Economic Systems xxx (xxxx) xxx

Contents lists available at ScienceDirect

Economic Systems

journal homepage: www.elsevier.com/locate/ecosys

Formalization, productivity, and hidden costs: Evidence from Vietnam

Le Phan[a], Hai Thanh Nguyen[b,*]

[a] Thanh Do University, Hanoi, Viet Nam
[b] Faculty of Business Administration, Ton Duc Thang University, Ho Chi Minh City, Viet Nam

ARTICLE INFO

Keywords:
Firm behavior
Formalization
Productivity
Hidden cost
Vietnam

ABSTRACT

This paper examines the micro-level benefits and (hidden) costs of the transition of Vietnam's informal household businesses into the formal sector during the period 2007–2015. On the benefit side, the paper finds that such a transition, or "formalization," leads to higher investment, greater capital stock, and a lasting increase in labor productivity, which ranges between 23 and 69 percent. There is no statistically significant increase in total factor productivity, indicating that the gain in labor productivity comes from capital deepening rather than genuine innovation. On the cost side, the paper finds evidence of a lasting *visibility* effect, meaning that household firms have to pay higher bribes and spend more time dealing with government red tape after formal registration. JEL codes: D21, E26, L25, 017

1. Introduction

Informality is a prevalent feature in most, if not all, developing countries. While definitions vary, the informal sector essentially comprises micro-firms, often household-based, whose business activities are unregulated by national laws on enterprises (Boyd, 2017). Most firms in emerging economies operate in the informal sector. In Vietnam, for instance, the number of informal household firms in 2019 was nearly eight times higher than the number of domestic private firms, and the former's contribution to GDP was more than three times greater than the latter's (Table 1).

At the macro-level, countries with larger informal sectors tend to have lower income per capita, more primitive financial systems, lower investment, and higher inequality (World Bank, 2022). At the micro-level, informal firms are frequently associated with low productivity, credit constraints, tax evasion, and a lack of social protection for their employees (Rand and Torm, 2012). For these reasons, most developing countries encourage informal firms to transition into the formal sector, a process known as "formalization."

The literature on informality to date has focused predominantly on its determinants and macro-level effects rather than on micro-level impacts, largely due to the scarcity of longitudinal data covering both formal and informal firms. Among the few studies that discuss the micro-level impacts of formalization, Farrell (2004) found that informality had a negative effect on firm-level productivity for a number of selected countries. Fajnzylber et al. (2009) concluded that formalization brought about a 20-percent increase in the profitability of Mexican firms. Rand and Torm (2012), who conducted the closest study to this paper, showed that formalization in Vietnam led to higher profitability and investment, along with a decrease in the use of casual labor. However, their study focused

* Corresponding author.
E-mail addresses: lphan@thanhdouni.edu.vn (L. Phan), nguyenhaithanh@tdtu.edu.vn (H.T. Nguyen).

https://doi.org/10.1016/j.ecosys.2025.101311
Received 3 April 2024; Received in revised form 18 April 2025; Accepted 22 April 2025

Please cite this article as: L. Phan and H.T. Nguyen, Formalization, productivity, and hidden costs: Evidence from Vietnam, Economic Systems, https://doi.org/10.1016/j.ecosys.2025.101311

Table 1
Number of firms and GDP contribution by economic sector, 2019.

Economic sector	Number of firms	Contribution to GDP (%)
State	2260	27.06
Collective	14,388	3.63
Domestic private	647,632	9.68
Household (*informal*)	5200,479	29.37
Foreign investment	18,762	20.35

Source: Vietnamese National Statistics Yearbook 2019, General Statistics Office of Vietnam, GSO (2019).

exclusively on the benefit side of formalization and used only two years of panel data from 2007 to 2009. This made it difficult to capture the micro-level impacts of formalization, since such impacts often take time to materialize.

Using the biennial surveys of small and medium-sized enterprises (SMEs) in Vietnam from 2007 to 2015, this paper is, to the best of the authors' knowledge, the first empirical study to shed light on both the micro-level benefits and costs of formalization. Specifically, this paper examines two research questions: (i) Does formalization bring about higher productivity for formalized firms? and (ii) Do informal firms incur higher hidden costs after formalization? Regarding (i), this paper finds that formalization leads to higher investment, greater capital stock, and an increase in labor productivity, which ranges between 23 and 69 percent. There is, however, no statistically significant impact on total factor productivity (TFP), meaning that the gain in labor productivity comes from capital deepening rather than genuine innovation. This finding is in line with McCaig and Nanowski (2019) and the stylized facts highlighted in Lay and Tafese, (2020). Regarding (ii), this paper finds that formal firms have to spend more time dealing with government red tape and have to pay higher bribes to officials than informal firms. Together, these findings indicate that while formalized firms are subject to higher (hidden) costs, the overall result of formalization is a net benefit in terms of labor productivity.

The rest of this paper is structured as follows. Section 2 discusses the theories of change underlying the impact of formalization on productivity and hidden costs. Section 3 reviews the background of policies on the informal sector in Vietnam. Section 4 introduces the models used to verify the relationship between formalization and productivity on the one hand and formalization and informal payments on the other. Section 5 describes the SME survey dataset and the selection of variables, including productivity, formality, and hidden costs. Section 6 discusses the empirical results and robustness checks. The last section summarizes the paper's key findings and contributions, and discusses areas for future research.

2. Theories of change

2.1. Formalization and productivity

There are several competing theories that seek to explain the widely observed productivity gap between informal and formal firms. The optimistic theory, proposed in de Soto (1989), considers informal firms to be potentially as productive as formal firms, but constrained by numerous barriers to official status: onerous government regulations, bureaucracy, and their inability to secure property rights and to access formal loans. The theory implies that once these barriers are removed, informal firms would register formally, which would subsequently improve their productivity and lead to their expansion.

The survival theory, on the other hand, proposes that informal work is a way for people with limited skills to make a living in economies with few job opportunities. These individuals would prefer formal jobs but are often rejected due to their lower productivity (Fields, 2004). Informal businesses are thus their means of survival, and it is not in the best interest of these businesses to formalize.

The *rational choice* theory considers formalization to be a conscious investment decision; that is, informal firms will not formalize if the expected costs of formalization outweigh its expected benefits (de Mel et al., 2011; Maloney, 2004; Ulyssea, 2018). Maloney (2004) explains the formal-informal productivity gap by arguing that businesses that would benefit from formalization are already formal, while smaller and less productive businesses choose to remain informal because they see little advantage in formalization. Ulyssea (2018) argues that informal businesses choose to be informal because they gain a significant cost advantage by avoiding taxes, regulations, and the pressure to make informal payments. Staying small helps reduce the chances of these businesses being discovered.

Table 2 lists the potential channels through which formalization can impact firm-level productivity. Formal businesses are often perceived as more legitimate and trustworthy by customers, suppliers, and partners, which can enable a business to improve its brand position and increase its unit value-added. Better access to credit leads to capital deepening and higher labor productivity or increased total factor productivity if it results in greater investment in R&D and skill upgrading. Government support programs and policies can help improve efficiency and total factor productivity.

On the other hand, formalization may result in formal and hidden costs that adversely affect firm-level productivity. Formal costs include initial registration fees and ongoing costs such as tax burdens and the costs of compliance associated with increased labor and administrative regulations. Hidden costs encompass the time spent dealing with red tape as well as bribery payments. These require formalized businesses to divert resources away from productive activities, which adversely affects their productivity.

The existence of both positive and negative impact channels means that the overall effect of formalization on productivity is ambiguous and varies across different country contexts. In the case of Vietnam, while several studies have found evidence of the

Table 2
Potential impacts of formalization on firm-level productivity.

Positive impact channels	Negative impact channels
Legitimacy and trust	Registration fees
Better access to credit	Increased tax burdens and filing costs
Government support programs and policies	Increased labor and administrative regulations, e.g., formal accounting procedures, formal registration of employees, and minimum wages
Eligibility to expand to multiple locations and secure property rights	Time spent dealing with red tape
Ability to increase sales through serving a new customer segment that requires tax receipts	Bribery payments

Source: Authors' compilation based on Bruhn and McKenzie (2018).

positive channels (Demenet et al., 2016; Rand and Torm, 2012), evidence on the net effect of formalization on productivity is unclear. Lay and Tafese, (2020) use descriptive statistics to argue that the average TFP of formalized firms is actually lower than that of those that remain informal in the period 2013–2015. This shows that some productive informal firms decide to remain informal, which is in line with the rational choice theory. Using three waves of the Vietnam Household Living Standards Survey, McCaig and Nanowski (2019) find that formalization does not have a significant impact on productivity after controlling for pre-formalization trends.

2.2. Formalization and hidden costs

The literature highlights two opposing potential effects of formalization on hidden costs. The bribes-to-hide effect predisposes some businesses to operate informally to avoid paying taxes. In order to maintain their informal status, they might bribe government officials. This practice will continue as long as the benefits of avoiding taxes outweigh the costs of these bribes (Rand and Torm, 2012; Svensson, 2003). By contrast, the visibility effect means that larger formal firms are more easily detected by corrupt officials and hence have to pay more bribes than informal firms. Which effect dominates is an empirical issue.

Previous studies on this topic in Vietnam seem to suggest that the visibility effect predominates. Rand and Torm (2012) use two rounds of the Vietnamese SME surveys from 2005 and 2007 to show a positive correlation between formalization and bribery payments. However, their study does not control for the effect of selection into formality. Using the SME surveys and the Provincial Competitiveness Index (PCI) data, Le et al. (2020) argue that corrupt officials actually welcome formalization because it exposes firms to more rigorous auditing and regulations, thereby allowing them to see the true size of businesses. Once corrupt officials have greater knowledge of firm performance, they use this information to extract bribes, in line with the visibility effect.

3. Background of policies on the informal sector in Vietnam

The concept of "informality" was introduced by the International Labour Organization (ILO) in 1972 (ILO, 1993). Prior to that, the dominant development discourse in the 1950s and '60 s focused on capital formation, infrastructure investment, and export promotion while leaving the labor market to self-regulate (Bangasser, 2000). The rationale was that, as capital investment accelerated economic growth, people working in "informal" activities would be absorbed into the "formal, modern" sectors of the economy. Informal employment was thus viewed as a residual and temporary problem that would disappear with the onset of economic expansion.

However, not only did this "residual" and "temporary" problem not disappear, it in fact grew larger and more visible throughout the 1960s in most developing countries. Demographic trends, together with massive urban migration, resulted in an ever-increasing number of people entering the urban labor market, outpacing the generation of formal and modern sector jobs available in developing countries. It was not until 1972 that the ILO finally brought the topic of informality to attention in development discourse. The informal sector, according to the ILO, consists of household businesses that are not considered separate legal entities from their owners, do not follow formal bookkeeping practices, and are insufficiently covered by formal arrangements, for example, national laws on enterprises.

This definition places all household businesses in Vietnam in the informal sector. Although recognized as autonomous economic units and as constituting an important part of the economy, household businesses have never been considered a formal enterprise form under the country's national laws on enterprises. From the government's Decree 88/2006 to the latest Decree 01/2021, household businesses have been regulated separately from other forms of businesses and include firms with the following characteristics: they are established by an individual who is a Vietnamese citizen or by a group or a household; they can register the business at one location only; they may use not more than ten employees; they do not have the status of a legal entity; and they are fully liable with all of their assets for their business activities.

As the Vietnamese economy developed and restrictions on household business activities were gradually relaxed, the number of household businesses rose dramatically, from 0.33 million in 1989–1.5 million in 1999, and 4.75 million in 2015. In 2019, they accounted for 29.37 percent of the country's GDP, three times that of formal domestic private enterprises and even surpassing the contributions of the state and foreign investment sectors. Household businesses are one of the driving forces that are promoting entrepreneurship and developing Vietnam's market economy. However, as small and fragmented businesses without the status of

legal entities, household businesses face restrictions in terms of their access to formal credit, legitimacy, ability to operate in multiple visible, permanent locations, and access to business training programs.

Given the importance of household businesses to the economy, since the early 2000s, the Vietnamese government has tried to incentivize these informal firms to convert into formal enterprises. The incentives offered include: (i) Free consultation and guidance on documents and procedures for enterprise establishment; (ii) Fee exemptions for enterprise registration and initial enterprise information provision; free evaluation and exemptions from fees and charges for the first business license for conditional business lines; license fee exemptions for three years from the date of issuance of the first enterprise registration certificate; (iii) Exemption from or reduction of corporate income tax for a specified period in accordance with the Corporate Income Tax Law; and (iv) Exemption from or reduction of land use fees for a specified period in accordance with the Land Law.

Yet, according to a recent survey by the General Department of Taxation, most household businesses remained reluctant to convert into formal enterprises due to the perceived high costs of conversion, which include: following formal bookkeeping practices; preparing financial statements; complying with formal tax procedures; spending time on complicated administrative procedures related to insurance and occupational and fire safety, or having to pay bribes to avoid being hassled by corrupt officials. In this context, this paper seeks to explore whether the benefits of formalization outweigh its perceived costs.

4. Empirical frameworks

To address the research questions of whether formalization leads to higher productivity and/or higher hidden costs for formalized firms, this paper uses the following baseline model:

$$Outcome_{it} = \alpha + \beta Formal_{it-1} + \omega \boldsymbol{X_{it}} + \delta_i + \rho + \epsilon_{it} \quad (1)$$

where $Outcome_{it}$ can indicate labor productivity, total factor productivity, time spent dealing with government red tape, or bribery payments of firm i at time t. The first two dependent variables represent different measures of productivity and contribute to answering the first research question. While easier to measure, labor productivity changes capture both innovations in firm efficiency and variation in capital deepening. For this reason, TFP, which takes account of changes in both labor and capital, is a better indicator of efficiency improvements. By analyzing both measures, this paper seeks to investigate the channels through which formalization in Vietnam affects firm-level productivity; that is, whether it takes effect through innovation or capital deepening.

The latter two dependent variables measure different types of hidden costs. One variable is the amount of time firms have to spend dealing with government red tape – a reason often cited in the Vietnamese media for the low rate of formalization. Based on qualitative evidence, formality status is associated with frequent government inspections and complex accounting/bureaucratic procedures. The additional time that formal firms have to spend dealing with government officials and regulations represents a hidden cost that may deter household businesses from formalizing.

The second measure of hidden costs is bribery payments to government officials. Corruption remains a serious issue in Vietnam, as can be seen from the country's consistently low ranking in the Corruption Perceptions Index throughout the period 2007–2015 (see Fig. 1). If it were true that formalization leads to more frequent government inspections – in line with the *visibility* effect, this might result in formal firms having to pay higher bribes than informal ones. On the other hand, some household businesses may decide to bribe government officials to evade paying taxes, as the accounting procedures for household businesses tend to be opaque (the bribes-to-hide effect). Whether formalization leads to higher bribery payments remains an open question for empirical investigation.

Source: Authors' compilation based on data from Transparency International (2024). *A higher ranking indicates more severe corruption.

The explanatory variable of interest is a binary variable indicating the formality status of firms. This indicator is 1 if the firm is formal; otherwise, it is 0. Some firms in the sample remain informal, while others formalize during different time periods (see

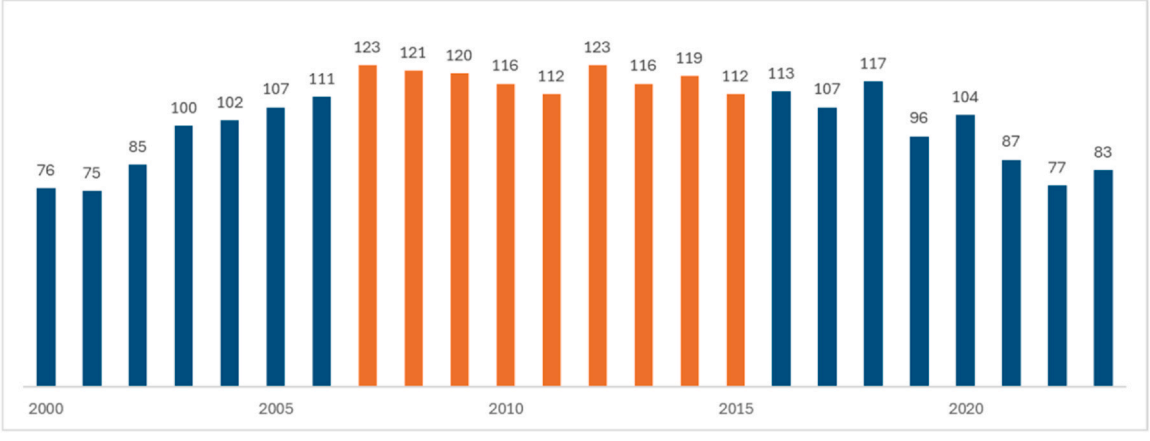

Fig. 1. Vietnam's ranking in the Corruption Perceptions Index.

Table 4). The formality status variable is included in the baseline models in lag one, as productivity changes often take time to materialize; this also alleviates the potential issue of reverse causality.

Regarding the vector of covariates $\boldsymbol{X_{it}}$, an important control variable is firm size. Since larger firms tend to be more productive than smaller ones, the coefficient of firm size is expected to be positive when productivity is the outcome variable. Similarly, larger firms are more likely to attract attention from government inspectors than smaller ones and thus may have to spend more time dealing with government red tape and pay higher amounts of bribes. This means that the coefficient of firm size is expected to also be positive when hidden costs are the outcome variable.

Other important firm-specific variables concern the owners' characteristics, such as gender, education, and business networks, which may affect both firm-level productivity and the decision to formalize. Rand and Tarp (2011), for example, showed that female owners tend to provide more generous non-wage benefits, which can affect their employees' productivity. Babbitt et al. (2015) found that female entrepreneurs in rural Indonesia are 20 percent more likely to join the formal sector than their male counterparts. Regarding the education of the owners, La Porta and Shleifer (2008) affirmed that formal firms are, on average, run by much better-educated owners than informal ones. Jaramillo (2009) found a positive association between the owner's level of education and the likelihood of formalization. In relation to business networks, Sheng et al. (2011) conducted a survey of Chinese firms and found that business ties have a positive impact on firm performance. More recently, Chien et al. (2019) demonstrated a positive association between firm performance and participation in research and development, commercial, or shared-director business networks.

Aside from the owners' characteristics, other firm-specific covariates include the workforce skill level, access to infrastructure, possession of property rights, and/or the intensity of inspections. The importance of a skilled workforce to firm performance has been emphasized in numerous previous studies (Atkinson and Mayo, 2010; Peri et al., 2015; Siepel et al., 2021). Similarly, infrastructure access can serve as a determinant of productivity due to its impacts on production techniques and supply chain management (Tybout, 2000).

Another important covariate is *property rights*, represented by the acquisition of a land use rights certificate. This is a peculiar feature of a socialist economy whereby land is considered to be collectively owned by the people, and thus individuals or businesses can acquire merely a certificate that grants them the right to use the land. Possession of such a certificate facilitates credit access, which can, in turn, affect investment and labor productivity. Malesky and Taussig (2009) also found that property rights have a stronger influence on formalization in Vietnam compared to other types of legal institutions.

Finally, the intensity of inspections, which indicates the frequency of government investigations into firm activities, may be negatively associated with the decision to formalize. Household firms may also decide to pay higher bribes in order to reduce the likelihood of being inspected.

In addition, the baseline models control for year, province, and two-digit sector fixed effects, represented by the vector ρ, to alleviate the issue of omitted variable bias. Nguyen et al. (2007) and Malesky and Taussig (2009) found that economic governance in Vietnam varied significantly across provinces and served as a determinant of formalization. The sector fixed effect, on the other hand, captures inherent sectoral differences, such as technological intensity, which are likely to affect firm-level productivity.

However, the baseline models do not adequately address the issue of self-selection and omitted variable bias. If any permanent differences exist in the outcomes between formal and informal firms, the estimated effect of formalization will be biased. Second, the baseline models fail to account for unobserved time-variant factors that may simultaneously affect the decision to formalize as well as productivity or the amount of hidden costs that firms have to incur. For example, Siepel et al. (2021) showed that it is a combination of various skills, such as a university degree in a science, technology, engineering, or mathematics (STEM) subject, creativity, or management, that determines firm performance; a variable that measures the share of trained workers would therefore fail to capture such nuances. In addition, the firm owners'/managers' business aptitude or experiences, which may influence their decision to formalize as well as firm performance, are not fully reflected in the observed variables such as educational attainment.

To address the above endogeneity issues, this paper employs two different empirical approaches: (i) the matched difference-in-difference (DiD); and (ii) the instrumental variable (IV) approaches. The main approach, matched DiD, resolves the issue of permanent average differences in outcomes between formal and informal firms that exist prior to formalization. Following the strategy outlined in Dettmann et al. (2020) (DGW), treated and control groups are matched one period prior to the (different) years of formalization based on combined statistical distance functions. These statistical distance functions incorporate the mean absolute differences for selected continuous variables and the matching coefficients for selected categorical variables, and have been shown to be superior to other distance measures such as propensity score matching or Mahalanobis distance (Dettmann et al., 2011). Using statistical distances, each treated unit is matched with its corresponding control(s) based on either nearest neighbor matching or radius matching. The former matches each treated unit with its closest control in terms of statistical distances, while the latter matches treated units with controls that fall within the predefined neighborhood of the treated units' statistical distances.[1] Once the matching process is complete, the mean differences in outcomes (productivity, hidden costs) between formalized firms and their corresponding controls are used to estimate the average treatment effect on the treated (ATT):

$$ATT = \frac{1}{N}\sum_{i=1}^{N}\left[\left(Y_{i,t_{0i+\delta_i}} - Y_{i,t_{0i}}\right) - \left(Y_{j,t_{0i+\delta_i}} - Y_{j,t_{0i}}\right)\right] \tag{1}$$

[1] Each approach has its advantages and disadvantages. Nearest neighbor matching is vulnerable to imprecise matching in cases where the nearest neighbors are far away from the treated units. For radius matching, the setting of the predefined radius can be quite arbitrary.

In Eq. (1), the date of formalization is denoted by t_{0i}, allowing for varying treatment time, while $t_{0i+\delta_i}$ indicates the individual duration from the time of formalization to outcome observation. Outcome differences for the treated firms i, $Y_{i,t_{0i+\delta_i}} - Y_{i,t_{0i}}$, are benchmarked against those for the respective controls j, $Y_{j,t_{0i+\delta_i}} - Y_{j,t_{0i}}$. The ATT is then calculated as the average of the individual comparisons.

The DGW estimator has been shown to be asymptotically unbiased under four conditions (Baker et al., 2022). First, the common support condition must be satisfied, as with all matching procedures. Second, there must be no spillover effect from the treated group to the control group at the time of matching. Third, treatment cannot be reversed; that is, a treated unit must remain treated for all the remaining time periods. Fourth, the conditional parallel trend assumption must be satisfied; that is, unobserved individual factors must be time-invariant for units with the same observed characteristics.

To meet the first condition, this paper conducts multiple tests for the matching procedures, such as the covariate imbalance test proposed by Leuven and Sianesi (2003), the nonparametric Kolmogorov–Smirnov tests, which compare the cumulative distributions between treated and control groups for *continuous* variables, and the chi-squared tests for *categorical* variables. To meet the third condition, firms that change their formality status more than once are removed from the final sample. In addition, this paper follows the best practices proposed by Baker et al. (2022), such as reporting the treatment timing of the formality status indicator and excluding post-treatment control variables when performing the matched DiD estimation.

For robustness checks, this paper uses another empirical approach, IV, to address the endogeneity issue of the selection on unobservables. Specifically, a firm's formality status is instrumented by the share of formal firms within the same year, province, and two-digit sector, excluding the firm of interest. This instrument is employed for the following reasons: First, if competitors in the same province decide to formalize, this is likely to influence the firm of interest's decision to formalize. The Staiger–Stock F-test, which examines partial correlations in first-stage regressions, can be used to validate the relevance of the selected instrument (Staiger and Stock, 1997). Second, since the selected instrument excludes the firm of interest and the owner's business networks have been controlled for, unobserved time-variant factors such as skill composition or business experience are unlikely to correlate with the instrument. In addition, to take into account economy-of-localization and economy-of-urbanization effects, this paper also instruments formality status with the proportion of formal firms in the same sector but outside the province in which the firm of interest is located. The results for the second IV are reported in Appendix 5.

5. Data and variable selection

The main dataset used in this paper is the biennial SME surveys for the period 2007–2015. This unique panel dataset contains information on formality status, firm performance, time spent dealing with government red tape, and bribery payments, which are rarely found in the literature.

This paper uses five rounds of SME surveys from 2007, 2009, 2011, 2013, and 2015. Each round covers approximately 2500 firms in ten cities/provinces, namely: Hanoi, Ha Tay, Haiphong, Phu Tho (Northern region); Nghe An, Khanh Hoa, Quang Nam (Central region); Lam Dong (Central Highlands region); Long An and Ho Chi Minh City (Southern region). The provinces are selected so as to ensure that (i) the surveys cover different geographical regions in Vietnam; and (ii) firms from both major urban cities and rural areas are included.

Regarding firm selection, the SME surveys cover micro-, small, and medium-sized firms: micro-firms have between 1 and 10 employees; small firms between 11 and 50 employees; and medium-sized firms between 51 and 300 employees. This categorization broadly follows the World Bank's definition of an SME and differs from the formal definition of SMEs stipulated in the Vietnamese government's Decree No. 90/2001 on supporting the development of small- and medium-sized enterprises.[2] Further, the surveys sample non-state[3] manufacturing firms based on ownership form, which includes household firms, private firms, collectives, partnerships, limited liability firms, and joint stock firms. Joint ventures between multinational enterprises (MNEs) and domestic firms are excluded due to the opaque involvement of the state or foreign investors in these firms' ownership structures. This exclusion is of little concern to this paper, as its focus is on household firms. Household firms in the SME surveys are a subsample of the quinquennial Economic Census conducted by the Vietnamese General Statistics Office (GSO), which gathers information on all registered household firms.

One caveat here is that, for each city/province, the SME surveys were confined to only those districts covered in the annual Vietnam Enterprise Surveys (VES). Both surveys are implemented by the GSO, with the latter covering only formal firms that have fixed professional premises. This sampling strategy means that household firms in the SME surveys operate in areas with many formal firms and are likely to be more competitive than other (non-surveyed) household firms that are clustered in areas with little or no presence of formal businesses.

To examine the effect of formalization, this paper uses a subset of firms that were household businesses in the base year 2007, as all informal firms in the SME surveys belong to the household sector. Firms that formalized during the period 2009–2015 form the treatment group, while informal firms that remain informal constitute the control group. In addition, unlike Rand and Torm (2012),

[2] According to Decree No. 90/2001, micro firms have between 1 and 10 employees, small firms between 11 and 100 employees, and medium-sized firms between 101 and 200 employees.

[3] This paper uses the term "non-state," rather than domestic private firms, to follow the GSO's official terminology and because the category non-state firms includes informal household firms in addition to formal private firms. Furthermore, the non-state firm types include a group of firms called "private firms"; the term "non-state" is therefore used to avoid confusion with this type of firm.

Table 3
Number of firms by formality status in the sample, 2007–2015.

	2007	2009	2011	2013	2015
Informal	1669	1323	1050	899	748
Formal	0	41	55	59	65
Total obs.	1669	1364	1105	958	813

Source: Authors' compilation.

who treat *registered*[4] household firms as formal firms, this paper considers *all* household firms as informal for two reasons. First, Vietnam's national laws on enterprises have never recognized household businesses as a *formal* enterprise form. Thus, the regulations of the Enterprise Law, to which formal firms are subject, such as accounting procedures, do not apply to household firms, regardless of their registration status. Second, households can use their land use rights certificates as collateral to obtain loans without the need to register; the advantages of registration for household businesses are thus unclear.

The data are then cleaned to remove duplicate firm-year observations and conflicting values for the selected variables; for example, negative value-added, revenues, employees, or productivity. A small number of firms that change their formality status more than once are also removed from the dataset. As shown in Table 3, the final dataset contains 5909 firm-year observations, with 220 observations for formal firms (the treatment group) and 5689 observations for informal firms (the control group).

The outcome variables of interest are productivity and hidden costs (time spent dealing with red tape and bribery payments). Productivity measures include both labor productivity and TFP. The former is calculated as real value-added divided by the number of employees. Gross value-added (GVA) is proxied by the sum of gross wages, pre-tax profits, and indirect taxes. To obtain real values of GVA, its nominal values are deflated using two-digit industry-specific producer price indexes with the base year 2010. The number of employees is measured as the year-end number of full-time production workers. Table 4

Table 4
Summary statistics.

Initially informal (household in 2007)	Formal – 2009		Formal – 2011		Formal – 2013		Formal – 2015	
	Yes	No	Yes	No	Yes	No	Yes	No
Value-added (log real VND million)	12.4	11.1***	12.5	11.1***	13.2	11.9***	13.3	11.8***
Labor productivity (log VND million/person)	10.1	9.7*	10.1	9.8	11.0	10.7***	11.2	10.6***
Log TFP	7.8	7.9	7.7	7.7	8.6	8.6	8.8	8.5***
Capital stock (VND million)	2772	1116***	4320	2171***	2843	1572***	3251	1433**
Investment (VND million)	49	160***	762	159***	659	153***	570	105***
Firm size (no. of full-time employees)	13	5***	15	4***	11	4***	11	4***
Owner's gender (male = 1, female = 0)	0.6	0.7	0.7	0.7	0.6	0.7	0.5	0.6
Owner's education (unskilled = 0, vocational level = 1; college or above = 2)	1.1	0.7***	1.3	0.9***	1.3	1.0***	1.3	1.0***
Owner's business networks (no. of business/govt contacts)	40	36	65	26**	45	30*	43	32**
Workforce skill level (share of unskilled production workers)	37 %	31 %	35 %	34 %	33 %	38 %	50 %	51 %
Property rights (land use rights cert. = 1; otherwise = 0)	0.5	0.7	0.5	0.8	0.7	0.8	0.8	0.9
Road/port/rail access (none = 0; all = 3)	2.1	1.6***	1.8	1.4***	2.1	1.7***	2.1	1.7***
High-tech sector (yes = 1; no = 0)	0.17	0.12	0.25	0.1***	0.25	0.1***	0.25	0.1***
Time costs of red tape (pct. of mgmt time)	1.7	0.8***	3.5	2.2***	3.4	2.1***	2.2	1.3***
Bribery payments (VND thousand)	7682	2921**	7787	2397**	6583	2298***	2571	2838

Notes: Mean estimates, by year and formality status. *T*-tests by formality status and with unequal-variance assumption are implemented for each year. *, **, *** indicate that the differences between formal firms and informal firms are significant at the 10-, 5-, and 1-percent levels, respectively.

[4] "Registered" means having a business registration certificate and a tax code.

Since changes in labor productivity reflect not only innovations in production methods but also variations in capital deepening, a more precise measure of productivity is TFP. In this paper, TFP is calculated using the method proposed by Ackerberg et al. (2015) (ACF). The ACF method yields nonlinear, robust generalized method-of-moments estimates that do not suffer from the functional dependence problems identified in the Olley and Pakes (1996) and Levinsohn and Petrin (2003) approaches (Ackerberg et al., 2015).

Table 4 shows that there are statistically significant differences in value-added, firm size, and labor productivity between formal and informal firms during the period 2009–2015. On average, household firms that formalize have higher value-added and are of a larger size than those that remain informal. The former's labor productivity is also significantly higher than the latter's for nearly all the sampled years. However, there seems to be no significant difference in TFP between formal and informal firms, except for the final year, 2015. Since the formalized household firms in the sample are still mostly small firms, the lack of evidence of significant technological upgrading after formalization should come as no surprise.

To examine the hypothesis that differences in labor productivity between formal and informal firms in Vietnam result from capital deepening rather than efficiency improvements, this paper examines two outcome variables related to capital deepening, namely firm-level capital stock and the amount of investment. Capital stock is measured in this study as fixed reproducible tangible assets (buildings, equipment, and machinery), thus excluding assets such as land, inventories, and financial assets. Investment in the SME surveys is reported as total investment and investment by subcategories; for example, land, buildings, equipment, R&D, human capital upgrading, or patents. For the purpose of this study, the total amount of investment is used. For both variables, it is clear from Table 4 that the average figures for formal firms are significantly higher than those for informal ones in all years from 2009 to 2015. This suggests that easier access to capital may serve as an important driver of formalization in Vietnam.

The two measures of hidden costs are the amount of time spent dealing with red tape and bribery payments. The former is based on the question: "What percentage of the management's working time each month is spent dealing with government regulations and officials (including taxes, permits, licenses, and business and trade regulations)?" while the latter is derived from the question: "Approximately how much did you pay in informal fees in total in a year?"[5]; For both types of hidden costs, formal firms incur a significantly higher burden on average than informal ones (see Table 4). For example, formal firms have to pay nearly three times more in bribery payments than informal firms. This can act as a barrier that discourages informal firms from formalization.

Since the decision to formalize is made by household business owners, it is important to control for the owners' characteristics. The owner's gender is a dummy variable equal to 1 if the owner is male and 0 otherwise. The owner's education level is a categorical variable equal to 0 if the owner is an unskilled worker (having no vocational training or higher educational degree); 1 if the owner has vocational training; and 2 if the owner has a university-level education and above. The owner's business networks are derived from the question: "Of the following groups, with approximately how many people do you currently have regular contact: businesspeople, bank officials, politicians, and civil servants?" Regular contact is defined as occurring at least once every three months and as being useful to the owners for their business operations.

The two-sample mean-comparison tests for formal and informal groups of firms show that the owner's gender plays an insignificant role in the decision to formalize (see Table 4). By contrast, the owner's education level is significantly higher in formal firms compared to informal ones. This is in line with La Porta and Shleifer (2008) and Jaramillo (2009), who demonstrated a positive correlation between the owners' educational attainment and the decision to formalize. Further, owners of formal firms tend to have wider business networks than those of informal firms, which can facilitate their formalization process.

Since the skill composition of a firm's workforce is likely to affect firm-level productivity, this paper controls for the shares of unskilled production workers employed in the firms. Somewhat surprisingly, there seems to be no significant difference in terms of workforce skill level between formal and informal firms, suggesting that the decision to formalize in Vietnam is not taken to obtain access to better-skilled workers. This is in line with La Porta and Shleifer (2008) and Rand and Torm (2012), who also found no difference in workers' educational levels between formal and informal firms.

Another important control variable is the property rights held by the firms, measured as a dummy equal to 1 if the firm has a land use rights certificate and 0 otherwise. Firms with land use rights certificates have an advantage in obtaining loans compared to those that do not have well-established property rights. This, in turn, facilitates the former's capital-deepening process and can affect firm-level productivity, in particular labor productivity. In the case of Vietnam, there seems to be no significant difference in terms of property rights between formal and informal firms; in each group, on average more than half of the firms possess land use rights certificates.

The infrastructure variable is constructed based on whether the firm has easy access to: (i) a main road (yes = 1, no = 0); (ii) rail (yes = 1, no = 0); or (iii) a port (yes = 1, no = 0). The ease of access is based on the subjective perception of the firm's owner. The three answers are then added up to give a variable measuring the firm's infrastructure access, the value of which ranges from 0 to 3. Table 4 shows that formal firms have, on average, significantly better access to main roads, railways, and ports than informal ones – a factor that is likely to impact productivity levels, thus highlighting the importance of controlling for this variable.

Further, as different sectors have varying levels of technological intensity and government support, this paper includes a dummy variable equal to 1 if the firm belongs to a medium and high-tech (MHT) sector and 0 otherwise. A full set of two-digit sector dummies is not used due to the inconsistent reporting of operating sectors. For example, a firm that manufactures wooden furniture might report its sector as 16 (wood and products of wood) in one year and 31 (furniture) in another. Since both sectors belong to the low-technology group, using a high-tech dummy variable helps alleviate this inconsistency issue.

[5] In Vietnamese, "informal fees" is implicitly understood as referring to bribes. Since paying bribes is illegal, answers to this question should be treated with caution, as household firms may understate their informal payments due to fear of retribution.

Table 5
Formality and labor productivity – Baseline estimates.

Dep. var: **Labor productivity**	(1) OLS	(2) Dynamic RE	(3) FE
(Lagged) Formal (yes = 1; no = 0)	0.238**	0.265***	0.389***
	(2.29)	(3.06)	(2.98)
(Lagged) Productivity		0.0960***	
		(4.77)	
(Lagged) Firm size (log)	0.0501*	0.0506*	0.0260
	(1.82)	(1.69)	(0.54)
Owner's gender	−0.0466	−0.0427	−0.0544
(male = 1; female = 0)	(−1.11)	(−1.01)	(−0.76)
Owner's education	0.0624	0.0607	0.00984
(unskilled = 0; vocational level = 1; college or above = 2)	(1.40)	(1.40)	(0.17)
Owner's business networks	0.000194	0.000164	−0.000102
(no. of business/govt contacts)	(1.47)	(1.27)	(−0.85)
Workforce skill level	−0.0239	−0.0275	−0.0473
(share of unskilled prod. workers)	(−0.49)	(−0.56)	(−0.85)
Property rights	0.0692	0.0613	0.0504
(land use rights certificate = 1; otherwise = 0)	(1.26)	(1.09)	(0.63)
Road/port/rail access	0.0388**	0.0388**	0.0190
(none = 0; all = 3)	(2.17)	(2.14)	(0.73)
Year dummies	Yes	Yes	Yes
Province dummies	Yes	Yes	Yes
MHT sector dummy	Yes	Yes	Yes
Observations	4164	4173	4173
R^2	0.152	0.203	0.163

Notes: OLS, RE, and FE estimates. *t*-statistics (reported in parentheses) are heteroskedasticity-robust. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The classification of the MHT sector is adapted from the United Nations Industrial Development Organization (UNIDO; 2010) and based on the two-digit International Standard Industrial Classification (ISIC) Revision 4 (see Appendix 1). The share of formal firms in the MHT sector is about 25 percent, while the figure for informal firms is significantly lower at about 10 percent. This low share is understandable given the focus of this study on household firms.

Finally, the key explanatory variable of interest to this paper is an indicator equal to 1 if a firm belongs to the formal sector and 0 otherwise. Firms that are not household establishments are considered to be part of the formal sector, which encompasses the following ownership forms: (i) sole proprietorship; (ii) partnership; (iii) collective/cooperative; (iv) limited liability company; (v) joint-stock company with state capital; (vi) joint-stock company without state capital; (vii) joint venture with foreign capital; (viii) state enterprise (central); and (ix) state enterprise (local). However, the dataset does not include any household firms that transform into categories (vii)–(ix) during the period 2007–2015.

6. Results

6.1. Productivity

Table 5 shows the baseline estimates for the relationship between formalization and labor productivity. Following Aw et al. (2000), the key explanatory variable of interest, *formal,* is included with one lag because the potential impact on productivity often takes time to manifest. Column (1) presents the ordinary least squares (OLS) specification with labor productivity (in logs) as the outcome variable.

Column (1) presents the following control variables: lagged firm size (in log), owner's characteristics (gender, education, business networks), workforce skill level, possession of property rights, infrastructure access, plus a full set of dummies for year, location, and MHT sector. The rationale for controlling for firm size one period prior to formalization is that firms may expand after entering the formal sector.[6] Under this specification, labor productivity is estimated to increase by 23.8 percent after household firms enter the formal sector, all else equal. Additionally, in line with the existing literature, larger firms are found to be more productive than smaller ones. Based on Table 5, it is estimated that a 10-percent increase in the number of full-time employees corresponds to a 0.5-percent increase in labor productivity. A one-point rise in infrastructure access (values ranging from 0 to 3) also corresponds to a 3.8-percent increase in the same outcome variable.

Differing from column (1), column (2) introduces the dynamic random effect (RE) model to take advantage of the panel data structure and to capture autocorrelation that persists over time. The key findings from the previous OLS estimates remain robust, with labor productivity estimated to rise by 26.5 percent after formalization, all else equal. A 10-percent increase in workforce size is associated with a 0.5-percent gain in labor productivity, while a one-point increase in infrastructure access leads to a 3.9-percent rise

[6] Appendix 2 shows the results with firm size not in lag. The main findings remain robust.

Table 6
Formality and TFP – Baseline estimates.

Dep. var: **TFP**	(1) OLS	(2) Dynamic RE	(3) FE
(Lagged) Formal (yes = 1, no = 0)	0.0444	0.103	0.276**
	(0.43)	(1.10)	(2.00)
(Lagged) TFP		0.0115	
		(0.37)	
(Lagged) Firm size (log)	0.0110	0.00243	−0.01
	(0.33)	(0.07)	(−0.19)
Owner's gender	0.00306	0.0149	−0.0536
(male = 1; female = 0)	(0.06)	(0.29)	(−0.73)
Owner's education	0.0733	0.0853	0.1092
(unskilled = 0; vocational level = 1; college or above = 2)	(1.08)	(1.23)	(1.56)
Owner's business networks	−0.00665	0.000903	0.0001
(no. of business/govt contacts)	(−0.12)	(1.25)	(0.12)
Workforce skill level	0.00104	0.00830	−0.0293
(share of unskilled prod. workers)	(1.52)	(0.14)	(−0.47)
Property rights	0.00675	−0.0246	0.0205
(land use rights certificate = 1; otherwise = 0)	(0.11)	(−0.36)	(0.23)
Road/port/rail access	0.0183	0.0191	0.0052
(none = 0; all = 3)	(0.84)	(0.86)	(0.20)
Year dummies	Yes	Yes	Yes
Province dummies	Yes	Yes	Yes
MHT sector dummy	Yes		Yes
Observations	2339	2339	2339
R^2	0.112	0.078	0.157

Notes: OLS, RE, and FE estimates. *t*-statistics (reported in parentheses) are heteroskedasticity-robust. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

in the same outcome variable.

Finally, column (3) displays the fixed effects (FE) estimates of the impact of formalization on labor productivity. The lagged outcome variable is excluded in this specification due to its correlation with the demeaned error term, which is a violation of the exogeneity assumption of OLS regression. The specification consists of the full set of control variables, plus dummies for year, location, and MHT sector to control for unobservable time-invariant factors specific to the firm, province, sector, and year that may influence both the outcome variable and the key explanatory variable of interest. Similar to the previous estimates, the coefficient of formality status is positive and significant at the 1-percent level, further confirming the positive association between formalization and labor productivity. Specifically, entering the formal sector corresponds to a 38.9-percent increase in labor productivity.

Somewhat surprisingly, none of the covariates besides formality status is significant at the 10-percent level under the FE specification. Two reasons can be put forward for this phenomenon: First, FE estimates capture within-group variation only. Nevertheless, household establishments, even after formalization, mostly remain micro- and small firms. The effects of the other covariates on productivity may therefore take longer to materialize. Second, and more importantly, the FE specification is not ideal for addressing the impact of formalization on productivity. While productivity tends to be highly persistent, the FE specification does not allow for the inclusion of the lagged outcome variable. It is also vulnerable to omitted variable bias, as fixed effects cannot take into account the influence of unobserved time-varying factors. For this reason, it is important to complement the baseline estimates here with other approaches that can minimize the possible endogeneity biases and are able to detect a longer-duration impact on productivity.

Since labor productivity captures not only efficiency improvements but also capital deepening, this paper also examines the relationship between formalization and TFP – a more precise measure of productivity (see Table 6 below). Column (1) shows the OLS estimates, and column (2) the dynamic random effect estimates, while column (3) presents the FE specifications. The outcome variable is TFP (in logs). The key explanatory variable of interest is the firms' formality status (in lag one).

Similar to Table 5, column (1) presents the following control variables: lagged firm size (in log), the owner's characteristics (gender, education, business networks), workforce skill level, possession of property rights, infrastructure access, plus a full set of dummies for year, location, and MHT sector. In line with the mean-comparison tests, which show no significant difference in terms of TFP between formal and informal firms, the coefficient of the formality status variable in column (1) is not significant at the 10-percent level, which suggests that the benefits of formalization stem from capital deepening rather than genuine productivity improvements.

Column (2) shows the dynamic random effect estimates in order to capture productivity autocorrelation that persists over time. Similar to the OLS estimate, the coefficient of the key explanatory variable, *formal*, is not significant at the 10-percent level.

Column (3) presents the FE estimates with the full set of control variables as in column (1) plus fixed effects for year, location, and MHT sector. The lagged outcome variable is excluded from the FE specification due to its correlation with the demeaned error terms. Formalization is found to have a positive and significant correlation with TFP. Specifically, TFP rises by 27.6 percent when a household firm formalizes, all else equal. This result, however, should be taken with a grain of salt, since the FE estimates do not take into account the dynamic factor as the OLS models do; thus the significance of formality status may well merely reflect the self-selection of highly productive household firms into the formal sector.

Table 7
Formality and productivity – Matched DiD estimates.

	Formalization treatment effect			
	(1) Labor productivity		(2) TFP	
	ATT	Standard errors	ATT	Standard errors
One period after formalization	0.55**	0.25	0.24	0.35
Two periods after formalization	0.69**	0.31	0.30	0.51
Total obs.	1669		1669	
Treated obs.	100		100	

Notes: Average treatment effect on the treated (ATT) using nearest neighbor matching (one match per observation). Reported standard errors are consistent bias-corrected estimators as proposed in Abadie and Imbens (2011). The dependent variable is labor/total factor productivity one period after formalization (Row 1) and two periods after formalization (Row 2). Matching is performed one period prior to formalization. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

To alleviate the potential issues of reverse causality and omitted variable bias, this paper applies a matched DiD approach to control for observable determining factors that may influence both the decision to formalize and the resulting productivity performance (see Table 7 below). The treated and control groups are matched one period[7] prior to the (varying) time of formalization, with the formality status indicator as the outcome variable, while the list of covariates comprises both continuous variables (productivity, firm size, owner's business networks, workforce skill level) and categorical variables (owner's gender and education level, possession of property rights, infrastructure access, and dummies for province and MHT sector). Labor productivity is used as a control variable for matching when the outcome variable of interest is labor productivity, and TFP is used for the matching procedure when the outcome variable is TFP. This paper investigates the impact of formalization both one period afterward (Row 1 in Table 7) and two periods afterward (Row 2 in Table 7) to shed light on a possible lasting impact and also to ensure the robustness of key results. Observations are matched using bias-corrected nearest neighbor matching. Further, following Dettmann et al. (2011), this paper employs aggregated statistical distance functions for the matching procedure, as these functions have been shown to better summarize similarities in differently scaled variables in small samples compared to the often-used measures such as propensity score, index score, or Mahalanobis distance.

The comparability of the matched groups can be seen through multiple tests presented in Appendix 3. These include the covariate imbalance test proposed by Leuven and Sianesi (2003), the nonparametric Kolmogorov–Smirnov (K–S) tests, which compare the cumulative distributions between treated and control groups for *continuous* variables, and the chi-squared tests for *categorical* variables. The covariate imbalance test measures the standardized percentage difference, or bias, between the means of the treated group and the control group for each of the matching variables in order to examine whether both groups have equal means. In addition, the test sheds light on the similarity of the variances in the treated and control groups for each matching variable. The results in Appendix 3 show that both the means and variances of the matching variables are balanced, regardless of whether nearest neighbor matching or radius matching is employed. The balance of the matching variables is also confirmed by the K–S tests and chi-squared tests.

In Table 7, the dependent variable in column (1) is labor productivity (in logs), and in column (2), it is TFP (in logs). Row A displays impact results one period after formalization, and Row B two periods after formalization. For column (1), the ATT results are positive and significant at the 5-percent level both one period after formalization and two periods afterward. Together with the baseline estimates, this outcome further confirms the significantly positive impact of formalization on labor productivity. Moreover, the magnitude of impact increases two periods after formalization, thus showing that formalization brings about a lasting impact.[8]

In column (2), by contrast, the ATT of formalization is not significant at the 10-percent level when the dependent variable is the log of TFP. This holds true for both time periods, showing that the formalization of household firms does not actually lead to efficiency improvements even after four years of formalization. This finding is in line with Do and Vu (2021), who found that formalization has no significant effect on human capital investment.

6.2. Hidden costs

Table 8 presents the baseline estimates of the impact of formalization on the percentage of management time that firms have to spend dealing with government regulations. Column (1) shows the pooled OLS estimates, and column (2) the dynamic random effect estimates, while column (3) presents the FE specifications. The outcome variable is the time costs of red tape, measured as a percentage point. The key explanatory variable of interest is the firms' formality status. While productivity improvements may take time to materialize, household firms can face higher hidden costs immediately upon entering the formal sector. For this reason, the *formal* variable is not lagged in Table 8.

For the baseline OLS estimates, column (1) presents the following control variables: firm size, owner's characteristics (gender, education, business networks), workforce skill level, possession of property rights, infrastructure access, plus a full set of dummies for year, location, and MHT sector. Firm size is found to be significantly and positively correlated with the time costs of red tape. A 10-

[7] One period is equivalent to two years.
[8] The result is robust when radius matching is used instead of nearest neighbor matching (see Appendix 4).

Table 8
Formality and the time costs of red tape – Baseline estimates.

Dep. var: **Time costs of red tape**	(1) OLS	(2) Dynamic RE	(3) FE
Formal (yes = 1, no = 0)	0.802***	0.821***	0.0832
	(2.69)	(3.59)	(0.20)
(Lagged) Time costs		0.00390	
		(1.41)	
Firm size (log)	0.178***	0.173***	0.103
	(3.12)	(2.87)	(0.82)
Owner's gender	−0.0481	−0.0478	−0.0882
(male = 1; female = 0)	(−0.62)	(−0.62)	(−0.65)
Owner's education	0.147*	0.120	0.123
(unskilled = 0; vocational level = 1; college or above = 2)	(1.81)	(1.50)	(1.05)
Owner's business networks	0.000584	0.000534	0.000547*
(no. of business/govt contacts)	(1.17)	(1.18)	(1.69)
Workforce skill level	0.152*	0.117	0.209*
(share of unskilled prod. workers)	(1.70)	(1.31)	(1.93)
Property rights	−0.281***	−0.265***	−0.0950
(land use rights certificate = 1; otherwise = 0)	(−3.13)	(−2.83)	(−0.55)
Road/port/rail access	0.0310	0.0268	0.0117
(none = 0; all = 3)	(0.95)	(0.78)	(0.25)
Year dummies	Yes	Yes	Yes
Province dummies	Yes	Yes	Yes
MHT sector dummy	Yes	Yes	Yes
Observations	2871	2871	2871
R^2	0.108	0.123	0.106

Notes: OLS, RE, and FE estimates. *t*-statistics (reported in parentheses) are heteroskedasticity-robust. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

percent increase in workforce size corresponds to a 2-percentage point increase in the time costs of red tape, which is significant considering that the average time costs of red tape are around 3 percent.

Further, the coefficient of the formality status variable is positive and significant at the 1-percent level. Specifically, the formalization of firms is associated with an increase in the time spent dealing with red tape of 0.802 percentage points. Given that the average time costs of red tape are around 3 percent of total management time, an increase of 0.8 percentage points represents a significant amount.

Column (2) introduces the dynamic random effect model to take advantage of the panel data structure and to capture autocorrelation that persists over time. The key findings from the previous OLS estimates remain robust, with the time costs of red tape being estimated to rise by 0.821 percent after formalization, all else equal.

Column (3) presents the FE estimates with the full set of control variables as in column (1) plus fixed effects for year, location, and MHT sector. The lagged outcome variable is excluded from the FE specification due to its correlation with the demeaned error terms. In contrast to the OLS estimates, formalization under the FE estimates does not have a positive and significant correlation with the time costs of red tape. The loss of statistical significance for formalization under FE estimation suggests that formalized household businesses may already face higher time costs due to red tape prior to formalization than informal household businesses. Such high time costs continue after formalization; hence within-firm estimates show no significant effect of formalization on the time costs of red tape.

In addition to the time costs of red tape, another common type of hidden cost is bribery payments. According to the Provincial Competitiveness Index (PCI) survey in 2015, 66 percent of surveyed firms in Vietnam had to make some form of bribery payments to government officials, and 11 percent had to spend 10 percent of their revenues on such informal fees. Table 9 provides the baseline estimates of the impact of formalization on the amount of bribery payments (in VND thousand). Column (1) shows the pooled OLS estimates, and column (2) the dynamic random effect estimates, while column (3) presents the FE specifications. The outcome variable is bribery payments, while the key explanatory variable of interest is the firms' formality status.

Similar to Table 8, column (1) presents the following control variables: firm size, owner's characteristics (gender, education, business networks), workforce skill level, possession of property rights, infrastructure access, plus a full set of dummies for year, location, and MHT sector. The coefficient of firm size is positive and significant at the 1-percent level. The magnitude is also significant: a 10-percent increase in the number of employees corresponds to an additional VND 25 million (approximately USD 1100) in bribery payments. The explanatory variable of interest, formality status, is also found to be positively correlated with bribery payments. Specifically, firms have to pay an extra VND 1.4 million (USD 61) on average after formalization. This finding highlights the problem of petty corruption for household firms in Vietnam.

The dynamic random effect and FE estimates further confirm the findings from the pooled OLS regressions. In columns (2) and (3), the coefficients of both formality status and firm size are positive and significant. Household firms have to pay an extra VND 2.5 million–VND 2.7 million in informal fees after formalization. A 10-percent increase in firm size is also associated with an increase in bribery payments of VND 19 million–VND 21 million. It should be noted that, since paying bribes is illegal, firms are likely to understate the true amount of bribery payments. The results presented here can thus be considered a lower bound for the impact of

Table 9
Formality and bribery payments – Baseline estimates.

Dep. var: **Bribery payments**	(1) OLS	(2) Dynamic RE	(3) FE
Formal (yes = 1, no = 0)	1375.3*	2705.4**	2530.0**
	(1.83)	(2.16)	(2.17)
(Lagged) Bribery payments		−0.0124	
		(−0.15)	
Firm size (log)	2476.2***	2071.3***	1949.9***
	(3.28)	(2.90)	(3.31)
Owner's gender	−330.9	503.1	256.6
(male = 1; female = 0)	(−0.54)	(0.73)	(0.47)
Owner's education	−247.0	−513.8	−535.4
(unskilled = 0; vocational level = 1; college or above = 2)	(−0.39)	(−1.19)	(−1.24)
Owner's business networks	4.183	4.472*	4.924**
(no. of business/govt contacts)	(1.41)	(1.71)	(2.05)
Workforce skill level	−591.6	296.1	241.3
(share of unskilled prod. workers)	(−0.60)	(0.71)	(0.57)
Property rights	373.5	1065.2	957.5
(land use rights certificate = 1; otherwise = 0)	(0.77)	(1.16)	(1.07)
Road/port/rail access	197.6	601.9**	527.1**
(none = 0; all = 3)	(1.64)	(2.32)	(2.48)
Year dummies	Yes	Yes	Yes
Province dummies	Yes	Yes	Yes
High-tech sector dummy	Yes	Yes	Yes
Observations	1484	1484	1484
R^2	0.079	0.306	0.173

Notes: OLS, RE, and FE estimates. *t*-statistics (reported in parentheses) are heteroskedasticity-robust. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

formalization on bribery payments. Further, the presence of such bribery and its connection to formal status confirms the *visibility* effect, and despite not being large, it nevertheless sends a negative message about the formal business environment in Vietnam.

Table 10 presents the matched DiD estimates of the effect of formalization on hidden costs. Similar to Table 7, the matching procedure is performed one period prior to the time of formalization, with the outcome variable being the formality status indicator and the list of covariates including continuous variables (productivity, firm size, hidden costs, workforce skill level) and categorical variables (owner's education, possession of property rights, infrastructure access, and dummies for province and MHT sector). The time costs of red tape are selected as a matching covariate when examining the effect of formalization on the time costs of red tape, and bribery payments are used when the effect of formalization on informal fees is of interest. This paper investigates the impact of formalization both one period afterward (Row 1 in Table 10) and two periods afterward (Row 2 in Table 10) to shed light on a possible lasting effect. Following Dettmann et al. (2011), this paper employs aggregated statistical distance functions for the matching procedure.

The dependent variable in column (1) is the time costs of red tape, measured as a percentage of management time spent dealing with government regulations, and in column (2), it is bribery payments (in VND million). For column (1), the ATT result from nearest neighbor matching shows that formalized firms' owners need to spend an additional 0.85 percent (Row 1) and 1.25 percent (Row 2) of their time dealing with government regulations compared to their informal counterparts, which is in line with the anecdotal evidence of informal firms' hesitancy to join the formal sector due to complex bureaucratic procedures.

In column (2), the ATT of formalization confirms another hidden cost that formalized household firms have to bear, namely bribery payments. Specifically, formalization leads to an increase in the amount of bribery payments of VND 10.4 million one period after formalization and VND 12 million (approximately USD 500) two periods afterward. The increase in bribery payments is in line with the *visibility* theory and shows that formalization has a lasting effect on hidden costs.

Table 10
Formality and hidden costs – Matched DiD estimates.

	Formalization treatment effect			
	Time costs of red tape		Bribery payments (VND million)	
	ATT	Standard error	ATT	Standard error
One period after formalization	0.85*	0.5	10.4*	5.7
Two periods after formalization	1.25*	0.6	12*	5.4
Total obs.	1422		1569	
Treated obs.	96		100	

Notes: Average treatment effect on the treated (ATT) using nearest neighbor matching (multiple matches with the same statistical distance per observation). Reported standard errors are consistent bias-corrected estimators as proposed in Abadie and Imbens (2011). The dependent variables are the time costs of red tape and bribery payments one period after formalization (Row 1) and two periods after formalization (Row 2). Matching is performed one period prior to formalization.

Table 11
Formalization and productivity – IV estimates.

	(1) Labor productivity	(2) TFP
Formality status	13.60***	16.63
	(4.65)	(1.29)
Firm size	−0.662***	−1.399
	(−4.25)	(−1.35)
Owner's gender	0.165*	0.400
(male = 1; female = 0)	(1.69)	(1.06)
Owner's education	−0.241*	−1.033
(unskilled = 0; vocational level = 1; college or above = 2)	(−1.80)	(−1.09)
Owner's business networks	−0.000530	−0.00854
(no. of business/govt contacts)	(−1.13)	(−1.08)
Workforce skill level	0.157*	0.326
(share of unskilled prod. workers)	(1.66)	(1.14)
Property rights	0.316***	0.254
(land use rights certificate = 1; otherwise = 0)	(3.15)	(1.09)
Road/port/rail access	0.0222	−0.112
(none = 0; all = 3)	(0.59)	(−0.80)
Constant	10.36***	11.01***
	(17.61)	(3.76)
Province dummies	Yes	Yes
MHT sector dummy	Yes	Yes
First-stage F-stat	21.24	19.68
Observations	4163	2338

Note: t-statistics reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

7. Robustness checks

While the matched DiD approach is useful for handling endogeneity biases due to reverse causality and selection on observables, it does not address the potential problem of selection on *unobservables*. For example, the owners'/managers' business aptitude and determination may be a firm-specific, time-variant factor that is not fully captured through the observed variables such as educational attainment or business networks. Thus, for robustness checks, this paper applies an IV identification strategy to alleviate the issue of selection on unobservables. The instrumental variable is the share of formal firms within the same province, two-digit sector, and year, excluding the observation of interest. This instrument is valid for two reasons. First, an individual household firm's decision to formalize is likely to be influenced by whether its nearby competitors are also formalizing (relevance condition). Second, unobserved firm-specific factors are unlikely to correlate with *other* firms' decisions to formalize, especially when "business networks" have already been controlled for (exogeneity condition).

Table 11 presents the IV estimates of the impact of formalization on productivity. The list of control variables in both columns (1) and (2) is similar to those in column (6) of Table 6. In column (1), the dependent variable is the log of labor productivity, while in column (2), it is the log of TFP. The key explanatory variable of interest is formality status, instrumented by the share of formal firms within the same province, two-digit sector, and year. The weak instrument test shows an F-statistic of 21.24, meaning that the selected instrument meets the relevance condition. Its coefficient in column (1) is positive and significant, which is in line with the baseline and matched DiD estimates and further confirms the positive impact of formalization on labor productivity. In column (2), while the selected instrument is still relevant with an F-statistic of 19.68, the coefficient of formality status is not significant at the 10-percent level. This is in line with the OLS and matched DiD estimates, which do not show any effect of formalization on TFP. Further, Appendix 5 shows that this remains consistent when the proportion of formal firms outside the province of the firm of interest is used as the instrument. Together, the results from Table 11 suggest that formalization leads to better capital deepening, which raises labor productivity, but not efficiency improvements (and hence has no effect on TFP).

The positive impact of formalization on capital deepening is further confirmed in Table 12. Columns (1) and (2) present the OLS and FE estimates, respectively, with the log of real capital stock as the dependent variable. Columns (3) and (4) show the OLS and FE estimates of the impact of formalization on the amount of firm investment, measured in VND thousand. The list of control variables is similar to those in column (6) of Table 5, with the exception of the lagged dependent variable, which is the lag of capital stock (column (1)) and the lag of investment (column (3)) instead of lagged productivity. Under all specifications, the explanatory variable of interest, formality status, is found to be positively and significantly correlated with both capital stock and investment. Specifically, joining the formal sector raises capital stock by 51–76 percent and increases the amount of investment by VND 310 million–VND 469 million (USD 15,000–23,000) after two years. This indicates that household firms have easier access to capital after joining the formal sector.

Table 12
Formalization and capital deepening.

	Capital stock (log real VND million)		Investment (in VND thousand)	
	(1) OLS	(2) FE	(3) OLS	(4) FE
Formality status	0.514***	0.768***	469182.7**	310313.8***
	(4.38)	(3.84)	(2.57)	(3.65)
Firm size		0.0796	183374.6***	114939.0***
		(1.06)	(5.99)	(3.55)
Owner's gender	0.153***	−0.163**	51328.3	−21016.4
(male = 1; female = 0)	(3.25)	(−1.99)	(1.59)	(−0.49)
Owner's education	0.130**	0.643***	61265.3	−5784.6
(unskilled = 0; vocational level = 1; college or above = 2)	(2.10)	(7.89)	(1.37)	(−0.16)
Owner's business networks	0.00142*	0.000758	1792.5	−42.49
(no. of business/govt contacts)	(1.84)	(1.12)	(1.17)	(−0.08)
Workforce skill level	0.104**	0.102*	−2453.8	−26915.2
(share of unskilled prod. workers)	(2.02)	(1.86)	(−0.07)	(−0.74)
Property rights	0.139**	0.356***	−20529.4	8976.2
(land use rights certificate = 1; otherwise = 0)	(2.44)	(3.95)	(−0.48)	(0.21)
Road/port/rail access	0.0579***	−0.0410	−14870.0	−5398.2
(none = 0; all = 3)	(2.77)	(−1.18)	(−1.02)	(−0.35)
Constant	4.321***	10.22***	−350734.3***	87805.7
	(17.11)	(55.01)	(−3.36)	(1.08)
Year dummies	Yes	No	Yes	Yes
Province dummies	Yes	Yes	Yes	Yes
High-tech sector dummy	Yes	Yes	Yes	Yes
Observations	2339	3299	1161	2047
R^2	0.592	0.066	0.228	0.035

Notes: OLS and FE estimates. *t*-statistics (reported in parentheses) are heteroskedasticity-robust. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

8. Conclusion

Many developing countries implement policies that encourage informal firms to join the formal sector. These policies will not work if the perceived costs of formalization outweigh its expected benefits. Most empirical studies to date, however, have not analyzed both the benefit and cost sides of formalization. This study attempts to fill this gap by examining the relationship between formalization, firm-level productivity, and hidden costs, using a panel dataset of Vietnamese firms during the period 2007–2015. Specifically, this paper focuses on two research questions: (i) Does formalization bring about *higher productivity* for formalized firms? and (ii) Do informal firms incur *higher hidden costs* after formalization?

The findings show that formalization is associated with a 23–69 percent increase in labor productivity but has no significant impact upon TFP. This means that the gain in labor productivity actually comes from capital deepening rather than efficiency improvements. Models with capital stock and investment as the dependent variables further confirm this assessment. These findings are understandable given that most informal household firms, even after formalization, are micro-firms with less than 10 employees; thus technological upgrades may be less of a priority for these firms than gaining easier access to credit.

In addition, this paper finds evidence of the *visibility* effect, meaning that household firms have to incur higher hidden costs after joining the formal sector. Managers/owners have to allocate additional time to dealing with government regulations and spend an extra VND 10 million–VND 12 million on bribery payments after their household businesses become formal.

This paper's main contribution is that it is, to the best of our knowledge, the first empirical study to shed light on both the micro-level benefit and cost sides of formalization. Previous micro-level studies, such as Farrell (2004), Fajnzylber et al. (2009), or Rand and Torm (2012), focus exclusively on the benefit side of formalization. This is understandable given the widespread lack of information on the productivity of household firms and the hidden costs they incur. Such a one-sided approach, however, does not help explain why informal firms are often hesitant to join the formal sector despite the potential benefits. By contrast, this paper is able to examine both the advantages and disadvantages of formalization thanks to the unique panel dataset that contains a vast wealth of information on the operations of informal firms.[9]

[9] Since this paper gathers evidence only on hidden costs as opposed to the total costs of formalization, we refrain from weighing up the costs and benefits of formalization.

This paper also opens up new directions for future research. One direction would be to investigate the *channels* through which formalization affects productivity. For example, how do formalized firms utilize their improved access to credit if they are not using it for R&D or human capital investment? Do they invest more in marketing to improve their reputation or expand into additional markets? The second direction would be to analyze the *long-term* effect of formalization on productivity. In our dataset, most formalized firms are still small-sized enterprises, but over time they may be able to expand in size, and the beneficial impact on their TFP may become more evident. Given the prevalence of informal firms, research in this direction could provide meaningful recommendations for policymakers in developing countries.

Data statement

Data in support of the findings of this study is available from the corresponding author upon reasonable request.

CRediT authorship contribution statement

Le Phan: Writing – original draft, Methodology, Formal analysis, Conceptualization. **Nguyen Hai Thanh:** Writing – review & editing, Visualization, Data curation.

Appendix

Appendix 1
Classification of manufacturing industries by technological intensity

ISIC Rev. 4 codes	Sector	Technology classification
10	Food products	Low technology
11	Beverages	
12	Tobacco products	
13	Textiles	
14	Wearing apparel	
15	Leather and related products	
16	Wood and products of wood	
17	Paper and paper products	
18	Printing and reproduction of recorded media	
19	Coke and refined petroleum products	
25	Fabricated metal products except weapons and ammunition	
31	Furniture	
22	Rubber and plastics products	Medium technology
23	Other non-metallic mineral products	
24	Basic metals	
32	Other manufacturing except medical and dental instruments	
20	Chemicals and chemical products	High technology
21	Pharmaceuticals	
26	Computer, electronic and optical products	
27	Electrical equipment	
28	Machinery and equipment n.e.c.	
29	Motor vehicles, trailers, and semi-trailers	
30	Other transport equipment	

Source: UNIDO (2010).

Appendix 2
Formality and productivity – Baseline estimates

	Labor productivity		**TFP**	
	OLS	FE	OLS	FE
(Lagged) formal (yes = 1, no = 0)	0.253**	0.407***	0.122	0.313**
	(2.40)	(3.18)	(1.18)	(2.38)
Firm size	0.0643**	−0.00656	−0.0723**	−0.280***
	(2.14)	(−0.14)	(−2.03)	(−5.00)
Owner's gender	−0.0435	−0.0556	−0.0196	−0.0521
(male = 1; female = 0)	(−1.03)	(−0.78)	(−0.42)	(−0.72)
Owner's education	0.0453	0.0108	0.123**	0.115*
(unskilled = 0; vocational level = 1; college or above = 2)	(1.03)	(0.18)	(2.19)	(1.67)
Owner's business networks	0.000217	−0.000106	0.00135*	0.000182
(no. of business/govt contacts)	(1.54)	(−0.88)	(1.83)	(0.20)
Workforce skill level	−0.0330	−0.0492	−0.00323	0.00801
(share of unskilled prod. workers)	(−0.66)	(−0.88)	(−0.06)	(0.13)

(*continued on next page*)

Appendix 2 (*continued*)

	Labor productivity		**TFP**	
	OLS	FE	OLS	FE
Property rights	0.0515	0.0487	0.0215	0.0140
(land use rights certificate = 1; otherwise = 0)	(0.93)	(0.61)	(0.33)	(0.16)
Road/port/rail access	0.0397**	0.0190	0.0416**	0.00938
(none = 0; all = 3)	(2.21)	(0.73)	(2.06)	(0.36)
Constant	9.994***	9.831***	8.123***	8.284***
	(94.88)	(86.05)	(64.14)	(54.95)
Year dummies	Yes	Yes	Yes	Yes
Province dummies	Yes	Yes	Yes	Yes
MHT sector dummy	Yes	Yes	Yes	Yes
Observations	4173	4173	3299	3299
R^2	0.138	0.164	0.110	0.166

Appendix 3

Quality checks for matching procedures

Labor productivity

Covariate imbalance test (nearest neighbor matching): Variable	Mean			*t*-test		V(T)/V(C)
	Treated	Control	% bias	t	p > |t|	
Owner's education	1.0263	0.97368	7.8	0.34	0.736	1.00
Owner's business networks	29.921	29.842	0.3	0.01	0.990	0.55
Workforce skill level	0.35687	0.34834	2.6	0.11	0.910	0.91
Property rights	0.63158	0.65789	−5.4	−0.24	0.814	.
Infrastructure access	1.6316	1.5789	4.6	0.20	0.841	0.95
Province dummies	52.763	51.316	5.6	0.24	0.809	1.02
MHT sector dummy	0.60526	0.57895	5.3	0.23	0.818	.
Labor productivity	9.5101	9.5269	−1.5	−0.07	0.947	1.19

* If variance ratio outside [0.52; 1.92]

Ps R2	LR chi2	p > chi2	MeanBias	MedBias	B	R	%Var
0.004	0.47	1.000	4.1	5.0	15.5	1.02	0

* If B > 25 %, R outside [0.5; 2]

1. K–S tests (continuous variables) and chi-squared tests (categorical variables) for equality of distribution functions (nearest neighbor matching):

Variable	Combined K–S difference/Pearson's chi-square	*P*-value
Owner's education	0.1176	0.943
Owner's business networks	0.1842	0.449
Workforce skill level	0.1053	0.973
Property rights	0.0574	0.811
Infrastructure access	0.1436	0.986
Province dummies	0.7733	0.999
MHT sector dummy	0.0545	0.815
Labor productivity	0.0789	1.000

TFP

Covariate imbalance test (nearest neighbor matching):Variable	Mean			*t*-test		V(T)/V(C)
	Treated	Control	% bias	t	p > |t|	
Owner's education	1.1667	1.0833	15.2	0.37	0.713	1.26
Owner's business networks	44.25	31.5	42.8	1.05	0.306	1.30
Workforce skill level	0.40618	0.30537	31.1	0.76	0.454	1.15

Property rights	0.6667	0.6667	0.0	0.00	1.00	.
Infrastructure access	2	2.0833	−8.5	−0.21	0.836	0.92
Province dummies	45	34.083	39.9	0.98	0.339	1.07
MHT sector dummy	0.5833	0.4167	32.4	0.79	0.436	.
Labor productivity	7.4472	7.3353	12.0	0.29	0.772	1.18

* If variance ratio outside [0.29; 3.47]

1. K–S tests (continuous variables) and chi-squared tests (categorical variables) for equality of distribution functions (nearest neighbor matching):

Variable	Combined K–S difference/Pearson's chi-square	*P*-value
Owner's education	0.2588	0.879
Owner's business networks	0.2500	0.769
Workforce skill level	0.2500	0.769
Property rights	0.0000	1.000
Infrastructure access	0.2222	0.974
Province dummies	4.8667	0.676
MHT sector dummy	0.6667	0.414
Labor productivity	0.2500	0.769

Labor productivity

1. Covariate imbalance test (radius matching):

Variable	Mean			*t*-test		V(T)/V(C)
	Treated	Control	% bias	t	p > |t|	
Owner's gender	0.77273	0.77273	0.0	−0.00	1.000	.
Owner's education	0.86364	0.86364	0.0	0.00	1.000	1.00
Owner's business networks	29.727	35.335	−22.2	−0.62	0.540	0.48
Workforce skill level	0.3423	0.35327	−3.1	−0.11	0.916	0.78
Property rights	0.77273	0.77273	0.0	−0.00	1.000	.
Infrastructure access	1.4091	1.4091	0.0	0.00	1.000	1.00
Province dummies	58	58	0.0	0.00	1.000	1.00
MHT sector dummy	0.59091	0.59091	0.0	0.00	1.000	.
Labor productivity	9.2643	9.5136	−21.1	−0.75	0.458	1.74

* If variance ratio outside [0.42; 2.41]

Ps R2	LR chi2	p > chi2	MeanBias	MedBias	B	R	%Var
0.020	1.19	0.999	5.2	0.0	32.3*	1.42	0

* If B > 25 %, R outside [0.5; 2]

2. K–S tests (continuous variables) and chi-squared tests (categorical variables) for equality of distribution functions (radius matching):

Variable	Combined K–S difference/Pearson's chi-square	P-value
Owner's gender	0.0007	0.979
Owner's education	0.2160	0.898
Owner's business networks	0.1391	0.889
Workforce skill level	0.2672	0.159
Property rights	0.1823	0.669
Infrastructure access	2.4474	0.485
Province dummies	1.3908	0.966
MHT sector dummy	0.2230	0.637
Labor productivity	0.2030	0.456

TFP

1. Covariate imbalance test (radius matching):

Variable	Mean			*t*-test		V(T)/V(C)
	Treated	Control	% bias	t	p > |t|	
Owner's business networks	29.727	35.335	−22.2	−0.62	0.540	0.48
Workforce skill level	0.3423	0.35327	−3.1	−0.11	0.916	0.78
Province dummies	58	58	0.0	0.00	1.000	1.00
MHT sector dummy	0.59091	0.59091	0.0	0.00	1.000	.

* If variance ratio outside [0.43; 2.31]

Ps R2	LR chi2	p > chi2	MeanBias	MedBias	B	R	%Var
0.020	1.19	0.999	5.2	0.0	32.3*	1.42	0

* If B > 25 %, R outside [0.5; 2]

2. K–S tests (continuous variables) and chi-squared tests (categorical variables) for equality of distribution functions (radius matching):

Variable	Combined K–S difference/Pearson's chi-square	P-value
Owner's business networks	0.1789	0.379
Workforce skill level	0.3193	0.012
Province dummies	13.8820	0.085
MHT sector dummy	0.4856	0.486

Time costs of red tape

1. Covariate imbalance test (nearest neighbor matching):

Variable	Mean			*t*-test		V(T)/V(C)
	Treated	Control	% bias	t	p > |t|	
Owner's gender	.8	.8	0.0	0.00	1.000	.
Owner's education	1.1	1.1	0.0	0.00	1.000	1.00
Owner's business networks	46.1	31.7	41.1	1.14	0.268	2.45
Workforce skill level	.31664	.24222	24.4	0.55	0.592	0.88
Inspection frequency	1.7	1.4	21.7	0.48	0.634	1.10
Firm size	11.484	10.002	77.9	1.74	0.099	0.98
Province dummies	45.9	48.3	−0.4	−0.21	0.836	0.84
MHT sector dummy	.5	.5	0.0	0.00	1.000	.
Time costs of red tape	2.71	3.15	−17.5	−0.39	0.701	0.28

* If variance ratio outside [0.25; 4.03]

Ps R2	LR chi2	p > chi2	MeanBias	MedBias	B	R	%Var
0.322	8.94	0.443	22.4	17.5	141.1*	0.51	0

* If B > 25 %, R outside [0.5; 2]

2. K–S tests (continuous variables) and chi-squared tests (categorical variables) for equality of distribution functions (nearest neighbor matching):

Variable	Combined K–S difference/Pearson's chi-square	*P*-value
Owner's gender	0.0000	1.000
Owner's education	0.0000	1.000
Owner's business networks	0.3000	0.660
Workforce skill level	0.3000	0.660
Inspection frequency	1.5333	0.821
Firm size	0.4000	0.294
Province dummies	1.3333	0.970
MHT sector dummy	0.0000	1.000
Time costs of red tape	0.3000	0.660

Bribery payments

1. Covariate imbalance test (nearest neighbor matching):

Variable	Mean			*t*-test		V(T)/V(C)
	Treated	Control	% bias	t	p > |t|	
Owner's gender	0.4444	0.8889	−100.8	−2.14	0.048	.
Owner's education	1.3333	1.3333	0.0	0.00	1.000	1.00
Owner's business networks	34.111	29.667	27.0	0.57	0.574	1.55
Workforce skill level	0.4939	0.2688	67.2	1.43	0.173	3.06
Firm size	11.492	11.087	23.5	0.50	0.626	0.95
Province dummies	52.333	64.667	−43.9	−0.93	0.365	0.76
Bribery payments	4777.8	4466.7	7.8	0.17	0.870	0.89

* If variance ratio outside [0.23; 4.43]

Ps R2	LR chi2	p > chi2	MeanBias	MedBias	B	R	%Var
0.417	10.40	0.167	38.6	27.0	132.8*	0.74	0

* If B > 25 %, R outside [0.5; 2]

2. K–S tests (continuous variables) and chi-squared tests (categorical variables) for equality of distribution functions (nearest neighbor matching):

Variable	Combined K–S difference/Pearson's chi-square	P–value
Owner's gender	4.0000	0.046
Owner's education	0.0000	1.000
Owner's business networks	0.2222	0.960
Workforce skill level	0.5556	0.075
Firm size	0.2222	0.960
Province dummies	7.6667	0.363
Bribery payments	0.3333	0.593

Appendix 4

Formality and productivity – Matched DiD estimates with radius matching

	Formalization treatment effect			
	(1) Labor productivity		(2) TFP	
	ATT	Standard errors	ATT	Standard errors
Radius matching	0.82**	0.38	0.37	0.28
Total obs.	1669		1669	
Treated obs.	100		100	

Notes: Average treatment effect on the treated (ATT) using radius matching (three matches per observation). Reported standard errors are consistent bias-corrected estimators as proposed in Abadie and Imbens (2011). The dependent variable is labor/total factor productivity two periods after formalization. Matching is performed one period prior to formalization. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix 5. Formalization and productivity – IV estimates

	(1) Labor productivity	(2) TFP
Formality status	4.472**	−2.330
	(2.44)	(−0.55)
Firm size	−0.0701	0.0530
	(−0.95)	(0.44)
Owner's gender	−0.00291	−0.104
(male = 1; female = 0)	(−0.04)	(−0.92)
Owner's education	0.0729	0.0930
(unskilled = 0; vocational level = 1; college or above = 2)	(1.07)	(1.23)
Owner's business networks	−0.000161	−0.000268
(no. of business/govt contacts)	(−0.61)	(−0.27)
Workforce skill level	−0.0424	−0.0354

(share of unskilled prod. workers)	(−0.73)	(−0.59)
Property rights	0.0391	0.00161
(land use rights certificate = 1; otherwise = 0)	(0.50)	(0.02)
Road/port/rail access	0.0201	0.00308
(none = 0; all = 3)	(0.72)	(0.10)
Constant	9.897***	7.777***
	(63.75)	(33.71)
Province dummies	Yes	Yes
MHT sector dummy	Yes	Yes
First-stage F-stat	14.68	11.80
Observations	4173	3299

Note: *t*-statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Formalization and hidden costs – IV estimates

	(1) Time costs of red tape	(2) Bribery payments
Formality status	11.25***	3058.9
	(3.03)	(0.38)
Firm size	0.0298	3939.3***
	(0.11)	(3.59)
Owner's gender	0.0756	2429.0
(male = 1; female = 0)	(0.22)	(1.51)
Owner's education	0.443*	−1462.1
(unskilled = 0; vocational level = 1; college or above = 2)	(1.72)	(−1.26)
Owner's business networks	0.000232	2.270
(no. of business/govt contacts)	(0.19)	(0.28)
Workforce skill level	−0.0575	−136.0
(share of unskilled prod. workers)	(−0.22)	(−0.08)
Property rights	−0.229	1362.8
(land use rights certificate = 1; otherwise = 0)	(−0.64)	(0.72)
Road/port/rail access	0.0388	
(none = 0; all = 3)	(0.31)	
Constant	1.556**	−6354.0**
	(2.48)	(−2.24)
Province dummies	Yes	Yes
MHT sector dummy	Yes	Yes
First-stage F-stat	19.19	2.37
Observations	4187	667

Note: *t*-statistics (reported in parentheses) are heteroskedasticity-robust. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

References

Abadie, A., Imbens, G.W., 2011. Bias-corrected matching estimators for average treatment effects. J. Bus. Econ. Stat. 29 (1). https://doi.org/10.1198/jbes.2009.07333

Ackerberg, D.A., Caves, K., Frazer, G., 2015. Identification properties of recent production function estimators. Econometrica 83 (6). https://doi.org/10.3982/ecta13408

Atkinson, R.D., Mayo, M., 2010. Technology, Engineering and Mathematics (STEM) Education. Refueling the US innovation economy: Fresh approaches to Science. The Information Technology & Innovation Foundation.

Aw, B.Y., Chung, S., Roberts, M.J., 2000. Productivity and turnover in the export market: Micro-level evidence from the Republic of Korea and Taiwan (China). World Bank Econ. Rev. 14 (1). https://doi.org/10.1093/wber/14.1.65

Babbitt, L.G., Brown, D., Mazaheri, N., 2015. Gender, entrepreneurship, and the formal-informal dilemma: Evidence from Indonesia. World Dev. 72. https://doi.org/10.1016/j.worlddev.2015.02.019

Baker, A.C., Larcker, D.F., Wang, C.C.Y., 2022. How much should we trust staggered difference-in-differences estimates? J. Financ. Econ. 144 (2). https://doi.org/10.1016/j.jfineco.2022.01.004

Bangasser, P.E., 2000. The ILO and the informal sector: An institutional history. International Labour Organization.

Boyd, T., 2017, June, June. The informal economy in developing nations: a hidden engine of growth. WIPO Mag.

Bruhn, M., McKenzie, D., 2018. Entry regulation and the formalization of microenterprises in developing countries. PEDL Policy Insight Ser. 2.

Chien, C.H., Welsh, A.H., Breunig, R., 2019. Approaches Anal. Micro-Driv. Aggreg. Product.

de Mel, S., McKenzie, D., Woodruff, C., 2011. What is the cost of formality? Experimentally estimating the demand for formalization. Am. Econ. J. 5 (2), 1221–1250. ⟨www.doingbusiness.org⟩.

de Soto, H., 1989. The other path. Harper & Row.

Demenet, A., Razafindrakoto, M., Roubaud, F., 2016. Do informal businesses gain from registration? Panel data evidence from Vietnam. World Dev. 84, 326–341.

Dettmann, E., Becker, C., Schmeißer, C., 2011. Distance functions for matching in small samples. Comput. Stat. Data Anal. 55 (5). https://doi.org/10.1016/j.csda.2010.11.022

Dettmann, E., Giebler, A., Weyh, A., 2020. Flexpaneldid: A Stata toolbox Causal Anal. varying Treat. Time Durat. ⟨www.iwh-halle.de⟩.

Do, T.K.T., Vu, H.Van, 2021. Does formalization increase firm investment in human capital? New evidence from Vietnam. Financ. Res. Lett. 42, 101889.

Fajnzylber, P., Maloney, W.F., Montes-Rojas, G.V., 2009. Releasing constraints to growth or pushing on a string? Policies and performance of Mexican micro-Firms. J. Dev. Stud. 45 (7), 1027–1047. https://doi.org/10.1080/00220380802264911

Farrell, D., 2004. The hidden dangers of the informal economy. McKinsey Q. 3.
Fields, G.S., 2004. A Guide multisector Labor Mark. Models.
GSO, 2019. Statistical Yearbook of Vietnam. Statistical Publishing House.
ILO, 1993. Resolution concerning statistics of employment in the informal sector, adopted by the Fifteenth International Conference of Labour Statisticians (January 1993). *Fifteen-.-. Int. Conf. Labour Stat.* 1 (Issue August).
Jaramillo, M., 2009. *Is. there Demand Form. Informal firms? Evid. Micro Downt. Lima.* Dtsch. Inst. F. üR. Entwickl. (DIE.
La Porta, R., Shleifer, A., 2008. The unofficial economy and economic development. Brook. Pap. Econ. Act. 2. https://doi.org/10.2139/ssrn.1304760
Lay, J., Tafese, T., 2020. Formalization and productivity: Firm-level evidence from Viet Nam 2020/164 UNU-WIDERhttps://doi.org/10.35188/UNU-WIDER/2020/921-1
Le, D.T., Malesky, E., Pham, A., 2020. The impact of local corruption on business tax registration and compliance: Evidence from Vietnam. J. Econ. Behav. Organ. 177, 762–786. https://doi.org/10.1016/j.jebo.2020.07.002
Leuven, E., Sianesi, B., 2003. PSMATCH2: Stata module to perform full Mahalanobis and Propensity Score Matching, version 1.2. 1. *Stat. Softw. Compon., Febr.*.
Levinsohn, J., Petrin, A., 2003. Estimating production functions using inputs to control for unobservables. Rev. Econ. Stud. 70 (2). https://doi.org/10.1111/1467-937X.00246
Malesky, E., Taussig, M., 2009. Out of the gray: The impact of provincial institutions on business formalization in Vietnam. J. East Asian Stud. 9 (2). https://doi.org/10.1017/S1598240800003003
Maloney, W.F., 2004. Informality revisited. World Dev. 32 (7), 1159–1178. https://doi.org/10.1016/j.worlddev.2004.01.008
McCaig, B., Nanowski, J., 2019. Business formalization in Vietnam. J. Dev. Stud. 55 (5), 805–821.
Nguyen, T.B., Albrecht, J., Vroman, S., Westbrook, M.D., 2007. A stocktaking of Vietnam's labor market policies. In: Felipe, J., Hasan, R. (Eds.), Labor market in Asia: Issues and perspectives. ADB.
Olley, G.S., Pakes, A., 1996. The dynamics of productivity in the telecommunications equipment industry. Econometrica 64 (6). https://doi.org/10.2307/2171831
Peri, G., Shih, K., Sparber, C., 2015. STEM workers, H-1B visas, and productivity in US cities. J. Labor Econ. 33 (S1). https://doi.org/10.1086/679061
Rand, J., Tarp, F., 2011. Does gender influence the provision of fringe benefits? Evidence from Vietnamese SMEs. Fem. Econ. 17 (1). https://doi.org/10.1080/13545701.2010.542003
Rand, J., Torm, N., 2012. The benefits of formalization: Evidence from Vietnamese manufacturing SMEs. World Dev. 40 (5).
Sheng, S., Zhou, K.Z., Li, J.J., 2011. The effects of business and political ties on firm performance: Evidence from China. J. Mark. 75 (1). https://doi.org/10.1509/jmkg.75.1.1
Siepel, J., Camerani, R., Masucci, M., 2021. Skills combinations and firm performance. Small Bus. Econ. 56 (4). https://doi.org/10.1007/s11187-019-00249-3
Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. Econometrica 65 (3). https://doi.org/10.2307/2171753
Svensson, J., 2003. Who must pay bribes and how much? Evidence from a cross section of firms. Q. J. Econ. 118 (1), 207–230.
Tybout, J.R., 2000. Manufacturing Firms in Developing Countries: How Well Do They Do, and Why? J. Econ. Lit. 38 (1). https://doi.org/10.1257/jel.38.1.11
Ulyssea, G., 2018. Firms, informality and development: Theory and evidence from Brazil. Am. Econ. Rev. 108 (8), 2015–2047. ⟨www.doingbusiness.org⟩.
World Bank, 2022. In: Ohnsorge, F., Yu, S. (Eds.), The long shadow of informality. World Bank Group.