Courts, Firms and Informality*

Job Market Paper

Peter Neis†

Toulouse School of Economics

November 13, 2022

This paper studies the impact of judicial efficiency on informality. In many developing countries most firms and workers operate and work informally outside the legal and regulatory systems. A better judicial system can lead to a higher share of formal firms, since it changes the marginal benefit of being formal. Similarly, the impact on the workforce can go either way as formal firms can be incentivized to hire fewer formal workers. This can lead to the paradoxical effect that a more efficient judicial environment results in a larger share of formal firms albeit with fewer formal workers. The paper introduces judicial efficiency in an equilibrium model. Judicial efficiency can affect firms' decisions through two channels. First, it can have a direct impact on a firm's productivity, this captures a general, better economic environment. Second, it can have an impact on the relative cost to hire formal workers compared to informal workers. The paper uses data from India to test the impact of the judiciary on the two margins of informality. In India, most firms and workers are operating informally while the judicial system is hampered by its large and increasing backlog. Results show a negative effect of court efficiency on the share of informal firms among small firms and a negative but not statistically significant effect on the share of informal workers. Consolidating these findings with the proposed model implies that the observed effects can be explained solely by a relative cost on formal workers in formal firms and one does not need to include a general TFP shock.

Keywords: Judicial efficiency, Informal labor, Informal Firms
JEL Codes: E26, J46, K40, O17

*I am indebted to Daniel Chen, Mohamed and Saleh and Shareen Joshi for their for their guidance and support over the last years. I thank Matteo Bobba, Lorenzo Casaburi, Mateo Montenegro, Stéphane Straub for helpful comments and suggestions. I am very grateful to Sandeep Bhupatiraju without whom the court case data would have never been collected. Finally, I thank participants of workshops and seminars at Toulouse, the World Bank, Zurich and from the Enter Jamborees in Tilburg and Barcelona and the Economics of Informality Conference.

†Toulouse School of Economics, Université Toulouse Capitole, Toulouse, France. E-mail: peter.neis@tse-fr.eu

1. Introduction

A recent World Bank report states that "the justice system can have a direct impact on formality, making formality not only attractive but also feasible" and goes on to call for a consistent, fair, and effective judicial system (Loayza 2018). Informality is one striking characteristic of most developing economies. In many countries, most of the workforce is informally employed and many, often small, enterprises are not registered.[1] Also, over the last decades, institutions have been at the heart of developing economics (see for instance Acemoglu and Johnson 2005). The judiciary has emerged as one important institution in society with strong influence on firms' behavior and output.

Firms are impacted by the judicial system. A dysfunctional judicial system can lead to underinvestment (Klein, Crawford, and Alchian 1978), inefficient investments (Boehm and Oberfield 2018), reduction in credit (Visaria 2009), reduced access to credit (Lilienfeld-Toal, Mookherjee, and Visaria 2012) and economic growth (Amirapu 2017). All these channels will impact a firm's profit maximization problem when deciding whether to register or not and to hire informal workers or not. Most of the mentioned channels, such as better access to credit due to a well-functioning judicial system or more investments due to clearer property rights, give a comparative advantage to registered firms. Therefore, by improving the judicial system, at the margin, some firms might decide to become formal to have access to the benefits of formality.

Firms deciding to register, have another choice to make. How much of the workforce should be employed formally and how much informally? Again, this decision is likely to be impacted by the efficiency of the judicial system. Different effects can occur and have opposite signs. It depends on who is benefiting most from efficient and neutral of courts. One can think of a better judicial system allowing workers more easily to file cases against their employers if these are not respecting their contract or are violating labor laws. This can lead to more formal firms wanting to hire informal workers. At the same time, a better judicial system, which is inclined towards protecting employers, can allow firms to enforce labor contracts (Naidu and Yuchtman 2013). If this effect dominates, a better judiciary system should lead to more formal workers. Therefore, the sign of the effect remains an empirical question.

But, despite much work on links between weak institutions and informality and despite the strong empirical and theoretical background between institutions in general and informality and the judiciary and firms' behavior, little is known about the direct influence of the judiciary on informality. This paper links court efficiency to informality. First, it extends the model from Ulyssea (2018). It adds court efficiency to firms' profit maximization process to provide a theoretical understanding of how firm behavior links

[1] In India, around 80% of workers are informally employed.

court efficiency and informality.[2] Then it tests the predictions of the model by combining case-level data from district courts with multiple Indian data sources on firms and workers. The analysis documents two different effects of a more efficient judiciary - the effect on share of firms being registered and the effect on share of informal workers in formal firms.[3]

It is the first study linking the judiciary to informality using microdata from courts to construct court efficiency measures and combining firm-level and individual-level survey data to construct informality aggregates. Furthermore, it is the first study analyzing the impact of court efficiency on the two margins of informality defined by Ulyssea (2018): firms decide whether they want to register with the authorities (extensive margin) and formal firms decide if they want to hire informal workers (intensive margin).

Cross-sectional 2SLS regressions on the district level show a strong negative link between court efficiency and the share of firms being informal. On the intensive margin of informality, a negative, but not statistically significant impact of court efficiency on the share of informal workers as well as the share of workers in formal firms being informal is observed. Together with the theoretical model, the results suggest that there is no direct effect on the production process and that courts are potentially inclined to back the employers' cases.

This paper contributes to several literatures. Friedman et al. (2000) compare 69 countries and find that informality is correlated with their measure of the legal systems. Using an instrumental variable approach, they find that the better the judiciary, the lower the part of the total output produced by informal firms. Johnson et al. (2000) look at the intensive margin. They find that the hiding of output of registered firms in former European communist states firm is not correlated with the perceived efficiency of courts. Dabla-Norris, Gradstein, and Inchauste (2008) use data from the World Business Environment Survey to conduct a cross-country study to see what influences the share of informality. They do not make the difference between intensive and extensive margin and while looking at the quality of the legal framework as a determinant of informality they refer to it mainly as enforcement capacity. Nevertheless, they find that legal institutions have an impact on the size of the informal sector. Assenova and Sorenson (2017) uses hierarchical linear models on firm-level data from 18 states in sub-Saharan Africa and a country-wide level of citizens' perception of distrust in the court system to analyze the extensive margin of informality. They find that the more the population trust in courts the higher the benefits from registration. Shapiro (2015) parameterize a job-search model, based on Ulyssea (2010), and finds that better formal institutions lead to higher benefits

[2]Importantly, this paper does not look at improved enforcement (i.e. an increase of the police force, a rise in fraud detection capacity, ...) but solely at the speed of district courts which is taken as an indicator for the quality of the dispute settlement mechanisms.

[3]In a future version, a full structural estimation of the model can allow counterfactual and welfare analysis.

from hiring formal workers.

Other studies did not look at the impact of the judicial system on informality but used the Indian context to study aspects of informality or the interplay between the judicial system and firms. Martin, Nataraj, and Harrison (2017) analyze the effect of small firm protections in India. Using the de-reservation started in the late 1990s they exploit variation in timing to measure the impact on employment. They check for impacts on unorganized production but do not find a significant effect. Still, there is some evidence for a shift from the unorganized to the organized sector. Bertrand, Hsieh, and Tsivanidis (2017) use a judgment in 2001 on the Industrial Disputes Act of 2001 to analyze the effect of weak labor protection on workforce and growth. They argue that since the judgment companies rely more on contract workers which they can fire whenever they want, and therefore, hire more workers. Saikia (2011) looks at patterns at the state-level for the unorganized manufacturing sector before and after the reforms of the late 90s. Unorganized manufacturing stays very concentrated in some states but declines in almost all sectors. Abraham (2018) compares the workforce of informal and formal workers in informal and formal enterprises in India over time. She points out the importance of states as determinants of wage inequalities, especially for informal workers in informal enterprises. Hsieh and Olken (2014) show that there is no bimodal distribution of firm size for informal and formal firms in India, Indonesia, and Mexico. Besley and Burgess (2004) use amendments to the Indian Industrial Disputes Act to create state-wide variation in pro-worker or pro-employer climate. They find that pro-worker changes lower output, employment, investment, and productivity in formal firms while increasing output in informal firms. Surprisingly, they find as a byproduct that court inefficiency is correlated with lower informal sector manufacturing. Boehm and Oberfield (2018) look at how average court congestion (and so contract enforcement) affects the allocation of input use in the manufacturing sector in India. They find that lower contract enforcement leads to lower cost shares of intermediate inputs in industries that rely more on relationship-specific intermediate inputs, to a higher share of standardized intermediate inputs, and more vertical integration for those plants using relationship-specific intermediate inputs.

The remaining of the paper is structured as follows. Chapter 2 gives a quick overview of informality in India and some background information on her judiciary system. The next chapter presents the data. This is followed by the theoretical model. Chapter 5 presents the empirical strategy and results. The last section concludes.

2. Background

2.1. Informality in India

With its 1.2 billion inhabitants is India not only the second largest country and the biggest democracy in the world but also the largest common law country. Albeit the impressive growth of its economy over the last four decades, India's economy is, as it is the case for many developing countries, marked by high shares of informality. In 2018, with a high 80%, the World Bank placed it third out of 39 countries with respect to informal employment as a percentage of total non-agricultural employment.

In this paper, informal workers are defined as all workers with no formal labor contract and no social security benefits of any sort. This includes two groups, self-employed workers who work on their own benefit and casual workers who are unofficially hired by a firm. In 2010, 75% workers were working informally.[4]

Firms in India can be classified into three categories. First, self-employed, who do not get help from anyone outside the own household. Self-employed do not benefit from social security benefits and are mainly poor. They represent the largest group among Indian firms. A second group are non-registered firms which hire casual labor outside the own household. These two groups together are aggregated as "informal" firms and represented 71% of all Indian firms in 2010. Finally, there are formal firms which are registered with an authority. These formal firms are, on average, larger and have higher profits than informal firms.

2.2. Courts in India

India has the largest common law system in the world. Its judicial system has several layers of court hierarchy (compare Figure 1). The highest court in India is the Supreme Court, which handles appeals from high courts, questions of fundamental rights and disputes between different organs of the state. 25 High Courts constitute the next lower level. Each of them has, in general, judiciary over one or two of the 28 states. High courts are mainly courts of appeal against decisions of lower courts but exercise also original (civil and criminal) jurisdiction for cases which do not fall into the jurisdiction of lower courts.

Some high courts have benches in other parts of a state (beside the main seat). The area of jurisdiction of a high court is divided into judicial districts. The highest court in a judicial district is the district and sessions court. Judges are called district judges when handling civil, and sessions judges when handling criminal cases. These district and sessions courts are the principal courts of original civil jurisdiction in a district. The

[4]While a distinction between self-employed and casual workers is certainly interesting in some settings, it does not have a first order importance when investigating the link between courts and informality.

district courts are supervised and administered by the corresponding high court. Today, there are 671 district courts in 739 districts.

India is known for its high number of pending cases and for its slow judiciary. Although India has put effort in decreasing the backlog and speeding up court decisions, the number of pending cases has continued to increase in recent years. Today there are more than 800 pending cases per judge in the lower judiciary, with new filings exceeding disposals. The Heritage Foundation writes about rule of law in India: "Property rights are generally enforced in major metropolitan areas, although titling in some other urban and rural areas remains unclear. The judiciary is independent, but lower-level courts are understaffed, lack technology, and are rife with corruption. Most citizens have great difficulty securing justice."[5]

3. Data

This study uses different data sources from around 2009/10. All variables are created or aggregated at the district level. Depending on the data used, the final number of districts in the analysis lies between 189 and 420.[6] Table C.1 gives an overview of the number of districts per source.[7] Appendix C explains how districts were selected and some variables created. Figure 2 gives a graphical overview of which districts are included in the final data set and which are not.

3.1. Firms and Informality

SUNAE: From July to June 2011, the 67th round of National Sample Survey (NSS) was conducted and included a survey on Unincorporated Non Agricultural Enterprises (Excluding Construction) (SUNAE). The survey covered 334,474 firms and collected economic and operational characteristics of enterprises in manufacturing, trade and other service sector (excluding construction). The survey excluded all enterprises which are incorporated i.e. registered under Companies Act, 1956 or Factories Act, 1948, government and public sector enterprises and cooperatives.

IEU 2009/10: The Indian National Sample Survey Office conducted the 66th round of National Sample Survey (NSS) between July 2009 and June 2010. This round included the Indian Employment and Unemployment survey (IEU), an all-India household survey aiming at providing comprehensive data on the labor force and activity profiles of the

[5] https://www.heritage.org/index/country/india, accessed 02.09.2020

[6] While all data exist as well for the year 2015, some of these more recent datasets are not in the public domain or do not have district identifiers. The process to get access to these datasets has been started and hopefully a future version of this paper will be using this more reliable data.

[7] While backlog and the average age of pending cases are available in 506 districts, Clearance rate and disposition time have only around 190 observations. This is due to some districts having zero measures in some years.

population, beside other goals. For this survey, 100 957 households and 459 784 persons were surveyed.

ASI 2009/10: The Indian National Sample Survey Office conducts the Annual Survey of Industries (ASI), and annually surveys registered factories all over India. ASI consists of two parts. First, a census part which broadly includes all units having 100 or more workers and all factories covered under joint returns. Second, a sample part, which covers 19% from the remaining units in each State × NIC-04-code cluster. In total, ASI 2009/10 covered 61,080 units (23,782 census and 37,298 sample).

From these three surveys, different moments of informality are created. Table E.1 gives an overview over the moments and the data sources. ASI and SUNAE are used to create a size distribution of respectively formal and informal firms. The combination of the two survey gives the share of firms being informal. On the intensive margin, the total share of informal workers and the share of informal workers in formal firms are created from IEU.

3.2. Courts

Court efficiency is calculated from case level data from District and Sessions Courts. This data was webscraped from the Indian eCourt smartphone application.[8] The eCourts website is a centralized project of the Indian government. It was first implemented in 2007 following the "National Policy and Action Plan for Implementation of Information and Communication Technology (ICT) in the Indian Judiciary - 2005" with the aim to make the judiciary more efficient and accessible by computerization of courts. Following this implementation, many court complexes were computerized and processes digitalized. The eCourt project claims that today all district courts are connected to the internet and that most pending cases are entered into the centralized database.[9] In total, the website provides cause lists and case status for more than 70 million pending and disposed cases.

The database includes the flow and part of the stock of cases for all district courts connected to it. This means it includes all cases which have been treated at least once since a court was connected to the eCourt system. Therefore, all new cases and all cases disposed after the connection to the eCourt service are automatically included. The longer a court is already connected, the higher is the likelihood that all or most pending cases have been treated at least once. Still, there might be very old cases pending in the courts for a long time without ever having been treated after the implementation of ICT in a court. The web-scraped data contains cases filed in almost all district courts from 1954 until early 2019.

Indian district courts are highly congested. In many courts thousands of cases are

[8]Thanks to the Data Science Justice Collaboratory for sharing the raw data.
[9]https://ecourts.gov.in/ecourts_home/static/about-us.php

pending and it can take years until a new case is finally treated. Most first instance cases related to business, industries and labor are filed in District and Session courts (see for instance Rao 2020). Although these courts handle both, civil and criminal cases, and that firms are mainly involved in civil cases, both types add to the backlog of the courts. Therefore, no differentiation is made in this paper. Using the case level data from District and Sessions courts, for a given year (e.g. 2009) different court efficiency measures are calculated on a district level.

There is no unique measure of court efficiency (see the reports from Glanfield, Hall, and Keilitz 2018 and European Commission 2020). Court efficiency can be measured around multiple dimensions like speed of courts, access to justice and fairness of judgments. This paper focuses on the speed of court as one important indicator. Even for this single aspect of court efficiency multiple ways to construct a measure are possible. In the following four different measures are introduced, all of them used in different contexts in the literature.

First, a raw measure of court efficiency is backlog. This is the number of pending cases in a court. In the context of this paper, backlog is defined as all pending cases in a court at the end of a civil year.

Second, Boehm and Oberfield (2018) use the average age of pending cases. For this, all pending cases at the end of a year are taken and the average of their age (current data - filing date) is calculated.

Third, the disposition time is the estimated length of proceedings. This is the estimated number of days it takes for a case to be resolved in a court. It is calculated as the number of unresolved cases divided by the number of resolved cases in a court at the end of a year multiplied with 365.

Lastly, the clearance rate indicates whether a court is disposing more cases than new cases are filed. It is defined as number of resolved cases divided by the number of incoming cases.

3.3. Additional Data Sources

In 2009, the World Bank conducted a subnational Doing Business report for India.[10] The section "Paying Taxes" contains information on the tax rates a medium-sized company must pay in 17 cities in different states during the fiscal year 2007. Table E.3 depicts the statutory tax rates on salaries and revenues for two different cities, Bengaluru and Noida. As it is shown, the tax rates differ only slightly in the different states, the only difference being a 0.5 percentage points higher State VAT in Bengaluru. This leads to the conclusion, that a firm has to pay 16.75% of social security and insurance contributions on gross salaries and around 44% on revenues.

Other data sources include the Reserve Bank of India for outstanding credit of sched-

[10] https://www.doingbusiness.org/en/data/exploreeconomies/india/sub/noida#DB_tax

uled banks per district and the 2011 census for population data which are included in the empirical section as covariates.

3.4. Summary Statistics

Table 1 gives an overview over the data. On average, 73 % of firms in a district are informal. The share is higher for small firms and lower for large firms. Informal workers represent on average 92% of the total workforce and 36% of the workforce in formal firms. Again, the share of informal workers is larger among small formal firms than larger formal firms.

Panel (a) of Figure 4 depicts firm size densities for formal and informal firms. Most informal firms have a size between one and five. Formal firms on the other hand have a flatter density curve, with still a sizable amount of firms with a size larger than 15. Panel (b) of the same figure shows that the large majority of workers work in small informal firms. Panel (c) presents the share of the total workforce which works formally or informally in firms of a given size. More than 75% of the workforce works informally in small firms. Only in firms with more than 20 workers the majority of worker is formally employed. Panel (d) looks at formal firms only. One can see that there are both formal and informal workers in firms of all sizes. While for small formal firms the majority of workers is informally hired, this is reversed for medium-sized firms. In firms with more than 20 workers, almost two third of workers are formally hired.

On the side of the judiciary, on average there are 3320 pending cases in District and Sessions courts per district. These pending cases are on average more than three years old. The average clearance rate in a district is 0.32, implying that more cases are filed in District and Sessions courts than resolved. The mean disposition time lies around 1.6 and the mean occupancy rate of courtrooms in Districts and Sessions courts from 2004 to 2008 per district was 0.36. To get an overview over the variation, Figure D.1 contains histograms for backlog, average age of pending cases, clearance rate and disposition time at the end of 2008 in district and sessions courts per district.

4. A model of court efficiency and firms' decisions

This section presents a model of firm behavior incorporating judicial efficiency. It is a model à la Melitz (2003) and essentially a slightly modified version of the model in Ulyssea (2018), adding court efficiency into firms' maximization process. Heterogeneous entrants can choose to become formal or informal. If they chose to be formal, they can choose to hire formal and/or informal workers. These choices map the extensive and intensive margin of informality into firms' decisions. By adding court efficiency, courts can have an impact on both decisions and therefore on both margins.

4.1. Firms

Firms have a productivity θ and produce a homogeneous good using labor as only input:

$$y(\theta, l) = \theta q(l).$$

the function $q(\cdot)$ is assumed to be increasing, concave, and twice continuously differentiable.

4.2. Informal firms

Informal firms can only hire informal workers. They do not pay (labor and production) taxes but might be caught by the government with some probability. This is modeled as a labor distortion $\tau_i(l)$. In case they are caught, they have to pay a fine proportional to their revenue. The probability is increasing and convex in the size of the firm (larger firms have a higher probability of being caught): $\tau_i', \tau_i'' > 0$. The profit function for informal firms can be written as:

$$\Pi_i(\theta, w) = \max_l \{\theta q(l) - \tau_i(l) w\}$$

where the price of the final good is normalized to 1. w denotes the wage. One important finding of Ulyssea (2018) is that similar formal and informal workers (similar in observables as education) have the same wage inside formal firms. This supports the hypothesis that informal workers have the same wage as formal workers, conditioning on their skills. Therefore, the wage w is assumed to be the same for all workers, wherever they work and whatever is their employment status.[11]

4.3. Formal firms

Formal firms can hire formal (l_f) and informal (l_i) workers. Workers are assumed to be homogeneous, but the firms face different marginal prices of hiring them due to taxes and enforcement. They have to pay a tax on their revenue τ_y and payroll taxes τ_w on formal workers but no taxes on informal workers. Nevertheless, by hiring informal workers, they face a probability of being caught or audited, which is increasing in the number of informal workers they hire. Therefore, the cost of hiring informal workers is given by $\tau_f(l_i)$ which is increasing and convex in the number of informal workers. Furthermore, the productivity depends on the firm's district speed of the courts b, the higher the speed the better the contract enforcement, access to credit, etc for formal firms. This is represented

[11] One can introduce worker heterogeneity, for instance high skilled and low skilled workers with different wages, independent from the employment status.

by $\eta(b)$, a positive, decreasing function (e.g $\eta(b) = 1 + \frac{1}{b}$)

$$\Pi_f(\theta, w, b) = \max_l \{(1 - \tau_y)\eta(b)\theta q(l) - C(l)\}$$

The cost for formal firms to hire informal workers is given by $\tau_f(l_i)w$ and the cost of hiring formal workers by $(1 + \tau_w)l_f\lambda(b)w$. Since the marginal cost for informal workers is increasing in l_i and the marginal cost for formal workers is constant in l_f, there exists a unique $\tilde{l}$ such that up to this $\tilde{l}$, a formal firm only hires informal workers and above $\tilde{l}$ only formal workers. Furthermore, with a better judicial system, firms might face higher risks of being sued by workers, for instance if they do not respect the labor code. At the same time, with a better judiciary firms might have more leverage to push formal workers to fulfill their contract. These opposite effects are captured through a wedge in the labor cost for formal workers, $\lambda(b)$. $\lambda(b)$ is assumed to be a positive function of the average age of pending cases. It remains an empirical question whether $\lambda(b)$ is greater or smaller than 1 (i.e. who is benefiting from a faster judiciary on the intensive margin.) Therefore, the total labor cost can be written as:

$$C(l) = \begin{cases} \tau_f(l)w & \text{for } l \leq \tilde{l} \\ \left[\tau_f(\tilde{l}) + (1 + \tau_w)(l - \tilde{l})\lambda(b)\right] w & \text{for } l > \tilde{l}. \end{cases}$$

Lastly, firms have to pay a fix cost of operation: $\bar{c}_s$ for $s = f, i$. Therefore, the profits for a firm in sector s net of fixed costs are given by:

$$\pi_s = \Pi_s(\theta, w, b) - \bar{c}_s.$$

4.4. Entry, exit and dynamics

All firms in sector $s = i, f$ have a probability of an exogenous death shock, δ_s. Once, they have observed θ, it stays constant. Firms are agnostic about a potential change in b. Therefore, the value function of a firm with productivity θ, which observes wage w and court speed b can be simply written as:

$$V_s(\theta, w, b) = \max\left\{0, \frac{\pi_s(\theta, w, b)}{\delta_s}\right\}$$

At each period, there is a mass of M potential entrants. Each potential entrant observes a noisy signal $v \sim G$ before entry. v is positively correlated with the firms productivity θ: after entry, a firms draws its productivity from a conditional cumulative distribution function $F(\theta|v)$. Therefore, the expected value of entry into sector $s = f, i$ for a potential

entrant with signal v is:

$$V_s^e(\theta, w, b) = \int V_s(\theta, w, b) dF(\theta|v)$$

Given v, each potential entrant decides either to enter and to become formal or informal or not to enter. If a firm enters, it has to pay an entry cost c_s^e for $s = f, i$. The entry cost to formal sectors is assumed to be larger for the formal than for the informal sector: $c_f^e > c_i^e$.

Formal firms are assumed to have only two choices: either to stay formal or to exit. Informal firms have the choice between staying informal, becoming formal and paying the difference between the entry costs of the formal and informal sectors ($c_f^e - c_i^e$) or to exit. After entry into sector s, a firms exits directly without producing if $\theta < \bar{\theta}_s$, where $\pi_s(\bar{\theta}_s, w) = 0$.

4.5. Households

A representative household supplies inelastically $\bar{L}$. The household has no disutility from working and gets all its utility from consumption. (The household consumes all its income, which is given by $w\bar{L} + \pi + T$. T are taxes which are transferred to the household.)

4.6. Equilibrium

Entry in the informal sector occurs if the net expected value of entering this sector is positive and larger than the expected value from entering the formal sector:

$$V_i^e(v, w, b) - c_i^e \geq \max\{0, V_f^e(v, w, b) - c_f^e\}$$

The entry condition for the formal sector is analog:

$$V_f^e(v, w, b) - c_f^e > \max\{0, V_i^e(v, w, b) - c_i^e\}$$

Let $\underline{v_s}$ denote the signal observed by the last firm to enter sector s. Then, the mass of entrants into the two sectors is respectively given by:

$$M_i = [G(\underline{v_f}) - G(\underline{v_i})]M \quad \text{and} \quad M_f = [1 - G(\underline{v_f})]M$$

A stationary competitive equilibrium is given by a set of wage, allocations, cutoffs and measures (w, L_s, $\underline{v_s}$, M_s, μ_s) for $s = i, f$, such that they remain constant over time and the following holds in every period:

1. Labor markets clear: $\bar{L} = L_i + L_f$

2. Zero profit cutoff conditions hold: $\theta \geq \bar{\theta}_s$, where $\pi_s(\bar{\theta}_s, w) = 0$

3. Entry conditions hold, with equality if M_s is positive:

$$V_i^e(\underline{v_i}, w, b) = c_i^e$$
$$V_f^e(\underline{v_f}, w, b) = V_i^e(\underline{v_f}, w, b) + (c_f^e - c_i^e)$$

4. Both sectors remain constant in size: $\mu_s = \frac{1-F(\bar{\theta}_s)}{\delta_s} M_s$

4.7. Aggregation and state revenues

The taxes paid by the firms and workers are direct revenues for the state or a district. These revenues are used for instance to pay for the judicial system. Now, it is interesting to check, how much revenues are impacted via taxes and how much it would cost to increase judicial efficiency. From there, we might be able to find an optimum (from a revenue neutral perspective). From the political economy side, it is also interesting to compare revenues at the district level with expenditure at the state or country level, which could explain why the central government might lack incentives of improving the judicial efficiency if the benefits of it are going to other agents.

4.8. Comparative statics / Predictions

What happens with a change in court efficiency? Depending on $\eta(b)$ and $\lambda(b)$ the model predicts different scenarios.

1. Case: $\eta(b)$ and $\lambda(b)$ are both constant in b:
A change in b has no impact on firms' behavior, nothing happens.

2. Case: $\eta(b)$ is constant $\lambda(b)$ increasing in b:
An increase in b from b to b' does not have a direct impact on the revenue of a formal firm. However, the cost of formal workers increases. Therefore the threshold $\tilde{l}(b)$ increases to $\tilde{l}(b')$. There is no direct effect on small formal firms with a firm size below the threshold $\tilde{l}(b)$. Larger formal firms hire less and less formal workers, therefore the average size of formal firms decreases and the average number of informal workers in large formal firms increases. At the same time, potential entrants have a lower expected value of becoming formal, therefore more firms choose to enter the informal sector. This leads to more firms being informal, the informal firms being larger and the overall effect on informal workers in formal firms being ambiguous. Since large formal firms hire less workers and $q(l)$ is a positive and concave function in the total number of workers, the revenue per worker for large firms is expected to increase.

3. Case: $\eta(b)$ is constant $\lambda(b)$ decreasing in b: An increase in b leads to lower costs for formal workers. This has not direct impact on revenues. It implies a decrease in $\tilde{(l)}$.

Large formal firms hire more and more formal workers. Small formal incumbents are not impacted. As a secondary effect, entrants see an increase in V_f^e, therefore more firms decide to enter the formal sector, this leads to more formal firms. The effect on informal workers in formal firms is nonetheless ambiguous, the effect on informal workers in large formal firms negative.

4. Case: $\eta(b)$ is decreasing in b $\lambda(b)$ constant: An increase in b decreases a formal firm's revenue. Therefore, the optimal number of workers employed by a formal firm decreases and profits decrease. At the margin, small firms prefer to enter the informal sector. This increases the share of firms being informal. The share of informal workers increases as well, since all informal firms hire only informally, and the remaining formal firms hire less, and therefore less formal workers. The effect on the share of informal workers in formal firms is unclear. Since small formal firms which hired only informally become informal, this decreases the share of informal workers among the remaining formal firms. At the same time the number of workers hired by these remaining formal firms decreases and therefore the share of formal workers among them as well.

5. Case: $\eta(b)$ is decreasing in b $\lambda(b)$ increasing in b: An increase in b leads to lower revenue and higher cost of formal workers. Therefore, $\tilde{l}$ increases, the optimal number of workers employed by formal firms decreases and the profit of formal firms decreases. The share of firms being informal increases and for large formal firms the share of informal workers increases. The total share of informal workers increases.

6. Case: $\eta(b)$ is decreasing in b $\lambda(b)$ decreasing in b: This is the most complicated case. An increase in b leads to lower revenues for formal firms and to lower costs of formal workers. The later implies that $\tilde{l}$ decreases. At the margin, small firms prefer to be informal. Therefore, the share of firms being informal increases. Large formal firms will hire less workers in general due to lower revenues and less informal workers due to a lower $\tilde{l}$. In total, it is unclear what would be the expected effect on the shares of informal workers in total, of informal workers in formal firms, of informal workers in large formal firms and of firms being formal.

Table 2 summarizes the expected effects in the six scenarios. The table shows that although some effects are ambiguous, in theory, the signs of $\eta'(b)$ and $\lambda'(b)$ can be identified from the data. This is the main objective of the next section.

5. Empirical Analysis

This section tries to investigate the two fundamental questions of the impact of court efficiency on informality. How do firms react to changes, on both the extensive and intensive margin? Furthermore, using different specifications, the functional forms of the two main parameters from the model can be tested.

The empirical approach uses a cross sectional analysis, using data on informality and

firms from 2009 and court efficiency from 2008. The standard regression is defined as

$$y_d = \alpha_r + \theta b_d + \gamma X_d + \epsilon_d \tag{1}$$

where y_d is the outcome of interest in district d in years 2009/10. This is either the share of informal firms of all firms of a certain size, the share of informal workers, the share of informal workers in formal firms of a certain size or the revenue per worker for formal firms. b_d is court efficiency in district d in 2008, where court efficiency is measured as either backlog, average age of pending cases, clearance rate or disposal time. X_d are additional district level controls, like the share of the population being literate. α_r are region fixed effects, controlling for unobservables common among all districts in a region.

While the outcome of interest is measured in 2009/10, court efficiency is taken from 2008. This is due to the potential set of information firms use to take their decisions. A firm does not know what the backlog of a given year will be, it can only observe b from last year and build expectations about today's b only based on last realized values.

5.1. OLS Results

For the extensive margin, Table 3 presents simple ordinary least squares regressions of the share of informal firms among all firms of a given size on the speed of courts. Here, firms are categorized in four different bins: firms with 1-2, 3-4, 5-10 and more than 10 workers. These simple regressions do not include any covariates and are run, for each firm size bin, separately for all four court efficiency indicators backlog, average age of pending cases, clearance rate and disposition time.

All four measures of court speed have a statistically significant effect on the two groups of small firms (with 1-2 and 3-4 workers). The regressions with average age of pending cases, clearance rate and disposition time as independent variables have expected signs, indicating that faster courts lead to more formal firms. For instant from Panel C, a one standard deviation change in the clearance rate leads to a roughly 8 percentage points decrease in the share of informal firms among very small firms with one or two workers. This effect is reversed for large firms with more than 10 workers, but for them the effect is smaller in magnitude and not statistically significant. In contrast with the three other measurements, Backlog has the opposite effect. While still statistically significant for smaller firms, this can be interpreted as backlog not being a good indicator of court efficiency.

For the intensive margin, OLS results are less clear. Column 1 of Table 4 shows regressions of the share of informal workers on the four court speed measures. As the clearance rate goes up, the share of informal workers goes down. This could be explained by more firms becoming formal. Therefore, it is more interesting to look only at the workers in formal firms. Column 2 shows that a higher clearance rate leads as well to

less informal workers in formal firms. This is especially astonishing since for firms which become formal at the margin, are, with a higher court efficiency, expected to be of relative low productivity and to be of small size. These small, low productive firms are likely to hire mainly informal workers. The negative effect on the share of informal workers in formal firms should come from incumbent, larger firms which decide to hire more formal workers.

For the intensive margin, the signs of the estimators are the same when including region fixed effects and controlling for the share of population being literate, the share of population being part of a scheduled caste and dummies for urban and semi-urban districts. However, the estimates in columns 2 and 4 of Table 4 are not statistically significant from zero.

5.2. IV Strategy

While the OLS estimates from Table 4 and Table 3 give an intuition on the correlation between informality and court efficiency, one needs to be cautious with a causal interpretation. The error term in Equation 1 can be correlated with the variable b_d. This endogeneity issue arises due to potential reverse causality. Districts with more formal firms and workers have more revenues, this gives the district potentially more budget for the judiciary. Although the organization of the lower judiciary lies with the high courts and not the districts, districts can have an influence on the speed of court, for instance via the infrastructure available in the district. More revenue allows districts as well to spend more money on enforcement which can have a direct influence on the number of new cases being filed.

Therefore, to be able to interpret the results as causal impacts of the judicial system on firm decisions, an instrumental variable approach is used. Rao (2020) shows that in Indian states, judges at districts courts rotate every one or two years. The assignment for the rotation is quasi random. Combined with a large number of vacant judge positions[12], this leads to a quasi-random movement of vacant positions. The lower judiciary is under the authority of the respective high court. These high courts recruit district judges, which will be placed in a lower court under the judiciary of the respective high court. These judges have to rotate every 1 or 2 years and (normally) only exit the judiciary of the high court if they are promoted to any high court or if they retire. For the rotation, every time, they are moved to a new district, where they have not been active before. The exogeneity assumption of the instrument would be violated if judges always get their preferred position and if at the same time the preferences of judges for district d are correlated with y_d. As pointed out by Rao (2020), the legal framework of the judiciary backs the exogeneity assumption as well: While high courts have organizational

[12]Around 20% of judge positions are vacant in lower courts in India

freedom over lower courts under their judiciary, the budget of the judiciary depends on the central government. Therefore, high courts cannot respond in the short term to low court efficiencies by opening new vacancies. As an additional test, starting from the judge assignment to district and sessions courts in 2004 and randomly assigning 2/3 of the judges to new positions every year until 2008, the distribution of the mean share of occupied courtrooms from 2004 to 2008 is statistically not different from the actual distribution.

Furthermore, it can be shown that the filling of a vacant position is highly predictive of the number of cases disposed of by a court. Therefore, the rotation of judges can be used as an instrument. Specifically, in this context, firms could consider the general environment and probably only react little or not at all to short term changes. Therefore, the instrument used is the average share of filled judge positions in a district over the last five years (2004 - 2008). Figure 5 contains four scatter plots between the instrument and the four measures of court efficiency and gives an overview over the variation and distribution of the instrument as well as the correlation with the court efficiency measures. Figure 6 plots the distribution of the instrumental variable (mean share of occupied courtrooms from 2004 to 2008) with the corresponding within district standard variation.

Since taking the means over five years is an arbitrary definition of the instrument, mainly motivated by capturing a long term trend in the judiciary, alternative instruments can be considered. In Figure D.4, the density of the original instrument is plotted together with densities for four alternative specifications: taking the mean over the last four years (2005 - 2008), six year (2003 - 2008), including 2009 and using the median instead of the mean over years 2004 - 2008. All four specifications based on the mean have similar densities with a peak between 10 and 25%. Only the median measure has more weight close to zero but stays comparable to the other measures.

Table E.2 depicts the correlation between the different specifications of the instrumental variable and the four measures of court efficiency. The correlation varies only slightly between the instrumental variables. However, between the court efficiency measures, the correlation changes largely. Backlog and average age of pending cases are only slightly positively correlated with the instruments. Disposition time is negatively correlated, and clearance rate has the highest, positive, correlation. Altogether, using the instrumental variable and clearance rate as endogenous court efficiency measure seems promising from a standpoint of strength of the instrument.

Table 5 includes multiple first stage regressions. These are OLS regressions of the potentially endogenous variable Clearance Rate on the instrumental variable courtrooms and other control variables. The statistical significance of the instrument is high across all specifications. The F statistic is always at least above 20. Table 6 uses one of the specifications and alters the instrument. The F statistic and adjusted R-squared remain similar whichever instrumental variables is used. The instrument is always positive and

statistically significant. Lastly, Figure 7 includes the partial regression and partial residual plots of the mean share of occupied courtrooms 2004 - 2008 while including the share of the population being literate and region fixed effects.

5.3. 2SLS Results

Before analyzing the 2SLS results in detail, Table 7 compares different specifications for regressing the share of informal firms among firms with 1-2 workers on clearance rate and several covariates. The estimated effect of court speed on informality is robust to the different specifications. The impact is always negative and, when including region-fixed effects statistically significant. For the specifications with state-fixed effects, the estimated effect is larger in magnitude, not or only slightly statistically significant, and has a relatively low first-stage F statistic. The explained variation is around the same, using region- or state-fixed effects. Therefore, in the following, the specification of column 7 is taken as standard. It includes region-fixed effects and controls for the share of the population which is literate, the share of the population being of a scheduled cast, and if the district is mainly rural, urban or mixed.

Table 8 shows the results of 2SLS estimations of the share of informal firms among all firms of a given size on court efficiency controlling for regional fixed effects and the above-mentioned covariates. In line with the OLS results, there is a significant negative impact of the clearance rate on the share of informal firms among the smallest firms in India. For larger firms, estimates are as well negative but not statistically significant. Overall, there is clear evidence that faster courts lead to a lower share of informality among firms, an effect mainly driven by very small firms.

This implies, that cases 1 ($\eta'(b) = \lambda'(b) = 0$), 3 ($\eta'(b) = 0$, $\lambda'(b) < 0$) and 4 ($\eta'(b) < 0$, $\lambda'(b) = 0$) from the comparative statics section are excluded.

On the worker's side, with the given data, the results are less clear. The 2SLS estimates in Table 9 have the same negative signs as the original OLS estimates. Faster courts seem to lead to fewer informal workers in general and fewer informal workers in formal firms. Both estimates are not statistically different from zero. This can be due to the small sample size of only 190 districts and, for the share of informal workers in formal firms, to the low explanatory power of the chosen covariates.

Table 10 looks at the effect of clearance rate on the share of informal firms in formal firms, split by size of the firms. To test the structural form of the theory, the last column with large firms (more than 20 workers) is the most interesting. The estimated effect is negative but, again, not statistically different from zero.

Although the effects are not statistically significant, the negative sign of the estimate in column four of Table 10 is in line with the predictions of cases 2, 5, and potentially 6 from the comparative statics section.

To distinguish between these different cases, Table 11 and Table 12 use respectively gross sales value per worker and ex-factory value of manufactured goods per worker in formal firms as dependent variable. In both tables, column 1 includes all formal firms. In both regressions is the estimated effect of court efficiency on the outcome negative albeit not significant. Therefore, one can interpret this with some caution as evidence for case 2 from the comparative statics section.

This is a surprising result as it implies that there is no direct effect on the production process, as suggested by the literature. Rather, court efficiency impacts firms via a cost on formal workers. Furthermore, given the model, this implies that in the bargaining process between workers and firms, firms are benefiting more from more efficient courts than workers.

6. Conclusion

This paper investigates the link between court efficiency and informality. Both court efficiency and informality can be found in many developing countries. The paper adds court efficiency to the model from Ulyssea (2018) to give a theoretical background on how courts can influence firm decisions to become formal or informal and to hire formal and informal workers. The impact of court speed on these two margins of informality is then investigated in the context of India. Court efficiency is approximated via four different measurements from which the clearance rate is used for the main estimations. Using 2SLS to address the endogeneity of reversed causality, the findings promote a strong negative link between court efficiency and the share of firms being informal. On the intensive margin of informality, the results suggest a negative impact of court efficiency on the share of informal workers as well as the share of workers in formal firms being informal. This second set of estimates needs to be investigated further using better data.

The results allow identifying the sign and shape of the two main parameters in the theoretical model. Given the implied first derivatives, the model then states that there is no direct effect on the production process and that courts are potentially inclined to back the employers' cases.

References

Abraham, Rosa (2018). "Informal Workers in Formal and Informal Enterprises in India: A Comparative Analysis". In: *accessed January* 20.

Acemoglu, Daron and Simon Johnson (2005). "Unbundling institutions". In: *Journal of political Economy* 113.5, pp. 949–995.

Amirapu, Amrit (2017). *Justice delayed is growth denied: The effect of slow courts on relationship-specific industries in India*. Tech. rep. School of Economics Discussion Papers.

Assenova, Valentina A and Olav Sorenson (2017). "Legitimacy and the benefits of firm formalization". In: *Organization Science* 28.5, pp. 804–818.

Bertrand, Marianne, Chang-Tai Hsieh, and Nick Tsivanidis (2017). *Contract labor and firm growth in india*. Tech. rep. Mimeo.

Besley, Timothy and Robin Burgess (2004). "Can labor regulation hinder economic performance? Evidence from India". In: *The Quarterly journal of economics* 119.1, pp. 91–134.

Boehm, Johannes and Ezra Oberfield (2018). *Misallocation in the Market for Inputs: Enforcement and the Organization of Production*. Tech. rep. National Bureau of Economic Research.

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

European Commission (2020). "The 2020 EU Justice Scoreboard". In: *COM (2020), 306 final*.

Friedman, Eric et al. (2000). "Dodging the grabbing hand: the determinants of unofficial activity in 69 countries". In: *Journal of public economics* 76.3, pp. 459–493.

Glanfield, Laurie, Dan Hall, and Ingo Keilitz (2018). *Global Measures of Court Performance*.

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

Johnson, Simon et al. (2000). "Why do firms hide? Bribes and unofficial activity after communism". In: *Journal of Public Economics* 76.3, pp. 495–520.

Klein, Benjamin, Robert G Crawford, and Armen A Alchian (1978). "Vertical integration, appropriable rents, and the competitive contracting process". In: *The journal of Law and Economics* 21.2, pp. 297–326.

Lilienfeld-Toal, Ulf von, Dilip Mookherjee, and Sujata Visaria (2012). "The distributive impact of reforms in credit enforcement: Evidence from Indian debt recovery tribunals". In: *Econometrica* 80.2, pp. 497–558.

Loayza, Norman (2018). "Informality: Why Is It So Widespread and How Can It Be Reduced?" In: *World Bank Research and Policy Briefs* 133110.

Martin, Leslie A, Shanthi Nataraj, and Ann E Harrison (2017). "In with the big, out with the small: Removing small-scale reservations in India". In: *American Economic Review* 107.2, pp. 354–86.

Naidu, Suresh and Noam Yuchtman (2013). "Coercive contract enforcement: law and the labor market in nineteenth century industrial Britain". In: *American Economic Review* 103.1, pp. 107–44.

Rao, Manaswini (2020). "Judges, Lenders, and the Bottom Line: Court-ing Firm Growth in India". Job Market Paper.

Saikia, Dilip (2011). "Unorganised manufacturing industries in India: A regional perspective". In: *African Journal of Marketing Management* 3.8, pp. 195–206.

Shapiro, Alan Finkelstein (2015). "Institutions, informal labor markets, and business cycle volatility". In: *economia*, pp. 77–112.

Ulyssea, Gabriel (2010). "Regulation of entry, labor market institutions and the informal sector". In: *Journal of Development Economics* 91.1, pp. 87–99.

– (2018). "Firms, informality, and development: Theory and evidence from Brazil". In: *American Economic Review* 108.8, pp. 2015–47.

Visaria, Sujata (2009). "Legal reform and loan repayment: The microeconomic impact of debt recovery tribunals in India". In: *American Economic Journal: Applied Economics* 1.3, pp. 59–81.

7. Graphs

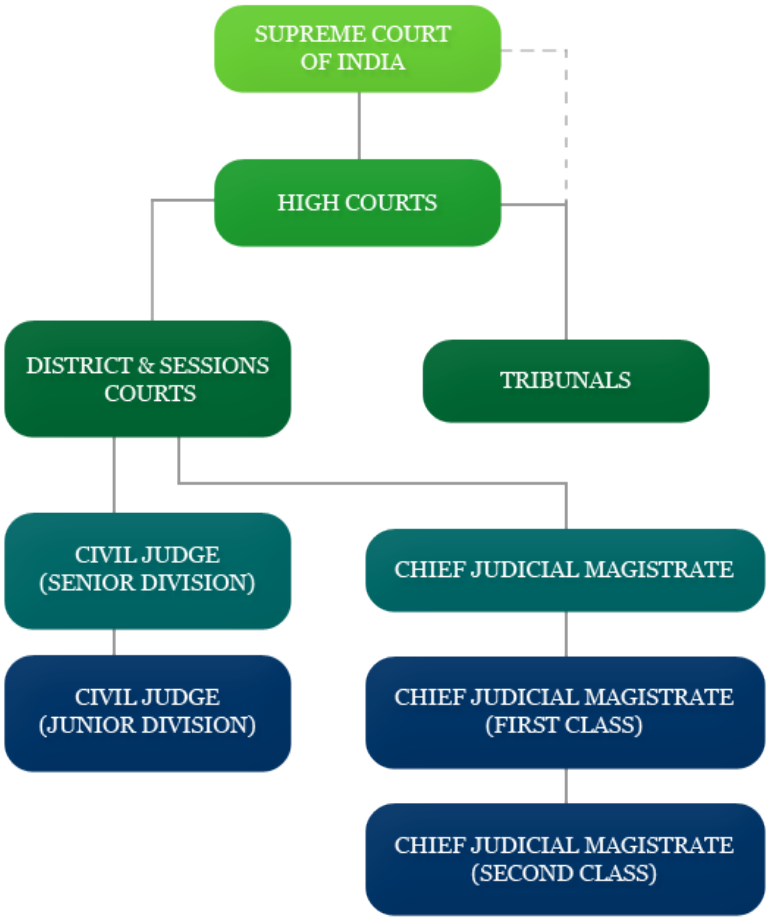


Figure 1: Overview over Judical Layers in India. Source: Daksh.

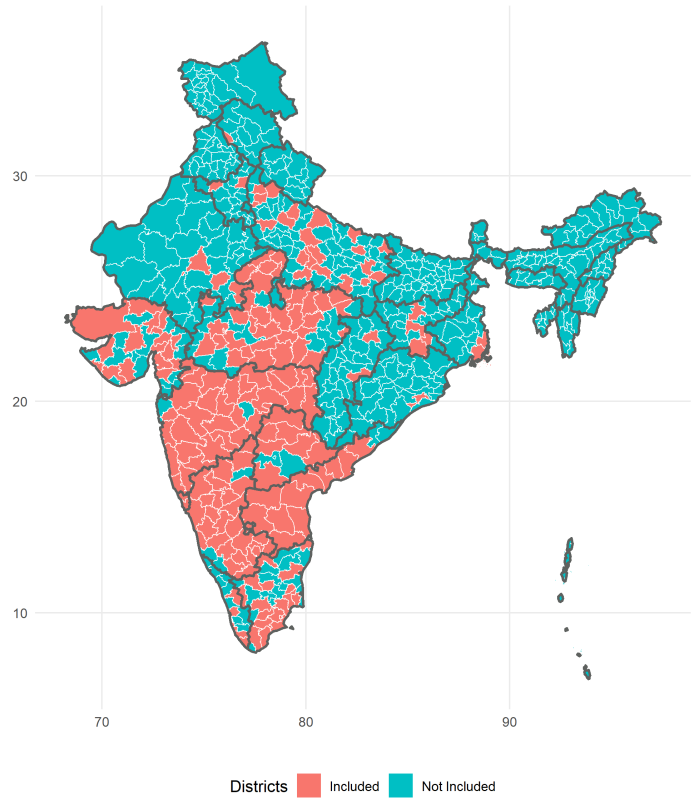
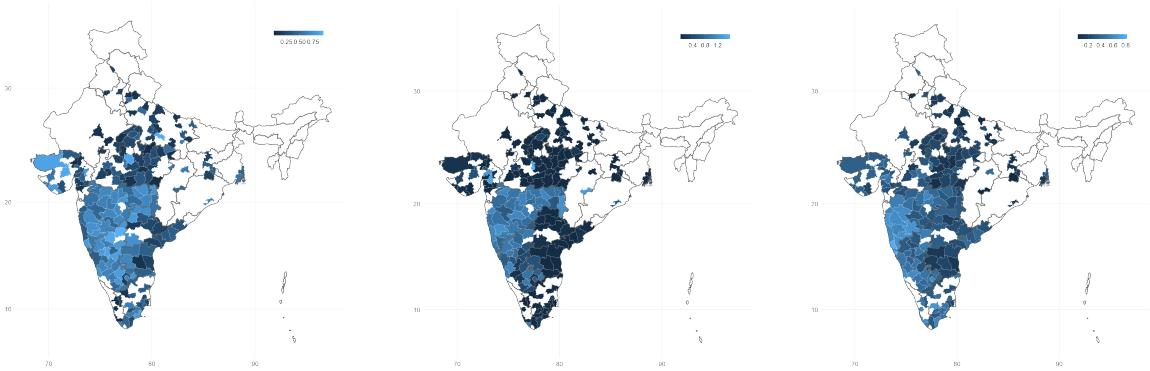
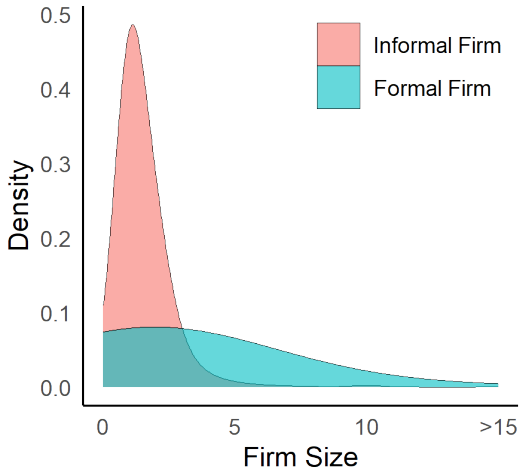


Figure 2: Map of India showing which districts are included in the analysis.

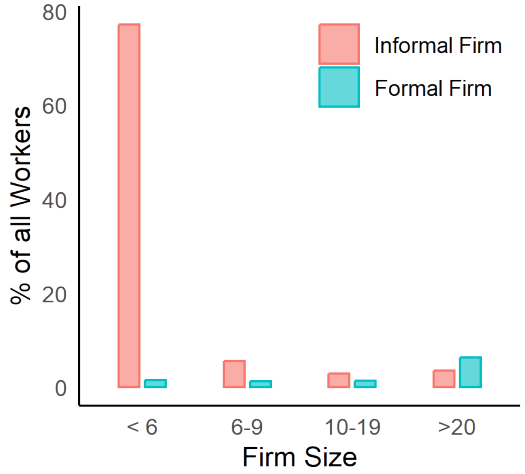


(a) Instrumental Variable (b) Clearance Rate (c) % Formal Firms

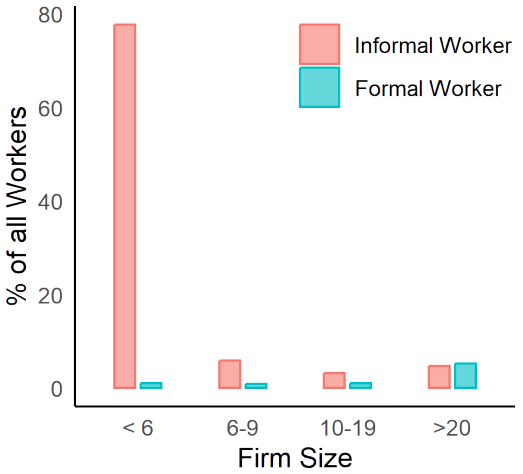
Figure 3



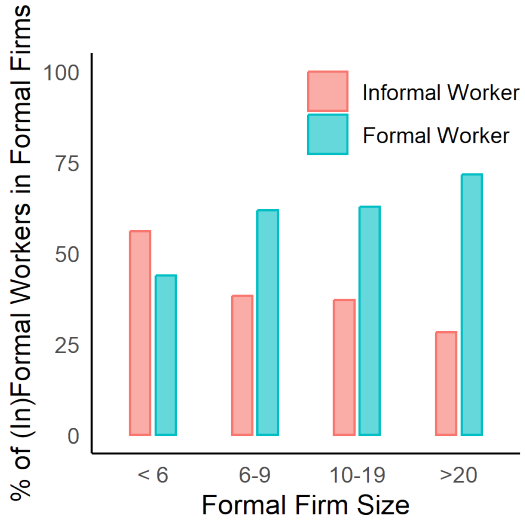
(a) Density of Firms by Firms' Informality Status and Size



(b) Share of Total Workforce by Firms' Size and Formality Status



(c) Share of Total Workforce by Workers' Formality Status and Firms' Size



(d) Share of Worker by Informality Status by Firms' Size

Figure 4: Descriptive Graphs on formal and informal firms and workers in India.

Figure 5: Scatter plots between the instrumental variable (Average share of filled court-rooms in district and sessions courts per district between 2004 and 2008) and the number of pending cases (top left), average age of pending cases (top right), clearance rate (bottom left) and disposition time (bottom right) in 2008.

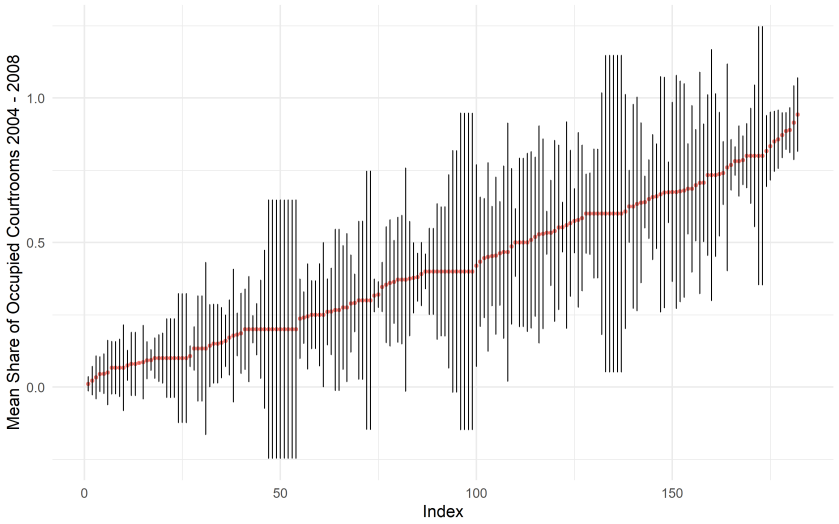


Figure 6: Distribution of instrumental variable. Red points are the mean share of occupied courtrooms in a district from 2004 to 2008. The gray bars represent the within district standard deviation of the share of occupied courtrooms in those five years.

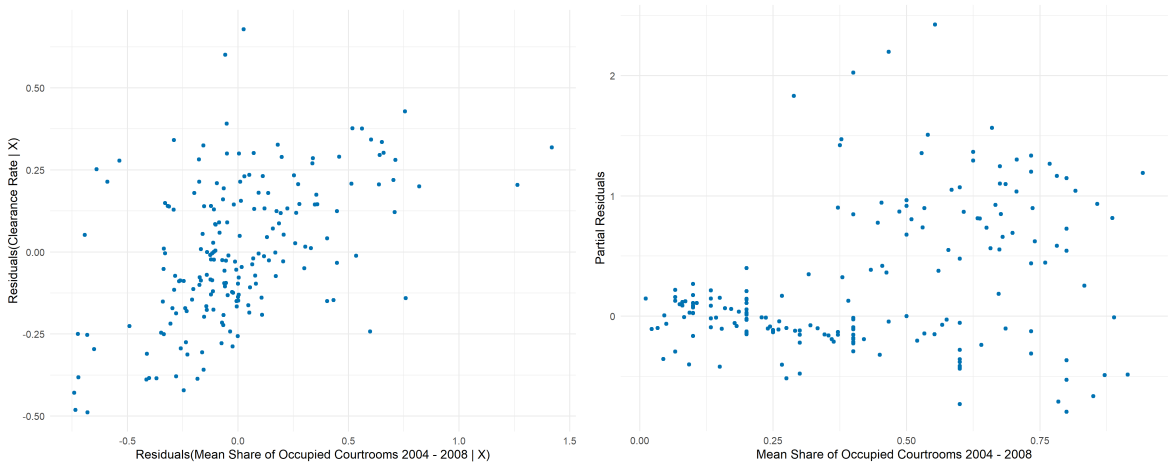


Figure 7: Partial regression (left panel) and partial residual (right panel) plots of the Mean Share of Occupied Courtrooms 2004 - 2008 in the first stage regression $Y \sim C + X$ where Y is the clearance rate per district in 2008, C is the Mean Share of Occupied Courtrooms 2004 - 2008 and the other covariates X are the share of the district population being literate and region fixed effects.

8. Tables

Table 1: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
Informal share of firms total	500	0.73	0.18	0.18	1.00
Informal share of firms of size 1-2	500	0.77	0.17	0.20	1.00
Informal share of firms of size 3-4	500	0.55	0.26	0.00	1.00
Informal share of firms of size 5-10	498	0.37	0.26	0.00	1.00
Informal share of firms of size >10	492	0.16	0.24	0.00	0.99
Share of inf. workers	497	0.92	0.07	0.45	1.00
Share of inf. workers in form. firms	497	0.36	0.22	0.00	1.00
Share of inf. workers in form. firms with < 6 workers	463	0.47	0.36	0.00	1.00
Share of inf. workers in form. firms with 6 - 9 workers	439	0.30	0.34	0.00	1.00
Share of inf. workers in form. firms with 10 - 20 workers	429	0.30	0.36	0.00	1.00
Share of inf. workers in form. firms with > 20 workers	465	0.27	0.29	0.00	1.00
Gross Sales Value in Form Firms / L (in million Rs)	474	2.28	3.65	0.02	71.41
Ex Factory Value of manufactured goods / L (in million Rs)	474	2.22	3.63	0.02	71.28
Backlog (in thousand)	420	3.32	9.74	0.001	157.58
Average Age of Pending Cases (years)	420	3.14	1.59	0.22	11.38
Clearance Rate	189	0.32	0.43	0.001	1.59
Disposition Time	192	1.63	2.89	0.001	15.08
Mean Occupancy in Courts 2004 - 2008	258	0.36	0.24	0.01	0.94
Population (in Lakh)	500	22.22	15.04	0.84	110.60
Share of Pop. SC	500	0.17	0.08	0.0005	0.50
Share of Pop. Litterate	500	0.62	0.10	0.34	0.89

Table 2: Comparative Statistics

	$\eta'(b)=0$ $\lambda'(b)=0$	$\eta'(b)=0$ $\lambda'(b)>0$	$\eta'(b)=0$ $\lambda'(b)<0$	$\eta'(b)<0$ $\lambda'(b)=0$	$\eta'(b)<0$ $\lambda'(b)>0$	$\eta'(b)<0$ $\lambda'(b)<0$
Sh. of inf. workers	.	+	-	-	+	?
Sh. of inf. workers in form. firms	.	?	?	?	?	?
Sh. of inf. w. in large form. firms	.	+	-	+	+	?
Sh. of firms being informal	.	+	-	-	+	?
$\tilde{l}$	.	+	-	.	+	-
Revenue / worker	.	+	.	-	-	-
Ex Factory Value of goods / worker	.	+	.	-	-	-

Note: This tables shows how different a positive change in b influences several moments of the data, depending on the functional form of $\eta(b)$ and $\lambda(b)$.

Table 3: OLS Regression of Firms' Formality Status on Efficiency Indicators by Firm Size

	Share of form. Firms of all Firms of Size				
	(1)	(2)	(3)	(4)	(5)
	All	$\leq$ 2 Workers	$>$ 2 Workers	$\leq$ 10 Workers	$>$ 10 Workers
Panel A					
Clearance Rate	0.108**	0.105*	0.0830*	0.110**	-0.0873
	[0.0287,0.187]	[0.0210,0.190]	[0.0123,0.154]	[0.0313,0.189]	[-0.245,0.0700]
Indep. Var. Mean	.29	.29	.32	.29	.35
Adj. R2	0.08	0.08	0.07	0.08	0.04
N	142,528	91,590	50,938	122,920	19,608
Panel B					
Avg. Age	-0.0139**	-0.0149**	-0.0115	-0.0138**	-0.0213
	[-0.0230,-0.00484]	[-0.0243,-0.00551]	[-0.0255,0.00255]	[-0.0228,-0.00478]	[-0.0552,0.0127]
Indep. Var. Mean	3.6	3.6	3.7	3.6	3.6
Adj. R2	0.06	0.05	0.08	0.06	0.04
N	281,552	187,173	94,379	247,124	34,428
Panel C					
Disposition Time	-3.459***	-2.968**	-4.268**	-3.482***	5.699
	[-5.336,-1.582]	[-4.911,-1.026]	[-7.023,-1.513]	[-5.358,-1.605]	[-3.061,14.46]
Indep. Var. Mean	.005	.005	.0049	.005	.0035
Adj. R2	0.06	0.06	0.06	0.06	0.02
N	146,535	93,614	52,921	125,874	20,661
Panel D					
Backlog	0.992*	0.837	0.838**	1.026*	0.294
	[0.213,1.770]	[-0.137,1.812]	[0.212,1.464]	[0.234,1.818]	[-0.462,1.050]
Indep. Var. Mean	.0069	.0065	.0093	.0069	.0087
Adj. R2	0.06	0.05	0.08	0.06	0.03
N	281,552	187,173	94,379	247,124	34,428
Region FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Dep. Var. Mean	0.30	0.25	0.55	0.29	0.78

Note: Backlog denotes the number of pending cases at the end of the civil year. Avg Age is the average age of all pending cases at the end of a year. Clearance Rate denotes the number of resolves cases divided by the number of incoming cases in a year. Disposition time is defined as the number of unresolved cases divided by the number of resolved cases in a given year. Stars *, ** and *** indicate that the p-value is below 0.1, 0.05 and 0.01 respectively.

Table 4: OLS Intensive Margin

	Share of Inf. Workers		Share of Inf Workers in Form Firms	
	(1)	(2)	(3)	(4)
Panel A				
Backlog	-0.0013^{***}	-0.00011	-0.0014	$-5.9e-05$
	(0.000343)	(0.000337)	(0.00109)	(0.00121)
Independent Var. s.d	9.771	9.771	9.771	9.771
Region FE		×		×
Covariates		×		×
Observations	417	417	417	417
Panel B				
Avg. Age	0.0015	0.0028	-0.0013	-0.0047
	(0.00214)	(0.00192)	(0.00671)	(0.00688)
Independent Var. s.d	1.592	1.592	1.592	1.592
Region FE		×		×
Covariates		×		×
Observations	417	417	417	417
Panel C				
Clearance Rate	-0.039^{**}	-0.013	-0.096^{**}	-0.062
	(0.0128)	(0.0134)	(0.0353)	(0.0444)
Independent Var. s.d	0.434	0.434	0.434	0.434
Region FE		×		×
Covariates		×		×
Observations	187	187	187	187
Panel D				
Disposition Time	0.0018	0.0028	0.0037	0.0014
	(0.00193)	(0.00179)	(0.00531)	(0.00591)
Independent Var. s.d	2.9	2.9	2.9	2.9
Region FE		×		×
Covariates		×		×
Observations	190	190	190	190

Note: Backlog denotes the number of pending cases at the end of the civil year. Avg Age is the average age of all pending cases at the end of a year. Clearance Rate denotes the number of resolves cases divided by the number of incoming cases in a year. Disposition time is defined as the number of unresolved cases divided by the number of resolved cases in a given year. Columns two and four include region fixed effects and control for the share of the population being literate, the share of the population being of a scheduled cast and if the district is mainly rural, urban or mixed. Stars *, ** and *** indicate that the p-value is below 0.1, 0.05 and 0.01 respectively.

Table 5: First Stage Regression

	Clearance Rate									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Courtrooms	1.011***	1.011***	1.011***	1.014***	0.992***	1.013***	0.998***	0.926***	0.809***	0.370***
	(0.105)	(0.105)	(0.098)	(0.098)	(0.105)	(0.100)	(0.100)	(0.102)	(0.138)	(0.119)
Constant	−0.098*	−0.098*	−0.098***	−0.023	−0.043	−0.033	0.201	−0.738***	−0.619***	−0.224
	(0.051)	(0.051)	(0.031)	(0.106)	(0.158)	(0.140)	(0.123)	(0.238)	(0.238)	(0.206)
Clustered SE		×	×	×	×	×	×	×	×	×
Robust SE			×	×	×	×	×	×	×	×
Rural FE				×	×	×	×	×	×	×
Bank Credits					×					
Population						×				
Share SC							×	×	×	×
Share Literate								×	×	×
Region FE									×	
State FE										×
F	92.56	92.56	106.33	35.45	22.38	26.81	29.77	38.97	27.34	
Observations	189	189	189	187	170	187	187	187	187	187
Adjusted R^2	0.328	0.328	0.328	0.322	0.329	0.318	0.380	0.434	0.495	0.669

Note: The dependent variable Clearance Rate is the number of resolved cases divided by the number of incoming cases in a year. Courtrooms is the average over the preceding five years of the shares of court rooms filled in District and Sessions courts. Rural is a dummy variable equal to one if the district is mainly rural. Half rural is a dummy variable equal to one if the district can not clearly be identified as rural or urban. Urban is omitted. Bank Credits denotes the outstanding credit of scheduled commercial banks in a district in 2009. Population is the population in a district in 2011. Share SC denotes the share of the population being part of a scheduled cast in 2011. Share Literate is the share of the population being literate in 2011. Clustered SE indicates that standard errors are clustered at the state level. Robust SE indicates heteroscedasticity robust standard errors. Region FE and State FE are dummy variables for each region and state respectively. Stars *, ** and *** indicate that the p-value is below 0.1, 0.05 and 0.01 respectively.

Table 6: First Stage Regression Robustness to Different Instrument Specifications

	Clearance Rate				
	(1)	(2)	(3)	(4)	(5)
Courtroom Occupancy Mean 2004 - 08	0.809***				
	(0.138)				
Courtroom Occupancy Mean 2005 - 08		0.789***			
		(0.126)			
Courtroom Occupancy Mean 2003 - 08			0.842***		
			(0.152)		
Courtroom Occupancy Mean 2004 - 09				0.868***	
				(0.138)	
Courtroom Occupancy Median 2004 - 08					0.549***
					(0.111)
Constant	−0.619***	−0.582**	−0.665***	−0.675***	−0.523**
	(0.238)	(0.232)	(0.239)	(0.232)	(0.249)
Region FE	×	×	×	×	×
Covariates	×	×	×	×	×
F	27.34	29.02	25.94	30.97	27.72
Observations	187	187	187	187	187
Adjusted R^2	0.495	0.511	0.493	0.511	0.463

Note: The dependent variable Clearance Rate is the number of resolved cases divided by the number of incoming cases in a year. The different courtroom occupancy measures calculate the mean or median of the share of filled courtrooms in District and Sessions courts over the indicated years. Share Literate is the share of the population being literate in 2011. Region FE includes dummy variables for regions. Covariates include the share of the population being literate, the share of the population being of a scheduled cast and dummies if the district is mainly rural, urban or mixed. Standard errors are robust to heteroscedasticity and clustered at the state level. Stars *, ** and *** indicate that the p-value is below 0.1, 0.05 and 0.01 respectively.

Table 7: 2SLS Regressions with different Covariates

	Share of informal firms among firms with 1-2 workers									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Clearance Rate	−0.264***	−0.110**	−0.175	−0.143***	−0.157***	−0.229*	−0.160***	−0.238*	−0.189***	−0.162***
	(0.042)	(0.054)		(0.048)	(0.047)	(0.128)	(0.050)	(0.134)	(0.054)	(0.049)
Constant	0.806***	0.840***	0.804	1.190***	1.136***	0.947***	1.180***	0.983***	1.218***	1.202***
	(0.017)	(0.017)		(0.076)	(0.075)	(0.131)	(0.095)	(0.150)	(0.109)	(0.095)
Share Literate				×	×	×	×	×	×	×
Share SC					×	×	×	×	×	×
Rural							×	×	×	×
Bank Credits									×	
Population										×
Region FE		×		×	×		×		×	×
State FE			×			×		×		
F	92.56	50.12	13.71	55.96	56.36	13.58	55.81	12.97	45.43	55.77
Observations	189	189	189	187	189	187	187	187	170	187
Adjusted R^2	0.170	0.449	0.491	0.504	0.500	0.447	0.493	0.433	0.487	0.491

Note: The dependent variable Clearance Rate is the number of resolved cases divided by the number of incoming cases in a year. Courtrooms is the average over the preceding five years of the shares of court rooms filled in District and Sessions courts. Rural is a dummy variable equal to one if the district is mainly rural. Half rural is a dummy variable equal to one if the district can not clearly be identified as rural or urban. Urban is omitted. Bank Credits denotes the outstanding credit of scheduled commercial banks in a district in 2009. Population is the population in a district in 2011. Share SC denotes the share of the population being part of a scheduled cast in 2011. Share Literate is the share of the population being literate in 2011. Region FE and State FE are dummy variables for each region and state respectively. Standard errors are robust to heteroscedasticity and clustered at the state level. Stars *, ** and *** indicate that the p-value is below 0.1, 0.05 and 0.01 respectively.

Table 8: 2SLS Regressions Extensive Margin with different Covariates

	Share of inf. Firms of all Firms of Size				
	(1)	(2)	(3)	(4)	(5)
	All	$\leq$ 2 Workers	$>$ 2 Workers	$\leq$ 10 Workers	$>$ 10 Workers
Panel A: OLS					
Clearance Rate	-0.108**	-0.105*	-0.0830*	-0.110**	0.0873
	[-0.187,-0.0287]	[-0.190,-0.0210]	[-0.154,-0.0123]	[-0.189,-0.0313]	[-0.0700,0.245]
Adj. R2	0.0220	0.0194	0.0295	0.0227	0.0207
Panel B: IV					
Clearance Rate	-0.163	-0.183	-0.133	-0.168	-0.0793
	[-0.268; -0.0573]	[-0.294; -0.0727]	[-0.259; -0.00706]	[-0.273; -0.0629]	[-0.282; 0.123]
First Stage F	325.37	291.58	290.09	321.04	112.86
Region FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
N	142,528	91,590	50,938	122,920	19,608

Note: The dependent variable Clearance Rate is the number of resolved cases divided by the number of incoming cases in a year. In the 2SLS regression, the average over the preceding five years of the shares of court rooms filled in District and Sessions courts is used as exogenous instrument. The regressions include region fixed effects and control for the share of the population being literate, the share of the population being of a scheduled cast and if the district is mainly rural, urban or mixed. Standard errors are robust to heteroscedasticity and clustered at the state level. Stars *, ** and *** indicate that the p-value is below 0.1, 0.05 and 0.01 respectively.

Table 9: 2SLS Regressions Intensive Margin with different Covariates

	Share of Form Workers	Share of Form Workers in Form Firms
	(1)	(2)
Panel A: OLS		
Clearance Rate	0.013	0.062
	(0.0134)	(0.0444)
Adj. R2	0.34	0.04
Panel B: 2SLS		
Clearance Rate	0.013	0.037
	(0.0252)	(0.0988)
First Stage F	55.81	55.81
Region FE	×	×
Controls	×	×
Observations	187	187

Note: The dependent variable Clearance Rate is the number of resolved cases divided by the number of incoming cases in a year. In the 2SLS regression, the average over the preceding five years of the shares of court rooms filled in District and Sessions courts is used as exogenous instrument. The regressions include region fixed effects and control for the share of the population being literate, the share of the population being of a scheduled cast and if the district is mainly rural, urban or mixed. Standard errors are robust to heteroscedasticity and clustered at the state level. Stars *, ** and *** indicate that the p-value is below 0.1, 0.05 and 0.01 respectively.

Table 10: 2SLS Regressions

	Share of inf. Workers in form. Firms of Size			
	< 6 Workers	$6-9$ Workers	$10-19$ Workers	> 20 Workers
	(1)	(2)	(3)	(4)
Panel A: OLS				
Clearance Rate	-0.08	-0.03	0.028	$-0.12^{\cdot}$
	(0.0794)	(0.0746)	(0.0749)	(0.0604)
Adj. R2	0	-0.02	0.04	0.01
Panel B: 2SLS				
Clearance Rate	-0.32^{*}	0.00058	0.31	-0.1
	(0.155)	(0.156)	(0.191)	(0.121)
Adj. R2	-0.05	-0.02	-0.05	0.01
First Stage F	63.41	51.3	47.28	54.51
Region FE	×	×	×	×
Controls	×	×	×	×
Observations	179	167	160	174

Note: The dependent variable Clearance Rate is the number of resolved cases divided by the number of incoming cases in a year. In the 2SLS regression, the average over the preceding five years of the shares of court rooms filled in District and Sessions courts is used as exogenous instrument. The regressions include region fixed effects and control for the share of the population being literate, the share of the population being of a scheduled cast and if the district is mainly rural, urban or mixed. Standard errors are robust to heteroscedasticity and clustered at the state level. Stars *, ** and *** indicate that the p-value is below 0.1, 0.05 and 0.01 respectively.

Table 11: 2SLS Regressions

	Gross Sales Value per Worker in Formal Firms of Size					
	Total	< 5 Workers	$5 - 10$ Workers	$11 - 20$ Workers	$21 - 50$ Workers	> 50 Workers
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: OLS						
Clearance Rate	−1.1	1.2	0.75·	−3.2	−0.34	−0.37
	(1.13)	(0.952)	(0.443)	(3.28)	(0.461)	(0.485)
Adj. R2	0.01	0.04	0.04	0.02	0.02	0.03
Panel B: 2SLS						
Clearance Rate	−3.2	3.6	0.7	−8.9	−0.49	−1.3
	(3.64)	(2.5)	(0.762)	(10)	(0.924)	(0.967)
Adj. R2	-0.01	-0.02	0.04	0	0.02	0.01
First Stage F	55.39	27.53	45.24	52.11	58.06	54.6
Region FE	×	×	×	×	×	×
Controls	×	×	×	×	×	×
Observations	184	111	160	154	161	179

Note: The dependent variable Clearance Rate is the number of resolved cases divided by the number of incoming cases in a year. In the 2SLS regression, the average over the preceding five years of the shares of court rooms filled in District and Sessions courts is used as exogenous instrument. The regressions include region fixed effects and control for the share of the population being literate, the share of the population being of a scheduled cast and if the district is mainly rural, urban or mixed. Standard errors are robust to heteroscedasticity and clustered at the state level. Stars *, ** and *** indicate that the p-value is below 0.1, 0.05 and 0.01 respectively.

Table 12: 2SLS Regressions

	Ex Factory Value of Manufactured Goods per Worker in Formal Firms of Size					
	Total	< 5 Workers	$5 - 10$ Workers	$11 - 20$ Workers	$21 - 50$ Workers	> 50 Workers
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: OLS						
Clearance Rate	−1.1	1.4	0.54	−3.5	−0.31	−0.34
	(1.13)	(0.953)	(0.455)	(3.27)	(0.527)	(0.447)
Adj. R2	0.02	0.05	0.03	0.03	0.01	0.05
Panel B: 2SLS						
Clearance Rate	−3.1	4	0.39	−9.4	0.18	−1.2
	(3.63)	(2.49)	(0.835)	(9.94)	(1.15)	(0.877)
Adj. R2	0	-0.02	0.03	0.01	0	0.03
First Stage F	55.39	27.53	45.24	52.11	58.06	54.6
Region FE	×	×	×	×	×	×
Controls	×	×	×	×	×	×
Observations	184	111	160	154	161	179

Note: The dependent variable Clearance Rate is the number of resolved cases divided by the number of incoming cases in a year. In the 2SLS regression, the average over the preceding five years of the shares of court rooms filled in District and Sessions courts is used as exogenous instrument. The regressions include region fixed effects and control for the share of the population being literate, the share of the population being of a scheduled cast and if the district is mainly rural, urban or mixed. Standard errors are robust to heteroscedasticity and clustered at the state level. Stars *, ** and *** indicate that the p-value is below 0.1, 0.05 and 0.01 respectively.

Appendix A Uniqueness of Threshold $\tilde{l}$

$\tilde{l}$ defines the threshold below which formal firms only hire informal firms and above which every additional worker will be formally hired. Economically, this is the point where the marginal cost of hiring one more informal worker becomes more expensive than the marginal cost of one more formal worker. Therefore, the threshold can be found by equalizing the marginal costs of informal and formal workers in a formal firm:

$$\underbrace{\tau_f'(\tilde{l})}_{\text{MC(informal worker)}} = \underbrace{(1+\tau_w)\lambda(b)}_{\text{MC(formal worker)}}$$

If $\tau_f(l)$ is a positive, stricly convex function, then $\tau_f'(l)$ is strictly monotonically increasing. This implies that an inverse $\left(\tau_f'\right)^{-1}(c)$ exists and it is strictly increasing in c where c is s.t $\tau_f'(l) = b$. Then we find that:

$$\tilde{l} = \left(\tau_f'\right)^{-1}((1+\tau_w)\lambda(b))$$

This means that $\tilde{l}$ is increasing in $(1+\tau_w)\lambda(b)$. Since $(1+\tau_w)$ is greater than 0, $\tilde{l}$ is increasing in $\lambda(b)$. $\lambda(b)$ is a positive but decreasing function in b, therefore $\tilde{l}$ decreases in $\lambda(b)$. The slower the judiciary, (the larger b), the higher $\tilde{l}$ and the more informal workers are hired in formal firms.

Note that although the threshold $\tilde{l}$ does not depend on the wage w, the shares of formal and informal workers in formal firms depend on the total number of workers hired and this can depend on the wage. For instance, if $\tilde{l} = 2$ and all firms are homogeneous, with a very high wage, all firms may only want to hire one worker, which then is informally hired. Then the share of informal workers in formal firms is equal to 1. On the other hand, with a very low wage (and the same $\tilde{l}$), firms would hire many workers, and all workers above 2 formally. This would lead to a low share of informal workers in formal firms.

Appendix B Uniqueness + Existence of Equilibrium

The mass of entrants is fixed by M and $\underline{\theta_s}$. Since $\underline{\theta_s}$ exists uniquely for $s = i, f$, the number of entrants is fixed. In a stationary economy, the number of firms per sector must stay constant over time, this implies that the number of entrants has to be equal to the number of exits per sector. Therefore, the number of entrants M_s and the probability of exit uniquely define the size of the sector μ_s.

For the wage w, there exists a unique wage s.t labor markets clear. The profit functions for both sectors a strictly decreasing in w. The optimal labor for the informal sector is

given by the first order condition:

$$l^{\star}(\theta, w, b) \text{ s.t. } \frac{\partial \Pi_s}{\partial l} = \frac{\partial[\theta q(l) - \tau_i(l)w]}{\partial l} = 0$$

This gives $\theta q'(l) - w\tau_i'(l) = 0$. $q(l)$ is an increasing and convex function in l. Therefore $q'(l)$ is a positive but decreasing function. $\tau_i(l)$ is increasing and concave and $\tau_i'(l)$ positive and increasing. This shows that $l^{\star}(\theta, w, b)$ is decreasing in w. (One easily verifies that the solution $l^{\star}(\theta, w, b)$ is a maximum, by checking the second order condition which is always negative.) The reasoning for the formal sector is similar, with the only difference that the marginal cost of labor is either positive and increasing or positive and constant. Together this shows that if the demand of labor exceeds its supply, the wage increases up to the unique point where the labor market clears. (The opposite is true if the demand falls short of the supply.) Therefore, there exists a unique wage w which clears the labor market.

Appendix C Data Appendix

C.1 Data cleaning

To have a uniform sample of districts and industries, several adjustments have been undertaken. On the district side, all north-eastern states (Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Naghaland, Tripura, Sikkim) are excluded. This is due to them being far away and not properly covered in some surveys. Additionally, all Union Territories (Andaman & Nicobar Islands, Chandigarh, Dadra and Nagar Haveli, Daman and Diu, Delhi, Lakshadweep, Puducherry) are left out. This leads to a total of 454 districts (out of a maximum of 617 districts.)

On the firm side, all public sector enterprises and cooperatives have been excluded (since SUNAE is not covering them.) Furthermore, only firms operating in the 4-digits National Industrial Classifications 2008 (NIC-2008) overlap of SUNAE, ASI and IEU surveys are considered, assuring that the different surveys do not consider different industries.

The states are put together in six different regions, defined as:

- Central: Madhya Pradesh, Chhattisgarh
- East: Bihar, West Bengal, Odisha, Jharkhand
- West: Gujarat, Maharashtra, Goa
- South: Karnataka, Tamil Nadu, Kerala, Andhra Pradesh, Telangana
- North: Rajasthan, Punjab, Haryana, Uttar Pradesh, Jammu and Kashmir, Himachal Pradesh, Uttarakhand, Delhi
- Northeast: Tripura, Meghalaya, Manipur, Assam, Arunachal Pradesh, Mizoram, Sikkim, Nagaland

Table C.1: Sample Selection Due to Intermediate Steps per Dataset in 2008 and 2013

Panel A: 2008

	Backlog	Avg Age	Clear. Rate	Disp. Time	IV	SUNAE	IEU	ASI
Initial	506	506	208	211	294	617	611	547
Remove NE + UT	454	454	193	196	266	511	507	483
Not in Dictionary	420	420	189	192	258	500	497	479

Panel B: 2013

	Backlog	Avg Age	Clear. Rate	Disp. Time	IV
Initial	523	523	493	495	505
Remove NE + UT	467	467	445	447	457
Not in Dictionary	428	428	415	417	

Note: Numbers are representing the number of districts. Backlog denotes the number of pending cases at the end of the civil year. Avg Age is the average age of all pending cases at the end of a year. Clearance Rate denotes the number of resolves cases divided by the number of incoming cases in a year. Disposition time is defined as the number of unresolved cases divided by the number of resolved cases in a given year. IV stands for instrumental variable and is the average or median over the preceding five years of the shares of court rooms filled in District and Sessions courts.

Appendix D Additional Figures

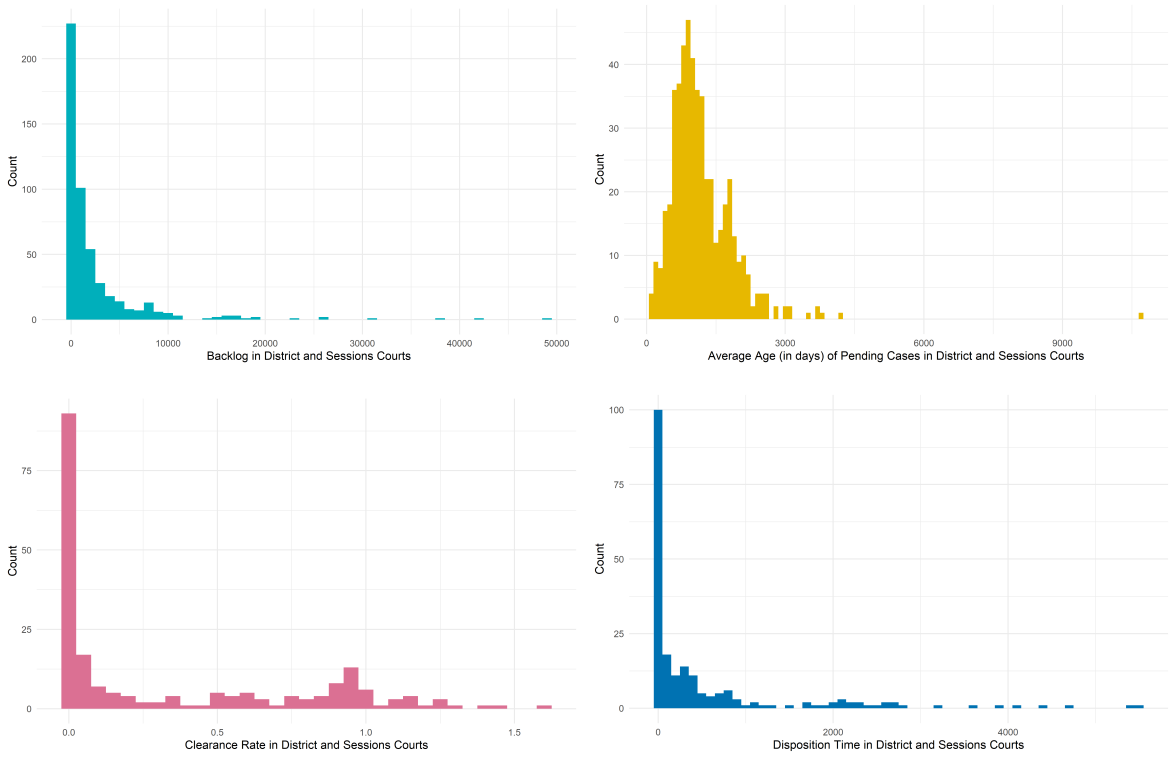


Figure D.1: Histograms of backlog (top left panel), average age of pending cases (top right panel), clearance rate (bottom left panel) and disposition time (bottom right panel) in district and sessions courts per district in 2008.

Figure D.2: Correlation between average age of pending cases in District and Session courts and share of informal firms (left panel), share of informal workers (center panel) and share of informal workers in formal firms (right panel) per district.

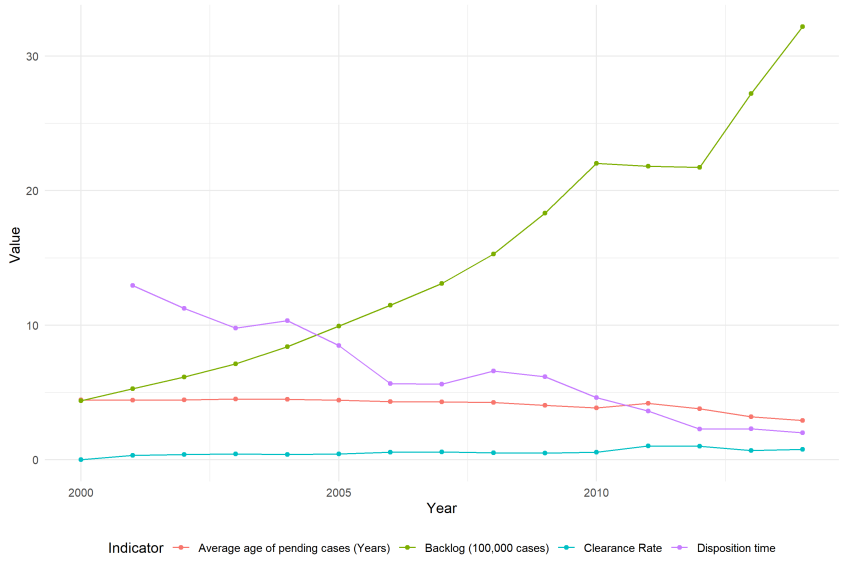


Figure D.3: Court Speed Indicators for District and Session courts in India 2000 - 2014

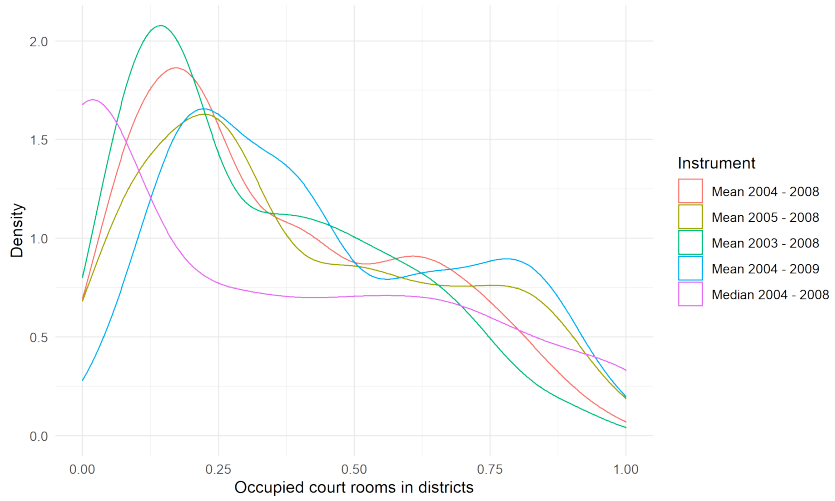


Figure D.4: Density plots of different specifications of the instrumental variable

Appendix E Additional Tables

Table E.1: Moments

Moment	Data Source
Share of formal firms which are of size	
$<$ 5 workers	ASI
5 - 10 workers	ASI
11 - 20 workers	ASI
21 - 50 workers	ASI
$>$ 50 workers	ASI
Share of informal firms which are of size	
$<$ 2 Workers	SUNAE
$<$ 5 Workers	SUNAE
Share of firms being informal	
1 - 2 workers	ASI + SUNAE
3 - 4 workers	ASI + SUNAE
5 - 10 workers	ASI + SUNAE
$>$ 10 workers	ASI + SUNAE
Share of inf. workers	
Total	IEU
In Formal Firms	IEU

Note:

Table E.2: Correlation between Court Efficiency Measures and Instrumental Variable 2008

	Backlog	Avg Age of Pending Cases	Clearance Rate	Disposition T
IV: Mean Share 2004-2008	0.14	0.05	0.56	-0.45
IV: Mean Share 2005-2008	0.13	0.03	0.60	-0.46
IV: Mean Share 2003-2008	0.15	0.07	0.53	-0.45
IV: Mean Share 2004-2009	0.15	0.04	0.63	-0.48
IV: Median Share 2004-2008	0.13	0.02	0.54	-0.39

Note: Backlog denotes the number of pending cases at the end of the civil year. Avg Age is the average age pending cases at the end of a year. Clearance Rate denotes the number of resolves cases divided by the n of incoming cases in a year. IV stands for instrumental variable and is the mean or median share per dist filled courtrooms in district in sessions courts over the indicated time span.

Table E.3: Statutory Tax Rates on Salaries and Revenues in two cities in India

Tax or mandatory contribution	**Tax Rate**	**Tax Base**
Bengaluru		
Social security contributions	12.00%	gross salaries
National insurance scheme	4.75%	gross salaries (state insurance contribution)
Central sales tax	2.00%	purchase price
Corporate income tax	30.00%	taxable profits
CENVAT (excise duty)	16.48%	value added (if trade with other states)
State VAT	12.50%	value added
Noida		
Social security contributions	12.00%	gross salaries
National insurance scheme	4.75%	gross salaries (state insurance contribution)
Central sales tax[13]	2.00%	purchase price
Corporate income tax	30.00%	taxable profits
CENVAT (excise duty)	16.48%	value added (if trade with other states)
State VAT	12.00%	value added

Note: In 2008, the central sales tax rate was decreased from 3% to 2%.

Table E.4: Parameters and Functions of the Model

Parameter	Description	Data Source	Value
c_f^e	Entry cost informal sector		
c_s^e	Entry cost formal sector		
$\bar{c}_i$	Operation cost informal sector		
$\bar{c}_f$	Operation cost formal sector		
δ_i	Exit shock informal sector		
δ_f	Exit shock formal sector		
τ_y	Production tax	World Bank?	
τ_w	Labor tax	World Bank?	
M	Mass of potential entrants		Per state?
b	Speed of court per district	eCourt	
w	Wage per district	Estimated	
M_s	Mass of entrants per sector	Estimated	
mu_s	Sector size	Estimated	
$\tilde{l}$	Informal worker threshold in formal firms	Estimated?	
$\underline{\theta}_s$	Productivity of last firm entering sector s		
L_s	Labor in sector s		

Function	Description	Form	Value
$\tau_i(l)$	Labor wedge: detection of informal firms		
$\tau_f(l)$	Labor wedge: detection of informal workers in formal firms		
$\eta(b)$	Productivity wedge due to court speed		
$\lambda(b)$	Labor wedge due to court speed		