

The Impact of Contract Enforcement Costs on Outsourcing and Aggregate Productivity

Johannes Boehm

▶ **To cite this version:**

Johannes Boehm. The Impact of Contract Enforcement Costs on Outsourcing and Aggregate Productivity. Review of Economics and Statistics, 2020, forthcoming, 10.1162/rest_a_00940 . hal-03566762

HAL Id: hal-03566762

https://sciencespo.hal.science/hal-03566762

Submitted on 11 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MISALLOCATION IN THE MARKET FOR INPUTS: ENFORCEMENT AND THE ORGANIZATION OF PRODUCTION*

JOHANNES BOEHM AND EZRA OBERFIELD

The strength of contract enforcement determines how firms source inputs and organize production. Using microdata on Indian manufacturing plants, we show that production and sourcing decisions appear systematically distorted in states with weaker enforcement. Specifically, we document that in industries that tend to rely more heavily on relationship-specific intermediate inputs, plants in states with more-congested courts shift their expenditures away from intermediate inputs and have a greater vertical span of production. To quantify the effect of these distortions on aggregate productivity, we construct a model in which plants have several ways of producing, each with different bundles of inputs. Weak enforcement exacerbates a holdup problem that arises when using inputs that require customization, distorting both the intensive and extensive margins of input use. The equilibrium organization of production and the network structure of input-output linkages arise endogenously from the producers' simultaneous cost-minimization decisions. We identify the structural parameters that govern enforcement frictions from cross-state variation in the first moments of producers' cost shares. A set of counterfactuals show that enforcement frictions lower aggregate productivity to an extent that is relevant on the macro scale. *JEL* Codes: E23, O11, F12.

I. INTRODUCTION

Weak contract enforcement hinders firm-to-firm trade and distorts production decisions. For example, a manager who

*We are grateful to Pol Antràs, several anonymous referees, David Baqaee, Hugo Hopenhayn, Sam Kortum, Eduardo Morales, Jacopo Ponticelli, Steve Redding, Matt Rognlie, Meredith Startz, Felix Tintelnot, Gustavo Ventura, Chris Woodruff, and various seminar participants for helpful suggestions. Anusha Guha, Hélène Maghin, and Juan Manuel Castro provided excellent research assistance. We are grateful to Daksh India for sharing data, and to CEPR/PEDL and the Sciences Po/Princeton partnership for financial support. Boehm thanks the International Economics Section at Princeton and Oberfield thanks the Stanford Institute for Economic Policy Research for hospitality during the later stages of the project. This document is an output from the research initiative "Private Enterprise Development in Low-Income Countries" (PEDL), a program funded jointly by the Centre for Economic Policy Research (CEPR) and the Department for International Development (DFID), contract reference PEDL_LOA_4006_Boehm. The views expressed are not necessarily those of CEPR or DFID. All mistakes are our own.

© The Author(s) 2020. Published by Oxford University Press on behalf of the President and Fellows of Harvard College. All rights reserved. For Permissions, please email: journals.permissions@oup.com

The Quarterly Journal of Economics (2020), 2007–2058. doi:10.1093/qje/qjaa020.
Advance Access publication on June 16, 2020.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

cannot rely on courts for timely and cheap enforcement may need to purchase low-quality substitutes from a family member, vertically integrate the production process, or switch to a different technique that avoids the bottleneck input. Regardless of the chosen alternative, she will find herself producing at a higher cost. Collectively, the micro distortions induced by weak enforcement alter the equilibrium network structure of production and reduce aggregate productivity.

This article studies theoretically and empirically how weak legal institutions—more precisely, slow contract enforcement due to congestion of the courts—shape the organization of production. We develop a framework that allows us to use detailed micro production data to quantify the impact of these frictions on aggregate productivity.

We study contract enforcement frictions in the context of the Indian manufacturing sector. India is a country with infamously slow and congested courts: the World Bank (2016) currently ranks India 172nd (out of 190) when it comes to the enforcement of contracts, behind countries such as the Democratic Republic of Congo (171st) and Zimbabwe (165th). Around 6 million of the 22 million pending cases are older than five years, and while India's Law Commission has been advocating vast reforms for several decades, these reforms have not been implemented, and pendency ratios have not decreased. At the same time, India's liberalization and growth has spurred demand for timely enforcement of contracts.

Using plant-level data from India's Annual Survey of Industries, we document several facts about how court congestion alters plants' input choices. Although there is an enormous amount of heterogeneity in the input bundles plants use even within narrowly defined (five-digit) industries,[1] the bundles differ in systematic ways related to the quality of courts. To focus on these differences, we differentiate inputs that are relatively homogeneous and standardized from those that require customization or are relationship-specific, using the classification from Rauch (1999). Users (or potential users) of relationship-specific inputs

1. Some of this heterogeneity reflects different organizational and technological choices. As an example, a plant that produces frozen chicken may purchase live chickens and slaughter and freeze them; or it may purchase chicken feed and raise, slaughter, and freeze the chicken at the same vertically integrated plant. Other examples indicate horizontal technological choices, for example, aluminum can be produced from bauxite or from aluminum scrap.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

are most likely to benefit from better formal enforcement of supplier contracts.

Our first fact is that in states where courts are more congested, plants in industries that typically rely on relationship-specific intermediate inputs have systematically lower cost shares of intermediate inputs. Second, we show that where courts are slower, the composition of plants' intermediate-input bundles is tilted toward homogeneous inputs. Third, we construct a measure of the vertical span of production of plants designed to capture the number of sequential production steps performed by the plant. We show that where courts are more congested, plants in industries that typically rely on relationship-specific inputs tend to have larger vertical spans of production; that is, more sequential production steps are performed within plants.

We use a variety of strategies to alleviate concerns that these patterns arise from omitted factors or through reverse causality. We control for a range of fixed effects and interactions with state and industry characteristics to ensure that court congestion is not standing in for other state characteristics. We also confirm these results using an instrumental variable strategy that exploits the historical origins and structure of the Indian judiciary. Newly created courts tend to be fast and accumulate backlogs over time. As a result, the oldest High Courts, which were set up by the British in the nineteenth century, are the most congested, and newer courts, which have often been created because of new states being carved out of existing ones, are typically faster. We therefore use the age of the High Court as an instrument for its congestion and argue that the nature of the political events surrounding the creation of new courts makes it less likely that the resulting congestion is correlated with unobserved determinants of plants' input mixes.

We construct a model to interpret these facts and quantitatively evaluate the ramifications of these distortions for aggregate productivity. Our model is written to speak to the patterns of intermediate-input use among Indian manufacturing plants. We see that plants use different mixes of intermediate inputs to produce the same output. It is likely that much of this heterogeneity—reflecting different organizational forms, technology, or variation in input prices—would arise even in the absence of distortions. The model incorporates a rich set of organizational, technological, and sourcing decisions so that we do not conflate distortions with other sources of heterogeneity.

Our model is a multi-industry general-equilibrium model of heterogeneous firms and intermediate-input linkages that form between them. Firms face a menu of technology/organizational choices (recipes) and draw suppliers along with match-specific productivities. Both primary inputs and relationship-specific inputs are subject to distortions that reflect weak contract enforcement. Each firm chooses the production technique and suppliers that minimize cost. The effective cost of an input depends on the match-specific productivity, the supplier's marginal cost, and the distortion. We model the enforcement distortion for each potential supplier as randomly drawn to reflect the idea that formal enforcement may only sometimes be relevant at the margin.[2] Weak enforcement has a direct impact on producers who use inputs that require contract enforcement, but may also lead firms to switch to suppliers with a higher cost or to an entirely different production technique with a different set of inputs. We think of changing the organization of production to increase the vertical span as one such option.

To make quantitative statements, we structurally estimate technological parameters and distributions of wedges that distort the use of relationship-specific intermediate inputs. Our identification strategy rests on two key properties of our model. First, the model has the implication that, among firms that in equilibrium use the same production recipe, average cost shares of each input among firms are invariant to factor prices but depend on distortions.[3] This is a weaker implication than the assumption typically made in the misallocation literature, which is that each plant's cost share is invariant to factor prices (Cobb-Douglas) and that all heterogeneity in cost shares is due to distortions. This property allows identification of distortions from variation in average cost shares across regions. We also make the conservative assumption that weak contract enforcement does not distort the use of homogeneous inputs. As a result, the structural estimating equations take a simple form that is similar to the

2. For example, formal enforcement may not be necessary if the buyer and supplier are engaged in a long-term relationship, are related, or share other social ties.

3. As in Houthakker (1955), the aggregate production function among those firms is Cobb-Douglas even though there may be considerable cross-sectional dispersion in cost shares. While Houthakker (1955) assumed that individual production functions are Leontief, we show that this result extends to any constant returns to scale production function in which inputs are complements.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

motivating reduced-form regressions. The identified structural wedges on relationship-specific intermediate inputs are therefore correlated with the observed congestion in the regional courts, in line with the results from the reduced-form regressions.

One issue that arises is that we must separate the impact of distortions from the possibility that plants in different states systematically use different recipes. This problem is closely related to those studied by Bonhomme and Manresa (2015) and Moon and Weidner (2015); we want to identify the effect of contract distortions and recipe technologies, but we do not directly observe which plants use which recipes. Like Bonhomme and Manresa (2015), we use an iterative procedure that alternates between (i) estimating recipe technologies and distortion parameters given an assignment of plants to recipes, and (ii) a clustering procedure that assigns plants to recipes given distortion parameters. Both large- and small-sample Monte Carlo simulations show that, to a reasonable degree of accuracy, the procedure is able to recover the distortion parameters.

Our results suggest that courts may be important in shaping aggregate productivity. Having estimated the model parameters, we conduct counterfactuals to investigate the role of contracting frictions. For each state we ask how much aggregate productivity of the manufacturing sector would rise if court congestion were reduced to be in line with the least congested state. On average across states, the boost to productivity is roughly 4%, and the gains for the states with the most congested courts are roughly 8%. Distortions on relationship-specific inputs are only imperfectly explained by our measured congestion, and if one were to halve the distortions on relationship-specific inputs, the average state would see a 7.1% increase in productivity (up to more than 20% for the most distorted state).

Our model builds on recent models of firm linkages in general equilibrium that include Oberfield (2018), Eaton, Kortum, and Kramarz (2015), Lim (2018), Lu, Mariscal, and Mejia (2013), Chaney (2014), Acemoglu and Azar (2020), Taschereau-Dumouchel (2020), and Tintelnot et al. (2017),[4] and uses aggregation techniques pioneered by Houthakker (1955) and Jones (2005). We model the technology choice and choice of organization concurrently with the sourcing decision, motivated by evidence

4. These are also closely related to models of global value chains and global sourcing such as Costinot, Vogel, and Wang (2012), Fally and Hillberry (2018), Antràs and de Gortari (2017), and Antràs, Fort, and Tintelnot (2017).

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

that increased access to intermediate inputs has a productivity-enhancing effect (e.g., Pavcnik 2002; Goldberg et al. 2010; Khandelwal and Topalova 2011; Bas and Strauss-Kahn 2015). As in Grossman and Helpman (2002), one producer's choice of organization depends on the industry environment and the choices of other producers.

Our article is also closely related to the literature on misallocation in developing countries (Restuccia and Rogerson 2008; Hsieh and Klenow 2009; Hopenhayn 2014). Several papers have studied distortions to the use of intermediate inputs, for example, Jones (2013), Bartelme and Gorodnichenko (2015), Caprettini and Ciccone (2015), Fadinger, Ghiglino, and Teteryatnikova (2016), Caliendo, Parro, and Tsyvinski (2017), Liu (2019), Osotimehin and Popov (2020), Baqaee and Farhi (2020), Bigio and La'O (2020), and Boehm (forthcoming). These papers typically posit industry-level production functions and use industry-level data. Our approach of identifying wedges from factor shares (in our case, intermediate-input expenditure shares) extends the work of Hsieh and Klenow (2009) along three key dimensions. First, we relate the estimated wedges to the quality of Indian state-level institutions, which allows us to draw policy conclusions from our exercise. Second, we confront the fact that firms produce in very different ways even in narrowly defined industries by explicitly modeling this heterogeneity; we allow firms to choose among several types of technologies (recipes) in the theory and identify these recipes in the data through the application of techniques from statistics/data mining.[5] Third, we identify wedges from systematic differences in first moments, which helps alleviate concerns about mismeasurement being interpreted as misallocation.[6] In fact, our model predicts that even in the absence of distortions, firms using the same broad technology would use inputs with varying intensities.

The article is related to the literature on legal institutions and economic development (La Porta et al. 1997; Djankov et al. 2003; Acemoglu and Johnson 2005; Nunn 2007; Levchenko 2007; Acemoglu, Antràs, and Helpman 2007; Laeven and Woodruff 2007; Mukoyama and Popov 2020; among

5. Similarly, Asturias and Rossbach (2019) cluster Chilean manufacturing firms based on their capital, labor, and intermediate-input usage and argue that if the clusters are interpreted as representing different technologies, ignoring this heterogeneity would lead to an overstatement of the extent of misallocation.

6. See Bils, Klenow, and Ruane (2017) and Rotemberg and White (2019).

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

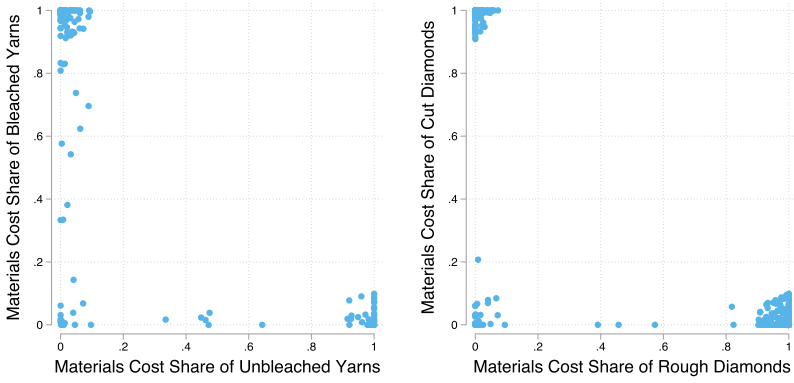
many others). Ponticelli and Alencar (2016) and Chemin (2012) argue that better courts reduce financial frictions. Amirapu (2017) shows that where district courts in India are more congested, firms in industries that relied on relationship-specific inputs grew faster. Johnson, McMillan, and Woodruff (2002) provide survey evidence that reduced trust in courts makes firms that rely on relationship-specific inputs less likely to switch suppliers. By embedding a contracting friction into a general equilibrium model, we explore its quantitative importance for aggregate outcomes. Boehm (forthcoming) characterizes the impact of weak enforcement on aggregate productivity, using cross-country differences in input-output tables to show that weak legal institutions have a larger effect on industry pairs that are more vulnerable to holdup problems.

II. Input Use among Indian Manufacturing Plants

II.A. Intermediate-Input Use

We use data from the 2000/01 to 2012/13 rounds of the Annual Survey of Industry (ASI), the official annual survey of India's formal manufacturing sector. The ASI is a panel that covers all establishments with more than 100 employees, and every year a fifth of all establishments with more than 20 employees (or more than 10 if they use power). The ASI's unique feature is that it contains detailed product-level information on each plant's intermediate inputs and outputs. Product codes are at the five-digit level, of which there are around 5,200 codes in their classification. The product classification remains largely unchanged during the years 2000/01 to 2009/10. The rounds 2010/11 to 2012/13 use a different (but similar) product classification, and we bring product-level data to the classification of the earlier years using the official concordance table published by the Ministry of Statistics. Online Appendix A contains more details on the data and a description of our sample.

One striking feature of the data is that even in narrowly defined industries, plants produce using very different input bundles. Figure I shows two examples that are particularly clear. Among producers of bleached cotton cloth and polished diamonds, output is made using different sets of inputs. Although we believe that much of the heterogeneity in organization and input bundles is not associated with inefficiencies and would arise naturally, Section II.C shows that some of the differences are systematically related to court congestion.



(A) Input mixes for Bleached Cotton Cloth (63303) (B) Input mixes for Polished Diamonds (92104)

FIGURE I

Heterogeneity in Input Mixes within Narrow Industries

The figure shows the variation in plants' input mixes for single-product plants that produce only bleached cotton cloth (Panel A) or polished diamonds (Panel B). Each dot represents a plant-year observation; the coordinates on the horizontal and vertical axes correspond to the materials cost share of two different types of intermediate inputs. For example, a plant on the top left of Panel B produces polished diamonds entirely from cut diamonds (therefore doing just the polishing); a plant on the bottom right produces polished diamonds entirely from rough diamonds (therefore doing both cutting and polishing themselves). Observations on the bottom left mostly produce their output from unbleached cloth (Panel A) or industrial diamonds (Panel B). Points have been jittered to improve readability.

Intermediate inputs vary in their degree to which buyers and sellers are subject to holdup problems. Producers of goods that are tailored to a particular buyer (relationship-specific) may find that buyers refuse to pay for the supplied good, knowing that they are useless to anyone but themselves (Iyer and Schoar 2008). We use the Rauch (1999) classification that divides goods into homogeneous goods (those that are traded on organized exchanges or for which a reference price exists) and relationship-specific goods (the remainder). Holdup problems are more likely to arise with relationship-specific inputs. At the same time, timely and cheap enforcement of contracts in a court of justice is a way to alleviate these holdup problems. Hence, firms rely more heavily on judicial institutions to enforce supplier contracts when trading goods belonging to the latter category (Johnson, McMillan, and Woodruff 2002).

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

II.B. Court Congestion in India

Among all the ills of the Indian judicial system, its slowness is perhaps the most apparent. As of 2017, about 9% of pending cases in district courts and 6% of pending cases in High Courts are older than 10 years.[7] Some cases make international headlines, such as in 2010, when the Bhopal District Court convicted eight executives of death by criminal negligence during the 1984 Bhopal gas leak, which killed thousands of people. The conviction took place some 25 years after the disaster; one of the eight executives had already passed away, and the remaining seven appealed the conviction.[8]

The slowness of the Indian courts is at least partly due to the uneven distribution of workload across its three tiers.[9] The lowest tier is the Subordinate (District) Courts, which have courthouses in district capitals and major cities.[10] The next tier are the High Courts, of which there generally exists one for each state and which have both appellate and original jurisdiction over cases originating from their state (and sometimes an adjacent union territory). High Courts also administer subordinate courts in their jurisdiction. The highest tier is the Supreme Court of India. All three tiers are heavily congested, with district courts facing the additional problems that judges are often inexperienced and are regarded as being less able or willing to make the right decision. While contract cases between firms should, in principle, be filed at the district level, litigants typically bypass this step by claiming an infringement of their fundamental rights or appealing to the constitution of India, in which case they are permitted to file the claim directly at a high court.[11] High Court judges, often taking a dim view of the subordinate judiciary,

7. Figures for district courts are from the National Judicial Data Grid from 2017. Figures for High Courts are based on authors' calculations from the Daksh data (see below).

8. James Lamont, "Painfully Slow Justice over Bhopal," *Financial Times*, June 7, 2010.

9. See Robinson (2016) for an overview of the Indian judiciary. Hazra and Debroy (2007) discuss its problems in relation to economic development.

10. Districts are the administrative divisions below states. Between 2001 and 2010 there were around 620 districts and 28 states in India. Union territories are small administrative divisions (typically cities or islands) that are under the rule of the federal government, as opposed to states, which have their own government.

11. Some High Courts, such as the High Courts of Bombay, Calcutta, Madras, and Delhi, even allow civil cases to be filed directly whenever the claim exceeds a certain value.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

tend to accommodate this practice. The result is that the Indian judiciary is relatively heavy in its upper levels, with only the simplest cases being dealt with in the subordinate courts.[12] For better or worse, the quality of the higher judiciary determines whether and how contracts can be enforced.

We construct a measure of court congestion from microdata on pending civil cases in High Courts, which the Indian nongovernmental organization Daksh collects from causelists and other court records (Narasappa and Vidyasagar 2016). These records show the status and age of pending and recently disposed cases, along with characteristics of the case, such as the act under which the claim was filed or a case type categorization. Our measure of high court congestion is the average age of pending civil cases in each court, at the end of the calendar year 2016. Whenever a high court has jurisdiction over two states and a separate bench in each of them (such as the Bombay High Court, which has jurisdiction over Maharashtra and Goa), we construct the statistic by state. We prefer this measure over existing measures of the speed of enforcement, such as pendency ratios published by the High Courts, which suffer from the problem that different high courts measure pendencies in vastly different ways (as emphasized by the Law Commission of India 2014).

The average age of pending civil cases varies substantially across high courts—from less than one year in Goa and Sikkim, to about four and a half years in Uttar Pradesh and West Bengal. The cross-state average is two and a half years. These differences can be seen in Figure II.

The problems of the Indian judiciary are not a recent phenomenon and have not gone unnoticed. Throughout the modern history of India as an independent nation, the Law Commission of India has pointed out the enormous backlogs and arrears of cases (14th report, 1958; 79th report, 1979; 120th report, 1987; 245th report, 2014), and suggested many policies to alleviate the situation. The vast majority of these proposals have not been adopted, and the few exceptions seem to have had little effect. Overall, the backlogs have slowly but continually accumulated.

12. Between 2010 and 2012, about 40% of all disposed cases in subordinate courts were related to traffic tickets, another 7% related to bounced cheques (Law Commission of India 2014).

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

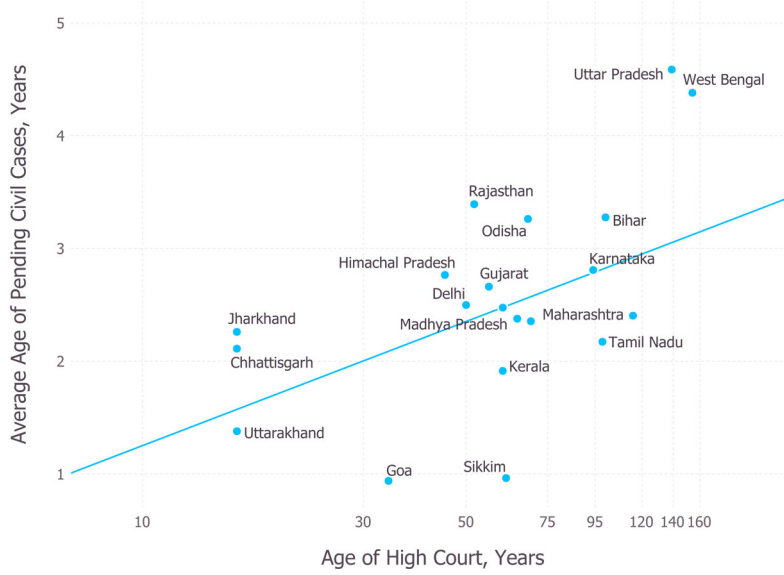


FIGURE II

Age of the High Court and Speed of Enforcement

The main explanation for why court speed varies so much across states lies in the history of India's political subdivisions. The first high courts (Madras, Bombay, and Calcutta) were set up by the British in the 1861 Indian High Courts Act and served as the precursor for India's post-Independence high courts. Upon Independence, India was divided into a number of federated states, with the constitution of India (1947) mandating a high court for each state. Throughout the twentieth century and beyond, India has frequently subdivided its states, often because of ethno-nationalist movements. These subdivisions were often accompanied with new high courts being set up, which then start without any existing backlog of cases.[13] The age of the high court is a strong determinant of its speed of enforcement (see Figure II; the F-statistic in that regression is 11.1). We later use the age of high courts as an instrument for its level of congestion.

13. Table A.1 in Online Appendix A summarizes the reasons for the high court being set up or the state being formed.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

II.C. Motivating Facts

We turn to documenting the correlation between court congestion and plants' intermediate input use. For the sake of comparability we restrict our attention here to single-product plants.[14]

FACT 1. In states with more-congested courts, cost shares of intermediate inputs are relatively lower in industries that tend to rely more on relationship-specific intermediate inputs.

Table I shows regressions of the plants' materials cost share on an interaction of court congestion (as measured by the average age of pending cases in the high court of the state in which the plant is located[15]) and the industry's reliance on relationship-specific inputs.[16] The interaction term has a negative and significant coefficient, showing that plants' materials cost shares decline more steeply with court congestion in industries that tend to rely more heavily on relationship-specific inputs. The magnitude in column (1) indicates that for each additional year of court congestion, plants' materials share of cost declines by 1.67 percentage points more in industries that rely on relationship-specific inputs than in industries that rely on standardized inputs.

A primary concern in this specification is that court congestion is standing in for the level of development, or that the level of development is correlated with the relative productivity of industries that rely on relationship-specific inputs. Column (1) includes district fixed effects. Column (2) controls for the interaction of relationship specificity with district income per capita, and column (3) adds controls for the interaction of relationship specificity with a variety of state characteristics, including measures of trust, corruption, linguistic fragmentation, and fragmentation

14. One difficulty that arises when studying multiproduct plants is that we do not observe which inputs are used to produce each product. Nevertheless, the results in this section are quantitatively similar when we include multiproduct plants and assign the plant to the category of its highest-revenue product.

15. In principle, firms can bring their cases to any court; the court decides whether it has jurisdiction over the case. We believe that in the case of India this practice is limited, as state borders are often also language borders. More generally, this would be a form of measurement error in the independent variable that would bias the coefficient toward 0.

16. Following Nunn (2007), we measure an industry's reliance on relationship-specific inputs at the national level by computing the fraction of intermediate-input expenditures spent on relationship-specific inputs across all plants in the industry. See Online Appendix A.1 for details.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

TABLE I

MATERIALS SHARES AND COURT CONGESTION (FACT 1)

	Dependent variable: Materials expenditure in total cost					
	(1)	(2)	(3)	(4)	(5)	(6)
Avg age of civil cases * Rel. spec.	−0.0167**	−0.0129*	−0.0118*	−0.0156+	−0.0201*	−0.0212**
	(0.0046)	(0.0051)	(0.0053)	(0.0085)	(0.0082)	(0.0078)
LogGDPC * Rel. spec.		0.0114	0.0102		0.00710	0.00556
		(0.0086)	(0.0091)		(0.0095)	(0.0096)
Rel. spec. × State controls			Yes			Yes
Five-digit industry FE	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	IV	IV	IV
R^2	0.480	0.482	0.484	0.480	0.482	0.484
Observations	208,527	199,544	196,748	208,527	199,544	196,748

Notes. Standard errors are in parentheses, clustered at the state × industry level. + $p < .10$, * $p < .05$, ** $p < .01$. "Rel. spec. × State controls" are interactions of trust, language Herfindahl, caste Herfindahl, and corruption with relationship specificity. The sample consists of single-product plants only.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

by caste. While the coefficients (reported in Online Appendix C) suggest that ethnolinguistic homogeneity facilitates the use of relationship-specific inputs, this appears to be orthogonal to court congestion. Finally, columns (4) to (6) employ an instrumental variables strategy that we discuss below in Section II.D.

FACT 2. In states with more-congested courts, intermediate input bundles are tilted toward standardized intermediate inputs.

Our first fact related court congestion to how plants divided their expenditures between intermediate and primary inputs. We next study how the composition of plants' intermediate-input baskets covaries with court congestion. Table II shows that in states where courts are faster, plants' intermediate-input baskets are tilted toward relationship-specific intermediate inputs. This correlation remains statistically significant when controlling for district income per capita and other state characteristics.

FACT 3. In states with more-congested courts, plants in industries that tend to rely more on relationship-specific intermediate inputs have larger vertical spans of production.

A low materials share suggests that a plant may be doing more consecutive steps in the production process themselves. For example, a car producer that assembles components may also manufacture those components in the same facility. The regressions in Table III show how court congestion is related to the vertical span of production, that is, to how many consecutive production steps are performed within the plant. We first construct a measure of the "vertical distance" between an output good ω to an input ω'. This is intended to capture the typical number of "steps" between the use of ω' and the production of ω, where we define a step to be the activity performed by a single plant.[17] Finally, for each single-product plant, our measure of vertical span is the expenditure-weighted average of the distance

17. We construct national input-output tables using our plant-level data. For each output good ω and input good ω', we take a weighted average of the number of steps along any path from ω' to ω, weighted by the product of the input-output shares along that path, excluding any path which cycles. This measure is similar to $Upstreamness_{ij}$ of Alfaro et al. (2019). Online Appendix B gives the precise mathematical definition of vertical distance.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

TABLE II

INPUT MIX AND COURT CONGESTION (FACT 2)

	Dependent variable: $\frac{X_j^R}{X_j^R+X_j^H}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Avg age of civil HC cases	−0.00547*	−0.00621**	−0.00530*	−0.0144**	−0.0146**	−0.0167**
	(0.0022)	(0.0023)	(0.0024)	(0.0044)	(0.0044)	(0.0045)
Log district GDP/capita		−0.00389	−0.00384		−0.00912+	−0.00980+
		(0.0045)	(0.0046)		(0.0051)	(0.0051)
State controls			Yes			Yes
Five-digit industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	IV	IV	IV
R^2	0.441	0.446	0.449	0.441	0.446	0.449
Observations	225,590	204,031	199,339	225,590	204,031	199,339

Notes. The dependent variable is the share of relationship-specific inputs in total materials cost of plant j. "State controls" are trust, language Herfindahl, caste Herfindahl, and corruption. The sample consists of single-product plants only. Standard errors are in parentheses, clustered at the state × industry level. + $p < .10$, * $p < .05$, ** $p < .01$.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

TABLE III

VERTICAL SPAN OF PLANTS AND COURT CONGESTION (FACT 3)

	Dependent variable: Vertical span					
	(1)	(2)	(3)	(4)	(5)	(6)
Avg age of civil cases * Rel. spec.	0.0195^{+}	0.0269*	0.0280*	0.0292	0.0314^{+}	0.0368*
	(0.011)	(0.012)	(0.012)	(0.019)	(0.018)	(0.018)
LogGDPC * Rel. spec.		0.0464*	0.0288		0.0491*	0.0330
		(0.022)	(0.024)		(0.023)	(0.024)
Rel. spec. × State controls			Yes			Yes
Five-digit industry FE	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	IV	IV	IV
R^2	0.443	0.451	0.453	0.443	0.451	0.453
Observations	163,334	156,191	154,021	163,334	156,191	154,021

Notes. Standard errors are in parentheses, clustered at the state × industry level. $^{+}$ $p < .10$, * $p < .05$, ** $p < .01$ "Rel. spec. × State controls" are interactions of trust, language Herfindahl, caste Herfindahl, and corruption with relationship specificity. The sample consists of single-product plants only.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

from the plant's output ω to its intermediate inputs:

$$\text{verticalSpan}_j = \sum_{\hat{\omega}\in\Omega} \frac{X_{j\hat{\omega}}}{\sum_{\tilde{\omega}\in\Omega} X_{j\tilde{\omega}}} \text{verticalDistance}_{\omega\hat{\omega}},$$

where $X_{j\hat{\omega}}$ is plant j's expenditure on input $\hat{\omega}$ and ω is j's output. A longer vertical span indicates that the plant uses inputs that are typically further upstream, and suggests that the plant is performing more steps in-house. Table III shows that plants' vertical spans of production increase more sharply with court congestion in industries that tend to rely more heavily on relationship-specific inputs.[18]

II.D. Endogeneity and the Historical Determinants of Indian Court Efficiency

The main caveat in the foregoing regressions is the concern that there are unobserved covariates of court congestion that may also affect the cost of plants' inputs, and thereby their input shares. The simplest version is reverse causality. In principle, the bias from reverse causality could be positive or negative. Suppose that, for exogenous reasons, a state had many plants that produced using relationship-specific inputs. The disputes that arise may cause the courts to be congested. Or alternatively, the state may respond to the disputes that arise by spending resources to reduce congestion. Either of these would be problematic for interpreting the regressions as a causal relationship.

Although we believe reverse causality is unlikely to arise—the fraction of cases related to firm-to-firm trade is relatively low[19]—it is difficult to rule out other factors that may influence both court congestion and usage of relationship-specific inputs.

18. Although the nature of our data allows us to speak to the activities of individual plants rather than of firms, it is likely that contracting frictions affect the boundaries of firms as well. While we do not know whether two plants belong to the same firm, our data contain an indicator of whether a plant is standalone or part of a multiplant firm. In Online Appendix C.2, we repeat Tables I, II, and III but compare standalone plants to plants that belong to multiplant firms. Note that interpreting these regressions is not straightforward because whether a plant is part of a multiplant firm is itself an endogenous outcome.

19. The fraction of cases that is related to the enforcement of supplier contracts is hard to pinpoint exactly because courts classify cases very broadly and in different ways across states. We know, however, that they account for less than 5%, 14%, and 7% of the pending cases in the High Courts of Allahabad, Bombay, and Calcutta, respectively.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

We therefore employ an instrumental variables strategy that uses the historical determinants of congestion. As discussed in Section II.B, courts have been continually accumulating backlogs throughout the twentieth century. At certain points, however, states were split or reorganized, mostly in response to ethno-nationalist movements. In the course of these reorganizations, new high courts were set up, which initially started with a clean slate but were, like existing courts, understaffed and started accumulating backlogs. The time since their founding—the court's age—is therefore a strong predictor for the current backlog, which in turn determines the present-day speed of enforcement. Our instrumental variable for the speed of enforcement is hence the (log) age of the high court, and the instrument for an interaction of an industry-level variable with court speed is the interaction of the industry-level variable with the log age of the court. Figure II shows the strong correlation between court age and speed of enforcement.

Columns (4) to (6) of Tables I, II, and III repeat the regressions while instrumenting for the speed of enforcement. The point estimates of the coefficient of the interaction term is usually slightly larger than the OLS estimates.

There are a few reasons that the exclusion restriction may be violated. We argue that two candidates would lead us to conclude that the true relationships are stronger than reported in the IV regressions. First, new states tend to be relatively poor and have low state capacity. Thus the usual concern that a high level of development causes firms to use more sophisticated technologies that use relationship-specific inputs would cause the newer states to have higher use of homogeneous inputs. Alternatively, it may be that when a state splits, many firms lose their suppliers and must switch. It may be easy to find a new supplier of homogeneous inputs, whereas it might be harder to find a supplier of relationship-specific inputs. This channel would also cause newer states to be more intensive in homogeneous inputs. In either case, the true relationship would be stronger than reported. A third reason that the exclusion restriction might be violated is that newly formed states might be more ethnically homogeneous, so that newer courts might be correlated with better informal enforcement. This is a particular concern here because ethno-nationalist conflicts are a primary reason that states split. Nevertheless, we can control for states' ethnic fragmentation—or more specifically the interaction of relationship specificity with

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

various measures of ethnic fragmentation—as we do in columns (3) and (6) of Tables I, II, and III. Controlling for these has little impact on the main coefficients of interest.

In addition to the cross-sectional regressions shown above, we bring evidence from time variation in the degree of court congestion in Online Appendix C.3. We first show results from two instances in which high courts set up new and fast benches in remote areas during our sample period. In addition, we use variation in court pendency ratios over time. In both cases we include fixed effects to identify the relationship from time variation in court congestion only. Although neither experiment is perfect—with the court expansion the sample is small and some estimates are imprecise, whereas with pendency ratios it is not clear what drives the variation—the results are consistent with our baseline cross-sectional evidence.

A final concern is that an industry's reliance on relationship-specific inputs is correlated with other industry characteristics such as capital intensity, skill intensity, upstreamness, or tradability. In Online Appendix C.1 we show that the estimates are robust to controlling for the respective interactions of court congestion with each of these industry characteristics.[20]

II.E. Inferring Properties of Contracting Frictions

What do these regressions tell us about the form of contracting frictions? The literature on contracting frictions (e.g., Antràs 2003) has emphasized that holdup problems may result in transactions in which the seller shades on the quantity or quality of the inputs. These can be modeled in different ways that would show up differently in the data. We explore the implications and empirical validity of these in Online Appendix D.

In particular, contracting frictions could be modeled as resulting in an increase in the shadow cost of an input (Hsieh and Klenow 2009). A testable implication of this formulation is that the distortion should raise the buyer's ratio of revenue to total cost. In our setting, we find that plants that are subject to larger wedges—those in industries that tend to use relationship-specific inputs in states with congested courts—have lower

20. Online Appendix C.4 and C.5 show the correlation of distortions with other establishment characteristics, including import decisions, age, size, and the number of products.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

revenue-cost ratios, in contrast to the prediction from a quantity distortion.

Alternatively, contracting frictions could raise the effective price of the input (i.e., an iceberg cost). To be consistent with our evidence, such a formulation would require that the elasticity of substitution between distorted and undistorted inputs is larger than 1. As we discuss in Online Appendix D, this contrasts with available evidence, both from the existing literature and from our setting.

Finally, it could be that distortions increase the effective cost of an input, but that the additional cost takes the form of additional expenditure on primary factors. For example, it may be that the buyer needs to pay workers to repair defective inputs. Such a formulation is consistent with the empirical evidence and we use it in the subsequent model.

III. Model

The previous section showed that imperfect contract enforcement alters the production decisions of manufacturing plants in India in systematic ways. We aim to quantify the impact of weak enforcement on the productivity of the manufacturing sector.

A commonly used approach to quantify the role of factor market frictions, pioneered by Hsieh and Klenow (2009), is to posit that plants in the same industry use a common Cobb-Douglas production function. This allows them to use dispersion of cost shares within industries to infer dispersion in marginal products of inputs. What gives this approach traction is the implication that in the absence of distortions, plants would have the same cost shares regardless of the factor prices they face. Deviations from this cost share would then be interpreted as the result of a distortion. This approach, however, extends poorly to the question of finding distortions in the use of intermediate inputs, where, as we have seen, variation in factor use along both intensive and extensive margins is widespread and unlikely to be driven entirely by distortions.[21] At the same time, departing from Cobb-Douglas would require reliable information on input prices that are comparable across plants and across locations, as well as a way to disentangle

21. We attempt such an exercise in Online Appendix H.3. There are several conceptual and practical obstacles. In any case, the observed variation in input use would lead to enormous implied distortions.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

distortions from differences in the level and factor bias of productivity. This is especially important in the context of intermediate inputs, where such information is typically unavailable or unreliable.

In this section we develop a model that we will use to evaluate the impact of contracting frictions. We will make two key identifying assumptions. First, our assumptions will imply that, at a certain level of aggregation, factor shares will be invariant to factor prices. This implication is related to but weaker than those made by Hsieh and Klenow (2009), which is that factor shares are invariant to factor prices at the firm level and that all dispersion is due to distortions. Second, we assume that imperfect contract enforcement distorts the use of relationship-specific inputs and of labor, but not the use of homogeneous inputs.[22] These two aspects of our model enable us to identify distortions in factor markets from factor shares alone (in our case, average factor shares, thereby not requiring price data), while still allowing plants to have production functions that deviate from Cobb-Douglas, and therefore have varying factor shares for reasons that are unrelated to distortions. We then estimate the importance of distortions on the use of relationship-specific intermediate inputs using these average factor shares. This ties back to the reduced-form evidence we have shown above: that patterns of plants' expenditure shares differ systematically across states, in a way that is correlated with a potential source of distortions.

The model gives a prominent role to two additional features that may be important for our quantification of contracting frictions: firm-to-firm trade and technology/organizational choice. First, there are many ways to avoid contracting frictions. Suppose a firm needed to use an input that required customization and therefore faced a holdup problem. In the face of weak formal contract enforcement, the firm might buy the intermediate input from a relative or rely on the repeated interactions of a long-term relationship. Such decisions, however, may come at a cost. A family member may not be the optimal supplier of an intermediate input, and if a firm is in a long-term relationship, it may pass up using new, more cost-effective inputs to remain in that

22. If contracting frictions also distorted the use of homogeneous inputs, the gains from reducing contracting friction would be larger than those we report.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

long-term relationship.[23] Such a firm's production cost is higher than it would be with better contract enforcement, even though there may not be a distortion associated with sourcing from that supplier. The higher production cost is therefore an indirect consequence of weak formal enforcement that arises when firms substitute away from suppliers with whom they would face a holdup problem. We can infer this indirect cost by incorporating this type of decision in the model, and estimating how easily firms can substitute across suppliers of the same input.

Second, as discussed in Section II.A, even in narrowly defined industries, firms produce in qualitatively different ways. As a simple example, consider two plants that produce polished diamonds, one that buys cut diamonds and another that buys rough diamonds and cuts and polishes them. It may well be the case that the latter's decision to have a larger vertical span of production and do both cutting and polishing was a consequence of a distortion. The two plants will use different production functions, and simply comparing their expenditure shares will not give a direct measure of the size of the distortion. The empirical implementation of our model must therefore be able to account for these differences when measuring the magnitude of the distortions.

III.A. The Environment

There is a set of industries Ω. For industry $\omega \in \Omega$, there is a mass of firms with measure J_ω that produce differentiated varieties. There is a representative household that inelastically provides a mass of labor with measure L and has nested CES preferences over all varieties in each industry, maximizing consumption of the bundle U defined as

$$U = \left[\sum_{\omega \in \Omega} \upsilon_\omega^{\frac{1}{\eta}} U_\omega^{\frac{\eta-1}{\eta}}\right]^{\frac{\eta}{\eta-1}}$$

$$U_\omega = \left[\int_0^{J_\omega} u_{\omega j}^{\frac{\varepsilon-1}{\varepsilon}} dj\right]^{\frac{\varepsilon}{\varepsilon-1}},$$

23. Lim (2018) documents that in each year, firms switch roughly 40% of suppliers, and Lu, Mariscal, and Mejia (2013) document ubiquitous switching of imported inputs among importers, suggesting gains from taking advantage of new opportunities that arise. Johnson, McMillan, and Woodruff (2002) show that those who distrust courts are less likely to switch suppliers, suggesting that weak enforcement inhibits this.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

where $u_{\omega j}$ is consumption of variety j in industry ω, υ_ω reflects the household's taste for goods in industry ω, η is the elasticity of substitution across industries, and ε is the elasticity of substitution across varieties within each industry.

There are several different ways to produce a good using different combinations of inputs. These different ways correspond to different production functions, which we call recipes. Denote the set of recipes to produce a good ω by ϱ_ω. Each recipe $\rho \in \varrho_\omega$ describes a production function $G_{\omega\rho}(\cdot)$ that can be used by any firm in industry ω to produce its good using labor l and a particular bundle of inputs $\hat{\Omega}^\rho$. We assume:

ASSUMPTION 1. For any recipe $\rho \in \varrho_\omega$, the production function $G_{\omega\rho}$ exhibits constant returns to scale and all inputs $(l, \hat{\Omega}^\rho)$ are complements.

Note that the different $G_{\omega\rho}$ may differ in their shape and their factor intensities and that they need not be Leontief.

Firm j in industry ω has random sets of productivity and supplier draws $\Phi_{j\rho}$ for each recipe $\rho \in \varrho_\omega$. We call these draws techniques. Each technique $\phi \in \Phi_{j\rho}$ is characterized by (i) a set of potential suppliers for each intermediate input, $\{S_{\hat{\omega}}(\phi)\}_{\hat{\omega}\in\hat{\Omega}^\rho}$; (ii) for each of those suppliers $s \in S_{\hat{\omega}}(\phi)$ an input-augmenting productivity and a distortion (which we discuss further below); and (iii) a labor-augmenting productivity $b_l(\phi)$. The input-augmenting productivity for supplier $s \in S_{\hat{\omega}}(\phi)$ consists of a match-specific component z_s that is specific to the supplier and a component $b_{\hat{\omega}}(\phi)$ that is common to all suppliers of input $\hat{\omega}$ in the set $S_{\hat{\omega}}(\phi)$.

Suppose that j produced its good using a technique of recipe ρ which used labor and intermediate inputs $\hat{\Omega}^\rho = \{\hat{\omega}_1, \ldots, \hat{\omega}_n\}$. If it chose to employ l units of labor and purchase $\{x_{s_1}, \ldots, x_{s_n}\}$ units of intermediate inputs from respective suppliers $s_1 \in S_{\hat{\omega}_1}, \ldots, s_n \in S_{\hat{\omega}_n}$, its output would be

$$y_j = G_{\omega\rho}\left(b_l l, b_{\hat{\omega}_1} z_{s_1} x_{s_1}, \ldots, b_{\hat{\omega}_n} z_{s_n} x_{s_n}\right).$$

Each technique is specific to the firm producing the output (the "buyer") and to the potential suppliers that might provide the intermediate inputs. In equilibrium, the firm chooses to produce using the technique and suppliers that is most cost-effective, which depends on the prices those suppliers charge and the input-augmenting productivities of the technique.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

Terms of trade among firms determine their choices of inputs, production decisions, and productivity. We also assume that sales of goods for intermediate use are priced at the supplier's marginal cost.[24] Firms engage in monopolistic competition when selling to the representative household and remit all profits to the household.

First, nature chooses the sets of techniques available to each firm. Then all firms simultaneously set prices and make their production decisions (i.e., choices of technique $\phi \in \bigcup_{\rho \in \mathcal{Q}_\omega} \Phi_{j\rho}$, suppliers, and inputs $l, x_{s_1}, \ldots, x_{s_n}$) to minimize cost, taking into account the decisions of others. All firms have perfect information about the economy's production possibilities and about other firms' choices. The probability distribution governing the set of techniques with which firm j can produce ($\{\Phi_{j\rho}\}_{\rho \in \mathcal{Q}_\omega}$) will be described below.

III.B. Contracting Enforcement

Enforcement of contracts facilitates the use of inputs that require customization and the use of labor. Imperfect enforcement introduces wedges between the effective cost to the buyer and the payment to the supplier. If an input requires customization, the supplier can shirk and provide a good that is imperfectly customized to the buyer. If this happens, the buyer needs to use extra labor to correct the defect. This is wasteful because the supplier has an absolute advantage in performing the customization. For each supplier, the buyer draws a random cost of enforcing the contract, which, by modulating the threat of enforcement, affects the equilibrium performance of the supplier and hence the extra labor the buyer needs to use to correct the defect. We discuss the full microfoundation in Online Appendix E.

24. One interpretation of this assumption is that in firm-to-firm trade, buyers have all of the bargaining power. For example, in a simpler environment, Oberfield (2018) characterizes an alternative market structure in which firm-to-firm trade is governed by bilateral trading contracts specifying a buyer, a supplier, a quantity of the supplier's good to be sold to the buyer, and a payment. Given a contracting arrangement, each entrepreneur makes her remaining production decisions to maximize profit. The economy is in equilibrium when the arrangement is such that no countable coalition of entrepreneurs would find it mutually beneficial to deviate by altering terms of trade among members of the coalition and/or dropping contracts with those not in the coalition. The terms of trade described here are one particular equilibrium in which buyers have all of the bargaining power.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

This game has a reduced form in which, for each potential supplier of a relationship-specific input $\hat{\omega}$, the buyer draws a random input wedge $t_x \in [1, \infty)$, from a distribution with CDF $T(t_x)$.[25] If the supplier's price is p_s per unit, the total cost to the buyer is $t_x p_s$, with p_s paid to the supplier and $(t_x - 1)p_s$ spent on extra labor. We model wedges as random and match specific because, for some suppliers (e.g., family members, those with whom the buyer is in a long-term relationship) the possibility of informal enforcement may mitigate any holdup problem; for others, the holdup problem may be more severe.

The distribution $T(t_x)$ summarizes the quality of enforcement. Perfect enforcement of contracts would imply that $t_x = 1$ for all suppliers. As discussed earlier, courts are not the only way to enforce contracts; contracts could be enforced informally through social punishments or reputation. t_x should be interpreted as the wedges that prevail after all forms of enforcement are exhausted.[26]

For completeness, we include the possibility that imperfect enforcement raises the cost of labor as well. If production is subject to the labor wedge t_l, then the firm needs to hire t_l workers to obtain one efficiency unit of labor, so that the effective cost of labor to the firm is wt_l. For simplicity we assume that t_l is the same across all firms.[27]

III.C. Production Decisions

For each technique, firm j draws a set of potential suppliers to provide each input. Each potential supplier $s \in S_{\hat{\omega}}(\phi)$ comes with an input-augmenting productivity draw and a wedge, so that the effective cost of using that supplier would be $\frac{t_{xs} p_s}{b_{\hat{\omega}}(\phi) z_s}$, which includes the cost of the extra labor needed to customize the

25. As we show in Online Appendix E, there is a one-to-one mapping between the enforcement cost and the equilibrium input wedge t_x.

26. For example, if formal enforcement would leave the wedge t_x^{formal} while informal enforcement would leave the wedge t_x^{informal}, then the parties would use whichever form of enforcement is better, that is, $t_x = \min\{t_x^{\text{formal}}, t_x^{\text{informal}}\}$. The argument extends in the obvious way if there are multiple ways of enforcing contracts informally. Improving the quality of courts would reduce the formal wedges and might alter the effectiveness of informal enforcement mechanisms if it worsens the informal arrangements that can be sustained.

27. Our counterfactuals focus on changes in the distribution of distortions that impede the use of relationship-specific intermediate inputs (T). Our identification strategy does not recover the labor wedge t_l.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

input. If j used technique ϕ, it would choose to use the supplier that delivered the lowest effective cost for input $\hat{\omega}$, so that its effective cost of that input would be:

$$\lambda_{\hat{\omega}}(\phi) \equiv \min_{s \in S_{\hat{\omega}}(\phi)} \frac{t_{xs}(\phi) p_s}{b_{\hat{\omega}}(\phi) z_s(\phi)}.$$

Similarly, the effective cost of labor when using technique ϕ is $\lambda_l(\phi) = \frac{t_l w}{b_l(\phi)}$. For the remainder, we normalize the wage to unity, $w = 1$.

The unit cost delivered by a technique depends on the effective cost of each input. Let $\mathcal{C}_{\omega\rho}(\cdot)$ be the unit cost function that is the dual of the production function $G_{\omega\rho}$, so that j's cost of producing one unit of output using technique ϕ would be $\mathcal{C}_{\omega\rho}(\lambda_l(\phi), \{\lambda_{\hat{\omega}}(\phi)\}_{\hat{\omega} \in \hat{\Omega}^\rho})$. Minimizing cost across all techniques, j's unit cost is

$$\min_{\rho \in \varrho(\omega)} \min_{\phi \in \Phi_{\omega j \rho}} \mathcal{C}_{\omega\rho}\left(\lambda_l(\phi), \{\lambda_{\hat{\omega}}(\phi)\}_{\hat{\omega} \in \hat{\Omega}^\rho}\right).$$

In words, firm j's unit cost equals that of the technique that delivers the lowest cost across all techniques of all recipes.

In this section, we specialize to particular functional-form assumptions. As we show below, with these assumptions, the model aggregates easily and allows us to use a transparent strategy to identify contracting frictions. The set of techniques available to each firm is random and governed by the following assumptions about the distributions of input-augmenting productivities.

ASSUMPTION 2. For a firm in industry ω,

i. Each supplier in the set $S_{\hat{\omega}}(\phi)$ is uniformly drawn from all firms that produce $\hat{\omega}$.

ii. For each technique ϕ that uses input $\hat{\omega}$, the number of suppliers in $S_{\hat{\omega}}(\phi)$ for whom the match-specific component of productivity is greater than z follows a Poisson distribution with mean

$$z^{-\zeta_{\hat{\omega}}}, \text{with } \zeta_{\hat{\omega}} = \begin{cases} \zeta_R, & \hat{\omega} \in \hat{\Omega}^\rho_R \\ \zeta_H, & \hat{\omega} \in \hat{\Omega}^\rho_H \end{cases},$$

where $\hat{\Omega}^\rho_R$ and $\hat{\Omega}^\rho_H$ are the subsets of inputs used in recipe ρ that are, respectively, relationship-specific and homogeneous.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

iii. Consider recipe $\rho \in \varrho_\omega$ which uses labor and the input bundle $\hat{\Omega}^\rho = (\hat{\omega}_1, \ldots, \hat{\omega}_n)$. For each plant, the number of techniques to produce using that recipe for which the common components of input-augmenting productivities strictly dominate[28] $b_l, b_{\hat{\omega}_1}, b_{\hat{\omega}_2}, \ldots, b_{\hat{\omega}_n}$ follows a Poisson distribution with mean

$$B_{\omega\rho} b_l^{-\beta_l^\rho} b_{\hat{\omega}_1}^{-\beta_{\hat{\omega}_1}^\rho} \ldots b_{\hat{\omega}_n}^{-\beta_{\hat{\omega}_n}^\rho}.$$

iv. There is a constant γ such that for each ω and each recipe $\rho \in \varrho_\omega$, $\beta_l^\rho + \beta_{\hat{\omega}_1}^\rho + \cdots + \beta_{\hat{\omega}_n}^\rho = \gamma$.

v. The following parameter restrictions hold for each $\hat{\omega}$: $\gamma > \varepsilon - 1$, $\gamma > \zeta_{\hat{\omega}} > \beta_{\hat{\omega}}^\rho$ where $\zeta_{\hat{\omega}}$ is ζ_R if $\hat{\omega}$ is relationship-specific or ζ_H if $\hat{\omega}$ is homogeneous.

Assumption 2.ii implies that above any threshold, the match-specific components of productivity follow a power law.[29] One implication is that the industry-level elasticity of substitution across groups of suppliers of the same input is $\zeta_{\hat{\omega}} + 1$. When there is more dispersion in these match-specific components of

28. We say that a vector $(x_0, x_1, \ldots, x_n)$ strictly dominates the vector $(y_0, y_1, \ldots, y_n)$ if $x_0 > y_0, x_1 > y_1, \ldots, x_n > y_n$.

29. This type of functional form assumption goes back to at least Houthakker (1955), and versions of it are also used by Kortum (1997), Jones (2005), Oberfield (2018), and Buera and Oberfield (2020). Note that the expected number of potential suppliers for an input is unbounded. Formally, an economy satisfying Assumption 2.ii can be thought of as the limit of a sequence of economies that satisfy more standard assumptions. Consider an economy in which firms were restricted to use only suppliers with a match-specific productivity greater than $\underline{z}$. Then the number of potential suppliers for each input of a technique would be given by a Poisson distribution with mean $\underline{z}^{-\zeta}$ and the match-specific productivity for each supplier would be drawn from a Pareto distribution with CDF $1 - (\frac{z}{\underline{z}})^{-\zeta}$. An economy satisfying Assumption 2.ii can be thought of as the limit of such an economy as $\underline{z} \to 0$. In this limit, the number of suppliers for each input of a technique grows arbitrarily large, but the match-specific productivity associated with any single supplier is drawn from an arbitrarily poor distribution. The limit is well-behaved because the probability of drawing a supplier with match-specific productivity greater than z does not change as $\underline{z} \to 0$. In principle, we could have allowed the level of the match-specific component of productivity draws to vary by input-output pair and recipe, or $Z_{\omega\rho\hat{\omega}} z^{-\zeta_\omega}$, reflecting the idea that industries are often concentrated geographically or ethnically, which may imply that a given output industry may face an unusually high number of good suppliers in the input industry relative to other output industries. However, it turns out that Assumptions 2.ii and 2.iii imply that any variation in $\{Z_{\omega\rho\hat{\omega}}\}$ would be absorbed into the constant $B_{\omega\rho}$, so we simply normalize each $Z_{\omega\rho\hat{\omega}}$ to unity.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

productivity (low $\zeta_{\hat{\omega}}$), a buyer is less likely to switch suppliers in response to changes in the supplier's price because it is likely that there is a larger gap between the best and second-best suppliers of an input. ζ_R will play a role quantitatively because it determines the likelihood that a buyer will have a close substitute if it faces a holdup problem with its best supplier of an input.

Assumption 2.iii says that the common components of input-augmenting productivities of a technique follow independent power laws. $B_{\omega\rho}$ summarizes the level of these productivity draws. We take these to be primitives, although a deeper model might model them as resulting endogenously from directed search or from the diffusion of technologies across entrepreneurs that know each other.

Assumption 2.iv says that for each recipe, the sum of the power law exponents is the same, equal to γ. We will show later that the industry-level elasticity of substitution across techniques is $\gamma + 1$.[30] The parameter restrictions are necessary to keep utility finite.

It will be useful to normalize the power law exponents by their sum. For recipe ρ, define

$$\alpha_L^\rho = \frac{\beta_l^\rho}{\gamma}, \qquad \alpha_{\hat{\omega}}^\rho \equiv \frac{\beta_{\hat{\omega}}^\rho}{\gamma}, \quad \hat{\omega} \in \hat{\Omega}^\rho.$$

Note that this implies that $\alpha_L^\rho + \sum_{\hat{\omega}\in\hat{\Omega}^\rho} \alpha_{\hat{\omega}}^\rho = 1$. Furthermore, for some results, it will be useful to define $\alpha_R^\rho = \sum_{\hat{\omega}\in\hat{\Omega}_R^\rho} \alpha_{\hat{\omega}}^\rho$ and $\alpha_H^\rho = \sum_{\hat{\omega}\in\hat{\Omega}_H^\rho} \alpha_{\hat{\omega}}^\rho$.

With these assumptions in hand, we now characterize the equilibrium. All proofs are contained in Online Appendix F.[31]

PROPOSITION 1. Under Assumptions 1 and 2, the fraction of firms in industry ω with unit cost greater than c is

$$e^{-\left(\frac{c}{C_\omega}\right)^\gamma},$$

30. It would be straightforward to allow different industries to have different values of γ. However, as we show below, our counterfactuals are insensitive to the value of γ. We therefore leave the γ constant across industries to reduce notational clutter.

31. We thank Matt Rognlie and Sam Kortum for suggestions that simplified the proof of Proposition 1.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

where

$$C_\omega = \left\{ \sum_{\rho \in \varrho_\omega} \kappa_{\omega\rho} B_{\omega\rho} \left(\left(t_x^*\right)^{\alpha_R^\rho} t_l^{\alpha_L^\rho} \prod_{\hat{\omega} \in \hat{\Omega}^\rho} C_{\hat{\omega}}^{\alpha_{\hat{\omega}}^\rho} \right)^{-\gamma} \right\}^{-\frac{1}{\gamma}}$$

$$t_x^* = \left(\int_1^\infty t_x^{-\zeta_R} dT\,(t_x) \right)^{-\frac{1}{\zeta_R}} \tag{1}$$

and $\kappa_{\omega\rho}$ is a constant that depends on technological parameters.

Proposition 1 shows that the distribution of cost among firms within each industry takes the simple form of a Weibull distribution with shape parameter γ and scale determined by C_ω, which we call the cost index for industry ω. Equation (1) relates industry ω's cost index to that of the industries that provide the inputs for each recipe and to t_x^* and t_l, which summarize the effect of imperfect enforcement on those that produce the inputs used in recipe ρ. t_x^* accounts for both the direct impact of the wedges—the wasted resources from holdup problems—and the indirect impact: wedges might cause firms to switch to a supplier with higher cost or lower productivity, or to a different technique altogether.

Equation (1) is a system of equations that implicitly determines each industry's cost index, $\{C_\omega\}_{\omega \in \Omega}$. Proposition 2 shows that these are sufficient to characterize aggregate productivity.

PROPOSITION 2. Under Assumptions 1 and 2, the household's aggregate consumption is

$$U = \left\{ \sum_{\omega \in \Omega} \upsilon_\omega J_\omega^{\frac{\eta-1}{\varepsilon-1}} \Gamma \left(1 - \frac{\varepsilon - 1}{\gamma}\right)^{\frac{\eta-1}{\varepsilon-1}} C_\omega^{1-\eta} \right\}^{\frac{1}{\eta-1}} L.$$

We next turn to industry-level expenditure shares. The next proposition characterizes the aggregate share of total expenditures (on both intermediate inputs and labor) that is spent on each input among all firms that use a particular recipe.

PROPOSITION 3. Suppose that Assumptions 1 and 2 hold. Among firms that in equilibrium produce using recipe ρ:

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

- the average and aggregate shares of expenditures spent on inputs from $\hat{\omega} \in \hat{\Omega}_R^\rho$ are both $\frac{\alpha_{\hat{\omega}}^\rho}{\bar{t}_x}$,
- the average and aggregate shares of expenditures spent on inputs from $\hat{\omega} \in \hat{\Omega}_H^\rho$ are $\alpha_{\hat{\omega}}^\rho$,
- the average and aggregate shares of expenditures spent on labor are $\alpha_L^\rho + \left(1 - \frac{1}{\bar{t}_x}\right)\alpha_R^\rho$,

where $\bar{t}_x \equiv \left[\int_1^\infty t_x^{-1} d\tilde{T}(t_x)\right]^{-1}$ and $\tilde{T}(t_x) \equiv \frac{\int_1^{t_x} t^{-\zeta_R} dT(t)}{\int_1^\infty t^{-\zeta_R} dT(t)}$.

Proposition 3 provides relatively simple expressions for the average and aggregate cost shares of each input among those that choose to use a particular recipe. These properties will be central to our identification procedure. Although there is micro-level heterogeneity in the cost shares among those using a particular recipe, the aggregate factor shares among those firms depend only on technological parameters, not on the relative prices of the inputs. Thus at the recipe level, there is a Cobb-Douglas aggregate production function. This extends the celebrated aggregation result of Houthakker (1955), who derived a similar result under the assumption that individual production functions are Leontief.[32] We require only that the production function exhibits constant returns to scale and that all inputs are complements.[33]

32. Jones (2005) builds on Houthakker (1955) but derives a different type of result. Jones first shows that if a single plant draws many Leontief production functions where factor-augmenting productivities are drawn from independent Pareto distributions, the envelope of those production functions is Cobb-Douglas. He shows numerically that the result extends beyond Leontief to CES production functions when the factors are complements. Note that these are not aggregation results; these results apply at the level of a single firm. Lagos (2006) and Mangin (2017) also build on Houthakker (1955) incorporating labor market search, while Growiec (2013) extends Jones's argument to microfound an aggregate CES production function.

33. Why complements? If inputs are complements, then an increase in the price of an input would have two offsetting effects on the aggregate cost share. The higher price raises each firm's cost share of that input. At the same time, firms that use that input more intensively are likely to shrink or switch to a technique that uses the input less intensively. When factor-augmenting productivities are drawn from independent Pareto distributions, these offset exactly and aggregate factor shares are unchanged. If inputs were substitutes, the two effects would push in the same direction, so that if the price of an input rose, its aggregate cost share would fall. Mathematically, if inputs were substitutes then the constant $\kappa_{\omega\rho}$ would diverge, as the arrival rate of techniques that deliver cost lower than c would be infinite for any c. While the assumption of constant returns to scale keeps the characterization of sourcing decisions tractable, the extension of Houthakker (1955) requires only homotheticity.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

Imperfect enforcement, on the other hand, reduces the expenditure share of relationship-specific inputs. The buyer's production decisions depend on each input's effective cost, whereas the expenditures reflect the actual payment to each supplier. Recall that imperfect enforcement means that the buyer's effective expenditure on a relationship-specific input is spent partly on payments to the supplier for the input and partly on labor to customize the good.[34]

III.D. Counterfactuals

The quality of contract enforcement can be summarized by the distribution of wedges T. Suppose that the quality of enforcement changed in such a way that the distribution of wedges changed from T to T'. How would this affect aggregate productivity? Taking J_ω and $B_{\omega\rho}$ as primitives, the following proposition shows how one can compute the impact of such a change.[35]

Let HH_ω be the share of the household's expenditure on goods from industry ω in the current equilibrium. Among those of type ω, let $R_{\omega\rho}$ be the share of total revenue of those that use recipe ρ in the current equilibrium.

PROPOSITION 4. If the quality of enforcement changed so that the distribution of wedges changes from T to T', the change in household utility would be

$$\frac{U'}{U} = \left(\sum_{\omega} HH_\omega \left(\frac{C'_\omega}{C_\omega} \right)^{1-\eta} \right)^{\frac{1}{\eta-1}}$$

and the change in industry efficiencies would satisfy the following system of equations

$$\left(\frac{C'_\omega}{C_\omega} \right)^{-\gamma} = \sum_{\rho \in \mathcal{Q}_\omega} R_{\omega\rho} \left[\left(\frac{t_x^{*\prime}}{t_x^*} \right)^{\alpha_R^\rho} \prod_{\hat{\omega} \in \hat{\Omega}^\rho} \left(\frac{C'_{\hat{\omega}}}{C_{\hat{\omega}}} \right)^{\alpha_{\hat{\omega}}^\rho} \right]^{-\gamma}. \tag{2}$$

34. The wedge due to imperfect enforcement and input-augmenting productivity affect a firm's unit cost in the same way. It is important, however, to model them separately because they affect expenditure shares in different ways. Larger wedges tend to reduce the share of expenditures on that input because some of the effective cost is paid to labor; lower input-augmenting productivities do not.

35. An interesting alternative exercise is asking what would happen if $\{J_\omega\}$ and $\{B_{\omega\rho}\}$ also responded to the change in T.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

The first part of the proposition states that to know how a change in court quality affects aggregate productivity, it is sufficient to know only the changes in industry cost indices, $\frac{C'_\omega}{C_\omega}$. In turn, the change in each industry's cost index depends on the weighted average over recipes of the change in the cost index of the industries that supply inputs along with the change in t_x^*, the statistic that summarizes the direct and indirect impact of the wedges that distort production using relationship-specific inputs. Equation (2) describes a system of equations that implicitly characterizes these changes in cost indices.

While Proposition 4 describes exactly how a change in enforcement would alter welfare, it is instructive to study a perturbation of the distribution of wedges to show which features of the economy are important for determining the first-order impact of a change in the quality of enforcement.

COROLLARY 1. The marginal welfare impact of a change in court quality is

$$d\log U = -\sum_{\omega\in\Omega} HH_\omega d\log C_\omega,$$

and the change in industry efficiencies can be summarized by the following system of equations:

$$d\log C_\omega = \sum_{\rho\in\varrho_\omega} R_{\omega\rho}\left[\alpha_R^\rho d\log t_x^* + \sum_{\hat{\omega}\in\hat{\Omega}^\rho} \alpha_{\hat{\omega}}^\rho d\log C_{\hat{\omega}}\right].$$

One implication is that, to a first order, the change in utility resulting from a change in the quality of enforcement does not depend on γ or η.[36]

IV. IDENTIFICATION AND ESTIMATION

Our main counterfactual of interest is how aggregate productivity and the organization of production would change if the quality of enforcement improved. We do this in several steps.

36. This can be viewed as an application of the envelope theorem. For small changes T, one can compute the impact on aggregate productivity holding fixed other choices, that is, holding fixed the technique each firm uses. Thus γ, which regulates substitution across techniques, does not matter to a first order.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

We first parameterize the model using information from the ASI under the assumption that the quality of enforcement varies by state. We then project the implied quality of enforcement for each state on our measures of court congestion. Finally, we compute the gains from reducing congestion to the level prevailing in the least congested state.

Our most important identifying assumption is that weak enforcement may introduce a wedge in the use of inputs that require customization and in the use of labor, but not in the use of standardized inputs. Given our scheme for identification, we view this as a conservative assumption. If the use of standardized inputs were also distorted by weak contract enforcement, then all of the wedges would be larger than the ones we infer.

The following proposition shows a set of moments that we can use in a GMM procedure to estimate the model parameters.

PROPOSITION 5. Let s_{Rj}, s_{Hj}, s_{Lj} be firm j's spending on relationship-specific inputs, homogeneous inputs, and labor, respectively, as shares of its revenue. Under Assumptions 1 and 2, the first moments of revenue shares among firms that produce ω that in equilibrium use recipe ρ satisfy:

$$\mathbb{E}\left[\bar{t}_x \frac{s_{Rj}}{\alpha_R^\rho} - \frac{s_{Hj}}{\alpha_H^\rho}\right] = 0$$

$$\mathbb{E}\left[\frac{s_{Lj} + s_{Rj}}{\alpha_L^\rho + \alpha_R^\rho} - \frac{s_{Hj}}{\alpha_H^\rho}\right] = 0.$$

ASSUMPTION 3. We impose that the following objects are the same across states: (i) the form of the production function for each recipe $\{G_{\omega\rho}\}$; (ii) the power law exponents for the input-augmenting productivity draws for techniques of each recipe $\{\beta_l^\rho, \beta_{\hat{\omega}_1}^\rho, \ldots, \beta_{\hat{\omega}_n}^\rho\}$; and (iii) the power law exponents for the match-specific productivity draws, ζ_R and ζ_H.

We allow all other features of preferences and technology to vary freely across states. This includes absolute and comparative advantages in recipes, $\{B_{\omega\rho}\}$; (ii) the measure of firms of each type $\{J_\omega\}$; (iii) the households' tastes, $\{\upsilon_\omega\}$; and most important, (iv) the quality of contract enforcement, T, and t_l.

We also impose a parametric form for the stochastic wedges that a firm draws for each supplier of a relationship-specific input.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

In particular, the wedge is drawn from a Pareto distribution, where the shape parameter is specific to a state.[37]

ASSUMPTION 4. The distribution of wedges in state d is $T_d(t_x) = 1 - t_x^{-\tau_d}$ for $t_x \geqslant 1$.

Our algorithm for identification is thus as follows:

i. Identify recipes, estimate technology parameters $\{\alpha_L^\rho, \alpha_H^\rho, \alpha_R^\rho\}_{\rho \in \varrho_\omega, \omega \in \Omega}$, and distortions to the cost of relationship-specific inputs for each state, $\bar{t}_x^d$. We use an iterative procedure to ensure that our recipe classification is consistent with the possibility of distortions that vary across states.
   a. Start with an initial guess of $\bar{t}_x^d$ for each state d.
   b. Identify recipes from plants' cost shares (see the next section for details), taking out the distortion to the cost shares of relationship-specific inputs $\bar{t}_x^d$.
   c. Use Proposition 5 to estimate the production parameters that are common across locations $\{\alpha_L^\rho, \alpha_H^\rho, \alpha_R^\rho\}_{\rho \in \varrho_\omega, \omega \in \Omega}$ and a new set of the state-specific variables, $\{\bar{t}_x^d\}$.
   d. Go back to step i.b until the $\bar{t}_x^d$ have converged.

ii. Compute t_x^* for each state. Assumption 4 implies that $\bar{t}_x = 1 + \frac{1}{\zeta_R + \tau_d}$ and $t_x^* = \left(\frac{\tau_d + \zeta_R}{\tau_d}\right)^{\frac{1}{\zeta_R}}$. We estimate ζ_R externally, and then use this along with our estimates of $\bar{t}_x$ to compute t_x^*.

iii. For the counterfactual, we also need values of the industry-level output elasticities of each input for each recipe, $\{\alpha_{\hat{\omega}}^\rho\}$. To do this, we pool data across states to estimate the remaining production function parameters, $\alpha_{\hat{\omega}}$, by using the aggregate expenditures. For example, if the sourced good $\hat{\omega}$ is relationship specific, then $\alpha_{\hat{\omega}}^\rho$ is equal to α_R^ρ multiplied by the ratio of total expenditure

37. The Pareto distribution has the attractive property that it is closed under minimization. Following the discussion in Section III.B, contracts might be enforced formally or informally. If the probability that the formal wedge is greater than t_x is $t_x^{-\tau_d^{\text{formal}}}$ and the probability that the informal wedge is greater than t_x is $t_x^{-\tau_d^{\text{informal}}}$, then the probability that the effective wedge is greater than t_x is $t_x^{-\tau_d}$, where $\tau_d = \tau_d^{\text{formal}} + \tau_d^{\text{informal}}$.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

on input $\hat{\omega}$ by those that use recipe ρ to total expenditure on relationship-specific inputs.

iv. For each state-recipe, directly measure the share of industry ω revenue earned by firms that in equilibrium use recipe ρ, $\{R_{\omega\rho}\}$. Similarly, directly measure for each state the share of final demand spent on industry ω, $\{HH_{\omega}\}$.

v. Calibrate η and γ externally.

In implementing this algorithm, we make several auxiliary assumptions that in principle could be relaxed. First, we assume that there is no trade across state borders. Although it would be fairly straightforward to incorporate interstate trade, we lack the relevant data.[38] A second assumption is that the recipes used by multiproduct firms and the distribution of wedges facing them are the same as those of single-product firms. This type of assumption, while strong, is standard in the literature, as we lack the data that indicates which inputs are used in the production of which products. It allows us to estimate wedge distribution parameters and the α's using single-product plants only, while still being able to make statements about the whole formal manufacturing sector by including multiproduct plants when we calculate revenue and expenditure shares $R_{\omega\rho}$ and HH_{ω}. We discuss the treatment of multiproduct plants in more detail in Section IV.C. Third, we treat service inputs and energy inputs as primary inputs.

IV.A. Defining Recipes

One of the salient facts of the Indian manufacturing data is that even within narrow industries, plants use vastly different combinations of intermediate inputs to produce the same output. Our model provides a natural way to think of these input-output combinations as different recipes $\rho \in \rho(\omega)$ that could be used to produce the same output ω. To estimate the model from the microdata, we need a procedure that classifies each plant-year observation into which recipe the plant is using.

38. To this point there is no comprehensive, publicly available data about cross-state trade in goods. The conventional wisdom has been that interstate trade is minimal, although the 2016–17 Economic Survey of India's Ministry of Finance challenges this conventional wisdom. In Online Appendix G.4 we explore how incorporating interstate trade might alter our counterfactuals. We make the assumption that 10% of inputs are sourced from out of state. We find that this has a minor quantitative impact on our counterfactual results.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

The idea that guides our classification is that for a given output good, similar input mixes should belong to the same recipe. We describe each plant j's input mix by the vector of its input expenditure shares, $(m_{j\hat{\omega}})_{\hat{\omega}\in\Omega} = \left(\frac{X_{j\hat{\omega}}}{\sum_{\hat{\omega}'} X_{j\hat{\omega}'}}\right)_{\hat{\omega}\in\Omega}$. Graphically, each vector corresponds to a point in the $|\Omega|$-dimensional hypercube that is lying on the hyperplane where the sum of all coordinates equals one. Our task is to find plants with similar input mixes, that is, clouds of points that are close to each other (according to some metric). In statistics, this task is known as cluster analysis, and there are a large number of algorithms that classify clusters based on distance, density, shape, and other criteria. Looking at the input mixes of plants in many different industries (see the examples of bleached cotton cloth and polished diamonds in Figure I) convinced us that these clusters do exist and have a meaningful economic interpretation.

Our preferred method is due to Ward (1963) and constructs clusters recursively, starting with the partition where every cluster is a singleton. In each step, two clusters are merged to minimize the sum of squared errors:

$$\min_{\rho_n \gtrdot \rho_{n-1}} \sum_{\rho\in\rho_n} \sum_{j\in\rho} \sum_{\hat{\omega}} \left(m_{j\hat{\omega}} - \overline{m}_{\rho\hat{\omega}}\right)^2,$$

where the ρ_n are partitions of J_ω, and in each step ρ_n runs over all partitions that are coarser than ρ_{n-1}. This method constructs a hierarchical set of partitions of J_ω: one for each desired number of clusters.

Our implementation of the clustering procedure to identify recipes defines the vector m_j as the concatenation of the vector of j's five-digit materials shares and its three-digit shares, to allow for the possibility of misclassification of inputs within three-digit baskets. We determine the number of potential recipes using the prediction strength method of Tibshirani and Walther (2005). This procedure is similar to cross-validation. The procedure divides the sample into two subsamples (A and B) and then clusters the data in each subsample. It then takes the cluster centroids from A and all of the datapoints from B and assigns them to clusters based on the centroids from A. Each pair of points in B thus can be categorized as being in the same cluster or different clusters based on the original clustering of the B subsample and based on the centroids in A. The procedure then tabulates the fraction of pairs of points for which the two agree. We choose a

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

TABLE IV

STATISTICS ON PRODUCTS AND RECIPES

	Count
Products (five-digit ASIC)	4,530
Products with $\geqslant$ three plants	3,573
Products with $\geqslant$ five plants	3,034
Recipes	26,776
Recipes with $\geqslant$ three plants	6,280
Recipes with $\geqslant$ five plants	2,574
Avg. plants per recipe	8.2
Std. dev. plants per recipe	79.4

Notes. "Products" are the five-digit product codes in our data, "recipes" are the output from our clustering procedure. Plant counts include only single-product plants.

TABLE V

SUMMARY STATISTICS ON RECIPES

	Mean	Std. dev.	Min	Max
Cost share of L	0.40	0.22	0.0002	0.999
Cost share of X_R	0.27	0.28	0	0.999
Cost share of X_H	0.33	0.30	0	0.998
Number of inputs with cost share $> 1\%$	4.4	4.6	1	37
Number of inputs with cost share $> 0.1\%$	6.4	12.6	1	205

threshold parameter of 0.95, meaning that the procedure chooses the largest number of clusters so that on average at least 95% of pairs of points agree. Although a lower threshold allows for more clusters—for example, a threshold of 1.00 is typically consistent with a single cluster—the relative number of clusters for each industry is driven by the shape of the point clouds in the data.[39] Table IV shows statistics on the number and size of clusters (recipes) that result from using this procedure, and Table V shows statistics on cost shares across recipes. We find that among plants we classify as using the same recipe, most of the expenditures are on the same set of inputs.[40] Figure III shows the resulting

39. In Online Appendix G.2 we show results for alternative values of the threshold parameter, which varies the degree of recipe fineness.

40. Our model imposes that all firms using the same recipe use the same bundle of inputs, but our classification algorithm does not. Formally, this can be understood in the context of our model as imposing that all plants that we classify as using the same empirical recipe actually use different recipes that share the same α_R, α_H, and α_L. We believe that this is a relatively innocuous assumption,

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

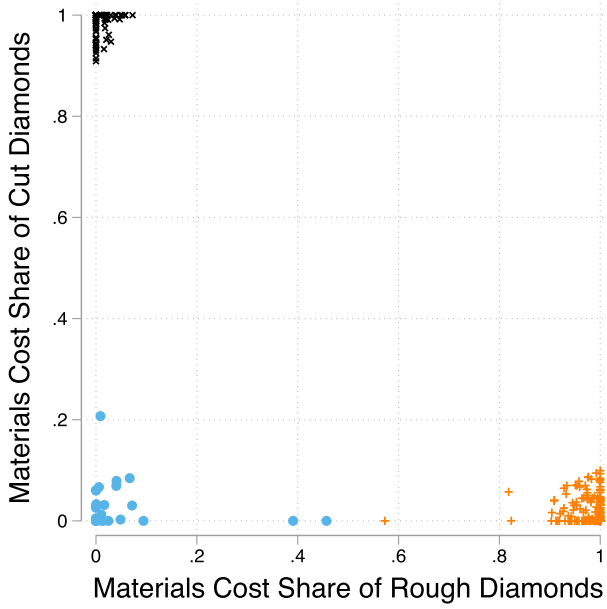


FIGURE III

Recipe Classification Example: Polished Diamonds (92104)

The figure shows the result from the recipe classification procedure (after correcting for wedged) for "polished diamonds" (92104). Observations that are classified as belonging to different recipes are tagged with different markers.

classification for the example of the producers of Polished Diamonds.

Once we have defined recipes, we assign plants to belong to a recipe with a probability that is proportional to the inverse Euclidean distance to the recipe center:

$$P(j \in \rho) = \frac{\frac{1}{\sqrt{\sum_{\omega'}(m_{j\omega'} - \overline{m}_{\rho\omega'})^2}}}{\sum_{\rho' \in \rho(\omega)} \frac{1}{\sqrt{\sum_{\omega'}(m_{j\omega'} - \overline{m}_{\rho'\omega'})^2}}}. \tag{3}$$

because inputs that are not being used by a plant but are used by some other plants using the same empirical recipe account for very low cost shares: on average (across recipes, firms, and inputs) for around 0.32% of the recipe's materials expenditure. It is likely that some of the differences in input bundles, especially those arising from inputs with very small cost shares, are due to respondents occasionally omitting unimportant inputs.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

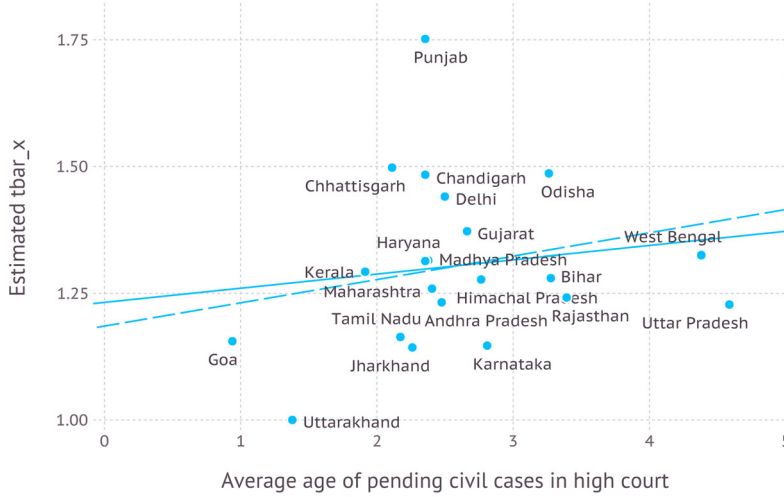


FIGURE IV

Correlation between $\bar{t}_x$ and Court Congestion

The figure shows the correlation between $\bar{t}_x^d$ and the average age of pending civil cases in the state's high court. The solid line is the OLS regression line; the dashed line is the fit of an IV regression where the age of pending cases is instrumented using the log age of the high court.

We use these assigned probabilities as weights in the estimation below.

IV.B. Estimation

We estimate the $\bar{t}_x^d$, α_R^ρ, α_H^ρ, and α_L^ρ from the moment conditions in Proposition 5 using our algorithm described above. To identify the level of $\bar{t}_x$, we set the smallest $\bar{t}_x$ to 1, thereby making the assumption that the least distorted state is undistorted.[41] We also exclude state-recipe pairs where the average share of relationship-specific inputs in sales exceeds that of homogeneous inputs by a factor of 100 (and vice versa).

Figure IV shows the estimated $\bar{t}_x^d$ and their correlation with high court congestion as measured by the average age of pending civil cases. Frictions are large: $\bar{t}_x$ exceeds 1.75 in the most heavily

41. We view this as conservative. Given the expenditure shares we see in the data, more severe distortions (larger $\bar{t}_x$) would raise the estimated output elasticities of relationship-specific inputs (higher α_R^ρ). This would amplify the responses to changes in enforcement.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

distorted state, and is close to 1.5 in several others. Some of that variation is explained by the congestion of courts. In states with slower courts, firms face larger distortions (higher $\bar{t}_x$) when sourcing relationship-specific intermediate inputs. The solid line is the fit of an OLS regression of $\bar{t}_x$ on court congestion; the dashed line is the fit of an IV regression where we instrument court congestion using the log age of the high court. The estimated IV slope coefficient is similar to the one in the OLS. This relationship between $\bar{t}_x$ and the age of pending court cases is closely related to the fact that intermediate-input bundles are tilted towards homogeneous inputs in states where enforcement is weak (Fact 2). The main difference here (beyond the fact that the $\bar{t}_x$ are coming from a nonlinear regression) is that the $\bar{t}_x$ are identified from within-recipe variation in the input mix, whereas Fact 2 is about within-product variation.

Before proceeding, it is worth discussing how we separate heterogeneity in production technology (i.e., recipes) from distortions. Given a set of recipes and a classification of plants into which recipes they use, identification of the α's and $\bar{t}_x$'s fundamentally comes from the assumption that the parameters that govern contracting frictions (the $\bar{t}_x$'s) vary by state, whereas the recipe technology parameters (the α's) are the same across states but vary across recipes.[42] Similar to the problem of estimation with grouped fixed effects studied by Bonhomme and Manresa (2015), we face the problem of estimating these parameters jointly with the number of recipes and the assignment of plant observations into recipes, that is, jointly with the clustering procedure with a choice of the number of clusters. Like Bonhomme and Manresa (2015), we do this by iterating between a clustering procedure and parameter estimation, so that, in our context, the recipe classification accounts for the possibility of distortions.

The theoretical results in Bonhomme and Manresa (2015) and Moon and Weidner (2015), who study similar estimation procedures in closely related models, suggest that our estimation procedure has the following properties. First, if we choose the right number of recipes, the estimated distortions will be

42. This is analogous to a nonlinear panel regression. In a panel regression, one might estimate individual fixed effects and time fixed effects. Here we are estimating state fixed effects (the distortions) and recipe fixed effects (technology parameters). The fact that the moment conditions are nonlinear does not change the basic logic.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

consistent as the number of plants grows large. Second, if we choose too few recipes, the estimated distortions may be biased, as the difference in recipe usage across states might be misinterpreted as distortions. Third, if we choose too many recipes, the estimated distortions will be consistent, as the additional recipes act like irrelevant extra regressors in a least squares regressions. Although our model is not nested by either of those, we confirm these large sample properties, and study the small-sample behavior of our estimator, using Monte Carlo simulations in Online Appendix G.3.

IV.C. Estimating and Calibrating the Remaining Parameters

To perform a counterfactual where we assess the aggregate effect of a change in the wedge distribution T, Proposition 4 tells us that we need to know the change in the moment t_x^* of the wedge distribution that is relevant for the industry's cost distribution, which depends on the parameter ζ_R and, under our parameterization of the wedge distribution, on its Pareto tail τ_d. We also need to know the parameters $\alpha_{\hat{\omega}}^{\rho}$, the within-industry sales shares $R_{\omega\rho}$ of each recipe, the household's expenditure shares HH_{ω}, and the elasticities γ and η.

1. Estimating ζ. The parameter ζ_R, which allows us to back out τ_d from $\bar{t}_x$, also governs the elasticity of substitution across sets of suppliers of the same input. Although our data do not indicate the identity of the supplier of an input, it does indicate whether it was purchased from a foreign or domestic supplier. We can thus estimate ζ_R using these two groups for each input-output pair:

$$\log\left(\frac{X_{\omega\hat{\omega}t}^{DOM}}{X_{\omega\hat{\omega}t}^{IMP}}\right) = \zeta \log(1+\iota_{\hat{\omega}t}) + \lambda_t + \lambda_{\omega\hat{\omega}} + \eta_{\omega\hat{\omega}t},$$

where $\iota_{\hat{\omega}t}$ is the import tariff on $\hat{\omega}$ at time t, and the λ's are fixed effects. Table VI shows the results for this regression. We use the point estimate of 0.218 for ζ_R.[43]

2. Calculating the $\alpha_{\hat{\omega}}$ from α_R and α_H. The elasticities $\alpha_{\hat{\omega}}$ can be recovered as the product of the input-type elasticity $\left(\alpha_R^{\rho}\text{ or }\alpha_H^{\rho}\right)$

43. Choosing a low ζ_R is conservative; see the discussion in Section IV.E. We conduct sensitivity checks in Online Appendix G.1.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

TABLE VI
ESTIMATING ζ

	Dependent variable: $\log\left(\frac{X^{DOM}_{\omega\hat{\omega}t}}{X^{IMP}_{\omega\hat{\omega}t}}\right)$		
	(1)	(2)	(3)
$\log(1+\iota_{\hat{\omega}t})$	0.617	0.218	1.209*
	(0.44)	(0.77)	(0.52)
Industry × Input FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Level	Five-digit	Five-digit	Five-digit
Sample	All inputs	R only	H only
R^2	0.601	0.580	0.623
Observations	23,692	12,002	11,690

Notes. Robust errors are in parentheses, clustered at the state × industry level. $^{+}$ $p < .10$, * $p < .05$, ** $p < .01$. Dependent variable is the log ratio of total expenditure on domestically sourced to total imported intermediate inputs of type $\hat{\omega}$ among producers of ω at time t. We only use census plants (which are surveyed every year) to reduce artificial fluctuations that result from sampling.

and the average cost share of plants that produce using recipe ρ:

$$\alpha^{\rho}_{\hat{\omega}} = \alpha^{\rho}_{R}\frac{\overline{m}_{\rho\hat{\omega}}}{\sum_{\omega'\in\Omega^{R}}\overline{m}_{\rho\hat{\omega}'}} \text{ if } \hat{\omega} \text{ relationship specific,}$$

$$\alpha^{\rho}_{\hat{\omega}} = \alpha^{\rho}_{H}\frac{\overline{m}_{\rho\hat{\omega}}}{\sum_{\omega'\in\Omega^{H}}\overline{m}_{\rho\hat{\omega}'}} \text{ if } \hat{\omega} \text{ homogeneous.}$$

3. Accounting for Multiproduct Plants. We target the demand aggregator's expenditure shares HH_{ω} and the recipe revenue shares $R_{\omega\rho}$ to represent the aggregate of India's formal manufacturing sector, which includes single- and multiproduct plants. We calculate HH_{ω} separately for each state as total sales of ω (from both single-product and multiproduct plants) minus total intermediate consumption (less imports) of ω (or 0, if this difference is negative), divided by the sum of this difference over all products.

To calculate the recipe revenue shares $R_{\omega\rho}$, we need to assign the revenue of multiproduct plants to a particular recipe. We assume that multiproduct plants produce each of their products ω using one or more recipes $\rho \in \rho_{\omega}$, with each of them accounting for a fraction $r_{j\omega\rho}$ of sales of ω. We estimate the $r_{j\omega\rho}$ by minimizing the Euclidean distance between the firm's observed vector of materials cost shares $m_{j\omega}$ and expected cost shares

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

(which depend on the $r_{j\omega\rho}$ as well as on $\bar{t}_x^d$). We then construct the recipe sales shares $R_{\omega\rho}$ using the output of single-product plants and the estimated output of multiproduct plants. This strategy is model consistent but requires that all recipes used by multiproduct firms are in the support of recipes used by single-product firms (otherwise we would not be able to detect them in the estimation/clustering procedure). Moreover, when calculating $R_{\omega\rho}$, we pool observations across states and years but weight each plant-year observation by the inverse of the number of times the plant shows up in the ASI. This weighting allows us to construct parameters that better represent the aggregate of India's formal manufacturing sector.

Finally, we calibrate $\eta = 1$ and $\gamma = 1$; we show in Online Appendix G.1 that our counterfactuals are insensitive to these parameters. Inputs that do not show up in our data as outputs (predominantly agricultural and mineral commodities) are assumed to have unchanged productivity distributions in the counterfactual simulation.

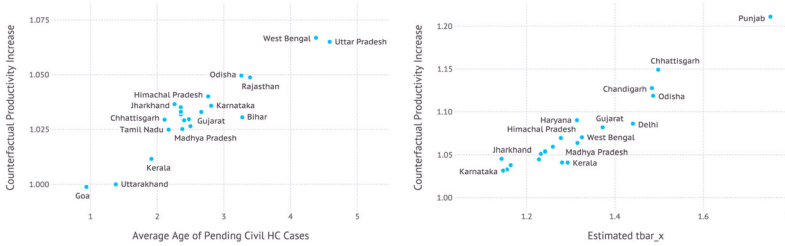
IV.D. Counterfactuals

We then perform two counterfactuals. In the first one, we reduce $\bar{t}_x^d$ for each state by the amount that is implied by the IV regression in Figure IV, down to a point where the average age of pending cases is one year (which is roughly the level of congestion enjoyed by the best state, Goa). Using our estimate for ζ_R, we back out τ_d at the original and counterfactual level and compute the change in the welfare-relevant moment $t_x^* = \left(\frac{\tau_d+\zeta_R}{\tau_d}\right)^{\frac{1}{\zeta_R}}$. We then compute the corresponding change in the household's utility aggregate as given by Proposition 4.

Figure V, Panel A shows the corresponding counterfactual increase in the household's consumption aggregate U. The numbers are in the range of 0% to 7%, suggesting that the gains from improving court speed can be large. That said, there are several sources of uncertainty around these estimates. The correlation between court congestion and identified wedges $\bar{t}_x$ is a correlation at the state level and is therefore not very precisely estimated. Furthermore, the wedges $\bar{t}_x$ are estimates themselves, and this moment will be less precisely estimated in states where we have fewer plant observations.

In the second counterfactual, we reduce the magnitude of distortions in such a way that the wedges $\bar{t}_x$ are half as large

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021



(A) Improving court speed to level of best court

(B) Halving wedges

FIGURE V

Counterfactual Increases in Aggregate Productivity

The figure shows the counterfactual increase in C when the wedges on relationship-specific inputs are reduced. In Panel A we reduce $\bar{t}_x$ according to the fraction of $\bar{t}_x$ that is explained by court congestion in a linear IV regression (Figure IV); in Panel B we cut the $\bar{t}_x$ in half.

as in the estimates. This counterfactual is independent from the correlation of wedges with the measured congestion in the high courts, and would thus reflect an improvement in the overall quality of contract enforcement. Figure V, Panel B shows the counterfactual increase in U for each state. U would increase by 7.1% on average and by more than 10% in the most distorted states. These welfare gains are more precisely estimated than those for the first counterfactual; however, they are still subject to random variation in the estimated wedge moments $\bar{t}_x$ that comes from the fact that we do not observe the universe of smaller plants, just a random sample of them.

IV.E. Direct and Indirect Costs of Distortions

The model captures the idea that buyer–supplier relationships differ in their susceptibility to imperfect contract enforcement. Some buyer–supplier pairs are able to enforce contracts informally or transact in a way that avoids holdup problems. To capture this in a parsimonious way, we assume that the buyer draws a distortion for each potential supplier from a distribution that deteriorates when formal enforcement is less reliable.

Our estimation procedure identifies $\bar{t}_x$ for each state, which is the (harmonic) average of the realized distortions. This summarizes the direct cost of weak enforcement on users of relationship-specific inputs. t_x^* captures the direct and indirect cost of the distortions. We infer the indirect cost by estimating

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

ζ_R, which indexes the probability that a firm has a comparable alternative supplier. Jensen's inequality implies that $\bar{t}_x \leqslant t_x^*$, which implies that the $\bar{t}_x$ is a lower bound for the total impact, and this lower bound is attained only if the distribution of distortions T is degenerate. In that case, a firm faces the same wedge for all suppliers and could not avoid a distortion by substituting to an alternative supplier; as a result, there would be no indirect cost.

The respective contributions of the direct and indirect effects depend on both the value of ζ_R and the levels of the distortion. Given our estimates and the range of $\bar{t}_x$ we observe, the direct effect comprises much more of the effect of weak enforcement, but the indirect effects are nontrivial. For the average state, the indirect impact accounts for $\frac{t_x^* - \bar{t}_x}{t_x^*} = 6\%$ of the total impact, and for the most distorted state the indirect impact accounts for about 23%.

IV.F. The Role of Heterogeneity

An alternative approach that is common in studies that measure misallocation using input-output tables is to posit industry-level Cobb-Douglas production functions and back out distortions from differences across states (or countries) in industry-level cost shares.[44] In our view, our approach has several advantages over this alternative.

First, a cursory look at the data indicates that plants are producing using different technologies (i.e., some appear to be more vertically integrated than others). We believe that allowing for several recipes facilitates making apples-to-apples comparisons when measuring the effect impact of distortions. Our identification strategy relies on comparing input expenditures among plants that, in equilibrium, use the same recipe. With the industry-level approach, differences in recipe composition across states could lead researchers to infer spurious distortions.

Second, the industry-level approach would miss the indirect effect of the distortions. Given our identification strategy and conditioning on the data, the fact that wedges differ across suppliers implies indirect productivity losses due to plants switching to alternative suppliers to avoid a distorted input. This raises the implied productivity loss from imperfect enforcement.

44. Our model nests such a model as a special case when there is a single recipe per industry and the distribution of wedges T is degenerate so that all suppliers of relationship-specific inputs face the same wedge.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

Our identification strategy consists of measuring $\{\bar{t}_x\}$ and then using these along with our functional-form assumptions to compute t_x^*. In the industry-level approach, there is typically only a direct effect.

Third, as discussed earlier, there is quite a bit of heterogeneity in cost shares across plants. This leads to the following situation: there are occasionally state-industry pairs that are dominated by a small number of plants, and the aggregate cost shares among these plants happen to deviate quite a bit from the industry average. If we used the industry-level cost shares to back out a wedge for that state-industry, we would conclude that the state-industry is severely distorted. In contrast, our estimating equations treat each individual plant as an observation, and the model has a structural error term for the cost share of each plant, which stems from different productivity and cost draws. Thus our approach allows for the possibility that such a state-industry has an extreme cost share because of its draw of a technique; we are not forced to conclude that the distortion for the state-industry is severe.

IV.G. Entry and Exit

In our benchmark model, we assumed that the number of producers of each product is given exogenously. Of course we would expect that if the formal enforcement improved, profitability would increase more in industries that rely more heavily on relationship-specific inputs. We do in fact observe that as formal enforcement deteriorates, the number of firms decreases relatively more in the industries that rely more heavily on relationship-specific inputs (see Table D.3 in Online Appendix D.3). Thus it is likely that the quality of enforcement affects firms' entry and exit decisions.

How would our estimates differ if firms could endogenously enter and exit? Our estimates of contracting frictions and technology would be unchanged because the strategy places no restrictions on the determinants of the set of firms. Nevertheless, entry and exit may be important for the counterfactuals we conduct. Exactly how depends on the specifics of entry and exit and the particular counterfactual. Consider the following two possibilities. First suppose firms were committed to enter but could choose which industry to enter. Then the additional adjustment after formal enforcement improved could raise aggregate productivity

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

even more than a benchmark model suggests.[45] Second, suppose that firms were committed to producing in a particular industry but choose which state to locate in. Here, if formal enforcement improves in one state, aggregate productivity in that state would rise as firms for other states moved in. However, aggregate productivity in other states would fall as firms moved out. This suggests that the response to nationwide improvement in formal enforcement would differ from the response to a single state improving enforcement. Furthermore, cross-sectional evidence about how the number of firms responds to changes in enforcement may not be informative about the economy's response to a nationwide improvement.

IV.H. Interpretation of Our Estimates

Our estimates suggest that lowering the degree of congestion in high courts to the level enjoyed by the most efficient states would bring about productivity increases of on average 4%. For several reasons this estimate is likely to be a lower bound for the overall productivity loss arising from poor contract enforcement. First, even the least congested Indian courts suffer from other problems that constitute imperfections in the enforcement of contracts that are not captured by our court quality measure (such as objectivity and incorruptibility). Second, if contracting frictions are also present for homogeneous intermediate inputs, our estimates of the productivity gains from improving courts are biased downward.[46] Third, holdup problems are particularly severe for services inputs, which account for a substantial fraction of cost in some industries; hence court improvement could lead to substantial additional cost reductions (Boehm forthcoming). Formal enforcement is also likely to be important for the smallest firms that may be less able to rely on reputation. Moreover, courts matter for economic outcomes beyond their role in mitigating holdups with suppliers, notably by improving contracting between managers

45. In Online Appendix G.5, we posit such a model and study how our counterfactual exercises might change. We show that if the household's elasticity of substitution across industries (η) is one, as we have assumed in our calibration, then the improvement in aggregate productivity is the same as in our baseline.

46. That said, Johnson, McMillan, and Woodruff (2002) bring survey evidence that the role of courts in determining supplier switching is about five times as large for relationship-specific inputs than for standardized inputs, suggesting that most of the productivity gains are correctly accounted for.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

and workers (Besley and Burgess 2004; Bloom, Sadun, and Reenen 2012), and improving access to financing (Visaria 2009).

V. Conclusion

This article studies the organization of production in the Indian manufacturing sector and how it relates to the quality of formal contract enforcement institutions. We make two main points: one about the within-industry heterogeneity and measurement of the organization of production, and a second one about how institutions shape intermediate-input use and aggregate productivity.

First, we show that there is considerable heterogeneity in technology and organization even within narrowly defined industries. These differences should be reflected in the level of aggregation at which researchers assume a homogeneous shape of the production function. We argue that information on input use can be helpful in understanding differences in organization and technology within industries: a plant that produces cotton cloth from raw cotton has a long vertical span of production and performs both spinning and weaving, whereas a plant that produces cotton cloth from cotton yarn will only perform the weaving and might therefore have different factor intensities. We provide a simple and flexible way of finding groups of firms that use the same production function (which we call recipes) using a statistical clustering algorithm.

Second, we find that the organization of production is shaped by the contracting environment. We find that slow enforcement of contracts impedes the use of relationship-specific materials. As a result, firms tilt their input basket toward the use of more standardized inputs, for which spot markets exist, and for which enforcement by courts is not necessary. We develop a multi-industry general-equilibrium model where firms source multiple inputs and choose the organizational form optimally to minimize the cost of production. We estimate the relative distortions associated with the use of relationship-specific inputs from first moments of cost shares, which, compared with the standard approaches of estimating input wedges, is more robust to potentially imprecise measurement of input use. Our results suggest that distortions associated with poor courts are sizable and that improving courts would increase welfare: reducing the average age of pending cases by a year would, on average, increase a state's aggregate productivity by about 2%.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

SCIENCES PO
PRINCETON UNIVERSITY

SUPPLEMENTARY MATERIAL

An Online Appendix for this article can be found at *The Quarterly Journal of Economics* online.

DATA AVAILABILITY

Data and code replicating tables and figures in this article can be found in Boehm and Oberfield (2020), in the Harvard Dataverse, doi: 10.7910/DVN/OGP2EK.

REFERENCES

Acemoglu, Daron, Pol Antràs, and Elhanan Helpman, "Contracts and Technology Adoption," *American Economic Review*, 97 (2007), 916–943.
Acemoglu, Daron, and Pablo D. Azar, "Endogenous Production Networks," *Econometrica*, 88 (2020), 33–82.
Acemoglu, Daron, and Simon Johnson, "Unbundling Institutions," *Journal of Political Economy*, 113 (2005), 949–995.
Alfaro, Laura, Davin Chor, Pol Antràs, and Paola Conconi, "Internalizing Global Value Chains: A Firm-Level Analysis," *Journal of Political Economy*, 127 (2019), 508–559.
Amirapu, Amrit, "Justice Delayed is Growth Denied: The Effect of Slow Courts on Relationship-Specific Industries in India," School of Economics, University of Kent, Technical Report, 2017.
Antràs, Pol, "Firms, Contracts, and Trade Structure," *Quarterly Journal of Economics*, 118 (2003), 1375–1418.
Antràs, Pol, and Alonso de Gortari, "On the Geography of Global Value Chains," NBER Working Paper no. 23456, 2017.
Antràs, Pol, Teresa C. Fort, and Felix Tintelnot, "The Margins of Global Sourcing: Theory and Evidence from US Firms," *American Economic Review*, 107 (2017), 2514–2564.
Asturias, Jose, and Jack Rossbach, "Grouped Variation in Factor Shares: An Application to Misallocation," Georgetown University, Technical report, 2019.
Baqaee, David Rezza, and Emmanuel Farhi, "Productivity and Misallocation in General Equilibrium," *Quarterly Journal of Economics*, 135 (2020), 105–163.
Bartelme, Dominick, and Yuriy Gorodnichenko, "Linkages and Economic Development," Mimeo, UC Berkeley, 2015.
Bas, Maria, and Vanessa Strauss-Kahn, "Input-Trade Liberalization, Export Prices and Quality Upgrading," *Journal of International Economics*, 95 (2015), 250–262.
Besley, Timothy, and Robin Burgess, "Can Labor Regulation Hinder Economic Performance? Evidence from India," *Quarterly Journal of Economics*, 119 (2004), 91–134.
Bigio, Saki, and Jennifer La'O, "Distortions in Production Networks," *Quarterly Journal of Economics*, 135 (2020), 2187–2253.
Bils, Mark, Peter J. Klenow, and Cian Ruane, "Misallocation or Mismeasurement?" Stanford King Center on Global Development Working Paper no. 599, 2017.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen, "The Organization of Firms across Countries," *Quarterly Journal of Economics*, 127 (2012), 1663–1705.
Boehm, Johannes, "The Impact of Contract Enforcement Costs on Outsourcing and Aggregate Productivity," *Review of Economics and Statistics*, forthcoming.
Boehm, Johannes, and Ezra Oberfield, "Replication Data for: 'Misallocation in the Market for Inputs: Enforcement and the Organization of Production'," (2020), Harvard Dataverse, doi: 10.7910/DVN/OGP2EK.
Bonhomme, Stéphane, and Elena Manresa, "Grouped Patterns of Heterogeneity in Panel Data," *Econometrica*, 83 (2015), 1147–1184.
Buera, Francisco J., and Ezra Oberfield, "The Global Diffusion of Ideas," *Econometrica*, 88 (2020), 83–114.
Caliendo, Lorenzo, Fernando Parro, and Aleh Tsyvinski, "Distortions and the Structure of the World Economy," NBER Working Paper no. 23332, 2017.
Caprettini, Bruno, and Antonio Ciccone, "Turnover Taxes and Productivity: Evidence from a Brazilian Tax Reform," University of Zurich, Technical report, 2015.
Chaney, Thomas, "The Network Structure of International Trade," *American Economic Review*, 104 (2014), 3600–3634.
Chemin, Matthieu, "Does Court Speed Shape Economic Activity? Evidence from a Court Reform in India," *Journal of Law, Economics, and Organization*, 28 (2012), 460–485.
Costinot, Arnaud, Jonathan Vogel, and Su Wang, "Global Supply Chains and Wage Inequality," *American Economic Review*, 102 (2012), 396–401.
Djankov, Simeon, Rafael La Porta, Florencio Lopez de Silanes, and Andrei Shleifer, "Courts," *Quarterly Journal of Economics*, 118 (2003), 453–517.
Eaton, Jonathan, Samuel Kortum, and Francis Kramarz, "Firm-to-Firm Trade: Imports, Exports, and the Labor Market," Brown University, Technical report, 2015.
Fadinger, Harald, Christian Ghiglino, and Mariya Teteryatnikova, "Income Differences and Input-Output Structure," University of Mannheim, Technical report, 2016.
Fally, Thibault, and Russell Hillberry, "A Coasian Model of International Production Chains," *Journal of International Economics*, 114 (2018), 299–315.
Goldberg, Pinelopi Koujianou, Amit Kumar Khandelwal, Nina Pavcnik, and Petia Topalova, "Imported Intermediate Inputs and Domestic Product Growth: Evidence from India," *Quarterly Journal of Economics*, 125 (2010), 1727–1767.
Grossman, Gene M., and Elhanan Helpman, "Integration versus Outsourcing in Industry Equilibrium," *Quarterly Journal of Economics*, 117 (2002), 85–120.
Growiec, Jakub, "A Microfoundation for Normalized CES Production Functions with Factor-Augmenting Technical Change," *Journal of Economic Dynamics and Control*, 37 (2013), 2336–2350.
Hazra, Arnab Kumar, and Bibek Debroy, *Judicial Reforms in India: Issues and Aspects* (Delhi: Academic Foundation, 2007).
Hopenhayn, Hugo A., "Firms, Misallocation, and Aggregate Productivity: A Review," *Annual Review of Economics*, 6 (2014), 735–770.
Houthakker, Hendrik, "The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis," *Review of Economic Studies*, 23 (1955), 27–31.
Hsieh, Chang-Tai, and Peter J. Klenow, "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 124 (2009), 1403–1448.
Iyer, Rajkamal, and Antoinette Schoar, "The Importance of Holdup in Contracting: Evidence from A Field Experiment," MIT, Technical report, 2008.
Johnson, Simon, John McMillan, and Christopher Woodruff, "Courts and Relational Contracts," *Journal of Law, Economics, and Organization*, 18 (2002), 221–277.
Jones, Charles I., "The Shape of Production Functions and the Direction of Technical Change," *Quarterly Journal of Economics*, 120 (2005), 517–549.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

———, "Input–Output Economics," in *Advances in Economics and Econometrics: vol. 2, Applied Economics: Tenth World Congress* (Cambridge: Cambridge University Press, 2013), 419.
Khandelwal, Amit, and Petia Topalova, "Trade Liberalization and Firm Productivity: The Case of India," *Review of Economics and Statistics*, 93 (2011), 995–1009.
Kortum, Samuel S., "Research, Patenting, and Technological Change," *Econometrica*, 65 (1997), 1389–1420.
La Porta, Rafael, Florencio Lopez de Silanes, Andrei Shleifer, and Robert W. Vishny, "Legal Determinants of External Finance," *Journal of Finance*, 52 (1997), 1131–1150.
Laeven, Luc, and Christopher Woodruff, "The Quality of the Legal System, Firm Ownership, and Firm Size," *Review of Economics and Statistics*, 89 (2007), 601–614.
Lagos, Ricardo, "A Model of TFP," *Review of Economic Studies*, 73 (2006), 983–1007.
Law Commission of India, "Report No. 245. Arrears and Backlog: Creating Additional Judicial (wo)manpower," Law Commission of India Technical Report, 2014.
Levchenko, Andrei A., "Institutional Quality and International Trade," *Review of Economic Studies*, 74 (2007), 791–819.
Lim, Kevin, "Endogenous Production Networks and the Business Cycle," University of Toronto, Technical report, 2018.
Liu, Ernest, "Industrial Policies in Production Networks," *Quarterly Journal of Economics*, 134 (2019), 1883–1948.
Lu, Dan, Asier Mariscal, and Luis Fernando Mejia, "Imports Switching and the Impact of Large Devaluation," University of Rochester, Technical report, 2013.
Mangin, Sephorah, "A Theory of Production, Matching, and Distribution," *Journal of Economic Theory*, 172 (2017), 376–409.
Moon, Hyungsik Roger, and Martin Weidner, "Linear Regression for Panel with Unknown Number of Factors as Interactive Fixed Effects," *Econometrica*, 83 (2015), 1543–1579.
Mukoyama, Toshihiko, and Latchezar Popov, "Industrialization and the Evolution of Enforcement Institutions," *Economic Theory*, 69 (2020), 745–788.
Narasappa, Harish, and Shruti Vidyasagar (eds.), *State of the Indian Judiciary: A Report by Daksh* (Lucknow: Eastern Book Company, 2016).
Nunn, Nathan, "Relationship-Specificity, Incomplete Contracts, and the Pattern of Trade," *Quarterly Journal of Economics*, 122 (2007), 569–600.
Oberfield, Ezra, "A Theory of Input-Output Architecture," *Econometrica*, 86 (2018), 559–589.
Osotimehin, Sophie, and Latchezar Popov, "Misallocation and Intersectoral Linkages," University of Quebec at Montreal, Mimeo, 2020.
Pavcnik, Nina, "Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants," *Review of Economic Studies*, 69 (2002), 245–276.
Ponticelli, Jacopo, and Leonardo S. Alencar, "Court Enforcement, Bank Loans, and Firm Investment: Evidence from a Bankruptcy Reform in Brazil," *Quarterly Journal of Economics*, 131 (2016), 1365–1413.
Rauch, James E., "Networks versus Markets in International Trade," *Journal of International Economics*, 48 (1999), 7–35.
Restuccia, Diego, and Richard Rogerson, "Policy Distortions and Aggregate Productivity with Heterogeneous Establishments," *Review of Economic Dynamics*, 4 (2008), 707–720.
Robinson, Nick, "Judicial Architecture and Capacity," in *The Oxford Handbook of the Indian Constitution*, Sujit Choudhry, Madhav Khosla, and Pratap Bhanu Mehta, eds. (Oxford: Oxford University Press, 2016), 330–348.

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021

Rotemberg, Martin, and T. Kirk White, "Measuring Cross-Country Differences in Misallocation," New York University, Technical report, 2019.
Taschereau-Dumouchel, Mathieu, "Cascades and Fluctuations in an Economy with an Endogenous Production Network," Cornell University, Technical report, 2020.
Tibshirani, Robert, and Guenther Walther, "Cluster Validation by Prediction Strength," *Journal of Computational and Graphical Statistics*, 14 (2005), 511–528.
Tintelnot, Felix, Ken Kikkawa, Magne Mogstad, and Emmanuel Dhyne, "Trade and Domestic Production Networks," Unpublished Manuscript, University of Chicago, 2017.
Visaria, Sujata, "Legal Reform and Loan Repayment: The Microeconomic Impact of Debt Recovery Tribunals in India," *American Economic Journal: Applied Economics*, 1 (2009), 59–81.
Ward, Joe H., Jr, "Hierarchical Grouping to Optimize an Objective Function," *Journal of the American Statistical Association*, 58 (1963), 236–244.
World Bank, *Doing Business 2017: Equal Opportunity for All* (Washington, DC: World Bank, 2016)

Downloaded from https://academic.oup.com/qje/article/135/4/2007/5858011 by Fondation Nationale Des Sciences Politiques user on 11 February 2021