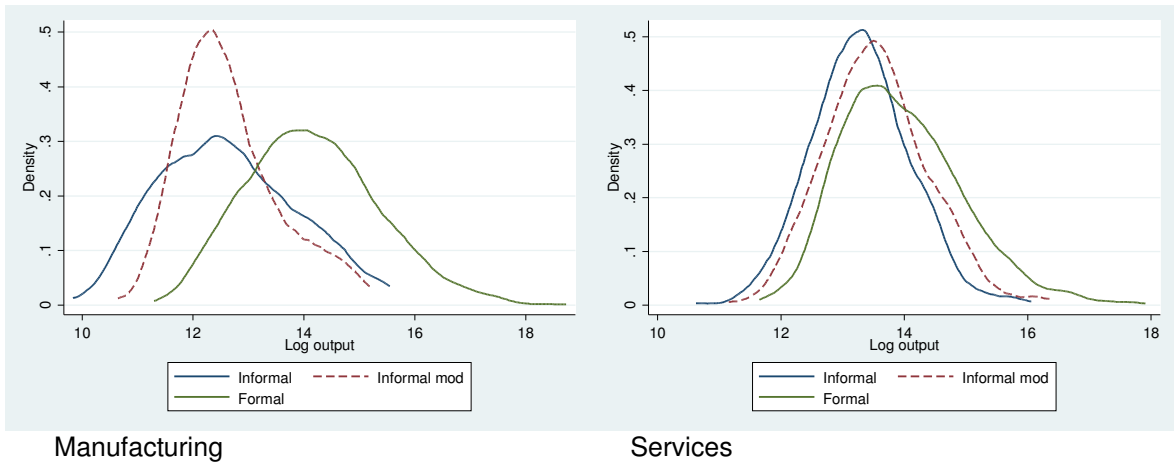


Table 1. Proportion of formal firms, by types of formality, 2001

	Manufacturing	Services	Total
Registration	0.887	0.962	0.943
Tax office	0.892	0.970	0.950
Social security	0.815	0.833	0.828
<i>n</i>	1288	3712	5000

Source: Calculated from the ERF dataset on small firms

Table 2. Reasons for informality at the time of start-up, 2001
(Proportion of firms)

	Type of informality		
	Registration	Tax	Social security
Expensive	0.307	0.268	0.539
Tedious	0.057	0.088	0.058
Useless	0.421	0.418	0.225
Other	0.215	0.227	0.178
<i>n</i>	228	194	929

Source: Calculated from the ERF dataset on small firms

Table 3. Degree of formality (social security), 2001
(Proportion/number of firms)

<i>Firm size</i>	Manufacturing	Services
<i>All observations</i>	0.815 1288	0.833 3712
<i>Firm size</i>		
1	0.648 162	0.791 746
2-4	0.780 649	0.815 2235
5-9	0.912 296	0.924 497
10-19	0.923 130	0.942 173
20-49	0.961 51	0.967 61
<i>Entrepreneur's gender</i>		
Male	0.862 1158	0.838 3498
Female	0.400 130	0.752 214
<i>Entrepreneur's education</i>		
Less than primary	0.162 37	0.696 69
Primary	0.789 593	0.826 1479
Secondary	0.834 175	0.802 607
High school	0.871 350	0.846 1213
University	0.940 133	0.901 344

Source: Calculated from the ERF dataset on small firms

Table 4. Descriptive statistics, 2001

Variable	Legend	Manufacturing		Services	
		Informal	Formal	Informal	Formal
<i>Firm characteristics</i>					
Output (log), value of sales	lq	12.586	14.247	13.311	13.990
Capital (log), value of machinery and equipment	lk	13.869	16.120	14.076	15.017
Inputs (log), raw materials, energy, etc.	lm	11.454	12.449	11.445	11.378
Labor (log), number of employees	ll	0.903	1.388	0.759	1.017
Labor ^a , number of employees		2.467	4.009	2.135	2.764
Labor productivity (log)	lp	11.503	12.607	12.249	12.827
Capital intensity (log), capital/labor ratio	lkint	12.947	14.729	13.304	13.995
Firm age ^a	fage	6.282	7.436	4.223	6.791
Technology (traditional/modern)	technology	0.332	0.528	0.418	0.513
Female employee share	femalesh	0.353	0.071	0.077	0.082
Young employee share	youngsh	0.308	0.257	0.260	0.231
Main consumer (household/formal enterprise)	blink	0.151	0.338	0.081	0.154
Main market (local/regional,national)	rlink	0.210	0.415	0.118	0.165
Log wages, all employees	lwage	11.780	12.127	11.791	12.029
Log wages, female employees	lfwage	10.493	11.350	11.342	11.488
Log wages, male employees	lmwage	11.031	10.893	11.073	11.075
Partnership (no/yes)	partners	0.203	0.402	0.261	0.319
Profit margin (profits/sales)	profitmar	0.449	0.445	0.544	0.602
Credit use in the last year (no/yes)	credit	0.231	0.260	0.271	0.281
Location in a cluster (no/yes)	cluster	0.143	0.271	0.097	0.178
<i>Entrepreneur characteristics</i>					
No diploma	educ0	0.130	0.006	0.034	0.016
Primary school	educ1	0.525	0.446	0.416	0.395
Secondary school	educ2	0.122	0.139	0.194	0.158
High school	educ3	0.189	0.290	0.302	0.332
University	educ4	0.034	0.119	0.055	0.100
Vocational training (no/yes)	vocation	0.038	0.167	0.040	0.084
Gender (male/female)	gender	0.328	0.050	0.085	0.052
Age ^a	Age ^a	33.248	37.226	30.958	35.905
Experience in the present job (years)	exper	12.148	13.861	6.465	11.297
Rural origin (no/yes)	rural	0.627	0.326	0.462	0.335
Sole economic activity	sole_act	0.870	0.964	0.950	0.960
Reason for establishing new firm, only option	aimtina	0.118	0.081	0.150	0.093
<i>Number of observations</i>		238	1050	620	3092

Item response rate is lower for some variables.

a Geometric average

Source: Calculated from the ERF dataset on small firms

Table 5. Determinants of informality, 2001
(Switching regression results)

	Manufacturing		Services	
	Coeff	Std dev	Coeff	Std dev
ll	-0.653	0.122 ***	-0.593	0.079 ***
gender	1.874	0.243 ***	0.225	0.207
age	-0.242	0.046 ***	-0.189	0.023 ***
age2	0.003	0.001 ***	0.002	0.000 ***
lexper	-0.171	0.115	-0.324	0.067 ***
lfage	-0.087	0.108	-0.159	0.076 **
rural	0.884	0.179 ***	0.474	0.104 ***
educ2	-0.324	0.292	-0.012	0.142
educ3	-0.632	0.243 ***	-0.381	0.128 ***
educ4	-0.957	0.485 **	-0.383	0.223 *
vocation	-0.907	0.395 **	-0.798	0.256 ***
aimtina	0.405	0.301	0.484	0.147 ***
sole_act	-1.305	0.332 ***	-0.378	0.234 *
cluster	-0.272	0.241	-0.500	0.163 ***
Constant	2.905	1.016 ***	3.438	0.584 ***
n	1224		3408	
Chi-square	185.4 ***		310.4 ***	
Pseudo R2	0.258		0.126	

Robust standard errors in brackets

*** p<0.01, ** p<0.05, * p<0.1

Table 6. Average effects of informality, 2001
(Matching propensity score estimation results)

Variable	Informal (Treated)	Formal (Untreated)		Number of observations	
		All	Control	Informal	Formal
<i>Manufacturing</i>					
lp	11.44	12.59 +	12.51 *	115	596
klint	12.93	14.72 +	14.21 *	179	967
lwage	11.74	12.12 +	12.15 *	62	690
lfwage	10.56	11.30 +	11.04	14	122
lmwage	11.01	10.88	11.11	64	726
partners	0.19	0.40 +	0.24	214	1009
femalesh	0.37	0.07 +	0.27 *	215	1009
youngsh	0.31	0.26 +	0.32	215	1009
blink	0.13	0.33 +	0.18	215	1009
rlink	0.20	0.42 +	0.20	215	1009
technology	0.34	0.52 +	0.36	215	1009
profitmar	0.43	0.44	0.57 *	132	619
credit	0.24	0.25	0.25	215	1009
cluster	0.14	0.27 +	0.12	215	1009
<i>Services</i>					
lp	12.25	12.81 +	12.84 *	362	1922
klint	13.25	13.99 +	13.71 *	483	2642
lwage	11.76	12.02 +	11.96 *	198	1567
lfwage	11.31	11.46	11.65 *	30	326
lmwage	11.04	11.07	11.15	203	1587
partners	0.25	0.32 +	0.24	533	2875
femalesh	0.07	0.08	0.09	533	2875
youngsh	0.26	0.23 +	0.27	533	2875
blink	0.08	0.15 +	0.08	533	2875
rlink	0.12	0.16 +	0.11	533	2875
technology	0.40	0.51 +	0.47 *	533	2875
profitmar	0.55	0.60 +	0.63 *	392	1974
credit	0.28	0.28	0.28	533	2875
cluster	0.10	0.18 +	0.09	533	2875

+ means the difference between informal and formal (all) firms is statistically significant at the 10 % level.

* means the difference between informal and formal (control group) firms is statistically significant at the ,
10 % level, biased-corrected bootstrap estimate based on 250 replications.

Table 7. Production function estimation results, 2001 (Switching regression results)

	Manufacturing						Services					
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
lk	0.2717*** [0.0601]	0.2193*** [0.0318]	0.2393*** [0.0561]	0.1786*** [0.0330]	0.1164** [0.0569]	0.1580*** [0.0364]	0.2743*** [0.0394]	0.2424*** [0.0173]	0.2578*** [0.0408]	0.2216*** [0.0166]	0.2229*** [0.0390]	0.2139*** [0.0164]
lm	0.2126*** [0.0682]	0.1876*** [0.0258]	0.1567** [0.0714]	0.1897*** [0.0249]	0.1449** [0.0669]	0.1817*** [0.0236]	0.2014*** [0.0277]	0.1683*** [0.0155]	0.1910*** [0.0281]	0.1412*** [0.0154]	0.1963*** [0.0282]	0.1310*** [0.0153]
lll	0.3086 [0.2005]	0.7922*** [0.0848]	0.3651* [0.2110]	0.6900*** [0.0924]	0.3733 [0.2318]	0.6839*** [0.1044]	0.3845*** [0.1255]	0.6298*** [0.0491]	0.3080** [0.1299]	0.5477*** [0.0501]	0.3900*** [0.1277]	0.5932*** [0.0604]
educ2			0.4147 [0.2528]	0.3413*** [0.1271]	0.4089* [0.2440]	0.3115** [0.1222]			0.1904 [0.1319]	0.3248*** [0.0712]	0.1713 [0.1324]	0.2882*** [0.0700]
educ3			0.8295*** [0.2616]	0.4325*** [0.1128]	0.8401*** [0.2698]	0.3800*** [0.1128]			0.1847 [0.1363]	0.5862*** [0.0601]	0.2143 [0.1385]	0.5113*** [0.0604]
educ4			0.6473 [0.6654]	0.5929*** [0.1565]	0.4371 [0.5773]	0.4938*** [0.1581]			0.4110 [0.2700]	0.7174*** [0.0989]	0.2819 [0.2823]	0.6159*** [0.1019]
vocation			-0.5526 [0.3808]	0.0020 [0.1252]	-0.5294 [0.4027]	-0.0109 [0.1227]			-0.3081 [0.2409]	-0.2249** [0.0918]	-0.1957 [0.2306]	-0.2193** [0.0907]
gender			-0.1341 [0.3609]	0.3970** [0.1646]	0.4464 [0.4545]	-0.0485 [0.1866]			-0.1379 [0.2506]	0.0873 [0.1136]	-0.5228 [0.3532]	-0.2042 [0.1649]
partners			0.0964 [0.2415]	0.1112 [0.0998]	0.0097 [0.2564]	0.0442 [0.1014]			0.2878** [0.1255]	0.3347*** [0.0616]	0.2413* [0.1237]	0.2512*** [0.0631]
technology					0.5119** [0.2059]	-0.0886 [0.0813]					0.2411** [0.1146]	0.0402 [0.0506]
femalesh					-0.7447* [0.4373]	0.8520*** [0.2424]					0.6367 [0.4813]	0.4627*** [0.1681]
youngsh					-0.3705 [0.3223]	-0.3509** [0.1573]					-0.4885*** [0.1650]	-0.4361*** [0.0930]
blink					0.2221 [0.2729]	0.4502*** [0.0931]					0.0801 [0.2041]	0.1157 [0.0719]
rlink					0.5234* [0.2897]	0.1079 [0.0895]					0.4140* [0.2176]	0.3786*** [0.0797]
Constant	5.5548*** [0.7614]	7.0668*** [0.4908]	6.7175*** [1.2233]	7.1531*** [0.5273]	7.8202*** [1.2809]	8.0111*** [0.5932]	6.4216*** [0.6064]	7.8386*** [0.3223]	6.8303*** [0.7170]	7.9797*** [0.3142]	7.8118*** [0.7923]	8.4284*** [0.3407]
rts	0.793	1.199	0.761	1.058	0.635	1.024	0.860	1.041	0.757	0.911	0.809	0.938
rho	0.4563	-0.4552**	0.3314	-0.3534**	0.3779	-0.3763**	0.3095	0.2589	0.2978	0.0058	0.1180	-0.2825
n	843		843		843		2442		2442		2442	
Log likelihood	-1552		-1529		-1500		-4684		-4595		-4548	
chi2	747.1***		932***		995.4***		738.5***		1048**		1098***	

Robust standard errors in brackets

*** p<0.01, ** p<0.05, * p<0.1

rts: returns to scale parameter

rho: correlation between error terms from the production function model and selection model

Table 8. Degree of formality (social security), 2008
(Proportion/number of firms)

<i>Firm size</i>	Manufacturing	Services
<i>All observations</i>	0.901 434	0.922 602
<i>Firm size</i>		
1-19	0.883 342	0.921 479
20-49	0.949 59	0.919 74
50-249	1.000 24	0.921 38
250+	1.000 9	1.000 11
<i>Entrepreneur's gender</i>		
Male	0.853 238	0.895 325
Female	1.000 7	0.941 17
<i>Entrepreneur's education</i>		
Primary or no diploma	0.776 58	0.850 60
Secondary	0.818 44	0.854 41
High school	0.846 65	0.884 121
Vocational school	0.917 12	0.889 9
University	0.953 64	0.953 106

Source: Calculated from the World Bank dataset

Table 9. Descriptive statistics, 2008

Variable	Legend	Manufacturing		Services	
		Informal	Formal	Informal	Formal
<i>Firm characteristics</i>					
Output (log), value of sales	lq	9.439	9.848	9.596	9.601
Capital (log), value of machinery and equipment	lk	10.817	11.722	10.699	10.809
Inputs (log), raw materials, energy, etc.	lm	9.167	9.890	9.302	8.747
Labor (log), number of employees	ll	1.801	2.102	2.010	1.905
Labor ^a , number of employees		6.055	8.181	7.464	6.718
Labor productivity (log)	lp	7.581	7.720	7.632	7.752
Capital intensity (log), capital/labor ratio	lkint	9.032	9.610	8.716	8.880
Firm age ^a	fage	9.642	11.499	8.719	10.493
Technology (traditional/modern)	technology	0.233	0.365	0.348	0.302
Large suppliers	dlsupp	0.605	0.619	0.532	0.582
Large buyers	dlbuy	0.442	0.499	0.362	0.377
Log wages, all employees	lwage	6.970	7.078	6.748	6.975
Profits relative to the market average (1-10 scale)	relprofit	4.595	5.672	4.956	5.385
Efficiency relative to the market (1-5 scale)	efficiency	2.698	3.060	2.674	3.064
Credit use in the last year (no/yes)	credit	0.349	0.361	0.277	0.334
Export intensity (exports/sales ratio)	expint	0.018	0.072	0.047	0.036
Corporate status (0 individual, 1 incorporated)	corporate	0.419	0.752	0.638	0.699
Wage payments through a bank (no/yes)	wagebank	0.023	0.321	0.128	0.254
Accounts payable (no/yes)	accpay	1.163	1.189	1.340	1.336
Account receivable (no/yes)	accrec	1.233	1.184	1.362	1.470
Training in the last year (no/yes)	train	0.186	0.250	0.277	0.300
Investment in the last year (no/yes)	invest	0.233	0.347	0.255	0.191
Use of subcontracting (no/yes)	subuse	0.070	0.111	0.170	0.164
Producing as subcontractor (no/yes)	subsup	0.163	0.127	0.234	0.158
<i>Respondents' characteristics</i>					
Primary or no diploma	educ1	0.371	0.214	0.257	0.166
Secondary school	educ2	0.229	0.171	0.171	0.114
High school	educ3	0.286	0.262	0.400	0.349
Vocational school	educ4	0.029	0.052	0.029	0.026
University	educ5	0.086	0.290	0.143	0.329
Gender (male/female)	gender	0.000	0.033	0.029	0.052
Age ^a	Age ^a	41.257	41.024	38.941	40.675
Experience in the present job (years)	exper	9.700	9.187 #	8.869	8.514
Level of trust in institutions (1-10)	trust	4.606	5.189	4.722	5.168
<i>Number of observations</i>		43	391	47	555

Item response rate is lower for some variables.

a Geometric average

Source: Calculated from the World Bank dataset

Table 10. Determinants of informality, 2008
(Switching regression results)

	Manufacturing		Services	
	Coeff	Std dev	Coeff	Std dev
ll	2.237	0.891 **	0.829	0.428 **
ll2	-0.587	0.229 ***	-0.131	0.082
gender	0.174	0.715	-0.132	0.526
age	0.068	0.098	0.042	0.109
age2	-0.001	0.001	-0.001	0.001
lexper	-0.181	0.235	0.417	0.203 **
lfage	-0.127	0.212	-0.488	0.209 **
educ2	-0.388	0.547	-0.376	0.574
educ3	-0.504	0.489	-0.510	0.451
educ4	-1.317	1.126		
educ5	-1.743	0.630 ***	-1.048	0.472 **
trust	-0.204	0.103 **	-0.183	0.097 *
Constant	-3.092	2.331	-1.620	2.375
n	393		514	
Chi-square	25.8 **		15.9	
Pseudo R2	0.109		0.058	

Robust standard errors in brackets

*** p<0.01, ** p<0.05, * p<0.1

Table 11. Average effects of informality, 2008
(Matching propensity score estimation results)

Variable	Informal (Treated)	Formal (Untreated)		Number of observations	
		All	Control	Informal	Formal
<i>Manufacturing</i>					
lp	7.55	7.82	7.53	29	257
klint	9.04	9.66 +	9.81	37	287
lwage	7.01	7.09	6.99	36	290
dlsupp	0.60	0.61	0.50	40	353
dlbuy	0.45	0.49	0.30 *	40	353
technology	0.25	0.37	0.20	40	341
relprofit	4.64	5.64 +	5.77 *	39	345
efficiency	2.75	3.05 +	2.88	40	349
expint	0.02	0.07	0.06 *	29	249
credit	0.35	0.37	0.28	40	348
corporate	0.43	0.75 +	0.75 *	40	353
wagebank	0.03	0.30 +	0.25 *	40	350
accpay	1.15	1.18	1.23	40	353
accrec	1.20	1.17	1.25	40	353
train	0.20	0.25	0.20	40	350
invest	0.23	0.33	0.43 *	40	348
subuse	0.08	0.11	0.08	40	344
subsup	0.18	0.13	0.08	40	342
<i>Services</i>					
lp	7.58	7.69	8.17 *	35	327
klint	8.60	8.91	8.57	40	320
lwage	6.72	6.96 +	6.91	42	356
dlsupp	0.54	0.59	0.59	46	468
dlbuy	0.37	0.37	0.26 *	46	468
technology	0.36	0.32	0.16 *	45	451
relprofit	4.95	5.35	5.23	44	461
efficiency	2.67	3.08 +	2.98	45	453
expint	0.04	0.03	0.05	37	323
credit	0.28	0.33	0.33	46	463
corporate	0.65	0.68	0.72	46	468
wagebank	0.11	0.24 +	0.24 *	46	446
accpay	1.35	1.34	1.46 *	46	468
accrec	1.35	1.47	1.48	46	468
train	0.28	0.31	0.33	46	466
invest	0.26	0.21	0.07 *	46	457
subuse	0.17	0.18	0.26 *	46	456
subsup	0.24	0.18	0.13	46	452

+ means the difference between informal and formal (all) firms is statistically significant at the 10 % level.

* means the difference between informal and formal (control group) firms is statistically significant at the , 10 % level, biased-corrected bootstrap estimate based on 250 replications.

Table 12. Production function estimation results, 2008
(Switching regression results)

	Manufacturing		Services	
lk	0.259*** [0.090]	0.245*** [0.092]	0.157** [0.071]	0.157** [0.074]
lm	0.219*** [0.078]	0.217*** [0.080]	0.155*** [0.048]	0.147*** [0.052]
lll	0.640*** [0.157]	0.663*** [0.168]	0.756*** [0.122]	0.731*** [0.128]
informal	-2.833*** [0.742]	-2.863*** [0.743]	-2.493*** [0.778]	-2.158** [1.174]
educ2		0.927* [0.530]		0.410 [0.449]
educ3		0.476 [0.455]		0.451 [0.464]
educ4		0.816 [0.686]		1.342** [0.661]
educ5		0.518 [0.396]		0.580 [0.421]
gender		0.496 [0.462]		-0.270 [0.409]
technology		0.059 [0.211]		0.285 [0.228]
dlsupp		-0.284 [0.269]		-0.046 [0.220]
dlbuy		0.319 [0.226]		-0.280 [0.272]
Constant	3.677*** [0.902]	3.288*** [1.110]	5.416*** [0.811]	5.067*** [0.899]
rts	1.119	1.126	1.069	1.037
n	228	217	248	242
Log likelihood	-509	-483	-574	-557
chi2	165.6***	196.7***	98.3***	118.0***

Robust standard errors in brackets

*** p<0.01, ** p<0.05, * p<0.1

rts: returns to scale parameter

Table 13. Share of formal employment by age group and employment status in urban areas, 2006

Age group	Regular	Casual	Employer	Self-employed	Unpaid family
<i>Manufacturing, male</i>					
15-19	0.298	0.158	1.000	0.111	0.159
20-24	0.668	0.093	0.717	0.369	0.204
25-29	0.789	0.103	0.688	0.410	0.375
30-34	0.850	0.150	0.772	0.509	0.452
35-39	0.845	0.128	0.880	0.487	0.497
40-44	0.857	0.179	0.871	0.589	0.000
45-49	0.731	0.166	0.699	0.501	0.000
50-54	0.609	0.087	0.686	0.430	0.564
55-59	0.563	0.055	0.653	0.390	0.000
60-64	0.452	0.057	0.395	0.346	0.000
65+	0.445	0.055	0.383	0.445	
<i>Services, male</i>					
15-19	0.274	0.085	0.428	0.042	0.086
20-24	0.637	0.086	0.640	0.339	0.187
25-29	0.804	0.098	0.800	0.470	0.239
30-34	0.862	0.087	0.844	0.535	0.426
35-39	0.885	0.095	0.865	0.560	0.313
40-44	0.886	0.123	0.889	0.598	0.314
45-49	0.867	0.058	0.760	0.530	0.228
50-54	0.751	0.045	0.620	0.417	0.186
55-59	0.641	0.099	0.583	0.341	0.050
60-64	0.594	0.018	0.384	0.356	0.392
65+	0.272	0.000	0.400	0.286	0.000
<i>Manufacturing, female</i>					
15-19	0.278	0.520	0.520	0.000	0.000
20-24	0.629	0.087	0.316	0.016	0.138
25-29	0.762	0.023	0.854	0.119	0.181
30-34	0.750	0.037	0.519	0.045	0.208
35-39	0.752	0.024	0.674	0.023	0.161
40-44	0.683	0.029	0.727	0.084	0.000
45-49	0.742	0.000	0.661	0.152	0.294
50-54	0.616	0.108	1.000	0.223	1.000
55-59	0.581	0.000	0.000	0.000	0.000
60-64	0.818	0.000	0.000		
65+	0.000	0.000	0.428	0.000	0.000
<i>Services, female</i>					
15-19	0.406	0.164	0.000	0.000	
20-24	0.751	0.177	1.000	0.346	0.208
25-29	0.872	0.048	0.796	0.518	0.217
30-34	0.885	0.070	0.810	0.489	0.233
35-39	0.860	0.035	0.876	0.408	0.210
40-44	0.821	0.071	0.902	0.349	0.170
45-49	0.731	0.082	0.641	0.341	0.095
50-54	0.702	0.000	0.657	0.300	0.129
55-59	0.562	0.051	0.437	0.364	0.187
60-64	0.652	0.000	0.458	0.000	0.000
65+	0.520	0.000	0.215	0.217	0.169

Source: HLFS, 2006

Table 14. Share of formal employment by firm size and employment status in urban areas, 2006

Firm size	Regular	Casual	Employer	Self-employed	Unpaid family
<i>Manufacturing, male</i>					
<10	0.452	0.103	0.728	0.476	0.244
10-24	0.612	0.186	0.812	1.000	0.478
25-49	0.767	0.234	0.845	0.370	
50+	0.919	0.611	0.892	1.000	
<i>Services, male</i>					
<10	0.539	0.058	0.776	0.491	0.174
10-24	0.830	0.138	0.854	0.943	0.493
25-49	0.901	0.229	0.765	0.866	
50+	0.966	0.631	0.798	0.488	
<i>Manufacturing, female</i>					
<10	0.285	0.008	0.622	0.073	0.131
10-24	0.361	0.051	0.803	1.000	0.475
25-49	0.636	0.000	0.660	0.602	
50+	0.844	0.589	0.865		
<i>Services, female</i>					
<10	0.481	0.023	0.768	0.402	0.164
10-24	0.861	0.314	0.863	1.000	0.512
25-49	0.914	0.289	0.937	0.148	
50+	0.969	0.442	0.687	0.000	

Source: HLFS, 2006

Table 15. Formal and informal average wages in manufacturing and services, 2006 (mean log wages)

Firm size	Manufacturing			Services		
	Informal	Formal	Wage diff (%)	Informal	Formal	Wage diff (%)
<i>Male workers</i>						
<10	5.97	6.27	0.30	5.87	6.33	0.46
10-24	6.12	6.37	0.25	6.07	6.55	0.48
25-49	6.16	6.38	0.22	6.14	6.63	0.50
50+	6.24	6.50	0.26	6.29	6.76	0.47
<i>Female workers</i>						
<10	5.31	6.21	0.91	5.62	6.26	0.64
10-24	5.96	6.28	0.32	5.91	6.52	0.61
25-49	5.95	6.25	0.30	5.94	6.60	0.66
50+	5.93	6.30	0.37	5.98	6.75	0.78

Source: HLFS, 2006

Table 16. Determinants of employment decision, urban regions, 2006
(Multinomial regression model, marginal effects)

	Unemp	Informal worker		Formal worker		Employer	Self-employed	Mean value
		Manuf	Services	Manuf	Services			
<i>Male</i>								
Child	0.013	0.000	0.020	-0.002	0.032	-0.021	-0.041	0.315
Age groups								
20-24	-0.305	-0.021	-0.033	0.060	0.070	0.152	0.076	0.101
25-29	-0.400	-0.034	-0.054	0.048	0.081	0.230	0.128	0.118
30-34	-0.410	-0.044	-0.066	0.029	0.072	0.305	0.114	0.110
35-39	-0.389	-0.048	-0.074	0.005	0.075	0.320	0.113	0.099
40-44	-0.379	-0.051	-0.076	-0.010	0.080	0.320	0.115	0.103
45-49	-0.339	-0.049	-0.072	-0.051	0.042	0.331	0.137	0.085
50-54	-0.264	-0.048	-0.064	-0.079	-0.042	0.322	0.175	0.078
55-59	-0.201	-0.051	-0.065	-0.093	-0.101	0.320	0.191	0.056
60-64	-0.088	-0.053	-0.069	-0.103	-0.127	0.280	0.161	0.038
65+	0.049	-0.063	-0.080	-0.119	-0.173	0.249	0.137	0.079
Educational level								
Primary	0.307	-0.005	-0.024	-0.081	-0.137	-0.036	-0.024	0.034
Secondary	0.191	0.017	-0.010	-0.071	-0.097	-0.028	-0.003	0.048
High school	-0.039	-0.048	-0.030	-0.033	0.140	0.021	-0.012	0.137
Vocational	-0.149	-0.035	-0.026	0.061	0.164	0.008	-0.024	0.109
University	-0.259	-0.059	-0.051	-0.014	0.403	0.026	-0.045	0.116
Single	0.103	0.013	0.018	-0.034	-0.049	-0.020	-0.031	0.286
Divorced	0.088	0.012	0.017	-0.054	-0.072	-0.012	0.022	0.021
Parent*family size	-0.067	0.030	0.020	-0.022	-0.010	0.001	0.047	0.878
Child*family size	-0.063	0.030	0.018	-0.016	-0.044	0.009	0.067	0.489
Any registered	0.689	-0.091	-0.120	-0.102	-0.151	-0.048	-0.177	0.546
Unemployed HH	0.091	-0.012	-0.013	-0.019	-0.031	-0.008	-0.008	0.163
<i>Female</i>								
Child	0.014	-0.005	-0.003	0.000	-0.004	0.000	-0.002	0.268
Age groups								
20-24	-0.041	0.001	0.004	0.001	0.010	0.001	0.023	0.119
25-29	-0.086	0.001	0.004	0.002	0.024	0.007	0.048	0.119
30-34	-0.124	0.001	0.005	0.002	0.031	0.017	0.067	0.112
35-39	-0.152	0.001	0.008	0.002	0.039	0.025	0.077	0.096
40-44	-0.114	-0.001	0.006	0.001	0.032	0.019	0.056	0.100
45-49	-0.068	-0.002	0.003	0.000	0.009	0.017	0.042	0.080
50-54	-0.036	-0.003	-0.002	0.000	0.002	0.011	0.028	0.072
55-59	-0.021	-0.004	-0.005	0.000	-0.004	0.010	0.025	0.052
60-64	0.004	-0.004	-0.007	0.000	-0.005	0.005	0.008	0.037
65+	0.020	-0.005	-0.011	-0.007	-0.010	0.005	0.008	0.092
Educational level								
Primary	0.022	-0.003	-0.006	-0.001	-0.009	-0.001	-0.003	0.170
Secondary	0.012	0.000	-0.004	0.000	-0.005	-0.001	-0.001	0.075
High school	-0.037	-0.002	0.002	0.000	0.035	0.001	0.000	0.103
Vocational	-0.078	-0.001	0.006	0.001	0.071	0.001	0.001	0.064
University	-0.274	-0.003	0.005	0.001	0.259	0.005	0.006	0.071
Single	-0.019	0.005	0.008	0.000	0.004	0.000	0.002	0.228
Divorced	-0.007	0.001	0.005	0.000	0.000	0.000	0.001	0.111
Parent*family size	0.009	-0.002	-0.002	0.000	-0.004	0.000	-0.001	0.908
Child*family size	-0.004	0.003	0.002	0.000	0.001	0.000	-0.001	0.426
Any registered	0.595	-0.106	-0.238	-0.008	-0.118	-0.009	-0.116	0.934
Unemployed HH	0.010	-0.001	-0.003	0.000	-0.004	0.000	-0.002	0.339

Table 17. Determinants of wages, 2006
(Multinomial participation decision corrected wage equation estimates)

	Male employees								Female employees							
	Manufacturing				Services				Manufacturing				Services			
	Informal		Formal		Informal		Formal		Informal		Formal		Informal		Formal	
	Coeff	Std dev	Coeff	Std dev	Coeff	Std dev	Coeff	Std dev	Coeff	Std dev	Coeff	Std dev	Coeff	Std dev	Coeff	Std dev
<i>Age groups</i>																
20-24	0.337	0.033 ***	0.324	0.037 ***	0.270	0.036 ***	0.161	0.032 ***	0.147	0.055 ***	0.215	0.041 ***	0.184	0.046 ***	0.268	
25-29	0.464	0.047 ***	0.435	0.044 ***	0.359	0.054 ***	0.204	0.038 ***	0.132	0.075 *	0.316	0.047 ***	0.293	0.058 ***	0.476	
30-34	0.445	0.054 ***	0.493	0.044 ***	0.373	0.064 ***	0.248	0.039 ***	0.001	0.076	0.352	0.047 ***	0.310	0.063 ***	0.546	
35-39	0.516	0.058 ***	0.536	0.046 ***	0.430	0.075 ***	0.228	0.043 ***	-0.021	0.085	0.341	0.047 ***	0.445	0.063 ***	0.562	
40-44	0.566	0.065 ***	0.621	0.047 ***	0.483	0.076 ***	0.261	0.044 ***	0.105	0.093	0.368	0.046 ***	0.420	0.063 ***	0.562	
45-49	0.686	0.071 ***	0.694	0.048 ***	0.542	0.076 ***	0.301	0.045 ***	0.089	0.125	0.317	0.055 ***	0.470	0.064 ***	0.611	
50-54	0.659	0.080 ***	0.787	0.049 ***	0.531	0.061 ***	0.515	0.039 ***	0.110	0.125	0.391	0.071 ***	0.418	0.073 ***	0.666	
55-59	0.513	0.096 ***	0.762	0.061 ***	0.459	0.060 ***	0.698	0.042 ***	-0.324	0.173 *	0.341	0.143 **	0.451	0.097 ***	0.727	
60-64	0.419	0.116 ***	0.846	0.098 ***	0.478	0.071 ***	0.814	0.056 ***	-0.550	0.310 *	0.623	0.211 ***	0.371	0.149 **	0.547	
65+	0.298	0.157 *	0.641	0.123 ***	0.291	0.083 ***	1.111	0.107 ***	0.433	0.373		***	0.708	0.310 **	0.694	
<i>Firm size</i>																
10-24	0.162	0.017 ***	0.080	0.015 ***	0.167	0.022 ***	0.111	0.010 ***	0.438	0.040 ***	0.115	0.041 ***	0.173	0.041 ***	0.118	
25-49	0.206	0.020 ***	0.073	0.013 ***	0.191	0.024 ***	0.132	0.009 ***	0.408	0.044 ***	0.078	0.036 **	0.227	0.043 ***	0.129	
50+	0.232	0.021 ***	0.132	0.011 ***	0.331	0.029 ***	0.248	0.007 ***	0.385	0.043 ***	0.127	0.032 ***	0.328	0.049 ***	0.231	
<i>Education</i>																
Primary school	0.069	0.043	0.032	0.065	0.106	0.053 **	-0.098	0.069	-0.061	0.092	0.144	0.092	0.032	0.073	-0.043	
Secondary school	-0.031	0.047	0.120	0.060 **	0.086	0.053	-0.117	0.066 *	-0.140	0.074 *	0.223	0.081 ***	0.079	0.056	0.137	
High school	0.157	0.104	0.353	0.075 ***	0.298	0.068 ***	-0.069	0.080	-0.078	0.137	0.457	0.095 ***	0.406	0.094 ***	0.486	
Vocational school	0.040	0.070	0.417	0.067 ***	0.199	0.066 ***	-0.008	0.074	-0.130	0.158	0.518	0.104 ***	0.501	0.103 ***	0.599	
University	0.515	0.146 ***	0.687	0.084 ***	0.528	0.095 ***	0.020	0.095	-0.121	0.256	0.952	0.133 ***	0.865	0.140 ***	0.804	
Working time	0.163	0.028 ***	-0.121	0.021 ***	0.298	0.024 ***	-0.101	0.012 ***	0.165	0.067 **	-0.030	0.048	0.445	0.030 ***	-0.095	
Full-time dummy	0.320	0.036 ***	0.616	0.041 ***	0.190	0.036 ***	0.191	0.016 ***	0.874	0.088 ***	0.666	0.070 ***	0.013	0.046	0.186	
<i>Selection variables</i>																
m0, unemployed	0.172	0.111	0.018	0.083	-0.411	0.143 ***	0.058	0.040	0.535	0.317 *	-0.107	0.119	0.460	0.210 **	0.082	
m1, informal, manuf	0.005	0.058	-1.799	0.215 ***	-0.830	0.230 ***	-1.567	0.145 ***	0.016	0.048	-0.911	0.209 ***	-0.165	0.241	-1.443	
m2, informal, services	0.017	0.286	0.925	0.212 ***	0.078	0.032 **	1.534	0.148 ***	0.612	0.454	0.110	0.255	0.082	0.030 ***	0.618	
m3, formal, manuf	-0.752	0.233 ***	0.053	0.029 *	-0.868	0.197 ***	0.631	0.127 ***	-0.305	0.383	0.067	0.025 ***	-0.111	0.265	1.097	
m4, formal, services	1.126	0.255 ***	-0.011	0.205	0.128	0.264	-0.236	0.027 ***	0.336	0.419	0.155	0.216	0.928	0.307 ***	-0.015	
m5, employer	-1.868	0.289 ***	-2.177	0.252 ***	-1.447	0.241 ***	-0.855	0.175 ***	-4.214	0.998 ***	-4.049	0.472 ***	-1.828	0.546 ***	-1.560	
m6, self-employed	0.954	0.198 ***	1.439	0.123 ***	0.435	0.175 **	1.122	0.086 ***	0.953	0.282 ***	0.838	0.161 ***	0.041	0.214	0.214	
Constant	4.780	0.262 ***	5.196	0.218 ***	3.124	0.259 ***	7.139	0.183 ***	4.492	0.454 ***	4.495	0.289 ***	3.595	0.267 ***	5.703	
n	5606		11968		6409		20407		1230		2161		2271		7321	
F-stat	65.6 ***		222.0 ***		92.3 ***		560.2 ***		35.9 ***		61.0 ***		39.5 ***		224.2	
R-squared	0.241		0.334		0.281		0.426		0.446		0.426		0.322		0.454	

***, **, and * mean statistically significant at the 1%, 5% and 10%, respectively.

Robust standard errors.

Table 18. Expected average (log) output
(conditional on informality status)

	Current status	
	Informal	Formal
<i>Manufacturing</i>		
Informal operation	12.62	13.14
Formal operation	12.67	14.15
Output differential	0.05	1.01
<i>Services</i>		
Informal operation	13.30	13.16
Formal operation	13.55	13.96
Output differential	0.25	0.80

Source: Calculated from the ERF dataset on small firms

Table 19. Effects of enforcing formality on urban employment and wages
(Simulations based on the 2006 data)

	Current		Total	All formal		Wages change (%)
	Informal	Formal		Employees		
				number	change (%)	
Mining	20	108	128	117	-0.089	0.09
Manufacturing	1341	2845	4186	3425	-0.182	0.13
EGW	3	90	93	91	-0.017	0.02
Construction	788	479	1267	1115	-0.120	0.28
Trade, H&R	2037	2693	4730	4336	-0.083	0.18
Transportation	464	699	1163	1073	-0.077	0.20
Finance, real estate	198	813	1011	972	-0.038	0.11
Other services	658	3007	3665	3537	-0.035	0.16
<i>Total (non-agriculture)</i>	<i>5508</i>	<i>10734</i>	<i>16242</i>	<i>14666</i>	<i>-0.097</i>	<i>0.18</i>

Table 20. Effects of enforcing formality on the composition of urban employment
(Simulations based on the 2006 data, no growth)

	Current	All formal	Change (%)
<i>Education level</i>			
Illiterate	227	56	-0.752
Primary	405	137	-0.661
Secondary	8606	6798	-0.210
High school	2231	2275	0.020
Vocational	2054	2202	0.072
University	2719	3198	0.176
<i>Age</i>			
15-24	2826	1789	-0.367
25-34	6100	5997	-0.017
35-44	4489	4569	0.018
45-54	2211	1951	-0.117
55-64	506	312	-0.384
65+	109	48	-0.558
<i>Gender</i>			
Male	13248	12163	-0.082
Female	2994	2504	-0.164
<i>Total</i>	<i>16242</i>	<i>14666</i>	<i>-0.097</i>

Table 21. Effects of enforcing formality on the composition of urban employment
(Simulations based on the 2006 data, 10 % growth)

	Current	All formal	Change (%)
<i>Education level</i>			
Illiterate	227	62	-0.727
Primary	405	151	-0.628
Secondary	8606	7478	-0.131
High school	2231	2503	0.122
Vocational	2054	2422	0.179
University	2719	3517	0.294
<i>Age</i>			
15-24	2826	1968	-0.304
25-34	6100	6597	0.081
35-44	4489	5026	0.119
45-54	2211	2146	-0.029
55-64	506	343	-0.323
65+	109	53	-0.514
<i>Gender</i>			
Male	13248	13379	0.010
Female	2994	2754	-0.080
<i>Total</i>	<i>16242</i>	<i>16133</i>	<i>-0.007</i>