Corruption and Firm Dynamics in India

Samarth Gupta [*]

September, 2016

Abstract

I present a theoretical model of how corruption affects dynamics of firm size evolution. I model corruption as a stochastic dynamic process which affects firms in each period. Extractive corruption (in the form of bribes or extortion by bureaucrats) reduces the optimal size of the firm. Costs of extractive corruption create incentives for firms to conduct club corruption in the form of graft and regime-influence. Club corruption appears as an entry cost for entrants leading to lower entry of new firms in the industry. The prediction of the model are tested on a data of manufacturing firms in India, where I show that states with higher corruption have higher skewness of industries and lower entry of new firms. The paper contributes to the causes of mis-allocation in Indian manufacturing industry.

1 Introduction

Corruption is endemic in developing countries and is now considered a stylized fact. It affects every facet of economic activities including firm outcomes and is present in each level of hierarchy of administration (Sukhtankar & Vaishnav (2015), Olken & Pande (2011)). Corruption at lower level of governance acts in an extractive manner where inspectors siphon-off some fraction of output (Niehaus & Sukthankar (2013); Djankov et al (2002);Olken & Barron (2009)). At the highest level, it appears in the form of an exclusive club where individuals with deep pockets can enjoy undue and unfair favors from the ruling regime (Fisman (2001); Sukhtankar (2012)).

At the same time, some empirical puzzles regarding firm dynamics in developing countries stand out. Studies, both recent and classic, have investigated the existence of huge number of

---

[*] Economics Department, Boston University

small firms in India (Tybout (2000) and Amirapu & Gechter (2015)). A wide literature exists on mis-allocation of TFP of firms (Hsieh & Klenow (2009); Aghion et al (2008); Dougherty et al (2011); Besley & Burgess (2002)) and redistribution of market shares from mid-sized firms to larger and smaller firms- defined as a missing middle (Tybout (2014) and Alfonso & Chari (2014)).

In this paper, I try to link the empirical puzzles regarding distribution of firms by size and facts regarding corruption using a model of firm dynamics under corruption. I model the two kinds of corruption described above as different forms of dynamic process. Extractive corruption affects firms every period through a stochastic process. It pushes the firm size lower every period creating left-skewed distribution of firm size. Graft, which I define as club corruption, acts as a one-time fixed cost to avail protection from extractive corruption. High entry cost of graft reduces turnover (Hopenhayhn 1992). I test the predictions of the model on ASI Data of firms from 1999-2010. Using corruption metrics of Transparency International from 2005, I show that skewness of firms is more positive in more corrupt states and there is lower entry of large firms (with more than 100 workers) in more corrupt states.

The paper is divided in the following manner. I present the model in section-2. Section-3 and 4 provide the data and empirical proof of the propositions of the model. I discuss some caveats in Section-5 and finally I conclude in Section-6.

2 Model

Before setting up the model, I will describe nature and type of corruption that I intend to model here.

I divide corruption activities into two categories:-

1 **Extractive Corruption** is the extortionary behavior of public officials whenever providing service to the rightful beneficiaries. This kind of corruption takes the form of a private costs over and above the specified costs for a public service. For example, when a new establishment requires an operating license from a government department, the public officer can withhold license until a bribe has been paid.

2 **Club Corruption** is the cases of grafts which lately have been creating headlines all over the world. These exist in various forms such as kickbacks for contracts, regime influencing and lobbying efforts. Examples of club corruption include Nira Radia Tape

Scandal in India, Dilma Rousseff's Car wash Scandal etc.

Main difference between Extractive and Club corruption is that while Extractive corruption is conducted by the public officials demanding bribes to provide a service, club corruption is offered by private individuals as a gift to tilt the scale in their favor.
Now I will set up the model of firm operation.

2.1 Primitives of the Model

Consider a firm with an infinite horizon with the following objective function in time period t:-

$$\pi_t(q;\tau) = p.(1-\tau_t)q - c(q)$$

where $\pi(.,.)$ is profit function
$c()$ is cost function where $c' > 0$ and $c" > 0$
and $\tau \in (0,1)$ is the degree of extractive corruption that a firm faces in period t; This way of modeling corruption is similar to how firms are treated in the mis-allocation literature (Hsieh & Kelnow (2009)). Given τ_t, a firm optimizes to produce :

$$q(\tau_t) = c'^{-1}(p.(1-\tau_t))$$

Thus, quantity is declining in level of extractive corruption.
Degree of extractive corruption is an equilibrium object and is determined as per the quality of institution.

2.2 Extractive Corruption

Let following function be the probability of punishing a corrupt bureaucrat:- $p : \mathcal{T} x \mathcal{K} \rightarrow [0,1]$ where $\tau \in [0,1]$, $\mathcal{K}$ is set of exogenous institution-specific parameter and $p(\tau;k) \in [0,1]$ is the probability of bureaucrat getting caught and punished for demanding extraction.
Let $\mathcal{K} = \{k_i | k_i \in [0, k^m]\}$ where k^m is the maximum value of institution quality.

Consider the following assumptions:-

Assumption-1: $p(0;k) = 0 \forall k \in \mathcal{K}$

Assumption-2: $\frac{\partial p}{\partial \tau} > 0$ and $\frac{\partial p}{\partial k} > 0$

Assumption-1 states that probability of a bureaucrat getting punished is 0 when there no bribe is demanded. Assumption-2 states that this probability increases as rent demanded by bureaucrat increases and better the quality of institution (higher k), higher is the probability of a corrupt bureaucrat being punished.

Consider the following regularity assumption to bound the probability of a bureaucrat getting punished:-

Assumption-3: If $\exists k$ such that for some $\bar{\tau} \in [0,1), p(\bar{\tau}, k) = 1$, then $\forall \tau > \bar{\tau}, p(\tau, k) = 1$

Now, consider the following optimization problem for a population of bureaucrat indexed by i. Bureaucrats are unaware of the exact value of k. However, bureaucrat i has a belief over institution quality. Let

$$U(\tau; k^e) = (1 - p(\tau, k))\tau R(\tau) - p(\tau, k)\Phi(\tau)$$

be the payoff function for the bureaucrat. R is the revenue of the firm from whom the bureaucrat is extracting rent, $\Phi(\tau)$ is the cost of punishment with $\Phi' > 0$ and k^e is the expected value of institution quality for a bureaucrat.

Assumption-4: $R(\tau) + \tau R' + \Phi' > 0$

Assumption-4 can be stated as gradient of cost ($\Phi(\tau)$) rises faster than gradient of benefit ($\tau R(\tau)$ with respect to τ.

Let τ^* be the optimal level of rent extraction given bureaucrat's belief over k. Assumptions 1-4 and Theorem 2.8.1 from Topkis (1998) imply that τ_t^* and k_t^e are inversely related.

Thus, optimal level of extraction for the bureaucrat i is inversely proportional to her belief over quality of the institution, k_t^*.

In the next period, given the successful extraction, bureaucrat would update her belief regarding k using Bayes' Rule. The process of updating belief regarding institution slope is demonstrated in the appendix since it is closely linked to the proof of the following Theorem.

Theorem-1: $\tau' > \tau'' \implies F(\tau|\tau') < F(\tau|\tau'')$ where F is the probability distribution function of extraction τ in period t+1 conditional on extraction in τ

Proof of the Theorem can be found in the appendix.
Theorem-1 states that conditional probability distribution of higher extraction First Order

Stochastically Dominates conditional probability distribution of lower extraction. Intuitively, on having successfully extracted a high rent in the first period, the bureaucrat updates her belief of quality of institution, k, to be lower. This reinforces her behavior to extract even higher.

2.3 Dynamics under Extractive Corruption

Let $G(.|q)$ be the conditional CDF of the quantity produced by the firm given firm's production q in period-t

Lemma-1: $q' > q'' \implies G(Q|q') < G(Q|q'')$
Proof: Firm's optimal output is given by

$$q^*(\phi) = c'^{-1}(\phi p)$$

Let $\phi' > \phi''$, $q' = q^*(\phi')$ and $q'' = q^*(\phi'')$. Since c' > 0, $c'^{-1} > 0$. This implies $q' > q$"

Now, consider distribution of output for a firm at t = 1.

$$Pr(Q < q) = Pr(c'^{-1}(\Phi p) < c'^{-1}(\phi p))$$

$$\implies Pr(Q < q) = Pr(\Phi < \phi)$$

For $t \geq 2, 3, 4...$

$$G(Q|q') = H(\Phi|\phi') < H(\Phi|\phi'') = G(Q|q'')$$

Hence, proved.

Proposition-1: In sectors with high initial corruption shock (low ϕ_0), distribution of firm by size would have higher mass on the left tail.
Proof: At T = 0, $q_0 = c' - 1(\phi_0 p)$
As ϕ_0 decreases, q_0 decreases.
By Lemma-1, $G(Q|q_0)$ is smaller compared to firms in sectors with low extractive corruption. By Law of Large Numbers, there will be larger mass on the left tail of the distribution of firms by size.

2.4 Club Corruption

As defined above, club corruption refers to graft and kickbacks that politicians and regime holders receive in exchange for providing benefits. In this section, I model club corruption as one time payment by firms to politicians to avoid any extractive corruption. Thus, the key trade-off is between a one-time fixed costs on one hand and continuing variable costs (of extractive corruption) on the other. [1]

For a firm in the club, $\tau = 0$. Thus, profits of the firm are $\pi(0)$. Let $\bar{C}$ be the payment made to avoid club corruption.

Assumption-5: Each firm has credit availability $C_f \sim H(0, \infty)$, where H is the distribution of firm's liquidity or cash availability.
Politicians maximize aggregate club fee subject to IC and Credit Constraint

$$Max_C C \int_C^{\infty} h(x)dx = C(1 - H(C))$$

Subject to

$$IC : \frac{\pi(0)}{1-\delta} - C \geq \frac{\pi(\bar{\tau})}{1-\delta}$$
$$CC : C_f \geq C$$

where IC is incentive compatibility constrain and CC is credit availability constraint.

2.5 Dynamics under Club Corruption

Theorem-2: Let $\bar{C}$ be the optimal club fee charged by the politician. $\bar{C}$ is weakly increasing in $\bar{\tau}$

Proof: Unconstrained Optimum is when $C^* = \frac{(1-H(C^*))}{(h(C^*))}$
If IC is satisfied with inequality at C^*, then $\bar{C} = C^*$

If IC holds with equality then $\bar{C} = C(\bar{\tau})$ where $C(\bar{\tau}) = \frac{\pi(0)-\pi(\bar{\tau^*})}{1-\delta}$

$$\bar{C} = \begin{cases} C(\bar{\tau}) & : C^* \geq C(\bar{\tau}) \\ C^* & : C^* < C(\bar{\tau}) \end{cases}$$

Let $\hat{\tau}$ be the solution to $C^* = \frac{\pi(0)-\pi(\hat{\tau}}{)}1 - \delta$ Since, $C(\bar{\tau})$ is increasing in $\bar{\tau}$ we get

---

[1] In reality, club corruption exists not simply to avoid extortionary demands by lower ranking public officials but to garner special status to the current regime. I abstract away from that.

$$\bar{C} = \begin{cases} C(\bar{\tau}) & : \hat{\tau} \geq \bar{\tau} \\ C^{*} & : \hat{\tau} < \bar{\tau} \end{cases}$$

Proposition-2: In sectors with higher corruption shocks, there will be lower entry in firms on the right tail of the distribution.
Proof: As extractive corruption increases, $\bar{\phi}$ decreases. If continuation payoff from outside the club decreases and entry cost of being in the club increases. By Hopenhayn (1992), increase in entry cost would decrease turnover.

3 Data

3.1 Industry-level Data

To measure outcomes on firms, I use data from Annual Survey of Industries (ASI) from fiscal years 1999 to 2010. The ASI is an annual survey of manufacturing establishments conducted by Ministry of Statistics and Programme Implementation (MOSPI) of Government of India. This survey covers factories in manufacturing sector registered under Factories Act. Registration under this Act is required for establishments with more than 10 employees if using power or more than 20 employees if not using power. ASI provides data on fixed capital, depreciation, labor expenditure, employment, type of ownership, gross output and gross revenue among other things and 4-digit industry code of the plant. The dataset has 582605 plant-year observations.

3.2 Corruption Data

Corruption is difficult to measure due to the criminal nature of the act. Although some measures exist, Academia does not seem to have arrived at a consensus on which corruption measure to be used as default. To address this problem, I use alternative measures of corruption. If the theory is correct, similar results should be obtained across different corruption measures. In the empirical tests of theorems, I use the following three measures as proxies for average initial shock of corruption:-

1 **Perception and Experience Based measure** of corruption across states from Transparency International (TI) 2005 report. In this survey, respondents from urban and rural sector were surveyed regarding whether they have faced any corruption

and what there perception is of corruption in public service. From these surveys, TI computed a composite score of corruption in each state.

2 **Per Capita Bribe Paid** in each state from TI 2008 report. This survey was a follow-up on TI 2005 survey. However, in this survey an emphasis was put on below poverty line households. From this survey, I use the estimate of per capita bribe paid by poor household in each state as a proxy of initial corruption shock.

3 **Transmission and Distribution Losses** from 2002-2003 for each state. This is the amount of power generated in each state but not paid for as a percentage of total power generated. Kocchar et al (2006) used this as a measure of corruption since it represents "politician's unwillingness to enforce laws, as well as viability and level of corruption in each state."

3.3 Regulatory Environment for firms in States

Due to the Federal structure of Indian constitution, state governments have independence in establishing the nature of policies that influence economic activities. In this paper, I use the following metrics to control for heterogeneity in state policies.

1 Product Market Regulation for each state. I use the measures used in Dougherty et al (2011) as they comprise of various factors which govern regulatory environment for firms in each state affecting variables like entry, exit and competition[2].

2 Labor Market Regulation for each state. The classic measures used to proxy labor market regulation in Indian states has been the index first developed in Besley & Burgess (2002), which considers amendments conducted by state governments in Industrial Dispute Act. However, I rely on the metric constructed by Dougherty et al (2011). It is a composite measure of 8 fields regarding labor laws including the Industrial Disputes Act[3].

3.4 Supply-Side Factors for firms

Considering each state as a localized market, I control for supply-side factors in each state.

---

[2] Factors include nature of state control, barriers to entrepreneurship and Administrative Regulations.

[3] Other factors include Factories Act, Shops Act, Contract Labor Act, Role of Inspectors, Register and Union Representation

1. Development Index from Fisman & Khanna (1998). This is a composite measure of infrastructure (both physical and social) of each state. This measure proxies the supply-side factors determining firm decisions.
2. Electricity Generated in each state from 2002 to 2008. Literature on endogenous location of firms (Carlton (1979)) suggests total power generated is a strong predictor of where firms want to locate. Since, the second proposition is essentially about entry of firms in each state, I use this as a control in the test of second proposition.

3.5 Demand-Side Factors for firms

Analogously, I control for demand-side factors in each state.

1. Log of Gross State Domestic Product (GSDP) in each state in each year. This measure proxies the level of demand factors influencing firm decisions. Since, the first hypothesis is essentially about the size of the firm in each state, log of GSDP would be used as a control in the test of first hypothesis.
2. Average Level of Urbanization for each state. Urbanization rates are available from Census 2001 and Census 2011. I use the mean of these two levels to depict concentration of high-value economic activity in each state.

Finally, I use a very recent and unique measure from Amirapu & Gechter (2015) of firm's propensity in each state to mis-report to avoid regulation .

4 Empirical Strategy and Results

4.1 Empirical Strategy for Proposition-1

Proposition-1 states that in sectors where initial corruption shocks are high, mass on the left tail of the distribution of firms by size would be higher. As a metric for mass on the left tail of the distribution, I use skewness of the distribution which is defined as third moment around mean. Positive (Negative) skewness of a distribution represents higher left tail (right tail) weight. For size of the firm, I use log of total gross sales value of a plant in a given year.

I construct measures of skewness for each 4 digit industry for each state in each year. I obtain 32759 observations for all 31 states in India. However, information on regulatory environment and supply-side factors is not available for north-eastern states, smaller states and Union Territories. I am restricted, thus, to 20393 observations. Table 1 details summary statistics on skewness measure. The mean of skewness is very small. Small mean of skewness can also be seen in Figure 1, which is the density of skewness pooled across all industry-state-year category. The density of skewness is very highly concentrated around the mean suggesting very low variance.

To test Proposition-1, I use the following empirical model:-

$$skew\_dist_{isy} = \beta z\_corr_s + \kappa X + \phi_I + \phi_y + \epsilon_{isy}$$

where

1 $skew\_dist_{isy}$ is skewness of industry i in state s and year y

2 z_corr_s is the standardized value of corruption score for state s. As mentioned above, I use three different measures of corruption- Perception Based Measure from Transparency International 2005, Per-capita bribe paid in each state from Transparency International 2008 and Transmission and Distribution Loss from 2002 to 2008.

3 X contains those state specific variables which affect size of the firm. I include index of product and labor market regulations in states (Dougherty et al 2011), standardized score for infrastructure quality (Fisman and Khanna 2008), log of gsdp in each state for each year and a measure of firm's propensity to avoid regulation (Amirapu and Gechter 2014).

4 ϕ_I and ϕ_y are industry and year fixed effect

Hypothesis: $\beta > 0$
I use robust standard errors.

Table 2 provides the result for the above model. Across all specifications, skewness of the industry is positively co-related with all three corruption measure confirming the results of the theory. Further, co-efficient on corruption measure, β, is strongly significant when perception based or per capita bribe paid measure is used. On using transmission and distribution loss, significance of the effect is lost. However, that is understandable since transmission and distribution loss is a composite measure of lack of institutional measure over and above corruption. Although, the magnitude of the effect appears to be small, the mean of skewness is -0.004 and very low variance as observed in Figure 1. Thus, given the distribution of skewness, the effect is large.

4.2 Empirical Strategy for Proposition-2

Proposition 2 states that in sectors with high corruption shock, there will be lower entry of large firms. In the ASI data, all firms with more than 100 employees are surveyed every year creating a census of "big" firms. This allows me to observe the universe of big firms.

To test this proposition, I first find the number of firms that entered in industry i in state s in year t and divide that by the number of existing firms in that industry-state-year cross-section. This gives me the entry rate with a total number of 16536 industry-state-year observations. As mentioned above, not all states have information regarding there regulatory environment and supply-side factors leaving me with 7258 observations. Table 3 details summary statistics on entry rate. Figure 2 presents the density of entry rate of firm pooled across industry-state-year observations. Bi-modality of distribution reflects high variance of the entry rate of firms.

I use the following model:-

$$entry\_rate_{isy} = \beta z\_corr_s + \kappa X + \phi_i + \phi_y + \mu_{isy}$$

where

1 $entry\_rate_{isy}$ is proportion of new firms entering in industry i in state s in year y

2 z_corr_s is standardized value of corruption in state s. I use three different measures of corruption- Perception Based Measure from Transparency International 2005, Per-capita bribe paid in each state from Transparency International 2008 and Transmission and Distribution Loss from 2002 to 2008.

3 X contains state specific variables and controls for confounding effect. These consist of factors that affect entry of a firm in a given state. I include average level of urbanization and total amount of electricity produced (Carlton (1979)), apart from product market and labor market regulations.

4 ϕ_I and ϕ_y are industry and year fixed effect

Hypothesis: $\beta < 0$
I use robust standard errors.

Table 4 provides results for the test of Proposition-2. The effect of standardized corruption score on entry rate remains negative in all specifications. However, magnitude of the coefficient on corruption, β, is small and insignificant in some specifications. Further, the effect is unstable and ranges from -0.001 (insignificant)to -0.052 (significant). Thus, inference of the effect remains an issue. However, the effect, when significant, is close to 10% of the mean rate of entry of firms.

It should also be noted that when perception based measure is used β becomes significant on using tau. This is a measure of firm's propensity of misreporting to avoid regulation in each state (Amirapu and Gechter 2015). Further, the co-efficient on tau is positive and strongly significant in all specifications. This is a peculiar effect which requires further exploration in future work.

5 Discussion

Propositions of theoretical model rely heavily on Theorem-1 which states that higher initial corruption shock leads to higher likelihood of corruption in next period. This way of modeling corruption should be contrasted to tollbooth theory of corruption (Djankov et al (2002)), which states that corruption is conducted using cumbersome entry-level regulations. However, while Djankov et al (2002) only considers entry level regulation, Theorem-1 reflects the continuing interaction between industry and government. In the models above, entry level regulation is captured by product market and labor market regulations. Thus, the results in the above models are over and above the effects recorded in Djankov et al (2002).

In the above empirical models, I use corruption measure at the state level. However, most corrupt states are poorer states with myriad of political and institutional shortcomings apart from corruption. This makes it likely that the corruption variable is picking up a state level effect which is not corruption. While this is a pertinent concern, it should be noted that the effect remains robust to different specifications and alternative measures of corruption. This adds weight to the theoretical arguments provided above.

In the empirical models, I allow errors to co-relate across all industry-state-year observations, thus, using robust standard errors. Alternatively, one could expect strong serial co-relation to occur across time within a state. Hence, an argument can be made (as per Bertrand et al 2003) to cluster at the state level. While this criticism is fully acknowledged, it should be recored that asymptotic properties of clustered standard errors depends on the number of clusters. Most of the regressors used above are estimated for at most 20 states which makes clustered standard errors unreliable and hence, inference of the effect becomes an issue.

6 Conclusion

Corruption is a serious issue and has for long been the focus of academic and policy interest. However, the mechanism of how corruption actually effects economic outcomes needs to be understood.

In the paper, I model corruption as a stochastic dynamic process which affects firm outcomes in heterogeneous manner. Using this formulation, I explain certain well-studied stylized facts regarding industry level dynamics, namely large number of small firms and higher persistence of big firms. I test the results of the model on industry-level data from India to show that states with higher level of corruption tend to have a higher mass on the left tail of firm distribution of firms by size and lower entry of big-sized firms which coincide with the stylized facts.

The effects are robust to alternative measures of corruption and various state specific controls. Further, the results go beyond the current studies which attribute mis-allocation in firm outcomes to regressive either state-level labor and industry laws (Dougherty et al (2011)) or deregulation in early 1990s (Aghion et al (2008)). Hence, past studies need to be revisited with a fresher perspective.

7 References

Aghion, Philippe, et al. "The unequal effects of liberalization: Evidence from dismantling the License Raj in India." The American Economic Review 98.4 (2008): 1397-1412.

Alfaro, Laura, and Anusha Chari. Deregulation, misallocation, and size: Evidence from india. No. w18650. National Bureau of Economic Research, 2012.

Besley, Timothy J., and Robin Burgess. "Can labour regulation hinder economic performance? Evidence from India." (2002).

Carlton, Dennis W. Why new firms locate where they do: An econometric model. Joint Center for Urban Studies of MIT and Harvard University, 1979.

Djankov, Simeon, et al. "The regulation of entry." Quarterly journal of Economics (2002): 1-37.

Dougherty, Sean, Verónica C. Frisancho Robles, and Kala Krishna. Employment protection legislation and plant-level productivity in India. No. w17693. National Bureau of Economic Research, 2011.

Fisman, Raymond, and Tarun Khanna. "Facilitating development: The role of business groups." World Development 32.4 (2004): 609-628.

Hopenhayn, Hugo A. "Entry, exit, and firm dynamics in long run equilibrium." Econometrica: Journal of the Econometric Society (1992): 1127-1150.

Hsieh, Chang-Tai, and Peter J. Klenow. Misallocation and manufacturing TFP in China and India. No. w13290. National Bureau of Economic Research, 2007.

Hsieh, Chang-Tai, and Benjamin A. Olken. "The missing “missing middle”." The Journal of Economic Perspectives 28.3 (2014): 89-108.

India Corruption Study 2005, Transparency International

India Corruption Study 2008, Transparency International

Jovanovic, Boyan. "Selection and the Evolution of Industry." Econometrica: Journal of the Econometric Society (1982): 649-670.

Kochhar, Kalpana, et al. "India's pattern of development: What happened, what follows?." Journal of Monetary Economics 53.5 (2006): 981-1019.

Niehaus, Paul, and Sandip Sukhtankar. "Corruption dynamics: The golden goose effect." American Economic Journal: Economic Policy 5.4 (2013): 230-269.

Olken, Benjamin A., and Rohini Pande. Corruption in developing countries. No. w17398. National Bureau of Economic Research, 2011.

Olken, Benjamin A., and Patrick Barron. "The simple economics of extortion: evidence from trucking in Aceh." (2007).

Pham, Hien Thu, and Shino Takayama. "Revisiting the "Missing Middle": Productivity Analysis." (2015).

Sukhtankar, Sandip, and Milan Vaishnav. "Corruption in India: Bridging Research Evidence and Policy Options." India Policy Forum. Vol. 11. 2015.

Sukhtankar, Sandip. "Sweetening the deal? political connections and sugar mills in India." American Economic Journal: Applied Economics 4.3 (2012): 43-63.

Tybout, James R. "Manufacturing firms in developing countries: How well do they do, and why?." Journal of Economic literature 38.1 (2000): 11-44.

Tybout, James. "The missing middle, revisited." Journal of Economic Perspectives 28.4 (2014): 235-36.

Shleifer, Andrei, and Robert W. Vishny. "Politicians and firms." The Quarterly Journal of Economics (1994): 995-1025.

8 Appendix

To prove Theorem-1, I will first establish the process of updating quality of institution, k, by the bureaucrat.

First note that $k' > k'' \implies \tau' < \tau''$
Let us assume $k \sim U[0, k^m]$ where k^m be the maximum quality of institution[4].
Consider a bureaucrat, who was successfully able to extract rent τ from a firm in period t. Denote by τ^* level of extraction when it is successful. In the following period, she updates k using Bayes' rule in the following manner:

$$Pr(k < K/\tau^*) = \frac{Pr(k < K)Pr(\tau^*/k < K)}{Pr(\tau^*)}$$

where $K \in [0, k^m)$

Under our assumptions,
$Pr(k < K) = \frac{1}{K}$, $Pr(\tau^*/k < K) = 1 - \tau\frac{K}{2}$ and $Pr(\tau^*) = 1 - \tau\frac{k^m}{2}$

Thus,

$$Pr(k < K/\tau) = \frac{1}{K}(\frac{1 - \tau\frac{K}{2}}{1 - \tau\frac{k^m}{2}})$$

Since, $K < k^m$, $(\frac{1-\tau\frac{K}{2}}{1-\tau\frac{k^m}{2}})$ is increasing in τ. Thus,

$$\tau' > \tau''$$

$$\implies (\frac{1 - \tau'\frac{K}{2}}{1 - \tau'\frac{k^m}{2}}) > (\frac{1 - \tau''\frac{K}{2}}{1 - \tau''\frac{k^m}{2}})$$

$$\implies Pr(k < K|\tau') > Pr(k < K|\tau'')$$

Given that belief over k and extraction level are inversely related, we get

$$F(\tau|\tau') < F(\tau|\tau'')$$

Hence, Theorem-1 is proven. Further, since $\phi = 1 - \tau$, the equivalent result we can obtain is $\phi' > \phi'' \implies H(\phi|\phi') < H(\phi|\phi'')$

---

[4] This is a simplifying assumption and proofs would follow for a more general functional form of distribution as well.

9 Figures

Figure 1: Skewness

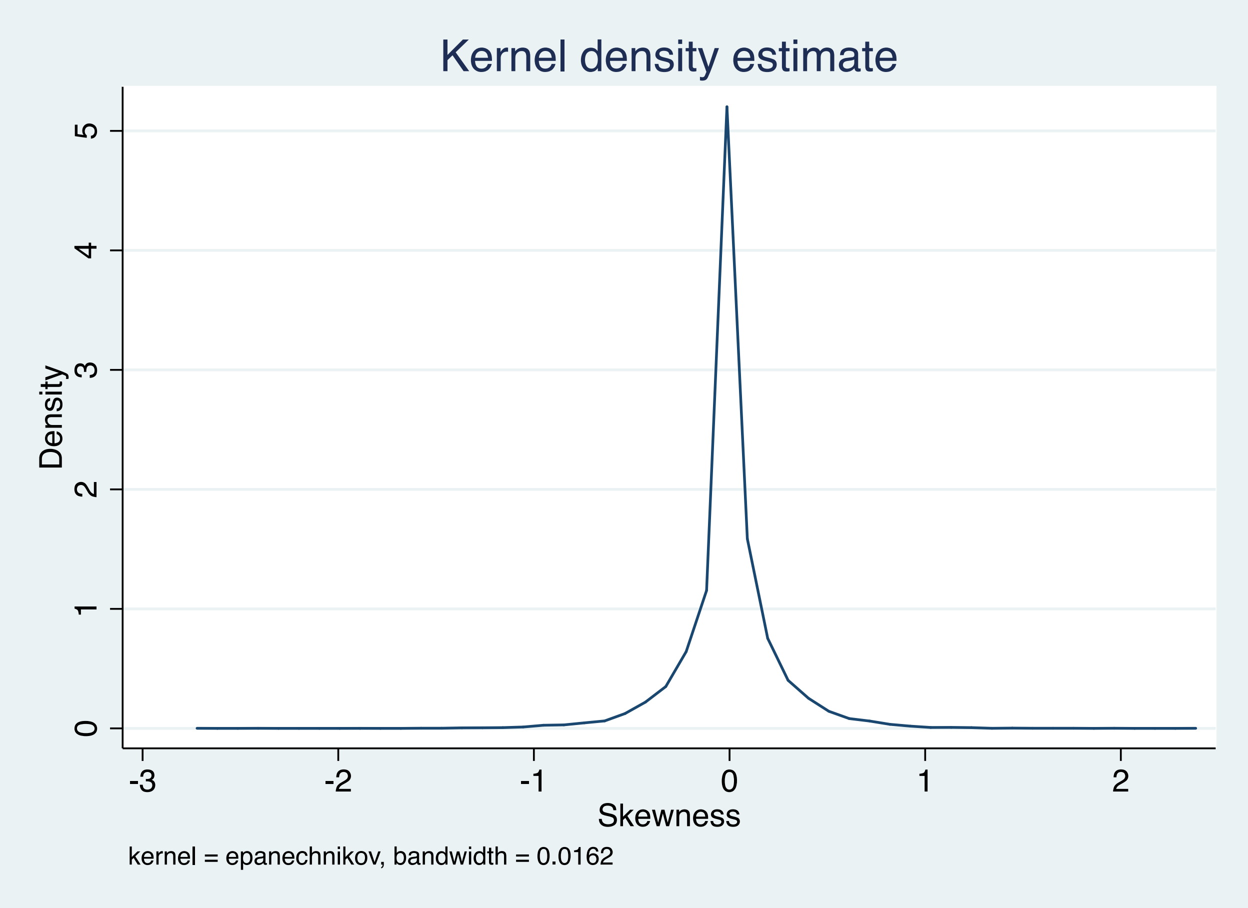
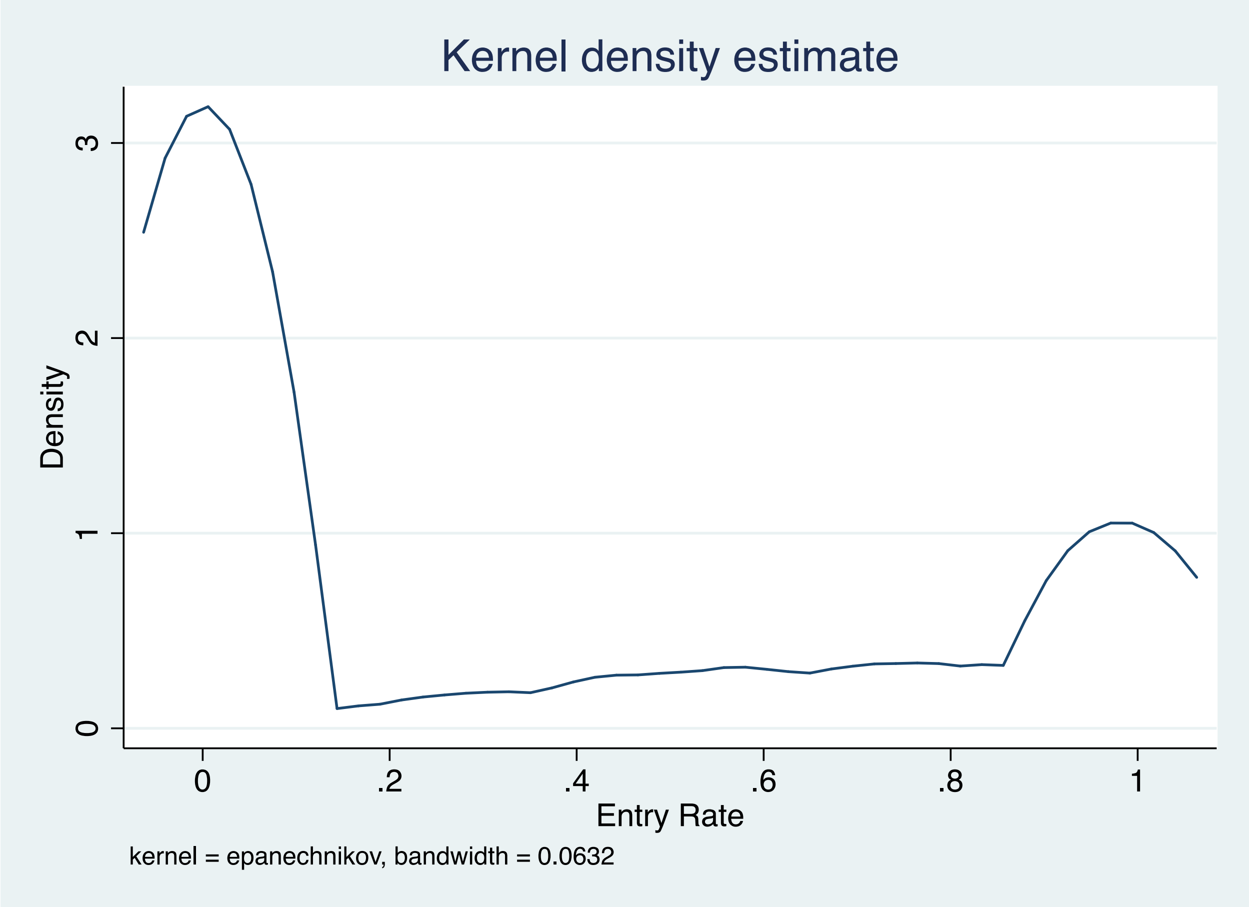


Figure 2: Entry Rate



10 Tables

Table 1: Skewness: Summary Statistics

Mean	Standard Deviation	Observations
0.000038	0.253	20393

Table 2: Skewness of Industry

	Skewness	Skewness	Skewness	Skewness	Skewness	Skewness
Corr_2005 (Standardized)	0.005*	0.005**				
	(0.003)	(0.003)				
Corr_2008 (Standardized)			0.006***	0.006***		
			(0.002)	(0.002)		
T & D Loss (Standardized)					0.007	0.007
					(0.005)	(0.005)
Prod. Mkt Regulations	0.003	0.002	0.002	0.002	0.005*	0.004
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Lab. Mkt Regulations	0.003	0.002	0.006**	0.006*	0.002	0.001
	(0.002)	(0.003)	(0.002)	(0.004)	(0.003)	(0.003)
Development Index	0.036***	0.040***	0.030***	0.027*	0.056***	0.060***
	(0.011)	(0.012)	(0.012)	(0.015)	(0.013)	(0.015)
log_gsdp	-0.025**	-0.032***	-0.039***	-0.035***	-0.039***	-0.047***
	(0.010)	(0.012)	(0.007)	(0.012)	(0.010)	(0.014)
tau (Amirapu & Gechter)		-0.003		0.001		-0.003
		(0.003)		(0.003)		(0.004)
Observations	20393	20393	17754	17754	12129	12129

I use robust standard errors in all specifications. All specifications use Industry and Year Fixed Effects. Development Index has been adopted from Fisman and Khanna (1998) and log of gsdp is from each year in each state at 2005 prices.

Table 3: Entry Rate: Summary Statistics

Mean	Standard Deviation	Observations
0.311	0.415	7258

Table 4: Entry Rate in Industry

	entry_rate	entry_rate	entry_rate	entry_rate	entry_rate	entry_rate
Corr_2005 (Standardized)	-0.001	-0.013*				
	(0.005)	(0.007)				
Corr_2008 (Standardized)			-0.027***	-0.029***		
			(0.005)	(0.005)		
T & D Loss (Standardized)					-0.038***	-0.052***
					(0.009)	(0.010)
Prod. Mkt Regulations	0.010**	0.012***	0.007	0.009*	0.006	0.009*
	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Lab. Market Regulations	-0.008	-0.004	-0.010	-0.002	-0.010*	-0.005
	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)	(0.006)
Development Index	0.190***	0.212***	0.248***	0.254***	0.173***	0.197***
	(0.025)	(0.027)	(0.030)	(0.030)	(0.026)	(0.026)
Avg. Urbanization	-0.006***	-0.006***	-0.005***	-0.005***	-0.007***	-0.008***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Electricity Prod.	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
tau (Amirapu & Gechter)		0.019***		0.016***		0.023***
		(0.006)		(0.006)		(0.006)
Observations	7258	7258	6323	6323	7258	7258

I use robust standard errors in all specifications. All specifications use Industry and Year Fixed Effects. Development Index has been adopted from Fisman and Khanna (1998). Average Urbanization is the mean of urbanized rates of each state in year 2001 and 2011. Electricity Produced is the total amount of electricity generate in each state from 2003 to 2008.