Axel Demenet, Mireille Razafindrakoto, François Roubaud   DIAL

Do informal businesses gain from registration?

Panel data evidence from Vietnam

DRAFT

May, 2013

Executive summary

This paper evaluates the impact of formalization on Household Businesses performance and mode of operation, a topic that has attracted recent attention in the literature. It capitalizes on a unique panel data, result of a five years research project, which allows (i) being representative of the urban informal sector, and (ii) measuring their performance and intermediate outcomes with great precision, reconstructing their often missing accounts. Focusing on already existing informal businesses that left the informal sector, we quantify a plausibly causal effect on performance, and identify the channels through which it occurs.

We find that leaving the informal sector has a significant positive impact on firms' performance, increasing annual value added by 22% on average. More importantly, we show that this improvement is made possible by associated changes in operating conditions: released from the constraints of informality, household businesses have better equipment, can increase their scale of operation, and operate in a more competitive environment. We confirm this result by looking at the dual problem: *informalization* (businesses that gave up registration but continued operating) has a symmetrical negative impact on performance and conditions.

These findings raise important questions concerning the types of businesses likely to formalize and the potential incentive policies targeted towards the informal sector.

Key words : Informal Sector, Household Business, Formalization, Matching

Corresponding author: axel.demenet@ensae.fr

1 Introduction

Should we hope that each of the Informal Production Units (IPU) that constitute a predominant share of developing economies will formalize in the medium run? The question is rarely asked directly. Considering the long-standing negative connotation of informality, the answer given by the literature -at least implicitly- makes no doubt. However, the extent to which micro-firms themselves would benefit from formalization remains unclear.

The question is yet a first-plan research topic. Firstly, it is closely related to the micro-determinants of informality: a large segment of literature defends the view of chosen informality, which implies that the overall size of the informal sector would depend on the perceived costs and benefits of each legal status. Furthermore, estimating the causal impact of registration is a necessary condition to the implementation of efficient policies towards informality. It is of particular interest in the case of Vietnam since encouraging formalization is one of the national priorities for the country's employment policy pointed out by the Ministry of Social Affairs. The country is symptomatic of such heterogeneous and predominant informal sector. Despite the rapid growth that occurred since the 1986 liberalization (Doi moi), and despite the new status of middle-income country according to the World Bank's classification, the informal sector is still the leading job provider in Hanoi and Ho Chi Minh City, accounting for around 50% of non-agricultural jobs.

If costs associated with formality have been extensively described (De Soto (1989) [19]; Djankov et al. (2002)[12]), measuring the gains of registration is not straightforward and raises two major problems. Firstly, given the very nature of the informal sector, quantitative data is everything but easy to produce (ILO (1993) [1], Razafindrakoto *et al.* (2009) [32]). The original sin of IPU -being unregistered- keeps them inherently away from statistical systems. They often operate without

fixed premises, outdoor or at home, which makes classical enterprises surveys (often census-type) inefficient in capturing for this phenomenon.

Secondly, even in the few countries where survey data was made available, evaluating a causal impact requires to overcome the fundamental problems of selection bias and endogeneity. On the one hand, selection bias arises from the fact that businesses that choose to formalize are not comparable with the ones that remained informal. Employing the impact evaluation terminology of Rubin (1974 [45], 1990 [46]), we can denote F the binary treatment variable (that takes the value $F = 1$ if the production unit formalized) and Y the outcome (that takes values Y^f and Y^{inf} for the "formalized" and "remained informal" outcomes). The causal impact of formalization can be written $\Delta = Y^f - Y^{inf}$. Since we do not observe the counterfactuals $\mathbb{E}(Y^f|F = 0)$ and $\mathbb{E}(Y^{inf}|F = 1)$, identification requires $(Y^{inf}, Y^f) \perp F$. This hypothesis is highly questionable as the potential outcome of the formalized Household Businesses (HB) would probably have been different from the non-formalized one, whatever their trajectory. Status *vis à vis* registration might be correlated with unobserved characteristics that also affect outcomes —for instance if more able HB heads are more likely to formalize. On the other hand, registration might be partly determined by outcomes, generating a simultaneity bias. This can be true if for instance higher profits lead to more visibility and therefore a higher probability to register. These forms of endogeneity have been largely documented (Maloney (2004) [33], De Paula and Scheinkman (2007) [47]), and need to be accounted for. Naive estimation of the impact of registration is likely to overestimate the gains.

This paper capitalizes on the panel data produced during the 5 years IRD/GSO[1] research project to approach the problem. A large scale survey has been conducted in 2007 and 2009 by the GSO in the two major cities, Hanoi and Ho Chi Minh City (HCMC). It is based on the mixed-survey methodology, whose principle is to identify the IPU heads in a first-phase Labour Force Survey, and to build a sampling frame of informal production units that are surveyed in a second phase. This methodology allows to capture outdoor unregistered businesses and to be representative of the informal sector (Hussmanns, 2008[42]; Roubaud and Séruzier (1991) [15]; Roubaud (2008) [14]). The data includes a total of 1,465 initially informal Household Businesses (HB), among which 147 did formalize between the two years, enabling us to identify the impact of registration on a rich set of intermediate and final outcome variables, controlling for a large number of observables.

The aim of the paper is threefold: (1) determine what types of businesses do change their legal status and choose registration, (2) measure the impact of formalization on performance for the existing businesses that chose to leave the informal sector, and (3) identify the channels through which this effect occurs by evaluating the impact of registration on their conditions of operation (that is, determine whether operating formally allows better access to equipment, operating on a larger scale, and the type of business environment).

In order to overcome selection bias and endogeneity of registration —and thus estimate a plausibly causal effect— several methods are applied. We make use of the panel nature of our data to produce Difference-in-Difference (DiD) estimates, both in a OLS and Fixed-Effect (FE) setup, to control for potential unobserved time-invariant characteristics. Furthermore, we use DiD Matching estimators to control for selection bias. None of the method is individually satisfactory, since they only eliminate partly endogeneity bias. Combining them allows measuring a plausibly causal impact, although some of selection on unobservables remains.

The major result are as follows. (1) Firms that registered belonged to the higher-end of informal household businesses, since their characteristics were close to already formal household businesses.

---

[1] French Institute of Research on Development and Vietnamese General Statistics Office. Fore more details on the project, see Cling *et al.*, 2010[23]

(2) There is a large and significant effect of formalization on a set of measures of conditions of operation. Formalizing is found to be associated with an improved access to electricity and Internet, to allow for strongly increasing size, improve premises and widen the use of formal accounts. Furthermore, firms that decided to register seem to operate in a more competitive environment, reporting more problems with customers and competitors. (3) This, in turn, improves significantly the economic performance of businesses. Those who chose to register increased their annual value added by 20% (lower bound of estimations). Supporting the results of McKenzie and Sakho (2010 [35]), we find that originally self-employed workers did not benefit from registration as much as bigger businesses, suggesting the existence of a threshold below which the choice is not relevant.

This paper is organized as follows: section 2 reviews literature on the topic, emphasizing the existing studies on formalization benefits. Section 3 presents the data, and a descriptive analysis of the formalized HBs' characteristics. Section 4 details the identification strategy and the estimation results. Section 5 provides further robustness checks.

2 Literature: what are the expected effects of formalization?

The central question of the effect of (in)formality on firms outcomes highly depends on the nature of informal activities and the motivations of individuals undermining their decision to operate in the informal sector. In the vision of the *Legalist School* (De Soto (1989) [19]), informality is a rational decision of workers willing to escape the bureaucratic costs of formality, and their outcomes would be lower if they had to comply with excessive regulation. Thus, informality would have a positive impact on outcomes. The opposite conclusions are given by the *Structuralist School* (Moser, 1978 [8], Portes et al., 1989 [40]), for whom informality stems from the strategy of cost-optimizing enterprises willing to outsource their production to unprotected local workers. In this case, informality may be an advantage in terms of flexibility for local micro-firms subordinated to internationalized enterprises, but is above all synonymous to unregulated, precarious and insecure work, and deprives workers from legal protection. For the *Dualist School* (Hart, 1973 [27]) that sees informal businesses as subsistence activities, the legal status can hardly be linked with the outcomes as small self-employed businesses are at stake that would be of low productivity in both cases.

It is now widely acknowledged that the three conceptions of informality can be simultaneously true given the heterogeneity of existing businesses. Progress has been made in recognizing the segmentation of the informal sector itself (OECD, 2009 [26]) where the higher, intermediate and lower segments represent respectively the Legalist, Structuralist and Dualist visions[2]. The impact of (in)formality on outcomes is thus unclear and largely depends of the segment considered.

The literature handled the problem in three manners. The first strand of literature aims at identifying the correlates of informality at the firm level. Although informative, this approach does not allow isolating the effect of informality on outcomes: given the selection issue, a direct comparison is required between (otherwise similar) formal and informal businesses. The second strand of literature, to which contributed several recent papers, compares firms that *are currently* formal and firms that *are currently* informal. An additional question handled in the third strand, to which this paper contributes, is the effect of formalization for already existing informal firms. A dynamic vision comparing firms that remained informal and firms that formalized is closer to the actual question for policy makers raised by the existence of a predominant informal sector.

The correlates of informality at the firm's level

Enlightening the characteristics associated with informality at the Production Units (PU) level is one of the more documented strands of literature about developing countries. Informality has been shown to imply a number of correlates that are associated with inferior production conditions and subsequently with reduced performance.

First, Informal Production Units (IPU) are generally small. Not only are they made (at least partly) of self-employed workers and subsistence businesses, but also their expansion can be inhibited by the fear of attracting the attention of the authorities (World Bank, 2007). Even if their purpose is not illegal and their official registration not always compulsory, they operate in a fuzzy legal framework that neither informal workers nor the police really knows, and often prefer remaining unnoticed. In Vietnam as in several other countries, registration is only compulsory above a certain threshold of size and/or activity that only a tiny minority of workers know. Most of them

---

[2]This is particularly true in the case of Vietnam where the multi-segmentation of the informal sector has been showed by Cling*et al.*, 2010 [23]. Diametrically opposite types of household business coexist, from the most precarious and insecure businesses operating outdoors (40% of them, supporting the Dualist conception) to the high-end professionals (10% of the total).

believe that they are illegal, whether they actually are or not, which may prevent small IPU from expanding when they have the opportunity to do so.

Second, one of the most cited point is that IPU allegedly suffer from low productivity, most noticeably in the view of Levy (2007 [29]) and after the work of La Porta and Schleifer (2008 [39]). Relying on basic correlations and a comparative analysis of "official" and "unofficial" firms in several developing countries, the latter describe Informal Production Units (IPU) as "extremely unproductive". This view must be taken with care given the weaknesses of their data: 85% of their "unofficial firms" have 2 workers or more, which is far from representative. Benjamin and Mbaye (2011 [7]) findings follow the same direction, documenting a labour productivity gap in West Africa.

By contrast, rigorous measurements of capital returns consistently find them to be very high among IPU, although highly dependent from the initial stock. Grimm, Krüger and Lay (2011 [31]) measured them for West Africa using 1-2-3 databases representative of the urban informal sector and found very high marginal returns at low levels of capital, consistently over 70 per cent, but rapidly decreasing with the level of capital down to 4-7 per cent. Other findings are consistent with this result: Göbel, Grimm and Lay (2012 [18]) repeated the exercise in Peruvian micro-firms using panel data; De Mel, Mc Kenzie and Woodruff's (2008 [10]) experiment of randomized grants in Sri Lanka showed returns to be on average of 55%–63% per year; Mc Kenzie and Woodruff (2010 [35]) also came out with 20–33 percent a month from a similar random allocation of grants in Mexico. The overall productivity gap between formal and informal firms is subject to caution: if labour productivity seems to actually suffer from it, capital returns are high in the small IPU, justifying the recent interest for microcredit.

Third, All results of the above-cited studies corroborate the fact that informal firms face significant constraints in access to credit. This explains the high level of returns to capital: IPU are poorly endowed entrepreneurs who operate in imperfect capital markets and a very risky environment. Because they have no legal existence and generally no collateral to offer, IPU are deprived from access to formal loans, which is only partly balanced by the existence of informal channels (see e.g. Udry, 1993 [50]).

Fourth, an explanation of the labour productivity gap often put forward is differences in production means. IPU suffer from unequal access to public services. Often lacking fix premises, they logically have less access to running water, electricity or telephone. Moreover, legal protection and contract enforcement can also be seen as a public good to which the absence of legal status prevents from accessing. Apart from differentiated access to public goods and services, IPU generally suffer from challenging access to intrants (Levy (2007) [29]). Inability to enter formal contractual relationships can be associated to difficulties in establishing long-term quality relations with suppliers. Consequently, the quality of goods produced by IPU is inferior (Banerji and Jain, 2007 [5]).

Fifth, formality of the production unit has an impact on the type of manpower and the labour cost. Logically enough, formal PU have a higher unit labour cost than informal ones, as they have to provide workers with social benefits and/or contribute to social protection schemes, and are unable to pay them below minimum wage. They may also have less flexibility to fire workers in case of negative shocks (Heckman and Pages, 2004 [20]). Almeida and Susanlı (2012) argue that informality allows firms greater flexibility in their employment decisions which, in turn, allows them to operate more efficiently. On the other side, formal contracts may act as risk pooling mechanisms that may attract a better quality manpower (Fajnzylber, Maloney and Montes-Rojas, 2009 [16]. From the workers' point of view, formality is also associated with a higher income. The wage gap has been measured with consistent results in South America (Maurizio, 2012 [34]) and Vietnam

(Rand and Torm, 2012b [44] and Nguyen, Nordman and Roubaud, 2013 [37]).

As useful as these findings are, they say little about the pure impact of the legal status on the firms outcomes and conditions of operation. Such identification requires to overcome the endogeneity of legal status. Indeed, IPU's underlying characteristics (such as head's ability, family background, wealth and social network) determine both outcomes and legal status: they make firms more productive and at the same time more likely to need formality. As in Lenvenson and Maloney (1998 [28]), IPU's performance can make them more likely to be detected by authorities and hence formalize. Ignoring this selection issue leads to biased results: it is obvious that the observed difference between both status will incorporate differences in these underlying characteristics that determine at the same time the outcomes and the legal status, leading to overestimate the pure impact of formalization from the firm's point of view.

Identifying the effect of legal status on IPU's outcomes

Acknowledging the endogeneity of legal status, several recent studies isolated its effect on performance. Fajnzylber, Maloney and Montes-Rojas (2011 [17]), examining the impact of choosing formality on revenue, employment and capital stock in Brazil, find a positive effect on the three dimensions. Newly created firms that opt for operating formally show higher levels of revenue and profits, employ more workers and are more capital intensive.

In a previous paper, Fajnzylber, Maloney and Monte Rojas (2009 [16]) showed in the context of Mexico that micro-firms that participating in credit markets, receive training, pay taxes and belong to business associations exhibit significantly higher profits (20 per cent increase). Moreover, tax registration together with access to credit increases the likelihood of firm survival, that is the probability to stay in business.

Although overall positive, the effect of registration on profits can be highly heterogeneous, depending on the type of businesses considered. McKenzie and Sakho (2010 [35]) measure the impact of formality (in the sense of tax registration) on firm profitability among Bolivian micro-firms, using the distance of a firm from the tax office as an instrument. They find it to be significant for the medium-sized firms of their sample, but surprisingly enough to be negative for small and large businesses.

The channels through which the alleged benefits occur remain subject to caution. Fajnzylber, Maloney and Montes-Rojas (2011, *op. cit.*) argue that it is not thanks to credit access or contracts with large firms, but through a shift in production means, becoming more capital intensive. In particular, the choice of a permanent location is supposed to facilitate capital and employment extension, and thus operate on a larger scale.

These finding are valid in the hypothesis of production units choosing between formality and informality, but are not looking at the impact of formalization on existing informal firms. This issue is yet of prior importance in the point of view of the policy maker confronted with an existing and predominant informal sector. The results may differ for already existing informal firms that choose to *formalize*, and they may occur through different channels. Much of the policy recommendations as regards informality are dealing with Household Businesses formalization (OECD 2009 [26], WTO 2009 [52], World Bank, 2008 [2]), and yet very few studies exist to measure the real benefit from the firms' point of view.

Dynamic comparison: what is the impact of formalisation of existing businesses?

The data requirements to directly measure the effect of already existing informal firms' registration are much higher and make studies scarce. The more frequently handled related question is whether all types of businesses are potentially concerned, and when does the decision take place. Fajnzylber, Maloney and Monte Rojas (2011, *op. cit.*) argue that formalization is not relevant for all types of businesses: the intrinsic characteristics of many IPU make them unlikely to ever grow large enough to need institutions and formalization. De Mel, McKenzie and Woodruff (2008 [10]) also question the potential of small IPU for income growth. Indeed, many of the self-employed subsistence businesses addressing local clients may not have the ambition, neither the possibility, to enlarge their scale of operation and enter the formal sector.

At least for a segment, the question is however relevant, but requires following the same units over time, and include both formalized and still informal units. Such panel data has been collected only twice, to the best of our knowledge, and both times in Vietnam: the large scale Household Business & Informal Sector Survey used in this paper on the one hand, and the Small and Medium Enterprise Survey by CIEM on the other hand. The latter has been exploited by Rand and Torm (2012a [43]) in a seminal paper that enlightens the effects of formalization for firms that chose to leave the informal sector. They documented a significant effect of registration on profits and investment, and additionally a decrease in the use of casual labour. However, their data suffers from several limitations. First, it is not representative of the Vietnamese informal sector: all IPU were included by on-site identification, and thus operate alongside formal enterprises. The average size of IPU is around 4 full time employees, which is far from the actual figure of 1.5 (Cling *et al.*, 2010 [23]). The possibility remains that their results are driven by a small segment of the informal sector, namely the biggest firms that correspond to the view of the Legalist School and deliberately chose to hide their output. Following McKenzie and Sakho (2010 *op.cit.*), one can argue that firms' owners of this segment are of higher ability and thus benefit more from formality (investment in particular). Second, the level of details does not allow investigating through which channels does the overall positive effect occur.

The present paper builds on an original and unparalleled representative panel data to contribute to the last strand of literature, and tentatively identify the channels that lead the positive effect of legal status on performance.

3   Data and descriptive statistics

The authors of this paper have undertaken, in response to a technical assistance request from the Vietnamese General Statistics Office (GSO), a five years research project that aimed at measuring the size and characteristics of the informal sector. This paper capitalizes on the panel data produced during this project, of which some methodological strong points should be stressed. It is based on micro data drawn from a mixed-survey that allows being representative of the urban informal sector, and applying definitions in line with international recommendations. The questionnaire allows measuring the outcomes of informal production units (IPU) with great precision, reconstructing their (often missing) accounts.

We adopt an operational definition of the informal sector in line with international recommendations[3] (ILO, 1993 [1]; OECD, 2002 [38]; SNA, 2008 [51]); that is, the whole set of unincorporated household businesses that are unregistered as regards business license. Other types of registration criteria are sometimes used (social security, tax code); since our data allows to apply all of them, we chose to conduct the core analysis with business registration in order to stay in line with the literature on informality in Vietnam, and to use alternative definitions for robustness checks.

3.1   Capturing Informality in survey data: HB&IS sample design

The original sin of informal production units -being unregistered- keeps them inherently away from statistical systems. They often operate without fixed premises, outdoor or at home, which prevents from constructing classical sampling frames of informal production units. This makes classical enterprises surveys (often census-type) inefficient in capturing this phenomenon (see ILO (1993) [1] and Razafindrakoto *et al.* (2009) [32]). The survey data used in this paper allows to overcome the absence of sampling frame by applying a mixed household/enterprise methodology. In a first phase, informal businesses heads are identified through a set of questions in the Labour Force Survey.[4] In a second phase, the Household Business and Informal Sector Survey (HB&IS) is conducted on a sample of those production units.

Information drawn from this kind of surveys is more comprehensive and representative of informality than when surveying separately households and production units (Ramilison, 2007 [13]; Hussmanns, 2008[42]). The identification of informal businesses is comprehensive since every household member declaring a self-employment and/or employer activity in the LFS that matched informality criteria will be included in the sampling frame. For more details about mixed-methodology and *1-2-3 surveys*, see for instance Roubaud and Séruzier (1991) [15] and ILO (2012) [21].

In keeping with this methodology, HB&IS survey has been conducted in Hanoi and Ho Chi Minh City (HCMC) in 2007 and 2009, including both formal and informal units to allow for comparisons. Descriptive results and detailed information about the survey quality have been edited in a book (Cling et al., 2010[23]). These two rounds of surveys constitute the first reliable estimates of informal employment and informal sector in Vietnam, and to our knowledge the first panel data representative of the urban informal sector worldwide.

---

[3]Confusion still prevails on the definition of informality. Several close notions are sometimes employed (moonlighted economy, underground, unregistered, shadow, etc.), and symmetrically the term "informality" is sometimes misused to refer to illegal activities (which are a tiny minority of unregistered household businesses).

[4]In 2009, of the 4,500 households surveyed by the LFS in Hanoi and HCMC, around 2,300 individuals are heads of a HB in their main or secondary job

The initial sample contains 2,606 observations in 2007 and 2,012 in 2012; overall, 648 units of the initial sample were not matched between the two years. The attrition process is all the more important to take into account that informal businesses are allegedly less lasting in business than formal ones (Fajnzylber, Maloney and Monte Rojas, 2009 [16]). The intrinsic precariousness of informal sector activities associated with the context of crisis could have led to a high attrition; if this process is non-random, the results obtained by applying panel methods to our data would be biased. Mortality rates (HB that stopped activity) are very high in both cities (14% and 19% in Hanoi and HCMC respectively), but thanks to the scrutiny of interviewers, the attrition rate is quite low (9.1%). [5] During the second survey, all the HBs that changed location within the cities or disappeared were tracked through family members, neighbours, etc. In order to check the randomness of the attrition process combined with mortality, we conducted Becketti, Gould, Lillard and Welch (1988 [6]) tests on the whole sample. We computed F-tests of the joint significance of the attrition dummy and the interaction control variables on all outcomes of interest, [6] and for none of them is it possible to reject the randomness of the attrition process.

After balancing the panel between years and restricting to non-missing values of control variables, we obtain a total of 3970 observations of both formal and informal household businesses (1,985 per year), respectively 1,972 in Hanoi and 1,998 and Ho-Chi-Minh City. Units of interest, that is household businesses that were initially informal in 2007, represent 73,2% of the sample (2930 observations). The rest of the paper focuses on this sub-sample of initially informal businesses, and compare treated (formalized) and control (remained informal) units. The identification relies on the businesses that were operating informally in 2007 and shifted to the formal sector (i.e. obtained a business license) before the second wave in 2009. 147 units in total followed this trajectory, which represents 10.07% of the total number of informal unit in our data.

3.2 Descriptive statistics: which types of businesses chose formality?

A first descriptive analysis comparing treated and control production units along outcomes and control variables shows that both groups were barely comparable in their characteristics, initial conditions and evolution.

3.2.1 Head's and Businesses' characteristics

As illustrated in table 1, the formalized HB were more often from Ho Chi Minh City, which is somehow coherent with the southern economic capital having a higher income per capita. A recurrent idea in the literature is that formality choice typically occurs when the firm is created, or that new businesses go through a first period of informality before getting registered: it appears in our data that already existing informal business that choose registration are not significantly older or younger than others (reaching an average of 8.4 years in 2009). This result suggests that formalizing remains an option for businesses that were created as informal ones long after the hypothetical test period. However, they differ by their sectorial repartition since most of the treated units operate in the trade sector, and significantly less than control ones in manufacture. Comparing formalized units with other also reveals interesting patterns as regards heads' characteristics: they were significantly more frequently led by males, with a better education (half less numerous to have a primary education or lower, and 80% having reached upper secondary school). We included the motivation for starting business as an additional proxy for head's ability, and it appears that the proportion of heads who started their business to be independent is higher among the formalized

---

[5]A complete descriptive analyses of informality dynamics can be found in Demenet *et al.*, 2010 [11].

[6]The idea is to determine whether the coefficients from the explanatory variables differ between households who are retained or not in the panel. Results are available on demand.

units. Taken together with a higher education level, this tends to indicate a higher ability of heads who chose to formalize.

Table 1: HB&IS panel: observations and basic control variables

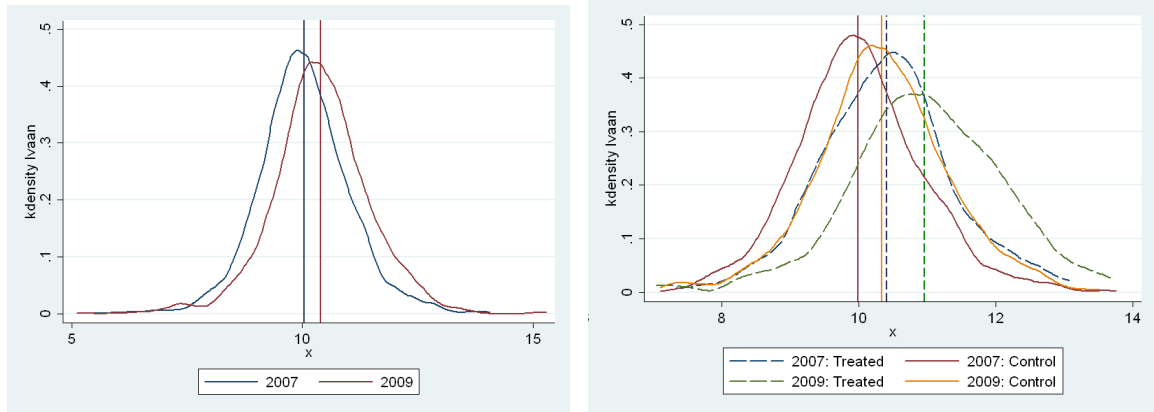|	2007			2009		
	Treated	Control	P-Value	Treated	Control	P-Value
Observations	147	1318		147		1318
Control variables A. IPU characteristics						
City: Hanoi	0.48	0.52	0.334	0.48	0.52	0.334
Duration	6.39	6.42	0.955	8.37	8.41	0.947
Industry: Manufacture	0.14	0.20	0.073	0.16	0.20	0.197
Industry: Trade	0.46	0.32	0.001	0.47	0.31	0.000
Industry: Services	0.41	0.48	0.081	0.37	0.49	0.006
Control variables B. Head characteristics						
Sex: male	0.52	0.42	0.027	0.50	0.42	0.058
Migrant	0.05	0.04	0.305	0.03	0.04	0.546
School: primary or lower	0.13	0.25	0.001	0.14	0.25	0.003
School: upper sec.	0.80	0.70	0.009	0.80	0.71	0.025
School: college or more	0.07	0.04	0.207	0.07	0.05	0.265
Age	44.25	43.82	0.643	45.88	45.6	0.765
Motivation: no work	0.29	0.34	0.214	0.13	0.30	0.000
Motivation: better income	0.27	0.22	0.172	0.30	0.23	0.077
Motivation: independent	0.34	0.24	0.007	0.44	0.32	0.003
Motivation: other/family	0.10	0.20	0.003	0.14	0.15	0.647

3.2.2 Initial performance and evolution

No consensus has been reached on the appropriate measure of outcomes for micro-enterprises in the literature. In the rest of the study, we focus on Value Added[7] defined as the difference between turnover and intermediate consumption (which includes cost of products, raw material, self-consumption, rent, utilities like electricity and water, and other expenses), and measured in log of nominal annual value. Such variable integrates labour income, profits of the entrepreneur and capital income. The three aspects are mingled in the case of self-employed workers or IPU limited to unpaid family workers, but not in the presence of paid workers. We use therefore annual net profits as an alternative outcome measure for robustness checks (difference between value added and total paid wages and taxes), closer to the entrepreneurial profits in a narrow sense.

Initially, formalized units were performing better than those that remained in the informal sector, and their value added increased much more (figure 1). The mean value of annual Value Added was 53,536,150 VND (approximately 2,560 USD) in 2007 for the treated group, significantly different from the figure of the control one (37,174,680 VND or 1780 USD). It appears graphically that the distribution was similar among groups near the tails, and that formalized units were initially more concentrated in the high-end middle levels. Between the two years, annual Value Added knew a small increase in nominal terms for the whole sample of informal firms. This evolution was actually entirely driven by the sub-sample of informal units that chose to enter the

---

[7]Such choice is made in Grimm, Krüger and Lay (2011 [31]), among others

formal sector: their added value grew (at constant prices) by 45.3% while it remained stable for other IPU. The difference in distribution is thus more pronounced in 2009: treated units improved their relative performance -the figures are 101,732,900 and 55,704,610 VND respectively for treated and controls.

(a) Distribution for the whole informal businesses sample

(b) Distribution among treated and control groups

Figure 1: Distribution of (log) annual Value Added

3.2.3 Operating conditions

Comparing treated and controls along the variables measuring their operating conditions (table 2) reveals a high heterogeneity. Informal units that did register were significantly more often equipped with water (36 vs. 53%), electricity (62 vs. 79%), phone (40 vs. 61%) and mobile phone (39 vs. 20%). This gap worsened after the treatment for electricity and phone variables. The proportion of treated with Internet access, that were comparable to the one of controls in 2007, gained 10 percentage points while it remained stable in the other group.

Table 2: Intermediate outcomes by treated and control observations

	2007			2009		
	Controls	Treated	P-value	Controls	Treated	P-value
Observations	1,321	148		1,321	148	
Variables A. Equipment						
Water	.36488	.53378	0.000	.34974	.57432	0.000
Electricity	.6215	.79054	0.000	.57229	.86486	0.000
Phone	.39516	.60811	0.000	.37699	.63514	0.000
Mob. Phone	.20363	.39189	0.000	.44436	.65541	0.000
Internet	.01136	.02703	0.110	.0265	.12162	0.000
Variables B. Scale of operation						
Size	1.5125	1.6689	0.172	1.4762	2.0473	0.000
Outside prem.	.39288	.16892	0.000	.37547	.07432	0.000
Borrowed	.08554	.07432	0.642	.08251	.12162	0.109
Investment	.14156	.16892	0.369	.23013	.22297	0.844
Accounts	.2975	.4527	0.000	.30053	.60811	0.000
Variables C. Expressed problems						
Supply	.09992	.10811	0.754	.09841	.14865	0.057
Customers	.24375	.22973	0.706	.25132	.31757	0.081
Competitors	.37774	.33108	0.266	.36109	.43919	0.062

If treated and controls were significantly different as regards type of premises in 2007 (39% of controls operating outside vs. 17% of treated only) and accounts (28 vs. 45% keeping formal accounts), it is worth noticing that they were comparable in terms of size (1.51 and 1.66 workers respectively), and in terms of proportion of units that borrowed and invested money in the past 12 month (these proportions being very low). The average size of treated units reached 2.05 in 2009, and the difference increased in terms of accounts and premises, but both groups remained comparable as regards borrowing and investment.

Last group of intermediate outcomes indicating the conditions of operation, variables of expressed problems are a self-assessment of the intensity of competitiveness by each PU's head. The proportion of treated and controls that expressed problems with supply of raw material, sales (due to lack of customer) and sales (due to competitors) was comparable in 2007, and increased slightly for the treated group in 2009, making formalized HB significantly different in terms of appreciation of the business environment at the 10% level.

All in all, treated observations were already performing better in 2007 in terms of value added, and the gap worsened. Moreover, they were operating with better equipment, less often outside, and kept more often formal accounts. The specificities of the formalized businesses group, made of initially better-off units, justify the need for a rigorous impact measurement taking selection bias into account.

4   Results and discussion

The main contribution of this paper is to focus on already existing informal businesses that left the informal sector, and to rely on the rich database to decompose the effect and identify the channels though which transition towards formality affects Household Businesses. In a first stage, we evaluate the impact of transition on final outcome, i.e. performance measured by annual value added. These results can be compared with those obtained by Rand and Torm (2012a) [43], and to a certain extent to the rich result of McKenzie and Sakho (2010 [35]). In a second step, we investigate more deeply on the channels through which these effects occur by estimating the impact of formalization on the firms' condition of operation: access to equipment (water, electricity, phone, Internet), scale of operation (size, premises, amount borrowed or invested in the past 12 months, type of accounts), and competitive intensity (problems with suppliers, customers, competitors).

Our result is that leaving the informal sector is associated with a better performance (the effect ranges from 20 to 22%). More specifically, we are able to prove that this positive effect already mentioned in the above-cited papers is made possible by the improvement in operating conditions associated with formalization. Getting registered is associated to a better access to equipment (electricity and Internet) and premises, increased size, more frequent formal accounts, and more intense competition. This makes firms more efficient by releasing many of the constraints associated to informality.

4.1   Impact of formalization on final outcomes

Our purpose is to estimate the causal effect of formalization on outcome measured as annual Value Added. We rely on the 147 units of our panel that were operating informally in 2007 and became formal before 2009, and compare this treated group with the control group of 1,321 businesses that remained informal. Since we do not observe the counterfactual outcomes $\mathbb{E}(Y^f | F = 0)$ and $\mathbb{E}(Y^{inf} | F = 1)$, naive estimates might be subject to both selection bias —better-off businesses self-select into formality— and endogeneity of the legal status —registration decision might be typically correlated with unobserved factors influencing outcomes (e.g. ability, social network).

We denote Y the outcome of interest, F a dummy variable representing the treatment (getting registered), t a time dummy, and X the set of control variables. To measure the effect of formalization on outcome controlling for other covariates, we first run the following regression to double-difference in an OLS setting:

$$Y_{it} = \beta_0 + \beta_1 t_{it} + \beta_2 F_{it} + \beta_3 (F_{it} * t_{it}) + \beta_4 X_{it} + \epsilon_{it} \qquad (1)$$

The effect we want to estimate is then β_3, coefficient of the interaction term between time and treatment. OLS regressions are simple and easily-implementable parametric methods to evaluate the impact of interest, but have several obvious limitations. First, the parametric specification of the outcome, supposed to be linearly dependent from the covariates. The unobserved residual has allegedly an additive and separable form. Thus, even if one has a complete and relevant set of control variables, the estimated ATT will be biased if the real specification of the interest variable is not linear. Secondly, it is sensitive to the distribution of covariates among treated and non-treated, since it is based on a linear extrapolation to build a counterfactual (Goussé *et al.*, 2010 [30]). With the regression $Y_i = \beta_0 X_i + \beta_1 F_i + \epsilon_i$, the average treated individual with covariates $\bar{X}^{F=1}$ will have as counterfactual the value $\bar{Y}^{F=0} + \beta_0(\bar{X}^{F=1} - \bar{X}^{F=0})$. If $\bar{X}^{F=1}$ and $\bar{X}^{F=0}$ are somehow close, the specification of the regression function will not influence the prediction of the average counterfactual. If the distribution of covariates differs, it will be affected. This is all the more problematic

that the number of covariates is high, which is precisely our case given the richness of our survey data.

A first solution is to control for unobserved time-invariant characteristics by introducing fixed effects. We can then control for the fact that initial conditions of explanatory variables may have an impact on the subsequent changes in performance or decision to formalize. We denote the individual fixed effect β_{0i} and re-write the previous model:

$$Y_{it} = \beta_{0i} + \beta_1 t_{it} + \beta_2 (F_{it} * t_{it}) + \beta_3 X_{it} + \epsilon_{it} \tag{2}$$

Fixed-effects models overcome part of the selection bias on unobservables. The counterpart is that variables that do not vary over time for each individual (e.g. structural characteristics: city, industry) are not used. Furthermore, it is unlikely that all the selection on unobservables is because of time-invariant factors; some factors explaining both registration and performance may remain.

Table 3 presents the results of difference-in-difference estimates of the effect of formalization on log annual value added for both OLS and fixed-effects models. We progressively include the two groups of control variables related to production units' characteristics (city, time in business and industry), and then to heads' characteristics (sex, migration, age and proxies for abilities: education and motivation). Our interaction variable of time and registration has a constant positive effect on our performance measure, ranging from 21.5% to 22.3%. It is significant at the 10% level with OLS dif-in-dif specifications and 5% level with fixed-effects. This tends to indicate that the selection effect leads to softly underestimate the gains of registration in terms of performance. This result, considering the whole sample, is consistent with what has been obtained in other papers, even if the observation unit is often bigger firms.

Table 3: Effect of formalization on performance for the whole sample: (DiD) OLS and FE estimates

	(1) OLS	(2) OLS	(3) OLS	(4) FE	(5) FE	(6) FE
Controls	*none*	*IPU char.*	*all*	*none*	*IPU char.*	*all*
Time	0.340***	0.322***	0.294***	0.340***	0.364***	0.361***
	(0.0389)	(0.0398)	(0.0386)	(0.0293)	(0.0354)	(0.0358)
Treated	0.414***	0.456***	0.340***			
	(0.0839)	(0.0837)	(0.0799)			
Formalization (T*t)	0.216*	0.216*	0.223*	0.216**	0.218**	0.215**
	(0.128)	(0.127)	(0.125)	(0.101)	(0.100)	(0.101)
Control variables A. IPU Characteristics						
City:Hanoi		0.0956**	0.0758*			
		(0.0372)	(0.0400)			
IPU duration ref: 0-3 years						
3-10 years		0.0614	0.0737*		-0.0399	-0.0338
		(0.0455)	(0.0446)		(0.0646)	(0.0652)
>10 years		0.0461	0.147***		-0.275**	-0.279**
		(0.0553)	(0.0551)		(0.136)	(0.137)
Industry ref: manufacture						
Trade		-0.450***	-0.332***		-0.395*	-0.399**
		(0.0586)	(0.0565)		(0.205)	(0.195)
Services		-0.292***	-0.227***		-0.0850	-0.0891
		(0.0537)	(0.0513)		(0.212)	(0.203)
Control variables B. Head Characteristics						
Sex			0.262***			0.138
			(0.0375)			(0.152)
Migration			0.0996			-0.243*
			(0.0841)			(0.133)
Schooling ref: Primary or less						
Upper second			0.0893*			0.0627
			(0.0473)			(0.0817)
College or tertiary			0.139			0.0964
			(0.0969)			(0.155)
Age ref: <35						
Age: 35-45			0.0514			-0.0803
			(0.0511)			(0.0965)
Age: 46-60			-0.106**			-0.0837
			(0.0534)			(0.115)
Age >60			-0.468***			-0.177
			(0.0824)			(0.169)
Motivation for starting IPU- ref:no work						
Reason: better income			0.282***			0.0999*
			(0.0496)			(0.0594)
Reason: independent			0.329***			0.0723
			(0.0456)			(0.0554)
Reason: other or family			-0.0531			-0.0653
			(0.0551)			(0.0662)
Constant	9.989***	10.19***	9.850***	10.03***	10.27***	10.22***
	(0.0269)	(0.0602)	(0.0803)	(0.0140)	(0.163)	(0.204)
Observations	2,930	2,930	2,930	2,930	2,930	2,930
R-squared	0.055	0.080	0.146	0.105	0.111	0.120
Number of id				1,465	1,465	1,465

Dependant variable: log of annual Value Added

Robust standard errors in parentheses

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

Conventional DiD estimators rely on the common trend assumption (treated and untreated would have followed parallel paths in the absence of treatment), which is not completely plausible

if pre-treatment characteristics associated with subsequent changes in performance variables are unbalanced between the treated and the untreated. Therefore, the next step of the identification strategy to reduce selection bias is to apply DiD matching, to rebuild acceptable counterfactuals and keep controlling for time-invariant sources of bias. DiD matching estimators have been shown by Smith and Todd (2005) [25] to perform best among the different matching methods. For more details about the theoretical aspect of matching, see Cochran and Rubin (1973) [9], Stuart and Rubin (2007) [49], and Imbens (2004) [22].

Noting p_i be the propensity score score of individual i (that is, probability to register conditional on observables) and d some distance measure, the first matching estimator of the ATT can be written when each treated observation is matched with its closest neighbour:

$$\Delta_M^{ATT} = \frac{1}{N_f} \sum_{F_i=1} [Y_i^f - \sum_{F_j=0} \omega_{N_{inf}}(i,j) Y_j^{inf}]$$

where N_f and N_{inf} are the number of formalized and non-formalized observations, and ω the weight function such as:

$$\omega_{N_{inf}}(i,j) Y_j^{inf} = \begin{cases} 1 \text{ if } d(p_i, p_j) = min_j d(p_i, p_j) \\ 0 \text{ else} \end{cases}$$

The effect is estimated as the average difference of pairwise observations; the closest in terms of propensity score receives a positive weight of 1. Matching individuals on M nearest neighbours (we take M=4 in the following results) allows reducing the variance but increases bias (since we match individuals that are less and less close); with uniform weights we take:

$$\omega_{N_{inf}}(i,j) Y_j^{inf} = \begin{cases} \frac{1}{M} \text{ if } j \in I_M(i) \\ 0 \text{ else} \end{cases}$$

Where $I_M(i)$ is the set of M non-treated individuals closest to observation i:

$$I_M(i) = \{j_1(i), ..., j_M(i)/F_{jk(i)} = 0 \text{ and}$$
$$d(p_i, p_{j1(i)}) \leq d(p_i, p_{j2(i)}) \leq ... \leq d(p_i, p_{jM(i)})$$
$$\leq d(p_i, p_{j1(i)}) \leq d(p_i, p_l) \forall l \notin I_M(i)\}$$

Taking the differential variables to control for time-invariant unobserved factors, we can write the DiD matching estimator (DiDM) as follow (Smith and Todd, 2005 [25]):

$$\Delta_{DiDM}^{ATT} = \frac{1}{N_{ft}} \sum_{F_i=1}^{N_{ft}} \{Y_{ti}^f - \hat{\mathbb{E}}(Y_{ti}^{inf}|p_i, F_i = 0)\} - \frac{1}{N_{ft'}} \sum_{F_j=1}^{N_{ft'}} \{Y_{t'j}^{inf}(X_j) - \hat{\mathbb{E}}(Y_{t'j}^{inf}|p_j, F_j = 0)\} \quad (3)$$

under conditions:

$$0 < P(F = 1|X) < 1$$

which is ensured by probit estimation of the score $P(X)$ and restriction to common support, and:

$$\mathbb{E}(Y_t^{inf} - Y_{t'}^{inf}|P(X), F = 1) = \mathbb{E}(Y_t^{inf} - Y_{t'}^{inf}|P(X), F = 0)$$

Table 4 presents the results of the difference-in-difference matching estimation of the effect of registration on annual value added. Three types of DiD matching are computed. The first one consists in reweighing the observations by the odds-ratio of the score estimated in the first step,

while restricting to common support. The two others are nearest neighbours (with 4 units) and radius matching. Inference with propensity score matching estimators is problematic: standard errors are underestimated because failing to take into account that the score is estimated in a first step. In addition, Abadie and Imbens (2006a [4] and 2006b [3]) demonstrated that bootstrap is not valid for inference in this case. Therefore, the reported S.E. in the case of reweighting and radius matching constitute lower bounds of the real values. We provide consistent S.E. in the case of nearest neighbours matching estimators with replacement, as in Abadie and Imbens (2006b) [3].

Table 4: Effect of formalization on performance: (DiD) Matching estimates

	(A) Reweighting	(B) Nearest neighbour (4)	(C) Radius matching
Coefficient	0.217**	0.217*	0.204*
Standard error	0.102	0.113	0.109
	90 non treated units are off common support		
	*** $p<0.01$, ** $p<0.05$, * $p<0.1$		

This additional reduction of selection bias on observables confirms the overall positive impact on performance, the estimated coefficient being in the same range than the previous specifications. However, the huge heterogeneity of informal sector businesses may lead to contrasted results according to the profile of firms. Subsistence small businesses operate alongside bigger units escaping regulation, and both types may have contrasted motivations for registration, and thus face different consequences. As in McKenzie and Sakho (2010 [35]) we chose to integrate this possibility by splitting the sample between self-employed workers and the rest of units including at least one worker (roughly half of the treated observations, according to their initial status).

Table 5 presents the coefficient of the effect of formalization on annual value added for 6 specifications. The two first specifications are an OLS setting including firstly the control variables of head's characteristics only, and secondly the PU characteristics as well. Models 3 and 4 do the same in a FE configuration, and finally columns 5 and 6 present the results for Matched double differences (reweighting and radius). The more selection on observable is accounted for, the stronger and more significant the effect; this result confirms the downward bias of naive benefits estimation. Moreover, it appears that the major part of previous results was driven by the initially bigger businesses (those having at least one worker, paid or not). While the effect becomes non-significant for sub-sample of initially self-employed workers in all models, it ranges from 35% to 48% for those having initially at least one worker.

Table 5: Disaggregating by size of production units

	(1) OLS controls: Head's char.	(2) OLS controls: all.	(3) FE controls: Head's char.	(4) FE controls: all.	(5) DiD matching: reweighting	(6) DiD matching: radius
Self-employed	0.164	0.144	0.151	0.130	0.101	0.072
	(0.155)	(0.152)	(0.136)	(0.139)	(0.138)	(0.151)
1+ worker	0.347*	0.393**	0.375**	0.370**	0.478***	0.483***
	(0.190)	(0.189)	(0.150)	(0.144)	(0.146)	(0.157)

s.e. in parenthesis

While interpreting these results, one should bear in mind that between the two years of the panel occurred a major macroeconomic crisis that affected the Vietnamese economy (Cling *et al.*, 2010 [24]), dividing the growth rate by 2. Even if the hardest part of the crisis was already behind at the time of the second round of the panel in 2009, it is not impossible that the performance of formal HB was more affected than the one of IHB —due to lower flexibility—, leading to underestimate the effect of formalization. Applying the same evaluation process within another macroeconomic context might show that registration has an even more characterized effect on performance. Furthermore, the nature of the data does not allow taking directly into account the time elapsed between the business registration and the outcome measured. The impact of formalization on performance has no reason to be constant over time and could be under-evaluated if a large share of Production Units registrations occurred too closely before the second survey.

4.2 Impact of formalization on intermediate variables: does it change Household Businesses conditions of operation?

Although showing a plausibly causal effect on final performance when firms leave the informal sector and corroborating the existing results, the previous approach does not tell much about the mechanisms through which this effect occurs. One of the main contributions of this paper is to check a series of hypothesis on these potential channels.

Tables 6 to 8 report the coefficient and significance of the formalization variable on a set of intermediate outcomes representing businesses' conditions of operation. The same methodology is applied, using specifications (1), (2) and (3). For OLS and FE specifications, we control alternatively for (i) the characteristics of HB's heads only (sex, migration, schooling, age class, motivation) and (ii) the characteristics of both the head and the Production Unit itself (city, time in business, type of industry). Most of the outcomes of interest (all except size) are modelled in the form of dummy variables. Therefore, OLS, FE and IV specifications are estimated as linear probability models (LPM), whose usual limitation (to produce unbounded results that may be outside the correct interval) does not apply in a DiD setting.

4.2.1 Access to equipment

Building on the views of Steel and Snodgrass (2008 [48]) and Levy (2007 [29]), the first hypothesis was that leaving the informal sector allows reducing the practical constraints that firms face in terms of access to adequate production means and public goods. This hypothesis is verified, as showed in table 5 that presents the estimation results on the PU's access to water, electricity, phone and Internet. Formalization has a strong and significant effect on access to equipment that (mostly) require legal existence: electricity and Internet. We found that treated units are respectively 13 and 8 per cent more likely to gain access to them between 2007 and 2009 than the control group.

Table 6: Impact of formalization on equipment rates

	Water	Electricity	Phone	Mob. Phone	Internet
DiD: OLS					
Controls: head char.	0.0563	0.117**	0.0383	0.0276	0.0796***
	(0.0590)	(0.0476)	(0.0582)	(0.0552)	(0.0298)
Controls: all	0.0567	0.124**	0.0454	0.0292	0.0797***
	(0.0604)	(0.0483)	(0.0599)	(0.0552)	(0.0299)
DiD: FE					
Controls: head char.	0.0480	0.117***	0.0371	0.0217	0.0786***
	(0.0490)	(0.0408)	(0.0524)	(0.0488)	(0.0280)
Controls: all	0.0515	0.119***	0.0424	0.0231	0.0787***
	(0.0487)	(0.0404)	(0.0519)	(0.0484)	(0.0280)
DiD Matching					
(A) Reweighting	0.0573	0.132***	0.0545	0.0249	0.0807***
	(0.0498)	(0.0418)	(0.0532)	(0.0494)	(0.0282)
(B) Nearest neighbour (4)	0.047	0.115**	0.088	-0.039	0.071**
	(0.061)	(0.0515)	(0.0619)	(0.0601)	(0.0298)
(C) Radius matching	0.066	0.135***	0.063	0.028	0.074**
	(0.0501)	(0.0418)	(0.0528)	(0.0495)	(0.0279)
Observations	2,938	2,938	2,938	2,938	2,938
Treated	296	296	296	296	296

Robust standard errors in parentheses

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

By contrast, formalized Production Units do not have improved access to water, phone or mobile phone. Mobile phones do not require legal existence, nor signing any contract, and are already widespread in Vietnam. Although they almost fully substitute to landlines, access to the latter do require legal existence, and the absence of effect is more surprising. Water supply was already much higher among treated observations before registration (table 3), and did not vary significantly more for the units that registered.

4.2.2 Scale of operation

The second type of results relates to the often alleged constraint that informality represent in terms of scale of operation. Operating without license logically means remaining unnoticed, and thus the possibility has been evoked that informality prevents firms to reach their optimal size (such hypothesis has been made in particular in Fajnzylber, Maloney and Montes-Rojas (2011 [17])). In order to test the possibility that formalization allows HB to operate on a larger scale, table 6 presents the effect of registration on size (discrete variable), probability to have outdoor premises, indicator of having borrowed money in the past 12 month, to have invested in the same period, and finally probability to hold formal accounts.

Table 7: Impact of formalization on scale of operation

	Size	Premises	Borrowed money	Invested	Holds accounts
DiD: OLS					
Controls: head char.	0.412***	-0.0663	0.0496	-0.0320	0.142**
	(0.153)	(0.0432)	(0.0360)	(0.0476)	(0.0586)
Controls: all	0.428***	-0.0783*	0.0508	-0.0308	0.153**
	(0.157)	(0.0431)	(0.0362)	(0.0482)	(0.0601)
DiD: FE					
Controls: head char.	0.408***	-0.0719**	0.0512	-0.0275	0.145***
	(0.102)	(0.0338)	(0.0339)	(0.0455)	(0.0512)
Controls: all	0.416***	-0.0765**	0.0519	-0.0326	0.153***
	(0.104)	(0.0335)	(0.0338)	(0.0459)	(0.0512)
DiD Matching					
(A) Reweighting	0.419***	-0.0650*	0.0524	-0.00980	0.126**
	(0.103)	(0.0346)	(0.0346)	(0.0442)	(0.0519)
(B) Nearest neighbour (4)	0.481***	-0.0828*	0.0946**	-0.0068	0.111
	(0.131)	(0.0393)	(0.0395)	(0.0591)	(0.0641)
(C) Radius matching	0.409***	-0.068*	0.0502	-0.0272	0.137**
	(0.106)	(0.0346)	(0.0341)	(0.0469)	(0.0527)
Observations	2,938	2,938	2,938	2,938	2,938
Treated	296	296	296	296	296

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Household businesses leaving the informal sector increased their size of about 40% more than control units, reaching a medium value of 2.04. This finding confirms the view of Fajnzylber, Maloney and Montes-Rojas (*op.cit.*). Leaving the informal sector is a condition for firms to increase their size, since it will be harder to stay unnoticed. Moreover, it may have an indirect causal effect by the better final performance and improved access to equipment, which could encourage businesses' heads to hire more workers. Getting registered also decreases the probability to operate outdoors (significant at the 10% level), indicating a better access to premises for registered PU. Controlling for endogeneity allows finding a small impact on the probability to have borrowed money during the past 12 month, significant at 10%. By contrast with the results of Rand and Torm (2012a) [43], this is not associated to a higher probability to invest during the same time period. The proportion of units that invested has almost doubled for both controls and treated (table 7), but did not increase faster for the latter. The possibility for the control group to access informal credit channels associated with the potentially remaining credit constraints for formalized businesses may explain why the treated units did not invest significantly more. Finally, formalized businesses are better managed: we find a well-determined effect on the probability to keep formal accounts, which is a necessary condition to operate more efficiently on a larger scale, and may also stem from the constraints associated with the possession of a license.

4.2.3 Intensity of competition

Finally, we estimate the impact of formalization on the heads self-assessment of their environment's competitiveness. Our outcome variables are in this case answers to a subjective appreciation on their situation, that is, "Do you have problems or difficulties in the following domains:"supply of raw materials (quantity or quality), sale of your production (lack of customers), sale of your production (too much competition). Our specifications indicate that the supply of materials did not change in a different manner for treated and controls (in particular, it did not improved for registered units), but controlling for endogeneity we find that PU heads report significantly (at 10%) more problems with customers, and above all between 12 and 17% stronger increase in reported problems with competitors.

Table 8: Impact of formalization on expressed problems

	Supply	Customers	Competitors
DiD: OLS			
Controls: head char.	0.0396	0.0754	0.122**
	(0.0400)	(0.0539)	(0.0577)
Controls: all	0.0423	0.0775	0.121**
	(0.0403)	(0.0542)	(0.0590)
DiD: FE			
Controls: head char.	0.0388	0.0738	0.114**
	(0.0409)	(0.0534)	(0.0581)
Controls: all	0.0428	0.0785	0.119**
	(0.0409)	(0.0532)	(0.0582)
DiD Matching			
(A) Reweighting	0.0358	0.0813	0.120**
	(0.0404)	(0.0531)	(0.0577)
(B) Nearest neighbour (4)	0.078	0.073	0.133
	(0.0475)	(0.0634)	(0.0723)
(C) Radius matching	0.0333	0.076	0.125*
	(0.0417)	(0.0549)	(0.0598)
Observations	2,938	2,938	2,938
Treated	296	296	296
	Robust standard errors in parentheses		
	*** p<0.01, ** p<0.05, * p<0.1		

All in all, our result show that formalized HB benefit from improved conditions of operation. On the one hand, registration facilitates accessing new equipment (electricity and Internet) and operating on a larger scale (bigger size, indoor premises, borrow money more often and hold formal accounts). On the other hand, it seems to be associated with a more competitive environment, since registered heads report more problems with competitors and customers.

In order to investigate on the reasons why the sub-sample of initially self-employed businesses do not benefit from registration as much as bigger businesses in terms of performance, we computed the effect of registration on the two separated groups (results in annex A). This supports our interpretation of improved conditions allowing in turn to increase economic performance, since the initially smaller businesses do not benefit as much from improvements: they do not have better access to electricity, and the effect on size increase is twice lower.

5 Robustness

The following section provides additional robustness checks of the results by answering potential concerns as regards performance measure and definition of formalization. The same approach is applied to alternative variables of both concepts. We find the effect to be consistent, although less strong, with a measure of net profits. Moreover, when defining formalization as tax registration (instead of business license), we find even more characterized results.

5.1 Alternative performance measure: impact of business registration on net profits

Table 11 presents the results of difference-in-differences estimations of the impact of formalization on net profits. Our previous measure of performance, based on the difference between turnover and intermediate consumption, integrates at the same time labour, capital and entrepreneur's incomes. Although relevant when considering businesses' performance as a whole, this concept can be completed by another variable capturing more specifically entrepreneurial profit. Therefore, we subtracted to annual value added the total paid wages and taxes to get a measure of net profits, and computed OLS and Fixed-effects double-differences integrating PU and head's characteristics. Results are provided for the whole sample and for the two sub-sample of self-employed workers and employers.

First, taking the whole population of formalized businesses, the effect remains significant at the 10% level when accounting for unobserved time-invariant heterogeneity. HB that chose to register increased in this case their net profits by 17.9% more than those remaining informal. The impact is however not as high as the improvement of value added. Two reasons may directly explain the somehow inferior effect on net profits. On the one hand, business registration is likely to imply directly an increased amount of tax payed —even if fully unregistered businesses also pay some taxes. On the other hand, one can reasonably assume thanks to the previous results that most of the positive impact of formalization is linked to the improvement of operating conditions that in turn allow to be more efficient. This improvement also implies increased charges for the production unit, in particular when deciding to operate on a larger scale and to hire workers. The cost of manpower can raise both because of the higher number of employees and because of the increase in the unit cost[8], and may in turn decrease net profits (but not value added).

Second, even more strikingly than with value added, the whole difference of performance is concentrated on PU that had at least one worker in 2007 with an impact of 34% while the difference is not significant for initially self-employed workers. This corroborates the idea that the smallest businesses take no direct advantage from registration in terms of performance because they do not benefit in the same terms from the improvement in operating conditions and the possibility to enlarge scale of operation. Their initial amount of capital may be too small —and the credit constraint too strong— to allow them making the necessary investments. This result supports the views of Levenson and Maloney (1998 [28]) and McKenzie and Woodruff (2006 [36]) for whom formality is potentially irrelevant for many micro-firms. Many informal enterprises made of self-employed workers never grow large enough to need the institutions to which formality gives access (credit ones noticeably) while they support direct costs of registration and increased competition.

---

[8]as showed in Rand and Torm (2012a [43], registration decreases the use of casual labour force.

Table 9: Effect of formalization on net profits: OLS and FE estimates by size of PU

	Whole sample		Self-employed		At least one worker	
	(2) OLS	(4) FE	OLS	FE	OLS	FE
Time	0.302***	0.361***	0.358***	0.411***	0.192***	0.241***
	(0.0373)	(0.0359)	(0.0414)	(0.0436)	(0.0676)	(0.0644)
Treated	0.302***		0.323***		0.136	
	(0.0802)		(0.0957)		(0.132)	
Formalization (T*t)	0.187	0.179*	0.0973	0.0824	0.343*	0.338**
	(0.125)	(0.102)	(0.150)	(0.138)	(0.199)	(0.148)
Controls A. IPU Characteristics						
City:Hanoi	0.0850**		0.0530		0.271***	
	(0.0388)		(0.0444)		(0.0691)	
IPU duration ref: 0-3 years						
3-10 years	0.0651	-0.0490	0.102**	-0.0362	0.0578	-0.0941
	(0.0434)	(0.0663)	(0.0495)	(0.0753)	(0.0759)	(0.133)
>10 years	0.126**	-0.235*	0.177***	-0.277*	0.0829	-0.106
	(0.0530)	(0.140)	(0.0606)	(0.158)	(0.0914)	(0.297)
Industry ref: manufacture						
Trade	-0.198***	-0.297	0.119*	-0.455**	-0.310***	-0.245
	(0.0535)	(0.210)	(0.0631)	(0.226)	(0.0854)	(0.355)
Services	-0.0972**	0.0143	0.226***	-0.263	-0.182**	0.496
	(0.0483)	(0.223)	(0.0611)	(0.245)	(0.0781)	(0.374)
Controls B. Head Characteristics						
Sex	0.228***	0.135	0.143***	-0.0132	0.0901	0.455
	(0.0366)	(0.153)	(0.0435)	(0.175)	(0.0664)	(0.289)
Migration	0.0877	-0.264*	-0.0686	-0.208	0.129	-0.345
	(0.0829)	(0.136)	(0.0965)	(0.159)	(0.122)	(0.239)
Schooling ref: Primary or less						
Upper second	0.0670	0.0680	0.112**	0.0676	-0.0816	0.106
	(0.0461)	(0.0817)	(0.0505)	(0.0994)	(0.0842)	(0.140)
College or tertiary	0.0754	0.0863	0.220**	0.158	-0.286*	-0.0232
	(0.0911)	(0.155)	(0.108)	(0.216)	(0.165)	(0.211)
Age ref: <35						
Age: 35-45	0.0472	-0.0705	-0.0507	-0.0914	0.136	0.00758
	(0.0506)	(0.0984)	(0.0561)	(0.118)	(0.0984)	(0.182)
Age: 46-60	-0.0937*	-0.0546	-0.202***	-0.0288	0.0446	-0.108
	(0.0530)	(0.116)	(0.0584)	(0.142)	(0.105)	(0.207)
Age >60	-0.430***	-0.122	-0.513***	-0.0337	-0.323**	-0.212
	(0.0815)	(0.164)	(0.0917)	(0.195)	(0.157)	(0.284)
Motivation for starting IPU- ref:no work						
Better income	0.248***	0.0845	0.172***	0.130*	0.117	-0.00836
	(0.0485)	(0.0610)	(0.0551)	(0.0707)	(0.0913)	(0.122)
independent	0.273***	0.0546	0.214***	0.0689	0.135	0.0350
	(0.0444)	(0.0555)	(0.0496)	(0.0651)	(0.0862)	(0.110)
Other or family	-0.0385	-0.0579	-0.0914	-0.101	-0.0260	0.0431
	(0.0536)	(0.0667)	(0.0556)	(0.0725)	(0.112)	(0.153)
Constant	9.734***	10.06***	9.360***	10.06***	10.37***	10.14***
	(0.0778)	(0.215)	(0.0837)	(0.235)	(0.154)	(0.405)
Observations	2,928	2,928	1,990	1,990	938	938
R-squared	0.120	0.113	0.136	0.147	0.088	0.081
Number of id		1,465		996		469

Dependant variable: log of annual Value Added

Robust standard errors in parentheses

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

5.2   Formalization and registration: is the operational definition relevant?

The statistical definition of informal production units has been adopted internationally in 1993 at the15th International Conference of Labour Statistics (ILO, 1993 [1]) and enlarged during the 90th International Conference of Labour in 2002 under a form that leave a certain degree of freedom to each country as regards the criterion used to define informality. The more frequent are related to registration, and we chose business license in the core analysis to stay in line with the literature on informality in Vietnam (Cling *et al.*, 2010 [23]). However, some concern may remain that a different form of registration may capture the same concept of formalization but lead to contrasted results. To address this concern and provide a larger operational definition of formalization, we reproduce the results with the alternative definition of tax registration (instead of business license). The two aspects are supposed to be highly correlated (in particular because business license is virtually mandatory to register for tax), but we find in our sample that among the 133 units that performed tax registration between the two years, only 85 registered for business license as well (while 49 registered only for taxes).

Tables 10 and 11 reproduce the analysis with this alternative definition of formalization; we focus on the sample of units that were not registered for tax payment in 2007 and define our treated observations as the units that got a tax registration between the two years. Table 10 provides OLS and FE difference-in-difference estimates of tax registration on performance (measured as annual value added). We find a highly significant (at 1% level) and strong effect, as treated units increased their performance 41 or 42% more than the control group, which confirms that the previous results were not sensitive to the choice of the operational definition. Regressions on the two separated sub-samples of self-employed and employers (not reproduced) show a similar pattern to what has been previously showed. The effect is in fact mainly driven by the biggest units, even if it is partly significant for self-employed.

Table 10: Effect of tax registration on annual value added

Controls	(1) OLS none	(2) OLS IPU char.	(3) OLS all	(4) FE none	(5) FE IPU char.	(6) FE all
Time	0.328***	0.307***	0.280***	0.333***	0.353***	0.351***
	(0.0380)	(0.0390)	(0.0378)	(0.0289)	(0.0351)	(0.0354)
Treated	0.484***	0.515***	0.435***			
	(0.111)	(0.109)	(0.101)			
Formalization (T*t)	0.424***	0.432***	0.419***	0.419***	0.422***	0.412***
	(0.163)	(0.161)	(0.156)	(0.122)	(0.120)	(0.120)
Controls A. IPU Characteristics						
City:Hanoi		0.0849**	0.0666*			
		(0.0370)	(0.0397)			
IPU duration ref: 0-3 years						
3-10 years		0.0732	0.0844*		-0.0281	-0.0221
		(0.0454)	(0.0444)		(0.0646)	(0.0652)
>10 years		0.0583	0.159***		-0.254*	-0.258*
		(0.0549)	(0.0547)		(0.135)	(0.136)
Industry ref: manufacture						
Trade		-0.444***	-0.332***		-0.408**	-0.412**
		(0.0584)	(0.0560)		(0.205)	(0.197)
Services		-0.285***	-0.223***		-0.0916	-0.0966
		(0.0536)	(0.0512)		(0.216)	(0.208)
Controls B. Head Characteristics						
Sex			0.262***			0.128
			(0.0372)			(0.149)
Migration			0.104			-0.241*
			(0.0855)			(0.132)
Schooling ref: Primary or less						
Upper second			0.0903*			0.0529
			(0.0472)			(0.0815)
College or tertiary			0.126			0.0785
			(0.0977)			(0.154)
Age ref: <35						
Age: 35-45			0.0521			-0.0786
			(0.0509)			(0.0956)
Age: 46-60			-0.106**			-0.0698
			(0.0535)			(0.115)
Age >60			-0.459***			-0.165
			(0.0818)			(0.167)
Motivation for starting IPU- ref:no work						
Reason: better income			0.300***			0.103*
			(0.0493)			(0.0596)
Reason: independent			0.340***			0.0677
			(0.0457)			(0.0554)
Reason: other or family			-0.0542			-0.0658
			(0.0551)			(0.0659)
Constant	10.00***	10.20***	9.845***	10.03***	10.27***	10.22***
	(0.0262)	(0.0600)	(0.0803)	(0.0140)	(0.165)	(0.205)
Observations	2,921	2,921	2,921	2,921	2,921	2,921
R-squared	0.059	0.084	0.152	0.109	0.115	0.124
Number of id				1,465	1,465	1,465

Dependant variable: log of annual Value Added

Robust standard errors in parentheses

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

Table 11 presents the estimated effect of tax-registration on our measures of conditions of operation. Overall, previous results are not only confirmed but also appear to be stronger with the alternative definition. We find a significant effect (with practically all specifications) that did not appear before on access to phone (at 1%) and mobile phone (at 10% with matching and fixed effects). Tax registration is probably more reliable in Vietnam as a proof of existence to benefit from basic equipment than business license, and it is possible that many businesses did register to overcome the constraint that represents informality as regard access to equipment. Tax registration is also associated with an increase of size that ranges from 60 to 77%, depending on the specification, much higher than the figures of 41% obtained with business license. The probability to operate outdoors is significantly (at 1%) decreased, although much of the effect disappears when controlling for endogeneity (many of the formalized businesses already operated). The proportion of units that borrowed or invested money is not significantly different (and the coefficient is even negative for the latter). As regards business environment, registering for tax is associated with a comparable effect on expressed problems with competitors. Moreover, the coefficient of suppliers problems takes an expected negative sign, although remaining unsignificant.

Table 11: Effect of tax registration on conditions of operation

| Controls: | OLS | | FE | | Matching |
	Head	all	Head	all	(reweighting)
Access to equipment					
Water	0.0978	0.101*	0.0668	0.0687	0.0764
	(0.0629)	(0.0612)	(0.0550)	(0.0549)	(0.0524)
Electricity	0.141***	0.140***	0.0957**	0.0967**	0.104***
	(0.0460)	(0.0460)	(0.0378)	(0.0382)	(0.0376)
Phone	0.222***	0.225***	0.173***	0.179***	0.191***
	(0.0610)	(0.0594)	(0.0485)	(0.0488)	(0.0477)
Mobile Phone	0.136**	0.136**	0.0790	0.0814*	0.0885*
	(0.0579)	(0.0581)	(0.0491)	(0.0490)	(0.0484)
Internet	0.155***	0.155***	0.145***	0.146***	0.146***
	(0.0391)	(0.0388)	(0.0334)	(0.0334)	(0.0323)
Scale of operation					
Size	0.760***	0.766***	0.603***	0.608***	0.604***
	(0.229)	(0.222)	(0.116)	(0.116)	(0.119)
Premises	-0.124***	-0.119***	-0.0777***	-0.0757***	-0.0772***
	(0.0423)	(0.0430)	(0.0275)	(0.0275)	(0.0276)
Borrowed money	0.0583	0.0556	0.0512	0.0505	0.0571*
	(0.0404)	(0.0405)	(0.0354)	(0.0354)	(0.0345)
Invested	-0.0761	-0.0815*	-0.0706	-0.0708	-0.0538
	(0.0506)	(0.0495)	(0.0476)	(0.0473)	(0.0457)
Holds accounts	0.227***	0.219***	0.159***	0.155***	0.143***
	(0.0598)	(0.0592)	(0.0530)	(0.0526)	(0.0534)
Intensity of competition					
Supply	0.0147	0.0109	-0.0165	-0.0185	-0.0340
	(0.0474)	(0.0470)	(0.0478)	(0.0473)	(0.0448)
Customers	0.108*	0.105*	0.0842	0.0835	0.0631
	(0.0592)	(0.0597)	(0.0584)	(0.0585)	(0.0580)
Competitors	0.162**	0.162**	0.128**	0.129**	0.112*
	(0.0632)	(0.0633)	(0.0587)	(0.0584)	(0.0577)
Observations	3,524	3,524	3,524	3,524	3,524
Treated	133	133	133	133	133
Robust standard errors in parentheses					
*** p<0.01, ** p<0.05, * p<0.1					

26

5.3 The dual problem: effect of informalization on originally registered businesses

As mentioned earlier, the HB&IS survey was conducted in a context of financial and economic crisis in Vietnam. As a consequence, alongside the units of our panel that chose to formalize, many of the already registered businesses chose to enter the informal sector, which seemed to play a *buffer* role. This phenomenon was not marginal: 122 units that were originally formal chose to *informalize*. Qualitative interviews conducted with some of these units in 2013 showed that they were going through difficult times, and concretely gave back their license (while they were still operating) to avoid paying tax and gain flexibility. This phenomenon interestingly completes the previous analysis as it can be conceived as its dual problem; we expect to find a negative effect of informalization on our parameters. Table 12 provides the coefficient of this new treatment variable (entering the informal sector, for originally formal businesses) compared to the control group of formal units that remained registered. It is to our knowledge the first causal measurement of the impact of informalization for already existing formal production units. Applying the same methodology to control for endogeneity and selection bias, we find that informalization has a significant negative impact on performance, reducing Value Added by 33 to 46%. It also significantly decreases the probability to access equipment, to hold formal accounts and to operate in fix premises.

Table 12: Effect of informalization on HB performances and operating conditions

	OLS	FE	DiD Matching
Value added	-0.336**	-0.337***	-0.460***
	(0.134)	(0.102)	(0.106)
Access to equipment			
Water	-0.0745	-0.0817	-0.120*
	(0.0689)	(0.0575)	(0.0709)
Electrivity	-0.0942*	-0.0902**	-0.0998***
	(0.0482)	(0.0352)	(0.0385)
Phone	-0.161**	-0.168***	-0.210***
	(0.0648)	(0.0536)	(0.0577)
Mobile Phone	-0.161**	-0.175***	-0.163**
	(0.0670)	(0.0642)	(0.0711)
Internet	-0.124***	-0.122***	-0.0846***
	(0.0379)	(0.0342)	(0.0324)
Scale of operation			
Size	-0.289	-0.287**	-0.146
	(0.225)	(0.132)	(0.198)
Premises	0.154***	0.150***	0.177***
	(0.0466)	(0.0357)	(0.0412)
Borrowed money	-0.0206	-0.0146	-0.0296
	(0.0415)	(0.0402)	(0.0428)
Invested	0.0829*	0.0903*	0.0721
	(0.0495)	(0.0485)	(0.0659)
Holds accounts	-0.121*	-0.117*	-0.156**
	(0.0671)	(0.0633)	(0.0732)
Intensity of competition			
Supply	-0.0192	-0.00342	0.0506
	(0.0541)	(0.0530)	(0.0619)
Customers	-0.0292	-0.0269	0.0143
	(0.0667)	(0.0619)	(0.0799)
Competitors	-0.109	-0.115*	0.0135
	(0.0713)	(0.0682)	(0.0807)
Observations	1,040	1,040	1,040
Treated units	122	122	122

6 Conclusion

In this paper, we addressed two policy-relevant questions on informality that remained unclear so far: the gains directly at stake when firms leave the informal sector on the one hand, and the channels through which it may occur on the other hand. Capitalizing on an exceptional panel micro data, product of a five years research partnership between the IRD (Institute of Research on Development) and the Vietnamese General Statistics Office, we evaluated the causal impact of exiting informality on several measures of operating conditions and performance.

Using a rich Diff-in-diff specification to control for unobserved time-invariant heterogeneity, selection bias and endogeneity of the legal status by fixed-effects and matching DiD estimators, it has been shown that:

- The production units that decided to register are a specific population, whose characteristics were already close to the ones of formal Household Businesses. We therefore support the view expressed (among others) by McKenzie and Woodruff (2006 [36]) that not all production units in this sector have vocation to register: many of them are subsistence activities, for which the point of getting a business license is not even raised. For some others though (mainly the intermediate and "professional" types in the terminology of Cling *et al.*, 2010 [23]) the question is relevant, and the benefits and mechanisms of leaving the informal sector for the business themselves were unclear so far.

- Formalized HB benefit from improved conditions of operation. On the one hand, registration facilitates accessing new equipment (electricity and Internet) and operating on a larger scale (increase size, access indoor premises, borrow money more often and hold formal accounts). On the other hand, it seems to be associated with a more competitive environment, since registered heads report more problems with competitors and customers. Overall, firms are more efficient by escaping many of the constraints associated to informality.

- This ability to operate on a larger scale in a more competitive environment, and to use a better equipment in turn increases their global performance. We estimate the impact of business registration to be at least 20% on annual value added (and 18% on net profits), and even stronger for the alternative measure of tax registration (41%).

Moreover, all of the above results are strongly differentiated by initial size of businesses. Self-employed workers who chose to register did not improve their operating conditions, and neither their value added nor their profits increased significantly. This confirms that a certain threshold of size exists for businesses to be able to benefit from legal existence and public goods.

These results have important policy implications: formalization of household businesses is seen as a way to increase government income by enlarging the tax base. It is also presented as a desirable trend in terms of regulatory power for the State: the more informality, the less sovereignty is an intuitive link. Therefore, policies aiming at encouraging formalization are often advocated for (see e.g. Cling *et al.*, 2010 [24] and World Bank, 2008 [2]). One immediate interpretation of this paper in terms of policy-making is to discuss the conditions of efficiency of such policies.

Firstly, it should be acknowledged that the lower-end group of informal PU, whose justification is mainly subsistence, will not choose to formalize by itself. Since they constitute a huge part of the economy, they should however not be deprived from regulator's attention, and should benefit from specifically targeted programs aiming at improving their operating conditions. The current policy seems to be the opposite in Vietnam: a law has been voted in 2008 prohibiting itinerant

sale, which denotes a will to display a face of emerging city (Pulliat, 2012 [41]) at the expense of the reality of predominant informal activities.

Secondly, for businesses who are concerned by the possibility of registration, encouraging it[9] is relevant since it is likely to improve their conditions of operation, and decisions could be taken to enhance the impact on performance and make the choice more attractive. In particular, we argue that the tax burden should be kept low when entering the formal sector, by applying for instance a low level flat-tax.

---

[9]Among the simplest policy to encourage formalization, a low hanging fruit would be to clarify the registration rules and improve information. Almost 90% of unregistered business are so because they don't know that they have to register (Cling *et al.*, 2010 [23]).

References

[1] Resolution concerning statistics of employment in the informal sector, fifteenth international conference of labour statistics. Technical report, International Labor Organization, January 1993.

[2] *Doing Business 2009.* The World Bank, September 2008.

[3] A. Abadie and G. W. Imbens. Large sample properties of matching estimators for average treatment effects. *Econometrica*, 74:235–267, 2006.

[4] Alberto Abadie and Guido W. Imbens. On the failure of the bootstrap for matching estimators. Working Paper 325, National Bureau of Economic Research, June 2006.

[5] Arup Banerji and Sanjay Jain. Quality dualism. *Journal of Development Economics*, 84(1):234 – 250, 2007.

[6] Becketti, Gould, Lillard, and Welch. The panel study of income dynamics after fourteen years: An evaluation. *Journal of Labor Economics*, 6(4):472–92, October 1988.

[7] Nancy C. Benjamin and Ahmadou Aly Mbaye. The informal sector, productivity, and enforcement in west africa: A firm-level analysis. *Review of Development Economics*, 16(4):664–680, 2012.

[8] Moser C.N. Informal sector or petty commodity production : dualism or independence in urban development. *World Development*, 6:1041–1064, 1978.

[9] W. G. Cochran and D. B Rubin. Controlling bias in observational studies: A review. *Sankhya: The Indian Journal of Statistics*, 16(A35):417–446., 1973.

[10] Suresh De Mel, David McKenzie, and Christopher Woodruff. Returns to capital in microenterprises: Evidence from a field experiment. *Quarterly Journal of Economics*, 123(4):1329 – 1372, 2008.

[11] Axel Demenet, Hien Nguyen Thi Thu, Chi Nguyen Huu, Mireille Razafindrakoto, and François Roubaud. Dynamics of the informal sector in hanoi and ho chi minh city : Main findings of the household business and informal sector survey 2007 and 2009. Technical Report 325, GSO, IRD, World Bank, DFID, June 2010.

[12] Simeon Djankov, Rafael La Porta, Florencio Lopez-De-Silanes, and Andrei Shleifer. The regulation of entry. *The Quarterly Journal of Economics*, 117(1):1–37, February 2002.

[13] Ramilison E. Measuring the informal sector in africa. Technical report, ECA, Addis Abbaba, 2007.

[14] Roubaud F. *"Mixed Surveys Using the Modular Approach," in ILO Manual on Surveys of Informal Employment and Informal Sector.* ILO, Geneva, 2008.

[15] Roubaud F. and M. Séruzier. Economie non enregistrée par la statistique et secteur informel dans les pays en développement. *Statéco*, 68, 1991.

[16] Pablo Fajnzylber, William Maloney, and Gabriel Montes-Rojas. Releasing constraints to growth or pushing on a string? policies and performance of mexican micro-firms. *The Journal of Development Studies*, 45(7):1027–1047, 2009.

[17] Pablo Fajnzylber, William F. Maloney, and Gabriel V. Montes-Rojas. Does formality improve micro-firm performance? quasi-experimental evidence from the brazilian simples program. *Journal of Development Economics*, 94(2):262 – 276, 2011.

[18] Kristin Göbel, Michael Grimm, and Jann Lay. Constrained firms, not subsistence activities: Evidence on capital returns and accumulation in peruvian microenterprises. IZA Discussion Papers 6866, Institute for the Study of Labor (IZA), September 2012.

[19] De Soto H. *The Other Path: The Invisible Revolution in the Third World.* New York: Harper and Row, 1989.

[20] James Heckman and Carmen Pages. Law and employment: Lessons from latin america and the caribbean. NBER Working Papers 10129, National Bureau of Economic Research, Inc, 2004.

[21] ILO. *"Mixed surveys", in Measuring informality: a Statistical Manual on the informal sector and informal employment (pre-edited).* June 2012.

[22] Guido W. Imbens. Nonparametric estimation of average treatment effects under exogeneity: A review. *The Review of Economics and Statistics*, 86(1):4–29, February 2004.

[23] Cling J.-P., Nguyen T. T. Huyen, Nguyen H. Chi, Phan T. N. Tram, Razafindrakoto M., and Roubaud F. *The Informal Sector in Vietnam: A focus on Hanoi and Ho Chi Minh City.* Hanoi: Editions The Gioi, 2010.

[24] Cling J.-P., Razafindrakoto M., and Roubaud F. Assessing the potential impact of the global crisis on the labour market and the informal sector in vietnam. *Journal of Economics & Development*, 38:16–25, June 2009.

[25] A. Smith Jeffrey and Petra E. Todd. Does matching overcome lalonde's critique of nonexperimental estimators? *Journal of Econometrics*, 125(1-2):305–353, 2005.

[26] Jutting J.P. and eds. J.R. de Laiglesia. *s Informal Normal? Towards more and better jobs in developing countries*. Paris, OECD Development Centre, 2009.

[27] Hart K. Informal income opportunities and urban employment in ghana. *The Journal of Modern African Studies*, 11(1):61–89, 1973.

[28] Alec R. Levenson and William F. Maloney. The informal sector, firm dynamics, and institutional participation. Policy Research Working Paper Series 1988, The World Bank, September 1998.

[29] Santiago Levy. *Good Intentions, Bad Outcomes: Informality, Productivity, and Growth in Mexico*. Brookings Institution Press, 2007.

[30] Goussé M., Murard E., Naouas A., and Silavong C. Les estimateurs de matching, une alternative à l'expérimentaiton aléatoire? ENSAE, Rapport de GT, 2010.

[31] Grimm M., Krüger J, and Lay J. Barriers to entry and returns to capital in informal activities: Evidence from sub-saharan africa. *Review of Income and Wealth*, 57:S27–S53, 2011.

[32] Razafindrakoto M., Roubaud F., and C. Torelli. Measuring the informal sector and informal employment: the experience drawn from 1-2-3 surveys in african countries. *African Statistical Journal*, 9:88–129, November 2009.

[33] William F. Maloney. Informality revisited. *World Development*, 32(7):1159–1178, July 2004.

[34] Roxana Maurizio. Labour informality in latin america: the case of argentina, chile, brazil and peru. Working Paper 978-1-907247-64-4, Brooks World Poverty Institute, 2012.

[35] David McKenzie and Yaye Seynabou Sakho. Does it pay firms to register for taxes? the impact of formality on firm profitability. *Journal of Development Economics*, 91(1):15 – 24, January 2010.

[36] Woodruff C. McKenzie, D. Do entry costs provide an empirical basis for poverty traps? evidence from mexican microenterprises. *Economic Development and Cultural Change*, 55:3–42, 2006.

[37] Huu Chi Nguyen, Christophe Jalil Nordman, and François Roubaud. Who suffers the penalty? a panel data analysis of earnings gaps in vietnam. IZA Discussion Papers 7149, Institute for the Study of Labor (IZA), January 2013.

[38] OECD, IMF, ILO, and CIS-STAT. Measuring the non-observed economy: A handbook. Technical report, International Labour Organization, 2002.

[39] Rafael La Porta and Andrei Shleifer. The unofficial economy and economic development. *Brookings Papers on Economic Activity*, 39(2 (Fall)):275–363, 2008.

[40] M. Castells Portes A. and L.A. Benton. *The Informal economy: Studies in advanced and less developed countries*. Baltimore MD, The John Hopkins University Press, 1989.

[41] Gwenn Pulliat. Se nourrir à hanoi : les recompositions du système alimentaire d'une ville émergente. *EchoGéo*, (21), October 2012.

[42] Hussmanns R. *Enquêtes par sondage sur les entreprises du secteur informel dans les pays en développement, In P. Guilbert, D. Haziza, A. Ruiz-Gazen and Y. Tillé (eds.), Méthodes d'enquêtes : applications aux enquêtes longitudinales, à la santé et aux enquêtes électorales dans les pays en développement*. Dunod, Paris., 2008.

[43] John Rand and Nina Torm. The benefits of formalization: Evidence from vietnamese manufacturing smes. *World Development*, 40(5):983–998, May 2012.

[44] John Rand and Nina Torm. The informal sector wage gap among vietnamese micro-firms. *Journal of the Asia Pacific Economy*, 17(4):560–577, 2012.

[45] Donald B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66:688–701, 1974.

[46] Donald B. Rubin. Comment: Neyman (1923) and causal inference in experiments and observational studies. *Statistical Science*, 5(4):472–480, 1990.

[47] Jose Scheinkman and Aureo de Paula. The informal sector. 2007 Meeting Papers 117, Society for Economic Dynamics, 2007.

[48] William F. Steel and Don Snodgrass. *"World Bank Region Analysis on the Informal Economy". In: Raising Productivity and Reducing Risk of Household Enterprises. Annex 1, "Diagnostic Methodology Framework,"*. Washington, DC: World Bank, 2008.

[49] E.A. Stuart and D.B. Rubin. *"Best Practices in Quasi-Experimental Designs: Matching methods for causal inference. Chapter 11 (pp. 155-176). In: Best Practices in Quantitative Social Science. J. Osborne (Ed.). Thousand Oaks, CA: Sage Publications."*. 2007.

[50] C. Udry. *Credit markets in Northern Nigeria: credit as insurance in a rural economy. In: Hoff, K., Braverman, A., Stiglitz, A. (Eds.), The Economics of Rural Organization*. World Bank, Oxford University Press, 1993.

[51] UNstats. System of national accounts 2008. Technical report, UN, EC, OECD, WB, IMF, 2008.

[52] WTO/ILO, Bacchetta M., Ernst E., and Bustamante Juana P. (eds). Globalization and informal jobs in developing countries. Technical report, World Trade Organization and International Labour Organization, 2009.

Appendix A: By which channels do formalization improve performance? Results by initial size

Table 13: Effect of business registration on conditions of operation by initial size

	Whole sample		Self-employed		At least one worker	
	OLS	FE	OLS	FE	OLS	FE
Access to equipment						
Water	0.0567	0.0515	0.0233	0.0390	0.0512	0.0478
	(0.0604)	(0.0487)	(0.0793)	(0.0618)	(0.0896)	(0.0809)
Electricity	0.124**	0.119***	0.0775	0.105*	0.141**	0.159***
	(0.0483)	(0.0404)	(0.0693)	(0.0562)	(0.0638)	(0.0611)
Phone	0.0454	0.0424	-0.0348	-0.00960	0.0859	0.0833
	(0.0599)	(0.0519)	(0.0778)	(0.0681)	(0.0859)	(0.0817)
Mob. Phone	0.0292	0.0231	0.0940	0.0996	-0.0684	-0.0823
	(0.0552)	(0.0484)	(0.0706)	(0.0625)	(0.0864)	(0.0785)
Internet	0.0797***	0.0787***	0.0669*	0.0669**	0.0783*	0.0752*
	(0.0299)	(0.0280)	(0.0400)	(0.0340)	(0.0430)	(0.0443)
Scale of operation						
Size	0.428***	0.416***	0.329***	0.326***	0.713**	0.661***
	(0.157)	(0.104)	(0.102)	(0.101)	(0.287)	(0.200)
Premises	-0.0783*	-0.0765**	-0.0500	-0.0790	-0.0456	-0.0589
	(0.0431)	(0.0335)	(0.0646)	(0.0505)	(0.0552)	(0.0417)
Borrowed money	0.0508	0.0519	0.0666	0.0746*	0.0248	0.0359
	(0.0362)	(0.0338)	(0.0487)	(0.0448)	(0.0528)	(0.0543)
Invested	-0.0308	-0.0326	-0.0486	-0.0479	0.0115	0.0102
	(0.0482)	(0.0459)	(0.0585)	(0.0580)	(0.0791)	(0.0752)
Holds accounts	0.153**	0.153***	0.147*	0.165**	0.146	0.135*
	(0.0601)	(0.0512)	(0.0759)	(0.0701)	(0.0896)	(0.0733)
Expressed problems						
Supply	0.0423	0.0428	0.0956*	0.0961*	-0.0104	0.00219
	(0.0403)	(0.0409)	(0.0511)	(0.0532)	(0.0654)	(0.0670)
Customers	0.0775	0.0785	-0.0185	-0.0236	0.213**	0.239***
	(0.0542)	(0.0532)	(0.0716)	(0.0732)	(0.0838)	(0.0759)
Competitors	0.121**	0.119**	0.0975	0.0998	0.164*	0.157*
	(0.0590)	(0.0582)	(0.0777)	(0.0839)	(0.0864)	(0.0836)
Observations	2,938	2,938	2,938	2,938	2,938	2,938
Treated	296	296	296	296	296	296
Robust standard errors in parentheses						
*** p<0.01, ** p<0.05, * p<0.1						