

The Effects of Formalization on Small and Medium-Sized Enterprise Tax Payments: Panel Evidence from Viet Nam

AMADOU BOLY*

Do firms pay more taxes after formalization? The answer to this question is nontrivial. Tax noncompliance can be a persistent behavior among formerly informal firms. Analyzing the relationship between formalization and tax payments can also be challenging if nonswitching and switching firms have different characteristics. I use a panel dataset built from five small and medium-sized enterprise surveys conducted in Viet Nam from 2005 to 2013. By comparing nonswitching informal firms to switchers, I show that switchers are more likely to pay taxes and to pay a higher amount, thereby confirming heterogeneity. By comparing switchers before and after formalization, I find that formalization increases tax payment likelihood by 20% and the tax amount paid by 93%. A control function approach indicates that my results are robust to potential endogeneity of formalization. Therefore, this paper provides supportive evidence for a key public policy rationale to promote formalization: increased tax revenues.

Keywords: formalization, taxes, Viet Nam

JEL codes: D22, O12, O17

I. Introduction

Existing research has mainly focused on the private costs and benefits of formalization for informal firms, with available evidence suggesting that formalization can have a positive effect on firm performance (McKenzie and Sakho 2010; Fajnzylber, Maloney, and Montes-Rojas 2011; Rand and Torm 2012; Bruhn and McKenzie 2014; Boly 2018).¹ In contrast to most previous studies, this paper

*Amadou Boly: African Development Bank, Abidjan, Côte d'Ivoire. E-mail: a.boly@afdb.org. This article is published here with due acknowledgement of UNU-WIDER, Helsinki, which commissioned the original research under the project *Structural Transformation and Inclusive Growth* in Viet Nam. I am grateful to the Development Economics Research Group at the University of Copenhagen for granting access to the Viet Nam data, in particular Finn Tarp, and to Christina Kinghan for constructing the panel dataset. I would also like to thank Robert Gillanders, Anthony Mveyange, and Robert D. Osei, the managing editor and the anonymous referees for helpful comments and suggestions. The views expressed here are those of the author and do not necessarily reflect those of the African Development Bank Group. ADB recognizes "Vietnam" as Viet Nam. The usual ADB disclaimer applies.

¹Formalization, which is viewed as a deliberate private decision by the firm taken after cost-benefit analysis, occurs when its perceived net benefits are positive (Maloney 2004; De Mel, McKenzie, and Woodruff 2013).

asks whether a government can also benefit from formalization through additional tax payments resulting from firms opting out of the informal sector.

Tax payments are costs to firms, and these costs will increase as a direct consequence of formalization only if there is (partial or full) compliance with formal tax regulations. Yet, noncompliance with formal tax or labor regulations can persist among formerly informal firms; noncompliance is a common phenomenon even among formal firms, including in member countries of the Organisation for Economic Co-operation and Development (OECD 2008).² In consequence, the claim that formalization increases government revenues is ultimately an empirical question, the answer to which can strengthen (or not) a key public policy rationale for promoting the formalization of small and medium-sized enterprises (SMEs).³

Analyzing the effects of formalization on informal firms is challenging due to potential firm heterogeneity. Heterogeneity can come from the fact that firms opting out of the informal sector have different characteristics (e.g., owner capabilities and firm preferences) compared with those that remain informal. Moreover, unobserved characteristics that affect firm outcomes may lead to formalization; if, for example, more successful firms become more visible, leading to a higher likelihood to register formally (McKenzie and Sakho 2010; Fajnzylber, Maloney, and Montes-Rojas 2011).

To study the relationship between formalization and tax payments, I applied regression analysis to a panel dataset compiled using five SME surveys from Viet Nam. I restricted the dataset to informal firms that remained informal throughout the survey periods and to informal firms that became formal at a given point in time (referred to as switching firms or formalized firms). Informality is a multidimensional concept that is difficult to define. However, for my purpose, formal firms are defined as those that registered to pay tax (obtained a tax code), which is a commonly used indicator of formality in the literature (McKenzie and Sakho 2010, Rand and Torm 2012).⁴

McKenzie and Sakho (2010) hypothesize that a profit-maximizing firm becomes formal if and only if the expected present discounted value of the net benefits from doing so outweighs the upfront costs:

$$\sum_{t=1}^T \delta^t E(\pi_{F,t} - \pi_{I,t}) + \theta_{law-abiding} > C_{Money} + C_{Time} + C_{Information}$$

where $\pi_{F,t}$ denotes the firm's profits if it is formally registered at time t , and $\pi_{I,t}$ denotes the firm's profits if it is not formally registered at time t . $\theta_{law-abiding}$ denotes the utility benefit to firm owners from obeying the law and feeling they are contributing to national welfare through paying taxes. C_{Money} , C_{Time} , and $C_{Information}$ denote the monetary, time, and information costs of registering, respectively.

²See also Basu, Chau, and Kanbur (2010) and Tedds (2010).

³See Bruhn and McKenzie (2014), who consider that the claim that formalization is socially optimal because it increases government revenues and reinforces a culture of respecting the rule of the law, requires more research.

⁴An operational definition based on business registration would be consistent with the International Labour Organization's (2003) definition of informal sector enterprises, which are basically "unregistered and/or small-scale private unincorporated enterprises that produce goods or services for sale or barter." Other criteria of informal employment include employment contract registration, provision of social security protection, and size of the

However, the absence of registration does not mean that the informal sector is not taxed, as a nonnegligible fraction of informal firms pay some sort of taxes in Viet Nam, albeit mostly local taxes (Cling, Razafindrakoto, and Roubaud 2011, p. 33).⁵ It is therefore important to ascertain that formalization leads to additional tax payments by switching firms beyond and above what was already paid.

Using the formal status variable (status equals 0 if a firm is informal and 1 if formal), I constructed a variable called *switcher*, which equals 1 for all years in which a switching firm has been observed in my sample, including the years before formalization, and 0 if the firm remained informal throughout the survey periods. The *switcher* variable, as firm-type fixed effects, allows for capturing heterogeneity between firms remaining informal and firms that switched out of the informal sector at some point. As a result, when included in my regression, the variable status, which takes a value of 1 only for switching firms after they have formalized, captures the net effect of formalization on switching firms.

By comparing nonswitching informal firms to switchers, I confirm heterogeneity between switching firms and nonswitching firms regarding tax payments: switchers paid a significantly higher amount of taxes than nonswitchers. Then, by comparing switchers before and after formalization, I find that formalization increases the likelihood of tax payment (by 20%) and the amount of taxes (by 93%) relative to preformalization levels. This significant increase in tax payments persists both in the short and long term. Using a control function approach, I show that my results are robust to potential endogeneity of formalization. I also find that the increase in tax payments is mainly driven by a significant increase in the payment of taxes such as license fees or import and export taxes, but not in the payment of revenue taxes that are arguably more difficult to collect. Firm size, previous performance, and compliance inspection all have a positive relationship with both the amount and likelihood of tax payments.

The remainder of this paper is organized as follows. Section II briefly presents an overview of the existing empirical literature on the effects of formalization. I describe my dataset in section III and explain the empirical approach in section IV. Section V presents the main results and section VI concludes.

II. Literature Review

The literature on the relationship between formalization and firm performance has mainly focused on the private benefits for a firm. Comparing

employer (Henley, Arabsheibani, and Carneiro 2009). However, compared to having a tax code, these other definitions may not be best fits, given that I am seeking whether formalization leads to additional tax payments.

⁵See also Olken and Singhal (2011) for a discussion of informal taxation in 10 developing economies, including Viet Nam.

firms that were created immediately before and after a business tax reduction and simplification scheme in Brazil, Fajnzylber, Maloney, and Montes-Rojas (2011) find that this scheme led to increased levels of registration and to higher revenues, profits, and employment among registered firms. Likewise, Sharma (2014) suggests that registration leads to significant gains in both sales and value added per employee in India. In Bolivia, McKenzie and Sakho (2010) analyze the effects of formalization on firm profits, using the physical distance between a firm and the tax office as an instrument. The assumption is that being closer to a tax office increases the probability of registration. Their findings suggest that the effect of tax registration is positive but heterogeneous. Related to the research question in the current paper, McKenzie and Sakho (2010) also show that registered firms are more likely to pay taxes, but not significantly more likely to be paying a larger share of their profits as taxes.

In contrast to most previous cross-section studies, Rand and Torm (2012) use the same panel data as in this study, but only for 2007 and 2009. Their results indicate that registration leads to an increase in profits and investments for Vietnamese SMEs. Boly (2018) shows that becoming formal can further increase gross profits and the value added of switchers compared with preformalization levels, both in the short and medium term. The present study is closest to Boly (2018) as it uses the same dataset and empirical approach, while focusing on analyzing the relationship between formalization and tax payments. As mentioned previously, the majority of the existing literature focuses on the private benefits of formalization for a firm, leaving an evidence gap on the potential social benefits, specifically those accruing to governments.

III. Data

The description of the dataset used in this study parallels that in Boly (2018), restricted to informal and switching firms. The dataset comes from SME surveys conducted in Viet Nam in 2005, 2007, 2009, 2011, and 2013. The surveys, which were conducted by the Central Institute for Economic Management and the University of Copenhagen, cover about 2,500 firms in each year. They were carried out in 10 locations in the cities of Ha Noi and Hai Phong, in Ho Chi Minh City, and in the rural provinces of Ha Tay, Khanh Hoa, Lam Dong, Long An, Nghe An, Phu Tho, and Quang Nam.

The population of nonstate manufacturing enterprises was based on two data sources from the General Statistics Office of Viet Nam (GSO): the Establishment Census 2002 (GSO 2004) and the Industrial Survey 2004–2006 (GSO 2008). A representative sample of registered household and nonhousehold firms was drawn from this population, using a stratified sampling procedure. The aim was to ensure the inclusion of an adequate number of enterprises in each province with different

ownership forms. For reasons of implementation, the survey was confined to specific areas in each province or city. In addition, the GSO enterprise census focused only on “visible” firms, which are those with fixed professional premises.

Informal household firms were included in the SME surveys based on random onsite identification within the survey districts observed by the enumerator. With such an identification approach, the informal firms included in the survey are those operating alongside officially registered enterprises. These informal firms may be relatively more competitive (and profitable) compared to informal firms clustering in areas with no or very few formal firms (Rand and Torm 2012). In this regard, the sample of informal firms in this study may not be fully representative of the informal sector as a whole in Viet Nam.

Despite the aforementioned weakness, my dataset remains unique by the number of survey years (5), number of firms, and focus on the informal sector. I restricted my sample to firms with at least two observations. I also excluded firms that were formal when initially entering the sample, given that my interest is in informal firms and whether they pay more taxes after formalization. The restricted sample is dominated by informal nonswitchers, which account for 66% of the total number of firms, while switchers account for 34% of the total number of firms (Appendix Table A1.1).⁶

IV. Empirical Approach

To examine the relationship between formalization and tax payments, I use the following specification:

$$y_{it} = \rho D_i^S + \gamma F_{it} + \beta X_{it} + \lambda_t + \mu_i + \varepsilon_{it} \quad (1)$$

where y_{it} is the dependent variable; D_i^S captures firm-type fixed effects (1 if switchers, 0 if nonswitchers); F_{it} indicates whether a firm has become formal or not (0 if a firm is informal, and 1 if the firm is formal); X_{it} represents additional control variables; λ_t denotes a full set of time dummy variables; i indexes individual firms; and t indexes time. As specified, my approach is comparable to a difference-in-difference approach with varying treatment years.

The difference-in-difference specification (between switchers and nonswitchers) rests on the parallel trend assumption in tax payments before switching. A graphical check of this assumption suggests that the parallel trend assumption holds well between 2005 and 2007 and between 2007 and 2009, but

⁶I assume that once a firm becomes formal it stays formal, and I recode its formality status accordingly. In other words, I do not allow a switcher to move back to the informal sector. This applies to about 4.2% of observations, a proportion that I consider negligible. The main justification is that once a firm enters tax authorities' records by acquiring a tax code, it becomes very difficult for the firm to move again into informality.

not between 2009 and 2011. As a robustness check, I run the above specification for different waves of the survey (see section V.C.3).

A key feature of my approach is the construction of a *switcher* variable denoted by D_i^S , using the variable F_{it} (0 if a firm is informal, and 1 if the firm is formal). If a firm has shifted out of the informal sector at any time in my survey periods, the *switcher* variable equals 1 for all years in which the firm has been observed in my sample (including the years before formalization); the *switcher* variable equals 0 if the firm remained informal throughout the survey periods. The (informal) nonswitcher group is used as control group in my regressions. The inclusion of firm-type fixed effects in my main regressions (using dummy variable D_i^S) enables me to account for time-invariant heterogeneity between nonswitching and switching firms. As a result, the variable F_{it} , which takes a value of 1 only for switching firms after they have formalized, picks up the net effects of formalization. However, a known limitation of the fixed-effects approach is that endogeneity due to time-varying omitted variables is still present, although the bias gets smaller than with cross-sectional data. I therefore discuss robustness to endogeneity in section V.C.2.

In addition to the previously mentioned variables, I include several control variables that could affect firms' decisions regarding tax payments. These control variables are summarized in Appendix Table A1.2 for the pooled sample and by firm type.⁷ The control variables include (i) gender of the owner or manager (1 if male, 0 otherwise); (ii) education level of the owner or manager (0 if secondary school has not been completed, 1 otherwise) to proxy for the owner's or manager's human capital; (iii) firm's previous performance using previous year's gross profits; (iv) number of regular full-time employees (in logarithmic form), as well as the square, to control for firm-size effects; (v) whether or not the firm holds a certificate of land use rights to proxy for property rights; (vi) government inspections (0 if the firm has not been subject to inspections in a given year, 1 otherwise); and (vii) dummy variables to control for location and time factors.

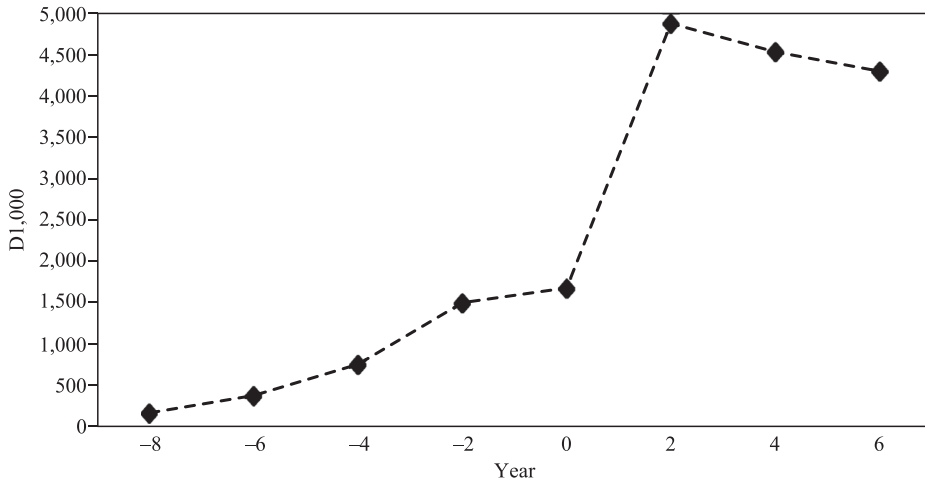
V. Results

After briefly discussing some summary statistics, I present results on the relationship between formalization and total tax payments, my main variable of interest, using a Tobit regression that considers both the binary participation decision and the amount paid.⁸

⁷See also Rand and Torm (2012).

⁸The total tax payments variable is constructed as the sum of various taxes and fees, such as taxes on revenues, license fees, import and export taxes, luxury good taxes, property taxes, and other taxes and fees.

Average Total Tax Payments—Evolution around Formalization Year



Note: Formalization year is equal to 0.

Source: Author's calculations.

A. Summary Statistics

As mentioned earlier, Appendix Table A1.2 describes the dependent variables (for the pooled sample) by firm type, that is, nonswitchers and switchers (and for switchers: overall, before switching, and after switching). I observe that nonswitchers paid some form of taxes 57% of the time, while switchers paid taxes 90% of the time (overall), 83% of the time before switching, and 97% of the time after switching. The differences are significant at the 1% level between nonswitchers and switchers before switching on the one hand, and between switchers before and after switching on the other hand.

The total tax payments made by switchers before switching (D1,778) are significantly higher than that of nonswitchers at the 1% level (D334).⁹ Switchers also pay a significantly higher amount of total taxes after joining the formal sector (D3,792). Overall, the average total tax payments of switchers (D2,883) is about 8.6 times higher than the total tax payments of nonswitchers (D334); switching from the informal to the formal sector resulted in tax payments increasing more than twofold from D1,778 to D3,792.

The figure uses an events-study graph to provide an illustration of the evolution of tax payments for switchers. In this graph, the year of formalization is set at 0 on the x-axis. Negative numbers refer to years in the preformalization period and positive numbers to years in the postformalization period. Although there is an

⁹All monetary values are in real 1,000 Vietnamese dong (D).

Table 1. Effects of Formalization on Total Tax Payments (Tobit)

Variables: Log (1 + taxes), (Real D1,000)	Model 1	Model 2
Switcher (from informal to formal)	2.36*** (0.21)	2.44*** (0.19)
Switcher (after formalization)	1.06*** (0.15)	
Time since becoming formal (dummy, 2 years or less)		1.39*** (0.14)
Time since becoming formal (dummy, 4 + years)		0.58*** (0.17)
Gender of owner or manager (female = 0, male = 1)	0.01 (0.13)	0.01 (0.13)
Owner or manager completed secondary school (no = 0, yes = 1)	0.19 (0.14)	0.21** (0.09)
Gross profits (previous year, log)	0.46*** (0.07)	0.46*** (0.08)
Firm size (log [1 + employment])	2.10*** (0.51)	2.10*** (0.43)
Firm size squared (log [1 + employment])	-0.38*** (0.12)	-0.38*** (0.11)
Share of female employees	0.39 (0.24)	0.39 (0.24)
Own land use right certificate, (no = 0, yes = 1)	0.22* (0.12)	0.23* (0.13)
Compliance inspections (no = 0, yes = 1)	0.73*** (0.10)	0.68*** (0.14)
Constant	-3.87*** (0.73)	-3.94*** (0.93)
Observations	4,795	4,795
Number of panels	1,306	1,306
Time dummies included	Yes	Yes
Province dummies included	Yes	Yes

D = dong.

Notes: Bootstrapped standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

Source: Author's calculations.

increasing trend in tax payments in the preformalization period (-8; -2), I observe a significant jump in the short term, postformalization period (0; +2); this increase in tax payments persists in the long term (+4; +6) compared to the preformalization period (-8; -2).

The preliminary analysis above therefore suggests that tax payments increase after formalization.

B. Panel Regression

I also study the relationship between formalization and tax payments using a random effects Tobit regression (left-censored at 0). The results in Table 1 (Model 1) show that switchers' total tax payments are significantly higher than nonswitchers.

I therefore find heterogeneity between switching and nonswitching firms; that is, switchers tend to pay higher taxes while in the informal sector than nonswitchers do. This heterogeneity has been typically assumed or differenced out in most of the previous studies on formalization.

The amount of switchers' tax payments, both before and after they formalized, is analyzed by looking at the coefficient of *switcher* (after formalization). I find that formalization increases switchers' tax amounts significantly compared to preformalization levels (Table 1, Model 1). For switchers, the mean marginal effects of formalization on the expected value of the censored outcome is about 93%, and the mean marginal effects of formalization on the expected value of the truncated outcome is about 73%, controlling for province and year fixed effects.

To study the effects of formalization over the short and long term, I use two dummy variables that reflect the length of time since formalization. The first dummy (short term) is 1 for firms that have been formal for 2 years or less, while the second dummy (long term) is for firms that have been formal for 4 years or more.¹⁰ As can be seen in Table 1 (Model 2), the increase in total tax payments is observed both in the short and long term. However, the coefficient for total tax payments in the short term (2 years or less after formalization) is significantly higher than the coefficient for tax payments in the long term (4 or more years after formalization), suggesting a decrease in total tax payments over time. Alternatively, the higher coefficients in the short term for total tax payments may be capturing initial entry costs into the formal sector.

Several other control variables are noteworthy in Table 1. First, lagged gross profits have a positive and significant relationship with tax payments at the 1% confidence level, indicating that more successful firms pay more taxes. Second, firm size also has a positive and significant relationship with tax payments, likely explained by the fact that it is more difficult for larger firms to hide their activities. Finally, undergoing at least one compliance inspection is positively related to the amount of total taxes.

C. Robustness Check

In this section, I test the robustness of my results to measurement errors or possible endogeneity.

¹⁰The time since switching to the formal sector is computed as follows. In the first survey year (2005), all firms were informal. The first switchers are recorded in 2007, with formalization having taken place between 2005 and 2007. For all switching firms, I assume that formalization took place in the year of the survey. As a result, for firms that were informal in 2005 but formal in 2007, the number of years since switching is 0 in 2007. For these firms that switched in 2007, the number of years since switching becomes 2 in 2009, 4 in 2011, and 6 in 2013. The same procedure is applied to firms that switched in 2009, 2011, or 2013.

1. Likelihood of Tax Payments

An advantage of the dataset used in this paper is the provision of hard evidence on tax payments made by informal sector firms. However, asking about tax payments through traditional survey techniques can be problematic as firms may decide to misreport.¹¹ Given these possible reporting errors, I conduct a robustness check by using a dummy variable that equals 1 when tax payments are strictly positive (see summary statistics in Appendix Table A1.2). This dummy dependent variable reflects the binary participation decision to pay or not to pay taxes; it is estimated using a random effects logit regression.

Table 2 shows that switchers' likelihood of paying taxes is significantly higher than nonswitchers (Model 1). The results also suggest that becoming formal is positively and significantly associated with an increased likelihood to pay taxes when looking at the coefficient of *switcher* (after formalization). Table 2 also shows that an increased likelihood of paying taxes is observed both in the short and long term (Model 2).

2. Endogeneity¹²

This section analyzes the robustness of my results to potential endogeneity of formalization, using a control function approach (Wooldridge 2015).

As a first step, I estimate a model of the endogenous explanatory variable:

$$F_{it} = 1 [\beta X_{it} + \omega I_{it} + v_{it}] = 1 [\delta Z_{it} + v_{it}] \quad (2)$$

where $1[\cdot]$ is the binary indicator function; F_{it} , the dependent variable, is a dummy variable that takes a value of 1 if a firm is formal and 0 otherwise; X_{it} are control variables described earlier in section IV; I_{it} corresponds to a set of exogenous variables that are omitted from equation (1) and that are partially correlated with formalization; $Z_{it} = (X_{it}, I_{it})$; and v_{it} is an error term.

As a second step, I compute generalized residuals, \hat{r}_{it} , based on results obtained from equation (2) as follows:

$$\hat{r}_{it} = F_{it}\lambda(\hat{\delta}Z_{it}) - (1 - F_{it})\lambda(-\hat{\delta}Z_{it}) \quad (3)$$

where $\lambda(\cdot) = \varphi(\cdot)/\Phi(\cdot)$ is the inverse Mills ratio.

As a third and final step, I reestimate equation (1) by adding \hat{r}_{it} as an additional regressor to control for endogeneity.

To construct I_{it} , I take the annual provincial-level averages for the following two binary variables: access to powered equipment (1 if with access, 0 otherwise),

¹¹For example, the World Bank Enterprise Surveys ask about "compliance of similar firms to yourself" to limit the threat of getting wrong responses.

¹²See also Boly (2018).

Table 2. Effects of Formalization on Total Tax Payments (Logit)

Variables: Dummy Variable (1 if Taxes > 0)	Model 1	Model 2
Switcher (from informal to formal)	1.72*** (0.21)	1.77*** (0.21)
Switcher (after formalization)	1.95*** (0.26)	
Time since becoming formal (dummy, 2 years)		3.32*** (0.50)
Time since becoming formal (dummy, 4 + years)		1.23*** (0.30)
Gender of owner or manager (female = 0, male = 1)	0.03 (0.13)	0.02 (0.13)
Owner or manager completed secondary school (no = 0, yes = 1)	0.10 (0.12)	0.12 (0.12)
Gross profits (previous year, log)	0.33*** (0.08)	0.33*** (0.08)
Firm size (log [1 + employment])	1.39*** (0.41)	1.36*** (0.41)
Firm size squared (log [1 + employment])	-0.29*** (0.11)	-0.29*** (0.11)
Share of female employees	0.37 (0.24)	0.37 (0.24)
Own land use right certificate (no = 0, yes = 1)	0.23* (0.12)	0.22* (0.13)
Compliance inspections (no = 0, yes = 1)	0.66*** (0.15)	0.64*** (0.15)
Constant	-4.07*** (0.82)	-4.13*** (0.82)
Observations	4,795	4,795
Number of panels	1,306	1,306
Time dummies included	Yes	Yes
Province dummies included	Yes	Yes

Note: *** p < 0.01, ** p < 0.05, * p < 0.1.

Source: Author's calculations.

and bribe payments (1 if the firm has made any bribe payments in a given year, 0 otherwise). Here, I restrict my sample to only always-formal firms and formalized firms (i.e., switchers once they have switched), with the latter included in the year following formalization. As noted in section III, informal firms in the sample were selected based on random onsite identification in the neighborhoods of formal firms. In Viet Nam, over 80% of formal firms consider that registration is beneficial, while nearly 50% of informal firms do not see any value to it (Cling, Razafindrakoto, and Roubaud 2012). My exclusive restriction assumption is that informal firms are more likely to formalize when they can observe formal firms' characteristics and potentially attribute those characteristics to formalization. Yet, these formal firms' observable characteristics will not affect informal firms' tax payment behavior.

Table 3. Effects of Formality on Total Tax Payments and Likelihood to Pay Taxes—Control Function Approach

Variables	Formal (No = 0, Yes = 1), Probit	Log Tax (Real D1,000), Tobit		Logit (1 if Taxes > 0)	
		(1)	(2)	(1)	(2)
Switcher (from informal to formal)		2.36*** (0.19)	2.45*** (0.19)	0.22*** (0.02)	0.23*** (0.02)
Switcher (after formalization)		1.06*** (0.14)		0.06*** (0.02)	
Time since becoming formal (dummy, 2 years)			1.44*** (0.18)		0.10*** (0.02)
Time since becoming formal (dummy, 4 + years)			0.62*** (0.22)		0.02 (0.02)
Generalized residuals		0.00 (0.17)	-0.07 (0.17)	0.03 (0.02)	0.02 (0.02)
Gender of owner or manager (female = 0, male = 1)	-0.12* (0.07)	0.01 (0.14)	0.01 (0.11)	0.01 (0.02)	0.01 (0.01)
Owner or manager completed secondary school (no = 0, yes = 1)	0.38*** (0.06)	0.19* (0.10)	0.21* (0.12)	0.01 (0.01)	0.01 (0.01)
Gross profits (previous year, log)	0.35*** (0.06)	0.46*** (0.09)	0.46*** (0.08)	0.04*** (0.01)	0.04*** (0.01)
Firm size (log [1 + employment])	1.01*** (0.22)	2.10*** (0.43)	2.11*** (0.45)	0.19*** (0.05)	0.20*** (0.05)
Firm size squared (log [1 + employment])	-0.16*** (0.05)	-0.38*** (0.11)	-0.38*** (0.12)	-0.04*** (0.01)	-0.04*** (0.01)
Share of female employees	-0.28** (0.12)	0.39 (0.27)	0.39* (0.24)	0.05* (0.03)	0.05* (0.03)
Own land use right certificate (no = 0, yes = 1)	0.01 (0.06)	0.22* (0.12)	0.23* (0.14)	0.03* (0.01)	0.03 (0.02)
Compliance inspections (no = 0, yes = 1)	-0.04 (0.06)	0.73*** (0.16)	0.68*** (0.13)	0.08*** (0.02)	0.08*** (0.02)
Access to machinery by formal firms (province-level average of dummy: 0 = no, 1 = yes)	8.70*** (0.91)				
Bribe paid by formal firms (province-level average of dummy: 0 = no, 1 = yes)	-0.70*** (0.12)				
Constant	-13.38*** (0.93)	-3.87*** (0.92)	-3.94*** (0.82)	-0.28*** (0.10)	-0.28*** (0.09)
Observations	4,795	4,795	4,795	4,795	4,795
Time dummies included	Yes	Yes	Yes	Yes	Yes
Province dummies included	No	Yes	Yes	Yes	Yes
Number of panels		1,306	1,306	1,306	1,306

D = dong.

Notes: Bootstrapped standard errors in parentheses; *** p < 0.01, ** p < 0.05, * p < 0.1.

Source: Author's calculations

The results of the first-step regression in the first columns of Table 3 (Formal, Probit) indicate that bribes paid by a formal firm have a negative and significant effect on the likelihood of formalization, while access to powered equipment has

a positive and significant effect. As a result, I_{it} fulfills a requirement of the control function approach that at least one exogenous variable that is omitted from equation (1) be partially correlated with the dependent variable in equation (2).

The results of the third step (in the control function approach described above) are also presented in Table 3, both for the total tax amount and the likelihood to pay tax. They suggest that endogeneity is not an issue. This provides supportive evidence that formalization can have a positive effect on switchers' total tax payments and that this effect can persist over time for the amount paid but not the likelihood to pay tax.

3. Unbalanced Panel

This paper uses an unbalanced panel, which includes only firms with at least two observations. However, if firms that remain informal are more likely to drop out of the panel (as formalized firms are more visible and easier to locate in subsequent rounds), then this would bias the estimates derived from firms that remain in the panel. As a first robustness check to the unbalanced panel issue, I run regressions using the full sample, or the sample of firms with at least three, four, or five observations. In the second check, I keep firms with at least two observations but vary the number of survey waves, starting with the first two waves in 2005 and 2007.

In all cases, the full model (including controls) is estimated. However, in Appendix Table A2.1 (varying sample size) and Appendix Table A2.2 (varying number of waves), I present the coefficients for only the variables *switcher (from informal to formal)* and *switcher (after formalization)* to save space. My results show that switching firms are different from informal nonswitchers and that the effect of formalization is positive and significant at conventional levels for all samples.

D. Discussion

To explore where the increase in tax payments after formalization comes from, I make a distinction between revenue taxes and other taxes, with the latter obtained by excluding revenue taxes from total tax payments. Such a distinction is motivated by the difficulty to monitor and collect income taxes compared with other types of taxes such as license fees or import and export taxes. This difficulty partly explains the use of presumptive taxes for taxing informal sector activities (Joshi, Prichard, and Heady 2014; Dube and Casale 2016).

Regression analysis indicates that both the likelihood of payments and the amount paid in revenue-related taxes and other taxes by switchers are significantly higher after formalization (Appendix Table A2.3). However, a large share of additional taxes after formalization appears to come from other types of taxes,

not revenue-related taxes. Indeed, switchers report only a small increase in revenue taxes paid before and after switching, based on the sample overall average (D1,152 before switching and D1,175 after switching); while other tax payments appear to increase significantly at more than double the amount paid in revenue-related taxes (D625 before switching and D2,616 after switching). Such a result could suggest that some firms formalize when they realize that in order to start importing or exporting, they need a tax identification number to make subsequent trade-related tax payments. To some extent, these results also confirm that revenue-related taxes can be challenging to collect, even after formalization, compared with other types of taxes such as license fees or import and export taxes. Arguably, collecting trade taxes mainly requires being able to control trade flows at major entry points (e.g., ports, airports, or land borders), while collecting income or sales taxes requires major investments in enforcement and compliance structures throughout the entire economy.

Overall compliance with tax obligations is another interesting aspect that can be considered when analyzing tax payments. To obtain an estimate of the degree of overall compliance, I compared reported total taxes paid to revenue levels by switchers, before and after formalization. However, a precise computation of the level of compliance would require a tax calculator, which is beyond the scope of this paper. The results on compliance should therefore be treated with caution.

I find that, using an overall average, the total tax-to-revenue ratio of switchers slightly increased from 1.26% before formalization to 1.44% after formalization.¹³ Using fractional regression, given that the ratio of total tax to revenue is between 0 and 1, I find that the increase in the compliance level is significant at the 1% level.¹⁴ Notably, the compliance rate of switchers, even after formalization, remains below that of always-formal firms, which stands at 3.47% in my sample, suggesting that switchers may still not be fully compliant with their tax obligations even after joining the formal sector.

VI. Concluding Remarks

Using a panel dataset consisting of five waves of SME surveys in Viet Nam, this paper analyzes the relationship between formalization and tax payments. Such an analysis can be challenging because of potential firm heterogeneity, due to the fact that firms choosing to formalize can have different underlying characteristics compared with the ones that remained informal.

¹³These levels are in line with the lower bound of estimates (1.3%–5% of annual income) provided by Demenet (2016, p. 40).

¹⁴The results of these calculations are available from the author upon request.

To control for unobserved heterogeneity, I created dummy variables that distinguish between two groups of firms in my sample: those that remain informal and those that switch to the formal sector at some point. As a result, when the variable *status*, which takes a value of 1 only for switching firms after they have formalized, is included in my regression, it picks up only the net effects of formalization on tax payments.

My results show that switching firms are different from informal nonswitching firms regarding total tax payments. Such heterogeneity is typically assumed in most previous studies on formalization but not explicitly assessed. After formalization, I observe a significant increase in the amount and likelihood of tax payments, both in the short and long term. These results are mainly driven by a significant increase in the payment of other types of taxes—such as license fees, import and export taxes, and property taxes—not in the payment of revenue taxes.

Overall, my results are supportive of government efforts to reduce the size of the informal sector by promoting formalization. Such efforts are likely to result in additional revenues for governments, which is an important assumption underpinning the public policy rationale for promoting formalization.

References

- Basu, Arnab K., Nancy H. Chau, and Ravi Kanbur. 2010. "Turning a Blind Eye: Costly Enforcement, Credible Commitment, and Minimum Wage Laws." *Economic Journal* 120 (543): 244–69.
- Boly, Amadou. 2018. "On the Short- and Medium-Term Effects of Formalisation: Panel Evidence from Vietnam." *Journal of Development Studies* 54 (4): 641–56.
- Bruhn, Miriam, and David McKenzie. 2014. "Entry Regulation and the Formalization of Microenterprises in Developing Countries." *World Bank Research Observer* 29 (2):186–201.
- Cling, Jean-Pierre, Mireille Razafindrakoto, and François Roubaud. 2011. "The Informal Economy in Viet Nam." Ha Noi: International Labour Organization.
- _____. 2012. "To Be or Not to Be Registered? Explanatory Factors behind Formalizing Non-Farm Household Businesses in Vietnam." *Journal of the Asia Pacific Economy* 17 (4): 632–52.
- De Mel, Suresh, David McKenzie, and Christopher Woodruff. 2013. "The Demand for, and Consequences of, Formalization among Informal Firms in Sri Lanka." *American Economic Journal: Applied Economics* 5 (2): 122–50.
- Demenet, Axel. 2016. "Insights into a Predominant and Dynamic Informal Sector: The Case of Vietnam." *Economies and Finances*, PSL Research University. English. <https://tel.archives-ouvertes.fr/tel-1587795>.
- Dube, Godwin, and Daniela Casale. 2016. "The Implementation of Informal Sector Taxation: Evidence from Selected African Countries." *eJournal of Tax Research* 14 (3): 601–23.
- Fajnzylber, Pablo, William F. Maloney, and Gabriel V. Montes-Rojas. 2011. "Does Formality Improve Micro-Firm Performance? Evidence from the Brazilian SIMPLES Program." *Journal of Development Economics* 94 (2): 262–76.

- General Statistics Office of Viet Nam (GSO). 2004. *Results of Establishment Census of Vietnam 2002: Vol. 2. Business Establishments*. Ha Noi: Statistical Publishing House.
- _____. 2008. *The Real Situation of Enterprises through the Results of Surveys Conducted in 2005, 2006, 2007*. Ha Noi: Statistical Publishing House.
- Henley, Andrew, G. Reza Arabsheibani, and Francisco G. Carneiro. 2009. "On Defining and Measuring the Informal Sector: Evidence from Brazil." *World Development* 37 (5): 992–1003.
- International Labour Organization. 2003. "Guidelines Concerning a Statistical Definition of Informal Employment." 17th International Conference of Labour Statisticians, Geneva.
- Joshi, Anuradha, Wilson Prichard, and Christopher Heady. 2014. "Taxing the Informal Economy: The Current State of Knowledge and Agendas for Future Research." *Journal of Development Studies* 50 (10): 1325–47.
- Maloney, William F. 2004. "Informality Revisited." *World Development* 32 (7): 1159–78.
- McKenzie, David, and Seynabou Sakho. 2010. "Does It Pay Firms to Register for Taxes? The Impact of Formality on Firm Profitability." *Journal of Development Economics* 91 (1): 15–24.
- Olken, Benjamin A., and Monica Singhal. 2011. "Informal Taxation." *American Economic Journal: Applied Economics* 3 (4): 1–28.
- Organisation for Economic Co-operation and Development (OECD). 2008. "Declaring Work or Staying Underground: Informal Employment in Seven OECD Countries." In *OECD Employment Outlook 2008*. Paris.
- Rand, John, and Nina Torm. 2012. "The Benefits of Formalization: Evidence from Vietnamese Manufacturing SMEs." *World Development* 40 (5): 983–98.
- Sharma, Smriti. 2014. "Benefits of a Registration Policy for Microenterprise Performance in India." *Small Business Economics* 42 (1): 153–64.
- Tedds, Lindsay M. 2010. "Keeping It Off the Books: An Empirical Investigation of Firms that Engage in Tax Evasion." *Applied Economics* 42 (19): 2459–73.
- Wooldridge, Jeffrey M. 2015. "Control Function Methods in Applied Econometrics." *Journal of Human Resources* 50 (2): 420–45.

Appendix 1

Table A1.1. Frequency of Sample Firm Types

Firm Type	Overall		Between	
	Frequency	Percentage	Frequency	Percentage
Informal nonswitchers	3,170	63	896	66
Switchers (informal to formal)	1,859	37	458	34
Total	5,029	100	1,354	100

Source: Author's calculations.

Table A1.2. Summary Statistics of Dependent and Independent Variables by Firm Type

Variables	Switchers (Informal to Formal)											
	Nonswitchers			Overall			Before Switching			After Switching		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Dependent												
Total taxes paid (real D1,000)	874	333.9	1,030	432	2,883	7,453	432	1,778	5,672	432	3,792	1,1113
Dummy total taxes (1 if total taxes > 0, 0 otherwise)	874	0.569	0.375	432	0.902	0.192	432	0.834	0.333	432	0.969	0.126
Independent												
Gender of owner or manager (female = 0, male = 1)	874	0.679	0.388	432	0.733	0.351	432	0.748	0.405	432	0.721	0.400
Owner or manager completed secondary school (1 if completed, 0 otherwise)	874	0.391	0.403	432	0.52	0.386	432	0.464	0.466	432	0.574	0.436
Gross profits (previous year)	874	23,182	27,309	432	52,155	61,918	432	44,181	110,186	432	57,961	66,845
Firm size (log [1 + employment])	874	3.39	3,006	432	6.637	6,048	432	7.069	7,567	432	6.171	5,602
Share of female employees	874	0.413	0.257	432	0.328	0.247	432	0.326	0.271	432	0.333	0.260
Own land use right certificate (no = 0, yes = 1)	874	0.734	0.348	432	0.685	0.37	423	0.671	0.440	432	0.697	0.405
Compliance inspections (no = 0, yes = 1)	874	0.132	0.205	432	0.306	0.266	432	0.381	0.427	432	0.241	0.333

D = dong, N = number of observations, SD = standard deviation.

Notes: The time series average of each variable is first calculated by firm before the average group statistics are computed. Although this table shows the amounts of taxes, gross profits, and the number of employees, I use their log transformations in regressions.

Source: Author's calculations.

Appendix 2

Table A2.1. Effects of Formalization on Total Tax Payments (Tobit)—Varying Sample Sizes

Variables: Log (1 + Taxes), (Real D1,000)	Minimum Observations in Panel				
	1	2	3	4	5
Switcher (from informal to formal)	2.31*** (0.18)	2.57*** (0.21)	2.60*** (0.21)	2.75*** (0.21)	2.72*** (0.31)
Switcher (after formalization)	1.14*** (0.15)	1.12*** (0.17)	1.09*** (0.20)	0.90*** (0.19)	0.75*** (0.23)
Constant	-3.29*** (0.94)	-3.07*** (0.83)	-3.63*** (0.99)	-4.25*** (1.20)	-3.48*** (1.31)
Observations	5,155	4,795	4,240	3,393	2,773
Number of panels	1,666	1,306	1,020	731	570
Other controls	Yes	Yes	Yes	Yes	Yes
Time dummies included	Yes	Yes	Yes	Yes	Yes
Province dummies included	Yes	Yes	Yes	Yes	Yes

D = dong.

Notes: Bootstrapped standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

Source: Author's calculations.

Table A2.2. Effects of Formalization on Total Tax Payments (Tobit)—Varying Survey Waves

Variables: Log (1 + Taxes), (Real D1,000)	2005–2007	2005–2009	2005–2011	2005–2013	2007–2009	2007–2011	2007–2013	2009–2011	2009–2013	2011–2013
Switcher (from informal to formal)	2.71*** (0.26)	2.82*** (0.23)	2.61*** (0.20)	2.57*** (0.21)	2.45*** (0.23)	2.27*** (0.21)	2.32*** (0.21)	1.67*** (0.36)	1.77*** (0.35)	0.79 (0.59)
Switcher (after formalization)	1.07*** (0.20)	1.54*** (0.14)	1.31*** (0.18)	1.12*** (0.17)	1.90*** (0.23)	1.58*** (0.17)	1.34*** (0.17)	2.32*** (0.33)	1.99*** (0.30)	2.40*** (0.57)
Constant	-4.23*** (1.33)	-3.67*** (0.95)	-3.81*** (1.01)	-3.07*** (0.83)	-6.67*** (1.17)	-6.04*** (0.87)	-4.41*** (1.06)	-6.40*** (1.12)	-4.43*** (1.18)	-2.35 (1.45)
Observations	1,962	3,017	3,979	4,795	2,128	3,090	3,906	2,017	2,833	1,778
Number of panels	1,075	1,204	1,306	1,306	1,203	1,305	1,305	1,157	1,157	970
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time dummies included	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province dummies included	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

D = dong.

Notes: Bootstrapped standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

Source: Author's calculations.

Table A2.3. Effects of Formalization on Revenue and Other Tax Payments

Variables	Revenue-Related Taxes		Other Taxes	
	Tobit	Logit (1 if Taxes > 0)	Tobit	Logit (1 if Taxes > 0)
Switcher (after formalization)	4.52*** (0.76)	1.01*** (0.23)	1.60*** (0.19)	3.19*** (0.45)
Gender of owner or manager (female = 0, male = 1)	-0.64 (0.79)	-0.15 (0.17)	0.03 (0.13)	0.36 (0.27)
Owner or manager completed secondary school (1 if completed, 0 otherwise)	0.30 (0.62)	0.05 (0.15)	0.23 (0.14)	-0.07 (0.28)
Gross profits (previous year)	1.26*** (0.40)	0.28** (0.11)	0.35*** (0.09)	0.31** (0.13)
Firm size (log [1 + employment])	8.78*** (2.65)	1.88*** (0.58)	2.76*** (0.61)	3.46*** (0.86)
Firm size squared (log [1 + employment])	-1.50*** (0.57)	-0.32*** (0.12)	-0.55*** (0.14)	-0.75*** (0.19)
Share of female employees	-0.40 (1.18)	-0.11 (0.28)	-0.08 (0.28)	-0.01 (0.57)
Own land use right certificate (no = 0, yes = 1)	-0.53 (0.57)	-0.09 (0.17)	-0.20 (0.12)	-0.26 (0.27)
Compliance inspections (no = 0, yes = 1)	0.65 (0.76)	0.15 (0.17)	0.36*** (0.13)	0.67** (0.30)
Constant	-20.96*** (4.66)	-4.53*** (1.20)	-1.93** (0.93)	-4.19*** (1.54)
Observations	1,744	1,744	1,744	1,744
Number of panels	432	432	432	432
Time dummies included	Yes	Yes	Yes	Yes
Province dummies included	Yes	Yes	Yes	Yes

Notes: Bootstrapped standard errors in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1.

Source: Author's calculations.