Courts, Firms and Informality*

Peter Neis

November 11, 2024

Abstract

In many developing countries, informality remains prevalent among firms and workers, while the judicial systems encounter challenges such as substantial backlogs and inefficiencies. This study investigates the potentially ambiguous effect of court efficiency on both margins of formality, formality of firms and workers. First, I introduce judicial efficiency in an equilibrium model. Court efficiency can influence firms' decisions through two distinct channels. Firstly, it may directly impact a firm's productivity, reflecting a more favorable economic environment. Secondly, it can influence the relative cost of hiring formal versus informal labor. By combining case-level data from Indian courts with survey data on firms and workers, I find a significant positive impact of court efficiency on firm formality, predominantly driven by small firms. Additionally, court efficiency reduces the proportion of casual labor in large firms, but does not impact other aspects of informal labor or firms' revenues. A wedge on the relative cost of formal compared to casual labor in formal firms is sufficient to explain these results.

Keywords: Judicial efficiency, Informal labor, Informal Firms, India

JEL Codes: E26, J46, K40, O17

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,

*Université Clermont Auvergne, CNRS, IRD, CERDI, F-63000, Clermont-Ferrand, France. E-mail: peter.neis@uca.fr.
I am indebted to Daniel Chen, Mohamed Saleh and Shareen Joshi for their guidance and support over the last years. I thank Emmanuelle Auriol, Simone Bertoli, Matteo Bobba, Lorenzo Casaburi, Matthieu Chemin, Clément Imbert, Mateo Montenegro, Stéphane Straub, and Zaki Wahhaj 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.

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

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

firm behavior links 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]

This 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).

Using 2SLS to address reversed causality, three main findings stand out. First, the results promote a strong positive impact of court efficiency on firm formality. T

Linking these result back to the model, allows identifying the functional form of the two parameters linked to court efficiency. First, there is no direct effect of court efficiency on the productivity of firms. Second, the relative cost of formal compared to casual labor decreases.

Cross-sectional 2SLS regressions on the district level show a strong positive impact of court efficiency and the probability of firms being formal. This impact is observable for small, but not for large firms. On the intensive margin, court efficiency decreases the share of casual labor in large firms. However, court efficiency does not impact other areas of informal labor and does not impact firms' revenues. 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

[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.

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 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. Section 2 gives an overview of different sources of informality and a broad presentation of India's judiciary. Section 3 presents the different data sources. Section 4 introduces the theoretical model. Section 5 presents the empirical strategy and section 6 the results. The last section concludes.

2 Background

2.1 Informality in India

India, with its vast population of 1.4 billion, not only is the second-largest country and the world's largest democracy but also stands as the largest common law jurisdiction. Despite the impressive growth witnessed in its economy over the past four decades, India, like many other developing nations, faces a significant challenge in grappling with a high incidence of informality. In 2018, the World Bank reported that a substantial 80% of its non-agricultural workforce was engaged in informal employment, ranking India third out of 39 countries in this aspect (Elgin et al. 2021).

In this paper, informal workers are defined as all individuals who lack formal labor contracts and any form of social security benefits. This category encompasses two groups. First, self-employed workers who operate independently for their own gain. These self-employed are predominantly associated with relatively low income levels. Second, casual workers, are hired unofficially by both formal and informal firms. As of 2010, approximately 75% of India's workforce was working on an informal

basis.[4]

Firms, on the other side, can be broadly classified into three categories. The first group encompasses self-employed individuals who rely solely on their own resources, without assistance from outside their own household. This group constitutes the largest proportion among Indian firms. The second group comprises non-registered firms, which hire casual labor from sources outside their immediate households. Combined, these two groups result in what is commonly referred to as "informal" firms, which accounted for 71% of all Indian firms in the year 2010. Lastly, the third group comprises formal firms that have been duly registered with the pertinent authorities. On average, these formal firms tend to be larger and yield higher profits in comparison to their informal counterparts.

2.2 Courts in India

India has the world's largest common law system, characterized by a multi-tiered court hierarchy (see Figure 2 for comparison). The apex judicial body in India is the Supreme Court, which handles appeals from high courts, questions of fundamental rights, and disputes between different organs of the state. The subsequent tier comprises 25 high courts, each exercising jurisdiction over one or more of the 28 states. Certain high courts have established benches in places beyond their primary locations. While these high courts are primarily appellate courts for decisions made by lower courts, they also exercise original (civil and criminal) jurisdiction for cases which do not fall under the jurisdiction of lower courts.

The jurisdictional area of a high court is divided into judicial districts, presided over by district and sessions courts. In these courts, judges handle both civil and criminal cases, and are called district judges when handling civil, and sessions judges when handling criminal cases. These district and sessions courts constitute the principal courts for original civil jurisdiction within a district, with their operations supervised and managed by the corresponding high court. Today, there are 671 district courts in 739 districts.

India's judicial system is known for its substantial backlog of pending cases in all levels of the

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

judicial hierarchy and a slow judicial process. Despite India's long-lasting efforts to reduce the backlog and expedite court proceedings, the number of pending cases has continued to rise in recent years. As of today, there are over 800 pending cases per judge in the lower judiciary, with new filings exceeding disposals. The Heritage Foundation describes the rule of law in India as follows: "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 the years 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 A8 gives an overview of the number of districts per source.[7] Appendix E explains how districts were selected and some variables created. Figure 4 gives a graphical overview of which districts are included in the final data set and which are not.

3.1 Firms

Information on informal and non-incorporated formal firms is obtained from the Survey on Unincorporated Non-Agricultural Enterprises (SUNAE) conducted under the 67th round of the Indian National Sample Survey. The survey excludes incorporated enterprises (i.e. registered under Companies Act, 1956 or Factories Act, 1948), construction, government and public sector enterprises, and cooperatives. In total, the survey covers 334,474 firms and includes economic and operational variables of enterprises in manufacturing, trade and other services.

Information on incorporated formal firms is obtained from the Annual Survey of Industries 2009/10

[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 unfortunately do not have district identifiers.

[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.

(ASI). 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 $\times$ four-digit national industrial classification cluster.[8] In total, ASI 2009/10 covered 61,080 firms (23,782 census and 37,298 sample).

Merging these two surveys, and considering the subset of industries covered in both, gives a total of 327,693 firms. Following the above definition of firm (in)formality, 29% of all the firms in this sample are classified as formal, while the remaining 71% are classified as informal. On average, informal firms employ 1.56 workers, whereas formal firms exhibit a twice as large workforce, employing an average of 3.27 workers. Panel A of Figure 3 displays density estimates of the size of formal and informal firms. One can observe that although there is a significant overlap between formal and informal firms, most informal firms have 5 or fewer workers, while the distribution for formal firms is less concentrated with a significant number of firms having a workforce larger than 5.

3.2 Workers

Data pertaining to workers is extracted from the Indian Employment and Unemployment Survey 2009/2010 (IEU), a nationwide household survey designed to provide comprehensive insights into the labor force and activity profiles of the population. This survey encompasses 100,957 households and 459,784 individuals. By focusing on workers within the overlap of sectors represented in ASI and SUNAE, the final sample comprises 118,133 workers. Within this sample, 9% are formally employed, while the remaining 91% are part of the informal labor force. Specifically, 63% are self-employed, and 28% are hired as casual workers.

Panels B and C of Figure Figure 3 present the percentage of the total workforce working in formal and informal firms, categorized by firm size, as well as the percentage of the total workforce classified as formal or informal, also categorized by firm size. Notably, these statistics show that a significant majority of workers are engaged in informal labor and within informal firms. Additionally, Panel D of

[8] Under the national industry classification, activities are first classified into broad sections (e.g. "Manufacturing"). These sections are then divided into 2-digit divisions (e.g. "13 - Manufacture of textiles"). The sections are constituted of 3-digit groups (e.g. "131 - Spinning, weaving and finishing of textiles"). Finally, each group consists of 4-digit classes (e.g. "1311 - Preparation and spinning of textile fibre"). See https://www.ncs.gov.in/Documents/NIC_Sector.pdf, accessed June 6, 2022

Figure Figure 3 displays the proportion of workers in formal firms, categorized by employment status and firm size. Remarkably, while small firms hire a slightly higher proportion of informal than formal workers, larger firms principally employ formal workers, though not exclusively.

From these three surveys, different statistics on workers and firms are created. Table A1 gives an overview over these moments and their respective data sources. ASI and SUNAE are employed to construct the size distribution of formal and informal firms, while the combined survey data yields insights into the overall share of informal firms. At the intensive margin, the total share of informal workers and the share of informal workers within formal firms are created from the IEU data.

3.3 Courts

Court efficiency is assessed through case-level data from District and Sessions Courts, obtained via web scraping from the Indian eCourts smartphone application.[9] The eCourts website is a centralized project initiated by the Indian government, commencing in 2007 following the "National Policy and Action Plan for Implementation of Information and Communication Technology (ICT) in the Indian Judiciary - 2005." The primary objective of this project was to enhance the efficiency and accessibility of the judiciary by implementing court computerization. Consequently, numerous court complexes underwent computerization, and processes were digitalized. The eCourts project claims that as of today, all district courts are presently interconnected with internet connectivity, and most pending cases have been entered into the centralized database.[10] As of now, the website provides cause lists and case statuses for more than 70 million pending and disposed cases.

This database includes both the flow and a segment of the stock of cases for all connected district courts, including those cases that have been treated at least once since the court's integration into the eCourts system. Hence, it covers all newly filed cases and those disposed of after connection to the eCourts service. The longer a court has been connected to the system, the greater the likelihood that most or all pending cases have undergone some form of treatment. However, it is plausible that certain very old cases have remained pending for an extended period without undergoing any treatment after

[9] Thanks to the Data Science Justice Collaboratory for sharing the raw data.

[10] https://ecourts.gov.in/ecourts_home/static/about-us.php, accessed March 27, 2022

the implementation of ICT in the court. The web-scraped data contains cases filed in nearly all district courts from 1954 until early 2019.

Indian district courts experience significant congestion, with many courts counting thousands of pending cases, leading to substantial delays before new cases can be treated. A considerable proportion of first instance cases related to business, industries, and labor are filed in District and Session courts (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. Hence, this paper does not differentiate between the two case types.

Figure Figure A1 displays the evolution of the number of filings in District and Session courts, accessible on eCourts, between 2000 and 2014 by case type. The left panel demonstrates an important increase in filings over time, for both civil and criminal cases. The right panel shows that the share of yearly filings by case type remained relatively stable over time, with civil cases consistently representing around 10% of all filings.

There is no singular, universally accepted measure of court efficiency (see for instance Glanfield, Hall, and Keilitz 2018 and European Commission 2020). Court efficiency can be evaluated from various perspectives, including the speed of court proceedings, access to justice, and fairness of judgments. This study focuses on the speed of court proceedings as a key indicator. Even within the realm of court speed, multiple measures can be used. In this paper, four distinct measures are introduced, all of which have been used in different contexts within the literature.

First, a raw measure of court efficiency is backlog, which denotes the number of pending cases in a court. For the purposes of this paper, backlog is defined as all cases that remain unresolved at the end of a civil year. Second, following Boehm and Oberfield (2018), the average age of pending cases is calculated as the average age of all pending cases at the end of a year, based on their filing date. Third, the disposition time represents 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 by dividing the number of unresolved cases by the number of resolved cases in a court at the end of a year, and then multiplying the result by 365. Lastly, the clearance rate indicates whether a court is resolving more cases than new cases being filed. It is defined as the ratio of the number of resolved cases and the number of incoming cases.

Figure A2 displays the evolution of these four court-speed indicators over time. Backlog increases significantly over the period 2000 to 2014. At the end of 2014, there are more than 3.2 million pending cases in Indian District and Session courts. The average age of pending cases decreases slightly from around 4.5 years in 2000 to 3 years in 2014. The disposition time decreases significantly, with the average time to disposal being around 2 years in 2014. Finally, the ratio of resolved to incoming cases (clearance rate) increases from 0.32 in 2000 to 0.76 in 2014. Taken together, although the court speed is increasing over time, there are still more incoming than resolved cases, leading to a continued congestion of courts.

3.4 Additional Data Sources

In 2009, the World Bank conducted a subnational Doing Business report for India.[11] 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 A7 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 to gross salaries and around 44% on revenues.

Other data sources include the Reserve Bank of India for outstanding credit of scheduled banks per district and the 2011 census for population data, which are included in the empirical section as covariates.

3.5 Aggregation and Summary Statistics

For the analysis in this paper, court efficiency measures and covariates are aggregated at the district level. Figure 4 gives an overview of the spatial coverage of the 189 districts which have data on clearance rate, and firm and worker informality. Appendix E describes the selection of districts and Table A8 displays why districts are not included in the final sample.

[11] https://www.doingbusiness.org/en/data/exploreeconomies/india/sub/noida#DB_tax, accessed September 15, 2021.

Table 1 gives an overview over the data. Informal workers represent, on average, 90% of the total workforce. 62% of them are self-employed and 28% are casual workers. On average, 70% of firms in a district are informal and hire 2.1 workers. There are, on average, 5090 pending cases in District and Sessions courts per district, and these cases are on average 3.16 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 A3 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.

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

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

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

firm 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 firm 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 $(w, L_s, \underline{v_s}, M_s, \mu_s)$ 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 Comparative statics / Predictions

What happens if court efficiency changes? Depending on the functional forms of $\eta(b)$ 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' maximization process. Firms do not change their behavior, therefore, observable outcome variables do not change either.

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 fewer 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 fewer 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: An increase in b leads to lower revenues for formal firms and to lower costs of formal workers. The latter 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 fewer 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.

5 Empirical Strategy

This section studies the two fundamental questions of the impact of court efficiency on informality. The empirical approach uses a cross-sectional analysis, using data on workers and firms from 2009/10 and court efficiency measures from 2008. The baseline regression is defined as

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

where y_d is the outcome of interest in district d in years 2009/10. This can be the formality status of a firm or worker, or a firms' revenues or sales per worker. 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 district level controls, especially the share of the population being literate and the share of the population belonging to a scheduled caste. α_r are region fixed effects.

The outcome of interest in this study is measured in 2009/10, while court efficiency data is obtained from the year 2008. This difference is due to the information set available to firms when making their decisions. Since firms ado not know about future realizations of court efficiency of a particular year, they can only observe past court efficiency. Thus, their expectations regarding present court efficiency are solely based on recent realized values.

Figure A7 displays correlations between one court efficiency measure (average age of pending cases) and district level informality shares. The top-left panel displays a positive correlation between informality of firms and the age of pending cases. The top-right panel shows a slightly negative rela-

tionship between the share of informal workers and the age of pending cases. Lastly, the bottom panel displays a negative relationship between the share of informal workers in formal firms and the age of pending cases. Taken together, higher court efficiency is associated with lower informality shares of firms but slightly higher informality shares of workers.

5.1 IV Strategy

Estimating Equation 1 via ordinary least squares (OLS) may yield biased results, and one needs to be cautious with a causal interpretation. The potential endogeneity issue arises from the correlation between the error term and court efficiency. This endogeneity is coming from the possibility of reverse causality. Specifically, districts with a higher concentration of formal firms and workers may generate more revenues, thereby providing the district with a potentially larger budget for the judiciary. Although the organization of the lower judiciary falls under the purview of high courts rather than the districts themselves, districts can still exert influence on court speed through factors such as available infrastructure. Increased revenue enables districts to allocate more resources to enforcement, which can directly impact the number of new cases filed.

Therefore, to be able to interpret the results as causal impacts of the judicial system on firm and worker outcomes, an instrumental variable approach is used. Rao (2020) shows that in Indian states, judges at districts courts rotate every one to two years. The assignment for the rotation is quasi random. Combined with many 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 a district are correlated with the outcome of interest in this district. As pointed out by Rao (2020), the legal framework of the judiciary backs the exogeneity assumption as well: While high

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

courts have organizational 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.

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 A4 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. The left panel Figure A5 plots the distribution of the instrumental variable (mean share of occupied courtrooms from 2004 to 2008) with the corresponding within district standard variation. The right panel of Figure A5 displays one simulated distribution, when 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, taking into account the aforementioned rules on the rotation scheme. When repeating this exercise 100 times, the simulated distribution of the mean share of occupied courtrooms from 2004 to 2008 is statistically not different from the actual distribution.

Taking the mean over five years is a quite arbitrary definition of the instrumental variable, mainly motivated by capturing a long term trend in the judiciary. However, alternative instruments can be considered. In Figure A6, 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 A2 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. Dispo-

sition time is negatively correlated, and clearance rate displays a large, positive correlation with the instrumental variable.

Column (1) of Table 3 displays the first stage regression of the mean share of occupied courtrooms 2004-2008 and clearance rate. The regression shows a strong positive relationship between the instrumental variable and the court efficiency measure. Columns (2) and (3) of Table 3 displays the reduced form regression of the instrumental variable on two outcomes of interest, the formality status of both firms and workers. Here, the instrumental variable is strongly and positively associated with the formality of firms. If the mean share of occupied courtrooms would switch from 0 to 1, firms would be 14.8 percentage points more likely to be formal. On the other hand, however, there is no statistically significant association between the instrumental variable and the informality of workers.

Table A3 displays the first stage regression of several alternative specifications of the instrumental variable on clearance rate. The results are similar in magnitude and statistical significance as for the baseline instrumental variable. Figure A8 displays partial regression (left panel) and partial residual (right panel) plots of the mean share of occupied courtrooms 2004 - 2008 while including the population share being literate, the population share of scheduled castes, and region fixed effects. Figure 5 is a graphical representation of the first and second stage of the instrumental variable regression of clearance rate on formality of firms.

The following section studies the causal relationship between court efficiency and firm and worker informality. Equation 1 is estimated via two stage least squares (2SLS). Standard errors are clustered at the state $\times$ NIC-4 digit level. 2SLS tables display the efficient first stage F statistic.

6 Results

Table 4 compares OLS estimates with two stage least squares estimations of the formality status of firms on clearance rate. It also displays the build-up to the preferred specification. The table does not show a major difference between OLS and 2SLS estimates. Also including dummies for urban and semi-urban districts, and the district's population share being literate and of scheduled caste does not change the over result. In column (6), the estimation indicates that if a district's clearance rate would

jump from 0 to 1, the probability of a firm being formal would increase by 16.3%.

Table 5 estimates the same relationship but by different firm sizes. Column (2) and column (3) separate very small firms with up to two workers and larger firms with more than two workers. Both groups are significantly positively impacted by court efficiency, although the very small firms exhibit a slightly larger elasticity. Columns (4) and (5) separate firms with up to 10 workers and firms with more workers. While the impact on firms with up to 10 workers is significant and similar in size as the earlier results, is the point estimate on larger firms less than half the size and not significantly different from zero. Taken together, court efficiency increases firm formality, and this effect comes primarily from small firms.

Table 6 display results of the estimation of worker informality on clearance rate. Column (1) presents the result on both types of worker informality, self-employed and casual workers, taken together. The point estimate is negative, but small and statistically not different from zero. Column (2) looks at the probability of being self-employed. The negative coefficient is again small and statistically not different from zero. Lastly, column (3) looks at the probability of being a casual worker. Here, again, the estimated positive impact is small and statistically not different from zero. Hence, there is no observable impact of court efficiency on the overall probability of being part of the informal workforce. Table A6

Table 7 investigates the link between court efficiency and casual labor further. Columns (1) to (3) exhibit the impact on casual workers by firm size. Most remarkably, there is a significant negative effect on casual labor in firms larger than 10 workers. If the clearance rate jumped from 0 to 1, the probability of a worker being hired casually instead of formally, would decrease by 20 percentage points. Columns (4) to (6) focus only on workers in formal firms. Here, no significant effect is observable. Hence, court efficiency does impact worker informality, but only for casual labor in large firms, which decreases the more efficient the courts.

Table A4, Table A5, and Table A6 display results of similar estimations as above, but with the other three court efficiency measures replacing clearance rate as explanatory variable.

Panel A of Table 8 display the impact of court efficiency on revenues per worker by firm size, while panel B displays the impact on the ex factory value of manufactured good per worker by firm size. None

of the estimated coefficients is statistically different from zero. However, two patterns can be observed. First, the coefficient for both revenues and manufactured value per worker is positive for small firms with up to 10 workers. Second, the coefficients for larger firms are much larger in magnitude and negative. Altogether, there is no observable impact of court efficiency on firms revenues.

6.1 Back to the model

Taken together, court efficiency increases firm formality, an effect mainly driven by small firms, decreases the share of casual labor in large firms, and does not impact firms' revenues.

These three results are aligned with the model predictions of the 3rd case in subsection 4.7. The positive impact on firm formality excludes cases 1, 2, and 5 (see Table 2). The positive impact on worker formality in large firms excludes cases 2, 4, and 5. Finally, the zero effect on revenues and manufactured goods per worker excludes cases 2, 4, 5, and 6. Therefore, the only functional forms, where the model predictions are aligned with the empirical results, are that $\eta'(b) = 0$ and $\lambda'(b) < 0$.

This implies, that in the model, court efficiency enters formal firm's maximization process solely through a wedge on the cost of formal workers. It does not impact directly firms' productivity through a better business environment, coming for instance from an overall increased access to credit. Rather, it impacts firms' cost function by decreasing the relative cost of hiring formal labor compared with casual labor. This makes hiring formal workers more attractive for larger firms and, consequently, it becomes more attractive for a firm to formalize. This result offers a new channel of how court efficiency impacts economics actors.

7 Conclusion

This paper investigates the relationship between court efficiency and informality, both prevalent in many developing countries. The paper adds court efficiency to the model from Ulyssea (2018) to give a theoretical background on how courts influence firms' decisions to become formal or informal, and to hire formal or casual labor.Specifically, court efficiency's impact on formal firms is explored, considering both direct effects on productivity and indirect effects on the cost function, influencing the

relative cost of formal labor.

The impact of court efficiency, and more precisely, court speed on these two margins of informality is then investigated in the context of India. First, using case level data from district and sessions courts, court speed is measured in four ways, with the clearance rate selected as the primary measure for estimations. Firm and worker informality is measured by combining household and enterprise surveys.

Using 2SLS to address reversed causality, three main findings stand out. First, the results promote a strong positive impact of court efficiency on firm formality. This impact is observable for small, but not for large firms. Second, court efficiency decreases the share of casual labor in large firms. Third, court efficiency does not impact other areas of informal labor and does not impact firms' revenues.

Linking these result back to the model, allows identifying the functional form of the two parameters linked to court efficiency. First, there is no direct effect of court efficiency on the productivity of firms. Second, the relative cost of formal compared to casual labor decreases.

Altogether, the paper combines a theoretical model with empirical estimates and proposes an additional channel of how courts can impact economic activity. Given the prevalence of inefficient judiciaries in many developing economies, these results offer valuable insights into the potential for enhancing court efficiency to reduce economic distortions.

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.

Elgin, Ceyhun et al. (2021). “Understanding informality”. In: *CERP Discussion Paper* 16497.

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.

Melitz, Marc (2003). "The impact of trade on aggregate industry productivity and intra-industry reallocations". In: *Econometrica* 71.6, pp. 1695–1725.

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.

Figures

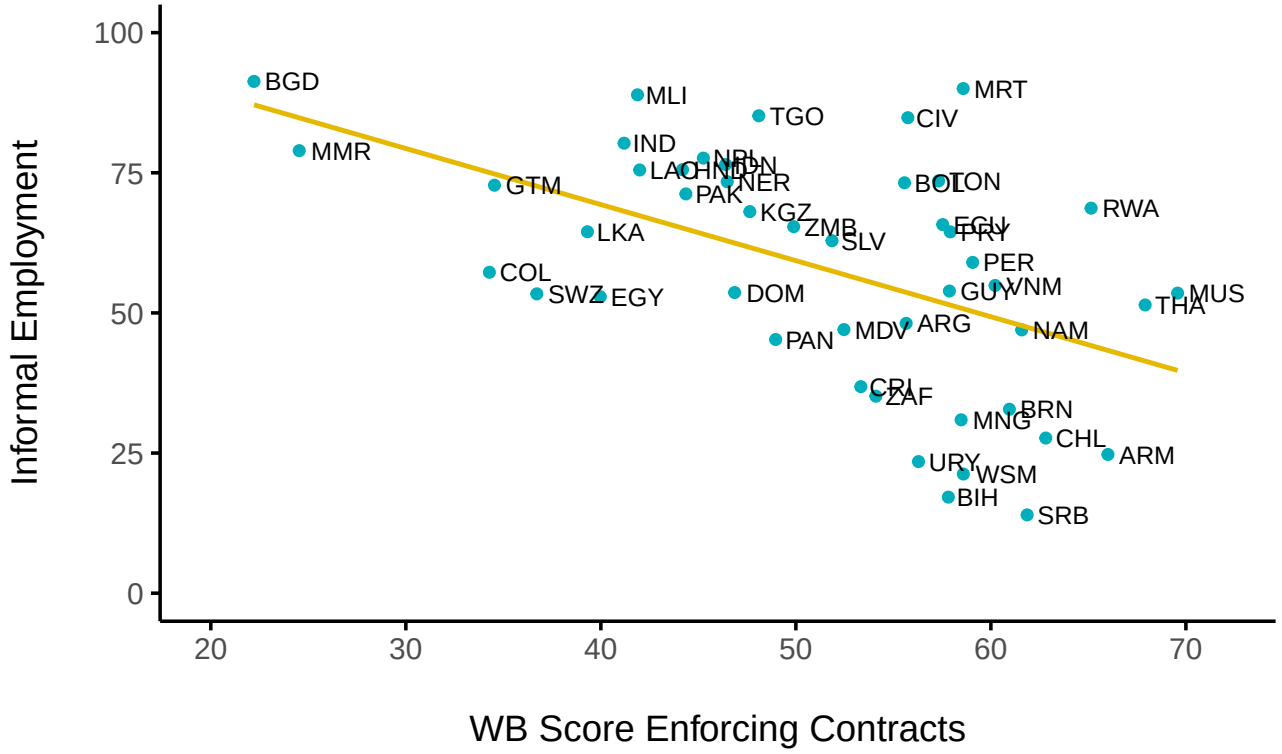


Figure 1: Correlation between Informal Employment and Contract Enforcement

Note: This figure displays on the x-axis country-level measurements of overall contract enforecement and on the y-axis the country's share of informal emplyoment. The data is based on the World Bank's Doing Business Report 2020.

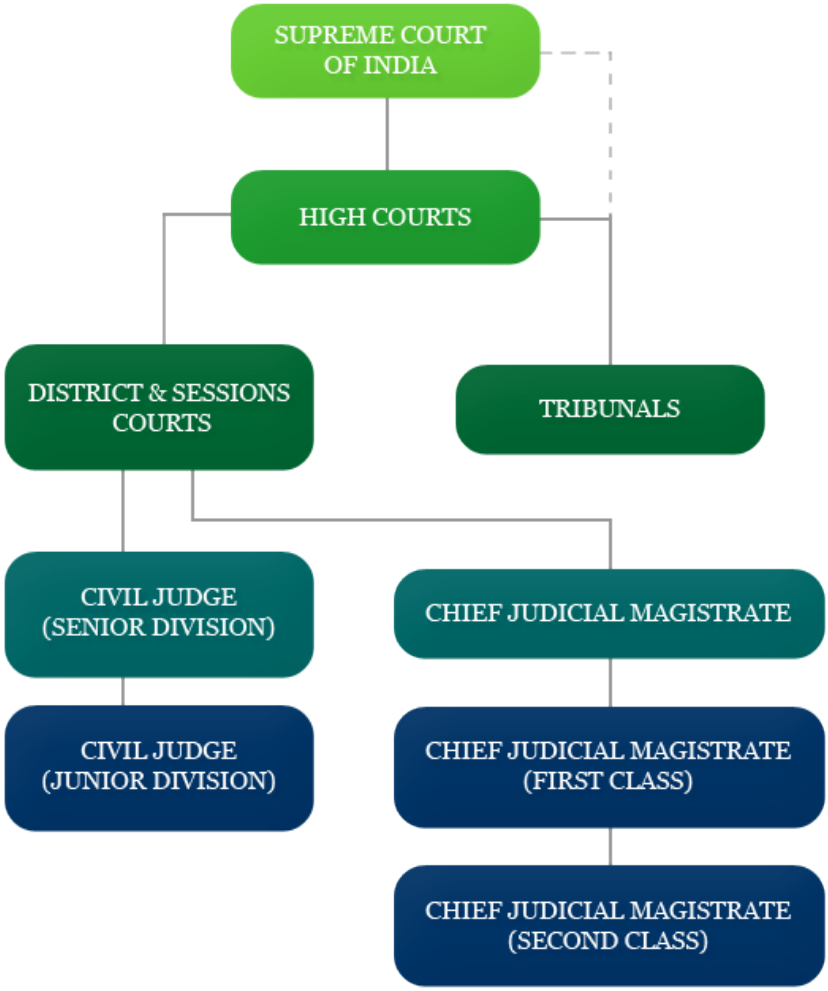
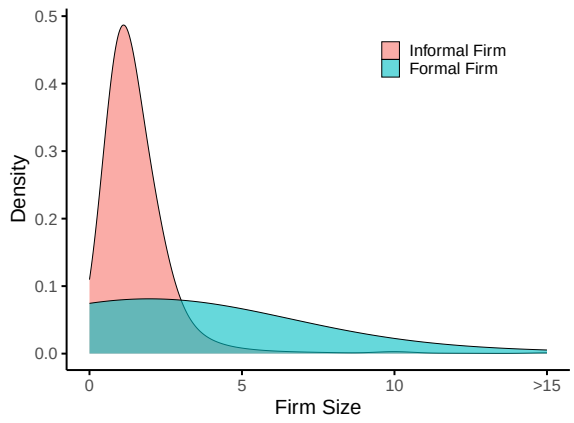
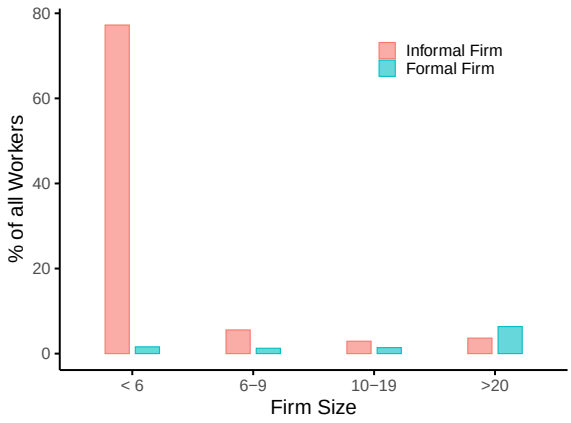


Figure 2: Hierarchy of the Indian Judicial System

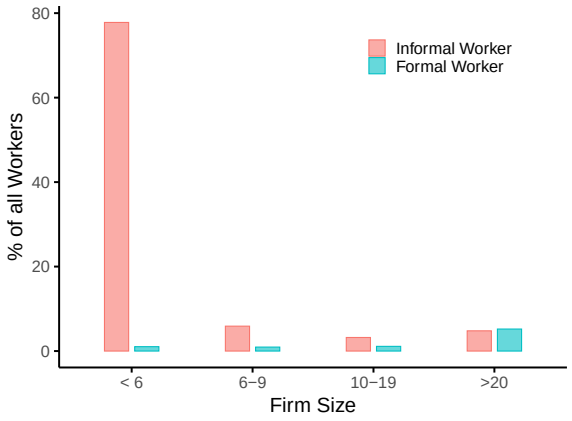
Note: This figure displays a stylized schema of the layers in the Indian judicial system. Source: Daksh (https://www.dakshindia.org/indian-courts/, accessed July 24 2023).



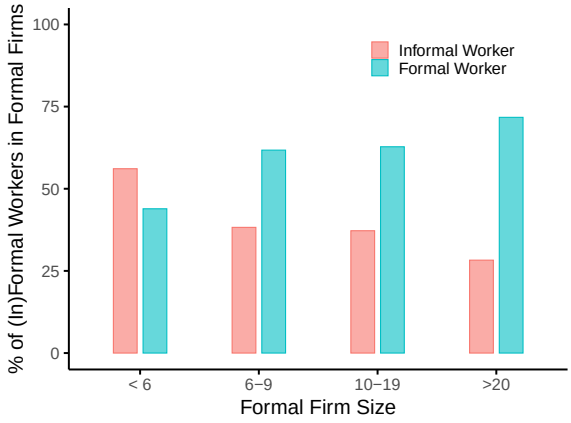
(a) Firm Size Density by Firm Status



(b) Workforce by Firm Size and Status



(c) Workforce by Firm Size and Worker Status



(d) (In)formal Workers in Formal Firms

Figure 3: Descriptive Graphs on formal and informal firms and workers in India.

Note: This figure displays descriptive stats of the (in)formality status and distributions of firms and workers. Panel (a) displays firm size distributions for both formal and informal firms. Panel (b) displays the share of all Indian workers by firms' size and firm formality status. Panel (c) displays the share of all Indian workers by firms' size and worker formality status. Panel (d) displays the share of (in)formal workforce in formal firms per firm size.

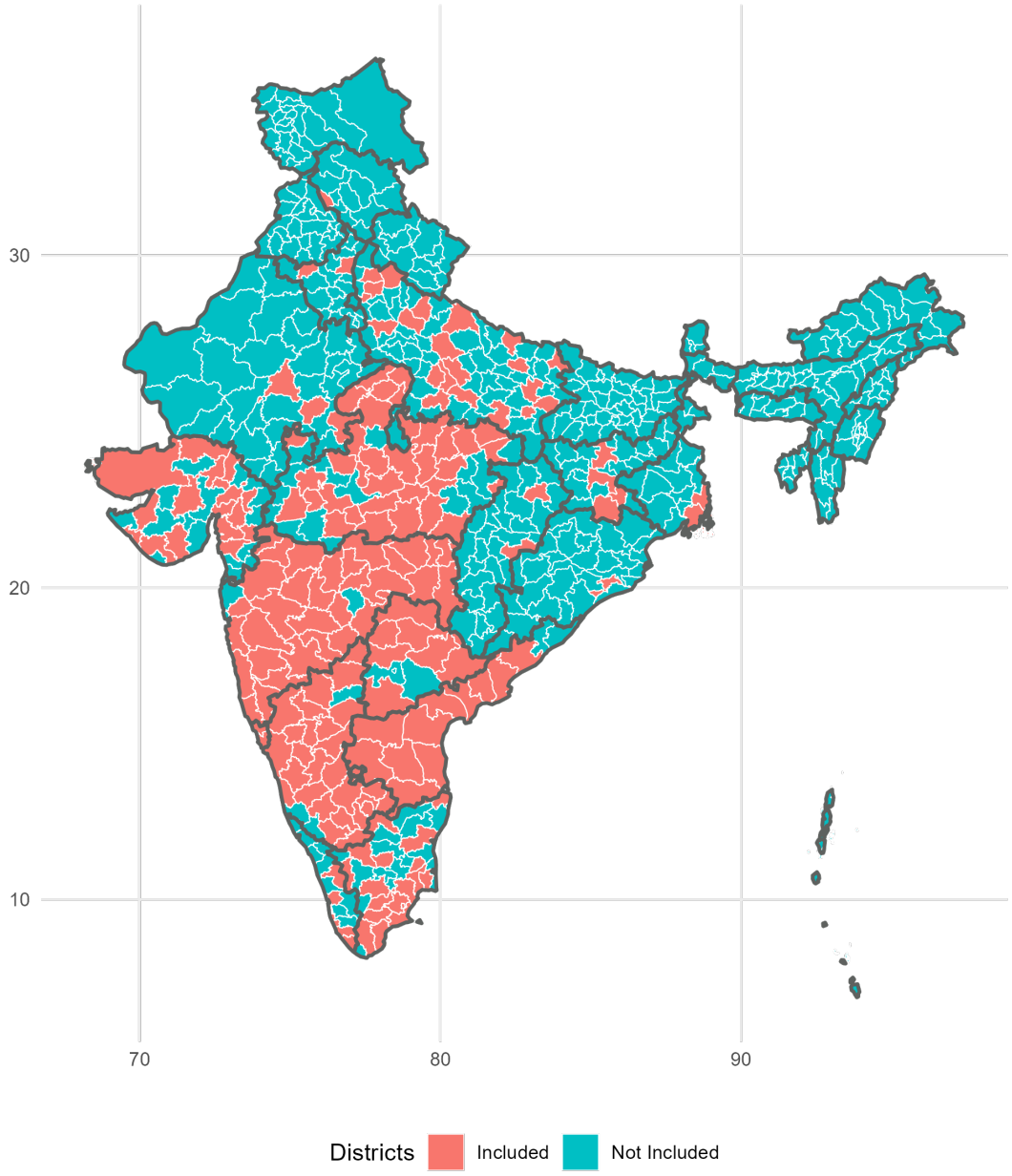
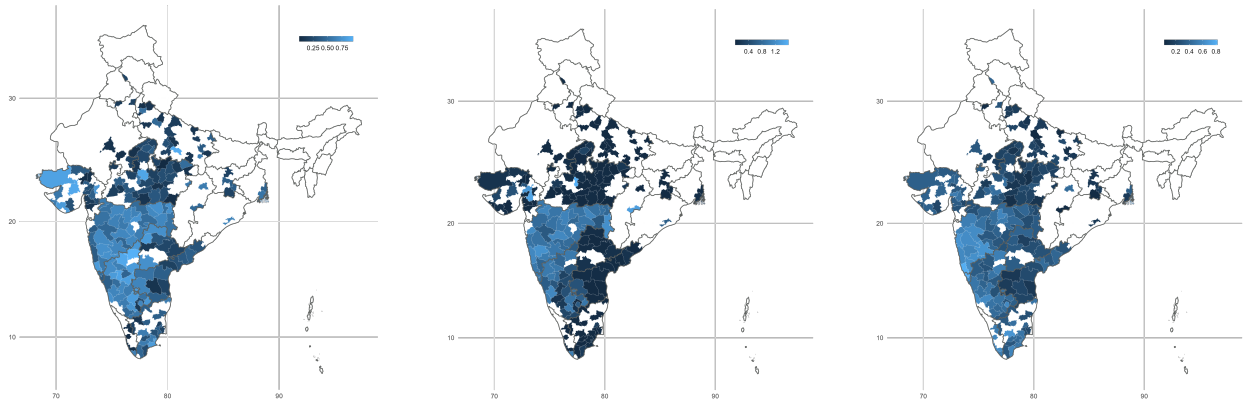


Figure 4: Spatial Coverage of Working Sample

Note: This figure displays a map of India, highlighting the spatial distribution of districts included in the analysis. Districts in red are included in the analysis. Districts in blue are excluded, due to lack of data in at least one of the main data sorces.



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

Figure 5: Spatial distribution of the Instrumental Variable, Clearance Rate, and the Share of Formal Firms in Districts

Note: This figure displays three maps of India at the district level. Panel (a) displays the average share of ocupied courtrooms in Indian District & Session Courts over the period 2004 - 2008. Panel (b) displays the clearance rate, defined as the number of resolved cases divided by the number of incoming cases in Indian District & Session Courts. Panel (c) displays the share of formal firms per district.

Tables

Table 1: Summary Statistics

	N	Mean	SD	Min	Max
Worker Informality (Worker Level)					
% Informal	65,291	0.90	0.30	0.00	1.00
% Self-employed	65,291	0.62	0.49	0.00	1.00
% Casual Workers	65,291	0.28	0.45	0.00	1.00
Firm Data (Firm Level)					
% Formal	190,990	0.30	0.46	0	1
Average Number of Workers	190,990	2.13	24.06	0	44,476
Gross Sales / Worker (million INR)	21,375	2.82	79.99	0	9,474
Value of Manufactured Goods / Worker (million INR)	21,375	2.79	80.06	0	9,474
District & Session Courts Data (District Level)					
Clearance Rate	189	0.32	0.43	.00071	1.6
Average Age of Pending Cases (Years)	256	3.16	1.50	.22	10
Disposition Time (Years)	192	1.63	2.89	.0011	15
Backlog (thsd.)	256	5.09	12.13	.003	158
Mean Share Occupied Courtrooms 2004-2008	258	0.36	0.24	.011	.94
Mean Share Occupied Courtrooms 2005-2008	258	0.33	0.23	.0093	.93
Mean Share Occupied Courtrooms 2003-2008	258	0.40	0.27	0	1
Mean Share Occupied Courtrooms 2004-2009	258	0.44	0.26	0	1
Median Share Occupied Courtrooms 2004-2008	258	0.33	0.33	0	1
Additional Variables (District Level)					
% Population Literate	258	0.64	0.09	.39	.87
% Population Scheduled Caste	258	0.17	0.07	.00086	.5
% Rural	255	0.83	0.38	0	1
% Urban	255	0.09	0.28	0	1
% Semi-Urban	255	0.09	0.28	0	1

Table 2: Model Predictions Depending on Parameters' Functional Forms

	$\eta'(b)=0$ $\lambda'(b)=0$	$\eta'(b)=0$ $\lambda'(b)>0$	$\eta'(b)=0$ $\lambda'(b)<0$	$\eta'(b)<0$ $\lambda'(b)=0$	$\eta'(b)<0$ $\lambda'(b)>0$	$\eta'(b)<0$ $\lambda'(b)<0$
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: First Stage And Reduced Form Regressions

	First Stage	Reduced Form	
	(1) Clearance Rate	(2) Formal Firm	(3) Informal Worker
Mean Share Occupied Courtrooms 2004-2008	0.809***	0.148**	-0.0110
	(0.133)	(0.0505)	(0.0307)
Region FE	✓	✓	✓
Covariates	✓	✓	✓
Observations	187	142,528	49,858

Note: This table displays the first stage and reduced forms of IV regressions of firms and workers' formality status in 2009 on the workers' district's clearance rate in 2008. The excluded instrument is the mean share of occupied courtrooms in district and session courts over the period 2004-2008 per district. Clearance rate is defined as the number of resolved cases divided by the number of incoming cases in district and session courts. Included covariates are the district's share of inhabitants being part of a scheduled caste in 2011, the district's share of inhabitants being literate, and two dummy variables equal to one if the district is mainly urban or semi-urban. *Region FE* represents dummies for six broad regions; Central, East, North, Northeast, South, and West. Robust standard errors are clustered at the industry-type $\times$ state level. Stars *, ** and *** indicate that the p-value is below 0.1, 0.05 and 0.01 respectively.

Table 4: Specification Build Up: Impact of Clearance Rate on Firms' Formality Status

	Formal Firm					
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV
Clearance Rate	0.129**	0.119	0.130**	0.130*	0.108**	0.163**
	(0.0394)	(0.0609)	(0.0405)	(0.0603)	(0.0403)	(0.0538)
Urban			0.0580**	0.0579**	-0.0227	-0.0207
			(0.0210)	(0.0211)	(0.0216)	(0.0209)
Semi-Urban			0.0557	0.0557	-0.000382	0.00485
			(0.0346)	(0.0348)	(0.0348)	(0.0347)
Share Literate					0.789***	0.747***
					(0.134)	(0.139)
Share SC					-0.214	-0.224
					(0.155)	(0.152)
Region FE	✓	✓	✓	✓	✓	✓
Mean Dep. Var.	0.32	0.32	0.32	0.32	0.32	0.32
First Stage F		31.99		31.64		37.23
Observations	143,404	143,404	142,528	142,528	142,528	142,528

Note: This table displays estimation results of OLS and IV regressions of a firm's formality status in 2009 on the firm's district's clearance rate in 2008. Columns 1, 3, and 5 display results from OLS regressions. Columns 2, 4, and 6 display results from instrumental variable regression. The excluded instrument is the mean share of occupied courtrooms in district and session courts over the period 2004-2008 per district. Clearance rate is defined as the number of resolved cases divided by the number of incoming cases in district and session courts. *Urban* is a dummy variable equal to one if the district is mainly urban. *Semi-Urban* is a dummy variable equal to one if the district is half-half urban and rural. *Share SC* denotes the district's share of inhabitants being part of a scheduled caste in 2011. *Share Literate* denotes the district's share of inhabitants being literate. *Region FE* represents dummies for six broad regions; Central, East, North, Northeast, South, and West. Robust standard errors are clustered at the industry-type $\times$ state level. Stars *, ** and *** indicate that the p-value is below 0.1, 0.05 and 0.01 respectively.

Table 5: Impact of Clearance Rate on Firms' Formality Status by Firm Size

	Formal Firm if Firm of Size				
	(1) All	(2) ≤ 2 Workers	(3) > 2 Workers	(4) ≤ 10 Workers	(5) > 10 Workers
Clearance Rate	0.163**	0.183**	0.133*	0.168**	0.0793
	(0.0538)	(0.0565)	(0.0643)	(0.0535)	(0.103)
Region FE	✓	✓	✓	✓	✓
Covariates	✓	✓	✓	✓	✓
Mean Dep. Var.	0.32	0.28	0.59	0.32	0.78
First Stage F	37.23	37.23	37.23	37.23	37.23
Observations	142,528	91,590	50,938	122,920	19,608

Note: This table displays estimation results of IV regressions of a firm's formality status in 2009 on the firm's district's clearance rate in 2008 by firms of a specific size. The excluded instrument is the mean share of occupied courtrooms in district and session courts over the period 2004-2008 per district. Clearance rate is defined as the number of resolved cases divided by the number of incoming cases in district and session courts. Included covariates are the district's share of inhabitants being part of a scheduled caste in 2011, the district's share of inhabitants being literate, and two dummy variables equal to one if the district is mainly urban or semi-urban. *Region FE* represents dummies for six broad regions; Central, East, North, Northeast, South, and West. Robust standard errors are clustered at the industry-type $\times$ state level. Stars *, ** and *** indicate that the p-value is below 0.1, 0.05 and 0.01 respectively.

Table 6: Impact of Clearance Rate on Workers' Formality Status

	Informal Worker (by Employment Type)		
	(1) Any Inf. Type	(2) Self-Employed	(3) Casual Worker
Clearance Rate	-0.0136	0.0636	-0.0772
	(0.0382)	(0.115)	(0.0973)
Region FE	✓	✓	✓
Covariates	✓	✓	✓
Mean Dep. Var.	0.90	0.61	0.28
First Stage F	37.23	37.23	37.23
Observations	49,858	49,858	49,858

Note: This table displays estimation results of IV regressions of workers' formality status in 2009 on the workers' district's clearance rate in 2008. The excluded instrument is the mean share of occupied courtrooms in district and session courts over the period 2004-2008 per district. Clearance rate is defined as the number of resolved cases divided by the number of incoming cases in district and session courts. Included covariates are the district's share of inhabitants being part of a scheduled caste in 2011, the district's share of inhabitants being literate, and two dummy variables equal to one if the district is mainly urban or semi-urban. *Region FE* represents dummies for six broad regions; Central, East, North, Northeast, South, and West. Robust standard errors are clustered at the industry-type $\times$ state level. Stars *, ** and *** indicate that the p-value is below 0.1, 0.05 and 0.01 respectively.

Table 7: Impact of Clearance Rate on Workers' Formality Status by Firm Size

	Casual Worker in Any Firm of Size			Casual Worker in Formal Firm of Size		
	(1) All	(2) < 10 Workers	(3) ≥ 10 Workers	(4) All	(5) < 10 Workers	(6) ≥ 10 Workers
Clearance Rate	-0.0470	0.0436	-0.200*	-0.0532	-0.114	-0.00231
	(0.0672)	(0.0586)	(0.0883)	(0.0700)	(0.136)	(0.0723)
Region FE	✓	✓	✓	✓	✓	✓
Covariates	✓	✓	✓	✓	✓	✓
Mean Dep. Var.	0.74	0.90	0.52	0.32	0.47	0.27
First Stage F	10.98	10.98	10.98	7.62	7.62	7.62
Observations	17,871	10,379	7,492	6,459	1,955	4,504

Note: This table displays estimation results of IV regressions of workers' formality status in 2009 on the workers' district's clearance rate in 2008. The excluded instrument is the mean share of occupied courtrooms in district and session courts over the period 2004-2008 per district. Clearance rate is defined as the number of resolved cases divided by the number of incoming cases in district and session courts. Included covariates are the district's share of inhabitants being part of a scheduled caste in 2011, the district's share of inhabitants being literate, and two dummy variables equal to one if the district is mainly urban or semi-urban. *Region FE* represents dummies for six broad regions; Central, East, North, Northeast, South, and West. Robust standard errors are clustered at the industry-type $\times$ state level. Stars *, ** and *** indicate that the p-value is below 0.1, 0.05 and 0.01 respectively.

Table 8: Impact of Clearance Rate on Formal Firms' Revenues per Worker

	Formal Firms of Size		
	(1) All	(2) ≤ 10 Workers	(3) > 10 Workers
Panel A: Sales / Worker (Million IRN)			
Clearance Rate	-6.841	1.782	-9.573
	(7.192)	(1.258)	(9.546)
Mean Dep. Var.	3.07	1.69	3.45
First Stage F	37.74	37.74	37.74
Observations	15,950	2,479	13,471
Panel B: Ex Factory Value of Manuf. Goods / Worker (Million IRN)			
Clearance Rate	-7.140	1.380	-9.840
	(7.206)	(1.368)	(9.558)
Mean Dep. Var.	3.05	1.71	3.42
First Stage F	37.74	37.74	37.74
Observations	15,950	2,479	13,471
Region FE	✓	✓	✓
Covariates	✓	✓	✓

Note: This table displays estimation results of IV regressions of formal firms' revenues per worker in 2009 on the firms' district's clearance rate in 2008 by firms of a specific size. Panel A displays the impact on gross sales value per worker in million Indian rupees. Panel B displays the impact on the ex factory value of manufactured goods per worker in million Indian rupees. The excluded instrument is the mean share of occupied courtrooms in district and session courts over the period 2004-2008 per district. Clearance rate is defined as the number of resolved cases divided by the number of incoming cases in district and session courts. Included covariates are the district's share of inhabitants being part of a scheduled caste in 2011, the district's share of inhabitants being literate, and two dummy variables equal to one if the district is mainly urban or semi-urban. *Region FE* represents dummies for six broad regions; Central, East, North, Northeast, South, and West. Robust standard errors are clustered at the industry-type $\times$ state level. Stars *, ** and *** indicate that the p-value is below 0.1, 0.05 and 0.01 respectively.

A Uniqueness of informal/formal worker threshold

$\tilde{l}$ 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, strictly 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 such that $\tau_f'(l) = b$. 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.

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 such that 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 for 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.

C Additional Figures

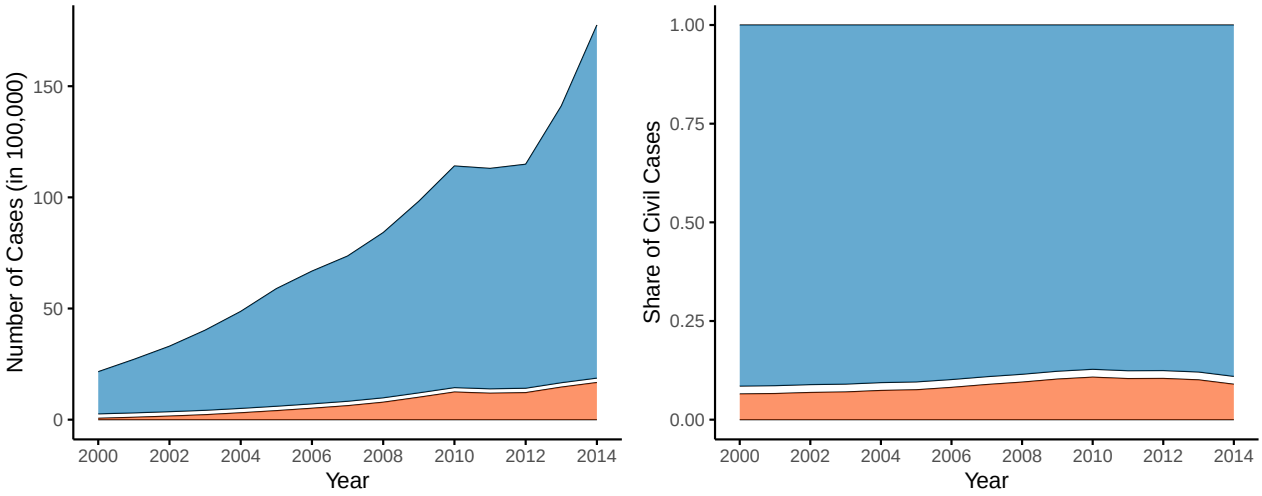


Figure A1: (Civil) Cases in District & Session Courts, 2000 - 2014

Note: This figure displays the evolution over time of the number of cases in district and session courts. The left panel displays the number of cases in 100,000. Its bottom part (in orange) represents the number of cicil cases per year (in 100,000). Its top part (in blue) represents the number of criminal cases per year (in 100,000). Both (orange and blue) together, rerpresent the total number of cases per year (in 100,000). The right panel displays the share of civil (bottom, orange) and criminal (top, blue) cases per year.

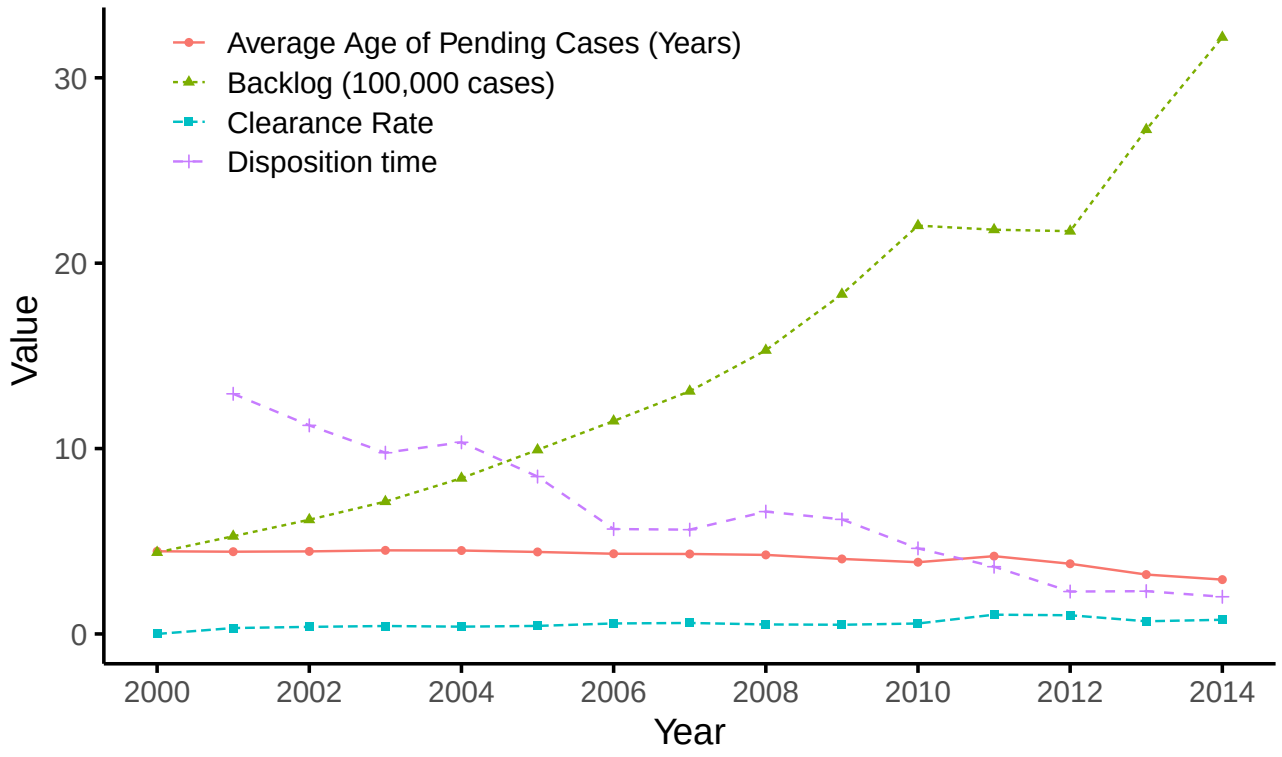


Figure A2: Court Speed Indicators for District & Session Courts, 2000 - 2014

Note: This figure displays the evolution over time of several court efficiency measures of district and session courts. The red line (cirqle) displays the average age of pending cases in years. The green line (triangle) displays the number of pending cases older than one year (backlog, in 100,000). The blue line (squares) displays the ratio of resolved cases and filed cases (clearance rate). The purple line (crosses) displays the ratio of pending cases and resolved cases (disposition time, in years)

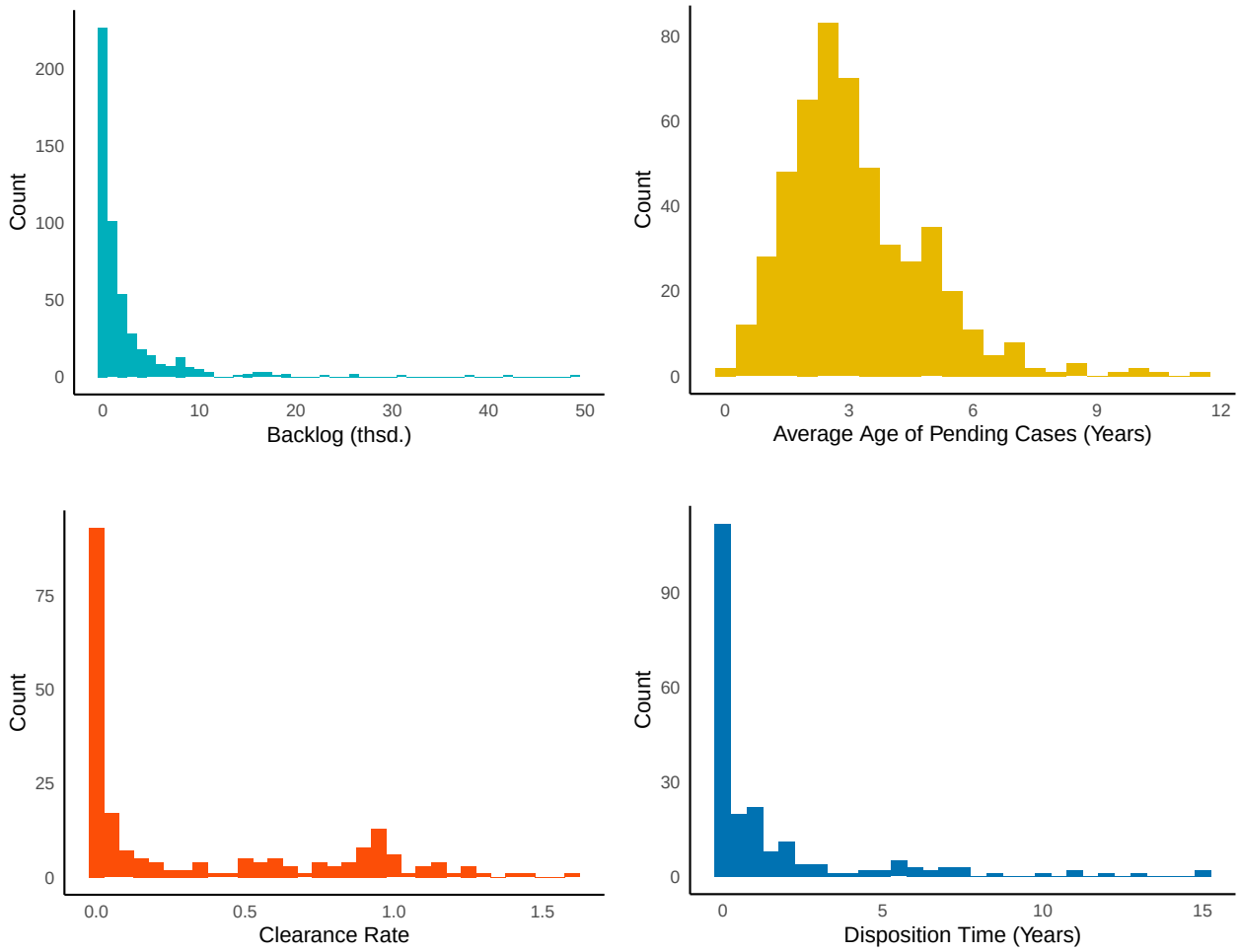


Figure A3: Histograms of District & Session Court Efficiency Measures in 2008

Note: This figure displays histrograms of several efficiency measures of Indian district & session courts per district in 2008. The top-left panel displays the the number of pending cases (in thousand) older than one year (backlog) per district at the end of 2008. The top-right panel displays the average age of pending cases (in years) per district at the end of 2008. The bottom-left panel displays the ratio of resolved cases and filed cases (clearance rate) per district in 2008. The bottom-right panel displays the ratio of pending cases and resolved cases (disposition time, per year) per district in 2008.

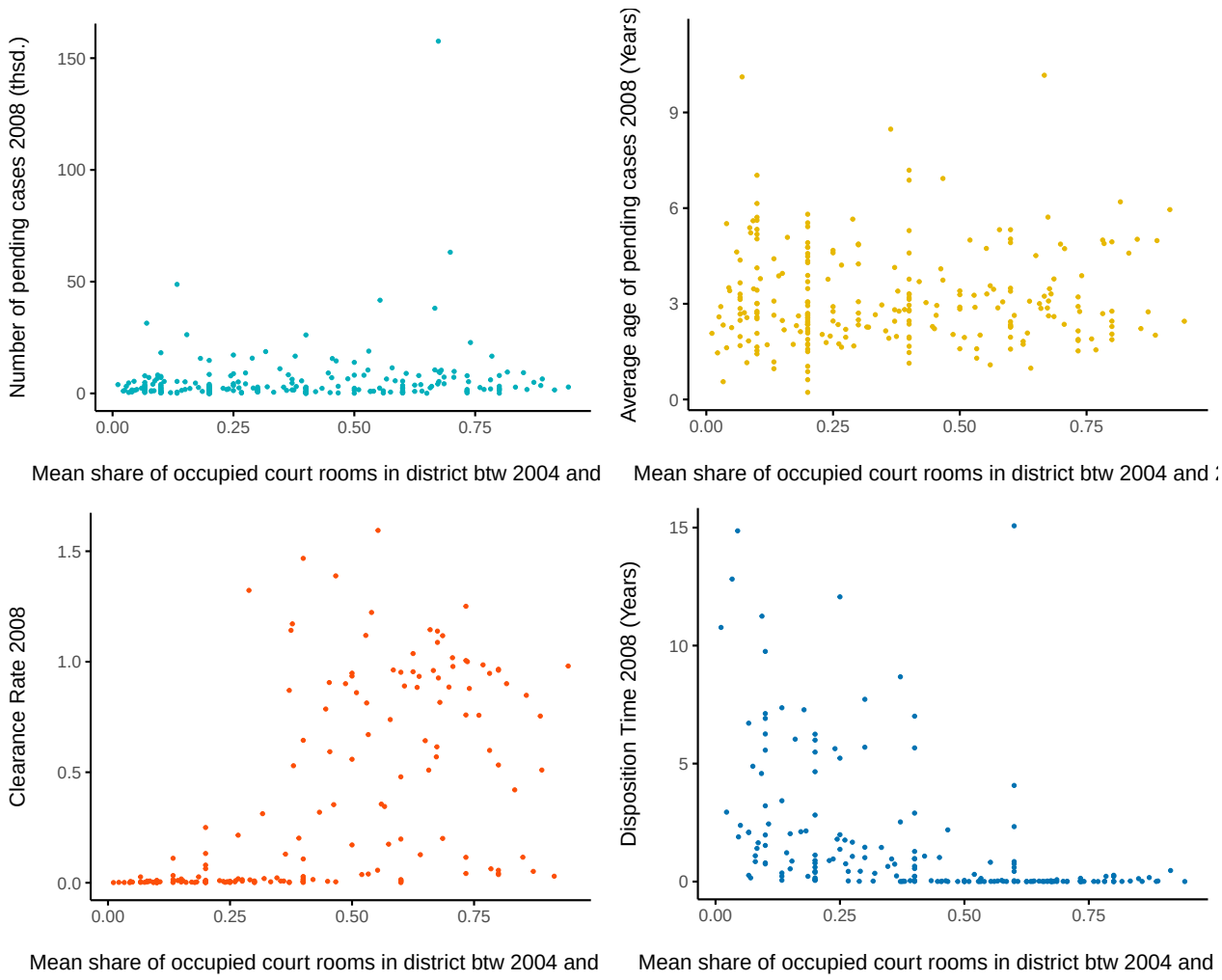


Figure A4: Correlation between the instrumental variable and court efficiency measures

Note: This figure displays scatter plots between the instrumental variable (average share of filled courtrooms in district and sessions courts per district between 2004 and 2008) on the x-axis and several measures of court efficiency in 2008 on the y-axis. The court efficiency measures are as follows. In the top-left panel, the number of pending cases in thousands. In the top right panel, the average age of pending cases in years. In the bottom-left panel, the clearance rate. In the bottom-right panel, the disposition time in years.

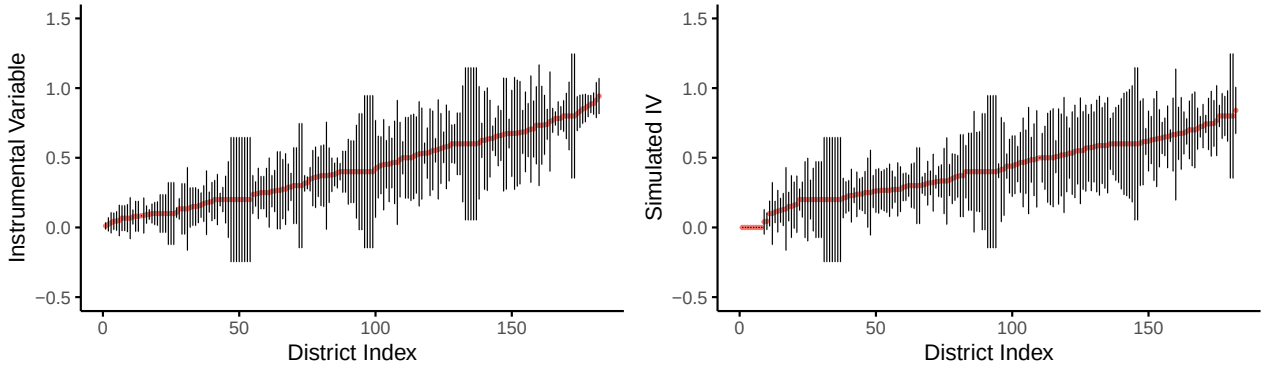


Figure A5: Distribution of Instrumental Variable across Districts

Note: This figure displays mean share of occupied courtrooms in district and session courts per district in the period 2004 to 2008, sorted in an ascending manner by the instrumnet's value. Red points represent the are the mean share of occupied courtrooms in district and session courts per district in the period 2004 to 2008. Gray bars represent the within district standard deviation of the share of occupied courtrooms in this five year period. The left panel displays the observed distribution. The right panel displays one simulated distribution of the instrumental variables, based on the initial occupanciy of court rooms in 2004 and consecutive random assignement of judges.

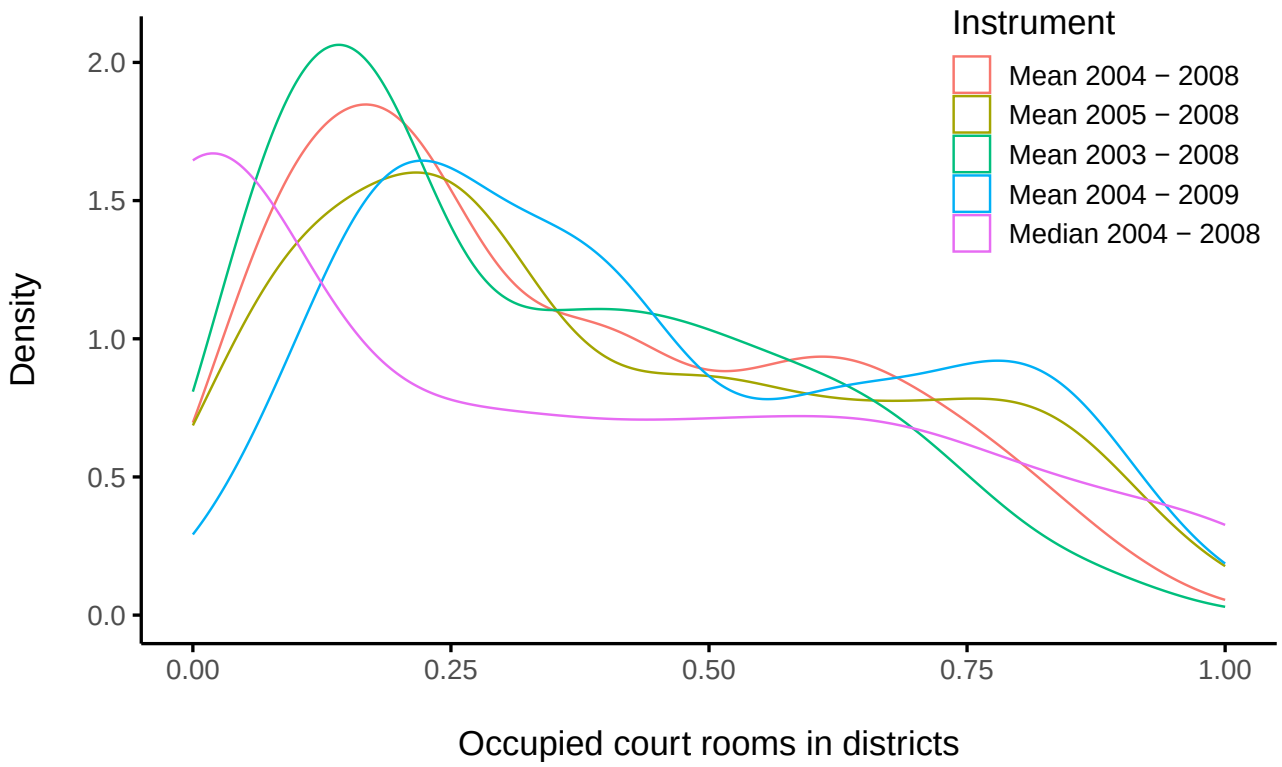


Figure A6: Densities of different instrumental variable specifications

Note: This figure displays densities of different specifications of the instrument variable. The baseline specification, in red, plots the average share of occupied courtrooms in district and session courts per district over the period 2004-2008. The light green line is calculated over the period 2005-2008, the dark green line over the period 2003-2008, and the blue line over the period 2004-2009. The purple line represents the median share of occupied courtrooms in district and session courts per district over the period 2004-2008.

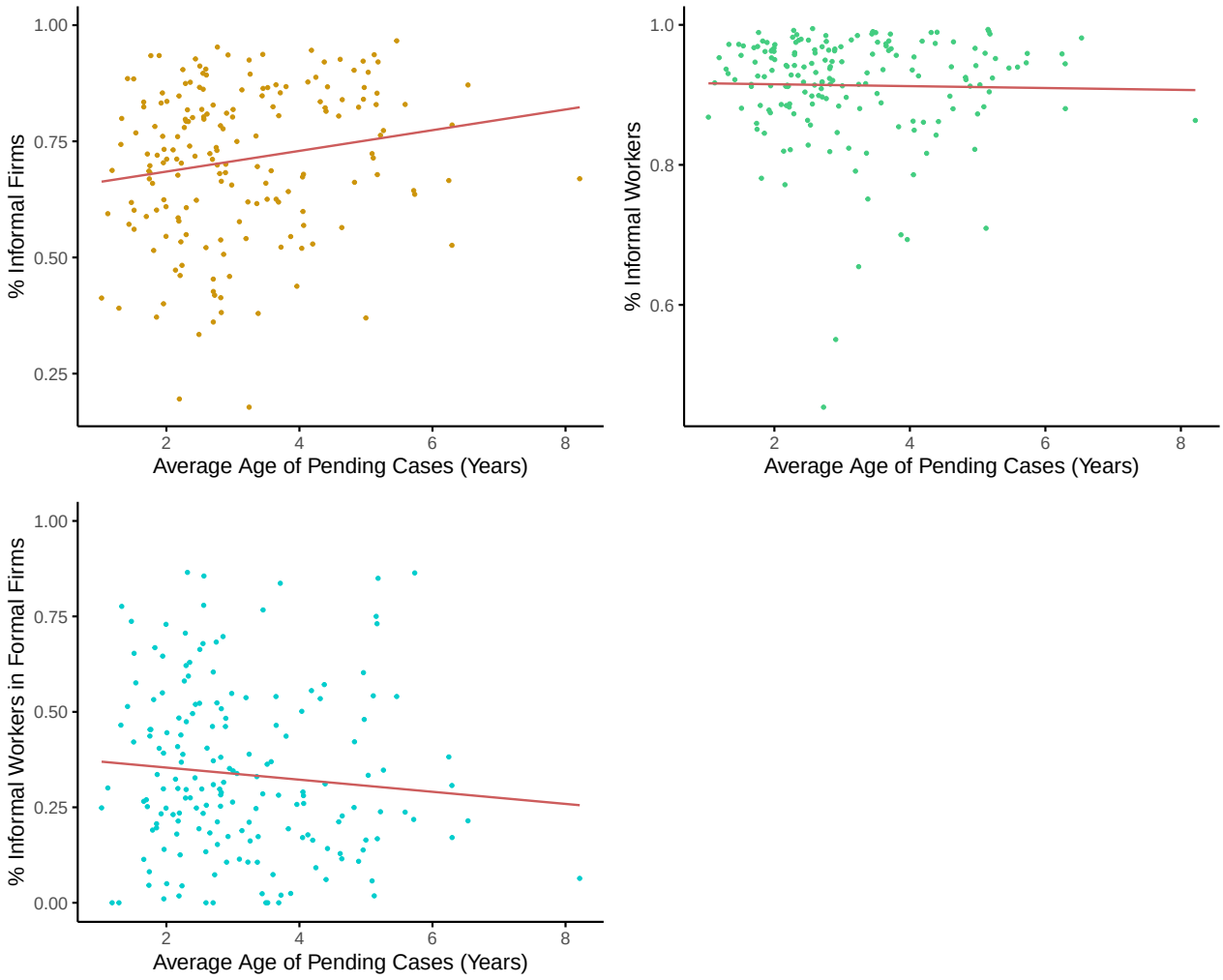


Figure A7: Average Age of Pending Cases in District & Session Courts and Informality Shares

Note: This figure displays the cooccurence of the avergae age of pending cases in district and session courts and different measures of informality per district in 2009. The measures of informality are as follows. In the top-left panel, the share of informal firms per district. In the top-right panel, the share of informal workers per district. In the bottom-left panel, the share of informal workers in formal firms per district. Solid red lines represent simple linear fits.

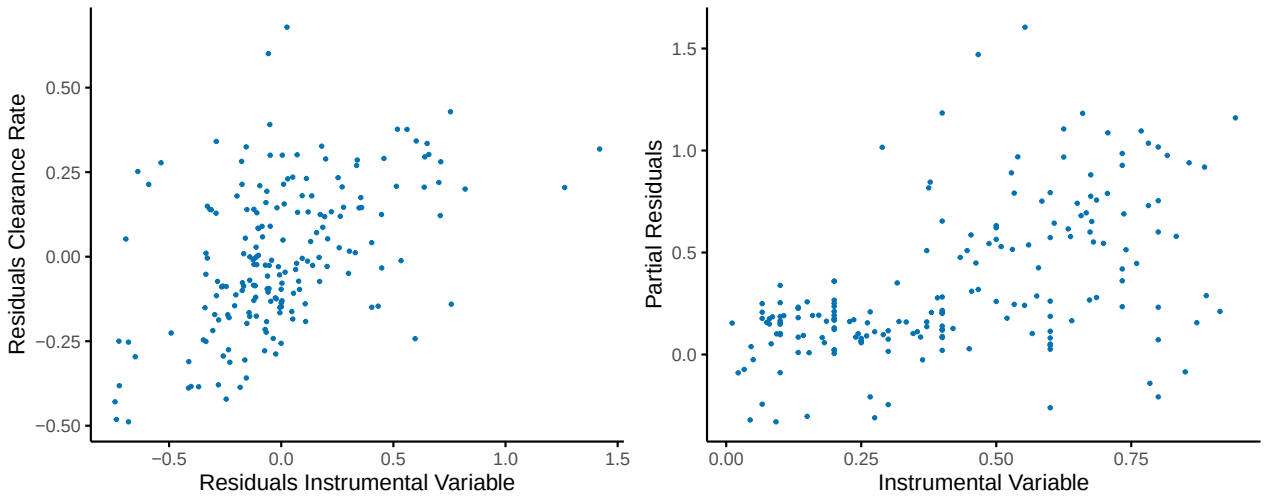


Figure A8: First Stage Relationship Between Clearance Rate and the Instrumental Variable

Note: This figure displays a partial regression (left panel) and partial residuals (right panel) of the following first stage regression. The first stage is regressing the clearance rate in district and session courts in 2008 on the mean share of occupied courtrooms of district and session courts over the period 2004 - 2008 (the instrument), the population share being literate, and region fixed effects. Observations are at the district level. The left panel displays on the x-axis the residuals from regressing the instrumental variable on the literacy share and region fixed effects, and on the y-axis the residuals from regressing the clearance rate on the same variables. The right panel displays on the x-axis the instrumental variable, and on the y-axis the sum of the first stage residuals and the predicted values based solely on the instrumental variable.

D Additional Tables

Table A1: 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 A2: Correlation between Court Efficiency Measures and Instrumental Variable 2008

	Backlog	Avg Age of Pending Cases	Clearance Rate	Disposition Time
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 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. IV stands for instrumental variable and is the mean or median share per district of filled courtrooms in district in sessions courts over the indicated time span.

Table A3: First Stage of Alternative Instruments

	Clearance Rate				
	(1)	(2)	(3)	(4)	(5)
Mean Share Occupied Courtrooms 2004-2008	0.809***				
	(0.133)				
Mean Share Occupied Courtrooms 2005-2008		0.789***			
		(0.121)			
Mean Share Occupied Courtrooms 2003-2008			0.842***		
			(0.145)		
Mean Share Occupied Courtrooms 2004-2009				0.868***	
				(0.133)	
Median Share Occupied Courtrooms 2004-2008					0.549***
					(0.106)
Region FE	✓	✓	✓	✓	✓
Covariates	✓	✓	✓	✓	✓
Observations	187	187	187	187	187

Note: This table displays first stages of IV regressions of firms' formality status in 2009 on the workers' district's clearance rate in 2008. Clearance rate is defined as the number of resolved cases divided by the number of incoming cases in district and session courts. Included covariates are the district's share of inhabitants being part of a scheduled caste in 2011, the district's share of inhabitants being literate, and two dummy variables equal to one if the district is mainly urban or semi-urban. *Region FE* represents dummies for six broad regions; Central, East, North, Northeast, South, and West. Robust standard errors are clustered at the industry-type $\times$ state level. Stars *, ** and *** indicate that the p-value is below 0.1, 0.05 and 0.01 respectively.

Table A4: Impact of Other Court Efficiency Measures on Firms' Formality Status by Firm Size

	Formal Firm if Firm of Size				
	(1) All	(2) $\leq$ 2 Workers	(3) $>$ 2 Workers	(4) $\leq$ 10 Workers	(5) $>$ 10 Workers
Panel A: Average Age of Pending Cases (Years)					
Average Age of Pending Cases	0.0854*	0.107**	0.0411	0.0904*	-0.0402
	(0.0347)	(0.0405)	(0.0281)	(0.0352)	(0.0524)
Mean Dep. Var.	0.30	0.26	0.56	0.30	0.78
First Stage F	2.17	2.17	2.17	2.17	2.17
Observations	189,158	122,224	66,934	163,384	25,774
Panel B: Disposition Time (Years)					
Disposition Time	-0.0437**	-0.0514**	-0.0250	-0.0457**	0.0197
	(0.0164)	(0.0180)	(0.0210)	(0.0164)	(0.0307)
Mean Dep. Var.	0.32	0.27	0.58	0.31	0.76
First Stage F	9.93	9.93	9.93	9.93	9.93
Observations	146,535	93,614	52,921	125,874	20,661
Panel C: Backlog (Thousand Cases)					
Backlog	0.00406**	0.00520**	0.00183	0.00431**	-0.00163
	(0.00152)	(0.00174)	(0.00123)	(0.00153)	(0.00226)
Mean Dep. Var.	0.30	0.26	0.56	0.30	0.78
First Stage F	3.64	3.64	3.64	3.64	3.64
Observations	189,158	122,224	66,934	163,384	25,774
Region FE	✓	✓	✓	✓	✓
Covariates	✓	✓	✓	✓	✓

Note: This table displays estimation results of IV regressions of a firm's formality status in 2009 on several court efficiency measures in the firm's district by firms of a specific size. The excluded instrument is the mean share of occupied courtrooms in district and session courts over the period 2004-2008 per district. *Average Age of Pending Cases* is the average age of all pending cases at the end of the civil year. *Disposition Time* is the ratio of the number of unresolved cases and the number of resolved cases in district and session courts. *Backlog* denotes the number of pending cases older than one year at the end of the civil year. Included covariates are the district's share of inhabitants being part of a scheduled caste in 2011, the district's share of inhabitants being literate, and two dummy variables equal to one if the district is mainly urban or semi-urban. *Region FE* represents dummies for six broad regions; Central, East, North, Northeast, South, and West. Robust standard errors are clustered at the industry-type $\times$ state level. Stars *, ** and *** indicate that the p-value is below 0.1, 0.05 and 0.01 respectively.

Table A5: Impact of Other Court Efficiency Measures on Workers' Formality Status

	Informal Worker (by Employment Type)		
	(1) Any Inf. Type	(2) Self-Employed	(3) Casual Worker
Panel A: Average Age of Pending Cases (Years)			
Average Age of Pending Cases	-0.0163	0.0391	-0.0554
	(0.0262)	(0.0783)	(0.0644)
Mean Dep. Var.	0.90	0.62	0.28
First Stage F	2.17	2.17	2.17
Observations	64,876	64,876	64,876
Panel B: Disposition Time (Years)			
Disposition Time	0.00508	-0.0105	0.0156
	(0.00878)	(0.0292)	(0.0247)
Mean Dep. Var.	0.90	0.61	0.28
First Stage F	9.93	9.93	9.93
Observations	50,992	50,992	50,992
Panel C: Backlog (Thousand Cases)			
Backlog	-0.000629	0.00151	-0.00214
	(0.000971)	(0.00313)	(0.00262)
Mean Dep. Var.	0.90	0.62	0.28
First Stage F	3.64	3.64	3.64
Observations	64,876	64,876	64,876
Region FE	✓	✓	✓
Covariates	✓	✓	✓

Note: This table displays estimation results of IV regressions of workers' formality status in 2009 on several court efficiency measures in the workers' district. The excluded instrument is the mean share of occupied courtrooms in district and session courts over the period 2004-2008 per district. *Average Age of Pending Cases* is the average age of all pending cases at the end of the civil year. *Disposition Time* is the ratio of the number of unresolved cases and the number of resolved cases in district and session courts. *Backlog* denotes the number of pending cases older than one year at the end of the civil year. Included covariates are the district's share of inhabitants being part of a scheduled caste in 2011, the district's share of inhabitants being literate, and two dummy variables equal to one if the district is mainly urban or semi-urban. *Region FE* represents dummies for six broad regions; Central, East, North, Northeast, South, and West. Robust standard errors are clustered at the industry-type $\times$ state level. Stars *, ** and *** indicate that the p-value is below 0.1, 0.05 and 0.01 respectively.

Table A6: Impact of Other Court Efficiency Measures on Workers' Formality Status

	Casual Worker in Any Firm of Size			Casual Worker in Formal Firm of Size		
	(1) All	(2) <10 Workers	(3) ≥ 10 Workers	(4) All	(5) <10 Workers	(6) ≥ 10 Workers
Panel A: Average Age of Pending Cases (Years)						
Average Age of Pending Cases	-0.0329	0.0184	-0.106*	-0.0213	-0.0494	-0.00238
	(0.0334)	(0.0297)	(0.0490)	(0.0257)	(0.0954)	(0.0230)
Mean Dep. Var.	0.75	0.90	0.52	0.31	0.47	0.26
First Stage F	9.76	9.76	9.76	0.00	0.00	0.00
Observations	22,522	13,284	9,238	8,045	2,463	5,582
Panel B: Disposition Time (Years)						
Disposition Time	0.0114	-0.00894	0.0400*	0.0128	0.0301	0.00185
	(0.0142)	(0.0135)	(0.0172)	(0.0144)	(0.0317)	(0.0139)
Mean Dep. Var.	0.74	0.90	0.52	0.31	0.47	0.26
First Stage F	4.71	4.71	4.71	0.23	0.23	0.23
Observations	18,309	10,621	7,688	6,594	1,974	4,620
Panel C: Backlog (Thousand Cases)						
Backlog	-0.00114	0.000784	-0.00304*	-0.000770	-0.00416	-0.0000789
	(0.00114)	(0.00125)	(0.00126)	(0.000931)	(0.00834)	(0.000760)
Mean Dep. Var.	0.75	0.90	0.52	0.31	0.47	0.26
First Stage F	0.03	0.03	0.03	0.20	0.20	0.20
Observations	22,522	13,284	9,238	8,045	2,463	5,582
Region FE	✓	✓	✓	✓	✓	✓
Covariates	✓	✓	✓	✓	✓	✓

Note: This table displays estimation results of IV regressions of workers' formality status in 2009 on several court efficiency measures in the workers' district by firms of a specific size. The excluded instrument is the mean share of occupied courtrooms in district and session courts over the period 2004-2008 per district. *Average Age of Pending Cases* is the average age of all pending cases at the end of the civil year. *Disposition Time* is the ratio of the number of unresolved cases and the number of resolved cases in district and session courts. *Backlog* denotes the number of pending cases older than one year at the end of the civil year. Included covariates are the district's share of inhabitants being part of a scheduled caste in 2011, the district's share of inhabitants being literate, and two dummy variables equal to one if the district is mainly urban or semi-urban. *Region FE* represents dummies for six broad regions; Central, East, North, Northeast, South, and West. Robust standard errors are clustered at the industry-type $\times$ state level. Stars *, ** and *** indicate that the p-value is below 0.1, 0.05 and 0.01 respectively.

Table A7: Statutory Tax Rates on Salaries and Revenues in Two Indian Cities

	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[a]	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

[a] In 2008, the central sales tax rate was decreased from 3% to 2%.

Note: This table displays taxes rates and mandatory contributions in Bengaluru and Noida in 2019.

E Data Appendix

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.

States are classified into six broad regions, defined as follows. *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 A8: Sample Selection Due to Intermediate Steps per Dataset and -Source

	Number of Districts							
	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
IV Defined	256	256	189	192	258	258	258	258

Note: This Table displays the number of districts covered by different data sources and data cleaning steps. 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 courtrooms filled in District and Sessions courts.