Journal of Economic Behavior and Organization 177 (2020) 762–786

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

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

The impact of local corruption on business tax registration and compliance: Evidence from Vietnam

Duong Trung Le[a], Edmund Malesky[b], Anh Pham[c,*]

[a] World Bank, United States
[b] Duke University, United States
[c] George Mason University, United States

ARTICLE INFO

Article history:
Received 5 December 2019
Revised 3 June 2020
Accepted 3 July 2020

JEL classification:
D73
H26
O17
O12

Keywords:
Corruption
Tax Compliance
Informality
Household Business

ABSTRACT

This paper studies how corruption affects two fundamental dimensions of a firm's tax compliance: the likelihood of tax registration (possession of a tax ID) and the tax compliance ratio (the ratio between the firm's tax payment and revenue). We explore a census covering all Vietnamese household businesses and leverage the differential exposure to corruption, depending upon which province similarly situated businesses are located within. Comparing household businesses in contiguous commune pairs that straddle provincial borders, we discover two seemingly contradictory results. We find that a household business that operates in a more corrupt province is more likely to possess a tax ID, even though it does not necessarily pay more in taxes. In fact, among firms that possess tax IDs, an increase in corruption is associated with a decrease in the tax compliance ratio. We suggest a plausible explanation for this pattern is that corrupt bureaucrats encourage tax-ID possession, because the registration form provides them with better business information to extract bribes. This mechanism implies that an increase in corruption should be associated with a smaller increase in tax-ID possession among more "visible" businesses. We test and find supporting empirical evidence for this prediction.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

The relationship between corruption in the form of bribery to government officials and tax compliance is an important but understudied policy problem for many developing countries. Convincing citizens to pay their taxes is critical for delivering public services. If corruption reduces the government's tax yield, it can severely impede economic development. While there is a large literature addressing corruption and tax compliance separately, very few papers examine the association between the two, especially in the context of firms. Within the business tax compliance context, the academic focus has been on formal firms, those that are often legally required to be registered for tax purposes.[1,2] Existing research on tax reg-

* Corresponding author.
E-mail addresses: dle6@worldbank.org (D.T. Le), eddy.malesky@duke.edu (E. Malesky), apham16@gmu.edu (A. Pham).

[1] Some examples are Alm et al. (2004); Chen et al. (2010); Dyreng et al. (2010); Desai and Dharmapala (2006); Graham and Tucker (2006).

[2] During the period of our study, tax-ID registration and business registration required two different and independent administrative procedures in Vietnam. Appendix A explains the differences between the two registration procedures in more detail. To simplify the terminology, we refer to formal firms as firms with tax IDs.

https://doi.org/10.1016/j.jebo.2020.07.002
0167-2681/© 2020 Elsevier B.V. All rights reserved.

istration, especially of smaller firms that are usually subject to less formal regulation is scarce, even though informal firms make up a large fraction of many developing economies.

This paper studies how corruption affects two important and related layers of Vietnamese household businesses' tax compliance: tax registration (i.e., the possession of valid tax identification) and the tax compliance ratio (measured by the ratio between tax payments and revenue).[3] Household businesses are small, located in only one location, and highly prevalent in Vietnam and many other developing countries. In total, they make up about 17–30% of total GDP and at least 30% of employment in Vietnam.[4,5] In Vietnam, a household business is required to register and obtain a tax identification document (henceforth, *tax ID*) and pay taxes to the government if its annual revenue exceeds a standard threshold of 100 million Vietnam Dong (VND).[6] In this paper, we focus particularly on the compliance behavior of businesses that must pay taxes, because their reported revenue exceeds the government's threshold.

We employ two main sources of data for the analysis: the Vietnamese Household Businesses census (henceforth, *VHBS*) and the Provincial Competitiveness Index (henceforth, *PCI*). We obtain tax-ID registration status from the 2017 VHBS census. We also utilize a subset of the VHBS—a nationally representative sample of household businesses—to obtain information about tax payments. Next, we extract provincial corruption indicators from the PCI, which consists of an annual set of composite governance index scores for each of the sixty three provinces in Vietnam.[7]

Following Dube et al. (2010), we compare household businesses located in neighboring communes (the third subnational tier of government) but on opposite sides of provincial borders (the first subnational tier of government) to take advantage of the discontinuous variation in the level of provincial corruption. Our identification strategy relies on the assumption that, conditional on firm-specific observable characteristics, household businesses operating in contiguous commune pairs are virtually identical except for their differential exposure to provincial corruption. The assumption is plausible because a commune is a highly granular administrative unit—a median commune size is just above five square miles. Importantly, provincial borders in Vietnam were often arbitrarily drawn in the past, slicing through localities with very similar histories and culture (Malesky, 2009).

To further address concerns related to the excludability assumption, we progressively control for other important provincial governance activities, commune geographic characteristics and economic conditions, and firm-level characteristics.[8] We find robust results across different sets of covariates, and also when extending the sample beyond the contiguous pairing to include commune pairs based on proximity. Additionally, we check for and find no evidence that household businesses move across provincial borders to avoid corruption.

It is important to acknowledge that our identification strategy does not rely on a randomized control trial or exogenous policy change. Instead, we leverage the spatial variation in corruption indicators across provincial borders and the granularity of communes. Even though our results are robust to controlling for other important governance dimensions, we recognize that our method may not fully isolate corruption from unobserved institutional features. Thus, our measure of corruption may be capturing general institutional quality.

Our paper delivers two seemingly contradictory results. First, a one-standard-deviation increase in corruption is associated with a 15 percent *increase* in the probability that a household business registers for a tax ID. Second, an increase in corruption is associated with a *decrease* in the tax compliance ratio among firms in possession of tax IDs.[9] Section 3.2.2 and Appendix B explain how our results are unlikely to be driven by misreporting of tax payments or revenue.

Next, we propose a theory for these seemingly counter-intuitive results. We suggest that corrupt bureaucrats have a strong incentive to encourage tax-ID registration because it reduces transaction costs in bribe extraction by clarifying firm ownership, location, size, and performance.[10] This leads to a positive relationship between corruption and tax-ID possession. Once a firm has a tax ID, bureaucrats can request bribes with much greater precision, offering firms the opportunity to lower their tax burdens by bribing the tax collector. This leads to a negative relationship between corruption and reported tax payments to the government.[11]

To test our theory, we probe an observable implication of our proposed mechanism. If tax-ID possession removes an asymmetric information advantage that the business possesses over the tax collector, tax IDs should be less important for bribe extraction if businesses are already more visible to bureaucrats. We find that an increase in corruption is associated with a smaller increase in tax-ID possession among more visible businesses—those hiring more workers or operating more hours.

[3] In Section 3.2, we discuss how misreporting in the survey would affect our empirical results.

[4] See Nguyen (2019), Central Institute for Economic Management (2017), or Truong et al. (2013).

[5] For comparison, Alm and Embaye (2013) estimate that the shadow economy is 37% for lower-middle-income countries, and 38% for low-income countries. Hsieh and Olken (2014) show that in India, Indonesia, and Mexico, 98%, 97%, and 92% of firms have fewer than 10 employees, and these small firms employ 65%, 54%, and 22% of the labor force.

[6] Approximately USD $4,300 in 2018.

[7] The PCI is constructed from an annual firm-perception survey on the quality of local governance. Provincial corruption is one of ten PCI subindices. We provide a detailed explanation of PCI subindices in Section 2. More general information about the PCI can be found at http://eng.pcivietnam.org/.

[8] Some examples of provincial governance activities are inspection rates and business services provided.

[9] We note that in this sample, corruption is not associated with tax-ID possession, so we do not condition on an endogenous variable.

[10] We acknowledge that firms might under report their performance. Nevertheless, from the bureaucrats' point of view, having some information is better than not having anything.

[11] In Section 5.2, we discuss several alternative mechanisms.

This paper is closely related to separate literatures on corruption and tax compliance. Despite their individual prominence, there have been very few studies connecting the two by empirically examining the effect of corruption on firm tax compliance. Alm et al. (2016) is a notable exception.[12] The authors use World Bank survey data of formal firms from thirty two countries, finding that larger bribes result in a lower level of tax compliance. Our paper complements Alm et al. (2016) by studying proximate businesses in a single country. We find consistent evidence specifically when examining formal businesses that possess tax IDs.

Our findings unites two separate strands of the corruption literature that have reached seemingly opposite conclusions regarding corruption and regulatory outcomes. First, the positive relationship between corruption and tax-ID possession is consistent with the argument that corruption can lead to certain desirable outcomes by motivating low-paid, public-sector employees.[13] Second, our result that corruption decreases tax compliance among businesses that possess tax IDs is consistent with another empirical literature advocating for undesirable outcomes of corruption.[14] Ultimately, our analysis demonstrates that even though more taxpayers are officially brought into the system in corrupt locations, overall revenue may not necessarily increase.

This paper also complements the prominent empirical literature on how policies affect business informality, which focuses predominantly on entry costs.[15] Recent papers on the ongoing costs and benefits of formal firms focus on tax reduction[16] or the intensity of government auditing.[17] Our analysis extends this literature by examining business informality in the context of corruption. Furthermore, by focusing on the relationship between governance and tax compliance, this study adds to the literature on the informal sector and corruption. This literature shows that becoming officially registered can benefit firms by increasing profits, investments, and value added.[18] Notably, Rand and Tarp (2012) examine the determinants of corruption and discuss how business visibility and formal registration affect corrupt behavior in Vietnam. Our research differs in the empirical design; we look at the inverse relationship, studying how corruption affects tax-compliance behaviors.

Last but not least, our paper advances scholarship by improving data quality in two important ways. First, the VHBS is a uniquely rich data set that allows us to also study whether a business enters the formal tax system by registering for a tax ID. This resolves a key weakness in the broader literature on business tax compliance that only examines the behavior of formally registered firms due to data availability. Secondly, the VHBS is a census, while the existing work on informal businesses mostly employs much smaller-scale survey data.

The rest of the paper is organized as follows. Section 2 describes the Vietnamese institutional background. Section 3 introduces the data. Section 4 presents our empirical approach. Section 5 discusses the empirical results and introduces plausible mechanisms that could explain these findings. Finally, Section 6 concludes.

2. Institutional background

2.1. Household businesses & the Tax system in Vietnam

In Vietnam, a household business is a firm run by a household or family that employs fewer than ten workers and operates in a fixed location. If a business does not meet the above criteria, it must be registered as an enterprise. An enterprise can have fewer or more than ten employees but must register and pay taxes regardless of its revenue. Importantly, enterprises must hire an accountant and prepare official accounting documents, whereas household firms do not face such stringent financial oversight. According to the law, household businesses must pay taxes only if their annual revenue is more than 100 million VND.[19] Officially, to pay taxes to the government, a business needs to obtain a tax ID from the Tax Office, which is under the Ministry of Finance, by filling out tax forms that record its income and tax payments.

There are costs and benefits associated with tax-ID registration. On the one hand, businesses with tax IDs are more exposed to regulation. They are more likely to have to pay taxes and to receive inspections by tax and regulatory officials than businesses without tax IDs. Non-compliant businesses also face fines for not possessing tax IDs.[20] On the other hand,

[12] Khan et al. (2016) examine how performance pay for tax collectors potentially raises revenues by increasing bargaining power of tax collectors compared to taxpayers.

[13] Méon and Weill (2010); Jeong et al. (2019); Weaver (2018); Dreher and Gassebner (2013); Dabla-Norris (2000); Laffont and N'Guessan (1999).

[14] Some examples are Ferraz et al. (2012); Fisman and Svensson (2007); Reinikka and Svensson (2005).

[15] Some examples are Bruhn (2010); Kaplan et al. (2011); de Mel et al. (2013).

[16] Fajnzylber et al. (2011); Rocha et al. (2018); Ulyssea (2018).

[17] Almeida and Carneiro (2009); Holland (2016, 2015).

[18] Rand and Torm (2012); Boly (2018); Rand and Tarp (2012); McKenzie and Seynabou Sakho (2010).

[19] See Article 1, Decree 92/2015/ND-CP.

[20] According to Circular 166/2013/TT-BTC, there is an initial one-time fine equivalent to 14 to 40 percent of average monthly Vietnamese GDP*per capita* for non-compliance. The amount varies depending on how late the business is in applying for a tax ID. Second, the fine for filling out tax forms late also increases for each day that they are late. Third, a penalty for paying taxes late is 0.05 percent per day for the first ninety days and 0.07% per day after ninety days on the amount due. Fourth, if convicted of evading taxes, businesses have to pay a penalty up to three times the amount of taxes that they owe, depending on how many violations the firm has on record. If the evaded amount is more than 300 million VND (or approximately $15,000 USD), the case will go to court, and the business owner has an option to either pay the penalty or face jail time.

businesses with tax IDs can expand their customer base more easily because they can work with customers who require official receipts.

Since 2014, the most common taxes for household businesses are the annual license tax, value added tax (VAT), and personal income tax (PIT). The license tax is a fixed amount per year based on revenue brackets.[21] If we translate the fixed amount to rates in each revenue bracket, they range from 0.1% to 0.3% of a business' revenue. VAT and PIT rates are set by the central government and vary by industry. VAT and PIT tax payments are calculated as a business' tax rate times its revenue. In practice, businesses do not distinguish between paying a VAT or a PIT, as both taxes have identical rules. The combined VAT and PIT rates are 1.5%, 4.5%, and 7%. Tax deduction is rare and only possible when the owners are terminally ill or when the business is temporarily closed. Thus, under reporting revenue is almost the only way that businesses avoid or evade taxes.[22]

Officially, Vietnam's first-tier administrative structure includes fifty eight provinces and five national-level municipalities. Below the first tier are 713 district-level units comprising of rural districts, towns, and provincial-level cities. There are over 11 thousands third-tier level units, including city wards, townships, and communes, which we refer collectively to as *communes*. District tax offices are subordinates of provincial tax offices, which are subordinates of the central tax office. There are no tax offices at the commune level.[23]

Officials overseeing household businesses are either the same or subordinates of officials overseeing formal enterprises. Provincial tax offices are responsible for the tax collection from formal enterprises that are large or operate in multiple locations.[24] District tax offices oversee the remaining of businesses, including 1) smaller formal enterprises that do not have activities in multiple locations, and 2) household businesses.[25] For this reason, small formal enterprises and household businesses are more likely to face similar tax governance at the district level, which is very likely to be influenced by the governance at the province level that oversees larger enterprises. Therefore, the corruption of officials overseeing formal enterprises also influences the corruption of officials overseeing household businesses. This institutional detail is important. As we discuss in Section 3, our corruption measure is derived from a survey of formal enterprises.

2.2. Subnational administration and opportunities for corruption

Two features of Vietnam's administrative structure are critical for our research design that exploits the variation in corruption across provinces. First, while Vietnam is constitutionally a unitary state under a single-party communist regime, fiscal and investment decentralization since 1989 have provided provincial leaders with enormous authority over businesses operating within their borders (Vu, 2016b). Consequently, corruption varies dramatically among the different provinces of Vietnam.[26] Second, fiscal transfers and bureaucratic promotion opportunities ensure that district- and commune-level authorities are clearly subordinate to provincial officials. For the most part, commune or district officials expect to be promoted within the province and not to Hanoi. This means that provincial-level authorities have direct control over the promotion of commune-and district-level subordinates and can even select the criteria by which they are evaluated. In addition, provinces directly control fiscal decisions and regulatory implementation and enforcement in subordinate districts and communes (Do and Iyer, 2008; Hoang and Nguyen, 2018; Painter, 2003). Thus, while corruption may vary within a province somewhat, the administrative structure ensures that variance in corruption among communes within a province is less than variance across provinces. Consequently, businesses in adjacent communes, but on opposite sides of provincial borders, are likely to have very different experiences regarding the scope and scale of bribery requests.

3. Data and variables

3.1. Provincial Competitive Index (PCI)

The PCI is an annual survey of over ten thousand formal, domestic, and private enterprises (not household businesses) that answer questions related to their experience with provincial governance (Malesky et al., 2019). Firms are selected using random sampling to mirror provincial populations and are stratified by firm age, legal type, and sector to ensure that a representative sample is collected. The annual PCI report presents the overall ranking of provincial governments. This paper uses the 2017 iteration of the PCI survey to match with the 2017 Household Business Census.[27]

The overall PCI consists of ten subindices, measuring economic governance dimensions that affect the private sector. The subindices are 1) entry costs (time and procedures required to register as a formal business), 2) the ability of firms to access

[21] See Article 4, Decree 139/2016/ND-CP.

[22] See Tax Law 71/2014/QH13.

[23] Kerkvliet and Marr (2004); Vu (2016a).

[24] See Article 3 in Circular 127/2015/TT-BTC. Also, according to Article 3 in Circular 127/2015/TT-BTC, the provincial tax offices are also responsible for formal enterprises that are state-owned, foreign-owned, large, having multiple locations, in financial services, banking, stocks, insurance, real estate, auditors, law, and mineral. Such firms make up less than 4% of the PCI sample.

[25] See Article 6 in Circular 92/2015/TT-BTC.

[26] We elaborate on this issue in Section 3.1.

[27] Table 16 in the Appendix shows that regression results using the corruption levels in 2015 and 2016 (*Corruption* score constructed from PCI 2015 and PCI 2016) are similar with the main results using the 2017 corruption score.

Table 1
Statistics of the 2017 PCI's Governance Subindices.

	Mean	SD	Min	Max
*Corruption**	*5.63*	*0.85*	*3.182*	*6.9*
Entry	7.80	0.50	6.25	8.926
Land	6.29	0.59	4.35	7.608
Transparency	6.30	0.36	5.33	7.251
Time	6.55	0.81	4.83	8.69
Bias SOE	5.13	0.77	3.33	6.419
Proactivity	5.53	0.75	3.63	7.072
Business Services	6.50	0.59	4.77	7.815
Labor Training	6.42	0.77	5.09	8.175
Legal Framework	5.89	0.67	4.02	7.203
Observations	63			

Note: The reconstructed corruption index used in this paper is derived from PCI's *Informal Charges* subindex (i.e., Corruption score = 11 - Informal Charges score). Essentially, a province with a higher (reconstructed) corruption score is more corrupt.

and secure land, 3) the transparency in access to necessary information, 4) informal charges (corruption), 5) time costs (the amount of time and frequency that firms deal with bureaucratic procedures and inspections), 6) policy biases toward state- and foreign-owned firms, 7) the proactivity of provincial leadership in solving business problems for private firms, 8) quality of business support services (e.g., policies in industrial and economic zones), 9) labor training and recruitment policies, and 10) the fairness and effectiveness of legal procedures for dispute resolution. These ten subindices are then aggregated into an annual provincial PCI aggregate index. Each subindex ranges from one to ten, so that the higher the aggregate overall PCI index and PCI subindices, the better the ranking of the provincial government's economic governance.

In this paper, we focus on the corruption subindex.[28] It is important to note that, following Section 2.2 above, the PCI corruption index captures a general culture of corruption at the province level, which includes the corruption level of tax collectors. We use a modified corruption subindex, which is obtained by subtracting the raw corruption subindex from 11 (i.e., modified corruption score = 11 - raw corruption score). This simple procedure helps ease the interpretation, allowing for a positive relationship between the treatment variable (corruption score) and the provincial-level experiences with corruption. In other words, a province with a higher corruption score means the province is more corrupt. Fig. 1 shows the PCI corruption ranking distribution in 2017. Table 1 provides the summary statistics of the subindices.

Table 1 and Fig. 1 show that there is great variation in provincial corruption scores. Specifically, the standard deviation of corruption is 0.85, which is about 20 percent of the mean of 4.63. In addition, as Section 2.2 described, PCI reported that the share of firms claiming bribery was a normal part of business ranged from 40% in the least corrupt province to 76 percent in the most corrupt province, and the share of firms whose total bribe payments amounted to over 10% of revenue ranged from less than 1% to 26% in the most corrupt location (Malesky et al., 2019).

The following indicators from the PCI survey were used to create the PCI corruption subindex (all are re-scaled so that a higher score represents greater corruption):

- Enterprises in my line of business usually have to pay for informal charges (% agree or totally agree).
- Percentage of firms paying over 10 percent of their revenue in informal charges.
- Administrative procedures are commonly used to extract bribes from firms (% strongly agree or agree).
- Informal charges delivered expected result (% usually or always).
- Informal charges are at acceptable levels (% strongly agree or agree).
- Percentage of firms paying bribes to a regulatory inspector.
- Share of firms paying informal charges in land administrative procedures (% Yes).
- Share of firms paying kick backs to win the local government's contract (% Yes).
- Share of firms claiming that bribes in court are necessary (% Yes).

It is important to note that the PCI survey targets only formal enterprises and not household businesses, so the PCI governance indices may not accurately capture the true governance quality that affects household businesses. The assumption we make in this analysis is that the overall provincial governance quality experienced by household businesses can be captured in the perceptions of formal enterprises located in the same province. This is a reasonable assumption because officials responsible for dealing with formal enterprises and household businesses in a province are ultimately responding and reporting to provincial leadership.[29] From the tax administration's perspective, as explained in Section 2.1, small enterprises and household businesses are more likely to face similar tax governance at the district level, which is very likely to be influenced by the governance at the province level that oversees larger enterprises. As a robustness check for this institutional claim, we reconstructed the corruption subindex for small firms under five employees to show that the corruption subindex that we use in the paper is similar to the experience of small firms, which, like household businesses, interact

[28] We employ the other subindices as control variables in our fully-specified empirical model.
[29] See Point 15 of Decree 78/2015/ND-CP.

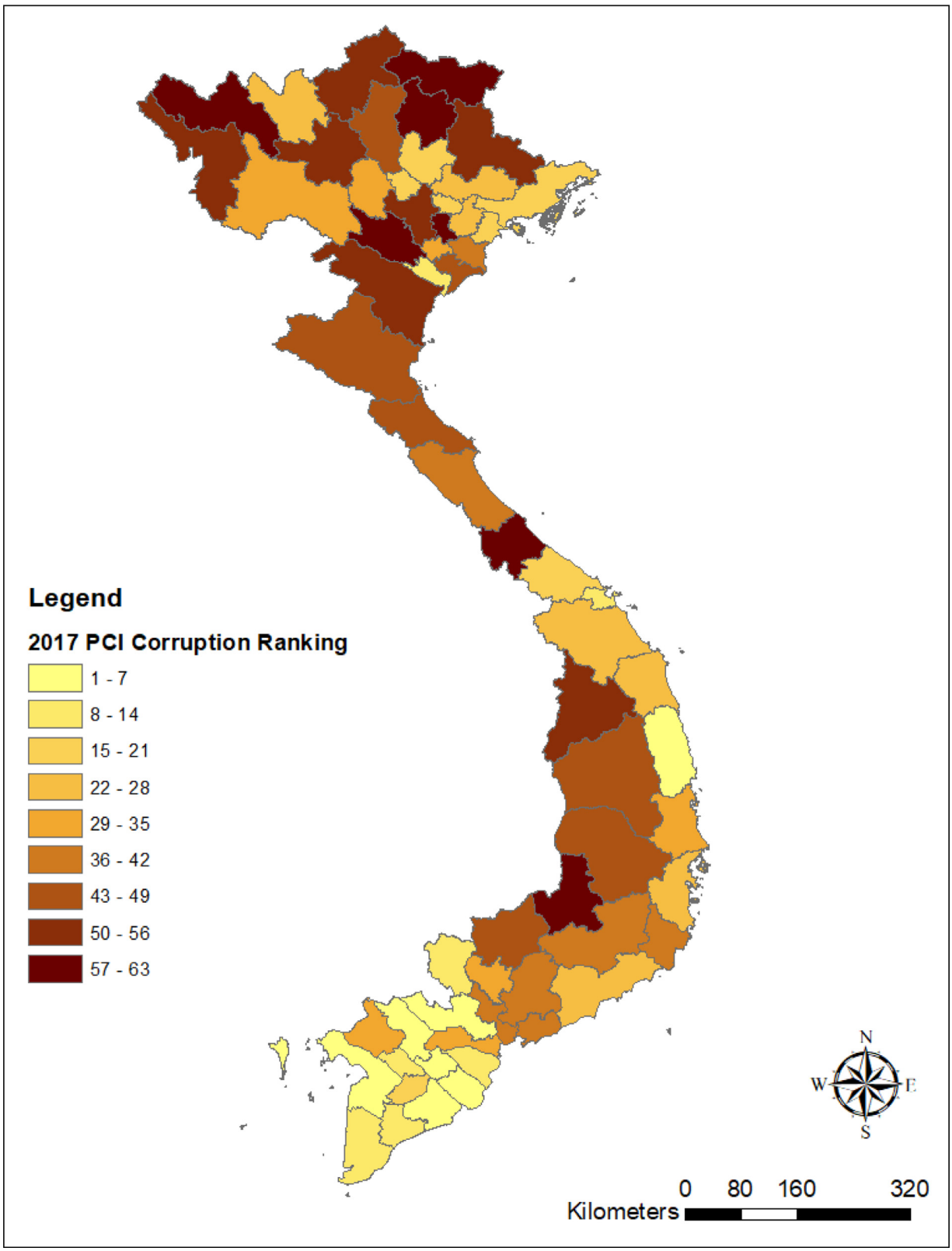


Fig. 1. Distribution of the 2017 PCI Corruption Index Ranking (Reconstructed).
Note: This figure shows the ranking distribution of the reconstructed 2017 corruption index across 63 provinces in Vietnam. The corruption index is derived from the Provincial Competitiveness Index (PCI)'s Informal Charges indicators. A province with a higher corruption ranking (i.e., darker shaded) is more corrupt.

predominantly with district level tax officials. The corruption subindex of small firms and the overall corruption index used in the paper are highly correlated (the correlation is 0.76 and statistically significant at one-percent confidence level). Note that since the sample of small firms in the PCI is not meant to be representative, we only do this exercise for robustness.

Another concern is that while our identification strategy explores the granularity of a commune (the smallest official administrative unit), our corruption index is only representative at the province level. Does corruption at the province level adequately capture the corruption that household businesses face at the commune level? First, as Section 2.2 highlighted, Vietnam's institutional structure ensures that officials of subordinate communes have very little room for independent action in terms of how much they can deviate from provincial corruption norms. While corruption may vary within a province somewhat, variance in corruption among communes within a province is less than variance across provinces. Consequently, firms in adjacent communes but on opposite sides of provincial borders, although operating in the same or very similar business markets, are likely to have very different experiences regarding the scope and scale of bribery requests. Second, if provincial corruption does not adequately capture communal or district corruption, our regression coefficients should be noisy and inconclusive. As we show in Section 5, we are able to estimate regression coefficients with precision, making us less worried about this concern.[30]

3.2. Household businesses in Vietnam

3.2.1. The census

We use the 2017 VHBS to examine the effect of corruption on tax-ID status. The census is collected by the General Statistical Office (GSO) of Vietnam to assess local and regional economic development and calculate the official statistical indicators such as Gross Domestic Product (GDP) at the national and local levels. This census is also used as the sampling frame for smaller surveys in subsequent years. VHBS includes only businesses and street vendors that operate in fixed locations. To collect VHBS, each commune is assigned a survey team whose members are residents living in that commune. Similar to the majority of "conventional" micro and small enterprise surveys in developing countries, all information collected in the VHBS survey is self-reported and does not come from audit data. Nevertheless, at the beginning of the survey and on tax-related questions, the surveyors remind businesses that the questions are for statistical analysis and not for any tax purposes, which should reduce concerns about misreporting.

The data contains information on household business's employment, operating revenues (reported for the first six months of 2017), assets, and tax registration status. Recall that we can locate businesses at the commune level, which is the smallest government administrative unit in Vietnam. By law, only businesses with annual revenue greater than 100 million VND (approximately $4,300 USD) must pay taxes. Since VHBS only reports revenue accumulated in the first six months instead of annually, we restrict the sample to firms that reported revenue in the first six months greater than 100 million VND in order to study tax compliance behavior.[31]

In terms of the level of reporting accuracy, specifically on tax registration status, we do not think that businesses purposely misreport. Businesses that claim to have tax IDs are also asked for their tax-ID numbers. A tax-ID number has ten digits, so it would be difficult to make one up on the spot. Thus, it is unlikely that a business that does not possess a tax ID would claim to have one. In addition, there is also no sensible incentive for businesses that possess tax IDs to claim otherwise. Therefore, any measurement error in the tax-ID possession variable would likely come from recall problems, which are unlikely to be correlated with provincial corruption levels.

3.2.2. The sample of businesses surveyed for tax information

In addition to the 2017 household business census, GSO also carried out an extensive survey that selected a random sample of businesses from VHBS. The survey consists of more detailed questions, including information about the amount of taxes that businesses paid. The sample is representative at the national and provincial level. We acknowledge that the sample of firms in the border communes are not representative of all household businesses in the country. However, we think that it is still helpful to examine the sub-population of firms that this sample represents. Note that this sample includes businesses with and without tax IDs, because the survey was rolled out at the same time as the census. Consequently, GSO did not know which firms had tax IDs and had to pay taxes when constructing the sample frame.

Following Alm et al. (2016), we calculate the ratio between a business's total tax payment and revenue in the first six months to measure tax avoidance or compliance.[32] The total tax payment in the survey is the tax amount that firms have to

[30] In a complementary exercise, we construct a corruption index at the commune level to test whether the variance of communal corruption between provinces differs from the variance of communal corruption within a province (using an F-test). The test implies that the variance in corruption among communes within a province is statistically smaller than that across provinces (F= 3.54; p-value= 0.00). In addition, we find that the variance of corruption among adjacent communes in different provinces is not statistically different from the variance of corruption in the sample of all border communes (F= 0.86; p-value= 0.71). These tests lend confidence to our intuition that there is a distinct difference in the levels of corruption that businesses experience across a provincial border. It is important to note the PCI stratified sample is representative at the provincial level but not at the commune level. Thus, our constructed communal corruption index is only intended for this test, but cannot be used in the broader research project.

[31] In the Appendix, we subsequently check and find no revenue "bunching" evidence under the 100 million VND threshold, which alleviates a concern that firms could manipulate their revenues and report below this official cutoff to avoid paying taxes entirely (Appendix F.1). We also additionally perform robustness checks using several other revenue thresholds (i.e., $\geq$ 50 million, $\geq$ 70 million, and $\geq$ 90 million VND) and find similar results (Appendix C.4).

[32] Alm et al. (2016) calculate the ratio between revenue reported for tax purposes and total revenue. Since we do not have revenue for tax purposes, we use the reported tax payments, which is a function of revenue for tax purposes.

Table 2
Summary statistics of household businesses' characteristics.

	Border Census					Border Sample				
	Obs	Mean	Std.Dev.	Min	Max	Obs	Mean	Std.Dev.	Min	Max
=1 if have a tax ID	387,427	0.205	0.403	0	1	13,034	0.302	0.459	0	1
log revenue	387,427	12.448	0.730	11.51	14.47	13,034	12.490	0.755	11.51	14.47
# workers	387,427	1.941	1.117	1	6	13,034	2.127	1.259	1	6
daily hours	386,508	10.752	3.387	1	24	12,980	10.982	3.007	1	24
=1 if restaurant	387,427	0.152	0.359	0	1	13,034	0.084	0.277	0	1
=1 if manufacturing	387,427	0.152	0.359	0	1	13,034	0.194	0.395	0	1
=1 if retail	387,427	0.474	0.499	0	1	13,034	0.481	0.500	0	1
=1 if wholesale	387,427	0.099	0.298	0	1	13,034	0.096	0.295	0	1
age of owner	387,427	44.388	10.338	16	99	13,034	44.037	10.337	19	93
=1 if owner is male	387,427	0.562	0.496	0	1	13,034	0.621	0.485	0	1
=1 if owner has no training	387,427	0.613	0.487	0	1	13,034	0.546	0.498	0	1
=1 if operate in owner's	387,426	0.681	0.466	0	1	13,033	0.723	0.448	0	1
tax payment/revenue (as-is)	8682	0.007	0.011	0	0.102	8682	0.007	0.011	0	0.102
tax payment/revenue (impute 0)	13,034	0.005	0.010	0	0.102	13,034	0.005	0.010	0	0.102

Note: This table presents descriptive statistics for the main outcome and independent variables used in the analysis. The*Border Census* includes all household businesses in VHBS located at border communes (i.e., communes straddling a provincial border). The*Border Sample* includes all household businesses in VHBS located at border communes (i.e., communes straddling a provincial border) that were additionally surveyed for tax information.

pay in the first six months. The license tax is an annual fixed amount based on revenue levels. Thus, the reported total tax payments in the first six months predominantly includes VAT and PIT, and sometimes the license tax, depending on how a business interprets the survey questions. As explained in Section 2.1, the VAT and PIT statutory tax rates vary by industry, and the license tax rate is relatively small compared to VAT and PIT statutory tax rates. Thus, within an industry, a higher ratio of tax payments by revenue implies higher compliance and lower avoidance.

An important discussion related to the appropriateness of the ratio between tax payments and revenue as a measure for business's tax compliance is how honestly firms report their tax payments and revenue in the survey. On the one hand, if these two indicators are reported with a high degree of accuracy, the ratio is an especially good measure, because there are almost no tax deductions for Vietnamese household businesses. The only way to avoid tax is to under report revenue to the tax authority. On the other hand, firms misreporting their tax amounts and revenue, whether deliberately or not, would potentially bias our subsequent results on tax compliance. In Appendix B, we discuss and address all potential concerns related to the reliability of the self-reported tax payments and revenue in VHBS. We also elaborate on how we believe that 1) firms are more likely to honestly report their tax payments and revenue in the survey, and 2) businesses under reporting revenues would most likely bias against our findings in Section 5. Additionally, besides the self-reported revenue provided by each business, the VHBS survey also asks the interviewers to offer their own independent assessments of what they think the business' revenue is. This measure allows us to calculate an alternative ratio between tax payments and revenue as a robustness exercise. As Appendix B discusses, the regressions using these two measures give very similar results.

Thirty-five percent of observations on tax payments are missing. However, among businesses not reporting tax payments, 97% report not having tax IDs. Therefore, it is likely that firms without tax IDs do not officially pay taxes to the government and thus leave the tax payment question unanswered. We address the missing-observation concern in the analysis by using both the original data (i.e., allowing for missing values in the reported tax payments) and a version where we impute missing tax observations with zero (in case firms that do not owe any taxes decide not to answer the tax-payment questions).

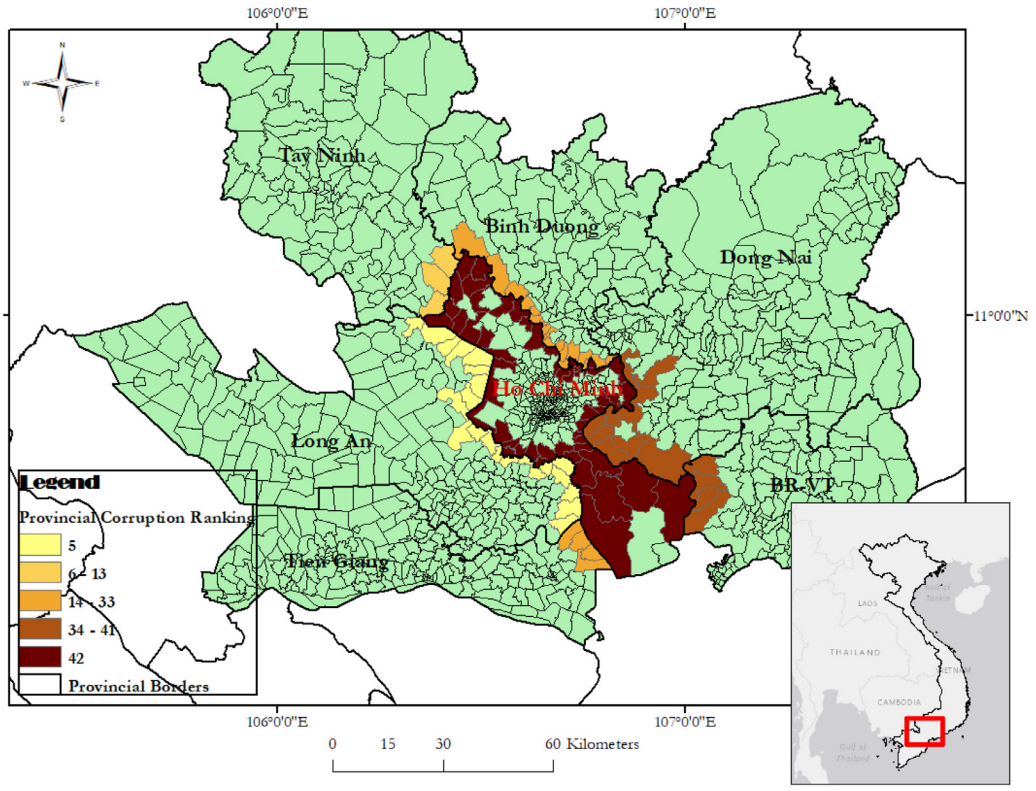
3.2.3. Summary statistics

As will be explained in Section 4, our main analysis sample consists of only household businesses in communes located at provincial borders. Table 2 shows the characteristics of businesses in the border communes with revenue greater than 100 million VND in the first six months. The first three columns are for all businesses in the border communes. The last three columns include businesses in the border communes that provided tax payment information. Among all businesses in the border communes, about 20% have tax IDs. The average number of employees is 1.94. An average business in the sample operates approximately eleven hours a day. For businesses that were asked questions about tax payments, these figures are 30%, 2.13, and eleven, respectively.[33]

4. Empirical method

Our method compares the likelihood of having obtained tax IDs and the tax compliance ratio among similar household businesses that locate in bordering communes–those geographically identical but exposed to different provincial corruption levels. Specifically, we compare businesses in two neighboring communes that belong to two different provinces. Fig. 2

[33] Restaurant and manufacturing sectors coincidently have the same values of mean and standard deviation because of rounding.



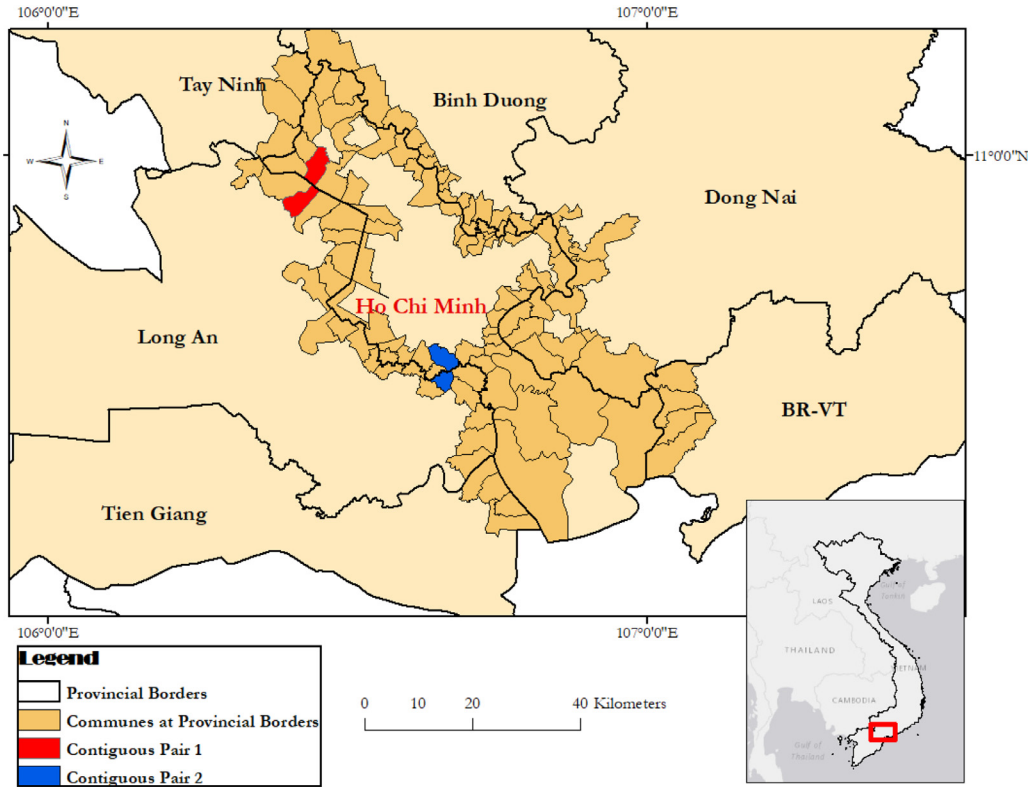


Fig. 2. Visualization of Contiguous Border Commune-Pair Fixed Effects.
*Note:*The above figures provide an illustrative explanation for the identification strategy, using the borders between Ho Chi Minh City and neighboring provinces as an example. The regression sample is restricted to household businesses located in communes straddling provincial borders.

Table 3
Corruption and tax-ID possession of household businesses.

	Dependent Variable: Possession of Tax IDs (0/1)				
	(1)	(2)	(3)	(4)	(5)
Corruption	0.0495***	0.0398**	0.0376**	0.0386**	0.0375**
	(0.012)	(0.0183)	(0.0181)	(0.0182)	(0.0175)
Observations	802,550	802,550	802,547	745,646	726,972
commune-pair FE	Yes	Yes	Yes	Yes	Yes
3-digit Industry FE	Yes	Yes	Yes	Yes	Yes
Governance controls		Yes	Yes	Yes	Yes
Firm's controls			Yes	Yes	Yes
Commune's policy controls				Yes	Yes
Commune's GIS controls				Yes	Yes
Night-light Index (2016)				Yes	Yes
Geographic controls				Yes	Yes
Formal sector activities controls					Yes

Two-way clustered standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
Note: This table shows the results from regression Eq. 1. The sample includes all businesses located in commune pairs straddling a provincial border. The dependent variable is a binary indicator of whether a household business possesses a tax ID. *Corruption* is the constructed index that measures the level of provincial corruption. Two-way standard errors are clustered at the provincial and provincial border levels.

shows a visual illustration of the method. For example, Ho Chi Minh City and Long An are two neighboring provinces but are ranked differently in their 2017 PCI corruption scores (top panel). First, we restrict the sample to a subset of firms in the communes at the border between each province. Second, following Dube et al. (2010), we employ a complete set of pairwise fixed effects for all contiguous commune pairs straddling a provincial border. This ensures that the comparison is drawn only from businesses located in adjacent communes. The right panel of Fig. 2 provides a graphical illustration; businesses in Long An and Ho Chi Minh City are compared separately in color pairs (blue communes and red communes). This ensures that culturally and geographically, the comparison pairs are as similar to one another as possible.

The regression equation of cross-sectional data takes the following form:

$$Y_{fcbp} = \beta_0 + \beta_b + \beta_{industry} + \beta_1 corruption_p + P_p + C_c + X_{fcbp} + \epsilon_{fcbp} \quad (1)$$

Y_{fcbp} is the dependent variable of firm f, in commune c, at border b, and province p. For the tax-ID registration outcome, the dependent Y_{fcbp} equals 1 if a business possesses a tax ID and 0 if it does not. For the tax payment outcome, the dependent variable is the tax compliance ratio. The variable $corruption_p$ is the corruption level of province p in 2017. The higher the value of the corruption variable is, the more corrupt the province is.

$\beta_{industry}$ represents 3-digit industry-specific fixed effects, which control for common shocks that affect a particular industry. β_b represents contiguous commune-pair fixed effects, ensuring that the corruption impact is derived from the comparison of businesses operating in the same commune pairs. P_p, C_c, and X_{fcbp} are covariate vectors controlling for the specific province, commune, and firm characteristics, respectively, that can influence tax-compliance outcomes independently from provincial corruption exposure. In Appendix D, we discuss the control variables in details.

We use two-way clustered standard errors at the province and provincial-border levels, following Dube et al. (2010) and recommendations from Cameron et al. (2011) and Cameron and Miller (2015). Clustering standard errors at the provincial level is recommended, because the treatment variable (the corruption score) occurs at the provincial level, but household firms are nested within those provinces and therefore their errors are likely to be correlated. In addition, our sample consists of all contiguous border commune-pairs. Thus, the presence of a single commune in multiple pairs along a provincial border induces a mechanical correlation across commune-pairs and potentially along an entire provincial border. At the same time, the border identification strategy may yield another violation of the i.i.d. assumption, as firms nested within contiguous border communes cannot be considered fully random draws.[34]

To test for the plausible mechanism in Section 5.2, the regression equation that examines the differential effects of corruption on the tax-ID status of firms that are more or less visible takes the following form:

$$Y_{fcbp} = \beta_0 + \beta_b + \beta_{industry} + \beta_1 Corruption_p + \beta_2 Corruption_p \times Visible_{fcbp} + P_p + C_c + X_{fcbp} + \epsilon_{fcbp} \quad (2)$$

For equation 2, $Visible_{fcbp}$ is measured by the firm's number of workers and hours of daily operation. Specifically, businesses that have more workers or operate longer are more visible to local tax collectors due to the size of their operations. Other variables and specifications in equation 2 are the same as those in equation 1.

Our identification strategy assumes that businesses in neighboring communes are identical other than that they face different corruption levels. The plausibility of our assumption is supported by the fact that a commune is a highly granular administrative unit–there are over eleven thousand communes in Vietnam and a median commune size is just above five

[34] In Appendix C.1 and C.2, we show that our results are robust to alternative specifications: 1) provincial level OLS regressions (Table 8) and 2) Tobit model (Table 9).

Table 4
Corruption and tax-ID possession of household businesses in the tax-survey sample.

	Dependent Variable: Possession of Tax IDs (0/1)				
	(1)	(2)	(3)	(4)	(5)
Corruption	0.0278*	0.0136	0.00734	0.00641	0.00317
	(0.0149)	(0.0272)	(0.0262)	(0.0258)	(0.0259)
Observations	26,545	26,545	26,542	23,484	23,484
commune-pair FE	Yes	Yes	Yes	Yes	Yes
3-digit Industry FE	Yes	Yes	Yes	Yes	Yes
Governance controls		Yes	Yes	Yes	Yes
Firm's controls			Yes	Yes	Yes
Commune's policy controls				Yes	Yes
Commune's GIS controls				Yes	Yes
Night-light Index (2016)				Yes	Yes
Only non-missing geographic controls				Yes	Yes
Geographic controls				Yes	Yes
Only non-missing formal sector activities				Yes	Yes
Formal sector activities controls					Yes

Two-way clustered standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
Note: This table shows the results from regression Eq. 1. The sample includes all businesses surveyed for tax liabilities that are located in commune pairs straddling a provincial border. The dependent variable is a binary indicator of whether household business possesses a tax ID. *Corruption* is the constructed index that measures the level of provincial corruption. Two-way standard errors are clustered at the provincial and provincial border levels.

square miles. And, as we noted in Section 2.2, provincial borders in Vietnam were often arbitrarily drawn in the past, dividing localities with very similar historical and cultural characteristics (Malesky, 2009). However, we recognize that because we have only cross-sectional data, we cannot control for all factors that differ across provinces; thus, we do not claim that our identification assumption is absolutely correct. Nevertheless, in Appendix D, we explain possible concerns of assumption violations and different ways that we attempt to alleviate these concerns. In addition, we progressively control for a host of control variables, X_{fcp}, at the firm, commune, and provincial levels to show in the Section 5 and the Appendix C that the results are highly robust across different econometric specifications.

5. Results

5.1. General results

Table 3 shows how corruption affects the probability of a household business possessing a tax ID. Column (1) shows results when we control for commune-pair fixed effects and industry fixed effects. Column (2) adds nine other PCI governance subindices to the regression to control for other measures of governance quality that differ across provincial borders. Columns (3), (4), and (5) gradually add in more control variables at the firm and commune levels to the regression. Table 3 shows that the corruption coefficients are consistent across different sets of controls.

We use the coefficient of the fully specified specification in column (5) of Table 3 to interpret the magnitude of the effect of corruption on the probability that businesses possess tax IDs. Specifically, a one-point increase in corruption increases the probability of businesses possessing tax IDs by 3.75 percentage points. This translates to an 18-percent increase from the mean probability.[35] Table 1 shows that the standard deviation of corruption is 0.85. Therefore, a one-standard-deviation increase in corruption is associated with a 15-percent increase in the probability that businesses possess tax IDs.

For the rest of the section, we utilize the sample of businesses that were surveyed for tax payment information to examine the relationship between corruption and the tax compliance ratio. To compare our results with Alm et al. (2016), we study this relationship among businesses with tax IDs. We need to do this analysis on a sample of businesses whose decisions to possess tax IDs are not related to corruption, because we do not want to condition our sample on an endogenous variable. Fortunately, Table 4 shows that among businesses that were surveyed for tax information, corruption is not associated with the probability of possessing a tax ID, conditional upon governance controls.[36] Therefore, we can perform analysis of how corruption is associated with the tax compliance ratio using all businesses with tax IDs in the sample of firms that were surveyed for tax payment information.

In Table 5, we examine how corruption is associated with the tax compliance ratio among all firms (with and without tax IDs) and among firms with tax IDs. The first two columns of Table 5 are results using available data. The last two columns are results when we impute missing tax payments with zero.

[35] Table 2 shows that approximately 20% of household business in the sample are registered for tax IDs.

[36] As we see in Table 2, businesses in the sample with tax payment information are somewhat different from the general population of businesses. For example, businesses in the tax payment sample have more workers than all businesses at the border. Columns (1) and (2) of Table 6 show that corruption is associated with a smaller increase in tax-ID possession among businesses with more workers.

Table 5
The effects of corruption on the tax compliance ratio.

Sample:	No Imputation		Impute Missing with 0	
	(1) All	(2) Have tax ID	(3) All	(4) Have tax ID
Corruption	−0.00156***	−0.00277**	−0.000218	−0.00233**
	(0.000583)	(0.00111)	(0.000433)	(0.00114)
Observations	15,385	6109	23,484	6341
Mean Dependent Variable	0.007	0.0148	0.005	0.0134

Two-way clustered standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
Note: This table shows the results from regression Eq. 1. The dependent variable in taxes compliance ratio. Control variables are the same as control variables in column (5) Table 3. Two-way standard errors are clustered at the provincial and provincial border levels.

Columns (2) and (4) of Table 5 show that among businesses that have tax IDs, a one-point increase in corruption is associated with a 0.28 percentage points or 0.23 percentage points decrease in the ratio of tax payments over revenue. The last row of the table shows that among firms with tax IDs, the average ratio is 1.48 percentage points (ignoring missing observations) and 1.50 percentage points (when missing observations are imputed with zero). Thus, a one-standard-deviation increase in corruption is associated with a 15% decrease in the ratio between tax payments and revenue. Therefore, tax compliance decreases among businesses with tax IDs.

Columns (1) and (3) show that among all firms, an increase in corruption is either associated with a decrease (without imputing 0 for missing observations) or not associated with the tax compliance ratio (results after imputing 0 for missing observations). Thus, we do not find evidence that the overall tax compliance increases. Additionally, Table 15 in the Appendix shows an increase corruption does not increase tax payment among all firms or firms with tax IDs.

The specifications thus far include firms located in contiguous communes that are on opposite sides of a provincial border. Tables 10, 11, and 12 in Appendix C show that the results are robust when we expand the sample to firms located in pairs of communes located within the proximity of one, two, and three kilometers from each other. Table 10 also shows that the results are robust after dropping observations in Hanoi and Ho Chi Minh City, the two major municipalities in Vietnam.

5.2. Plausible mechanisms

Our main findings from the previous section indicate that a household business that operates in a more corrupt province is more likely to possess a tax ID, while it pays relatively less in taxes. This section discusses several possible mechanisms for these two seemingly contradictory empirical results.[37] All of these mechanisms depend on two actors: household businesses and local tax collectors. A local tax collector is responsible for collecting taxes from household businesses and needs to meet a tax revenue target. All mechanisms assume the official salary of these tax collectors' salaries are low and they supplement their income with bribe payments. Based on these assumptions, we discuss below how corruption affects tax collectors' and household businesses' behaviors.

Our preferred mechanism for explaining the empirical findings pertains to the apparent asymmetric information regarding household business's revenue: corrupt bureaucrats encourage tax-ID possession because the registration form provides them with business' information, which facilitates bribe extraction. Specifically, in the household business sector in Vietnam and many other developing countries, household businesses are very small, transactions are predominantly cash based, receipts are rarely collected, and accounting standards are poor and seldom implemented. Thus, the household business possesses an asymmetric informational advantage over the tax collector. In many cases, the tax collector actually does not have any idea about businesses' existence, much less their business activity.

Tax IDs provide a partial solution to the tax collector's asymmetric information problem. Businesses with tax IDs must report their contact information and business performance with simple but standardized accounting rules. To maximize take-home income in a corrupt province, therefore the incentive of the tax collector is to coerce as many household businesses as possible into the formal tax system. Doing this not only provides tax collectors with a larger pool of taxpayers to fulfill the tax revenue target, but also provides clearer information on business location and performance, reducing her asymmetric informational disadvantage in negotiations. We recognize that businesses can also vastly under report their performances in the tax forms. Nevertheless, from a tax collector's point of view, having some information is better than having nothing at all. Anecdotally, our conversations with several tax collectors during the course of the research suggest that businesses with tax IDs are inspected much more often.

However, once the business location and part of their performance are known, the incentive for the corrupt bureaucrat is to negotiate more vigorously with the household business. This will result in a larger bribe payment for the bureaucrat and a higher proportion of bribery in the firm's tax bill. Because a rational firm will not pay a total tax bill (i.e., the total

[37] Appendix E discusses how we alleviate the concern that these "seemingly contradictory" results on tax-ID possession and tax-compliance ratio are the artifact of being based on two different datasets.

sum of tax and bribe payments) that is greater than their true tax obligation, this will result in lower tax payments and seemingly lower tax compliance in the corrupt location.

Apart from the above testable mechanism, there are several other plausible alternative explanations for our empirical discoveries. First, under the assumptions that corruption is prevalent and that the corrupt tax official pockets a portion of the firms' tax payments in the form of bribes, she would need to enlarge the pool of formal taxpayers to ensure that she still meets the tax revenue collection target assigned to her. As a result, there is an incentive for the tax collector to exert more effort in coercing firms to register and pay some taxes. This potentially lead to the same empirical observations in the paper: higher corruption is positively associated with tax-ID possession likelihood but negatively with tax payment.

Second, since tax collectors want to enroll businesses in the formal tax system, either because of the "asymmetric information" or the "enlarging the pool" mechanisms explained above, tax collectors would need higher bribes to let businesses stay in the informal sector. Thus, the bribes requested may be too high in a more corrupt province, so that more household businesses decide to register for tax IDs. Once businesses enroll in the formal tax system, they bribe the tax collectors to pay less in taxes if the bribe request is not unreasonably high.

Third, it could also be that businesses who are exposed to an environment where corruption is more widespread would anticipate that they can bribe more easily to reduce their tax payment. Consequently, they would be more willing to register and gain the benefits associated with the official enrollments in the tax system. This is true as long as the benefits can outweigh the cost in paying bribes.

Finally, the "Greasing The Wheel" hypothesis claims that firms are better able to overcome bureaucratic obstacles to fulfill basic regulatory requirements, such as business registration and licensing, by paying bribes. This could lead to higher tax-ID possession in more corrupt provinces. As Dreher and Gassebner (2013) point out, corruption should only be positively correlated with tax-ID possession in locations where the regulatory burden is excessively high. In our case, acquiring a tax-ID is not an especially burdensome activity for household business in our study period. It takes a maximum of 3 days and there is no formal fee (Circular 95/2016/TT-BTC). In addition, the "Greasing the Wheels" hypothesis is a derivation of the optimal level of corruption literature (see Laffont and N'Guessan (1999); Dabla-Norris (2000)), which proposes a quadratic relationship (inverted U-shape) between corruption and bureaucratic compliance. In other words, firms are only willing to pay bribes to formalize up to a point. When the bribery tax for evading regulatory burden is too high (beyond an inflection point), firms should be less likely to formalize. We address the alternative "Greasing" story in two ways. First, our fully specified models control for two forms of regulatory burden at the provincial level (Entry Costs and Post-Entry Regulatory Burden), which are also measured by the Provincial Competitiveness Index. Consequently, these models capture the effect of corruption over and above the provincial regulatory burden. Secondly, in Appendix F Table 22, we test to see whether there is a quadratic relationship between provincial corruption and tax-ID possession. We do not find evidence for a quadratic relationship (inverted U-shape) to support the "Greasing Story."

Except for the "Greasing The Wheel" hypothesis, it is possible that one or more of the remaining mechanisms described above occur simultaneously in Vietnam. However, only the asymmetric information mechanism between tax collectors and household businesses generates a testable prediction. This is because the scale of the asymmetric information problem differs by firm type. Highly visible businesses, whose size and scale of activities are well known, should be less able to hide their existence or performance from tax collectors than less visible businesses. In essence, information on tax-ID registration forms is less valuable for tax collectors when businesses are already more visible. We predict that if the asymmetric information mechanism correctly depicts the causes underlying our empirical results, an increase in corruption would be associated with a smaller increase in tax-ID possession among more visible businesses. We empirically test for this prediction in the next subsection.

5.3. Heterogeneity by firms' visibility

We first define the level of firms' visibility along two measurable dimensions: 1) the number of employees (including the owner); and 2) and the number of operating hours per day. Both of these dimensions would render the household business's existence more obvious to outsiders, specifically to the tax collectors, who can more easily observe commuting patterns of employees and shopping patterns of customers.

We then examine how corruption affects tax-ID possession rates differently along the two firm visibility dimensions discussed above by estimating Eq. 2. In Table 6 and Fig. 3, columns (1) and (3) show regression results for the base regression model. Columns (2) and (4) apply more stringent specifications where, instead of including provincial and communal observables in the regressions, we use commune fixed effects to control for all time-invariant commune's characteristics, including corruption. All four columns show that the interaction terms between *Corruption* × *Visible* are negative and statistically significant.[38] Fig. 3 graphs the marginal effects of corruption in columns (1) and (3) given each visibility level (top panel) and density of visibility variables (bottom panel). From Fig. 3, we see that the marginal effect of corruption on the possession of tax IDs decreases as businesses are more visible (more number of employees or operating longer hours). Thus, an increase in corruption is associated with a smaller increase in tax-ID possession among more visible businesses. Indeed, the relationship between corruption and tax ID possession is only statistically significant among less visible firms.

[38] The results still hold if we exclude firms with more than two employees.

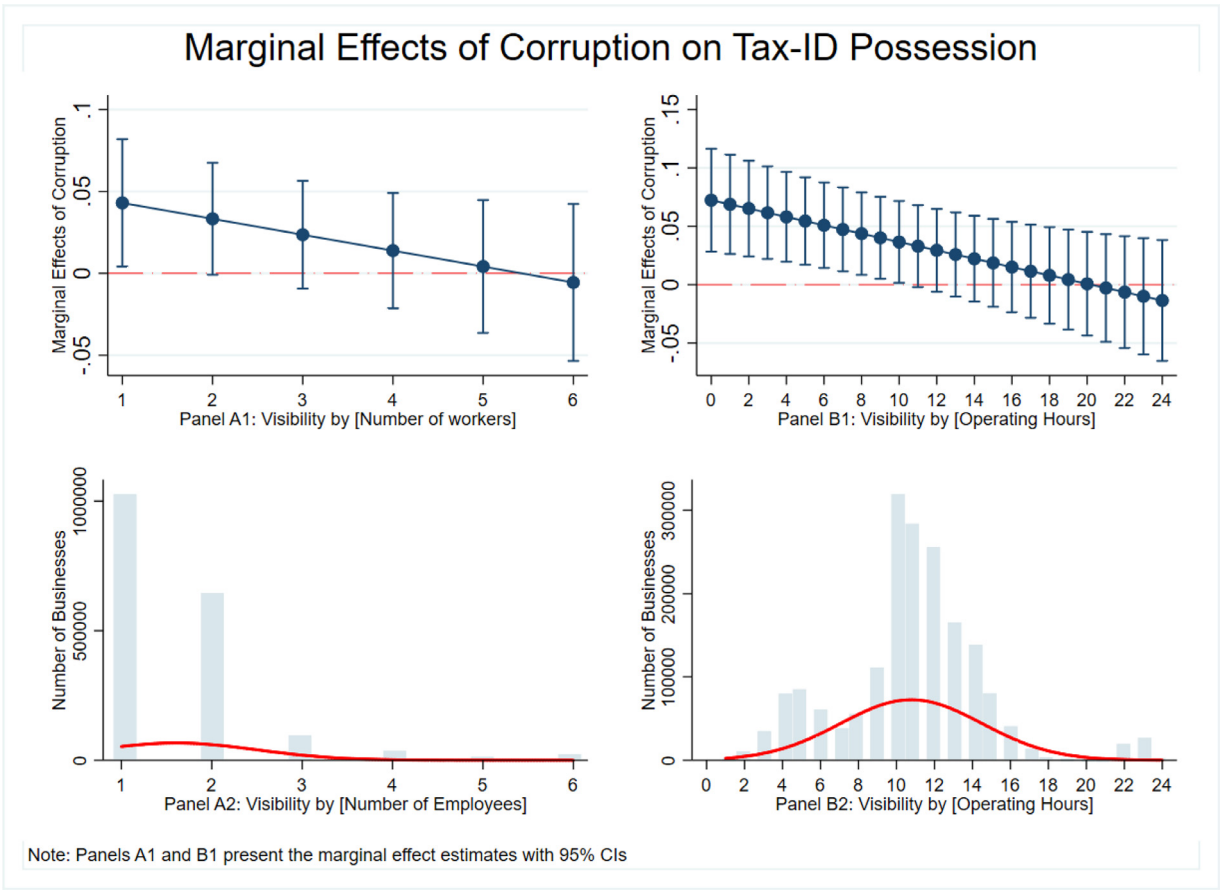


Fig. 3. Marginal Effects of Corruption on Tax-ID Possessions.
Note: The top panel shows the marginal effects and 95% confidence intervals of corruption on tax ID possessions by visibility levels using the regressions in columns (1) and (3) of Table 6. The bottom panel shows the distribution of the visibility variables.

Table 6
Heterogeneity of tax IDs possession likelihood by firms' visibility.

Dependent Variable: tax-ID Possession (0/1)				
Heterogeneity by:	Number of Employees		Operating Hours	
	(1)	(2)	(3)	(4)
Visible	0.114***	0.126***	0.0294***	0.0274***
	(0.0312)	(0.0300)	(0.0069)	(0.00636)
Corruption	0.0527**		0.0724***	
	(0.0234)		(0.0225)	
Corruption × *Visible*	−0.00972*	−0.0119**	−0.00358**	−0.00306**
	(0.00567)	(0.00550)	(0.00138)	(0.00123)
Observations	726,972	781,155	725,168	779,177
Commune FE	No	Yes	No	Yes

Two-way clustered standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
Note: This table shows the results from regression Eq. 2. The dependent variable is a binary indicator of whether a business possesses a tax ID. The set of control variables are the same with the specification in column (5) of Table 3. *Visible* is measured by the number of workers in column (1) and the number of operating hours per day in column (2). Two-way standard errors are clustered at the provincial and provincial border levels.

We use the coefficients in columns (1) and (3) of Table 6 to interpret the magnitude of coefficients of *Corruption* × *Visible* interaction terms. The first thing to notice is the positive and significant coefficient on the individual term *Visible*. This estimate implies that, in the absence of corruption, visible businesses are significantly more likely to have tax IDs.[39] This

[39] Particularly, in the absence of corruption, a business employing an additional worker is 11.4 percent more likely to possess a tax ID (column 1). Likewise, a business that operates an additional hour per day is 2.94% more likely to have a tax ID (column 3).

result is sensible, as there is no bribery cost associated with being transparent for a household business facing no corruption. Secondly, the coefficient on the corruption term is also consistent with our prediction that corrupt tax collectors will try to induce firms to formalize. When visibility is zero, a one point increase in corruption is associated with 5.3% (column 1) and 7.2% (column 3) higher tax-ID possession. Finally, the coefficient on the interaction term *Corruption* × *Visible*, which we interpret as the change in the marginal effect of corruption on tax-ID possession as visibility increases, is negative and statistically significant for both measures of visibility. Specifically, Table 6 shows that a household business employing one additional worker from one to two employees (i.e., a 52% increase from the mean of 1.94 workers; column 1) or operating one additional hour per day from 10 hours to 11 hours (i.e., a 9% increase from the mean of 10.75 h; column 3) is respectively 22.6% and 9.78% less likely to possess a tax ID when the level of provincial corruption increases by one index-point.[40] This result corroborates our theory that tax collectors are attempting to overcome the asymmetric information dilemma they face with household firms which we discussed in the previous sub-section.

We acknowledge that number of employees and number of daily operating hours might not be exogenous to the level of provincial corruption. Nevertheless, the result still provides a window into how the distribution of tax-ID possession likelihood differs across firms characterized by different degrees of visibility as corruption changes. Furthermore, we believe that the heterogeneity results in visibility are not driven by saturation of more visible firms because, on average, only about 20 percent of businesses claim to possess a tax ID.[41] Thus, the heterogeneity results by visibility provide supportive evidence for the asymmetric information mechanism. Critically, however, we cannot rule out the three alternative mechanisms, since we cannot empirically test them.

6. Conclusion

This paper studies the relationship of provincial corruption and tax compliance of household businesses in Vietnam. We examine the sample of businesses that are required to pay taxes by law. Exploiting border discontinuities in the levels of provincial corruption, we compare household businesses located in contiguous commune pairs that straddle provincial borders. Perhaps surprisingly, we find a higher tax registration likelihood among businesses operating in more corrupt provinces. However, we find no evidence that firms pay more in taxes in these locations. In fact, using a sample of firms that answer tax payment questions, we find that an increase in corruption is associated with a significant decrease in the tax compliance ratio among businesses that possess tax IDs.

We present several plausible mechanisms that can explain the two seemingly contradictory results. A mechanism that provides a testable prediction is the asymmetric information mechanism, which states that tax collectors do not have perfect information about businesses' existence or performance to extract bribes. Tax registration forms provide tax collectors with business's information. This mechanism implies that the tax registration form is less valuable for tax collectors when businesses are more visible to the bureaucrats. We provide empirical results consistent with this prediction.

This paper contributes to a growing literature on firm formality. We show that local corruption is a factor that influences an informal business's decision to become formal. Our study also contributes to the firm tax compliance literature that has concentrated mostly on the behavior of formal firms that are usually already registered in the tax system.

One methodological caveat is related to the nature of our cross-sectional setting. Even though we empirically probe for the validity of the excludability assumption by controlling for an extensive set of confounding components, including other observed governance dimensions, we recognize that our measure may not fully isolate corruption from unobserved factors related to institutional quality. However, under the absence of an ideal experiment that would allow us to cleanly infer the causality of corruption on tax compliance, we believe our analysis provides important evidence that illustrates the intricate relationship between these two fundamental policy problems.

Declaration of Competing Interest

We declare that we have no pecuniary or other personal interests, direct or indirect, in any organizations.

Acknowledgments

We would like to especially thank Ha Nguyen for invaluable suggestions. We also thank James Alm, Ritam Chaurey, Pauline Leung, Lara Loewenstein, Ruchi Singh, Sita Slavov, Yukun Sun, Tianyang Xi, Xuan Wang, Abigail Wozniak, and various seminar and conference participants for helpful comments.

[40] From Eq. 2, the marginal effect of corruption at any given visibility level is $\beta_1 + \beta_2 \times visible$. Thus, the 22.6% and 9.78% decreases in tax-ID possession likelihood are obtained from dividing β_2 by $(\beta_1 + \beta_2 \times Visible)$. Specifically, $22.6\% \approx 0.00972/(0.0527 - 0.00972 \times 1)$ (column 1) and $9.78\% \approx 0.00358/(0.0724 - 0.00358 \times 10)$ (column 3).

[41] Specifically, among businesses that operate more than 7 h a day (the majority of firms in our sample), tax-ID possession among these firms fluctuates around 21–24%. On average, businesses with more workers are more likely to have tax-IDs at baseline. However, possession of tax IDs at baseline is small. The percentage of businesses with 1 and 2 workers (the majority of firms in our sample) having tax IDs are 16% and 20%, respectively. Therefore, we are not worried that the possibility of saturation of tax-ID possessions among larger businesses would drive our results.

Appendix A. Business registration versus tax-ID registration

In Vietnam, business registration and tax-ID registration are two separate procedures. A household business needs to obtain a business registration card if its revenue exceeds a threshold set by the provincial government. After a business registers, it has to pay taxes only if its revenue exceeds the threshold set by the central government of 100 million VND. Having tax-IDs is necessary to pay taxes. To obtain tax IDs, businesses need to show business registration cards, but businesses do not need to show tax-IDs when registering their businesses. So practically, all businesses having tax IDs also have business registration cards, but not all businesses with business registration cards (registered businesses) have tax-IDs.

Appendix B. The tax compliance measure: further explanation

Following Section 3.2.2, this section elaborates on the appropriateness of the tax compliance measurement that we employ in this paper. Recall that we follow Alm et al. (2016) and calculate the ratio between a business's tax payment and its total revenue as a measure for the business's tax compliance. How reliable is the use of this ratio as a measure of tax compliance, given that both of the VHBS indicators on tax payment and operating revenue are self-reported? We address potential concerns below.

In terms of the reported tax payment, we believe that firms have no incentive to purposely misreport the tax amount that they officially pay to the government, especially when we know that anecdotally, tax collectors give businesses receipts after they pay their taxes. The receipts only include the tax amount and does not have information on a business's tax rate or revenue. However, some businesses might include the bribe payment to the tax collectors when reporting tax payment in our survey because they might not pay attention to the tax receipts and might not distinguish between the amounts paid as official tax contribution and as bribe. If so, the ratio between tax payment and revenue in the survey will make it appear that corruption increases the size of tax payment, which the opposite of our empirical finding among businesses possessing tax IDs. To that extent, we are confident that our empirical finding of corruption decreasing tax compliance is not driven by businesses misreporting tax payment.

In terms of the reported revenue, as explained in Section 2.1, under-reporting revenue for tax purposes is relatively easier in Vietnam than in developed countries because cash transactions are the norm. In fact, it is the only way that household businesses and tax collectors collude to lower a business's tax payment because there is almost no tax deduction. Nevertheless, we have several reasons to believe that firms are more likely to honestly report their revenue in our survey data, rather than just reporting what legally match with their tax payment to the tax authorities.

First, our data is a survey data and is not the tax administrative data. At the beginning of the survey and before the section asked about taxes, the instructions specifically state that interviewers need to remind business owners that the survey is for statistical and is not for tax purposes. Unlike district tax collectors who might not live in the commune, interviewers are residents in the commune, who would have a better sense of what is going on. Therefore, businesses would be more willing to reveal their actual revenue in our survey data than when reporting to the tax authority.

Second, it is not immediately obvious that businesses know how to create a revenue figure to legally match with their legal tax payment and tax rates. It is because tax collectors decide the tax amount that a business pays at the beginning of the year. The tax receipts that businesses receive after paying taxes only record the amount paid but no revenue figures. In addition, anecdotally, there are many household businesses that do not know their actual tax rates.

Third, under the scenario where businesses know their tax rates and do choose to misreport in the survey (so that the reported revenue is consistent with the amount of tax that they pay), our calculated ratio between tax payment and revenue would then have to be very similar to a business's combined VAT and PIT statutory tax rate (because as Section 2.1 mentioned, the base of the tax rate is the revenue and there is no tax deduction). However, about 90 percent of the calculated tax-revenue ratio in the survey were below a business's combined VAT and PIT statutory tax rate. For example, among businesses in the industry that are supposed to pay 1.5 percent in VAT and PIT, the median of the calculated ratio in our survey is 0.3 percent.

Fourth, it is also possible that businesses in corrupt provinces under-report revenue in the survey more because they purposely evade more in corrupt provinces, so they will be more likely to under-report revenue in the survey to be consistent with their under-reported tax payment. If so, we would see higher ratios between tax payment and revenue in more corrupt places. This implies higher tax compliance in more corrupt places (while in practice those places might have lower tax compliance). Thus, under-reporting revenue in this scenario would bias the results toward a finding that higher corruption is associated with more tax compliance. However, we find that higher corruption is associated with less tax compliance, which makes us less worried about the issue of under-reporting would bias our results.

Finally, besides the self-reported revenue provided by each business, the VHBS survey also asks the interviewers to offer their own independent assessments of what they think the business revenue is. This measure allows us to calculate an alternative ratio between tax payment and revenue as a robustness exercise. Table 7 shows the regression results using the alternative measure of the tax compliance ratio. We see that the results in Table 7 are consistent with the results in Table 5 in the main analysis when we use self-reported revenue. Taken altogether, we believe that the ratio between tax payment and revenue in our data is a reasonable measure of the household business's tax compliance.

Table 7
The effect of corruption on tax compliance ratio, an alternative measure.

Sample	Impute Missing with 0		No Imputation	
	(1)	(2)	(3)	(4)
	All	Have tax ID	All	Have tax ID
Corruption	−0.00110**	−0.00203**	−2.78e−05	−0.00169*
	(0.000475)	(0.000897)	(0.000353)	(0.000911)
Observations	15,378	6,097	23,478	6,329

Note: This table shows the results from regression Eq. 1. The dependent variable is tax compliance ratio, measured by the ratio between tax payment and revenue independently assessed by surveyors. Control variables are the same as control variables in column (5) Table 3. Two-way standard errors are clustered at the provincial and provincial border levels.

Appendix C. Robustness checks: alternative samples and indicators

C1. Robustness check with provincial level OLS regressions

Table 8
Estimations with provincial level regressions.

	tax-ID possession (1)	tax-ID possession (2)	tax-compliance ratio (3)	tax-compliance ratio (4)
Corruption	0.0252	0.0351	−0.00109**	−0.00132*
S.E.	(0.0154)	(0.0234)	(0.000514)	(0.000772)
Coef's p-value	1.636	1.500	−2.121	−1.710
R-squared	0.043	0.307	0.070	0.336
Observations	62	62	62	62
Other governance controls	No	Yes	No	Yes

Robust standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
Note: This table shows the provincial level regression outputs as a robustness check to the main results in Tables 3–6.

C2. Robustness check with Tobit regressions

Table 9
Estimations with Tobit model.

	Dep. Var: Tax Compliance Ratio	
	(1)	(2)
Corrupt	−0.00192***	−0.00176***
	−0.000309	−0.000295
Constant	0.0251***	0.0235***
	−0.00171	−0.0065
Observations	5,116	5,116
Model	Tobit	Tobit
Sample	Firms with Tax IDs at border	Firms with Tax IDs at border
1-d Industry FE	No	Yes

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
Note: This table shows outputs for the tax-compliance ratio outcome variable from Tobit regressions. The model uses the same sample as in the main text. To avoid over-specification bias, the regression controls for 1-digit industry FE (column 2).

C3. Robustness checks with proximity pairings

Recall that the main specification in this paper includes firms operating in contiguous commune pairs straddling a provincial border. That is, two communes in a specific eligible pair are always being 0-kilometer away from each other in the nearest-possible-distance sense. In this exercise, we test for the robustness of our main result by relaxing the strict underlying assumption on locational proximity between communes. Specifically, we reconstruct our robustness regression samples that allow for the eligible paired communes to locate within 1-km, 2-km, and 3-km from each other, again in nearest possible distance sense. Thus, our main specification-pairing communes that are adjacent to each others-can be thought

Table 10
The effects of corruption on possession of Tax IDs.

	All Provinces			Drop HN & HCM			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Corruption	0.0367**	0.0302	0.0330*	0.0556***	0.0545***	0.0511***	0.0555***
	(0.0174)	(0.0187)	(0.0183)	(0.0176)	(0.0178)	(0.0163)	(0.0157)
Observations	1,014,823	1,713,015	2,703,889	631,949	872,932	1,460,557	2,310,229
R-squared	0.198	0.208	0.205	0.211	0.209	0.212	0.211
Sample	≤ 1 km	≤ 2km	≤ 3 km	0 km	≤ 1 km	≤ 2 km	≤ 3 km

Two-way clustered standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
Note: This tables shows the results from regression Eq. 1. The dependent variable is a binary indicator equalling one if the business possesses a tax ID. Control variables are the same as control variables in column (5) Table 3. Columns (4) through (7) exclude Hanoi and Ho Chi Minh city from the regressions. Columns (1) and (5) include firms in communes that are least 1 km away from each other. Columns (2) and (6) include firms in communes that are least 2 km away from each other. Columns (3) and (7) include firms in communes that are least 3 km away from each other. Two-way standard errors are clustered at the provincial and provincial border levels.

Table 11
Robustness results for alternative pairwise regression samples: Tax Compliance Ratio–No Imputation.

Panel A: All firms				
	(1)	(2)	(3)	(4)
Corruption	−0.00165***	−0.00166***	−0.00156**	−0.00150*
	(0.000597)	(0.000608)	(0.000676)	(0.000807)
Observations	21,447	36,684	58,428	13,898
Panel B: Firms with Tax IDs				
Corruption	−0.00246**	−0.00309***	−0.00272**	−0.00231**
	(0.00112)	(0.00108)	(0.00115)	(0.00109)
Observations	8281	13,488	21,713	5562
Sample	≤ 1 km	≤ 2 km	≤ 3 km	0 km; drop HN & HCM

Two-way clustered standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
Note: This tables shows the results from regression Eq. 1. The dependent variable is the ratio between tax payment and tax revenue (missing values are dropped). The set of control variables is the same as in column (5) in Table 3. Columns (4) excludes businesses in Hanoi and Ho Chi Minh city. Two-way standard errors are clustered at the provincial and provincial border levels.

Table 12
Robustness results for alternative pairwise regression samples: Tax Compliance Ratio–Imputed sample.

Panel A: All firms				
	(1)	(2)	(3)	(4)
Corruption	−0.00035	−0.00045	−0.000456	−0.000269
	(0.00041)	(0.00043)	(0.00045)	(0.00047)
Observations	32,530	54,826	86,607	21,658
Panel B: Firms with Tax IDs				
Corruption	−0.00187*	−0.00237**	−0.00216**	−0.00188*
	(0.00110)	(0.00101)	(0.00105)	(0.00109)
Observations	8598	14,032	22,545	5758
Sample	≤ 1 km	≤ 2 km	≤ 3 km	0km; drop HN & HCM

Two-way clustered standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
Note: This table shows results from regression Eq. 1. The dependent variable is the ratio between tax payment and revenue. Missing tax and/or revenue observations are imputed with 0. The set of control variables is the same as in in column (5) in Table 3. Columns (4) excludes businesses in Hanoi and Ho Chi Minh city from the regressions. Two-way standard errors are clustered at the provincial and provincial border levels.

of as a special sub-sample of each of the above robustness samples. In Table 10, we also show the results when we drop Hanoi and Ho Chi Minh city (the two major cities in Vietnam) from the regression. As Tables 10 and 11 show, the results in the main analyses generally hold and are consistent across different samples.

C4. Robustness check with alternative revenue thresholds

C5. Robustness check with tax payment indicator

Table 13
Estimations with Alternative Revenue Thresholds—Tax-ID Possession.

	Dependent Variable: Possession of Tax ID				
	(1)	(2)	(3)	(4)	(5)
Corruption	0.0480***	0.0454***	0.0406***	0.0409***	0.0377**
	(0.0117)	(0.0116)	(0.0105)	(0.00959)	(0.0172)
Observations	852,675	852,672	792,245	771,538	771,538
Revenue Threshold			$\geq$ *90 million VND*		
Corruption	0.0397***	0.0374***	0.0324***	0.0322***	0.0277*
	(0.0112)	(0.0110)	(0.0103)	(0.00936)	(0.0161)
Observations	1,062,206	1,062,202	985,543	957,936	957,936
Revenue Threshold			$\geq$ *70 million VND*		
Corruption	0.0340***	0.0320***	0.0274***	0.0269***	0.0223
	(0.0108)	(0.0105)	(0.0100)	(0.00923)	(0.0150)
Observations	1,257,471	1,257,467	1,165,179	1,130,195	1,130,195
Revenue Threshold			$\geq$ *50 million VND*		
Commune-pair FE	Yes	Yes	Yes	Yes	Yes
3-digit Industry FE	Yes	Yes	Yes	Yes	Yes
Firm's controls			Yes	Yes	Yes
Commune's policy controls				Yes	Yes
Commune's GIS controls				Yes	Yes
Night-light Index (2016)				Yes	Yes
Geographic controls				Yes	Yes
Formal sector activities Controls					Yes
Governance Controls					Yes

Robust standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
Note: This table reports robustness-check results from regressions where revenue threshold is set at $\geq$50 million, $\geq$70 million, and $\geq$90 million VND. The outcome variable tax-ID possession likelihood.

Table 14
Estimations with Alternative Revenue Thresholds—Tax-compliance Ratio.

	Dependent Variable: tax-compliance ratio					
Sample:	All			Having tax-ID		
	(1)	(2)	(3)	(4)	(5)	(6)
Corruption	−0.00136	−0.00147*	−0.00157***	−0.00125	−0.00182	−0.00310***
	(0.000876)	(0.000737)	(0.000578)	(0.00132)	(0.00128)	(0.00110)
R-squared	0.341	0.341	0.340	0.413	0.415	0.427
Observations	20,869	18,792	16,054	7545	7020	6280
Rev Threshold	$\geq$50M	$\geq$70M	$\geq$90M	$\geq$50M	$\geq$70M	$\geq$90M

Robust standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
Note: This table reports robustness-check results from regressions where revenue threshold is set at $\geq$50 million, $\geq$70 million, and $\geq$90 million VND. The outcome variable is tax-compliance ratio.

Table 15
The Effects of Corruption on Tax Payment (alternative indicator).

	Dropped Missing Values		Imputed Sample	
Sample:	All	have tax ID	All	have tax ID
	(1)	(2)	(3)	(4)
Corruption	−107.2	279.2	138.7	407.8
	(201.8)	(475.9)	(153.6)	(441.6)
Observations	15,385	6109	23,484	6341

Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
Note: This table shows results from regression Eq. 1. The dependent variable is tax payment in absolute term. Columns (1) and (2) exclude missing tax and/or revenue observations. Columns (3) and (4) imputes missing values with 0. The set of control variables is the same as in column (5) in Table 3. Two-way standard errors are clustered at the provincial and provincial border levels.

C6. Robustness check with corruption indices from PCI 2015 and PCI 2016

Table 16
The Effect of Corruption on Tax-ID Possession and Tax Compliance Ratio.

	Tax-ID (1)	Tax Compliance Ratio (2)	Tax-ID (3)	Tax Compliance Ratio (4)
Corruption 2015	0.0302***	−0.000671*		
	(0.0113)	(0.000361)		
Corruption 2016			0.0446***	0.000104
			(0.0146)	(0.000348)
Observations	781,161	6939	781,161	6939
R-squared	0.189	0.440	0.190	0.439
commune-pair FE	Yes	Yes	Yes	Yes
3-digit Industry FE	Yes	Yes	Yes	Yes

Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
Note: Robustness checks to the main result. Regressions use *Corruption* scores constructed using PCI 2015 (columns 1–2) and PCI 2016 (columns 3–4).

Table 17
Smoothness of geographical characteristics at provincial border.

	(1) Area	(2) Elevation	(3) Slope	(4) Rainfall	(5) Temperature	(6) Light
Higher Score	−6.241	−30.57	−1.439	−0.486	0.132	−0.335
	(4.130)	(35.94)	(1.343)	(2.900)	(0.198)	(1.347)
Observations	9646	9172	9172	9169	9169	9646
R-squared	0.205	0.264	0.333	0.038	0.622	0.017
Commune Controls	Yes	Yes	Yes	Yes	Yes	Yes
Sample	At Border	At Border	At Border	At Border	At Border	At Border

Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
Note: This table shows results from regression Eq. D.1. The outcome variables are communes' geographic (area, average elevation, average terrain steepness) and average climate-related indicators (temperature and rainfall). Robust standard errors clustered at the provincial level.

Appendix D. Threats to causal identification

First, one reasonable concern could be that neighboring border communes belonging to two different provinces can face different aspects of governance quality other than corruption. To alleviate as much of this concern as possible, we include all nine other PCI subindices listed in Section 3 in our regressions to control for different aspects of governance quality. For example, one might think that corrupt local officials audit more frequently because they want to extract more rents. To address this, we control for the "Time Costs" subindex, which includes components such as the frequency of inspections or inspection hours. Even though we think that the PCI subindices capture the most important aspects of governance quality that affect the business environment, we recognize that they are not exhaustive. Thus, without a randomized control trial or an exogenous policy change in corruption, provincial corruption in our case may represent general institutional quality at the province level.

Second, since provincial borders could be set by natural landmarks, such as a river or a mountain, some might worry that geographic conditions could be different for neighboring communes across the borders, which may differentially affect the economic activities and governance. We address this concern by controlling for a host of communal geographical characteristics: area (in km square), latitude, longitude, elevation, temperature, precipitation, slope, whether a river runs through the commune, and the degree to which the commune is suitable for rice, tea, coconut, sugar, and coffee. In addition, we test for whether the geographical characteristics are similar in communes in less corrupt provinces versus communes in more corrupt provinces. Specifically, we run the following regression equation that employs a restricted sample of only the firms located in the border communes:

$$characteristics_{commune} = \alpha_0 + \alpha_1 HigherScore_p + \epsilon_{commune} \tag{D.1}$$

where $HigherScore_p$ is an indicator that equals 1 if a commune is in a province with a higher corruption level than its neighboring commune. If the two neighboring communes with different provincial corruption levels have different geographical characteristics, we would expect α_1 to be statistically different from 0. Tables 17 and 18 show that the two neighboring communes in two different provinces have virtually identical geographical characteristics and agricultural suitability.

Third, a possible confounding factor could be that different provinces have different economic policies to incentivize investment, which may be associated with both corruption and the probability that a firm obtains a tax ID. To control for the possible bias caused by different spatial economic policies, we control for variables indicating the presence of industrial and economic zones as well as economically disadvantaged areas that receive preferential tax treatment and transfers.

Fourth, local governance may affect local infrastructure and economic activities, which in turn may affect household business's tax compliance. We control for local infrastructure and economic activities such as 2016 night-time light brightness, number of enterprises, median assets, number of employees, and profits of enterprises at the commune level in 2014 in our regressions. We also control for household business characteristics such as owner's gender, age, education level, and the type of location where the firm operates (e.g., at the owner's house or the local market).

Table 18
Smoothness of agricultural endowments (i.e., Agricultural suitability) at provincial borders.

	(1) Tea	(2) Rice	(3) Coconut	(4) Coffee	(5) Sugar
Higher Score	0.0406	1.700	−0.433	−0.500	1.364
	(1.834)	(1.541)	(0.685)	(0.633)	(1.236)
Observations	9172	9172	9172	9172	9172
R-squared	0.305	0.390	0.773	0.590	0.202
Commune Controls	Yes	Yes	Yes	Yes	Yes
Sample	At Border	At Border	At Border	At Border	At Border

Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
Note: This table shows results from regression Eq. D.1. The outcome variables are communes' agricultural endowments based on land and soil structures (i.e., the level of suitability in growing agricultural products such as tea, rice, coconut, coffee, and sugar). Robust standard errors clustered at the provincial level.

Table 19
A test for whether firms move to a less corrupt province.

	Dependent Variable: Firm Count (log)				
	(1)	(2)	(3)	(4)	(5)
At Border Commune	−0.0163	−0.0555	−0.0473	−0.0504	−0.0392
	(0.0312)	(0.0477)	(0.0377)	(0.0386)	(0.0398)
Observations	7746	1399	3012	6658	10,224
R-squared	0.512	0.538	0.469	0.429	0.408
Commune Controls	Yes	Yes	Yes	Yes	Yes
Border FE	Yes	Yes	Yes	Yes	Yes
Cluster	2-way	2-way	2-way	2-way	2-way
Bandwidth	neighbor	5 km	10 km	20 km	30 km

Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
Note: This table shows the results from regression Eq. D.2. The dependent variable is the number of firms in a commune (log-transformed). The predictor variable is a dummy equaling 1 if the firm locates in a border commune. Two-way standard errors are clustered at the provincial and provincial border levels.

Fifth, some may have concerns about reverse causality. It could be that household business characteristics determine both provincial governance as well as tax-ID and liabilities status. For example, better-run household businesses could also have more collective power to change local public-sector governance and corruption. These better-run household businesses are more likely to obtain a tax ID or pay more taxes. However, in Vietnam, this is unlikely to be the case. Household businesses in a province do not have much of a collective voice in the decision-making process of the provincial policy makers. Furthermore, in Vietnam's single-party system, provincial leaders are not popularly elected by, and hence do not respond to, the local population. Rather, they are always appointed by, promoted by, and report to the central government in Hanoi (Vu, 2016b; Painter, 2003).

Sixth, firms could strategically move across a border to avoid corruption. This is, however, highly unlikely within the context of Vietnamese household businesses. Recall that our sample includes predominantly tiny household businesses, each employing under two workers on average (including the owner). About 70% of businesses operate inside their premises (Table 2). It is not likely that these businesses would move out of their houses and potentially rent business spaces in a different province to do business in response to corruption. In addition, the conjecture that businesses move across provinces because of corruption has a clear observable implication. If firms move to a less corrupt province, communes at the border of a less corrupt province would have a higher firm density than the interior communes in that province; the most cost-effective option for firms to minimize moving costs is to simply shift business activities right across the provincial border. Thus, to test if firms move across the border because of corruption, we restrict the sample to communes in less corrupt provinces. We then regress the log-transformed count of household businesses, controlling for the commune area in km square, on a dummy variable that equals 1 if the commune locates at the border and 0 if the commune neighbors a border commune or is within 5, 10, 20, or 30 km of a commune's centroid to the border (but not the border communes). The regression equation is:

$$Logfirm Count_c = \lambda_0 + \lambda_1 CommuneAtBorder_c + \lambda_2 CommuneArea + X_c + \epsilon_c \quad (D.2)$$

where $CommuneAtBorder_c$ is a dummy equal to 1 if the commune locates at the provincial border. Control variables X_c in this regression are the same as control variables at the commune level in regression 1 and 2. If firms move across the border, we expect λ_1 to be greater than 0. Table 19 shows that, controlling for commune's area, the number of household businesses in the border communes of a less corrupt province are not statistically higher than the interior communes in that province (i.e., λ_1 is not statistically greater than 0).

Appendix E. The "seemingly contradictory" results of corruption are not likely the artifact of being based on two different datasets.

A reasonable concern is that the "seemingly contradictory" results of corruption on tax-ID possession and tax compliance ratio are due to the fact that they are based on two different datasets. In this subsection, we provide our explanation and several empirical tests to alleviate that concern.

Specifically, we first elaborate on the sampling design of VHBS, and explain how selection bias should not be a threat to the detailed tax-survey sample. Consequently, our findings on tax-ID possession and tax-compliance ratio should not be driven by selection into the tax-survey sample. We then perform an extrapolation exercise, which suggests that if we were to observe tax compliance ratio for all businesses in the VHBS's border communes, the association between corruption and the tax compliance ratio might even be more negative. In this sample, the association between corruption and tax-ID possession is positive. Hence, our respective findings on the relationships between 1) corruption and tax-ID registration (positive), and 2) between corruption and tax-compliance ratio (negative) should remain valid.

E1. No selection bias in the detailed tax-survey sample

In addition to its survey on tax-ID statuses of all household businesses in the country (i.e., the VHBS census), the General Statistical Office of Vietnam (GSO) also chooses a subset of businesses to ask a tax-liability question. This is the sample we use to calculate the tax compliance ratio in the paper. This sample is supposed to be randomly selected, so that the GSO can meaningfully study the current statuses and trends in different detailed dimensions of firm's tax liability in the country. We do not know of any anecdotal evidence or reason that businesses would self-select into this detailed tax-survey.

Second, Table 20 below shows that provincial corruption is not statistically associated with the probability of being included in the tax-survey sample (column 1), nor with the item nonresponse rate on tax-liability questions (conditional on the firm possessing a tax-ID and being selected for the tax-survey sample; column 2). Therefore, the likelihood of being selected for the tax-survey or not responding to tax-liability questions are not be systematically correlated with the treatment variable ("Corruption"), and thus should not bias the observed effects of corruption on the tax compliance ratio.

E2. The extrapolation exercise

While we are comfortable that selection bias is not driving the differential results, we do find that our focus on border communes does lead to some differences in observable characteristics of the firm samples (tax-survey versus VHBS census in the border communes). As we show in the summary statistics (Table 2), businesses in the VHBS in border communes on average have fewer workers and are less likely to be in manufacturing than those in the tax-survey sample. There are also differences across businesses in the two samples in the restaurant-industry proportion, male ownership, background training of firm owners, and the share of businesses operating at the owners' premises. Note that Table 2 summarizes all firm control variables that are available in the dataset.

Below, we empirically test for the interaction effects of these firm-specific characteristics and corruption on the tax-compliance outcome using the tax-survey sample. This exercise should shed light on the direction of changes in our estimated coefficients if we were to observe firms' tax liability (and thus were able estimate corruption effect on tax compliance ratio) for the whole census. To do so, we run Eq. 2 with the tax compliance ratio as the dependent variable. We are interested in the coefficients of the interaction terms between corruption and all available firm observable characteristics.

Using the tax-survey sample, Table 21 shows that corruption tends to have a less pronounced negative effect on the tax compliance ratio among businesses with more workers (column 1) and businesses in manufacturing (column 3). We find no evidence of heterogeneity on the relationship between corruption and tax compliance ratio along other dimensions of firm observables upon which the tax-survey sample and the VHBS Census for border communes differ.

Building off of this finding, it can be reasonably extrapolated that if corruption decreases the tax-compliance ratio in the tax-survey sample, it should also decrease the tax-compliance ratio in full census for border communes. In fact, the negative

Table 20
Corruption, inclusion in the tax-survey, and nonresponse rate.

	(1) in the tax-survey sample	(2) nonresponse
Corruption	−0.00215	0.000156
	(0.00210)	(0.0144)
Observations	879,458	5642
R-squared	0.000	0.000
Sample	at the border	border tax survey& taxid=1

Robust standard errors in parentheses $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors are clustered at the province levels. The dependent variable in column (1) equals 1 if a business is in the tax-survey sample and 0 otherwise. Column (2) consists of tax-registered businesses that are in the tax-survey sample. The dependent variable in column (2) equals 1 a business does not answer the tax-liability question and 0 otherwise.

Table 21
Interaction effects of corruption on Tax compliance ratio.

	(1) #workers	(2) restaurant	(3) manufacturing	(4) owner is male	(5) owner has no training	(6) operate in owner's house
Interaction w/corruption	0.000750***	0.000836	0.00301***	0.000578	−0.000499	−7.71e−05
	(0.000257)	(0.00140)	(0.00102)	(0.000764)	(0.000649)	(0.000575)
Observations	6109	6109	6109	6109	6109	6109

Note: Two-way clustered standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The sample consists of household businesses with tax IDs that answer tax liability questions. This table shows results from regression Eq. 2. The dependent variables are in the tax compliance ratio. The set of control variables is the same as in column (5) in Table 3. Reported in the table are the estimated coefficients associated with the interaction terms between corruption and respective business characteristics in the column headings (for all available observable characteristics that we have in the data). Two-way standard errors are clustered at the provincial and provincial border levels.

effect might even be larger due to the fact that an average business in the VHBS census is slightly smaller and less likely to be in manufacturing than one in the tax survey. This is another piece of supporting evidence to our claim that corruption can have inverse consequences on tax registration and tax compliance ratio, i.e., a "seemingly contradictory" result that we theoretically explain for in the mechanism section.

Appendix F. Other graphs and tables

F1. No Bunching Around Revenue Threshold

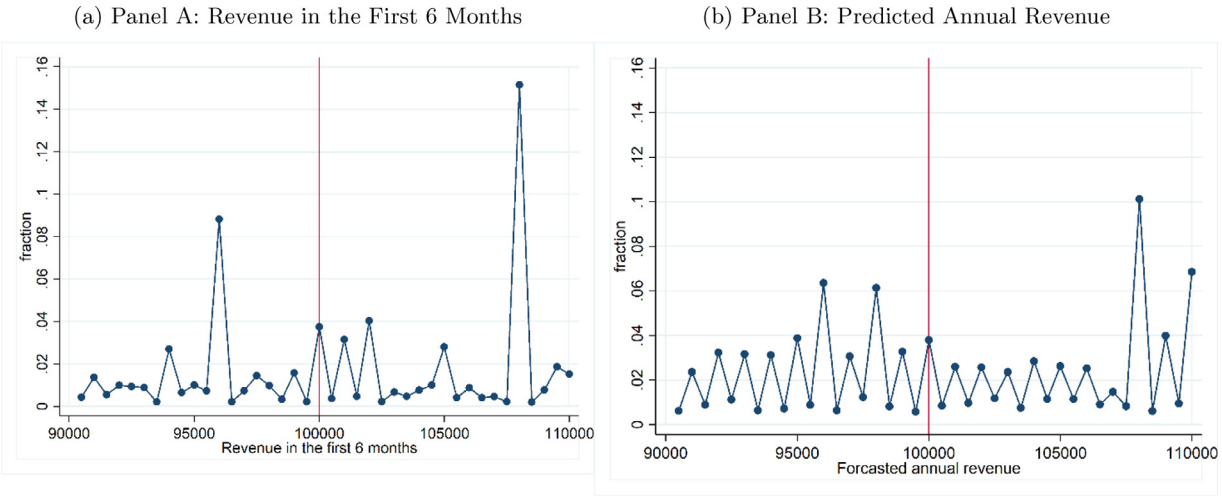


Fig. 4. Distribution of Revenue Around the 100 Million VND Threshold.
Note: this figure shows the distribution of businesses by reported revenue in the first six month (Panel A) and forecasted revenue in the last six months of 2017 (Panel B). In this figure, we do not see evidence of manipulation around the 100 million VND revenue threshold.

F2. Testing the "Greasing The Wheels" Hypothesis

Table 22
Testing the "Greasing The Wheels" Hypothesis.

	=1 if have a tax ID (1)	=1 if have a tax ID (2)
Corruption	−0.0397	−0.0493
	(0.126)	(0.153)
Corruption squared	0.00843	0.00842
	(0.0120)	(0.0143)
Observations	781,161	726,972
R-squared	0.191	0.204

Note: Two-way standard errors are clustered at the provincial and provincial border levels. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Column (1) only includes provincial border fixed effects. Column (2) includes the same set of control variables as those in column (5) in Table 3.

References

Alm, J., Blackwell, C., McKee, M., 2004. Audit selection and firm compliance with a broad-based sales tax. Natl. Tax J. 57 (2, Part 1), 209–227. doi:10.17310/ntj.2004.2.04.

Alm, J., Embaye, A., 2013. Using dynamic panel methods to estimate shadow economies around the world, 1984–2006:. Publ. Financ. Rev. doi:10.1177/1091142113482353.

Alm, J., Martinez-Vazquez, J., McClellan, C., 2016. Corruption and firm tax evasion. J. Econ. Behav. Org. 124, 146–163. doi:10.1016/j.jebo.2015.10.006.

Almeida, R., Carneiro, P., 2009. Enforcement of labor regulation and firm size. J. Comp. Econ. 37 (1), 28–46. doi:10.1016/j.jce.2008.09.004.

Boly, A., 2018. On the short- and medium-term effects of formalisation: panel evidence from Vietnam. J. Dev. Stud. 54 (4), 641–656. doi:10.1080/00220388.2017.1342817.

Bruhn, M., 2010. License to sell: the effect of business registration reform on entrepreneurial activity in Mexico. Rev. Econ. Stat. 93 (1), 382–386. doi:10.1162/REST_a_00059.

Cameron, A.C., Gelbach, J.B., Miller, D.L., 2011. Robust inference with multiway clustering. J. Bus. Econ. Stat. 29 (2), 238–249. doi:10.1198/jbes.2010.07136.

Cameron, A.C., Miller, D.L., 2015. A practitioner's guide to cluster-robust inference. J. Hum. Res. 50 (2), 317–372. doi:10.3368/jhr.50.2.317.

Central Institute for Economic Management, 2017. Executive summary - a study on formalization of household business in vietnam. [Online]. Available from: http://www.mekongbiz.org/wp-content/uploads/2017/07/Executive-Summary-Formalization-of-household-business-in-Vietnam.pdf.

Chen, S., Chen, X., Cheng, Q., Shevlin, T., 2010. Are family firms more tax aggressive than non-family firms? J. Financ. Econ. 95 (1), 41–61. doi:10.1016/j.jfineco.2009.02.003.

Dabla-Norris, E., 2000. A Game-Theoretic Analysis of Corruption in Bureaucracies, 2000–2106. IMF eLibrary. ISBN: 9781451852882 Library Catalog: www.elibrary.imf.org

Desai, M.A., Dharmapala, D., 2006. Corporate tax avoidance and high-powered incentives. J. Financ. Econ. 79 (1), 145–179. doi:10.1016/j.jfineco.2005.02.002.

Do, Q., Iyer, L., 2008. Land titling and rural transition in Vietnam. Econ. Dev. Cult. Change 56 (3), 531–579. doi:10.1086/533549.

Dreher, A., Gassebner, M., 2013. Greasing the wheels? the impact of regulations and corruption on firm entry. Public Choice 155 (1), 413–432. doi:10.1007/s11127-011-9871-2.

Dube, A., Lester, T.W., Reich, M., 2010. Minimum wage effects across state borders: estimates using contiguous counties. Rev. Econ. Stat. 92 (4), 945–964. doi:10.1162/REST_a_00039.

Dyreng, S., Hanlon, M., Maydew, E.L., 2010. The effects of executives on corporate tax avoidance. Account. Rev. 85 (4), 1163–1189.

Fajnzylber, P., Maloney, W.F., Montes-Rojas, G.V., 2011. Does formality improve micro-firm performance? Evidence from the Brazilian SIMPLES program. J. Dev. Econ. 94 (2), 262–276. doi:10.1016/j.jdeveco.2010.01.009.

Ferraz, C., Finan, F., Moreira, D.B., 2012. Corrupting learning: evidence from missing federal education funds in Brazil. J. Publ. Econ. 96 (9), 712–726. doi:10.1016/j.jpubeco.2012.05.012.

Fisman, R., Svensson, J., 2007. Are corruption and taxation really harmful to growth? Firm level evidence. J. Dev. Econ. 83 (1), 63–75. doi:10.1016/j.jdeveco.2005.09.009.

Graham, J.R., Tucker, A.L., 2006. Tax shelters and corporate debt policy. J. Financ. Econ. 81 (3), 563–594. doi:10.1016/j.jfineco.2005.09.002.

Hoang, K.L., Nguyen, T.H., 2018. Vietnam public expenditure review in the period of 2007–2017 and some recommendations. VNU J. Sci.: Econ. Bus. 34 (4).

Holland, A.C., 2015. The distributive politics of enforcement. Am. J. Polit. Sci. 59 (2), 357–371. doi:10.1111/ajps.12125. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajps.12125

Holland, A.C., 2016. Forbearance. Am. Polit. Sci. Rev. 110 (2), 232–246. doi:10.1017/S0003055416000083.

Hsieh, C.-T., Olken, B.A., 2014. The missing "missing middle". J. Econ. Perspect. 28 (3), 89–108. doi:10.1257/jep.28.3.89.

Jeong, D., Shenoy, A., Zimmermann, L., 2019. Are Transparency and Accountability Enough? Open Corruption and Why It Exists. Working Paper, 41.

Kaplan, D.S., Piedra, E., Seira, E., 2011. Entry regulation and business start-ups: evidence from Mexico. J. Publ. Econ. 95 (11), 1501–1515. doi:10.1016/j.jpubeco.2011.03.007.

Kerkvliet, T.J., Marr, D.G., 2004. Beyond Hanoi: Local Government in Vietnam. ISEAS / NIAS. Library Catalog: bookshop.iseas.edu.sg

Khan, A.Q., Khwaja, A.I., Olken, B.A., 2016. Tax farming redux: experimental evidence on performance pay for tax collectors. Q. J. Econ. 131 (1), 219–271. doi:10.1093/qje/qjv042.

Laffont, J.-J., N'Guessan, T., 1999. Competition and corruption in an agency relationship. J. Dev. Econ. 60 (2), 271–295. doi:10.1016/S0304-3878(99)00056-5.

Malesky, E., 2009. Gerrymandering-Vietnamese style: escaping the partial reform equilibrium in a nondemocratic regime. J. Polit. 71 (1), 132–159. doi:10.1017/s0022381608090099.

Malesky, E., Phan, T.N., Pham, N.T., 2019. The Vietnam Provincial Competitiveness Index: Measuring Economic Governance for Private Sector Development, 2018 Final Report. Vietnam Chamber of Commerce and Industry and United States Agency for International Development: Hanoi, VN.

McKenzie, D., Seynabou Sakho, Y., 2010. Does it pay firms to register for taxes? The impact of formality on firm profitability. J. Dev. Econ. 91 (1), 15–24. doi:10.1016/j.jdeveco.2009.02.003.

de Mel, S., McKenzie, D., Woodruff, C., 2013. The demand for, and consequences of, formalization among informal firms in Sri Lanka. Am. Econ. J.: Appl. Econ. 5 (2), 122–150. doi:10.1257/app.5.2.122.

Méon, P.-G., Weill, L., 2010. Is corruption an efficient Grease? World Dev. 38 (3), 244–259. doi:10.1016/j.worlddev.2009.06.004.

Nguyen, H.T., 2019. How large is Vietnam's informal economy? Econ. Affairs 39 (1), 81–100. doi:10.1111/ecaf.12328. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ecaf.12328

Painter, M., 2003. Public administration reform in vietnam: problems and prospects. Publ. Admin. Dev. 23 (3), 259–271.
Rand, J., Tarp, F., 2012. Firm-level corruption in Vietnam. Econ. Dev. Cult. Change 60 (3), 571–595. doi:10.1086/664022. Publisher: The University of Chicago Press
Rand, J., Torm, N., 2012. The benefits of formalization: evidence from Vietnamese manufacturing SMEs. World Dev. 40 (5), 983–998. doi:10.1016/j.worlddev.2011.09.004.
Reinikka, R., Svensson, J., 2005. Fighting Corruption to Improve Schooling: evidence from a Newspaper Campaign in Uganda. J. Eur. Econ. Assoc. 3 (2–3), 259–267. doi:10.1162/jeea.2005.3.2-3.259.
Rocha, R., Ulyssea, G., Rachter, L., 2018. Do lower taxes reduce informality? Evidence from Brazil. J. Dev. Econ. 134, 28–49. doi:10.1016/j.jdeveco.2018.04.003.
Truong, T., Tran, T., Pham, D., Le, B., 2013. The non-farm household business sector in vietnam. [Online]. Available from: https://www.economica.vn/Portals/0/Documents/fd921a86edc2a4407dc0ad2543d0e92d.pdf.
Ulyssea, G., 2018. Firms, informality, and development: theory and evidence from Brazil. Am. Econ. Rev. 108 (8), 2015–2047. doi:10.1257/aer.20141745.
Vu, T.T.A., 2016. Vietnam: decentralization amidst fragmentation. J. Southeast Asian Econ. 33 (2), 188–209. Publisher: Institute of Southeast Asian Studies (ISEAS)
Vu, T.T.A., 2016. Vietnam: decentralization amidst fragmentation. J. Southeast Asian Econ. 188–208.
Weaver, J., 2018. Jobs for Sale: Corruption and Misallocation in Hiring. Working Paper.