Boly, Amadou

Working Paper

On the effects of formalization on taxes and wages: Panel evidence from Vietnam

WIDER Working Paper, No. 2015/042

Provided in Cooperation with:
United Nations University (UNU), World Institute for Development Economics Research (WIDER)

Suggested Citation: Boly, Amadou (2015) : On the effects of formalization on taxes and wages: Panel evidence from Vietnam, WIDER Working Paper, No. 2015/042, ISBN 978-92-9230-927-5, The United Nations University World Institute for Development Economics Research (UNU-WIDER), Helsinki, https://doi.org/10.35188/UNU-WIDER/2015/927-5

This Version is available at:
https://hdl.handle.net/10419/126305

UNITED NATIONS UNIVERSITY
UNU-WIDER

WIDER Working Paper 2015/042

On the effects of formalization on taxes and wages

Panel evidence from Vietnam

Amadou Boly*

April 2015

Abstract: Based on a unique panel dataset consisting of both formal and informal firms surveyed every other year from 2005 to 2013, this paper explores the benefits of formalization to the government and firm employees in Vietnam. We find that formalization benefits the government by increasing the amount and the likelihood of tax payments. Formalization also increases the wage level paid by firms that shift out of the informal sector, as well as the share of wages in value added. Our results are therefore supportive of governments' efforts to reduce the size of the informal sector by promoting formalization.

Keywords: Informal sector, benefits of formalization
JEL classification: D21, O12, O17

Acknowledgements: I would like to thank the Development Economics Research Group at the University of Copenhagen for granting access to the Vietnam data; in particular Finn Tarp; and Christina Kinghan for constructing the panel dataset. The views expressed here do not necessarily reflect those of the United Nations University.

*UNU-WIDER; boly@wider.unu.edu.

This study has been prepared within the UNU-WIDER project 'Structural Transformation and Inclusive Growth in Vietnam', co-ordinated by Finn Tarp.

ISSN 1798-7237 ISBN 978-92-9230-927-5

Typescript prepared by Lesley Ellen for UNU-WIDER.

UNU-WIDER gratefully acknowledges the financial contributions to the research programme from the governments of Denmark, Finland, Sweden, and the United Kingdom.

UNU-WIDER, Katajanokanlaituri 6 B, 00160 Helsinki, Finland, wider.unu.edu

1 Introduction

Existing research has mainly focused on the private costs and benefits of formalization for informal firms (see Bruhn and McKenzie 2014). Formalization, viewed as a deliberate private decision by the firm after cost-benefit analyses, occurs only when its perceived net benefits are positive (see e.g. Maloney 2004; de Mel et al. 2011). This paper explores whether there are additional beneficiaries from formalization besides the firm itself, namely: (i) the government, through additional tax revenues or higher likelihood of tax payment by formalized firms, and (ii) the employees, through higher wages or share of wages in value added for firms shifting out of the informal sector.[1] Although scarce, empirical evidence on such 'social' benefits of formalization can strengthen public policy rationales for promoting formalization.[2]

However, analysis of the effects of formalization on informal firms is challenging due to potential selection bias and endogeneity. Selection bias can arise from the fact that firms choosing to formalize have different underlying characteristics, e.g. owner's abilities or firm preferences, compared to the ones that remain informal. In addition, formalization might be correlated with unobserved characteristics that affect firm outcomes. For instance, registration might be partly determined by performance if more successful firms become more visible, leading to a higher probability to formalize in order to avoid paying fines and/or bribes (see e.g. McKenzie and Sakho 2010; Fajnzylber et al. 2011). In summary, formal and informal firms may simply not be comparable due to firm heterogeneity.

To address the above challenges, we use a panel dataset constructed from five surveys of small- and medium-sized enterprises (SMEs) in Vietnam (conducted in 2005, 2007, 2009, 2011 and 2013). We define formal firms as those that are registered to pay taxes (i.e. have a tax code), a common indicator of formality in the literature (Fajnzylber et al. 2009; McKenzie and Sakho 2010; Rand and Torm 2012). Using the formal status variable (namely *Status*: 0 if a firm is informal, and 1 if formal), we construct our main variable of interest, *Switcher*. The latter variable equals 1 if a firm left the informal sector, irrespective of the year it became formal, and 0 if the firm remained formal or informal during the survey periods. Furthermore, we create two dummy variables to make a distinction between firms that remained formal (formal non-switcher) or informal (informal non-switcher) throughout the sample. By differentiating between formal, informal, and switching firms, we are able to account for time-invariant unobserved heterogeneity. Finally, by interacting the variables *Status* and *Switcher*, we obtain the effects of formalization on firm performance before and after switching to the formal sector.

Our findings are supportive of heterogeneity between switching firms and informal non-switching firms, even before formalization, both in terms of tax and wage payments. Such heterogeneity is typically assumed only or differenced out in most previous studies. Compared to those of informal non-switchers, the tax payments of switchers is 4.3 times higher, while their likelihood to pay tax is significantly higher by 9 per cent. Becoming formal multiplies the tax payments of switchers by 2.6 times and increases the likelihood to pay tax by 15 per cent. Finally, we show that formalization increases wage levels by 38 per cent, as well as the share of wages in value added. Our results indicate that formalization is beneficial for both the government and the firm's employees; they

[1] It is worth stressing that we are not analysing the wage gap between the formal and the informal sectors, but the wage gap between (i) informal non-switchers and informal switchers, and (ii) informal switchers before and after switching.

[2] According to Bruhn and McKenzie (2014), a key public policy rationale for formalization that requires more research is the claim that that formalization is socially optimal because it increases government revenues.

are therefore supportive of government's efforts to reduce the size of the informal sector by promoting formalization.

The remainder of this paper is organized as follows. Section 2 briefly presents an overview of the existing literature on the impact of formalization. In Section 3, we describe our dataset. Section 4 discusses the econometric approach, while Section 5 presents the main empirical results. We conclude in Section 6.

2 Literature review

The literature on the impact of formalization on firm performance has mainly focused on private benefits for the firm. It can be divided into two main categories: non-experimental econometric and experimental studies. The first category mainly uses cross-sectional data and relies on one or a combination of methods such as difference-in-difference (before and after an exogenous event), matching, instrumental variables, or regression discontinuity. The majority of these studies find that formalization has a positive impact on firm performance (see e.g. McKenzie and Sakho 2010; Fajnzylber et al. 2011; Rand and Torm 2012). Experimental studies on the other hand suggest that the costs of formalization typically outweigh the benefits, as many firms remain informal despite incentives to formalize (see e.g. Jaramillo 2009; de Andrade et al. 2013; De Mel et al. 2013).

Using firm-level data from Mexico, Fajnzylber et al. (2009) show that being formal increases profits by at least 20 per cent, assuming that formality status is determined by a set of observable variables (matching) or through a specific functional form in the estimation equation (control function).[3] If selection into formality is based partly on unobserved characteristics, however, this may lead to overestimating the effects of formalization (McKenzie and Sakho 2010). Fajnzylber et al. (2011) use regression discontinuity and difference-in-difference to compare firms that were born immediately before and after a business tax reduction and simplification scheme (SIMPLES) in Brazil. They find that this reform leads to increased levels of registration and to higher revenues, profits, and employment among registered firms. This paper concentrates on newly created firms that opt for operating formally, not existing informal-sector firms; the results could therefore reflect self-selection at formal sector entry.

McKenzie and Sakho (2010) estimate the impact of tax registration on firm profits in Bolivia, by using the distance between firm and registration office as an instrument for registration status.[4] The assumption is that being closer to a tax office increases the probability of registration. They find that the overall impact of tax registration is positive but heterogeneous; it leads to higher profits for medium-size firms in their sample, but has negative impact on small and large firms. They also find that owners of larger informal firms have higher entrepreneurial ability than owners of larger formal firms, in contrast to the mainstream view (see for instance La Porta and Shleifer 2008).

An exception to the use of cross-section data is Rand and Torm (2012) who use a matched double-difference with the same panel data as in this study, but for 2007 and 2009. Their results suggest that registration leads to an increase in profits and investments for Vietnamese SMEs. Rand and

[3] Likewise, Sharma (2014) finds, through propensity score matching, that registration leads to significant gains in sales per employee and value added per employee in India.

[4] See also de Vries (2010) who controls for self-selection by using the degree of value-added tax compliance among the firm's suppliers and buyers as instrument. He finds large differences in productivity when comparing formal retailers to informal ones in Brazil.

Torm (2012) also look at the impact of formalization on employees and find that formalization leads to an improvement in employment quality by decreasing the use of casual labour (indicating higher compliance with labour regulations). However, they find no evidence that formalization leads to a higher share of wages in total value added, a variable also analysed in the present paper.

Compared to Rand and Torm (2012), this paper analyses the potential benefits of formalization for government and employees. For employees, we consider wages and the share of wages in total value added. In contrast to Rand and Torm (2012), we find some evidence of a positive effect of formalization on the share of wages in total value added. In addition, by extending the panel dataset up to five observations per firm, the present study makes an explicit distinction between formal, formalized (switcher) and informal firms, allowing to account for time-invariant unobserved heterogeneity. The empirical approach is further explained in Section 4.

A second strand of the literature on the effects of formalization on firm performance uses the experimental approach. This recent experimental evidence suggests that the costs of formalization outweigh the benefits, resulting in many firms remaining informal despite incentives to formalize. De Andrade et al. (2013) conducted a field experiment in the city of Belo Horizonte in Brazil, to examine government actions that promote registration of informal firms. Firms were randomly assigned to a control group or one of four treatment groups: the first received information about how to formalize; the second received information about how to formalize and free registration costs along with the use of an accountant for a year; the third group was assigned to receive an enforcement visit from a municipal inspector; and the fourth group was assigned to have a neighbouring firm receive an enforcement visit to see if enforcement had spillovers. Receiving an inspection increased registration probability by 21 to 27 percentage points, but the three other interventions had no effect. This might suggest that informal firms formalize mostly when forced to do so.

De Mel et al. (2013) provide evidence suggesting that firms become formal as the related benefits increase. In a field experiment in Sri Lanka, the authors found that simply reimbursing the direct costs of registration had no effect on formalization. Yet, 20 per cent of firms registered when offered an amount equivalent to between one-half and one month of the median firm's profits, and 47 per cent registered when offered payments corresponding to two months of the median firm's profits. In follow-up surveys, firms that formalized were found to have higher profits, but this result was driven by a few fast-growing firms; formalizing had no effect on the profits of the majority.

Jaramillo (2009) reports an experiment in Lima, Peru where registration was promoted by subsidizing the full money cost of formalization and providing guidance through the process. Although most firms report greater disadvantages than advantages of being informal, only one out of four firms opted to formalize despite the incentive, suggesting that formalization is simply not desirable for some firms.

As mentioned previously, the majority of the existing literature focuses on private benefits from formalization by analysing the impact of formalization on the firm itself, leaving aside potential 'social' benefits, namely those accruing to government or employees.

3 Data

Our dataset comes from SME surveys conducted in Vietnam in 2005, 2007, 2009, 2011, and 2013.[5] The surveys, covering about 2500 firms in each year, were carried out in ten locations; namely the cities of Hanoi, Hai Phong, and Ho Chi Minh City, and rural provinces of Ha Tay1, Phu Tho, Nghe An, Quang Nam, Khanh Hoa, Lam Dong, and Long An.

The population of non-state manufacturing enterprises was based on two data sources from the General Statistics Office of Vietnam (GSO): the Establishment Census from 2002 (GSO 2004) and the Industrial Survey 2004–06 (GSO 2008). A representative sample of registered household and non-household 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, such as officially registered households, private firms, cooperatives, or limited liability companies. For reasons of implementation, the survey was confined to specific areas in each province/city. In addition, the GSO enterprise census focused only on 'visible' firms (those with fixed professional premises), which resulted in an underestimation of household firms.

Informal household firms were included in the SME survey 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 were those operating alongside officially registered enterprises. These informal firms may be relatively more competitive (and profitable) compared to informal firms clustering in areas with none or very few formal firms (see Rand and Torm 2012). In this regard, the sample of informal firms may not be fully representative of the informal sector as a whole in Vietnam.

Despite the above weakness, our dataset remains unique by the number of survey years (five) and the number of firms. We keep only firms with at least two observations in our sample for a total of more than 11,900 observations (3343 firms) in the dataset. At the panel level, in Table 1, the sample is dominated by formal non-switcher firms, which account for 60 per cent of the total number of firms, followed by informal non-switchers (27 per cent), and switchers (14 per cent).

Table 1: Frequency of firm types

	Overall		Between	
Firm type	Frequency	Per cent	Frequency	Per cent
Informal non-switcher	3,170	26.6	896	26.8
Switcher (informal to formal)	1,859	15.6	458	13.7
Formal non- switcher	6,894	57.8	1,989	59.5
Total	11,923	100.0	3,343	100.0

Source: Based on author's computations.

Tables A1 and A2 in the Appendix describe the dependent and independent variables per year.

5 These surveys were conducted by the Central Institute for Economic Management and the University of Copenhagen as part of a research project funded by Danish International Development Assistance.

4 Econometric approach

We now turn to the regression analyses to examine the effects of formalization on government (tax payments) and employees (wages). We estimate the following model using ordinary least squares (OLS):

$$Y_{it} = \beta X_{it} + \rho D_{it}^{Firm\ type} + w_{it} \quad (1)$$

As explained below, $D_{it}^{Firm\ type}$ is a dummy variable for the firm type (formal, informal, and switching firms).

When the error term is modelled as $w_{\mathrm{it}} = \mu_{\mathrm{i}} + \varepsilon_{\mathrm{it}}$, with $Var(\varepsilon_{\mathrm{it}}) = \sigma_{\varepsilon}^2$ and $Var(\mu_{\mathrm{i}}) = \sigma_{u}^2$ the above equation can be estimated using random-effect model to exploit the panel nature of the data:

$$Y_{it} = \beta X_{it} + \rho D_{it}^{Firm\ type} + \mu_i + \varepsilon_{it} \quad (2)$$

Where $\mathrm{Y_{it}}$ corresponds to tax payments (log); a dummy variable equals to 1 when tax payments is superior to 0, 0 else; total wages (log); total wages per employee (log); share of wages in value added; and share of wages in value added (log).

The average tax payment and the likelihood to pay tax are higher for switchers compared to informal non-switchers, but lower compared to formal non-switchers (see Table A3). The same results apply to wages and share of wages in value added (both total amount and per capita).

As highlighted in the introduction, the main difficulty in identifying the impact of formality on firm outcome variables is that formal, formalized, and informal firms may simply not be comparable due to unobserved firm heterogeneity. To address this issue, we control for firm-level fixed effects relative to non-switcher (formal and informal) and switcher status. Namely, we use the formal status variable *Status* (0 if a firm is informal, and 1 if the firm is formal) to construct our main variable of interest, *Switcher*, which equals 1 if a firm in our sample left the informal sector, irrespective of the year; 0 if the firm remained formal or informal during the survey periods.[6] We create two additional dummy variables to make a further distinction between firms that remain formal (formal non-switcher), and those that remain informal (informal non-switcher) throughout the survey periods; the latter group is used as control group in our regressions.

The inclusion of firm-level fixed effects in our regression model (by using a dummy variable, $D_{it}^{Firm\ type}$, for each type of firm) enables us to account for time-invariant unobserved heterogeneity between formal, informal, and switching firms. Finally, by interacting the variables *Status* and *Switcher*, we obtain the net effects of formalization on firm outcomes before and after switching to the formal sector. Such an approach amounts to a least square dummy variable model relative to firm type that we first estimate using OLS.

We then turn to panel regression by using a random effects model. It can be noted that the use of random effects (instead of fixed effects model in our regression analysis) is driven by the fact that our primary variables of interest are time-constant (i.e. being a formal, informal, and switching firm). A possible downside of random effects modelling relates to the requirement that the firm-

[6] In other words, the variable *Switcher* identifies firms that shifted out of the informal sector at a given point in time.

specific effect (μ_i) is uncorrelated with the explanatory variables. For a robustness check, we apply Mundlak's correction for possible violation of the independence assumption between the covariates and the error term (see e.g. Mundlak 1978; Bell and Jones 2015). Specifically, this is achieved through the inclusion of panel-group means of time-varying (continuous) variables.

In addition to the main variables of interest relating to formality status, we include several covariates, namely: (i) the gender of the owner/manager; (ii) the education level of the owner/manager; (iii) the number of regular full-time employees in log (also the square); (iv) the share of production and service workers over all types of employees; (v) the share of female workers in total regular employment; (vi) the number of government inspections; (vii) whether or not the firm owns a certificate of land use rights (CLUR); and (viii) location, industry, and time dummies.[7] A summary of the control variables is given in Table A2.

We now discuss each control variable in some detail. Unless otherwise mentioned, the average statistics discussed below are significantly different among the three groups of firms.

First, the gender of the owner/manager (0 if female, 1 else) is included as female owners have been found more likely to provide fringe benefits such as annual leave, social benefits, and health insurance (Rand and Tarp 2011), which in turn may affect firm profits (Rand and Torm 2012). The share of male-headed firms is highest among switching firms at 72 per cent, compared to 67 per cent among firms that remain informal, and 62 per cent among firms that were always formal in the sample.

Second, the education level of the owner/manager (0 if secondary school education not completed, 1 else) is used to proxy owner's/manager's human capital. Gennaioli et al. (2013) document large productivity gaps between firms run by educated versus uneducated managers, while Jaramillo (2009) finds post-secondary education to predict formalization. In relation to firm category, 42 per cent of owners/managers in informal non-switching firms have completed secondary school education, compared to 52 per cent in switching firms and 72 per cent in formal non-switching firms. These percentages are significantly different and highlight human capital differences between these three types of firms.

Third, the number of regular full-time employees (in log), as well as the square, are included to control for firm size effects, given that the costs and benefits of becoming formal are likely to vary according to firm size (McKenzie and Sakho 2010). The average size of firms (here number of full-time workers) is 5.5 for informal non-switchers and 7.3 for switchers; the difference is not significantly different between the two groups.[8] Yet, the average size in informal and formalized firms is significantly lower compared to that of formal non-switchers (23.78).

Fourth, the share of production and service workers (as opposed to white-collar workers) measures the average skill level in the firm, which can have an impact on firm performance (Rand and Torm 2012). This share is similar between switchers and formal non-switchers at about 69 per cent, and is significantly higher compared to the share of production workers in informal non-switching firms.

---

[7] The choice of covariates is derived mainly from Rand and Torm (2012).

[8] As noted previously, the sampling strategy may have led to an overrepresentation of relatively more competitive (and profitable) informal firms, given the relatively large average size of the informal firms in the sample (compared to 1.5 in Cling et al. 2010).

Fifth, the share of female workers has been shown to depress wage levels in firms, thereby affecting performance (Larsen et al. 2011), and this might result from women being less productive, being more likely to work in less productive enterprises, or being discriminated against. Consequently, the exact mechanism through which the share of female workers can impact firm performance remains unclear. The average share of female workers is comparatively higher for firms remaining in the informal sector at 41 per cent, relative to firms opting out of the informal sector (34 per cent) and incumbent formal firms (37 per cent).

Sixth, whether or not the firm owns a CLUR is used to proxy property rights. Rand and Torm (2012) typically control for this variable in their empirical model, based on the fact that household firms in Vietnam generally are able to use their CLUR as collateral for a loan, thereby easing potential financial constraints for increased investments and performance. The percentage of firms owing a CLUR is 73 per cent for informal non-switchers, 69 per cent for switchers, and 53 per cent for formal non-switchers respectively.

Seventh, for firms in Peru, Jaramillo (2009) finds inspection visits to be a major disadvantage of formalization, which negatively impacts the decision to formalize, being rated even above paying taxes. While inspections are likely to increase compliance with costly government regulations and affect profits (Rand and Torm 2012), they also increase the probability of registering (de Andrade et al. 2013). The government inspection variable takes value 0 if the firm has received no inspection in a given year, 1 if the number of inspection is equal or more than 1. Regarding government inspections, only 13 per cent of the informal non-switchers received a compliance visit. This rate goes to 28 per cent for switchers and up to 47 per cent for formal non-switchers, indicating that formal firms are more 'visible'.

Finally, dummy variables are used to control for industry, location, and time factors. The industry dummy variable equals 0 if the firm is in low-technology manufacturing, and 1 if the firm is in the medium-low and medium-high technology category.[9] The share of firms in the medium-high technology sector is lowest among firms that remain informal. Location dummies account for the fact that Vietnamese provinces are relatively autonomous, and have implemented centrally planned initiatives with different pace and enthusiasm (Nguyen et al. 2007; Rand and Torm 2012). Time dummies are included to control for potential time effects.

5 Results

In this section, we present results relative to the impact of formalization on the government and on employees respectively through tax payments and wages.

5.1 On the impact of formalization on tax payments

Table A4 presents OLS and random effects regressions for the log and the probability of tax payments. The probability of tax payments is used as a robustness check, in case of possible reporting errors in the tax amounts (see e.g. Rand and Torm 2012).

We find that switching firms are different from informal non-switching firms as they pay a higher amount of tax and have a higher likelihood to do so. Such a difference between informal non-switcher and switcher is explicitly captured in this paper while typically assumed or differenced out in most of the previous studies on formalization. Looking at the coefficient of *Switcher*, our results

[9] We use OECD technology classification.

indicate that the tax payments of switchers is 4.3 times higher, compared to those of informal non-switchers.[10] Likewise, the likelihood of switchers to pay tax is significantly higher by 9 per cent. However, by comparing the regression coefficients between switchers and formal non-switchers, it can be noted that firms that switch from informal to formal pay less tax and have a lower probability to do so, compared to non-switching formal firms. As raising additional revenues for governments is a key rationale for promoting formalization, this result suggests that governments should adjust their tax collection previsions downward if using incumbent formal firms as the reference group. Furthermore, the costs of bringing informal sector firms into the tax net should be taken into account.

We then compared tax payment of switcher before and after they formalized, by looking at the coefficient of '*Switcher* (after formalization)'. The evidence suggests that becoming formal leads to a significant increase in tax payments for switchers—about 2.6 times higher than the amount paid while they were still informal. Formalization also has a positive impact on the likelihood to pay tax, with a marginal increase of 15 per cent. These results give support to the public policy rationale that formalization is socially optimal because it can increase government revenues.

Several other control variables are noteworthy in Table A6. First, firm size has a positive impact on the amount and probability of tax payment. A potential explanation is that it becomes more difficult for larger firms to hide their activities. Second, receiving at least one compliance inspection has a positive impact on the likelihood to pay tax, but not on the amount. Third, owning a CLUR increases the probability and the amount of tax payments, as the firm is easier to locate. Finally, whether the owner or manager of the firm has completed secondary school matters positively.

5.2 On the effects of formalization on wages

We first start by describing differences between informal non-switchers and switchers. The results on wages, presented in Table A5, suggest that switchers tend to pay a higher total amount of wages to their employees, compared to informal non-switchers. Yet, when divided by the number of employees, the difference is no longer significant. An alternative way to look at the impact of formalization on wages is through the share of wages in value added (see Table A6). We find also that the share of wages in value added is significantly higher for switchers compared to informal switchers. Our results are therefore supportive of a wage gap between informal non-switchers and switchers.

Shifting out of the informal sector increases the total wages paid to employees by switchers, by 63 per cent. There is also a significant difference (+38 per cent) in wages per employee for switchers, before and after formalization. Such a result adds to evidence provided by Rand and Torm (2012), which indicates that formalization can improve employment quality by decreasing the use of casual labour.

The result of formalization on the share of wages in value added is significant and positive although not very robust. Precisely, as shown in Table A6, the results are significant for the OLS regressions (at 1 per cent level), and for the generalized least squares (GLS) regressions (10 per cent level and 1 per cent level when log-transformed), but insignificant with Mundlak's approach. Such a result,

---

[10] The estimates are based on the 'standard' random effects regression, unless otherwise specified. As we are using a semi-logarithmic functional form, we estimate the effect of a dummy variable coefficient on the dependent variable as: $g^* = \exp\left(\hat{c} - \frac{1}{2}V(\hat{c})\right) - 1$,

where $\hat{c}$ is the dummy variable coefficient and $V(\hat{c})$ its variance (see Halvorsen and Palmquist 1980; Kennedy 1981).

however, contrasts with Rand and Torm (2012) who find no evidence that formalization increases the share of wages in total value added, possibly due to the weakness of trade unions and collective bargaining among Vietnamese SMEs.

6 Conclusion

Using a panel dataset consisting of five waves of SME surveys in Vietnam, this paper analyses the impact of formalization on additional beneficiaries besides the firm itself, namely the government and the employees. Such an analysis can be challenging because of potential selection bias and endogeneity, due to the fact that firms choosing to formalize can have different underlying characteristics, e.g. owner's abilities or firm preferences, compared to the ones that remained informal.

To control for unobserved heterogeneity, we created dummy variables that distinguish between three groups of firms in our sample: those that remain informal, those that switch to the formal sector, and those that remain formal. Moreover, by interacting the variables for firm status (formal vs informal) and the dummy variable for firms that shift out of the informal sector, we obtain the net effects of formalization on informal switchers, after they opt out of the informal sector.

Our results show that switching firms are different from informal non-switching firms, both in terms of tax and wage payments. Such heterogeneity is typically assumed in most previous studies on formalization. Becoming formal leads to a significant increase in the amount and the likelihood of paying taxes. Relative to wages, we find that that formalization leads to increased wage levels as well as share of wages in value added. The previous results are supportive of governments' efforts to reduce the size of the informal sector by promoting formalization. As shown in this paper, such efforts are not only likely to provide additional revenues to governments, they are also likely to benefit employees through better employment conditions (i.e. higher wage payments).

References

Bell, A., and K. Jones (2015). 'Explaining Fixed Effects: Random Effects Modelling of Time-Series Cross-Sectional and Panel Data'. *Political Science Research and Methods*, 3(1): 133–453.

Bruhn, M., and D. McKenzie (2014). 'Entry Regulation and the Formalization of Microenterprises in Developing Countries'. *The World Bank Research Observer,* 29(2): 186–201.

Cling, J.P., M. Razafindrakoto, F. Rouubaud, H.T.T. Nguyen, C.H. Nguyen, and T.T.N. Phan (2010). *The Informal Sector in Vietnam: A Focus on Hanoi and Ho Chi Minh City*. Hanoi: The Gioi Editions.

de Andrade, G.H., M. Bruhn, and D. McKenzie (2013). 'A Helping Hand or the Long Arm of the Law? Experimental Evidence on what Governments can do to Formalize Firms'. IZA Discussion Paper 7402. Bonn: IZA.

de Mel, S., D. McKenzie, and C. Woodruff (2011). 'What is the Cost of Formality? Experimentally Estimating the Demand for Formalization'. Mimeo.

de Mel, S., D. McKenzie, and C. Woodruff (2013). 'The Demand for, and Consequences of, Formalization among Informal Firms in Sri Lanka'. *American Economic Journal: Applied Economics,* 5(2): 122–50.

de Vries, G.J. (2010). 'Small Retailers in Brazil: Are Formal Firms really more Productive?'. *The Journal of Development Studies*, 46(8): 1345–66.

Fajnzylber, P., W.F. Maloney and G.V. Montes-Rojas (2009). 'Releasing Constraints to Growth or Pushing on a String? Policies and Performance of Mexican Micro-Firms'. *The Journal of Development Studies*, 45(7): 1027–47.

Fajnzylber, P., W.F. Maloney, and G.V. Montes-Rojas (2011). 'Does Formality Improve Micro-Firm Performance? Evidence from the Brazilian SIMPLES Program'. *Journal of Development Economics*, 94(2): 262–76.

Gennaioli, N., R. La Porta, F. Lopez-de-Silanes, and A. Shleifer (2013). 'Human Capital and Regional Development'. *The Quarterly Journal of Economics*, 105: 164.

GSO (2004). *Results of Establishment Census of Vietnam 2002: Vol. 2. Business Establishments*. Hanoi: Statistical Publishing House.

GSO (2008). *The Real Situation of Enterprises through the Results of Surveys Conducted in 2005, 2006, 2007*. Hanoi: Statistical Publishing House.

Halvorsen, R., and R. Palmquist (1980). 'The Interpretation of Dummy Variables in Semilogarithmic Equations'. *American Economic Review*, 70(3): 474–75.

Jaramillo, M. (2009). 'Is there Demand for Formality among Informal Firms?: Evidence from Microfirms in Downtown Lima'. Discussion Paper 12/2009. Bonn: German Development Institute.

Kennedy, P.E. (1981). 'Estimation with Correctly Interpreted Dummy Variables in Semilogarithmic Equations'. *American Economic Review*, 71(4).

La Porta, R., and A. Shleifer (2008). 'The Unofficial Economy and Economic Development'. *Brookings Papers on Economic Activity*, 39(2) 275–363.

Larsen, A.F., J. Rand, and N. Torm (2011). 'Do Recruitment Ties Affect Wages? An Analysis using Matched Employer–Employee Data from Vietnam'. *Review of Development Economics*, 15(3): 541–55.

Maloney, W.F. (2004). 'Informality Revisited'. *World Development*, 32(7): 1159–78.

McKenzie, D., and Y. 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.

Mundlak, Y. (1978). 'On the Pooling of Time Series and Cross Section Data'. *Econometrica*, 46: 69–85.

Nguyen, B.T., J. Albrecht, S. Vroman, and M.D. Westbrook (2007). 'A Quantile Regression Decomposition of Urban–Rural Inequality in Vietnam'. *Journal of Development Economics*, 83(2): 466–90.

Rand, J., and F. Tarp (2011). 'Does Gender Influence the Provision of Fringe Benefits? Evidence from Vietnamese SMEs'. *Feminist Economics*, 17(1): 59–87.

Rand, J., and N. Torm (2012). 'The Benefits of Formalization: Evidence from Vietnamese Manufacturing SMEs'. *World Development*, 40(5): 983–98.

Sharma, S. (2014). 'Benefits of a Registration Policy for Microenterprise Performance in India'. *Small Business Economics*, 42(1): 153–64.

Appendix

Table A1: Summary statistics of dependent variables

Variable	Year	Obs	Mean	Sd	Min	Max
Log tax (real 1,000 VND)	2005	2,294	6.60	3.27	0	14.81
	2007	2,555	6.61	3.26	0	14.29
	2009	2,536	7.04	3.79	0	16.25
	2011	2,435	7.38	3.44	0	16.71
	2013	2,095	7.10	3.29	0	15.98
	All	11,915	6.94	3.44	0	16.71
Dummy tax (1 if tax payment>0, 0 else)	2005	2,294	0.85	0.35	0	1
	2007	2,555	0.85	0.35	0	1
	2009	2,536	0.83	0.37	0	1
	2011	2,435	0.89	0.31	0	1
	2013	2,095	0.90	0.31	0	1
	All	11,915	0.86	0.34	0	1
Log wages (log, real 1,000 VND)	2005	2,295	8.62	4.27	0	16.31
	2007	2,555	8.64	4.33	0	16.43
	2009	2,539	8.52	4.57	0	15.86
	2011	2,434	8.49	4.70	0	15.24
	2013	2,094	8.37	4.60	0	16.59
	All	11,917	8.54	4.50	0	16.59
Log wages per employee (log, real 1,000 VND)	2005	1,878	8.25	0.68	4.84	12.46
	2007	2,081	8.38	0.64	4.09	11.27
	2009	2,007	8.59	0.67	5.27	11.98
	2011	1,886	8.75	0.60	5.40	11.48
	2013	1,633	8.66	0.61	4.33	12.34
	All	9,485	8.52	0.67	4.09	12.46
Share of wages in value added	2005	2,254	0.36	0.24	0	0.86
	2007	2,531	0.32	0.22	0	0.86
	2009	2,526	0.33	0.23	0	0.86
	2011	2,411	0.33	0.23	0	0.86
	2013	2,027	0.34	0.24	0	0.85
	All	11,749	0.33	0.23	0	0.86
Log share of wages in value added	2005	2,254	0.29	0.18	0	0.62
	2007	2,531	0.26	0.17	0	0.62
	2009	2,526	0.27	0.18	0	0.62
	2011	2,411	0.27	0.18	0	0.62
	2013	2,027	0.28	0.19	0	0.61
	All	11,749	0.27	0.18	0	0.62

Source: Based on author's computations.

Table A2: Summary statistics of independent variables

Variable	Year	Obs	Mean	Sd	Min	Max
Share of female employees	2005	2,296	0.36	0.28	0.00	1.00
	2007	2,555	0.37	0.27	0.00	1.00
	2009	2,539	0.37	0.27	0.00	1.00
	2011	2,435	0.37	0.26	0.00	1.00
	2013	2,097	0.39	0.26	0.00	1.00
	All	11,922	0.37	0.27	0.00	1.00
Share of production workers	2005	2,296	0.83	0.17	0.00	1.00
	2007	2,555	0.65	0.22	0.00	0.99
	2009	2,539	0.65	0.21	0.00	0.98
	2011	2,435	0.63	0.22	0.00	0.97
	2013	2,097	0.60	0.23	0.00	1.00
	All	11,922	0.67	0.22	0.00	1.00
Firm size (log (1+employment))	2005	2,296	2.18	1.02	0.69	6.80
	2007	2,555	2.14	1.04	0.00	7.17
	2009	2,539	2.08	1.00	0.00	6.22
	2011	2,435	2.03	1.02	0.00	5.77
	2013	2,097	1.95	0.99	0.00	7.44
	All	11,922	2.08	1.02	0.00	7.44
Firm size square (log (1+employment))	2005	2,296	5.80	5.69	0.48	46.27
	2007	2,555	5.66	5.83	0.00	51.42
	2009	2,539	5.32	5.29	0.00	38.65
	2011	2,435	5.16	5.33	0.00	33.35
	2013	2,097	4.79	5.06	0.00	55.34
	All	11,922	5.36	5.47	0.00	55.34
Gender of owner/manager (female=0, male=1)	2005	2,296	0.69	0.46	0.00	1.00
	2007	2,555	0.67	0.47	0.00	1.00
	2009	2,539	0.66	0.47	0.00	1.00
	2011	2,435	0.63	0.48	0.00	1.00
	2013	2,096	0.61	0.49	0.00	1.00
	All	11,921	0.65	0.48	0.00	1.00
Own land use right certificate, CLUR (no=0, yes=1)	2005	2,101	0.60	0.49	0.00	1.00
	2007	2,555	0.57	0.50	0.00	1.00
	2009	2,539	0.61	0.49	0.00	1.00
	2011	2,435	0.66	0.47	0.00	1.00
	2013	2,098	0.70	0.46	0.00	1.00
	All	11,728	0.63	0.48	0.00	1.00
Compliance inspections (no = 0, yes = 1)	2005	2,296	0.47	0.50	0.00	1.00
	2007	2,555	0.58	0.49	0.00	1.00
	2009	2,539	0.57	0.49	0.00	1.00
	2011	2,435	0.06	0.24	0.00	1.00
	2013	2,098	0.01	0.10	0.00	1.00
	All	11,923	0.35	0.48	0.00	1.00

Variable	Year	Obs	Mean	Sd	Min	Max
Owner/manager completed secondary school	2005	2,296	0.57	0.50	0.00	1.00
	2007	2,555	0.55	0.50	0.00	1.00
	2009	2,539	0.58	0.49	0.00	1.00
	2011	2,435	0.62	0.49	0.00	1.00
	2013	2,098	0.70	0.46	0.00	1.00
	All	11,923	0.60	0.49	0.00	1.00
Medium high-tech sector dummy	2005	2,291	0.36	0.48	0.00	1.00
	2007	2,555	0.34	0.47	0.00	1.00
	2009	2,539	0.34	0.47	0.00	1.00
	2011	2,435	0.34	0.47	0.00	1.00
	2013	2,095	0.33	0.47	0.00	1.00
	All	11,915	0.34	0.47	0.00	1.00

Source: Based on author's computations.

Table A3: Summary statistics of dependent and independent variables, by firm types

	Informal non-switcher			Switcher (informal to formal)			Formal non-switcher		
Variables	N	Mean	Sd	N	Mean	Sd	N	Mean	Sd
Dependent									
Log tax (real 1,000 VND)	896	3.37	2.66	458	6.18	1.90	1,989	8.92	1.63
Dummy tax (1 if tax payment>0, 0 else)	896	0.59	0.38	458	0.89	0.21	1,989	0.99	0.06
Log wages (log, real 1,000 VND)	896	4.58	4.45	458	8.41	3.23	1,989	10.56	2.38
Log wages per employee (log, real 1000 VND)	551	8.03	0.70	427	8.30	0.47	1,964	8.63	0.48
Share of wages in value added	895	0.19	0.21	458	0.35	0.18	1,989	0.40	0.15
Log share of wages in value added	895	0.16	0.17	458	0.29	0.14	1,989	0.33	0.11
Independent									
Formal non-switcher	896	0	0	458	0	0	1,989	1	0
Switcher (from informal to formal)	896	0	0	458	1	0	1,989	0	0
Switcher (after formalization)	856	0	0	458	0.56	0.21	1,989	0	0
Share of female employees	896	0.41	0.26	458	0.34	0.25	1,989	0.37	0.22
Share of production workers	896	0.61	0.19	458	0.70	0.15	1,989	0.69	0.14
Number of workers (full-time)*	896	5.53	32.92	458	7.30	11.25	1,989	23.78	49.42
Gender of owner/manager (female=0, male=1)	896	0.67	0.40	458	0.72	0.36	1,989	0.62	0.40
Own land use right certificate, CLUR (no=0, yes=1)	896	0.73	0.36	458	0.69	0.37	1,989	0.53	0.41
Compliance inspections (no = 0, yes = 1)	896	0.13	0.20	458	0.30	0.27	1,989	0.47	0.31
Owner/manager completed secondary school	896	0.42	0.41	458	0.52	0.39	1,989	0.72	0.37
Medium high-tech sector dummy	896	0.23	0.41	458	0.32	0.46	1,989	0.39	0.46

Notes: The time-series average of each variable is first calculated by firm, before the average group statistics are computed. *Although we are showing the number of employees in this table, we use the log in our regressions.

Source: Based on author's computations.

Table A4: Effects of formality on government (tax payment)

	Log tax (real 1,000 VND)			Tax dummy (1 if > 0)
Variables	OLS	GLS	MUNDLAK	LOGIT
Formal non-switcher	3.78***	3.98***	3.45***	4.66***
	(0.08)	(0.11)	(0.09)	(0.25)
Switcher (from informal to formal)	1.46***	1.68***	1.46***	1.32***
	(0.12)	(0.15)	(0.10)	(0.18)
Switcher (after formalization)	1.48***	1.30***	1.30***	2.31***
	(0.11)	(0.12)	(0.08)	(0.23)
Share of female employees	-0.03	0.04	0.25**	0.33
	(0.09)	(0.10)	(0.12)	(0.21)
Share of production workers	-0.92***	-0.54***	0.00	-0.16
	(0.10)	(0.10)	(0.10)	(0.20)
Firm size (log (1+employement))	1.85***	1.46***	0.71***	1.58***
	(0.10)	(0.11)	(0.13)	(0.25)
Firm size square (log (1+employment))	-0.13***	-0.09***	-0.05**	-0.24***
	(0.02)	(0.02)	(0.02)	(0.05)
Gender of owner/manager (female=0, male=1)	-0.12***	-0.11**	-0.06	-0.06
	(0.04)	(0.04)	(0.06)	(0.12)
Own land use right certificate (no=0, yes=1)	0.11***	0.12***	0.20***	0.26**
	(0.04)	(0.04)	(0.05)	(0.11)
Compliance inspections (no = 0, yes = 1)	0.07*	-0.01	-0.06	0.29**
	(0.04)	(0.04)	(0.04)	(0.12)
Owner/manager completed secondary school	0.29***	0.29***	0.15***	0.27**
	(0.04)	(0.05)	(0.06)	(0.11)
Medium high-tech sector dummy	0.17***	0.15***	0.11*	0.33**
	(0.04)	(0.05)	(0.06)	(0.15)
Constant	1.86***	2.13***	1.78***	-1.01***
	(0.14)	(0.17)	(0.20)	(0.39)
Observations	11,601	11,601	11,601	11,601
R-squared	0.66			
Time dummies included	Yes	Yes	Yes	Yes
Province dummies included	Yes	Yes	Yes	Yes
Number of panels		3,303	3,303	3,303

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

In Mundlak's model, the panel means of independent variables are included in the regression, except: Formal non-switcher, Switcher (from informal to formal), Switcher (after formalization).

Source: Based on author's computations

Table A5: Effects of formality on employees (wages)

	Log wages (real 1,000 VND)			Log wages per employee (real 1,000 VND)		
Variables	OLS	GLS	MUNDLAK	OLS	GLS	MUNDLAK
Formal non-switcher	2.00***	2.34***	1.58***	0.30***	0.38***	0.25***
	(0.11)	(0.15)	(0.12)	(0.03)	(0.03)	(0.03)
Switcher (from informal to formal)	1.54***	1.85***	1.33***	0.01	0.04	-0.01
	(0.13)	(0.17)	(0.13)	(0.03)	(0.04)	(0.03)
Switcher (after formalization)	0.63***	0.49***	0.42***	0.29***	0.32***	0.32***
	(0.13)	(0.12)	(0.10)	(0.03)	(0.03)	(0.03)
Share of female employees	-2.32***	-1.71***	-1.03***	-0.37***	-0.31***	-0.14***
	(0.10)	(0.13)	(0.14)	(0.03)	(0.03)	(0.04)
Share of production workers	1.29***	0.72***	0.44***	-0.25***	-0.27***	-0.17***
	(0.14)	(0.14)	(0.12)	(0.04)	(0.05)	(0.04)
Firm size (log (1+employement))	6.51***	5.81***	4.84***	0.53***	0.42***	0.02
	(0.16)	(0.19)	(0.15)	(0.04)	(0.05)	(0.05)
Firm size square (log (1+employment))	-0.78***	-0.68***	-0.58***	-0.06***	-0.05***	-0.02***
	(0.03)	(0.03)	(0.03)	(0.01)	(0.01)	(0.01)
Gender of owner/manager (female=0, male=1)	-0.10*	-0.07	-0.10	-0.02	-0.02	-0.01
	(0.05)	(0.05)	(0.06)	(0.01)	(0.01)	(0.02)
Own land use right certificate (no=0, yes=1)	-0.14***	-0.16***	-0.12**	-0.02*	-0.02*	0.01
	(0.05)	(0.05)	(0.06)	(0.01)	(0.01)	(0.02)
Compliance inspections (no = 0, yes = 1)	-0.07	-0.08*	-0.07	-0.11***	-0.10***	-0.08***
	(0.05)	(0.04)	(0.05)	(0.01)	(0.01)	(0.01)
Owner/manager completed secondary school	0.56***	0.43***	0.21***	0.15***	0.14***	0.04**
	(0.06)	(0.06)	(0.06)	(0.01)	(0.02)	(0.02)
Medium high-tech sector dummy	0.30***	0.36***	0.19***	-0.00	0.00	-0.02
	(0.05)	(0.07)	(0.07)	(0.01)	(0.02)	(0.02)
Constant	-2.33***	-1.41***	-3.25***	7.83***	7.95***	7.46***
	(0.19)	(0.25)	(0.26)	(0.06)	(0.07)	(0.07)
Observations	11,604	11,604	11,604	9,240	9,240	9,240
R-squared	0.69			0.25		
Time dummies included	Yes	Yes	Yes	Yes	Yes	Yes
Province dummies included	Yes	Yes	Yes	Yes	Yes	Yes
Number of panels		3,303	3,303		2,896	2,896

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

In Mundlak's model, the panel means of independent variables are included in the regression, except: Formal non-switcher, Switcher (from informal to formal), Switcher (after formalization).

Source: Based on author's computations.

Table A6: Effects of formality on employees (share of wage in value-added)

	Share of wages in value added			Log share of wages in value added		
Variables	OLS	GLS	MUNDLAK	OLS	GLS	MUNDLAK
Formal non-switcher	0.06***	0.06***	0.05***	0.05***	0.05***	0.04***
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
Switcher (from informal to formal)	0.06***	0.07***	0.05***	0.05***	0.06***	0.05***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Switcher (after formalization)	0.02***	0.01*	0.01	0.02***	0.01**	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Share of female employees	-0.08***	-0.07***	-0.06***	-0.07***	-0.06***	-0.05***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Share of production workers	0.14***	0.09***	0.04***	0.11***	0.07***	0.04***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Firm size (log (1+employement))	0.32***	0.30***	0.25***	0.25***	0.24***	0.20***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Firm size square (log (1+employment))	-0.04***	-0.04***	-0.03***	-0.03***	-0.03***	-0.02***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Gender of owner/manager (female=0, male=1)	0.00	0.00	-0.00	0.00	0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Own land use right certificate (no=0, yes=1)	-0.02***	-0.02***	-0.02***	-0.01***	-0.01***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Compliance inspections (no = 0, yes = 1)	-0.03***	-0.03***	-0.03***	-0.02***	-0.02***	-0.02***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Owner/manager completed secondary school	0.02***	0.01***	0.01*	0.01***	0.01***	0.01*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Medium high-tech sector dummy	0.02***	0.02***	0.02***	0.02***	0.02***	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	-0.19***	-0.15***	-0.26***	-0.14***	-0.11***	-0.19***
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
Observations	11,445	11,445	11,445	11,445	11,445	11,445
R-squared	0.44			0.46		
Time dummies included	Yes	Yes	Yes	Yes	Yes	Yes
Province dummies included	Yes	Yes	Yes	Yes	Yes	Yes
Number of panels		3,303	3,303		3,303	3,303

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

In Mundlak's model, the panel means of independent variables are included in the regression, except: Formal non-switcher, Switcher (from informal to formal), Switcher (after formalization).

Source: Based on author's computations.