Value-Added Tax, Cascading Sales Tax, and Informality

Yue Zhou*

November 2020

Abstract

Informality is large and persistent in developing countries. This paper studies whether replacing a cascading sales tax with a value-added tax (VAT) can contribute to reducing informality. The paper builds a structural model that allows identifying the effect of VAT adoption on informality. It then estimates the model using a representative sample of formal and informal Indian manufacturing firms. The paper shows that the effect of VAT adoption on the extensive margin on formalization is small (only 0.6% of informal firms register after the introduction of VAT). However, VAT adoption has a large effect on formal sector output (the intensive margin), which expands by nearly 12%. If output is kept constant, adoption of VAT would lead to a loss of revenues, compared with the cascading sales tax. However, the paper shows that VAT adoption increases labor income and formal sector profits. This formalization process, in turn, leads to higher corporate and income tax revenues, which offset the other revenue losses. The results of the paper support the idea that VAT adoption favors more productive formal firms and contributes to growth-enhancing structural transformation.

Keywords: value-added tax; informality; registration; development.

JEL codes: D22, E26, H21, H25, H26.

* The Graduate Institute Geneva, Geneva, 1202. Email: yue.zhou@graduateinstitute.ch.
I am extremely grateful to Yi Huang and Ugo Panizza for the invaluable guidance, supervision, and encouragement. I also thank Maurizio Bussolo, Luming Chen, Ding Ding, Paul Mathieu, Tianshi Mu, Siddharth Sharma, Hans Timmer for helpful comments and suggestions, and Sandra Gain for editorial help. All errors remain my own.

1. Introduction

As informal firms do not pay taxes, a large informal sector constrains domestic revenue mobilization and tends to favor small, low-productivity firms. Thus, tackling informality can contribute to promoting state capacity and growth-enhancing structural change.

This paper uses a structural model and Indian data to study how the adoption of a value-added tax (VAT) affects informality. To the best of my knowledge, this is the first paper that quantifies the effect on informality of a VAT tax reform. A key finding of the paper is that most informal firms are too unproductive to react to tax policy changes. Even if taxes were brought to zero, more than 90% of informal firms would remain unregistered. However, the paper also finds that VAT adoption has an important positive impact on the most productive firms and hence contributes to the expansion of formal sector output.

It has been shown that due to its self-enforcing nature, value-added taxation can reduce tax avoidance (Pomeranz, 2015). However, studying the link between VAT adoption and informality is daunting because VAT reforms are often implemented together with other policy measures aimed at simplifying registration procedures and lowering registration costs. Thus, it is difficult to separate the effects on informality from those of complementary policies. Moreover, studying the effect of the VAT on informality requires micro-level data for informal sector firms. Such data are hard to collect because informal firms tend to be small and, by definition, unregistered.

I develop a structural model that builds on De Paula and Scheinkman (2010) and then take the model to the data to study how VAT adoption affected informality in India. The model assumes two stages of production characterized by upstream and downstream firms. Firms with different observed levels of productivity choose between operating in the formal or informal sector by comparing the benefits of avoiding taxes with the risk of being detected and punished for operating in the informal sector. As I assume that the probability of

detection increases with firm size and that the punishment is large, only firms that do not expect to become large choose to remain informal. I assume that there are no restrictions to trading across sectors: formal and informal firms can trade with both types of firms.

In the model, exit is determined by random idiosyncratic productivity shocks and, in steady state, the number of entrants in each sector equals the number of firms that exit. Unlike De Paula and Scheinkman (2010), my model includes entry and exit costs. This feature endogenizes the sizes of the formal and informal sectors (and the size of the economy), as firms' choices are driven by entry cost, tax rates, and the tax regime. These elements allow for structural estimation of the impact of VAT adoption on informality.

A key prediction of the model is that VAT adoption does not have a large effect on the total number of formal firms, but it has an important effect on the incentive of productive firms to register and expand. As a consequence, VAT adoption leads to an expansion of formal sector output.

India is an ideal laboratory for studying the effect of VAT adoption on informality because over 2002-14, Indian states gradually replaced sales taxes with value-added taxation (in 2017, India adopted a broad-based goods and services tax). The stated objectives of this tax reform included encouraging tax compliance by informal firms and boosting the manufacturing sector (Rao, Mukherjee, and Bagchi, 2019; Arora et al., 2017).

India is one of the few countries that collects detailed data on informal firms. I source my data from the Indian Unorganized Manufacturing Enterprises (UME) survey, which provides a representative sample of informal firms in the manufacturing sector. UME is based on the economic census, which is a complete count of all economic units, including establishments and firms that are not registered with the tax authorities.

After estimating the model, I conduct a counterfactual analysis aimed at comparing the VAT with the cascading sales tax, while keeping the tax rate constant. I find that only 0.3% of informal firms (a 2% increase in the number of formal firms) would register after the cascading tax is replaced with the VAT. However, the effect of this tax reform is larger for more productive firms, leading to a nearly 12% expansion of formal sector output. I also study the effect of the tax reform on firms that were already in the formal sector. I find that value added taxation is particularly beneficial for large firms. This result is driven by the fact that large firms tend to buy a larger share of inputs from formal suppliers. Hence, they benefit more from the possibility to deduct the VAT paid on their purchases.

If tax rates and formal economic activity are kept constant, moving from a cascading sales tax to a VAT will lead to lower tax revenues. However, other things are not constant: VAT adoption has a positive effect on labor income and formal sector profits. I show that higher corporate and income tax revenues more than offset the loss of tax revenues associated with the transition from the cascading sales tax to the VAT. According to my estimates, a switch to the VAT will lead to a 27% drop in sales tax revenues but a 1% decrease in total tax revenues.

My paper is closely related to several strands of the literature on VAT adoption and informality.[1] Relevant papers include Pomeranz (2015), who shows that the VAT facilitates tax enforcement by generating paper trails; Rocha et al. (2020), who show that lower tax rates and registration costs promote the formalization of Brazilian firms; and Emran and Stiglitz (2005), who study the welfare gains associated with replacing duties with VAT. The paper which is closest to my work is Briand and Hoseini (2020). These authors also study the link between VAT adoption and informality in India. The key

[1] Keen and Lockwood (2010), Keen (2014) discuss the non-cascading feature of the VAT. Bachi et al. (1994) and Chelliah (1994) discuss how VAT adoption would streamline the cascading taxes in India. Fajgelbaum et al. (2019), Gadenne et al. (2019), Gerard and Naritomi (2018), Gerard et al. (2018), Gordon and Li (2009), Papp and Takats (2008), Prado (2011) discuss the inter-sectoral trade structure and the welfare effect after the adoption of VAT.

difference between my work and theirs is that I develop a structural model and counterfactual analysis, while their estimates are based on reduced form regressions.

The paper is also related to the literature that portrays informal firms as capitalists-in-waiting who cannot join the formal sector because of high entry costs (for example, De Soto, 1989; Tokman, 2007). This literature suggests that formalizing informal firms by reducing entry costs will unleash the energy of the informal sector.[2] My results are closer to those of Levy (2008), Maloney (2004), La Porta and Schleifer (2014), and Ulyssea (2018), who show that low-return firms have limited incentives to register, even if taxes are reduced or enforcement tightened.

The remainder of the paper is organized as follows: Section 2 describes the main characteristics of the informal sector in India; Section 3 builds a structural model with two stages of production and VAT; Section 4 structurally estimates the model; Section 5 presents the counterfactual analysis; and Section 6 concludes.

2. Stylized Facts on Informality

2.1 Data

In this paper, I use the firm-level annual survey of industries (ASI) and the unorganized manufacturing enterprises (UME) survey, which provide representative samples of formal and informal firms in India's manufacturing sector. The ASI consists of manufacturing establishments registered under the factories Act, which covers all the manufacturing firms employing at least 20 workers or more than 10 workers when using industrial power. Each

[2] Bruhn and Mckenzie (2014), Bruhn (2013), Dabla-Norris et al. (2008), De Andrade et al. (2013), De Giorgi and Rahman (2013), De Mel and McKenzie (2013), Hsieh and Klenow (2009) study the causes of informality, while Fotoniata and Moutos (2013), Gelb et al. (2009), Banerji and Jain (2007), Ihrig and Moe (2004), Kaplan and Piedra (2007), Loayza (2006), Medvedev and Oviedo (2013), Steel and Snodfrass (2008) study the consequence of being informal.

year, all establishments with more than 100 employees and at least 12% of the others are surveyed with a representative sample at the state and 4-digit industry code level.

The UME surveys are conducted every five years. The sample design of UME is based on the latest Economic Census, a complete count of all economic units in India, synchronized with the house listing operations of the Population Census. The two surveys are complementary because UME covers all the manufacturing firms that are not covered by ASI.

Combining the two datasets, I construct a representative sample of all the manufacturing firms in India. Table 1 compares financial variables between registered and unregistered firms in the fiscal year of 2015.

Complementing the firm-level datasets, I use the transaction-level data in the state of West Bengal to calculate the inter-sectoral trade structure. VAT-paying firms in West Bengal must report their purchases from, and sales to, other VAT-paying firms. The transaction data contain information on 4.8 million transactions among formal firms during 2010 and 2016. Each supplier-client pair observation contains the tax code of both parties conducting the transactions.

[Table 1]

2.2 Stylized Facts

I use the combined ASI-UME dataset for 2015 and restrict the sample to manufacturing firms in West Bengal with more than two employees. This is to eliminate the household establishments that operate occasionally and are not typically included in the tax base. For a description of the complete sample, see Appendix C.

Several stylized facts of informality shows that: (i) informal firms are pervasive but only produce a small portion of total output; (ii) most informal firms are small relative to an average formal firm; and (iii) formal firms mainly trade with other formal firms, and the same is true for informal firms.

The size of the informal sector. Informal firms are prevalent in the Indian manufacturing sector, particularly in downstream industries. Following Antras and Chore (2013), I calculate the downstream index for each 4-digit NIC industry using the Indian input-output table for 2015. I categorize industries with a downstream index above the median value as downstream industries and the rest as upstream industries.[3] Figure 1 shows that 95% of firms in the manufacturing sector are informal. The share of informal firms is 98% in the downstream industries vis-à-vis 92% in the upstream industries. Despite their high share, however, informal firms only produce 21% of total output (measured by the revenue) in the manufacturing sector. The share of informal output is 31% in the downstream industries compared to 16% in the upstream industries. The result is consistent with the finding in Schneider and Melina (2018) that the informal sector produces 18% of GDP in India. In addition, 77% of workers are employed in the informal sector.

[Figure 1]

Distribution of informal firms. I rank all the informal firms by their output and select the firms at the 5^{th}, 25^{th}, 50^{th}, 75^{th}, and 95^{th} percentiles to calculate their relative size to an average formal firm.

On average, an informal firm produces 1.5% and 2.4% as much as an average formal firm in the downstream and upstream industries, respectively. The median informal firms only produce 0.5%-0.6% as much as an average formal firm while the 75^{th} percentile informal firm produces less than 1.5% of an average formal firm (Figure 2).

[3] Appendix A documents the construction of the downstream index at 4-digit industry level.

[Figure 2]

Figure 3 illustrates the output distribution of formal firms vis-à-vis informal firms. Most informal firms produce less than the small formal firm in the 10^{th} percentile. Farrell (2004) documents a bunching problem in which a productive informal firm restricts its revenue below a certain threshold to avoid being detected. Appendix E plots the histogram of the size of informal firms and does not find evidence of bunching near that threshold.

[Figure 3]

Trade structure. A less documented feature of the informal sector is its trade pattern. A key constraint of analyzing domestic trade is the lack of inter-firm transaction data. Combining the transaction-level data with firm-level data, I calculate the share of inputs from formal suppliers and the share of outputs to formal clients.[4]

Figure 4 illustrates a segmented trade structure. Formal firms purchase 80% of inputs from formal suppliers while informal firms purchase 29% of inputs from formal suppliers. In addition, formal firms sell 48% of total output to formal buyers while informal firms sell 16% of total output to the formal buyers.

[Figure 4]

2.3 Descriptive Evidence

This section provides a descriptive analysis on how the adoption of VAT affects informality. Since 2003, Indian states gradually adopted the VAT to replace the existing sales tax, with ten states before 2005, 19 states in 2005, and 4 more states between 2006

[4] For the detailed calculation procedure, see Appendix B.

and 2008. Appendix C provides detailed description of the VAT reform. The variations in the timing of VAT adoption provide a natural experiment to study the effect of VAT reforms across Indian states.

I aggregate the firm-level sample into state-level sample and state-industry-level sample. The state-level sample contains the number of formal (informal) firms, formal (informal) sector output measured by total revenue, number of workers employed in the formal (informal) sector in 2000, 2005, 2010, 2015.

Table 2 provides summary statistics of informality indicators in the state-level sample. On average, less than 7% of observations are registered with tax authorities, varying from 0.1% in Arunachal Pradesh to 60% in Daman & Diu. However, the formal sector accounts for more than 70% of total output and employs 28% of total workers. The share of formal firms changes from 7.4% to 5.8% after VAT adoption, but the t-statistics is insignificant. The share of formal sector output increases significantly from 56% to 84% and the share of formal sector employment increases from 26% to 31%. State-industry-level sample shows similar results except that the share of formal firms significantly increases by 5% after VAT adoption. The share of formal sector output and employment significantly expand by 18.4% and 7.8%, respectively. Appendix C2 provides empirical evidence tests on the impact of VAT adoption on the share of informality using a diff-in-diff specification.

3. Theory and Model

In this section, I develop a structural model that builds on De Paula and Scheinkman (2010) to jointly determine the sizes of the formal and informal sectors.

The model assumes two stages of production characterized by upstream and downstream firms. Upstream firms with different observed productivity utilize labor to produce intermediate goods for downstream firms. I assume that the labor market is competitive,

and the intermediate goods market involves monopolistic competition. Firms choose between operating in the formal or informal sector by comparing the benefits of avoiding the value-added tax and extra labor costs with the risks of being detected and punished for operating in the informal sector.

Downstream firms with different productivity use intermediate goods to produce homogeneous final products. I assume the final goods market is competitive so downstream firms are price takers. Downstream firms face the same trade-off between operating in the formal or informal sector.

To enter the formal and informal sectors, firms have to pay a one-time fixed cost which is higher for the formal sector. In each period, there is a fixed number of potential entrants with heterogeneous productivity available to join both sectors. Successful entrants will replace "dead" firms that exit the market after being hit by random productivity shocks. In the equilibrium, the number of productive entrants equals the number of productive firms that exit. Note that upstream and downstream firms can have positive profits because the number of firms is limited in the equilibrium.

I induce the VAT in the model as an ad valorem tax which applies to each sale of formal firms. The difference between a VAT and a cascading sales tax is the tax on intermediate goods. While the VAT removes taxes on inputs, the cascading sales tax still applies to intermediate goods. The tax incidence is shared by upstream and downstream firms depending on the elasticity of supply and demand.

3.1 Downstream Firms

Downstream formal and informal firms use the same technology to produce final goods. I assume a constant elasticity of substitution (CES) production function, which is increasing, concave, and twice continuously differentiable.

A formal firm's cost minimization problem is given by:

$$min \int_0^N x(i)p^f(i)[1-\tau(i)]di \quad (1)$$

$$s.t.\ y_0 = \theta^d \left[\int_0^N x(i)^\rho \beta(i) di\right]^{\frac{\alpha}{\rho}} \quad (2)$$

where θ^d is the observed productivity of the downstream firms, $x(i)$ is the quantity of inputs from the upstream supplier *i*, $p^f(i)$ is the price paid by to the upstream supplier *i*, y_0 is total output that maximizes the profit. I use $\beta(i)$ to capture the difference in quality between formal and informal inputs.[5] A smaller $\beta(i)$ means lower quality of product from supplier *i*. $\beta(i)$ is normalized to 1 for all formal suppliers and β for all informal suppliers. $\frac{1}{1-\rho}$ is the elasticity of substitution parameter, and α is a production coefficient between 0 and 1. In addition to the VAT, formal firms pay the VAT at the rate of τ and the corporate tax on their profits at the rate of τ^{corp}.

Informal firms avoid paying taxes but will lose all the profits if they are detected. I add an additional cost b_i on informal firms' total inputs, which is associated with the size of informal firms. While a more general form of b_i can be adopted as a function of firm size, for simplicity I assume b_i is infinite if the observed productivity of an informal firm exceeds that of the smallest formal firm and is one if the informal firm is below the threshold, as in De Paula and Scheinkman (2010).

The cost minimization problem for an informal firm is given by:

[5] One can also rewrite the production function as $y_0 = \theta^d \{\int_0^N \left[x(i)\beta(i)^{1/\rho}\right]^\rho di\}^{\frac{\alpha}{\rho}}$. $\beta(i)$ can be interpreted as the share of inputs that is usable. Because informal firms produce lower quality goods, $\beta(i)$ is smaller for informal firms than formal ones.

$$min \int_0^N b_i x^i(i) p^i(i) di \quad (3)$$

$$s.t.\ y_0^i = \theta^d [\int_0^N x^i(i)^\rho \beta(i) di]^{\frac{\alpha}{\rho}} \quad (4)$$

3.2 Upstream Firms

The upstream market is characterized by firms that produce intermediate goods with labor. I assume a simple linear production function of labor so that a firm's sales to different buyers are separable. I also assume formal firms pay additional labor costs, such as benefits, pensions, and social security. The upstream market is monopolistic competition in which firms have market power to set the prices of their products.

When an upstream formal firm sell to a downstream formal firm, its profit maximization problem is given by:

$$max(1 - \tau^{corp})[(1 - \tau) p_f^f x_f^f - w(1 + \tau_l) l(x_f^f)] \quad (5)$$

$$s.t. x_f^f = \theta^u l \quad (6)$$

When an upstream formal firm sell to a downstream informal firm, its profit maximization problem is:

$$max(1 - \tau^{corp})[(1 - \tau) p_f^i x_f^i - w(1 + \tau_l) l(x_f^i)] \quad (7)$$

$$s.t. x_f^i = \theta^u l \quad (8)$$

where p_f^f and p_f^i are the demand function of the downstream formal and informal firms, respectively. τ_l is the additional labor cost, θ^u is observed productivity, and l is the number of workers hired by the upstream firm.

When an upstream informal firm sell to a downstream formal firm, the profit maximization problem is given by:

$$max\, p_i^f x_i^f - wb_i l(x_i^f) \quad (9)$$

$$s.t. x_i^f = \theta^u l \quad (10)$$

When an upstream informal firm sell to a downstream informal firm, the profit maximization problem is given by:

$$max\, p_i^i x_i^i - wb_i l(x_i^i) \quad (11)$$

$$s.t. x_i^f = \theta^u l \quad (12)$$

where b_i is the punishment on informal firms using the same assumption as in the downstream industries.

3.3 Entry and Exit

I assume that the market involves endogenous entries but exogenous exits. There are M potential entrants into the downstream and upstream industries in each period. Potential entrants observe their pre-entry productivity θ^d that follows the distribution $\theta^d \sim G^{\mathrm{d}}$. In each period, there is an ad-hoc productivity shock after which a fixed proportion of firms will randomly exit the market. The exit rate is different for the formal and informal sectors, reflecting their different levels of resilience to productivity shocks. Hence, the value function based on the observed productivity θ can be written as:

$$\widetilde{V_s} = ma\, x\left\{\frac{\pi_s(\theta)}{1-\kappa_s}, 0\right\} \quad (13)$$

where $\pi_s(\theta)$ is the profit of the firm with productivity θ in sector s, κ_s is the probability of exit for sector s, and the discount rate is normalized to 1.

I assume the realized productivity v follows a random process: $v = \varepsilon\theta$, and remains constant over the following periods. ε is a random variable. The expected present value function is given by:

$$V_s = \int \widetilde{V_s} dF(v|\theta) \quad (14)$$

where $F(v|\theta)$ is the cumulative distribution function of realized productivity conditional on the observed productivity.

Firms will enter the formal sector if the expected value of being formal is larger than the that of being informal plus the difference of entry costs between the downstream formal sector and the downstream informal sector. Hence, a necessary condition for being formal is:

$$V_f > V_i + E_f - E_i, with\ \overline{V_f}\ as\ the\ cutoff\ present\ value \quad (15)$$

If a firm's expected present value is less than the entry cost into the formal sector E_f but higher than the entry cost of the informal sector E_i, then it can choose to enter the informal sector. The necessary condition for being informal is:

$$E_i < V_i < E_f \quad (16)$$

However, if the firm is surprised with a low productivity draw such that the present value based on the realized productivity is less than the corresponding entry costs, it will exit the market immediately. This condition can be written as:

$$V_s^r < E_s \quad (17)$$

where V_s^r is the realized present value of the firm in sector s, and E_s is the entry cost.

3.4 Equilibrium

To close the model, it is necessary to specify the consumer demand. I assume a representative household with a utility function: $U = u(x)$, and that total consumption x is the sum of labor income $(1 + \tau_l)wL^f + wL^i$, firm profit Π, and total tax T. Fines on the informal firms are treated as sink costs or a compensation of law enforcement. I also assume the total labor supply is fixed at $L = L^f + L^i$. All the consumption components are endogenously determined by the model except the fixed total labor supply L, and the total consumption is calculated as:

$$consumption = (1 + \tau_l)wL^f + wL^i + \Pi + T \quad (18)$$

The equilibrium is determined by the following conditions: (i) intermediate goods market clears, (ii) final goods market clears, (iii) labor market clears, and (iv) stationary equilibrium.

Intermediate goods market clears. The demand of intermediate goods by buyer i from supplier j must equal the supply from supplier j to buyer i. This determines the optimal prices charged by upstream suppliers to downstream buyers.

Final goods market clears. The total supply of final goods by downstream firms equals the total demand by final consumers. Since the model does not allow for investment, household income is equal to consumption of final goods.

Labor market clears. The total labor employed by upstream firm must equal the total labor supply by households. $l(x_1^f) + l(x_2^f) + l(x_1^i) + l(x_2^i) = L$. This condition determines the equilibrium wage.

Stationary equilibrium. The stationary equilibrium requires that all aggregate variables remain constant. Therefore, the number of entrants in each sector equals the number of exits in the corresponding sector.

$$M_f^d = \frac{1 - F^d(\overline{v_f^d} | \overline{\theta_f^d})}{\kappa_f} M[G^d(\overline{\theta_f^d}) - G^d(\overline{\theta_i^d})] \quad (19)$$

$$M_i^d = \frac{F^d(\overline{v_f^d} | \overline{\theta_f^d})}{\kappa_i} M \cdot [G^d(\overline{\theta_i^d}) - G^d(\underline{\theta_i^d})] \quad (20)$$

$$M_f^u = \frac{1 - F^u(\overline{v_f^u} | \overline{\theta_f^u})}{\kappa_f} M[G^u(\overline{\theta_f^u}) - G^u(\overline{\theta_i^u})] \quad (21)$$

$$M_i^u = \frac{F^u(\overline{v_f^u} | \overline{\theta_f^u})}{\kappa_i} M \cdot [G^u(\overline{\theta_i^u}) - G^d(\underline{\theta_i^u})] \quad (22)$$

In equation (18), M is the total number of potential entrants. $\overline{\theta_f^d}$ is the upper bound of downstream formal firms' observed productivity, $\overline{\theta_i^{\mathrm{d}}}$ is the upper bound of downstream informal firms' observed productivity, $\underline{\theta_i^d}$ is the lower bound of downstream informal firms' observed productivity, $\overline{\mathrm{v_f^d}}$ is the upper bound of downstream formal firms' realized productivity, and $\mathrm{M}_f^{\mathrm{d}}$ is the total number of incumbent formal firms in the downstream. The same denotation applies to the upstream firms.

In addition, $\mathrm{M}[G^d\left(\overline{\theta_{\mathrm{f}}^{\mathrm{d}}}\right) - G^d(\overline{\theta_{\mathrm{i}}^{\mathrm{d}}})]$ is the ex-ante number of downstream firms who decide to enter the formal sector before drawing their realized productivity, $1 - \mathrm{F}^{\mathrm{d}}(\overline{\mathrm{v}_{\mathrm{f}}^{\mathrm{d}}}|\overline{\theta_{\mathrm{f}}^{\mathrm{d}}})$ is the share of successful entrants into the formal sector over the number of firms who decide to enter. Thus, equation (18) requires that the number of successful entrants into the downstream formal sector in each period equals the total number firms that exit.

Figure 5 illustrates the framework by plotting the relationship between firm productivity and the counterfactual present values of firms entering the formal and informal sectors (net of entry costs). The blue line shows the value of firms entering the informal sector for each level of productivity, and the orange line similarly shows the value of firms entering the formal sector.

There is a positive relationship between productivity and the present value, but the shape of the relationship differs between the formal and informal sectors. Firms with the lowest levels of productivity have a negative present value of being formal because their net profits are not enough to cover the fixed costs of formality. But they have a positive present value of being informal because they can avoid paying taxes and the high entry costs. Hence, these low-productivity firms will choose to enter the informal sector.

As productivity rises, the present value of being formal rises faster than that of being informal because informal firms face higher risks of being detected and punished. There is threshold of productivity below which firms choose to enter the informal sector and above which firms choose to enter the formal sector.

[Figure 5]

Since the model does not generate an analytical solution, I run the estimation based on the numerical solution of the model. The detailed model solution is shown in Appendix D.

4. Calibration and Estimation

This section specifies the functional form and discusses the estimation of the full model in an equilibrium setting. The equilibrium outcome should be able to capture the key stylized facts stated in the previous section. I use the simulated method of moment (SMM) to estimate the parameters in the model.

4.1 Parameterization

I first specify the distribution function of the observed productivity $G(\cdot)$ and the realized productivity $F(\cdot)$. The distribution of observed productivity $G(\cdot)$ is assumed to follow the Pareto distribution. I allow the distribution of the observed productivity varying between the downstream and upstream industries, which is given by:

$$G^d(\theta \geq x) = \begin{cases} (\frac{v_0}{x})^{\xi^d} & for\ x \geq v_0 \\ 1 & for\ for\ x \leq v_0 \end{cases} \quad (23)$$

$$G^u(\theta \geq x) = \begin{cases} (\frac{v_0}{x})^{\xi^u} & for\ x \geq v_0 \\ 1 & for\ for\ x \leq v_0 \end{cases} \quad (24)$$

where v_0 is the minimum possible value of the Pareto distribution, and ξ is the shape parameter. The only difference of the pre-entry productivity distribution between the upstream and the downstream is the parameter ξ, which governs the right tail of the distribution. This setting requires the same minimum size of firm in both downstream and upstream firms while the maximum size of firms (large firms) can differ. This is consistent with the data where I limit the minimum size of firms to be 3 employees while I do not set a limit for the maximum firm size.

It should be note that the realized productivity distribution follows a simple process $v = \theta\varepsilon$. I assume the misperception of productivity ε has a log-normal distribution with mean

zero and variance of σ^2. Therefore, realized productivity v is the product of a log-normal and a Pareto random variable, which follows the Pareto-Lognormal distribution. The three-parameter distribution has a log-normal body and a Pareto right tail. Luttmer (2007) and Ulyssea (2018) prove that the Pareto-Lognormal distribution fits many salient features of firm size distribution in developing countries.

4.2 Calibration and Estimation of Parameters

The model parameters are divided into two sets. The first set includes three parameters which are calibrated from the real data. These parameters are:

$$\psi = \{\tau, \tau^{corp}, \kappa_f\}$$

The VAT rate in West Bengal is 12.5% in 2015, and the effective corporate tax rate is to 30%.[6] The exit rate of formal firms is calibrated from ASI annual data (2014-15). After controlling for the industry fixed effect, the exit rate of formal firm is calibrated at 8%. The Pareto distribution scale parameter (v_0) is set to 0.7 so that the minimum number of employees is three. This parameter only affects the scale of the key variables.

The second set includes 12 parameters, which are estimated by SMM. These parameters are:

$$\varphi = \{\beta, \xi^d, \xi^u, E_i^d, E_f^d, E_i^u, E_f^u, \alpha, \rho, \tau_l, \kappa_i, \sigma\}$$

The SMM finds a set of parameters so that it generates a set of moments that can be fit to the real data. For any given set of parameters, the model will generate a full set of observations that simulates firms' behavior, including tax status, outputs, and trade

[6] https://tradingeconomics.com/india/corporate-tax-rate

structure. I then use this set of observations to generate a set of moments and compare it with the set of moments from the real data. The SMM estimator is the set of parameters that best approximates the moments calculated from the real data.

First, I bootstrap the real data 500 times to generate 500 sets of bootstrapped datasets. Using these datasets, I calculate a vector of moments for each dataset and take the mean and the standard deviation of all vectors of moments.

$$\widehat{m_N} = \frac{1}{N}\sum_{i=1}^{N} m_i \quad (25)$$

where m_i is the vector of moments in the i[th] dataset and $\widehat{m_N}$ is the mean of all the vectors of moments.

Next, I choose an initial set of parameters φ_0 and generate a vector of moments from the model. Then, I replicate the process for 50 times and calculate the mean of all vectors of moments.

$$\widetilde{m_N} = \frac{1}{N}\sum_{i=1}^{N} \widetilde{m_i}(\psi, \varphi_0) \quad (26)$$

where $\widetilde{m_i}(\psi, \varphi_0)$ is the vector of moments from the i[th] replication and $\widetilde{m_N}$ is the mean of all vectors of moments.

Finally, I denote the distance between the moments $\widehat{m_N}$ from the real data and the moments $\widetilde{m_N}$ from the model as $g(\psi, \varphi_0) = \widehat{m_N} - \widetilde{m_i}(\psi, \varphi_0)$. Then the SMM estimator is:

$$\hat{\varphi} = argmin\{g(\psi, \varphi_0)'\widehat{W_N}g(\psi, \varphi_0)\}$$

where $\widehat{W_N}$ is a positive, semi-definite $r \times r$ matrix and r is the length of the vector of moments $\widehat{m_N}$. Under suitable regularity conditions, the estimator is consistent and asymptotically normal and the weighting matrix $\widehat{W_N}$ is chosen optimally in order to minimize the asymptotic covariance. I use the variance of real moments as $\widehat{W_N}$. Appendix G provides a detailed discussion of the SMM estimator.

I choose 26 moments to estimate these parameters, including: (i) the share of informal firms in the downstream industries, the share informal firms in the upstream industries, the share of informal sector output in the downstream industries, and the share of informal sector output in the upstream industries (4 moments); (ii) the ratio of output by the 5^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th} percentile informal firms over the average formal firm output in the downstream industries, and the same ratio in the upstream industries (10 moments); the ratio of output by the 5^{th}, 10^{th} percentile informal firms over the average formal firm output in the downstream industries, and the same ratio in the upstream industries (4 moments); (iii) the share of input from formal suppliers in the formal sector, the share of input from formal suppliers in the informal sector, the share of output to the formal buyers in the formal sector, and the share of output to the formal buyers in the informal sector (4 moments); (iv) the standard deviation of outputs divided by the mean output (coefficient of variance) in the downstream and upstream (2 moments); (v) the share of intermediate goods over total expense (1 moment); and (vi) the share of labor employed by the informal sector (1 moment).

I provide an intuitive analysis regarding the determination of the parameters based on these moments. The distribution of informal and formal firms in the downstream and upstream industries determines the Pareto shape parameter. A more clustered distribution of informal firms indicates a larger Pareto shape parameter. The share of informal firms in the downstream and upstream industries determines the entry costs into the formal sector, in which higher entry costs lead to a larger informal sector. The relative size of large informal firms helps to estimate the entry costs and the exit probability separately (Ulyssea, 2018).

A higher entry cost into the formal sector increases the size of informal firms that barely enter while the exit probability of informal firms increases the size of all informal firms.

The trade structure determines the quality of formal input β. Higher quality of formal input induces both formal and informal buyers to purchase more from formal suppliers. The trade structure and the share of labor employed by the informal sector jointly determine the extra cost of formal worker τ_l. Less purchase from formal suppliers and a larger share of workers in the informal sector indicate a higher extra cost of labor in the formal sector. The share of labor in the informal sector separates τ_l from β. The relative size of informal and small formal firms helps separate the exit parameter κ_i and the entry costs. While a higher exit rate of informal firms reduce the size of all the informal firms, a higher entry cost into the formal sector only reduces the size of small formal firms.

4.3 Estimates and Model Fit

Table 3 shows the calibration and estimation results. The entry cost into the formal sector is 1.1 and 1.8 times of the average annual profit of a formal firm in the downstream and upstream industries, respectively. This suggests that it is more difficult to enter the formal sector in the downstream than in the upstream, which is consistent with the larger share of informal firms in the downstream industries. Moreover, the entry costs into the formal sector is 90 and 37 times greater than the average annual profit of an informal firm in the downstream and upstream industries, respectively. This suggests that informal firms are too unproductive to overcome the entry cost into the formal sector. The Pareto shape parameter is 2.35 and 1.4 in the downstream and upstream industries, respectively.

The quality parameter of the formal inputs β is 1.15, which indicates that the quality of formal inputs is 15% higher than that of informal inputs. The difference in labor cost between the formal and informal sector is 10.7, meaning that formal firms pay 11 times higher on a worker than informal firms do.

Table 4 compares the moments generated from the real data and from the model using the above parameters. The model captures the share of informal firms and the share of informal output, as well as the trade structure. But the model does not fit well with the relative size of small formal firms, partly due to the constraints or advantages of small formal firms.

The model predicts that the share of informal firms in downstream and upstream industries is 97% and 93% respectively, with less than 1% deviation from the real data. Around 30% of the downstream output and 14% of the upstream output are produced by informal firms, which is similar to the real data. In general, the model captures the pervasive number of informal firms that contribute a small fraction of total output.

In the model, the relative size of an informal firm in each quantile is 0.22%, 0.31%, and 0.66% of the average formal firm in downstream industries. This is similar to the moments from the real data. For upstream informal firms, the model predicts the relative size in each quantile at 0.09%, 0.19, 0.72% of the average formal firm, which deviates from the moments in the real data by one or two standard deviations.[7] The model captures the relative size of the large informal firm at the 95th percentile. For small formal firms, the model prediction deviates one or two standard deviations from the real data.

The model predicts a segmented trade structure in which formal firms purchase 80% of their inputs from formal suppliers while informal firms purchase 43% of their inputs from formal suppliers. The share of sales to formal buyers is 48% and 28% in the model for formal and informal firms, respectively, compared to 48% and 16% in the real data. In general, the model captures the trade structure of formal firms but overestimates the informal sector's trade with formal partners.

[Table 4]

[7] Appendix G reports the standard deviation of these moments.

4.4 Identifying Different Type of Informal Firms

The introduction describes different perspectives of the informal sector, which imply different firm behavior in response to tax reforms. This section discusses which perspective best explains the response of informal firms regarding VAT adoption.

I categorize the informal firms into three groups, including reservoir, parasite, and unproductive groups. The reservoir-type firms are those that will become formal after all tax burdens (VAT and corporate tax) are removed. The parasite-type firms are those that will remain informal but can survive as formal when all tax burdens are removed. The unproductive-type firms are those that can only survive as informal. When tax rate lower to zero, the present value of being formal is negative but that of being informal is positive.

Figure 6 shows that most informal firms are too unproductive to formalize even if tax rates drop to zero. Less than 3.5% and 0.1% of informal firms in each production stage will formalize when tax rates drop to zero. Only 5.1% and 0.3% of informal firms can survive as formal firms in downstream and upstream industries, respectively. The unproductive firms make up the majority of informal firms and are too unproductive to survive as formal ones even after the tax burdens are removed. The prevalence of unproductive informal firms suggests that without a significant productivity increase, informal firms will not be able to survive and compete in the formal sector even tax rates drop to zero.

[Figure 6]

5. Counterfactual Analysis

In the counterfactual analysis, I compare the VAT with the cascading sales tax at the same tax rate. Tax rate is set to 7.5%, which is consistent with the statutory sales tax rate in the

state of West Bengal. I also consider the other two scenarios in which VAT rate is 9.5% and 10.1%, which generates the same number of formal firms (informality-neutral) or the same level of tax revenue after VAT adoption (revenue-neutral), respectively.

5.1 Informality: Number and Output

Table 5 shows the share of informality after VAT adoption. When replacing the cascading sales tax with the VAT, the share of formal firms increases slightly from 5.5% to 5.7% and the share of formal outputs rises from 87% to 89%. In addition, the share of workers in the formal sector also rises from 11.3% to 13.3%, suggesting a significant reallocation of labor from the informal sector to the formal one (workers employed in the formal sector increase by 17% while workers in the informal sector decrease by 2%[8]). If the VAT rate rises to the informality-neutral scenario, the share of formal firms in the economy remains similar at 5.6% which accounts for 89% of total outputs. If the VAT rate rises further to the revenue-neutral scenario, the share of formal firms is 5.5% and the share of formal output is 89%.

[Table 5]

However, the share of informality neglects the absolute change number and output. Table 6 shows that about 0.6% of informal firms become formal after VAT adoption, which corresponds to a 2% increase in the number of formal firms if the tax rate remains the same. However, VAT adoption has a significant impact on the formal sector output. Formal sector output increases by 12% while the informal sector output shrinks slightly by 2.8%. In the informality-neutral scenario, the number of formal firms remains unchanged (but the number of informal firms increases by 0.2%). The output of the formal and informal sector increases by 10% and -2%, respectively. In the revenue-neutral scenario, the output of formal and informal firms increases 10% and 1%, respectively. The results suggest that

[8]Note that the model assumes fixed labor supply.

VAT adoption does not have a large effect on the total number of formal firms, but it has an important effect on the incentive of productive firms to expand production.

[Table 6]

The expansion of formal sector output is mainly driven by large and high-productivity firms.[9] In downstream industries, small firms below the median size grow less than 2.5% while firms at the 90th percentile grow by 4.3%. In upstream industries, firms at the 10th percentile and 90th percentile grow by 9% and 15%, respectively (Figure 7).

This is because large firms tend to purchase a larger share of inputs from formal suppliers. Figure 7 shows that small formal firms at the 10th percentile purchase 53% of inputs from formal suppliers while large firms at the 90th percentile purchase 67% of inputs from formal suppliers. Similarly, small upstream formal firms at the 10th percentile sell 39% of their outputs to the formal sector while large firms at the 90th percentile sell 46% of outputs to the formal sector. When VAT removes tax costs on formal intermediate goods, large downstream firms deduct more taxes paid on their purchases, while large upstream firms sell more to downstream firms.

[Figure 7]

Now that a segmented trade structure between formal and informal sectors has been found, the extent to which the segmented trade can be attributed to the VAT can be evaluated. Table 7 compares the share of inputs from and outputs to the formal partners. In the baseline scenario, the formal sector purchases 74% of inputs from formal suppliers and sells 46% of outputs to formal buyers. After VAT adoption at the same tax rate, the share of inputs from formal suppliers rises to 78% and the share of output to formal buyers rises

[9] One concern is that the expansion of formal firms is driven by new entrants. To deal with this concern, I replicate analysis in the informality-neutral scenario so that the extensive margin of the formal sector does not change. The results hold in the informality scenario.

to 48%. Varying the VAT tax rate (informality-neutral rate or revenue-neutral rate) does not significantly affect the magnitude of change either.

[Table 7]

5.2 Taxation, Corporate Profits, and Household Income

One concern of the existence of a large informal sector is the potential loss of tax revenue which could have been used to finance public goods. Besley and Persson (2013) emphasize the difference in tax revenues to GDP ratios across developing and industrialized countries and argue that the main question in taxation and development is how a government can go from raising around 10% of GDP in taxes to raising around 40%.

VAT adoption generates less sales tax revenues than the cascading sales tax if the tax rate remains the same. However, the adoption of VAT has a positive effect on labor income and formal corporate profits. Table 8 shows that labor income of formal workers increases by 20% after VAT adoption while informal workers' income remains unchanged.

As for corporate profits, formal sector profits increase by 8% after VAT adoption while informal sector profits shrink by 2%. For informal firms, there is a blurred border between corporate profit and individual wage. A loss of corporate profit negatively affects informal workers despite an unchanged wage income. Since the profit loss of the informal sector is less than 2% after VAT adoption, it suggests a limited welfare loss of informal households.

[Table 8]

Higher corporate and income tax revenues offset the loss of other tax revenues (Table 8). If the composition of tax revenues remains unchanged, a switch to the VAT will lead to a

27% drop in sales tax revenues but a 1% decrease in total tax revenue. As tax rate increases, the net gain in total tax revenue becomes more significant (Figure 8).

[Figure 8]

6. Conclusions

In the past thirty years, more than 50 countries adopted value added taxation. In some of these countries, VAT was adopted with the explicit objective to improve compliance and reduce informality. Yet, we do not know much about VAT's impact on informality.

This paper presents a structural framework aimed at studying the relationship between VAT adoption and informality. Using data for Indian manufacturing firms, the paper shows that, although VAT does not have a large impact on informal firms with low productivity, it leads to a significant increase in formal sector output. Specifically, the paper finds that substituting a cascading sales tax with VAT would have a relatively small effect on the number of formal firms (a 2% increase, according to my estimates), but a much larger effect on formal sector output (a 12% increase, according to my estimates). I also find that VAT is particularly beneficial for larger firms which tend to buy a larger share of inputs from formal suppliers.

While the results of my papers are in line with the literature that finds that unproductive firms have limited incentives to register, even if taxes are reduced or enforcement tightened (Levy, 2008, Maloney, 2004, La Porta and Schleifer, 2014, and Ulyssea, 2018), my findings suggest that VAT adoption can be beneficial for productive firms. By streamlining the tax system to avoid double taxations at every stage of production, the VAT promotes the expansion of productive formal firms, leading them to absorb labor and capital from the informal sector.

It is through this reallocation process that VAT adoption can have a positive effect on economic growth and structural transformation.

Reference

Antràs, Pol, and Davin Chor. "Organizing the global value chain." *Econometrica* 81.6 (2013): 2127-2204.

Arora, Anchal, et al. "Catalysing MSME Entrepreneurship in India: Capital." *Technology and Public Policy (KPMG and IMC, June 2017), https://assets. kpmg. com/content/dam/kpmg/in/pdf/2017/06/Catalysing-MSME-Entrepreneurship-India. pdf (accessed February 15, 2018)* (2017).

Bagchi, Amaresh, et al. "Reform of domestic trade taxes in india: issues and options." *National Institute of Public Finance and Policy, New Delhi* (1994).

Banerji, Arup, and Sanjay Jain. "Quality dualism." *Journal of Development Economics* 84.1 (2007): 234-250.

Benjamin, Nancy C., and Ahmadou Aly Mbaye. "The Informal Sector, Productivity, and Enforcement in West Africa: A Firm-level Analysis." *Review of Development Economics* 16.4 (2012): 664-680.

Besley, Timothy, and Torsten Persson. "Taxation and development." *Handbook of public economics*. Vol. 5. Elsevier, 2013. 51-110.

Hoseini, Mohammad, and Océane Briand. "Production efficiency and self-enforcement in value-added tax: Evidence from state-level reform in India." *Journal of Development Economics* (2020): 102462.

Bruhn, Miriam, and David McKenzie. "Entry regulation and the formalization of microenterprises in developing countries." *The World Bank Research Observer* 29.2 (2014): 186-201.

Bruhn, Miriam. "A tale of two species: Revisiting the effect of registration reform on informal business owners in Mexico." *Journal of Development Economics* 103 (2013): 275-283.

Chelliah, Raja J. "Agenda for Comprehensive Tax Reform." *Indian Economic Journal* 41.3 (1994): 38.

Dabla-Norris, Era, Mark Gradstein, and Gabriela Inchauste. "What causes firms to hide output? The determinants of informality." *Journal of development economics* 85.1-2 (2008): 1-27.

De Andrade, Gustavo Henrique, Miriam Bruhn, and David McKenzie. *A helping hand or the long arm of the law? Experimental evidence on what governments can do to formalize firms*. The World Bank, 2013.

De Giorgi, Giacomo, and Aminur Rahman. "SME's Registration: Evidence from an RCT in Bangladesh." *Economics Letters* 120.3 (2013): 573-578.

De Mel, Suresh, David McKenzie, and Christopher Woodruff. "The demand for, and consequences of, formalization among informal firms in Sri Lanka." *American Economic Journal: Applied Economics* 5.2 (2013): 122-50.

De Paula, Aureo, and Jose A. Scheinkman. "Value-added taxes, chain effects, and informality." *American Economic Journal: Macroeconomics* 2.4 (2010): 195-221.

De Soto, Hernando. "The other path." (1989).

Fajgelbaum, Pablo D., et al. "State taxes and spatial misallocation." *The Review of Economic Studies* 86.1 (2019): 333-376.

Farrell, Diana. "hidden dangers." *The McKinsey Quarterly* 3 (2004): 26.

Fotoniata, Eugenia, and Thomas Moutos. "Product Quality, Informality, and Child Labor." *Review of Development Economics* 17.2 (2013): 268-283.

Gadenne, Lucie, Tushar K. Nandi, and Roland Rathelot. "Taxation and Supplier Networks: Evidence from India." (2019).

Gelb, Alan, et al. "To formalize or not to formalize? Comparisons of microenterprise data from Southern and East Africa." *Comparisons of Microenterprise Data from Southern and East Africa (July 20, 2009)* (2009).

Gérard, François, and Joana Naritomi. "Value Added Tax in developing countries: Lessons from recent research." *IGC Growth Brief Series* 15 (2018).

Gerard, François, Joana Naritomi, and Arthur Seibold. *Tax systems and inter-firm trade: evidence from the VAT in Brazil*. Tech. rep., Mimeo London School of Economics, 2018.

Gordon, Roger, and Wei Li. "Tax structures in developing countries: Many puzzles and a possible explanation." *Journal of public Economics* 93.7-8 (2009): 855-866.

Hsieh, Chang-Tai, and Peter J. Klenow. "Misallocation and manufacturing TFP in China and India." *The Quarterly journal of economics* 124.4 (2009): 1403-1448.

Ihrig, Jane, and Karine S. Moe. "Lurking in the shadows: the informal sector and government policy." *Journal of Development Economics* 73.2 (2004): 541-557.

Kaplan, David S., Eduardo Piedra, and Enrique Seira. *Entry regulation and business start-ups: Evidence from Mexico*. The World Bank, 2007.

Keen, Michael, and Ben Lockwood. "The value added tax: Its causes and consequences." *Journal of Development Economics* 92.2 (2010): 138-151.

Keen, Michael. "Targeting, cascading and indirect tax design." *Indian Growth and Development Review* (2014).

La Porta, Rafael, and Andrei Shleifer. "Informality and development." *Journal of Economic Perspectives* 28.3 (2014): 109-26.

Levy, Santiago. *Good intentions, bad outcomes: Social policy, informality, and economic growth in Mexico*. Brookings Institution Press, 2010.

Loayza, Norman V., and Jamele Rigolini. *Informality trends and cycles*. The World Bank, 2006.

Luttmer, Erzo GJ. "Selection, growth, and the size distribution of firms." *The Quarterly Journal of Economics* 122.3 (2007): 1103-1144.

Maloney, William F. "Informality revisited." *World development* 32.7 (2004): 1159-1178.

McFadden, Daniel. "A method of simulated moments for estimation of discrete response models without numerical integration." *Econometrica: Journal of the Econometric Society* (1989): 995-1026.

Medina, Leandro, and Friedrich Schneider. "Shadow economies around the world: what did we learn over the last 20 years?." (2018).

Medvedev, Denis, and Ana María Oviedo. *Informality and profitability: Evidence from a new firm survey in Ecuador*. The World Bank, 2013.

Papp, Tamás K., and Elöd Takáts. *Tax Rate Cuts and Tax Compliance--The Laffer Curve Revisited (EPub)*. International Monetary Fund, 2008.

Pomeranz, Dina. "No taxation without information: Deterrence and self-enforcement in the value added tax." *American Economic Review* 105.8 (2015): 2539-69.

Porta, Rafael La, and Andrei Shleifer. *The unofficial economy and economic development*. No. w14520. National Bureau of Economic Research, 2008.

Prado, Mauricio. "Government policy in the formal and informal sectors." *European Economic Review* 55.8 (2011): 1120-1136.

Rao, R. Kavita, Sacchidananda Mukherjee, and Amaresh Bagchi. *Goods and Services Tax in India*. Cambridge University Press, 2019.

Rocha, Rudi, Gabriel Ulyssea, and Laísa Rachter. "Do lower taxes reduce informality? Evidence from Brazil." *Journal of development economics* 134 (2018): 28-49.

Singh, Kanhaiya, and M. R. Saluja. "Input–Output Table for India 2013–2014: Based on the new series of national accounts statistics and supply and the use table." *Margin: The Journal of Applied Economic Research* 12.2 (2018): 197-223.

Steel, W., and Donald Snodgrass. "Raising productivity and reducing risks of household enterprises: diagnostic methodology framework." *World Bank Africa Region Analysis on the Informal Economy with the support of the WIEGO network* (2008).

Ulyssea, Gabriel. "Firms, informality, and development: Theory and evidence from Brazil." *American Economic Review* 108.8 (2018): 2015-47.

Figure 1: Share of Informal Sector

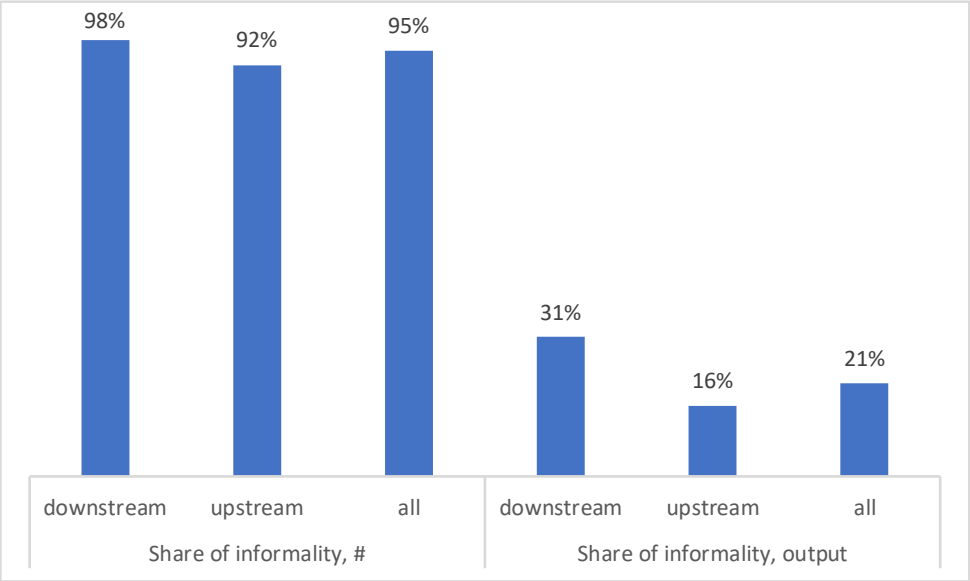
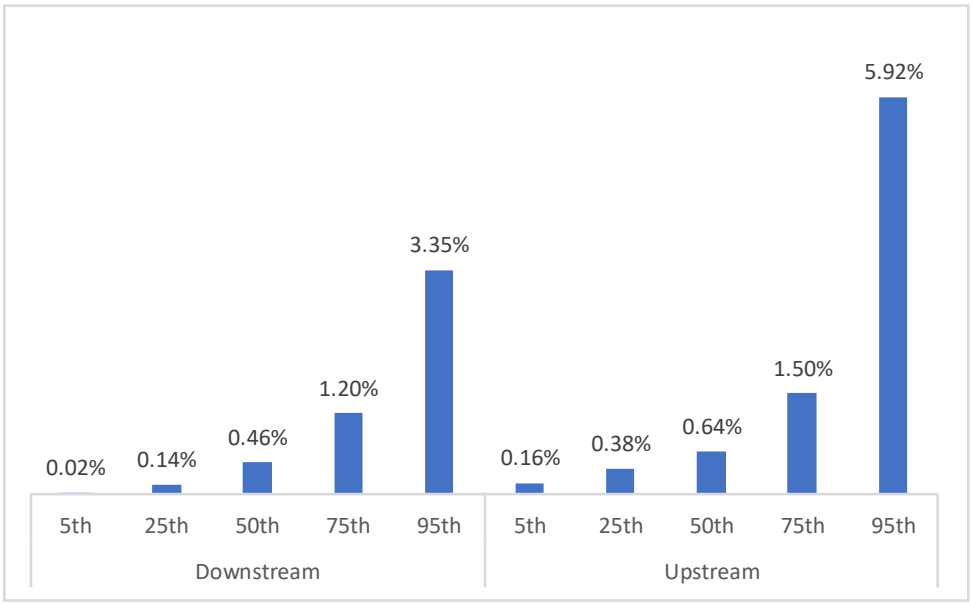


Figure 2: Sizes of Informal Firms Relative to an Average Formal Firm



Note: this figure plots the sizes of informal firms relative to an average formal firm. I select the informal firms at the 5^{th}, 25^{th}, 50^{th}, 75^{th}, and 95^{th} percentiles in terms of annual revenue. Then I calculate the relative size of these firms over the average annual revenue of formal firms.

Figure 3: Distribution of Formal and Informal Output

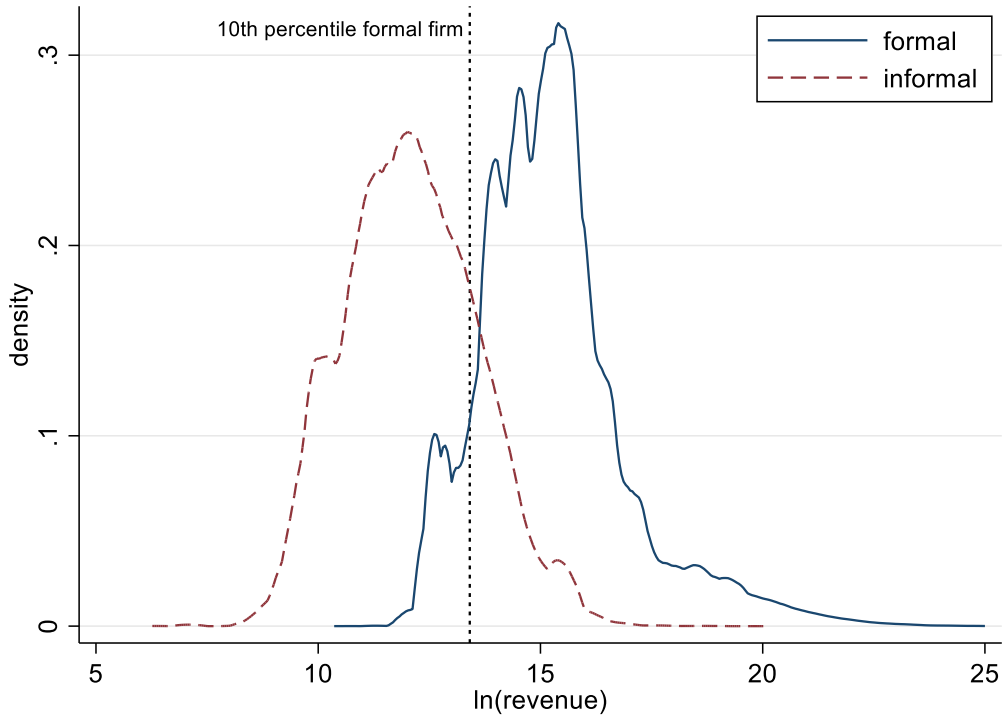


Figure 4: Trade Pattern

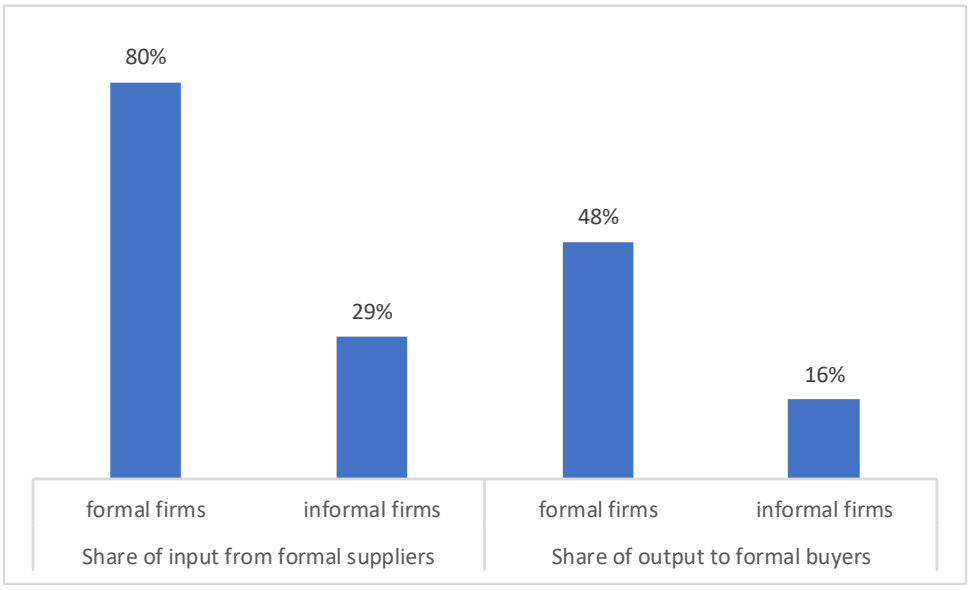
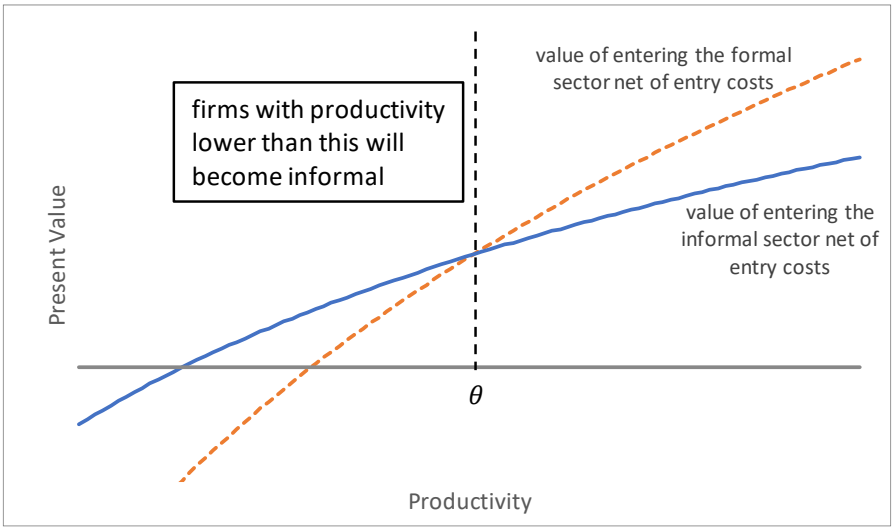
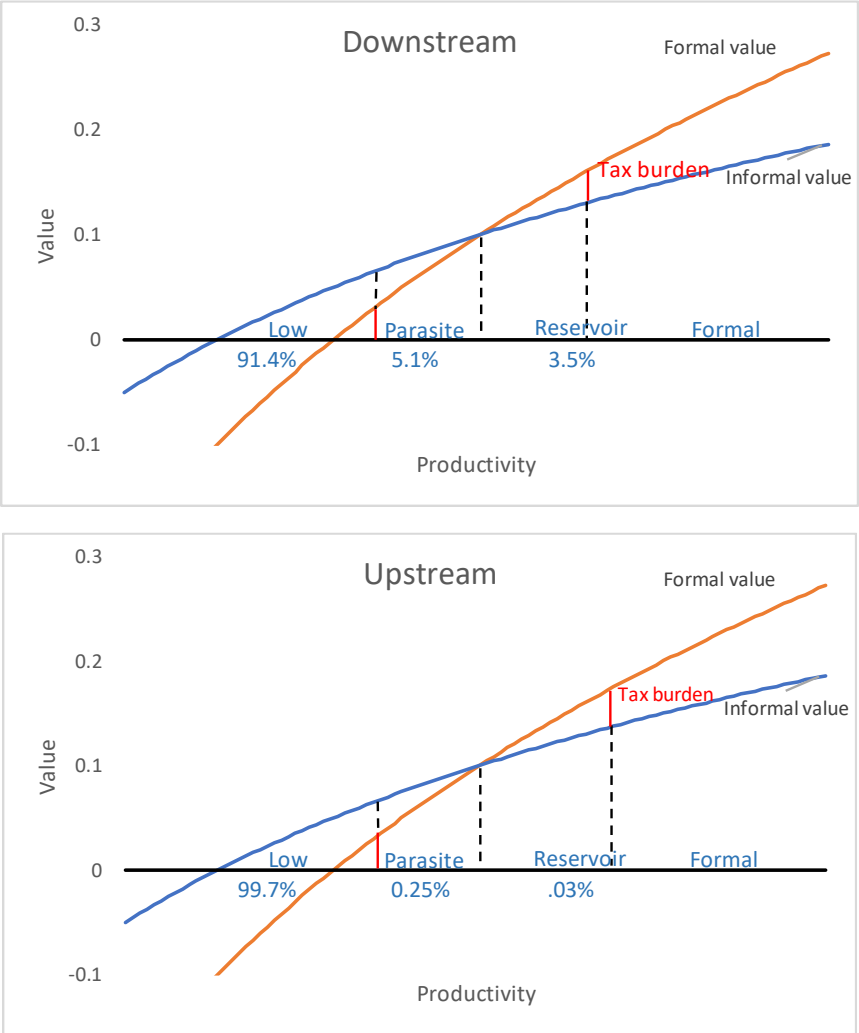


Figure 5: Illustrative Analysis of Formalization Choice



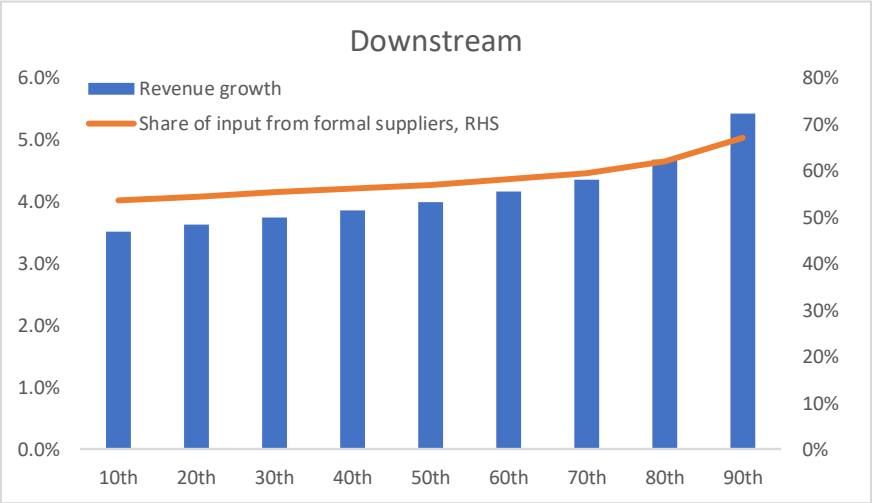
Note: this graph illustrates the framework by plotting the relationship between firm productivity and the present value of entering the formal and informal sector (net of entry costs). The blue line shows the value of entering the informal sector for each level of productivity, and the orange line similarly shows the value of entering the formal sector. The number of formal and informal firms depends on the distribution of productivity levels in the economy. Everything else being the same, an economy that has many low productivity firms (that is, those with productivity below θ) will have a relatively high level of informality.

Figure 6: Types of Informal Firms



Note: this graph illustrates the framework by plotting the relationship between firm productivity and the present value of entering the formal and informal sector (net of entry costs). The blue line shows the value of entering the informal sector for each level of productivity, and the orange line similarly shows the value of entering the formal sector. The number of formal and informal firms depends on the distribution of productivity levels in the economy. Everything else being the same, an economy that has many low productivity firms (that is, those with productivity below θ^1) will have a relatively high level of informality.

Figure 7: Revenue Growth by Size after VAT Adoption



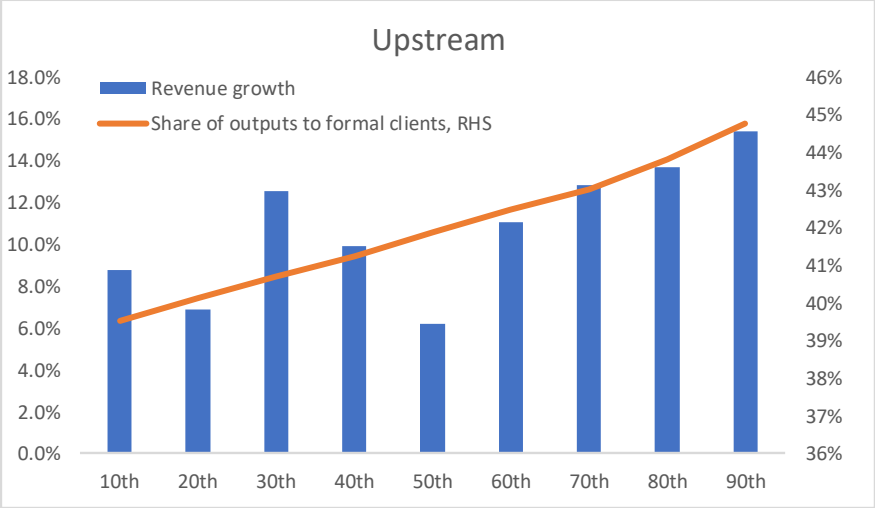
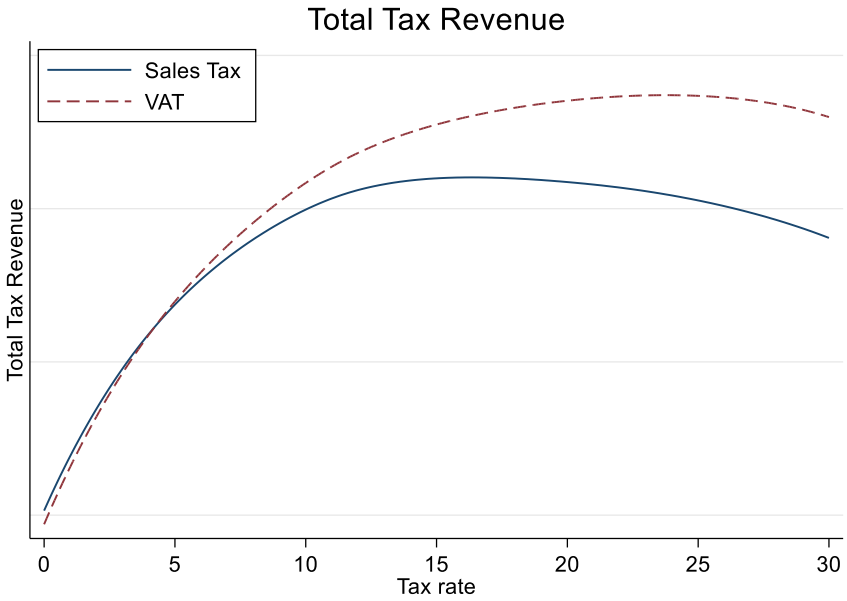


Figure 8: Tax Revenue and Tax Rate: VAT vs. Sales Tax



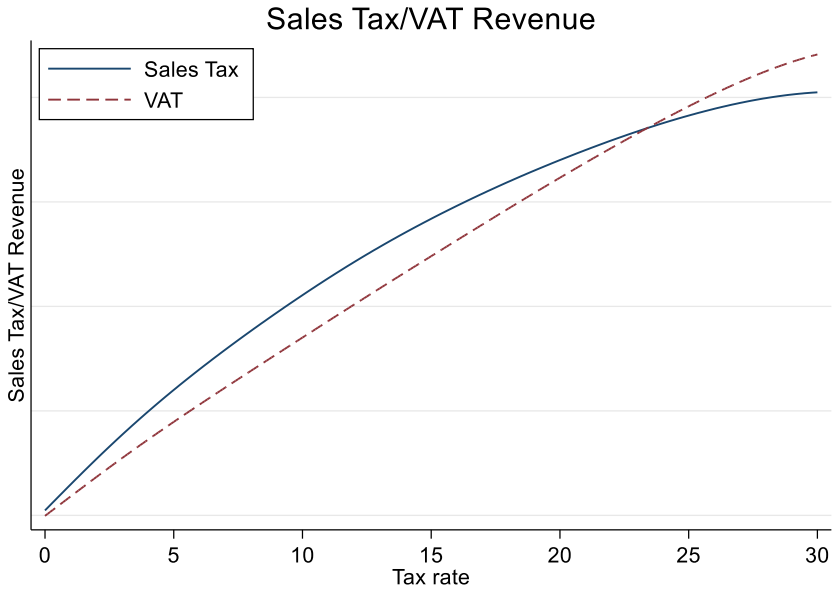


Table 1: ASI and UME Surveys (2015)

	Number	**Total labor, *million***	**Total revenue, *billion Rupees***	**Avg labor**	**Avg revenue, *million Rupees***
Tax registration					
Formal	817,944	19	68,718	25	84
Informal	5,325,092	26	6,224	5	1
Survey					
ASI	203,459	13	64,566	86	317
UME	5,939,577	31	10,376	5	2
Size					
>10 workers	591,592	21	67,094	39	113
<10 workers	5,551,444	24	7,848	4	1

Source: Pooled ASI and UME (2015).

Note: ASI is Annual Survey of Industries, UME is Unorganized Manufacturing Enterprises Survey. Revenue is expressed in Indian Rupee of 2015.

Table 2: Summary Statistics of State-Level Sample

	Obs	Mean	SD	Before	After	Diff
State-year sample						
Share of forma firms	138	0.065	0.117	0.074	0.058	-0.016
Share of formal output	138	0.710	0.309	0.559	0.838	0.279***
Share of formal labor	138	0.286	0.235	0.259	0.309	0.050*
State-industry-year sample						
Share of forma firms	1614	0.261	0.376	0.232	0.281	0.049***
Share of formal output	1589	0.677	0.336	0.570	0.754	0.184***
Share of formal labor	1598	0.460	0.380	0.415	0.493	0.078***

Source: Pooled ASI and UME (2000-2015).

Note: *** p<0.01, ** p<0.05, * p<0.1. ASI is Annual Survey of Industries, UME is Unorganized Manufacturing Enterprises Survey. Before refers to before VAT adoption, after refers to after VAT adoption. There are 35 states in India, Lakshwadadeep and Mizoram do not have registered manufacturing enterprises in 2015.

Table 3: Calibration and Estimation

Parameter	Description	Value	SE
Calibrate			
τ	VAT rate	0.125	-
τ^{corp}	Corporate tax rate	0.3	-
κ_f	Formal exit probability	0.08	-
SMM Estimate			
β	Formal input quality	1.15	0.09
ξ^d	Pareto shape parameter (downstream)	2.35	0.03
ξ^u	Pareto shape parameter (upstream)	1.40	0.03
E_i^d	Downstream informal entry cost scaled by avg downstream informal profit	0.51	0.01
E_f^d	Downstream formal entry cost scaled by avg downstream formal profit	1.08	0.04
E_i^u	Upstream formal entry cost scaled by avg upstream informal profit	1.22	0.06
E_f^u	Upstream informal entry cost scaled by avg upstream formal profit	1.80	0.16
α	production function coefficient	0.62	0.01
ρ	Substitution parameter	0.64	0.00
τ_l	wage gap	10.7	1.10
κ_i	Informal exit probability	0.15	0.01
σ	Post-entry shock variance	0.10	0.01

Table 4: Model Fit

Moments		**Data**	**Model**
Share of informality, #	downstream	98%	98%
	upstream	92%	93%
Share of informality, outputs	downstream	31%	30%
	upstream	16%	14%
Share of input from formal suppliers	formal firms	80%	80%
	informal firms	29%	43%
Share of output to formal buyers	formal firms	48%	48%
	informal firms	16%	28%
Distribution of downstream informal firm: relative size towards avg formal firm	5th percentile	0.02%	0.06%
	25th percentile	0.14%	0.22%
	50th percentile	0.46%	0.31%
	75th percentile	1.20%	0.66%
	95th percentile	3.35%	3.33%
Distribution of upstream informal firm: relative size towards avg formal firm	5th percentile	0.16%	0.05%
	25th percentile	0.38%	0.09%
	50th percentile	0.64%	0.19%
	75th percentile	1.50%	0.72%
	95th percentile	5.92%	7.92%
Distribution of downstream formal firm: relative size over avg formal firm	5th percentile	7.91%	15.15%
	10th percentile	9.70%	16.12%
Distribution of upstream formal firm: relative size over avg formal firm	5th percentile	1.42%	0.65%
	10th percentile	2.07%	0.72%
Standard deviation divided by mean	downstream	24.5	27.8
	upstream	24.4	37.8
Inputs/revenue	downstream	71%	53%
Share of informal labor	upstream	77%	81%

Table 5: Share of Informality

(F=Formal, I=Informal)

	Number Share		**Revenue Share**		**Labor Share**	
	F	I	F	I	F	I
ST, 7.5%	5.54%	94.46%	87.34%	12.66%	11.34%	88.66%
VAT, 7.5%	5.68%	94.32%	89.21%	10.79%	13.30%	86.70%
VAT, 17.5%	5.10%	94.90%	86.98%	13.02%	11.61%	88.39%
VAT, informality neutral	5.55%	94.45%	88.91%	11.09%	12.97%	87.03%
VAT, revenue neutral	5.52%	94.48%	88.72%	11.28%	12.88%	87.12%

Note: ST is cascading sales tax, VAT is value-added tax. Informality neutral rate is 9.5% so that the number of formal firms is the same to the benchmark scenario of cascading sales tax at the rate of 7.5%. Revenue-neutral rate is 10.1% so that the total value-added tax revenue is the same as the total cascading sales tax revenue.

Table 6: Change of the Formal and Informal Sector

(F=Formal, I=Informal)

	Number		**Revenue**	
	F	I	F	I
ST, 7.5%	0%	0%	0%	0%
VAT, 7.5%	2.0%	-0.6%	11.9%	-2.8%
VAT, 17.5%	-7.3%	1.3%	1.3%	6.4%
VAT, informality neutral	-0.1%	-0.2%	9.9%	-1.8%
VAT, revenue neutral	0.0%	0.4%	10.4%	0.6%

Note: ST is cascading sales tax, VAT is value-added tax. Informality neutral rate is 9.5% so that the number of formal firms is the same to the benchmark scenario of cascading sales tax at the rate of 7.5%. Revenue-neutral rate is 10.1% so that the total value-added tax revenue is the same as the total cascading sales tax revenue.

The change of number or revenue is the percent change from the benchmark scenario of cascading sales tax at the tax rate of 7.5%.

Table 7: Trade Pattern: VAT vs. Sales Tax

(F=Formal, I=Informal)

	Inputs		**Outputs**	
	F from F	I from F	F to F	I to F
ST, 7.5%	74.1%	44.9%	46.2%	29.9%
VAT, 7.5%	78.0%	44.9%	47.7%	29.9%
VAT, 17.5%	77.9%	38.3%	43.0%	24.7%
VAT, informality neutral	78.4%	43.9%	46.8%	28.8%
VAT, revenue neutral	76.3%	41.9%	46.5%	28.5%

Note: ST is cascading sales tax, VAT is value-added tax. Informality neutral rate is 9.5% so that the number of formal firms is the same to the benchmark scenario of cascading sales tax at the rate of 7.5%. Revenue-neutral rate is 10.1% so that the total value-added tax revenue is the same as the total cascading sales tax revenue.

Table 8: Change in Tax Revenue, Labor Income, and Corporate Profits

(F=Formal, I=Informal)

	VAT/ST	**Total tax**	**HH income**		**Corporate profit**	
	F	F	F	I	F	I
ST, 7.5%	0%	0%	0%	0%	0%	0%
VAT, 7.5%	-27%	-1%	20%	0%	8%	-2%
VAT, 17.5%	73%	25%	-1%	-4%	-1%	7%
VAT, informality neutral	-7%	4%	15%	-1%	6%	-2%
VAT, revenue neutral	0%	7%	15%	-1%	7%	0%

Note: ST is cascading sales tax, VAT is value-added tax. Informality neutral rate is 9.5% so that the number of formal firms is the same to the benchmark scenario of cascading sales tax at the rate of 7.5%. Revenue-neutral rate is 10.1% so that the total value-added tax revenue is the same as the total cascading sales tax revenue.

The change of number or revenue is the percent change from the benchmark scenario of cascading sales tax at the tax rate of 7.5%.

Appendix A: Downstream Index

Using the input-output table (IO table), Atras and Chor (2013) propose a measure to capture the "downstreamness" of an industry in the value chain. I provide a sketch of their measurement and document the application to the Indian IO table. The basic input-output identity is:

$$Y_i = F_i + Z_i$$

Where Y_i is the total output in industry i, F_i is the output of i that goes towards final use, Z_i is the use of i's output as inputs to other industries. In a world with N industries, this identity can be expanded as:

$$Y_i = F_i + \sum_{j=1}^{N} d_{ij}F_j + \sum_{j=1}^{N}\sum_{k=1}^{N} d_{ik}d_{kj}F_j + \sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N} d_{il}d_{lk}d_{kj}F_j + \cdots$$

Where d_{ij} for a pair of industries is the amount of i used as an input in producing one-dollar worth of industry j's output. $\sum_{j=1}^{N} d_{ij}F_j$ captures the value of i's "direct use" as an input by industry j to produce output that immediately goes to final use. The remaining terms $\sum_{j=1}^{N}\sum_{k=1}^{N} d_{ik}d_{kj}F_j + \sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N} d_{il}d_{lk}d_{kj}F_j + \cdots$ that involve higher-order summations reflect the "indirect use" of i as an input. These indirect inputs enter the value chain for future production, at least two production stages away from final use.

The downstream index measures the "distance" of an industry to the final products consumed by final consumers. In practice, the index is calculated by dividing the direct use of industry i's product as an input for final-use production by the total use of i's product as inputs for other industries. The higher the ratio for a given industry i, the more intensive is its use as a direct input for final-use production, so that the industry is closer to the

downstream stage of the value chain. Conversely, a lower ratio indicates that most of the contribution of input i to production processes occurs indirectly, and the industry is located more in the upstream stage of the value chain.

Based on their methodology, I use the Indian Input-Output Table to calculate the relative position of each industry. The IO table has been constructed for the year 2013-14 consistent with the national accounts estimates given the national accounts statistics (NSA) 2015. The IO table contains 140 rows (products) and 67 columns (sector), which have been collapsed and expanded to make the 130*130 input-output table by Singh and Saluja (2016)[10].

First, I calculate the N*N direct requirement matrix D, the N*1 final-use vector F by summing over the value of each industry i's output purchased for consumption and investment by private or government entities, and the N*1 output vector Y as the summation of all entries in row i in the IO table. Then I calculate the direct use of each industry as DF, and the total input use for each industry as $Y - F$, I divide the ith element of the direct use vector DF by the corresponding ith element of the input vector Y-F, which generates the downstream index of industry i.

Table A1 shows the downstream index and the share of informality in the largest ten industries. Informal firms are prevalent among the miscellaneous food producers, textile producers, and leather producers, whose share of informal firms are 89%, 94%, and 93% (and 11%, 8%, and 31% of total output), respectively. At the same time, these industries are also closer to the final consumers with their downstream indices higher than average. On the other side, there are less informal firms among petroleum producers and metallic mineral producers (Iron, steel and ferroalloys). The share of informal firms is 48% and 27% (less than 1% of total output) for these two industries respectively. These two industries are mainly operating in the upstream production stages.

[10] For detailed construction of the symmetric IO table, see Singh and Saluja (2016).

Table A1: Share of Informal Firms in The Largest 10 Manufacturing Industries

Industry	Downstream Index	Share of informal firm	Share of informal sector output
Petroleum products	0.85	48%	0%
Miscellaneous food products	1.20	89%	17%
Iron, steel and ferro alloys	0.65	27%	0%
Motor vehicles	0.37	63%	1%
Miscellaneous manufacturing	0.76	89%	11%
Miscellaneous metal products	0.44	80%	21%
Drugs and medicines	1.69	55%	0%
Cotton textiles	1.12	94%	8%
Plastic products	1.00	64%	4%
Leather and leather products	1.20	93%	31%

Source: Annual survey of industries (2015), Unorganized Manufacturing Enterprises survey (2015).

An interesting phenomenon is a weakly positive correlation between downstream index and the share of informality (Figure A1). The correlation between downstream index and informality might not be a coincident. Briand and Hoseini (2020) find that the informality is more salient in downstream industries. They calculate the backward linkage as a proxy for the position of an industry in the production chain. The larger the backward linkage, the longer chain of inputs, which is similar to the downstream index in my paper. They argue that a cascading sales tax distorts the product prices, which tends to accumulate as the value chain moves towards the downstream industries. Since downstream industries typically have strong backward linkages due to more production stages, their product prices deviate more from optimal prices as a result of cumulative taxes. Higher prices of formal products leave bigger space for cheaper informal products whose prices are not distorted by tax wedges. Hence, informal firms are more prevalent in the downstream industries than upstream ones. However, the linkage of downstream index and informality is beyond the scope of this paper. I will treat the difference of informality between the upstream and the downstream as stylized facts without causal analysis. For an illustrative description of tax cascading, see Appendix C.

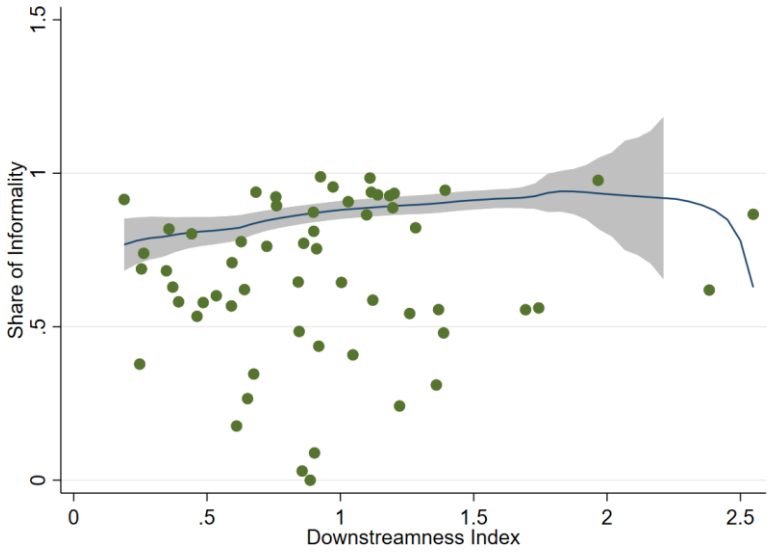


Figure A1: Share of Informal Firms and Downstream Index. Source: Annual survey of industries (2015), Unorganized Manufacturing Enterprises survey (2015).

Table A2: Downstream Index and The Share of Informal Firms

Industry	**Downstream Index**	**Share of informal firm**	**Share of informal sector output**
Petroleum products	0.85	48%	0%
Miscellaneous food products	1.20	89%	17%
Iron, steel and ferro alloys	0.65	27%	0%
Motor vehicles	0.37	63%	1%
Miscellaneous manufacturing	0.76	89%	11%
Miscellaneous metal products	0.44	80%	21%
Drugs and medicines	1.69	55%	0%
Cotton textiles	1.12	94%	8%
Plastic products	1.00	64%	4%
Leather and leather products	1.20	93%	31%
Non-ferrous basic metals	0.25	38%	1%
Other non-metallic mineral prods.	0.90	87%	23%
Cement	1.22	24%	0%
Other non-electrical machinery	0.84	65%	4%
Organic heavy chemicals	0.36	82%	0%
Motorcycles and scooters	0.86	3%	0%
Edible oils other than vanaspati	1.26	54%	1%
Beverages	1.28	82%	2%
Industrial machinery(others)	0.68	94%	51%

Electrical industrial Machinery	0.64	62%	2%
Paper, paper prods. & newsprint	0.59	71%	4%
Sugar	1.11	98%	13%
Soaps, cosmetics & glycerin	0.91	75%	3%
Fertilizers	1.37	55%	0%
Synthetic fibers, resin	0.92	43%	1%
Iron and steel casting & forging	0.25	69%	1%
Rubber products	1.12	58%	2%
Miscellaneous textile products	1.39	94%	48%
Art silk, synthetic fiber textiles	1.10	86%	1%
Electrical wires & cables	0.63	77%	2%
Tractors and agri. implements	2.38	62%	4%
Printing and publishing	0.26	74%	21%
Tobacco products	1.97	98%	14%
Wood and wood products	1.19	93%	46%
Structural clay products	0.90	79%	43%
Other chemicals	0.68	35%	1%
Electrical appliances	0.39	58%	2%
Paints, varnishes and lacquers	0.90	9%	0%
Silk textiles	0.93	99%	21%
Pesticides	1.36	31%	0%
Leather footwear	0.97	96%	17%
Furniture and fixtures-wooden	0.76	92%	77%
Readymade garments	0.19	91%	13%
Communication equipment	0.61	18%	1%
Batteries	0.59	57%	13%
Electronic equipment (incl.TV)	0.46	53%	0%
Hand tools, hardware	0.86	77%	16%
Hydrogenated oil(vanaspati)	1.74	56%	0%
Other electrical Machinery	0.72	76%	7%
Industrial machinery (F & T)	0.53	60%	9%
Machine tools	0.35	68%	1%
Carpet weaving	1.03	91%	13%
Jute, hemp, mesta textiles	1.39	48%	2%
Bicycles, cycle-rickshaw	0.49	58%	8%
Woolen textiles	1.05	41%	4%
Other transport equipment	2.55	86%	4%
Ships and boats	4.52	75%	3%
Khandsari, boora	1.14	93%	4%
Rail equipment	0.89	0%	0%

Source: Annual survey of industries (2015), Unorganized Manufacturing Enterprises survey (2015).

Appendix B: Calculating the Share of Formal Input and Output

First, I restrict the firm-level data to the manufacturing sector in West Bengal province whose transaction data is available from the local tax authority. Next, I divide firms into four categories, including downstream formal firms, downstream informal firms, upstream formal firms, and upstream informal firms. Then I aggregate the total revenue and total expense for each category using the sample weights of each firm. Revenue F denotes total revenue of formal firms, Revenue I denotes total revenue of informal firms, Expense F denotes total expense of formal firms, Expense I denotes total expense of informal firms.

Share of sales from formal sellers to formal buyers. Among the total revenue of formal firms (Revenue F), I need to calculate the share of revenue from formal firms to formal buyers. Revenue FF denotes the revenue from formal sellers to formal buyers. Here, I use the transaction level data from the tax authority that covers all the transactions of formal firms in the province. Formal firms report all their sales to other formal firms as required by the regulation. I aggregate the total sales from formal sellers to formal buyers (Revenue FF) and divide it by the total revenue of formal firms (Revenue F). This yields the share of sales from formal sellers to the formal buyers.

Share of inputs of formal firms from formal sellers. Among the total inputs of formal firms (Expense F), I need to calculate the share of inputs of formal firms from formal sellers. Expense FF denotes the total inputs of formal firms from formal sellers. Formal firms have to report all their purchases from other formal firms to the tax authority. I aggregate the total inputs of formal buyers from other formal sellers (Expense FF) and divide it by the total inputs of formal firms (Expense F). This yields the share of inputs of formal firms from formal sellers.

Share of sales from informal firms to formal buyers. Among the total revenue of informal firms (Revenue I), I need to calculate the share of revenue from informal firms to formal

buyers. Revenue IF denotes the total revenue from informal sellers to formal buyers. Since I do not have the transaction level data for informal firms, I cannot use the same aggregation to get the total revenue from informal firms to formal buyers. However, I have calculated the total revenue from formal firms to formal buyers (Revenue FF) and the total inputs of formal firms (Expense F). Note that formal firms purchase inputs from either formal sellers or informal sellers, which means Expense F=Revenue FF + Revenue IF. I use the identity and distract the inputs of formal buyers (Expense F) by the total sales from formal sellers to formal buyers (Revenue FF). This yields the total sales from informal sellers to the formal buyers (Revenue IF). Divided sales from informal sellers to the formal buyers (Revenue IF) by total sales of informal firms (Revenue I) yields to share of sales from informal sellers to the formal buyers.

Share of inputs of informal firms from formal sellers. Among the total inputs of informal firms (Expense I), I need to calculate the share of inputs of informal firms from formal sellers. Since there is no transaction level data from the informal firms, aggregation is not possible for the informal firms. But I have calculated the total inputs of informal firms (Expense I), total sales of formal sellers (Revenue F) and the total sales from formal sellers to the formal buyers (Revenue FF). I calculate the total sales of formal sellers to the informal buyers (Revenue FI) using the identity (Revenue FI + Revenue FF = Revenue F). Dividing the revenue from formal sellers to the informal buyers (Revenue FI) by the total inputs of informal buyers (Expense I), I calculate the share of inputs of informal buyers from formal sellers.

Note that households also purchase products from upstream suppliers. I assume a fixed share of upstream sales goes to household, which equals the share of final consumption in the IO table.

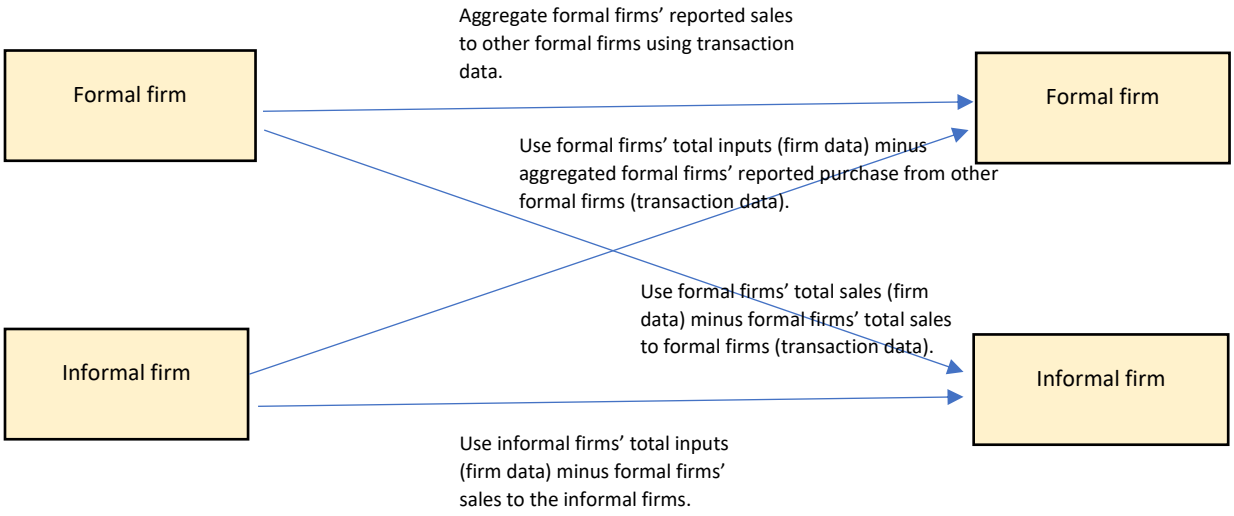


Figure B1. Measuring Transaction between Formal and Informal Sectors

Appendix C1: Background of Indian Tax Reform

India adopted VAT between 2003 and 2014 with variation across states. Both direct tax and indirect tax are levied in India. Direct taxes are mainly collected by the central government with the exception of profession and property taxes; indirect taxes are levied by both the central and the state governments. The service tax, custom duties and union excise duties which include all manufacturing products are the prerogative of the federal government. The only indirect tax at state level is the VAT which comprises more than half of total states' own tax revenue.

Prior to the adoption of VAT, the main source of tax revenue among states is the sales tax which explains the reluctance of states to reform it despite its flaws being regularly underlined. The sales tax in Indian states were generally cascading even under the two general rules that intend to differentiate between final and intermediate consumption. According to the "physical ingredient" rule, an "input" is a material or component that is physically incorporated in the goods destined to be sold. According to the "direct use" rule, items directly used in the production of goods are exempted from sales taxes. However, these rules are neither easy to apply nor guaranteeing the non-taxation of business inputs. there exists heterogeneity in tax treatment of inputs among the different Indian states: raw materials are exempted in only a few states (Delhi and Gujarat) but due to the widespread adoption of the "physical ingredient" rule, fuels, tools, machinery and equipment are also not considered as inputs in these states. Bachi (1994) and Ring (1998) estimate that 30%-40% of sales tax revenue is collected from intermediate inputs.

The VAT reform is accompanied with other policies such as lower registration costs, a change of registration threshold, and a change of tax rate. Figure below shows the VAT rates after reform are higher than the existing sales tax rates in most Indian states. At the same time, the threshold of VAT registration is also higher than that of sales tax registration, which suggests higher tolerance of informality under the VAT regime.

Figure C1: Tax Rate before and after VAT Reform

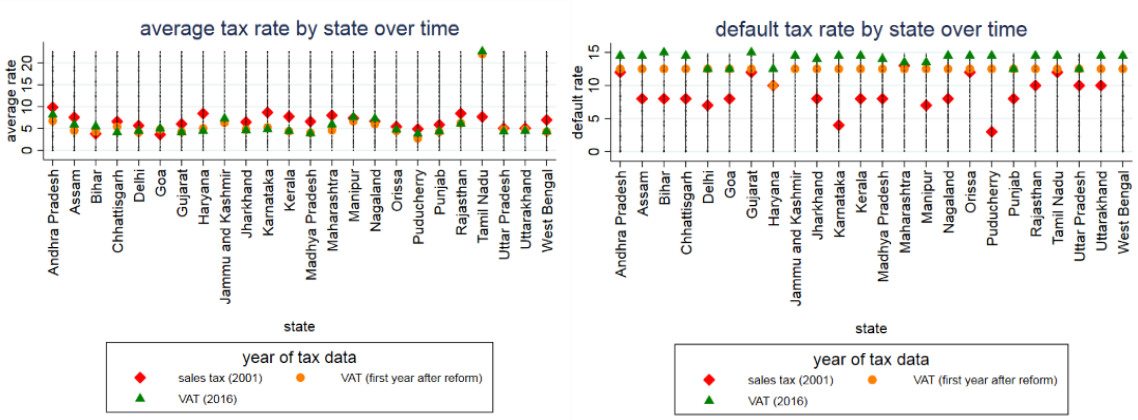
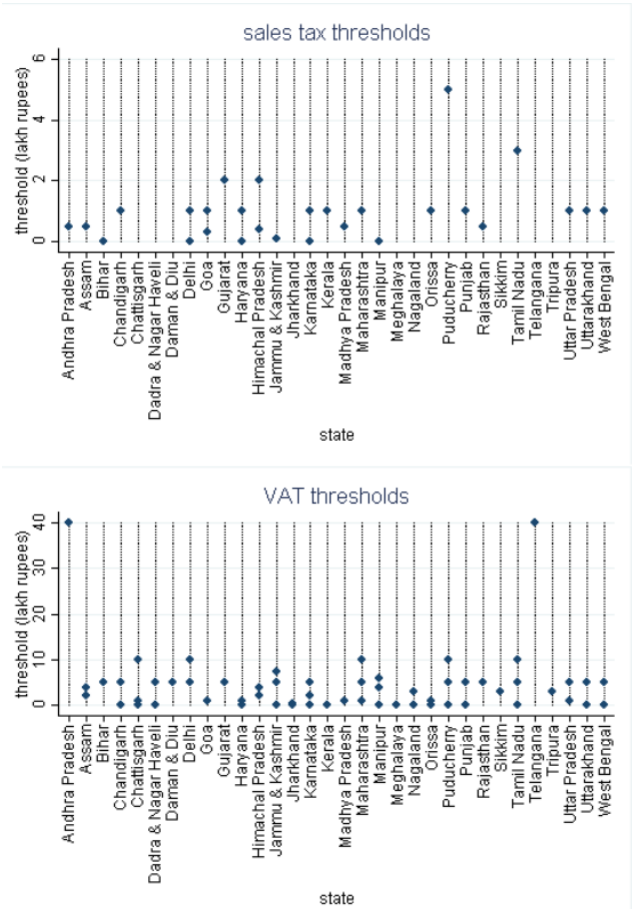


Figure C2: Tax Registration Threshold before and after VAT Reform



In the context of West Bengal, it is a large state in east part of India with 90 million residents. The total GDP in West Bengal is $210 billion in 2019-2020, which accounts for 7% of the national GDP. West Bengal implemented VAT in 2003 with tax regime border. West Bengal implemented VAT since June 1st, 2003.[11] All firms with a turnover of more than 500,000 INR (threshold) are required to remit tax to the state. Firms with a turnover of less than 5 million INR (border) can opt to remit tax under a simplified tax scheme under which they only pay a 0.025% tax on their total sales. However, firms in the simplified scheme cannot deduct tax paid by their suppliers from their tax liabilities.

Firms face different VAT rates depending on the goods they sell: 75% of them sell goods belonging to the "reduced" tax schedule and taxed at 4%, 21% sell goods in the "main" tax schedule taxed at 12.5%, and the remainder of firms are taxed at super-reduced rates of 0%-1%. In 2014, the VAT rates of the main and reduced schedules increased by 1 percentage points. For simplicity, this paper does not distinguish different tax schedule and assume that all firms pay a flat VAT rate at 12.5%.

Appendix C2: Illustration on Tax Cascading and Its Distortive Effect

VAT is an indirect tax levied on the value-added at each stage of the value chain. A firm's VAT tax base is the value of output less any of the costs of inputs that have already been taxed. Contrary to the sales tax that suffers from enforcement problem, VAT introduces the third-party reporting, verifiable paper trails, and whistle-blowers that play an important role in facilitating tax enforcement.

Non-cascading sales tax only applies to the final consumers rather than the business owners who purchase the goods as their inputs. In principle, only final consumers pay the sales tax on their purchases. Business owners who purchase inputs for resales should issue resale

[11] https://www.rediff.com/money/2003/apr/08vat2.htm

certificates or exemption certificate to their sellers so that they do not pay tax on the purchases.

Cascading sales tax. However, sales tax often suffers from enforcement problems. Sellers cannot always distinguish final consumers from business owners when sell to these customers—and they have little incentive to find out. In addition, tax authorities prefer collecting tax at the earlier stage of production from large-scale manufacturers or wholesalers rather than the later stage from a mass of retailers. If a seller does not impose a sales tax on consumer purchases, it incurs the tax evasion; if a seller imposes a tax on business purchases, the tax "cascades", which means that the sales tax is paid twice by both the seller and the buyer.

To illustrate the price distortion between formal inputs and informal inputs, consider a 3-stage production chain as in Table J1. The output price of upstream product 1 is α_1 under VAT and cascading ST. For the midstream firms, let γ_1 denotes the units of upstream product 1 used to produce 1 unit of midstream product 2. Assume the markup is α_2 and tax rate is τ. Hence, the output price of midstream product 2 is markup plus intermediate input and tax. VAT applies to the upstream product and the original purchase price of product 1 is $\alpha_2 + (1+\tau)\alpha_1\gamma_1$. However, midstream firms can get tax credit of $\tau\alpha_1\gamma_1$ if they purchase from formal sellers. Hence, the net-of-tax output price of midstream firms is also $\alpha_2 + \alpha_1\gamma_1$. For the cascading sales tax, the input price is $\alpha_1\gamma_1$. Since there is no tax deduction, their net-of-tax output price is $\alpha_2 + (1+\tau)\alpha_1\gamma_1$. Hence, the output price of midstream product is distorted under cascading ST, which will be carried to the downstream who uses product 2 as its input.

Table C1: Illustration of Exclusive Price of Output under VAT and Cascading Sales Tax

	Markup	Output prices, cascading sales tax	Output prices, VAT
Upstream, product 1	α_1	α_1	α_1
Midstream, product 2	α_2	$\alpha_2 + (1+\tau)\alpha_1\gamma_1$	$\alpha_2 + \alpha_1\gamma_1$
Downstream, product 3	α_3	$\alpha_3 + (1+\tau)[\alpha_2 + (1+\tau)\alpha_1\gamma_1]\gamma_2$	$\alpha_3 + \alpha_2\gamma_2 + \alpha_1\gamma_1\gamma_2$

Source: Briand and Hoseini, 2020

Since the input prices are distorted under cascading sales tax, firms will also change their input combination to maximize their profits. Assume downstream producers can select input materials from both formal and informal upstream sellers, cascading sales tax raises the price of intermediate goods from formal suppliers compared to informal ones. Hence, downstream firms will increase their purchase from the informal sellers despite a better quality of the formal products, leading to a productivity loss and social welfare loss.

Cascading sales tax has a more significant on the input choice between formal and informal goods. Higher cascading tax significantly reduces the share of formal goods used in the formal firms' production, implying an uneven advantage of informal goods due to tax evasion. The true distortion of cascading sales tax takes the form of excessive use of informal goods in the production that might reduce the overall production efficiency rather than the larger share of informal firms. Under cascading sales tax, firms either favor purchase cheaper inputs from informal sellers or pay a higher expense on formal inputs that squeezes the employment or the investment. The tax builds up over each successive stage of production, incurring market distortions and deadweight loss (Keen, 2014).

Appendix C3: A Natural Experiment in Indian VAT Reform

I use the diff-in-diff specification to identify the effect of VAT reform on informality across Indian states. The model to estimate is given by:

$$y_{st} = \beta_0 + \beta_1 \cdot VAT_{st} + \alpha_s + \tau_t + \varepsilon_{st}$$

where y_{st} is one of the formality measurement (share of formal firms, share of formal sector output, share of workers employed in the formal sector) in state s and year t, VAT_{st} is a VAT adoption dummy that equals one when state s adopts VAT in year t, α_s and τ_t are state dummy and year dummy, respectively.

Table C2 reports the results. The adoption of VAT does not have a significant impact on the extensive margin (column 1) of informality. This is consistent with the findings of Briand and Hoseini (2020) that on average the change in the share of formal firms is close to zero although there exists significant heterogeneity between industries.

However, VAT adoption increases intensive margin of the formal sector. The share of formal output increases by 10 percentage points (column 2) and the share of workers employed in the formal sector increases by 8 percentage points (column 3) after VAT reforms. This suggests that the VAT benefits the incumbent formal firms and expands their production and employment.

Table C2: Diff-in-Diff Results

VARIABLES	(1) ShareN	(2) ShareY	(3) ShareL
VAT	0.016	0.104*	0.084*
	(0.023)	(0.059)	(0.046)
Constant	0.070***	0.434***	0.258***
	(0.005)	(0.036)	(0.020)
State FE	Y	Y	Y
Year FE	Y	Y	Y
Observations	138	138	138
R-squared	0.859	0.764	0.817
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1			

Note: The state-level sample is aggregated from firm-level ASI-UME sample in 2000, 2005, 2010, and 2015. The dependent variables in columns 1 to 3 are the share of formal firms, the share of formal output, and the share of workers employed in the formal sector. VAT is a dummy variable that equals 1 if a state adopted VAT and 0 if otherwise.

I then estimate the above equation but replace the dependent variable with log number of formal (informal) firms, the log formal (informal) sector output, and the log workers employed in the formal (informal) sector. This will identify the impact of VAT adoption on the formal and informal sector, respectively.

The adoption of VAT has insignificant effects on the number of formal firms, but has significant effects on the formal sector output and the formal sector workers. Formal sector output increases by 70% and formal sector workers increases by 44% after VAT adoption. On the contrary, VAT adoption has little impact on the informal sector, including the number of informal firms, the informal sector output, and the informal sector employment.

Table C3: VAT Adoption and Informality in Indian States

VARIABLES	(1) logN	(2) logY	(3) logL	(4) logNinf	(5) logYinf	(6) logLinf
VAT	0.441	0.703**	0.699**	-0.089	-0.027	-0.120
	(0.303)	(0.352)	(0.293)	(0.155)	(0.251)	(0.173)
Constant	8.030***	25.015***	10.627***	11.257***	25.389***	12.018***
	(0.120)	(0.202)	(0.169)	(0.058)	(0.113)	(0.068)
State FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Observations	138	138	138	140	140	140
R-squared	0.932	0.942	0.937	0.987	0.945	0.984

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

Source: ASI and UME (2000, 2005, 2010, 2015).

Note: The dependent variables in columns 1 to 6 are log number of formal firms, log formal sector output, log formal sector employment, log number of informal firms, log informal sector output, log informal sector employment.

Then I aggregate firm-level data to 2-digit state-industry-year data to increases the identification power as in Briand and Hoseini (2020). The model is given by:

$$y_{sit} = \beta_0 + \beta_1 \cdot VAT_{st} + \alpha_{si} + \tau_t + \varepsilon_{st}$$

where y_{sit} is one of the formality measurement (share of formal firms, share of formal sector output, share of workers employed in the formal sector) in state s and year t, VAT_{st} is a VAT adoption dummy that equals one when state s adopts VAT in year t, α_{si} and τ_t are state-industry dummy and year dummy, respectively.

Table C4 reports the results. After VAT adoption, the share of formal firms increases by 3 percentage points while the share of formal sector output increases by 4.3 percentage points. The share of formal sector employment expands by 4 percentage points. The results are consistent with the previous state-year level estimation.

Table C4: Diff-in-Diff Results at State-Industry Level

VARIABLES	(1) ShareN	(2) ShareY	(3) ShareL
VAT	0.029**	0.043***	0.040**
	(0.012)	(0.014)	(0.017)
Constant	0.161***	0.651***	0.434***
	(0.004)	(0.008)	(0.007)
State-Industry FE	Y	Y	Y
Year FE	Y	Y	Y
Observations	2,100	2,027	2,046
R-squared	0.886	0.914	0.883
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1			

Note: The state-level sample is aggregated from firm-level ASI-UME sample in 2000, 2005, 2010, and 2015. The dependent variables in columns 1 to 3 are the share of formal firms, the share of formal output, and the share of workers employed in the formal sector. VAT is a dummy variable that equals 1 if a state adopted VAT and 0 if otherwise.

Appendix C4: Global VAT Adoption

The adoption of VAT in India is not an exception. In fact, more than 160 countries have adopted VAT, which accounts for more than 20% of the global tax revenue (Keen and Lockwood, 2010). Many of the recent tax reforms are characterized by the adoption of VAT, such as the GST reform in India that intends to unify the national VAT tax code. The trend was particularly significant in the past thirty years during which more than 50 countries adopted VAT to replace trade duties or sales tax.

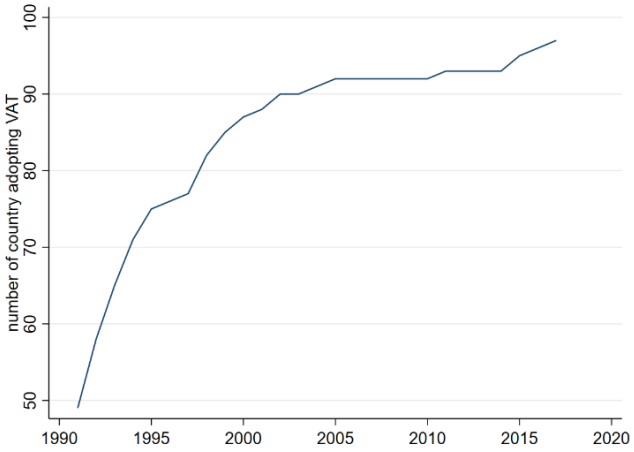


Figure C3: Number of Countries that Adopted VAT. Source: E&Y VAT report (2018).

VAT rates varies significantly across regions and countries, which confounds the adoption of VAT regime and adjustment of VAT rate. VAT rates are highest among high-income countries with the median tax rate at 20%, while the lowest VAT rate is only 5% in some

countries. The simple diff-in-diff method is not able to answer whether it is the VAT regime adoption or simply the tax rate adjustment that affects informality.

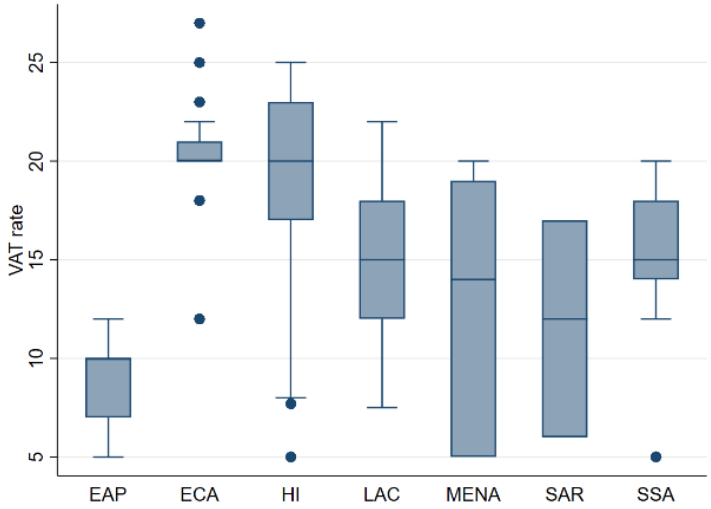


Figure C4: VAT Rate Difference by Region. Source: E&Y VAT report (2018).

Appendix C4: International Experience on VAT adoption and Informality

Since more than 50 countries adopted VAT since 1990, a global experience will be a good starting point to study the impact of VAT adoption on informality. Using a country-year panel data from Medina and Schneider (2018), I first implement the traditional diff-in-diff method to identify the shock after VAT adoption since 1990. I estimate the following equation:

$$Inf_{it} = \beta_0 + \alpha_i + \alpha_t + \beta_1 \cdot VAT\ adoption_{it} + \varepsilon_{it}$$

Where Inf_{it} is the share of output by the informal sector in country i and year t, α_i is country fixed effect and α_t is year fixed effect, $VAT\ adoption_{it}$ is a dummy variable that equals 1 after country i adopts VAT and 0 if a country has not or does not adopt VAT. The coefficient of interest is β_1, which captures the net effect of VAT adoption on the share of

informal outputs. The sample includes 109 developing countries from 1991 to 2017, out of which 55 countries adopted VAT during the period.

The adoption of VAT would lead to a drop of the share of informal outputs by 0.33 percentage points on average as shown in column 1. To capture the time lag for VAT to take effect, I augment the interaction term between the number of years before and after VAT adoption into the baseline regression[12]. On average, the drop in informality takes more than 5 years to materialize (column 2). The share of informality does not change in the first four years after VAT adoption but decreases by 0.49 percentage points after 5 years.

Column 3 and 4 replace the share of informality with log value of informal outputs as the dependent variable. This will capture the change in the total volume of informal output. The adoption of VAT has a negative impact on the informal outputs by 4.6% in developing countries. The effect is significant only in the longer term.

[12] I augment seven interaction terms into the equation to compare the evolvement of informality against that prior to VAT adoption. I interact the treatment group dummy with five dummy variables corresponding to the i th year after VAT adoption. If VAT adoption generates a consecutive decrease of informality in the next five years, it would be captured by these interaction terms.

Table C5: Net Effect of VAT Adoption on Informality

VARIABLES	(1) Baseline	(2) Post reform	(3) Baseline	(4) Post reform
	Share of informal output		Ln(Informal output)	
VAT adoption effect	-0.330*		-0.046**	
	(0.192)		(0.019)	
Treatment*1 year before VAT adoption		0.144		0.006
		(0.353)		(0.034)
Treatment*0 year after VAT adoption		0.024		0.020
		(0.358)		(0.034)
Treatment*2 year after VAT adoption		-0.104		-0.006
		(0.365)		(0.035)
Treatment*3 year after VAT adoption		0.090		-0.024
		(0.372)		(0.035)
Treatment*4 year after VAT adoption		-0.205		-0.046
		(0.384)		(0.036)
Treatment*5 year after VAT adoption		0.290		-0.042
		(0.386)		(0.037)
Treatment*5 year after VAT adoption		-0.490**		-0.048**
		(0.241)		(0.024)
Country FE	Y	Y	Y	Y
Year FE	Y	Y	Y	y
Observations	2,916	2,916	2,825	2,825
R-squared	0.966	0.966	0.987	0.987

Source: Medina and Schneider (2018) and World Development Indicators.

Appendix D: Model Solution

Downstream formal firm

Formal buyers receive tax credit from formal input at the rate of τ but not from informal input. The formal inputs give tax credit at the rate of τ. Here I impose both VAT and corporate tax into the model.[13] Corporate tax is imposed on the net profit of formal firms in both the downstream and the upstream. VAT is levied on the gross margin at the downstream with tax deduction if a formal buyer purchases from a formal supplier. Informal buyers and final consumers cannot deduct tax of intermediate input. Compared to corporate tax, VAT is levied on sales of formal firms but can be deducted by purchasing from other formal sellers. This creates an incentive for formal buyers to purchase from formal sellers than informal sellers if the price difference between formal input and informal input is less than the tax deduction, which might lead to the market segmentation between formal and informal sector as depicted in the stylized facts.

Solving equation (1) and (2), the demand function of $x(i)$ for upstream firm i is:

$$x^f(i) = [1-\tau(i)]^{\frac{1}{\rho-1}}\beta(i)^{-\frac{1}{\rho-1}}[\frac{y_0^f}{\theta^d}]^{\frac{1}{\alpha}} \cdot \left[\frac{p^f(i)}{P^f}\right]^{\frac{1}{\rho-1}} \quad (27)$$

With price index for downstream formal firms P^f:

$$P^f = [\int_0^N [(1-\tau(i))p^f(i)]^{\frac{\rho}{\rho-1}}\beta(i)^{-\frac{1}{\rho-1}}di]^{\frac{\rho-1}{\rho}} \quad (28)$$

[13] https://www.business-standard.com/article/pf/10-taxes-you-should-know-about-114041100175_1.html

$$P^f = [\int_{\underline{\theta^u}}^{\overline{\theta^u}} N[p_i^f]^{\frac{\rho}{\rho-1}} dF(\delta) + \int_{\overline{\theta^u}}^{\infty} N[(1-\tau)p_f^f]^{\frac{\rho}{\rho-1}} \beta^{-\frac{1}{\rho-1}} dF(\theta^u)]^{\frac{\rho-1}{\rho}} \quad (29)$$

Here, θ^u denotes the productivity of an upstream firm that sells to the buyer. I assume that downstream firms can observe the pre-entry productivity of upstream firms and form an initial expectation of the tax status of upstream suppliers, which coincides with the pre-entry productivity in the equilibrium. Specifically, downstream firms expect an upstream firm being formal if its pre-entry productivity exceeds the upper bar of $\overline{\theta^u}$, and an upstream firm being informal if its productivity exceeds the lower bar of $\underline{\theta^u}$, but is below the higher bar of $\overline{\theta^u}$. In the equilibrium, $\underline{\theta^u}$ is the realized lower bar of pre-entry productivity of informal firms, and $\overline{\theta^u}$ is the realized lower bar of pre-entry productivity of formal firms.

Then firms choose the optimal output to maximize their profits:

$$max(1-\tau^{corp})\{y_0 - \int_0^N x(i)p^f(i)[1-\tau(i)]di\} \quad (30)$$

Which yields the optimal output as:

$$y_0^f = \theta^d[\frac{(1-\tau)\theta^d\alpha}{Pf}]^{\frac{\alpha}{1-\alpha}} \quad (31)$$

Hence, the net profit of formal downstream firm is:

$$\pi_f^d = (1-\tau^{corp})\{y_0^f - \int_0^N x(i)p^f(i)[1-\tau(i)]di\} \quad (32)$$

Downstream informal firms

Solving the cost minimization problem from equation (3) and (4) yields the demand function of $x(i)$ for an upstream informal firm i:

$$x^i(i) = \beta(i)^{-\frac{1}{\rho-1}} \cdot \frac{y_0^{i\frac{1}{\alpha}}}{\theta^d} \cdot \left[\frac{p^i(i)}{P^i}\right]^{\frac{1}{\rho-1}} \quad (33)$$

With price index P^ias:

$$P^i = [\int_0^N [p^i(i)]^{\frac{\rho}{\rho-1}} \beta(i)^{-\frac{1}{\rho-1}} di]^{\frac{\rho-1}{\rho}} \quad (34)$$

$$P^i = \{\int_{\underline{\theta^u}}^{\overline{\theta^u}} N\left[p_i^i\right]^{\frac{\rho}{\rho-1}} dF(\theta^u) + \int_{\overline{\theta^u}}^{\infty} N\left[p_f^i\right]^{\frac{\rho}{\rho-1}} \beta^{-\frac{1}{\rho-1}} dF(\theta^u)\}^{\frac{\rho-1}{\rho}} \quad (35)$$

Then, downstream firm choose the optimal output to maximize its profit:

$$max\{y_0 - b_i \int_0^N x(i) p^f(i)[1-\tau(i)] di\} \quad (36)$$

And the optimal output is:

$$y_0^i = \frac{\theta^d}{b_i} [\frac{\theta^d \alpha}{P^i}]^{\frac{\alpha}{1-\alpha}} \quad (37)$$

Hence, the maximized profit of a downstream informal firm is:

$$\pi_i^d = y_0^i - \int_0^N b_i x^i(i) p^i(i) di \quad (38)$$

Upstream formal firms

Solving the maximization problem in equation (5) and (6), the optimal output for formal firm i to a formal buyer is given by:

$$x_f^f = [\frac{w(1+\tau_l)}{\theta^u \rho \beta p^f}]^{\frac{1}{\rho-1}} \left[\frac{y_0^f}{\theta^d}\right]^{\frac{1}{\alpha}} \quad (39)$$

Solving equation (7) and (8), the optimal output for a formal seller to an informal buyer is:

$$x_f^i = [\frac{w(1+\tau_l)}{(1-\tau)\theta^u \rho \beta P^i}]^{\frac{1}{\rho-1}} \left[\frac{y_0^i}{\theta^d}\right]^{\frac{1}{\alpha}} \quad (40)$$

Equation (39) and (40) show the optimal sale of a formal supplier to a single formal firm and a single informal firm. We assume upstream firms can observe the pre-entry productivity of downstream firms, based on which upstream firms form an initial expectation of the tax status of downstream buyers. The realized tax status in the equilibrium coincides with the initial guess. Specifically, upstream firms expect a downstream firm being formal if its productivity exceeds the upper bar of $\overline{\theta^d}$; and an downstream firm being informal if its productivity exceeds the lower bar of $\underline{\theta^d}$, but is below the higher bar of $\overline{\theta^d}$. In the equilibrium, $\underline{\theta^d}$ is the realized lower bar of pre-entry productivity of informal firms, and $\overline{\theta^d}$ is the realized lower bar of pre-entry productivity of formal firms. With the initial guess of the tax status of downstream firms, the profit of an upstream formal firm is:

$$\pi_f^u = \int_{\underline{\theta^d}}^{\overline{\theta^d}} x_f^i p_f^i - w(1+\tau_l)(\frac{x_f^i}{\theta^u})dF(\theta^d) + \int_{\overline{\theta^d}}^{\infty}(1-\tau)x_f^f p_f^f - w(1+\tau_l)(\frac{x_f^f}{\theta^u})dF(\theta^d) \quad (41)$$

Upstream informal firms

Solving equation (9) and (10), the optimal sales to a formal buyer is:

$$x_i^f = [\frac{w(1+b_i)}{\theta^u \rho \beta P^f}]^{\frac{1}{\rho-1}} \left[\frac{y_0^f}{\theta^d}\right]^{\frac{1}{\alpha}} \quad (42)$$

The optimal sale to an informal buyer is:

$$x_i^i = [\frac{w(1+b_i)}{\theta^u \rho P^i}]^{\frac{1}{\rho-1}} \left[\frac{y_0^i}{\theta^d}\right]^{\frac{1}{\alpha}} \quad (43)$$

Equation (11) and (12) show the optimal sale of an informal supplier to a single formal buyer and a single informal buyer. Upstream informal firms can also observe the pre-entry productivity of downstream firms, based on which they form an initial expectation of the tax status of downstream buyers. The realized tax status in the equilibrium coincides with the initial guess. With the initial guess of the tax status of downstream firms, the profit of an upstream informal firm is:

$$\pi_i^u = \int_{\underline{\theta^d}}^{\overline{\theta^d}} x_f^i p_f^i - w(\frac{x_i^i}{\theta^u})dF(\theta^d) + \int_{\overline{\theta^d}}^{\infty} x_i^f p_i^f - w(\frac{x_i^f}{\theta^u})dF(\theta^d) \quad (44)$$

Appendix E: A Granular Look at the Informal Sector

I plot the share of informal employment and the share of informal firms by states, sorted by their GDP per capita from high to low. As Ulyssea (2018) suggests, the informal employment and the informal firms do not necessarily move in the same direction. I show similar result using Indian data. The share of informal employment surpasses 95% among Mizoram, Lakshadweep, and Andaman, but barely reached 40% in Daman&Diu and Arunachal. The share of informal firm, on the other hand, is the highest among Manipur, Andaman, and Meghalaya, but the lowest among Daman&Diu, Dadra, and Goa. The distribution of informal workers and informal firms varies largely across states, but firm informality and employment informality do not necessarily move in the same direction. An example is Dehli, who ranks the 4th lowest share of informal firms, but only the 18th lowest share of informal employment.

Figure E1 also shows that the correlation between income level and the share of informality is not obvious. Rich provinces such as Lakshadweep also have a large share of informal firms or informal employment, implying multiple causes of informality in Indian states.

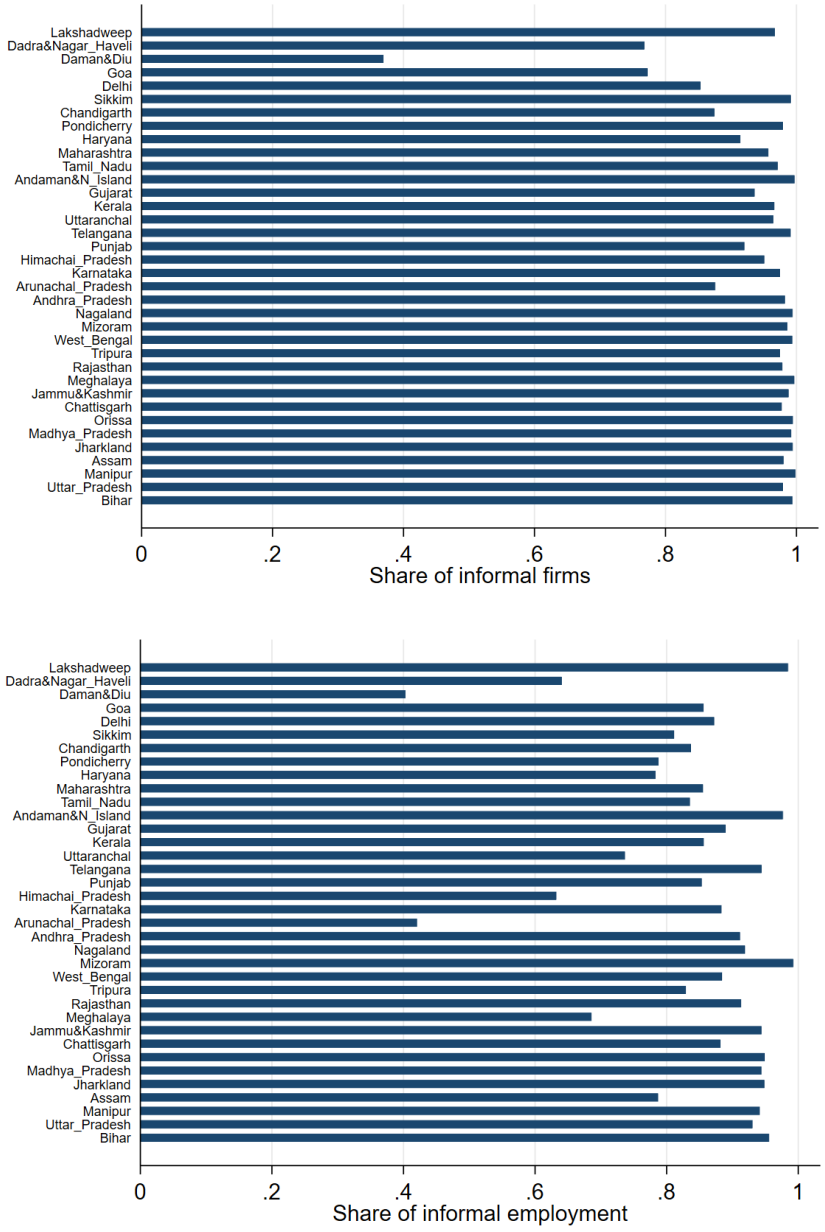


Figure E1: Share of Informal Firms by Province. Source: Annual survey of industries (2015), Unorganized Manufacturing Enterprises survey (2015), Labor force survey (2017-18).

I use the subsample of firms less than 10 employees in UME survey to plot the distribution of informal firms across its number of employees. The share of informal firms decrease as

their size grows. The share of informal firms is 99% in the one-person establishments, and the number drops to 96%, 88%, 84% for firms with employees of 2, 3, 4 respectively. For firms larger than 5 employees, informal firms only comprise less than 80% of the number of firms.

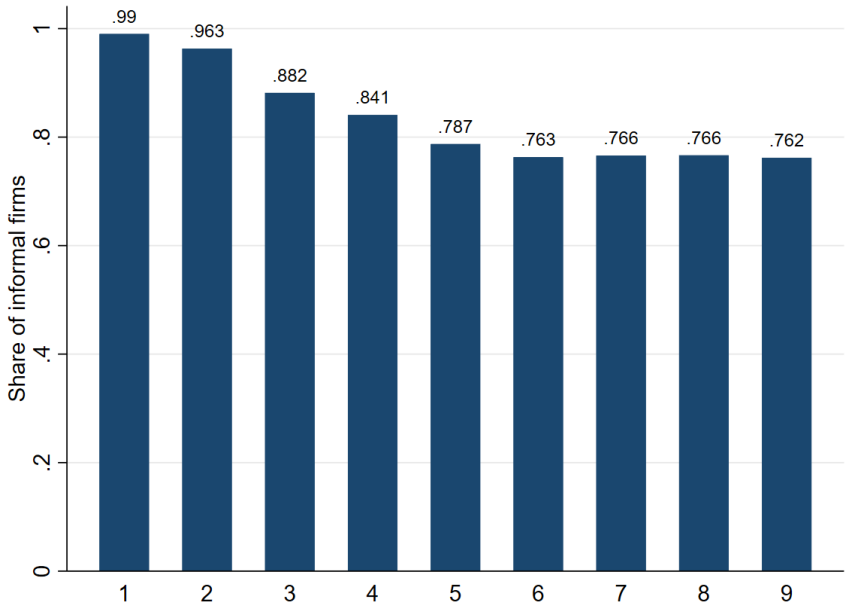


Figure E2: Share of Informal Firms by Number of Employees. Source: Annual survey of industries (2015), Unorganized Manufacturing Enterprises survey (2015).

In the combined dataset of ASI and UME, the average size of an informal firm is much smaller than a formal one despite the large number of informal firms. Firms less than three workers account for more than 85% of the total number of firms, but only generate 6% of total output.

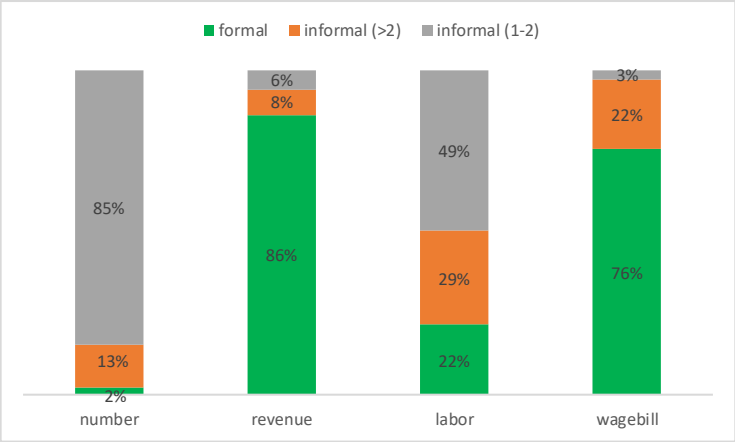


Figure E3: Composition of Manufacturing Sector. Source: Annual survey of industries (2015), Unorganized Manufacturing Enterprises survey (2015).

The average revenue of an informal firm larger than 3 employees is 14 million Rs, which is 11% of a formal one. For informal firms with less than 3 employees, the average revenue is only 1.3% of a formal firm. The average number of employees is 29.9, 4.8, and 1.3 for formal firms, informal firms larger than 3 employees, and informal firms smaller than 3 employees respectively.

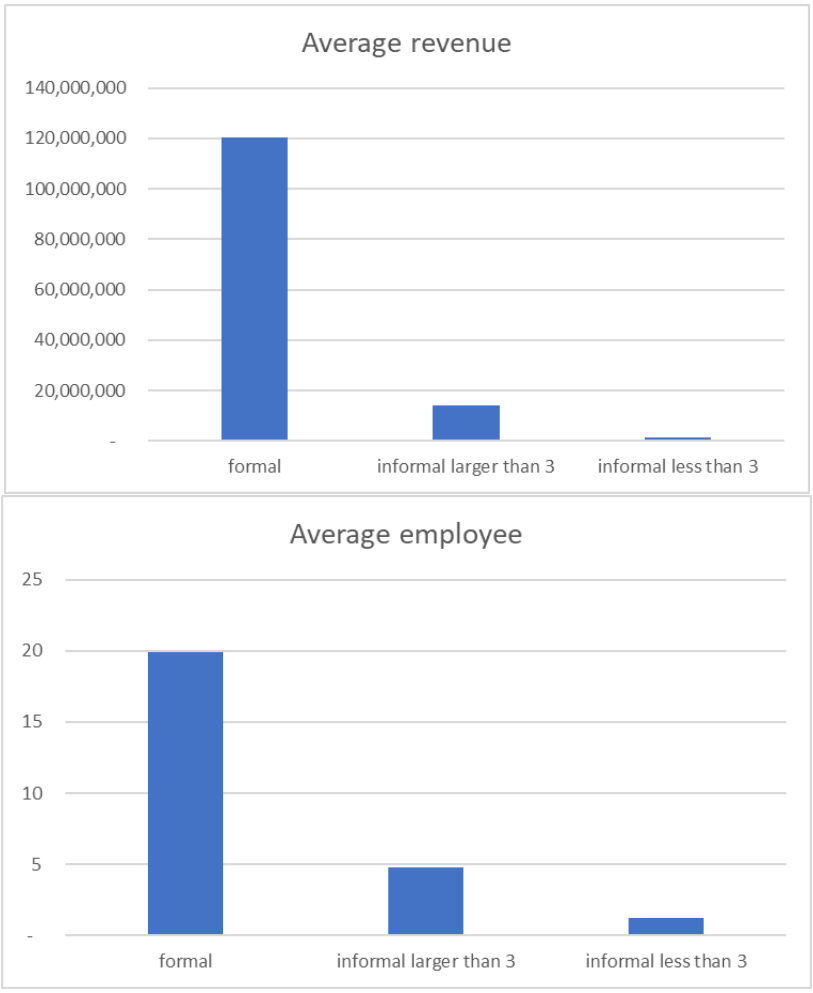


Figure E4: Average Firm Size by Number of Employees. Source: Annual survey of industries (2015), Unorganized Manufacturing Enterprises survey (2015).

Figure E5 plots the size distribution of informal firms whose annual revenue is between two and eight million Rupees. In the state of West Bengal, firms with annual revenue above five million Rupees are required to register with tax authorities. If informal firms want to keep their annual revenue below this level to avoid punishment, there should be a bunching near the threshold. However, we do not see a significant number of informal firms with annual revenue near the threshold.

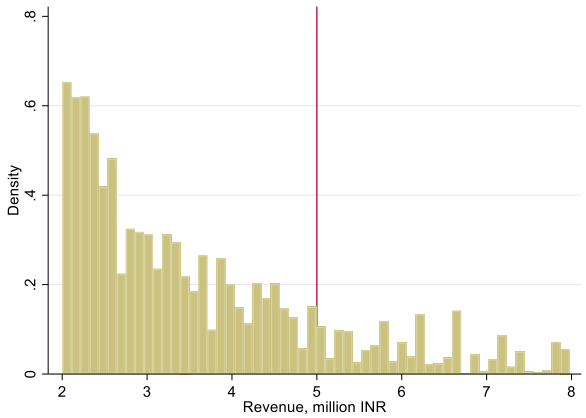


Figure E5: Size Distribution of Informal Firms. Source: Annual survey of industries (2015), Unorganized Manufacturing Enterprises survey (2015).

Table E1: Firm Characteristics by Size

Size	Value ln(VA per worker)	ln(K per worker)	ln(wage)
Informal firms			
1	10.62	11.52	9.27
	(0.99)	(0.99)	(1.25)
5	11.19	11.77	10.52
	(0.89)	(0.89)	(0.88)
10	11.89	13.04	11.22
	(1.31)	(1.31)	(0.54)
Formal firms			
20	12.74	13.76	11.64
	(0.89)	(0.89)	(0.66)
50	12.87	13.30	11.88
	(1.24)	(1.24)	(0.69)
100	13.02	13.32	12.07
	(1.18)	(1.18)	(0.76)
>100	13.14	13.93	12.03
	(1.11)	(1.11)	(0.83)

Source: Annual survey of industries (2015), Unorganized Manufacturing Enterprises survey (2015).

To qualtify the difference between formal firms and informal firms, I estimate the following reduced form model using the combined ASI and UME firm-level datasets.

$$y_{isj} = formal_{isj} + \alpha_s + \gamma_j + \varepsilon_{isj}$$

Where y_{isj} is the characteristics of firm i in sector s and state j, including log revenue, log number of employees, log value added, and log wage rate. $formal_{isj}$ is the dummy variable that equals 1 if firm i is registered in the tax authorities. I also control the ownership dummy, sector fixed effect and the state fixed effect.

Table E2 shows that on average, a formal firm is larger than an informal firm in terms of revenue, number of employees, wage rate, and value-added. The difference between the formal and the informal firms are more salient in the downstream than the upstream. Table E4 augments an interaction term between the formal dummy with the downstream index. It shows that the difference of revenue, labor, value-added, and wage rate between the formal and the informal sector are more salient in the downstream sectors. This implies stronger distortion in the downstream which raises the entry threshold of the formal sector. The results are robust when I use the unweighted sample that is representative of the manufacturing sector in India or when I restrict the sample to West Bengal province.

Table E2: Difference between Formal and Informal Firms

VARIABLES	(1) ln(revenue)	(2) ln(labor)	(3) ln(VA)	(4) ln(wage rate)
Formal	1.468***	0.535***	0.947***	0.490***
	(0.018)	(0.006)	(0.016)	(0.010)
Industry FE	Y	Y	Y	Y
State FE	Y	Y	Y	Y
Ownership FE	Y	Y	Y	Y
Observations	51,529	60,696	49,705	59,661
R-squared	0.486	0.447	0.460	0.307

Source: Annual survey of industries (2015), Unorganized Manufacturing Enterprises survey (2015).

Table E3: Difference between Formal and Informal Firms by Production Stages

VARIABLES	(1) ln(revenue)	(2) ln(labor)	(3) ln(VA)	(4) ln(wage rate)
Formal	0.914***	0.347***	0.445***	0.400***
	(0.043)	(0.015)	(0.038)	(0.023)
Formal*downstream index	0.631***	0.214***	0.572***	0.102***
	(0.044)	(0.016)	(0.039)	(0.024)
Downstream index	0.043	0.073	0.099	-0.074
	(0.138)	(0.051)	(0.123)	(0.089)
Industry FE	Y	Y	Y	Y
State FE	Y	Y	Y	Y
Ownership FE	Y	Y	Y	Y
Observations	51,529	60,696	49,705	59,661
R-squared	0.531	0.561	0.473	0.542

Source: Annual survey of industries (2015), Unorganized Manufacturing Enterprises survey (2015).

Hsieh and Klenow (2009) show that the size of registered Indian industrial firms grows slowly even after long time of operation. The relative employment of a formal firm is less than 20% of a US firm. I find that this is particularly true in the informal sector where the firm size grows even slower than the formal sector. Table E4 shows the correlation between the firm age and the firm size. An informal firm produces 9% less and employs similar employees compared to a new informal firm after 10 years of operation. The interaction term between the formal dummy and age is significantly positive, suggesting that formal firms grow faster but still slow at an absolute sense than informal ones. For a 10-year old formal firm, it employs only 4% more workers and does not increase its output than a new formal firm.

Table E4: Difference between Formal and Informal firms by Age

VARIABLES	(1) ln(revenue)	(2) ln(labor)	(3) ln(VA)	(4) ln(wage rate)
Formal	1.386***	0.491***	0.940***	0.437***
	(0.024)	(0.009)	(0.021)	(0.013)
Formal*age	0.008***	0.004***	0.002	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
Age	-0.009***	-0.001***	-0.007***	-0.003***
	(0.001)	(0.000)	(0.001)	(0.000)
Industry FE	Y	Y	Y	Y
State FE	Y	Y	Y	Y
Ownership FE	Y	Y	Y	Y
Observations	51,529	60,696	49,705	59,661
R-squared	0.529	0.562	0.472	0.544

Source: Annual survey of industries (2015), Unorganized Manufacturing Enterprises survey (2015).

The informal sector in India is persistent. For firms over 10 years old, more than 97% are informal, suggesting a persistent informality of firms who evade tax obligations. The large share of informality in old firms coincides with the observation in Hsieh and Klenow (2014) that that the average 40-year old plant in the US employs more than seven times as many workers as a corresponding 5 years or younger plant, while in India and Mexico, surviving firms are just twice as large over the same age range. However, this does not mean that the exit rate is low for informal firms. Literature has shown that informal firms often operate discontinuously, which means that these informal firms would quit the market when facing negative shocks, but would return to the market when the shock is over.

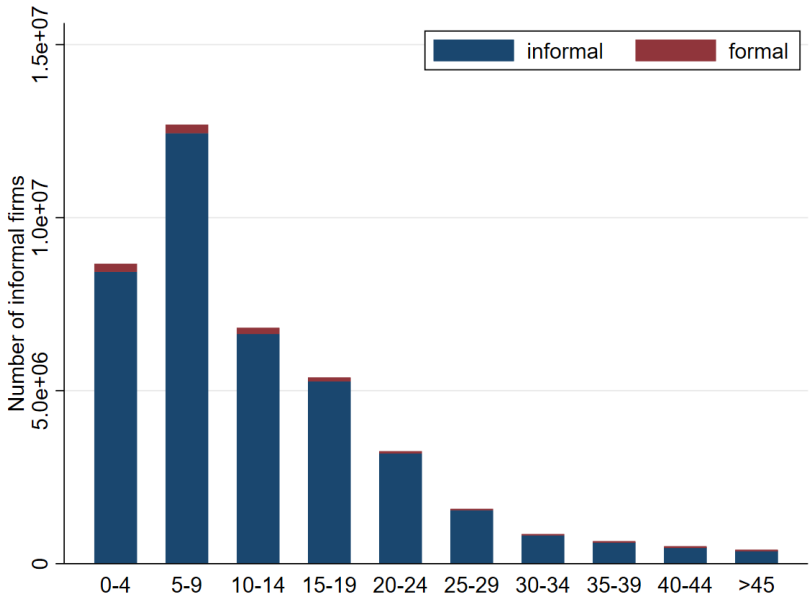


Figure E6: Distribution of Informal Firms by Age. Source: Annual survey of industries (2015), Unorganized Manufacturing Enterprises survey (2015).

The practice of informal firms also differs significantly from their formal counterparts. I use only the UME data and restrict the sample to firms with 3-10 employees in all sectors. The informal firms are less likely to keep a bank account than formal ones. 55% of informal firms maintain bank accounts compared to 93% for formal firms. As for the regular (book) account, only 10% of informal firms maintain a regular account while more than 70% of formal firms do so. Besides, informal firms are less likely to use computer and internet compared to formal ones. On the other hand, both types of firms have limited access to finance with 10% of informal firms and 8% of formal firms listing access to finance as stheir major challenge. At the same time, only 2% of both type of firms receives government assistance. Surprisingly, 11% of informal firms report that they took some work on contract base while only 4% of formal firms report so.

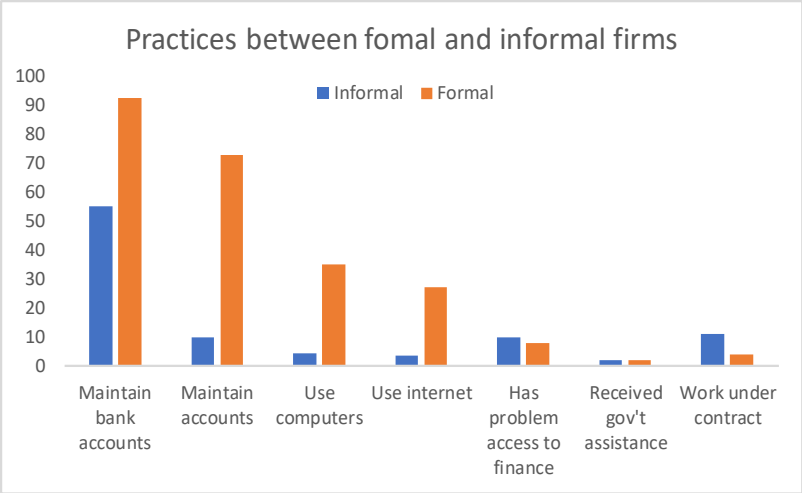


Figure E7: Comparison of Practice between Formal and Informal Firms. Source: Unorganized Manufacturing Enterprises survey (2015).

The informal practice of operation also reflects the low productivity of small informal firms. Figure E8 compares the distribution of value added per worker (in logs) in India's formal and informal sectors. Formal firms have on average higher value added per worker than informal firms, which is captured by the rightward crest of the distribution of formal firms. Value added per worker and capital per worker grow as a firm hires more employees.

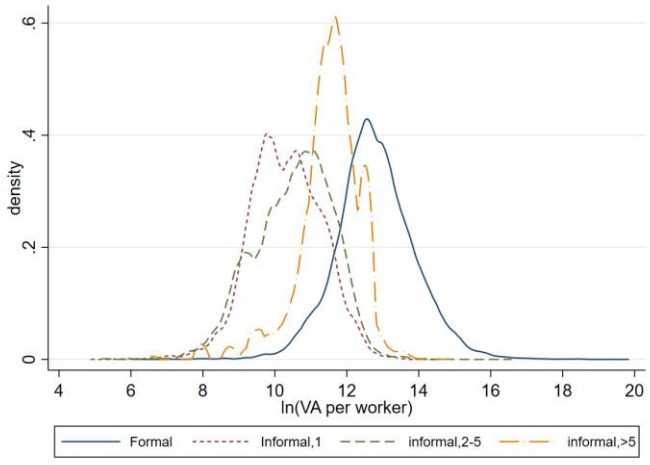


Figure E8: Distribution of Value-added Per Worker. Source: Annual survey of industries (2015), Unorganized Manufacturing Enterprises survey (2015).

Appendix F1: More Discussions on the Definition of Informality

Third, this paper is related to the literature on defining the informal sector. A prior question in the study of informality is its definition. Current literature includes three types of informality concept, including the size criteria, the compliance criteria, and behavior criteria. The size criteria is adopted by ILO which defines an informal firm as an unregistered firm where the owner is an individual or a household whose capital is not separable from that of the firm and for which there is not reliable accounting that could permit retracing the operations of the firm (ILO, 2002).

Dabla-Norris et al. (2008) and Benjamin and Mbaye (2012) find a strong correlation between firm size and its informality but they also argue that size might not be the only criterion of informality. Many micro firms are actually formal enterprises with tax registration and operational practices that are usually adopted only by formal firms (Gelb et al., 2009). Besides, ILO only sets up the upper bound of size of an informal firm at 10 and leaves the countries the job to clarify the actual size in their national surveys. This means that the size criterion will ignore the fact of tax compliance of many small firms.

The second criterion identifies the informality if their business activities are not registered with the appropriate authority. Gelb et al. (2009), Steel and Snodgrass (2008) use registration with tax authorities as the criterion in defining informal firms. They argue that informal sector is an entity unknown to the fiscal authorities, and hence does not comply with tax regulations.

But this criterion is also not enough to qualify a firm as formal as there is no clear cut as the definition of registration with appropriate authorities varies by country (La Porta and Shleifer, 2008). Benjamin and Mbaye (2012) demonstrate that even very small, informal firms without a known location are sometimes identified and listed by the tax department

in Senegal. Even if informal firms are not always recorded at the level of central authorities, they are recorded at the local level where they also pay taxes.

The third criterion is treating informality as a continuum characteristic. There is strong heterogeneity even within the informal sector. Benjamin and Mbaye (2012) state that whatever the criteria are used, it is difficult to define the informal sector in a dichotomized manner. The criteria of size, registration, payment of taxes. do not sufficiently discriminate between formal and informal firms. So, the informal sector appears to be a continuum of situations defined by a set of factors that determine the place of each firm the formality scale. This definition often implies the most sophisticated are the best target for programs aimed at increasing registration or formal taxation (Medvedev and Oliedo, 2013)

However, this method is prone to subjective perception of "formal practices". Identifying the practices that only apply to the formal sector is subjective, and giving weights to each of the practices is even more tricky. There is no guidance as to the selection of appropriate practices to define informality. While one might think maintaining a bank account is an important practice for the formal sector, others might regard contract-based work as the primary characteristic of the formal sector.

Appendix F2: Other Data Sources on the Informal Sector

There are several data sets that contains information on the informal sector. However, given the complexity of informality, these data sets have different coverage and serve different research topics.

Investment Climate Assessment (ICA) is conducted by the World Bank Group that assesses the investment environment for each member country. It includes a category labeled informal, which aims to capture the degree of informality in doing business. However, these are largely based on official lists firms. While these surveys may include many

informal firms, or even the majority of informal firms, they mainly rely on a sample frame of formally listed firms and do not include household enterprises.

World Bank Enterprise Surveys are recently including more data from firms sampled in well-recognized informal market places.[14] But there remain issues of representativeness of the overall informal sector. Informal marketplaces can cluster businesses by type for a variety of reasons while this approach also leaves out enterprises conducted in dwellings or with no fixed location.

Annual Survey of Industries (ASI) covers manufacturing firms with more than 10 workers that use electrical power, or 20 workers that do not use electrical power are required to register under the Factories Act; these firms are considered organized, or formal, and are covered by ASI.

Unorganized Manufacturing Enterprises (UME) Survey is conducted by National Sample Survey Organization (NSSO). All firms below employment thresholds of ASI are not required to register under the Factories Act, and unregistered firms are considered unorganized, or informal, and are covered by NSSO.

Schneider Index is a country-level index calculated by Medina and Schneider (2018), which is a ratio of the informal sector value-added to the entire economy (GDP).

[14] https://www.enterprisesurveys.org/en/data

Appendix G: SMM Estimator

Following Ulyssea (2018), I consider S=50 simulated data sets containing a mass of M=30,000 potential entrants each. For each potential entrant, I draw an observed productivity and a misperception of productivity. The stochastic components of the model are drawn only once in the beginning of the procedure and are kept fixed during the estimation.

To save time on computational time, I use 200 equally spaced grid as a representative sample of the whole population. Firms with productivity near one grid are categorized into the same grid. Then I calculate the economic behavior of these 200 representative firms. The behavior of firms in the whole sample are then recovered by a spline proxy to map their productivity into economic outcomes.

One difficulty when estimating discrete choice models using simulation-based method is that simulated choices will be a step function of the parameter vector given the random raw. Since these discontinuities are inherited by the objective function, this also precludes the use of derivative-based methods, which are faster and more accurate than derivative-free methods or random search algorithms (Bruins et al., 2015). To overcome these challenges, I use the following smoothing function to correct for the choppiness of the policy function.

$$h(\bar{V}(\varphi), m, \lambda) = \frac{\overline{V_m}(\varphi)/\lambda}{1 + \sum_k \overline{V_k}(\varphi)/\lambda}$$

Where $\bar{V}(\varphi)$ is the set of net payoffs attached to the choices of firms. As the smoothing parameter λ goes to zero, $h(\cdot)$ goes to one if alternative m provides the highest payoff, and zero otherwise. When choosing λ, I follow the method by Ulyssea (2018) to balance biase and smoothness. I set $\lambda = .05$ and $S = 50$.

The estimator is given by:

$$\hat{\varphi} = argmin_{\varphi} Q(\varphi) = \{g_{NS}(\varphi)' \widehat{W}_{g_{NS}} g_{NS}(\varphi)\}$$

Where $g_{NS}(\varphi) = \hat{m}_N - \tilde{m}_s(\varphi)$, and I omit the conditioning arguments for notational convenience.

The conditions for consistency are close to the ones for extremum estimator. For a extensive discussion of the weight matrix, see Ulyssea (2018), Newey and McFadden (1994), Gourinchas and Parker (2002).

Table G1: Moments

Moments		**Data**	**SD**
Share of informality, #	downstream	98%	0.38%
	upstream	92%	0.64%
Share of informality, outputs	downstream	31%	3.08%
	upstream	16%	2.90%
Share of input from formal suppliers	formal firms	80%	0.14%
	informal firms	29%	1.71%
Share of output to formal buyers	formal firms	48%	0.16%
	informal firms	16%	0.81%
Distribution of downstream informal firm: relative size towards avg formal firm	5th percentile	0.02%	0.01%
	25th percentile	0.14%	0.06%
	50th percentile	0.46%	0.14%
	75th percentile	1.20%	0.32%
	95th percentile	3.35%	0.86%
Distribution of upstream informal firm: relative size towards avg formal firm	5th percentile	0.16%	0.06%
	25th percentile	0.38%	0.11%
	50th percentile	0.64%	0.21%
	75th percentile	1.50%	0.46%
	95th percentile	5.92%	1.96%
Distribution of downstream formal firm: relative size over avg formal firm	5th percentile	7.91%	4.88%
	10th percentile	9.70%	5.85%
Distribution of upstream formal firm: relative size over avg formal firm	5th percentile	1.42%	0.51%
	10th percentile	2.07%	0.68%

Standard deviation divided by mean	downstream	24.5	2.0
	upstream	24.4	2.3
Inputs/revenue	downstream	71%	2.19%
Share of informal labor	upstream	77%	2.05%

In the model, I assume that firms form their expectation on the cutoff productivity between the formal and informal sectors, beyond which a firm will choose to formalize. In the equilibrium, the realized cutoff productivity confirms firms' expectation. Similarly, firms have an expectation of the cutoff productivity of the informal sector below which a firm will quit the market. Table G2 shows the expected cutoff productivity and the realized productivity in the equilibrium, which confirms that firms' expectation of the cutoff productivity is correct.

Table G2: Expected and Realized Cutoff Productivity

Parameter	Description	Expected productivity	Realized productivity
delta low	Least productive upstream informal firm	2	2.04
delta high	Least productive upstream formal firm	4.4	4.42
theta low	Least productive downstream informal firm	1	1.01
theta high	Least productive downstream formal firm	2.8	2.83

Appendix H1: Traditional Policies on Informality

Table H1: Comparison between Traditional Policies towards Informality

		Carrot	**Stick**	**TFP**
Number	Downstrem	0.954	0.937	0.970
	Upstream	0.892	0.894	0.902
	All	0.922	0.915	0.935
Output	Downstrem	0.243	0.106	0.268
	Upstream	0.198	0.124	0.226
	All	0.207	0.121	0.232
Avg size	Downstrem	0.015	0.008	0.011
	Upstream	0.030	0.017	0.032
	All	0.027	0.015	0.029
Trade	F from F	0.761	0.828	0.719
	I from F	0.290	0.416	0.265
	F to F	0.534	0.574	0.488
	I to F	0.269	0.333	0.253

Source: Annual survey of industries (2015), Unorganized Manufacturing Enterprises survey (2015).

The impacts of "carrot" policy

Lowering the entry cost into the formal sector will encourage the ready-to-formalize firms to register which reduce the extensive margin of informality. In this part, I modify the entry cost parameters in the upstream and the downstream and produce the counterfactual behavior of firms and its implications on the macro variables. Specifically, I set the baseline scenario where the entry costs are at the original value. Then I set the entry cost parameters at 50% of their original value and run the model to get a new set of simulation results. Then I increase the entry cost by 0.5% to 50.5% of their original value and get the second set of simulation results. I repeat the above procedure so that the entry cost ranged from 50% of their original value to its original value. Then I plot the variables of interest against the entry cost to visualize their relationship.

Table H1 summarizes the impacts of "carrot policy" on informality. Lowering entry costs by 50% leads to a moderate reduction of informality in the downstream where the share of informal firms decreases from 97% to 95% and the share of informal output drops from 28% to 24%. The informality rate in the upstream hardly change when entry costs are cut half. Most of the "reservoir" type of firms (99%) and a small fraction of "parasite" firms (33%) formalized in the downstream after entry costs lower. The change is less significant in the upstream where only 13% of "reservoir" firms and 2% "parasite" firms formalize.[15]

When the entry cost is high, the profit of being formal has to be large enough compared to being formal so that a firm with a given productivity has the incentive to formalize. Otherwise, the gain from being formal is not sufficient to cover the entry cost and tax liabilities. Intuitively, lower entry costs induce productive firms who opt to remain informal to formalize so that they can expand production without being worried about the punishment on the informal sector. Since the distribution of downstream productivity is more concentrated than that of the upstream as shown in

Table 2: Summary Statistics of State-Level Sample

	Obs	**Mean**	**SD**	**Before**	**After**	**Diff**
State-year sample						
Share of forma firms	138	0.065	0.117	0.074	0.058	-0.016
Share of formal output	138	0.710	0.309	0.559	0.838	0.279***
Share of formal labor	138	0.286	0.235	0.259	0.309	0.050*
State-industry-year sample						
Share of forma firms	1614	0.261	0.376	0.232	0.281	0.049***
Share of formal output	1589	0.677	0.336	0.570	0.754	0.184***
Share of formal labor	1598	0.460	0.380	0.415	0.493	0.078***

Source: Pooled ASI and UME (2000-2015).

[15] The smaller share of informal firms can either reflect a growing number of formal firms or a shrinking number of informal firms. Hence, I calculate the status of firms before and after the change of entry cost. I compare the baseline scenario and the "carrot" scenario with 50% of entry costs, the results show that 2.5% of the downstream informal firms switched to be formal, which leads to an increase of the number of formal firms in the downstream by 423%. To a lesser extent, the number of formal firms in the upstream increased by 13% as a result of 0.1% of informal firms switch to the formal sector.

Note: *** p<0.01, ** p<0.05, * p<0.1. ASI is Annual Survey of Industries, UME is Unorganized Manufacturing Enterprises Survey. Before refers to before VAT adoption, after refers to after VAT adoption. There are 35 states in India, Lakshwadadeep and Mizoram do not have registered manufacturing enterprises in 2015.

Table 3, there are more downstream informal firms near the margin of formalization. Lowering the entry costs by the same magnitude would encourage more downstream informal firms to enter the formal sector than in the upstream.

Table H2: Share of Firms that Formalize after Policy Changes

	Downstream			**Upstream**		
-Carrot	98.9%	33.0%	0.0%	12.5%	1.7%	0.0%
-Stick	98.9%	70.6%	0.0%	12.5%	1.7%	0.0%
-TFP	98.9%	25.9%	0.0%	0.0%	0.0%	0.0%

Source: Annual survey of industries (2015), Unorganized Manufacturing Enterprises survey (2015).

However, "carrot policy" incurs higher intensive margin of informality. The newly formalized firms are less productive than the incumbent ones. These newly formalized firms maintain their trade network with their informal counterparts, which drives down the total share of inputs in the formal sector. This is also reflected in the higher share of outputs to the formal sector among informal sellers.

"Carrot policy" raises tax revenue by enlarging the size of the formal sector. Since the "carrot policy" reduces the deadweight loss of being formal, informal firms find it more profitable to formalize, enjoying the tax deduction under VAT regime and avoiding fine on the informal sector. If the entry costs are reduced to 50% of its current value, total tax revenue could be 8% higher than the current level. Hence, by removing the entry costs of the formal sector, policy makers would enjoy simultaneously a lower level of informality and a higher level of tax revenue.

Besides the benefits of taxation, lowering the entry costs can increase corporate profits at the cost of household income. Compared to the baseline scenario, lowering the entry costs by 50% would lead to 8% increase of corporate tax but a decrease in household income by 13%. The loss of household income results from a lower equilibrium wage as informal

firms convert to formal ones. Because formal firms face higher labor costs, firms tend to hire less workers after formalization. Hence, part of workers who used to work informally now lose their jobs and income.

It is worth noting that reducing entry costs includes but not limited to streamlining the registration process and reduce the registration fee. Many countries have reduced the registration fee, simplified its taxation system, and implemented the "one-stop shop" to streamline the registration process for firms. However, De Mel (2013) and Bruhn and McKenzie (2014) show that waiving the registration fee is not enough to reduce the entry costs so that firms are encouraged to formalize. The ongoing costs of being registered, including but not limited to tax payment and labor market regulation, deter firms from registering even if the process of doing so is simplified. Andrade et al. (2013) and De Giorgi and Rahman (2013) reach the same conclusion using the experiment in Brazil where the government distributed a glossy brochure to informal firms and waived the registration fee. They found firms that received this brochure were not more likely to register over the following year. Gordon and Li (2009) show the significant benefit from tax evasion by not reporting cash transactions. However, when the subsidy increased to the highest level (approximately two months' profits of an informal firm's profit), almost half of them registered.

Hence it is critical to identify the entry costs into the formal sector and create a level-playing field for both formal and informal firms. For example, in India, size-dependent labor regulation that prevents large firms from laying off workers may retard the growth of high-potential firms by reducing their incentives to invest in expanding their size (Hseih and Klenow 2009). The resulting low-wage equilibrium would have too many small and informal firms (given fixed cost of registering). This section shows that identifying and removing the underlying sources of entry costs could significantly reduce informality while also increasing tax revenue and social welfare.

The impacts of "stick" policy

The "parasite" view of informality implies a stronger enforcement to punish informality. In this view, firms opt to stay informal so that they can avoid tax liabilities. In this part, I modify the punishment parameters on the informal firms and produce the counterfactual behavior of firms and of aggregated macro variables. Specifically, I set the baseline scenario where the entry costs are at the original value. Then I set the punishment parameters in the upstream and the downstream at 200% of their original value and run the model to get a new set of simulation results. Then I decrease the punishment by 0.5% to 199.5% of their original value and get the second set of simulation results. I repeat the above procedure so that the punishment ranges from 80% of their original value to 300% of its original value.

Table H2 shows that the share of informal firms decreases from 97% to 94% when enforcement on informality doubles, which corresponds to a reduction of informal output from 28% to 10%. The switch is much less significant in the upstream. In fact, 5% of informal firms quit the market and the rest reduce their production in response to the stricter enforcement. This is consistent with Andrade et al. (2013) and Bruhn and McKenzie (2014) that stronger punishment is highly efficient in reducing the extensive margin of informality.

As for the intensive margin of informality, the "stick policy" leads to a slight increase in the share of formal inputs by formal firms. When the cost of input becomes higher for informal firms, informal firms choose either to formalize or to transmit the punishment into higher product prices. Informal products become less attractive in price so downstream (formal and informal) firms will purchase less from informal sellers. In fact, informal firms will significantly increase their purchase from formal sellers.

"Stick" policy increases the cost of being informal, which increases tax revenue by a significant amount but at a severe cost of household income. When the cost of informal

firms doubles, the tax base enlarges as more firms formalize, which increases the tax revenue by more than 87% (Table). Less informal firms reduce competition in the market and increases the profit of incumbent firms by 70%. While stricter punishment does alter the balance of formality vis-à-vis the informality, entering the formal sector would incur heavy entry costs which prohibits the informal firms from doing so. Informal firms would rather exit the market rather than entering the formal sector. Household income drops by 74%, which reflects the loss in the informal sector employment as strict punishment weighs on the small informal business owners. As many informal firms quit or reduce their production, the demand for labor shrinks which results in a lower equilibrium wage. Given the fixed labor supply in the model, household income decreases as a result of stricter punishment on the informal sector.

The impact of development policy

The "weak" view suggests improving firm productivity by employing better technology, enhancing management skills, and improving doing business environment. Some informal firms find it more beneficiary to enter the formal sector while less productive firms-in-waiting will enter the informal sector. When the TFP increases by 50%, the share of number of informal firms and the share of informal outputs hardly change (Table). But both the formal and the informal sector becomes larger after productivity improves. In fact, the number of formal firms increases 65% in the downstream and 92% in the upstream. The overall number of informal firms will also increase 68% and 94% in the downstream and the upstream respectively.

The intensive margin of informality increases as productivity improves. Formal firms now purchase 72% from other formal sellers compared to 81% in the baseline results. Formal firms sell slightly more to other formal firms. This reflects a larger improvement of productivity in the informal sector than the formal sector.

The most significant improvement comes from the increase of household income. Since both formal and informal sectors enlarge production, their labor demand increases and pushes up the equilibrium wage, which in turn raises household income by 20 times if productivity raises 50%. The improvement in household income is accompanied with 23% higher tax revenue and 26% increase in corporate profit.

In sum, both the "reservoir" view and the "parasite" view of informality is plausible in the model set up. The former view indicates that potentially productive informal firms are kept out of the formal sector due to high entry costs. Hence, policy makers should implement "carrot" policy by lowering the entry costs and induce the informal firms to formalize. The latter view indicates that informal firms are productive enough to enter the formal sector but intentionally stay informal to avoid regulatory liabilities. Hence, policy makers should strengthen enforcement or put "stick" on the informal sector so that the informal firms are forced to formalize.

While I show that both "carrot" and "stick" policies are effective in reducing informality and raising tax revenue, they have converse implications on the social welfare. The "carrot" policy can improve social welfare through higher level of tax revenue and higher level of firm profit. Informal firms in the "reservoir" now find it more profitable to register while less productive firms can still maintain its operation. On the contrary, the "stick" policy leads to welfare loss because it shrivels the profit of all informal firms despite the gain of tax revenue. When the loss of firm profits and household incomes is larger than the gain of tax revenue, this would lead to a loss of social welfare despite higher level of formality.

Appendix H2: Transmission of Formalization

Besides the revenue gains, literature suggests an efficiency gain from VAT adoption. De Paula and Scheinkman (2010) show the chain effect in which formalization can transmit through the value chain. This will improve the efficiency of tax administration because tax

enforcement can only target one or several production stages that is key to transmit formalization to others. This section tests this hypothesis by showing an example of implementing strong enforcement on informal firms at one stage of the production stage. Specifically, I double the cost of inputs of either downstream informal firms or upstream informal firms and study its impact on informality. The changes of informality is shown in Table L1.

When strong enforcement is implemented to the downstream informal firms, it creates a reduction of downstream informal firms by 5%, which correspond to a stark drop of informal outputs by 53% in the downstream. The increase of formal firms is significant at 20% in the downstream, but only accounts for a 3% increase in the downstream formal outputs. But the formalization in the downstream does not transmit to the upstream. The number of both formal and informal firms shrink by 5% and 3% in the upstream respectively.

Newly formalized downstream firms have lower productivity than the incumbent formal firms. They are not productive enough to purchase heavily from formal sellers. Instead, they tend to maintain their informal network and purchase cheaply from informal sellers. For the informal sellers, they do not need to formalize to maintain their trade network even if their partners are now formal.

However, when strong enforcement is implemented to the upstream informal firms, this generate a significant chain effect and reduce informality in both production stages. The number of formal firms grows 27% and 22% in the downstream and the upstream respectively, corresponding to 14% and 86% increase in formal outputs. However, more formal buyers in the downstream also purchase more from upstream informal firms, which generates an overall positive increase in the upstream informal firms and their outputs.

When the upstream firms are forced to formalize, they can issue invoice and tax credit to their buyers. Their informal buyers, in turn, find it more profitable to formalize and take advantage of tax credit from formal sellers. This implies that tax credit from formal inputs provides a strong incentive for downstream buyers to formalize if their suppliers formalize.

In comparison to VAT, I replicate the above simulations targeting the downstream or the upstream informal firms under the cascading sales tax regime. But neither generates a chain effect in terms of formal outputs. When implementing the strong enforcement to the upstream informal firms, upstream formal firms expand outputs by 55% but downstream formal firms produce 19% less outputs. Under sales tax, the tax burden of formal inputs is bear by the downstream formal firms. In turn, downstream formal firms purchase more from informal suppliers at cheaper prices. When strict enforcement is implemented to the upstream informal firms, they will raise intermediate input prices and partially transmit the burden to the downstream sector. Hence, both formal and informal firms in the downstream suffer from higher input costs on their informal inputs.

Table H3: Transmission of Informality

			Number	
Change from baseline	**Downstream formal**	**Downstream informal**	**Upstream formal**	**Upstream informal**
Imposing stricter enforcement on informal firms				
VAT, 5%: on downstream	20%	-5%	-5%	-5%
VAT, 5%: on upstream	27%	10%	22%	10%
ST, 5%: on upstream	36%	7%	23%	6%

			Revenue	
Change from baseline	**Downstream formal**	**Downstream informal**	**Upstream formal**	**Upstream informal**
Imposing stricter enforcement on informal firms				
VAT, 5%: on downstream	3%	-53%	-5%	-3%
VAT, 5%: on upstream	14%	-10%	86%	11%
ST, 5%: on upstream	-19%	-39%	55%	6%

Source: Annual survey of industries (2015), Unorganized Manufacturing Enterprises survey (2015).

Note: ST is cascading sales tax, VAT is value-added tax. Informality neutral rate is 9.5% so that the number of formal firms is the same to the benchmark scenario of cascading sales tax at the rate of 7.5%. Revenue-neutral rate is 10.1% so that the total value-added tax revenue is the same as the total cascading sales tax revenue.

The change of number or revenue is the percent change from the benchmark scenario of cascading sales tax at the tax rate of 7.5%.