

Endogenous markups, input misallocation and geographical supplier access

Gregor Singer[‡]

17th June, 2019

[LINK TO LATEST VERSION](#)

Abstract

Inefficient allocation of inputs across firms has gained a prominent role in explaining development outcomes. Yet, inferring the costs of misallocation is challenging. Ignored firm heterogeneity from technology and demand biases the inferred costs from misallocation upwards or downwards. This paper develops and estimates a structural model that disentangles fundamental heterogeneity on the demand side from input misallocation distortions. Counterfactual analysis is performed by comparing equilibria in an oligopolistic setting with differentiated products. This enables comparative statics for a rich set of outcomes at any level of aggregation, as well as estimating their uncertainty. Instead of the usual TFP “accounting” approach that relies on aggregate production functions, exact consumer and producer welfare measures are used and aggregate inputs allowed to adjust. Using plant quantity and price data from the Indian iron and steel industry, I find no losses in aggregate labour or aggregate material productivity from misallocation as both total output and total input use are affected. Welfare losses, however, are large, equivalent to 31% of sales. The incidence is higher on consumers than producers, driven by higher prices. Perhaps surprisingly, welfare losses due to misallocation in material input markets are 90% larger than those from misallocation of labour. Geographical access to the relevant input suppliers through the transportation network is found to be a significant driver of material misallocation. A one standard deviation increase in access to suppliers reduces the material distortion by a third of its standard deviation. I argue that these distortions represent differences in indirect costs of trade.

JEL: L61, D24, L13, R12, D61, O18, R40

Keywords: misallocation, endogenous markups, structural estimation, welfare, materials, steel, supplier access

[‡]First draft: December 2016. London School of Economics and Political Science. E-mail: g.a.singer@lse.ac.uk. I thank Antony Millner, Ben Groom, Gharad Bryan, Francesco Caselli, Swati Dhingra, Steve Gibbons, Paul Heidhues, John Haltiwanger, Pasquale Schiraldi, Daniel Sturm, Lindsay Relihan, Martin Rotemberg, André Romahn, Sefi Roth and Eric Verhoogen for valuable comments and discussion as well as seminar participants at LSE in 2016, 2017, 2018, and 2019, at the World Congress for Environmental and Resource Economists 2018, the Congress of the European Economic Association 2018, the Econometric Society European Winter meetings 2018, Symposium of the Spanish Economics Association 2018, the Royal Economic Society PhD meetings 2018, a DICE seminar 2019, and the Royal Economic Society Annual Conference 2019. All errors are my own.

1 Introduction

Misallocation of production factors between firms can result in large losses of aggregate income. In their seminal paper, [Hsieh and Klenow \(2009\)](#) estimate aggregate TFP losses of 40%-60% in Indian manufacturing. Distortions such as preferential access to credit, labour regulations that depend on firm size, or political connections have been identified, amongst others, as causes for input misallocation in this emerging literature.¹

In this paper, I address a problem of ignored heterogeneity across firms when calculating misallocation losses. In the growing misallocation literature following [Hsieh and Klenow \(2009\)](#), ignored heterogeneity across plants in terms of production or demand is conflated with misallocation distortions in a non-trivial way. As a result, inferred welfare costs from input misallocation can be upwards *or* downwards biased. In the application of this paper, ignoring heterogeneity in markups would understate the true costs of input misallocation by up to 27%. The main contribution of this paper is that I disentangle plant level demand and production heterogeneity from input distortions. A second contribution is that I am able to distinguish between different aggregate outcomes and the incidence on consumers and producers, as well as estimating the uncertainty around misallocation losses. As a third contribution, this is the first paper to present evidence that these precisely estimated input distortions capture differences in geographical access to suppliers.

I focus on input misallocation between cast iron producers in India.² India's manufacturing sector is an interesting case for studying misallocation considering debates³ in the literature about the contribution of reforms to resource allocation. Misallocation could have also played a role in India's slow structural transformation compared to China and other East Asian nations despite its deep economic reforms in the early 90s ([Bhagwati and Panagariya, 2014](#)).

While the aggregate consequences of misallocation are usually expressed in aggregate TFP losses, we have not been able to say much about other margins. India has created a Resource Efficiency Panel in 2015 and sectoral material productivity is more generally receiving growing attention from an industrial competitiveness and environmental agenda (e.g. [OECD, 2015](#); [European Commission, 2013](#)). The substantial emissions of the steel sector are in part determined by its aggregate material productivity. Most research focuses on how innovation and technology diffusion can improve sectoral material productivity, but there is

¹See e.g. [Gopinath et al. \(2017\)](#) or [Midrigan and Xu \(2014\)](#), [Garicano et al. \(2016\)](#), and [Akcigit et al. \(2018\)](#) respectively. See [Restuccia and Rogerson \(2017\)](#) or [Hopenhayn \(2014a\)](#) for recent surveys. Appendix A presents a brief overview of additional related literature.

²Cast iron is an important product in India's manufacturing sector. It has one of the highest shares of any single product in manufacturing output. India's steel sector has a 15% share in total manufacturing value added which is one of the highest in the world ([UNIDO, 2016](#)).

³See e.g. [Bollard et al. \(2013\)](#); [Harrison et al. \(2013\)](#) vs. [Nishida et al. \(2014, 2015\)](#).

little evidence on potential allocative gains. With the methodology in this paper, I can recover the effects of misallocation on input productivities, which I can distinguish from consumer welfare or producer profits, and can thus shed light on this question.

It is worth to briefly conceptualise the misallocation distortions and to motivate why we need to do disentangle those from fundamental heterogeneity in the first place, before I introduce the nature of the counterfactuals and the role of supplier access. The input distortions (or “wedges”) are usually measured by the plant level gaps between the marginal revenue product (MRP) of an input and the input price. Distortions can rationalise the existence of these gaps and represent additional unobserved costs for a particular plant from using that particular input. These additional cost could arise through policies like taxes and subsidies, information frictions, transaction costs, corruption, extortion, indirect trade costs or other input constraints that vary across firms. The bundle of all these potential input distortions represent the gap between the MRP and input prices. In the absence of these distortions the MRP should be equal to the input price. In the presence of the distortions there is misallocation. Intuitively, moving a unit of an input from a low gap and low MRP plant to a plant with a larger gap and higher MRP increases aggregate output through a more efficient allocation. In the privately optimal equilibrium, plants that face larger distortions underutilise the input compared to the socially optimal outcome.

The reason why we should disentangle the distortions from fundamental heterogeneity is because the MRP and these gaps are not directly observed.⁴ The distortions (or the gaps) are determined by the demand elasticities (or markups), output elasticities and the revenue share of an input.⁵ Even if we were correct on average and some distortions are over- while others underestimated, I show that mismeasurement still impacts welfare conclusions. It matters which plants face which distortions, for example whether it is the more productive plants facing the more severe distortions.⁶ Any deviations of the assumed demand or output elasticities

⁴There has been some debate on the role of data management and measurement error. See [Rotemberg and White \(2017\)](#) and [Bils et al. \(2017\)](#) respectively.

⁵To organise thoughts, take the variance of MRP which is often used as a statistic of misallocation along the rationale described above, assuming constant input prices. Consider a simple profit maximisation problem $PQ - vX$, where either prices P depend on quantity Q or quantity on prices, and maximising with respect to input X . Without parametric functional form assumptions, we can write the first order condition in terms of the variance of the logged MRP of an input. It is equal to the variance of a combination of the inverse demand elasticity η , the output elasticity α and data on inputs and revenues:

$$\text{Var}[\log MRP] = \text{Var}\left[\log(1 + \eta) + \log(\alpha) + \log\left(\frac{PQ}{X}\right)\right]$$

Typically, the first two terms are assumed constant (within a sector). If, in reality, demand elasticities (or markups) or output elasticities are not constant, variation in measured MRP no longer imply input distortions, but could capture demand or production heterogeneity. Vice versa, a constant MRP does not imply the absence of input distortions or misallocation. Once we account for heterogeneity, the variance in MRP can go up or down, depending on the correlation between the terms.

⁶[Restuccia and Rogerson \(2008\)](#) show that the correlation between firm TFP and distortion matters.

from the real ones are captured by the input distortions, while they in fact are differences in demand or production technique.⁷ The literature following [Hsieh and Klenow \(2009\)](#), for example, assumes constant production elasticities within 4-digit sectors, and constant demand elasticities dictated by CES. There is, however, a large body of evidence that markups can vary across firms even within narrow industries.⁸ Depending on the relationship between the true markups, true output elasticities and the true input distortions, the variation of input distortions across plants can increase *or* decrease when it is mismeasured. As a result, ignoring heterogeneity biases the inferred costs from misallocation upwards or downwards.

One might argue that capturing ignored heterogeneity in the inferred input distortions could be desirable, at least when it comes to heterogeneity in markups. We would capture the bundle of distortions on the input and demand side that represent deviations from a CES framework, such as excess market power. Along the lines of the theory of the second best, we only care about the joint effect of all distortions. If the firm with the additional benefits on inputs (overusing input) also has more market power (under-producing), then these distortions could offset each other in a second best world. However, when we capture ignored demand heterogeneity in input distortions, we preclude any economic evaluation and welfare analysis of the distortions. This is because a model that is based on constant demand elasticities, such as CES, cannot be used to evaluate welfare losses if we believe that we capture variable demand elasticities. We would have to change the primitives of the model as well. On the contrary, with the approach in this paper, where I isolate the input distortions from variation in demand elasticities, we learn about whether the theory of the second best applies. In a second best world, removing only the input distortions for the counterfactual would lead to a *decrease* in welfare, a clear prediction that I can test and will reject.

The way that I disentangle misallocation from fundamental heterogeneity is through a combination of product level focus on the production side and estimating flexible endogenous demand elasticity, facilitated by using detailed quantity and price data on both outputs and inputs for a panel of plants. This is the first paper to estimate structural demand and production systems to combine it into a welfare framework to disentangle input distortions from endogenous markups.⁹

⁷This is also illustrated by the fact that the equation for input distortions in [Hsieh and Klenow \(2009\)](#) is the same as the equation for markups in the popular [De Loecker and Warzynski \(2012\)](#).

⁸See e.g. [Nevo \(2001\)](#), [De Loecker et al. \(2016\)](#) or [Hottman et al. \(2016\)](#).

⁹Some recent work incorporate separate but *exogenous* markup variation, e.g. [Ho and Ruzic \(2017\)](#) for different sectors, or [Lenzu and Manaresi \(2018\)](#), [Tortarolo and Zarate \(2018\)](#), [Haltiwanger et al. \(2018\)](#) and [Eslava and Haltiwanger \(2019\)](#) at the firm level. [Haltiwanger et al. \(2018\)](#) also show that some assumptions of the [Hsieh and Klenow \(2009\)](#) model do not hold using more detailed data from the US and argue that sharper estimates of distortions that are isolated from heterogeneity on the demand and production side indeed hold more informative signals. Related are also [Bayer et al. \(2018\)](#) and [Liang \(2017\)](#). [Peters \(2013\)](#) and [Edmond et al. \(2018\)](#) on the other hand focus on endogenous markups, but are not separating it from input distortions

On the production side, I estimate production functions for a single product (cast iron), which is much narrower than the usual production functions that are assumed to be identical for all plants within 4- or 2-digit sectors.¹⁰ The resulting product level output elasticities are much less likely to ignore production heterogeneity. Importantly, I am also able to address output and input price bias by estimating gross output production functions using observed quantities of outputs and inputs.¹¹ I recover plant level total factor productivities (TFPQ) based on the proxy method (Olley and Pakes, 1996; Wooldridge, 2009).

On the demand and output side, I observe heterogeneity across plants in the data – even for this single cast iron product category. There is significant variation in output prices, both in gross prices and prices net of sales tax, excise duty and other distributional and transport fees. This suggests a setting with differentiated products and different demand conditions. I build on the Berry et al. (1995, 1999) random utility mixed model framework embedded in an oligopolistic setting, where I identify the demand parameters using production cost shifters as instruments. The estimated demand elasticities (and markups) are flexible and endogenous.¹²

Having a clean measure of input distortions is only a first step, as it does not tell us anything about the welfare costs of misallocation, which requires a counterfactual analysis. What is an appropriate counterfactual? This naturally depends on the question being asked. I construct counterfactuals, where I remove input distortions such that there are no gaps between MRP and input prices, and search for an equilibrium where all firms are best responding to each other. The estimated welfare gains can be interpreted as the gains from removing misallocation distortions that would be obtained under an oligopolistic market environment, as opposed to comparing it to a socially planned allocation for example (Behrens et al., 2018). The counterfactuals and the misallocation losses are determined by the change in the estimated distortions and by the endogenous changes in prices and quantities of all plants.

There are five further features of the counterfactual analysis worth highlighting that set this study apart from most of the input misallocation literature. First, demand elasticities, markups and marginal cost pass-through are endogenous, as they depend on the prices of

empirically. Baqaee and Farhi (2017) study misallocation in general equilibrium focusing on markups. Also related is De Loecker and Scott (2016), who estimate markups from the production and the demand side to compare them.

¹⁰I estimate production functions for single product plants in a single 7-digit product category. Boehm and Oberfield (2018) argue that there is significant production heterogeneity even within narrowly defined 4- or 2-digit sectors in India.

¹¹When using deflated revenue, we conflate markups with physical output, stressed e.g. by Gandhi et al. (2016); Marin and Voigtländer (Forthcoming); Foster et al. (2008). When using deflated input expenditures instead of observed input quantities, we are likely to conflate quality with quantity, see Katayama et al. (2009); De Loecker (2014) and also Kugler and Verhoogen (2012).

¹²The demand elasticities are more flexible than in the Kimball (1995) model, used e.g. in Klenow and Willis (2016) or Edmond et al. (2018), where they are strictly decreasing (i.e. less elastic) in output share.

all plants and demand and production fundamentals. The counterfactuals are significant changes to prices and the economy, and restricting markups to exogenous factual levels biases misallocation estimates.¹³ Second, *aggregate* inputs are allowed to adjust endogenously in all equilibria. Allocative efficiency gains not only tend to increase total output, but may also affect total input use.¹⁴ Third, I can account for observed input price differences, due to local labour markets, or input quality difference, for example.¹⁵ Fourth, all comparative statics are at the plant level, which permits a rich analysis of outcomes at any level of aggregation. Fifth, I can recover standard errors of all estimated distortions, welfare losses and other comparative statics, which is novel to the input misallocation literature. This is because I avoid calibration, and instead estimate the model and every parameter directly.¹⁶ In addition, I am able to show which fundamental parameters are driving the uncertainty in estimated misallocation losses. It turns out that the estimated returns to scale are a main driver of the size of the misallocation losses, which underscores the importance of estimating it as well as providing estimates of uncertainty.

There is a trade-off between focusing on a single industry which delivers less biased misallocation costs and analysing misallocation for the entire manufacturing sector. While the focus in the literature has been on the latter, this paper emphasises the former to maximise the signal in the distortions. Not only could the misallocation costs vary substantially across sectors, but the potential bias when ignoring heterogeneity makes it difficult to draw meaningful conclusions. Ultimately, with increasing availability of quantity (and price) data on inputs and outputs, we can get a better grip on misallocation costs for growing parts of the economy.¹⁷

The estimated welfare costs from labour and material input misallocation are substantial and equivalent to around 31% of the sales from the plants in the sample.¹⁸ This is also

¹³Markup adjustment turn out to be important. I find markups adjust in the counterfactual for individual plants to a degree that is comparable with the original deviations of plant markups from the average markup.

¹⁴Recent work by [Catherine et al. \(2018\)](#) on collateral constraints and investment shows that the aggregate input changes are important to such an analysis. Yet, the key literature restricts aggregate inputs to be constant. I account for aggregate input changes and assume that inputs are elastically supplied given that I analyse a small industry.

¹⁵The size of the misallocation losses are also determined by input prices. [Hsieh and Klenow \(2009\)](#) need to assume constant factor prices across firms for their counterfactual analysis. Conceptually, the input distortions are separate from input prices in the literature, otherwise they would show up in input expenditures. I am able to reduce bias stemming from this source by using observed input prices. See [Cheng and Morrow \(2018\)](#), for example on factor price differences due to local labour markets in China.

¹⁶By bootstrapping from the estimated parameters' covariance structures I can provide confidence intervals around misallocation distortions or any other outcome.

¹⁷There can be spillovers into other sectors. [Jones \(2011, 2013\)](#) studies complementarities between sectors through input-output links. [Behrens et al. \(2018\)](#) construct a general equilibrium model and show that distortions in one sector can impact distortions in other sectors.

¹⁸For the counterfactual analysis I remove misallocation distortions in input materials, labour or both, but any distortions in capital use from the factual are preserved in the counterfactual. This is because a large fraction of static capital distortions might actually be inherent adjustment costs (time-to-build) to

evidence that the economy with variable markups and input distortion does not constitute a second best. The estimated counterfactual gains in compensating variation for consumers are larger than the gains in firm profits, driven by price decreases from cost pass-through. To get a sense of the bias in welfare costs from ignoring demand heterogeneity, I pretend that demand elasticities are constant, infer the wrong distortions, and use my model to calculate the bias. I find that the estimated welfare cost is between 13% to 27% lower (depending on the counterfactual) when ignoring variable markups, so we would underestimate misallocation costs in this case.

Surprisingly, aggregate input productivities are hardly affected from misallocation. Even when defining a standard Cobb Douglas aggregate production function, there are no aggregate TFP gains. This seems at first unexpected because of the large literature on TFP gains. However, I show that we can realign the results with the literature when their TFP gains are not interpreted as pure production side productivity gains, but instead as welfare gains consistent with their implicit underlying demand model. Those TFP (i.e. welfare) results are comparable with the welfare results in this paper.¹⁹

The total welfare costs of misallocation of materials are around 90% larger than from misallocation of labour, and the difference is statistically significant. This is a surprising result, given that we usually think of materials as a more flexible input than labour and therefore associated with fewer distortions. While the literature often abstracts from intermediates entirely with value added production functions, distortions in materials markets appear to be important and costly, at least for the cast iron industry in India.

In the last part of the paper, I ask, what is causing these costly input material distortions and misallocation losses? The steel industry relies on extensive shipping of heavy and bulky material inputs. Geography and transport infrastructure, particularly railroads, are thus important features in the production process. Any issues with sourcing are likely to play a role in the documented misallocation from material input distortions. I begin the last part by analysing the issues in Indian freight transport. Sourcing inputs through the transportation network in India is characterised by frequent delays and uncertainty. Infrastructure is outdated, leading to breakdowns, freight trains share congested tracks with passenger trains, and there are state border checkpoints for tax purposes that delay shipments, to name a few. With longer sourcing routes, these issues are likely to become more severe.

changes in the capital stock. As [Asker et al. \(2014\)](#) show capital could be much more optimally allocated in a dynamic sense. [David and Venkateswaran \(2017\)](#) address this by explicitly modelling capital dynamically with adjustment costs.

¹⁹Furthermore, note that the counterfactual gains in this paper incorporate the full effects of misallocation including changes in aggregate inputs. When holding aggregate inputs fixed, any output gains would be necessarily attributed to aggregate TFP.

Shipping fees also increase with longer sourcing routes. The advantage of this analysis is that I observe plant specific input prices that are measured at the factory gate. They explicitly contain input shipping fees and the model accounts for differences in these observed input prices. Any differences in estimated input distortions are therefore net of input shipping fees and the measure of supplier access is not simply picking up those. In the trade literature, shipping fees are often called direct trade costs. There is a large body of evidence that suggest that direct trade costs cannot account for the implied total costs of trade ([Anderson and Van Wincoop, 2004](#)). There are indirect costs of trade, such as search costs, contracting costs or costs associated with delay and uncertainty. The freight transport issues in India, such as uncertainty, are examples of such indirect costs of trade.²⁰ While the estimated input distortions are net of shipping fees, they would certainly capture differences in indirect trade costs. And if indirect trade costs increase with poorer geographical access to suppliers, the estimated material input distortions should decrease with better supplier access. This paper tests this hypothesis.

I construct a measure of supplier access by combining the cost to reach suppliers with the size of supplier industries in the around 540 districts of India. The cost to reach suppliers is based on the fastest path between any bilateral pairs of districts. I collect geo-located data on the entire rail and road infrastructure in India. Using information on the types of rails and roads (e.g. motorways vs. tertiary roads), I construct a weighted network graph to compute the fastest path between district pairs. Over the sample period from 2000 to 2012, there was hardly any variation in the railway infrastructure, which is the main mode of transport for this industry. Therefore, the infrastructure component is time invariant in the measure of supplier access. Even if there was time-variation, placement of railway infrastructure, as well as the location choice of plants is non-random.²¹ To address these issues, I condition the analysis on district fixed effects, and therefore use the variation in supplier access that is driven by the expansion and contraction of supplier industries in distant districts over time.

A one standard deviation increase in supplier access is associated with almost a third of a standard deviation decrease in the input material distortion. The estimate is robust to using lagged supplier access, addressing potential reverse causality concerns. To provide more evidence for the causality of the relationship, I present three types of placebo test. The first shows that supplier access has no significant relationship with the estimated labour distortion. The second shows that a measure of access to irrelevant “supplier” industries, e.g. textiles,

²⁰While storage can smooth over uncertainties, it is costly and therefore also part of indirect costs of trade.

²¹Endogenous infrastructure placement is a central challenge in the literature on the effects of infrastructure investments. There are some strategies to address this. [Faber \(2014\)](#), for example, constructs a hypothetical infrastructure network based on construction costs and a minimum spanning tree. [Banerjee et al. \(2012\)](#) use the areas on the straight line between start and end point of a transportation link.

rubber or food, is not related to the material input distortion. The third placebo test shows that access to markets on the output side cannot explain the distortion on the input side. It is thus only the access to relevant input suppliers that is associated with the material input distortions that drive misallocation losses.

The policy implications of this analysis are nuanced. Differences in shipping fees mirror the geographic reality that shipping goods across space is costly. While theoretically beneficial for manufacturing, it is also costly to build infrastructure to equalise shipping fees for all plants. The findings of this paper are conditional on the observed differences in input shipping fees, however. The distortions pick up the indirect trade costs, and those are in turn driven by supplier access. Addressing reliability, delay and shipping uncertainty, for example, could reduce the impact of supplier access on indirect trade costs, and could thus reduce misallocation losses.

This paper contributes to a rapidly growing literature on the size and determinants of misallocation in manufacturing industries, which has focused on capital and labour so far, for which I provide a better overview in Appendix A. Most of the literature on measuring misallocation abstracts from intermediate inputs entirely by using value added production functions.²² There is even less work on potential determinants of misallocation of material inputs. As an exception, [Boehm and Oberfield \(2018\)](#) study misallocation of intermediate inputs in India due to court congestion that generates a hold up problem in their model. In current work in progress, [Hornbeck and Rotemberg \(2019\)](#) examine the impact of railroad expansion on input reallocation in historic US manufacturing using a growth decomposition.²³ This paper aims to address the literature gap in misallocation of input materials.

The last part of the paper is also related to the literature studying the effect of transport networks on productivity and welfare.²⁴ Perhaps most closely related are papers that examine the effects of changes in market access.²⁵ [Donaldson and Hornbeck \(2016\)](#) analyse the contribution of historical railroads on agricultural productivity in the US. [Alder \(2017\)](#) compares the road construction projects in India with a counterfactual following a Chinese infrastructure model. [Allen and Atkin \(2016\)](#) estimate the impact of road expansion in

²²Exceptions are [Jones \(2013\)](#) or [Dias et al. \(2016\)](#), for example.

²³The citation of [Hornbeck and Rotemberg \(2019\)](#) refers to a presentation at LSE and conversations with one of the authors. At the time of writing, there was no working paper version available.

²⁴There is a longer literature that analyses the effects of transport infrastructure starting with [Fogel \(1964\)](#), which as received growing attention more recently. [Allen and Arkolakis \(2014\)](#), [Allen and Arkolakis \(2019\)](#) and [Fajgelbaum and Schaal \(2017\)](#) develop frameworks for estimating the impact of transport infrastructure investment on welfare in spatial equilibrium, where the latter two also account for traffic congestion. See [Redding and Rossi-Hansberg \(2017\)](#) for a review.

²⁵[Redding and Venables \(2004\)](#) provide a theoretical foundation of both market access on the output side and supplier access on the input side. [Redding \(2010\)](#) and references therein provide a summary of the earlier literature on market access.

India on the volatility of grain prices and farmer production choices, and [Huang and Xiong \(2018\)](#) analyse Chinese road expansion. As in this paper, they use geo-located transport infrastructure data to compute intra-national trade costs and a measure of market access.

Two aspects set this paper apart from this literature. First, these studies use access to output markets where the entire economic activity of other regions are taken into account. In contrast, I am using a measure of supplier access on the input side. Moreover, I only take access to *relevant* potential suppliers into account. Naturally, to study distortions on intermediate inputs, the access to these suppliers instead of market access is the object of interest. Second, while these studies rely on variation in infrastructure construction for changes in market access, I use growth in distant supplier industries for variation in supplier access. Appendix A relates this paper also to the literature that estimates the intra-national costs of trade from price differences instead of using fastest path algorithms ([Fackler and Goodwin, 2001](#); [Anderson and Van Wincoop, 2004](#); [Atkin and Donaldson, 2015](#); [Donaldson, 2018](#); [Asturias et al., 2018](#)).

The remainder of the paper begins by setting up the model and estimation strategy in Section 2. First I set up the firm problem and how they interact. This allows me to derive an expression for the input distortions which depends on production and demand parameters. I then present the production function estimation, demand estimation and welfare framework before I describe the counterfactual equilibria used for comparative statics. Section 3 presents the plant data along with some descriptive statistics as well as the construction of a weighted network graph from geo-located rail and road data. Section 4 briefly discusses the results for the production and demand estimation as well as descriptive statistics on the estimated input distortions. Section 5 analyses misallocation losses from the counterfactual exercise. In Section 6, I start with a discussion of the relevant issues in Indian freight transportation, which are likely to be captured as indirect trade costs in the estimated input distortions. I then construct a measure of supplier access, discuss the identification strategy and the results along with placebo tests. Section 7 concludes.

2 Model and estimation strategy

This section sets up the model in order to identify the misallocation distortions, estimate the structural production and demand side parameters, and calculate the counterfactuals. With a slight abuse of terminology, I use the term firms and plants interchangeably. To fix ideas, we should think of single plant and single product firms, which reflects the data I use.

2.1 Firm behaviour and input misallocation

2.1.1 Market structure

Suppose that firms interact in a market where each firm $j \in J$ is a single product firm and is selling a differentiated product j .²⁶ The firms compete strategically on prices in a Bertrand-Nash fashion to maximise profits in each market (period) t separately. In the Bertrand-Nash equilibrium, all firms are individually profit maximising and best responding to each other.

Due to product differentiation, as well as heterogeneous cost structures, firms charge different prices with different markups in equilibrium. Product differentiation is consistent with the data in the application of this paper, where prices vary across cast iron plants and are only weakly correlated with quantities. This requires differentiation, for example in terms of quality or geography. The elasticity of one firm's demand to another firms' prices varies by firm-pair. Therefore, firms with a unique high quality product can behave like a monopolist in a high quality segment. Similarly, this framework also allows for regional oligopolies where outsider firms' prices have little impact on local demand. These features will be captured by endogenous demand (cross-) elasticities as described in Section 2.3.

2.1.2 Firm profit maximisation

Firms maximise profits according to:

$$\max_{P_{jt}} P_{jt} Q_{jt}(\mathbf{P}_t) - C(Q_{jt}(\mathbf{P}_t), \mathbf{c}_{jt})$$

where P_{jt} is the output price of firm j at time t , and Q_{jt} is the output quantity which depends on the vector of output prices of all firms \mathbf{P}_t .²⁷ The equilibrium output prices and quantities are not necessarily equal to realised prices and quantities. This is because we introduce an unforeseeable zero mean shock to production later, which pins down realised quantities. The equilibrium strategies of the (risk-neutral) firms do not depend on the shock, which occurs after all choices have been made. The cost function $C(\cdot)$ depends on output quantity and the vector of firm-time specific cost parameters and shifters \mathbf{c}_{jt} . The first order condition for the

²⁶In the application, we can include multi-product firms that produce this differentiated product for the demand estimation.

²⁷Both output prices and input prices are at the factory gate. That is output prices are net of taxes, excise duties and shipping fees. On the input side, shipping fees are captured by the price of material inputs.

firm's profit maximisation problem is:

$$Q_{jt}(\mathbf{P}_t) + \frac{\partial Q_{jt}(\mathbf{P}_t)}{\partial P_{jt}} \left(P_{jt} - MC_{jt}(Q_{jt}(\mathbf{P}_t), \mathbf{c}_{jt}) \right) = 0;$$

where I use the standard definition of the marginal costs MC_{jt} which is allowed to change with output quantity. We can rewrite this as:

$$\frac{1}{1 + \eta_{jt}(\mathbf{P}_t)} = \frac{P_{jt}}{MC_{jt}(Q_{jt}(\mathbf{P}_t), \mathbf{c}_{jt})} \quad (1)$$

where the inverse demand elasticity is defined as $\eta_{jt} \equiv \frac{\partial P_{jt}}{\partial Q_{jt}(\mathbf{P}_t)} \frac{Q_{jt}(\mathbf{P}_t)}{P_{jt}}$, which is allowed to be firm and time specific and is endogenous depending on the prices of all firms. Equation (1) is the familiar relationship between the price elasticity of demand and the markup of prices over marginal costs used in the Learner index and reflects market power. The degree of market power falls with more elastic demand.

2.1.3 Input cost minimisation and input distortions τ

The input distortions enter in the cost minimisation problem. Before defining the firms problem, I highlight two assumptions. First, I follow recent literature (e.g. [De Loecker et al., 2016](#)) and frame the input cost minimisation problem as a short term cost minimisation of achieving the firm's required output Q_{jt} by choosing labour L_{jt} and materials M_{jt} conditional on installed capital K_{jt} . This avoids specifying a dynamic condition for capital optimisation, but it also precludes analysing distortions in capital inputs, as there is no capital demand condition to derive capital distortions from. As [Asker et al. \(2014\)](#) show, using a static condition for capital, a dynamic input, can be misleading when inferring distortions. Hence I only analyse misallocation in material and labour markets and preserve the mix of capital misallocation and adjustment costs contained in the unobserved rental rate across factual and counterfactual scenarios.²⁸

Second, inputs are elastically supplied. That is, firms are assumed to be price takers on the input side, consistent with a setting where they are relatively small players in the labour market and material input market. Appendix B discusses this assumption. Importantly, I show that input market power (monopsony power) would be captured by the input distortions. I present evidence that input market power is not likely in this setting, in favour of the

²⁸From a meta perspective, I believe we could interpret such adjustment costs as misallocation *if* they are not an inherent feature of production. That is if they are possible to change. Adjustment costs for materials, if they exist, are likely to be small and reducible, as the period of analysis are years.

assumption of elastic input supply. The firms minimise short run costs according to:

$$\begin{aligned} \min_{L_{jt}, M_{jt}} \quad & r_{jt}K_{jt} + \tau_{jt}^L w_{jt}L_{jt} + \tau_{jt}^M P_{jt}^M M_{jt} \\ \text{s.t.} \quad & F(K_{jt}, L_{jt}, M_{jt})\Omega_{jt} \geq Q_{jt} \end{aligned}$$

where the input prices are the rental rate r_{jt} , wages w_{jt} and materials price P_{jt}^M . Variation in input prices across firms can arise through using inputs of different quality, like more expensive materials or higher skilled workers.²⁹ The production function $F(\cdot)$ is assumed to have the same structure for all firms, since all firms produce the same 7-digit product, which is substantially narrower than a 4-digit sector. If a firm uses higher quality inputs, it does not produce more outputs, but higher quality outputs (with higher prices), such that the physical relationship of the weight of outputs and inputs is the same for high and low quality products. Differences in this relationship across firms are captured by firm specific Hicks-neutral total factor productivities Ω_{jt} .

Finally, the τ_{jt}^L and τ_{jt}^M are material and labour cost multipliers, that differ across firms and capture input misallocation. I follow the key literature in modelling this as a wedge that captures a range of distortions, as in [Chari et al. \(2007\)](#), [Restuccia and Rogerson \(2008\)](#) or [Hsieh and Klenow \(2009\)](#). Essentially, firms are assumed to behave optimally *given* their individual environment, constraints and distortions, which will enable us to infer the distortions τ . I next discuss the interpretation of these distortions.

2.1.4 Interpretation and identification of input distortions τ

Firms that face a higher τ_{jt}^M have higher additional costs associated with purchasing input materials, which drives wedges into the efficient allocation of inputs. What are these input distortion? The τ can be interpreted as plant specific “taxes” or “subsidies” on the particular input. They can be actual taxes due to e.g. firm size dependent policies regarding labour taxation³⁰, input subsidies for a subset of goods, or land and property rights regulation that affects firms differently³¹. They can also be advantages and windfalls through political connections ([Faccio, 2006](#); [Akcigit et al., 2018](#)), resulting in firms spending more than optimal amounts on certain inputs, i.e. a low τ . Other elements contained in τ are differential

²⁹As described in Section 5.1, since I measure labour L_{jt} in worked hours, I perform a robustness check where labour is measured as wage bill to capture skills, which can increase output quantity Q_{jt} not just sales price. More expensive high quality materials (per tonne) on the other hand should not increase output quantity, but rather the sales price.

³⁰See [Garicano et al. \(2016\)](#) for a study of such policies in France.

³¹E.g. [Duranton et al. \(2015\)](#). In the agricultural context see [Adamopoulos and Restuccia \(2014a\)](#); [Chari et al. \(2017\)](#); [Chen et al. \(2017\)](#).

overhead costs (e.g. legal and administration costs) associated with the input, market or informational frictions (David et al., 2016; Bloom et al., 2013; Allen, 2014), enforcement frictions (Boehm and Oberfield, 2018) and further plant specific barriers or advantages of using the input. Similarly, a model with constrained input access results in the same first order conditions.³² Importantly, differences in indirect trade costs such as uncertainty, delay or search costs would also be captured by the material input distortions. In Section 6, I present evidence that the material input distortions capture these indirect trade costs from differences in supplier accesses. Anything that incentivises or constrains the use of the input away from the optimum is captured in τ .

Using the definitions of the material and labour elasticities of output, $\alpha_{jt}^M \equiv \frac{\partial Q_{jt}}{\partial M_{jt}} \frac{M_{jt}}{Q_{jt}}$ and $\alpha_{jt}^L \equiv \frac{\partial Q_{jt}}{\partial L_{jt}} \frac{L_{jt}}{Q_{jt}}$, and that the Lagrange multiplier of the minimisation problem is the marginal cost of production MC_{jt} , the cost-minimising conditions for labour and materials $X \in \{L, M\}$ are:

$$\tau_{jt}^X P_{jt}^X - MC_{jt} \alpha_{jt}^X \frac{Q_{jt}}{X_{jt}} = 0 \quad (2)$$

Combining both first order conditions (1) and (2) yields the expression for τ_{jt}^M and τ_{jt}^L , which are identified if the parameters on the right hand side are identified:

$$\begin{aligned} \tau_{jt}^M &= (\eta_{jt} + 1) \alpha_{jt}^M \frac{P_{jt} Q_{jt}}{P_{jt}^M M_{jt}} \\ \tau_{jt}^L &= (\eta_{jt} + 1) \alpha_{jt}^L \frac{P_{jt} Q_{jt}}{w_{jt} L_{jt}} \end{aligned} \quad (3)$$

In standard imperfect competition models where $\tau_{jt}^M = 1$ (and analogously $\tau_{jt}^L = 1$), the markup adjusted output elasticity is equal to the expenditure share of the input, i.e. $(\eta_{jt} + 1) \alpha_{jt}^M = \frac{P_{jt}^M M_{jt}}{P_{jt} Q_{jt}}$, or the marginal revenue product of material is equal to the input price, so $MRPM_{jt} \equiv (\eta_{jt} + 1) \alpha_{jt}^M P_{jt} Q_{jt} / M_{jt} = P_{jt}^M$. We can rationalise measured gaps between the $MRPM_{jt}$ and input price P_{jt}^M with the distortions τ_{jt}^M . The τ_{jt}^M are associated with misallocation. Intuitively, reallocating inputs from a low τ_{jt}^M firm to a high τ_{jt}^M firm increases aggregate sales with the same amount of inputs used, because of the higher adjusted $MRPM_{jt}$ of the latter firm. By turning this logic around, we can infer the wedges τ_{jt}^M through the variation in the ratio of the estimated $MRPM_{jt}$ and observed material prices P_{jt}^M . Once we have recovered the τ_{jt}^M , we can ask how costly they are. This is the point of the

³²That is suppose the constrained cost minimisation is:

$$\min_{L_{jt}, M_{jt}} r_{jt} K_{jt} + w_{jt} L_{jt} + P_{jt}^M M_{jt} \quad s.t. \quad M_{jt} \leq \bar{M}_{jt} \quad \text{with multiplier } (\tau_{jt}^M - 1)$$

with the corresponding constraints for labour. See Peters (2013) for example.

counterfactual, where I remove the τ_{jt}^M (or τ_{jt}^L). Finally, Equation (3) demonstrates that any ignored heterogeneity in output elasticities α_{jt}^M or markups $1/(\eta_{jt} + 1)$ would be shifted to the left hand side and be captured by the input distortions. The misattribution problem has raised doubts in the recent review by [Restuccia and Rogerson \(2017\)](#) amongst others. This paper aims to disentangle this fundamental heterogeneity from input distortions.

2.1.5 Markup variation, misallocation and theory of the second best

I analyse misallocation from input distortions. That is, markups are allowed to vary in both the factual observed world and the counterfactual scenarios, and are determined by the estimated demand fundamentals and all prices and quantities. There is an alternative literature, that views variable markups as distortions compared to a CES world. This literature, predominately in trade, often analyses variation and changes in markups and market share reallocation in response to trade shocks.³³ One might argue that lumping the bundle of market power and input distortion together could be interesting along the lines of the theory of the second best: demand elasticities and input distortions could (partially) offset each other, such that we only care about the joint bundle. It might be tempting to conclude that the literature applying a [Hsieh and Klenow \(2009\)](#) approach does exactly that, as a constant markup is assumed and the inferred distortions implicitly capture the bundle of input distortions and idiosyncratic markups.³⁴ However, their CES demand framework is inconsistent with variable markups, so we cannot calculate welfare (or TFP) losses if we believe that the inferred distortions also capture variable markups, as it would necessarily change other parts of the model.

On the contrary, in this paper, we learn whether the theory of the second best applies by isolating the input distortions. In the counterfactual that removes all input distortions but allows variation in demand elasticities, welfare gains include any effects from previously offsetting distortions. If the theory of the second best applied, welfare would go down when input distortions are eliminated as they would have previously offset variation in demand elasticities.³⁵ I find welfare gains from removing input distortions, thus the joint presence of

³³For so-called “pro-competitive” effects of trade, see e.g. [Edmond et al. \(2015\)](#); [Arkolakis et al. \(2018\)](#). Typically the question is whether trade shrinks the variation in markups across firms due to increased output market competition. In models with CES demand, markups are constant. Models with variable demand elasticities introduce markups that vary across firms. When taking the stance that the variation in markups (i.e. demand elasticities) is not socially optimal, then we can also think of markup variation as additional misallocation of market shares. As in [Dhingra and Morrow \(Forthcoming\)](#), private markups are then not equal to socially optimal constant markups. See also [Behrens et al. \(2018\)](#) who quantify the welfare gap between equilibrium and the optimum allocation under monopolistic competition with heterogeneous sectors and firms.

³⁴To see that, take the well-known contributions of both [Hsieh and Klenow \(2009\)](#) and [De Loecker and Warzynski \(2012\)](#) that infer input wedges and markups respectively from the same first order condition equation. [Hsieh and Klenow \(2009\)](#) assume η_{jt} to be a constant scalar, while [De Loecker and Warzynski \(2012\)](#) implicitly assume $\tau_{jt}^M = \tau_{jt}^L$ or simply an absent τ .

³⁵Note that there is only one output, so one demand elasticity per firm, while there are multiple inputs, so

input distortions and variable markups does not constitute a second best outcome.

Before I describe how I perform the counterfactual estimation and the implications for welfare, I explain how I identify and estimate the output elasticities in the next section, and the demand elasticities thereafter. The demand framework also pins down the welfare framework used for counterfactual analysis.

2.2 Estimating production elasticities and output shocks

We can rewrite the production function in the cost minimisation problem in logarithmic form where lower case variables indicate logarithms.

$$q_{jt} = f(k_{jt}, l_{jt}, m_{jt}) + \omega_{jt}$$

2.2.1 Unexpected output shock and functional form assumption

I incorporate an additional error term ϵ into the entire structural model, so that the estimation is consistent with firm behaviour and the Bertrand competition framework throughout. I provide the details in Appendix C.1. During or after production, once input choices have been made, an unanticipated multiplicative shock to expected firm output occurs ($\exp(\epsilon_{jt})$) and defines realised, observed output Q_{jt}^r based on anticipated equilibrium output Q_{jt} :

$$Q_{jt}^r = Q_{jt} \exp(\epsilon_{jt}) \quad (4)$$

For the baseline estimation and counterfactual analysis, I follow the standard in the literature and assume a Cobb-Douglas production function:

$$q_{jt}^r = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \beta_m m_{jt} + \omega_{jt} + \epsilon_{jt} \quad (5)$$

The advantage of Cobb-Douglas production functions is that we can derive a simple closed form analytical solution for the conditional input demand functions which dramatically eases the search for equilibria. In Appendix C.3, I use a more flexible translog production function instead. The average production elasticities for this specification are reassuringly close to the Cobb-Douglas estimates.

multiple input distortions. If input distortions differ by input, then a single demand elasticity cannot offset all input distortions.

2.2.2 Control function approach for identification

There are two well-known challenges with estimating production functions. They stem from unobserved productivity ω_{jt} and generate a simultaneity bias and a selection bias, as explained in more detail in Appendix C.1. In order to estimate the production function consistently and address these concerns, I make a set of assumptions that was first introduced by Olley and Pakes (1996), and later refined by Levinsohn and Petrin (2003), Wooldridge (2009) and Akerberg et al. (2015), and commonly referred to as the proxy method or control function approach. The strategy is to use a control function for unobserved productivity to recover it, instead of for example, simply instrumenting for input choices.

A detailed description of my adaptation of this approach and the required assumptions are carefully explained in Appendix C.1. The population moment equations used for identification, where Θ is the vector of all structural parameters, are:

$$E \left(\begin{array}{c|c} \epsilon_{jt}(\Theta) & \mathbf{\Gamma}_{jt} \\ (\epsilon_{jt} + \zeta_{jt})(\Theta) & \mathbf{\Gamma}_{jt-1} \end{array} \right) = 0$$

where ϵ_{jt} is the unforeseen production shock, which is uncorrelated with the current period choices and information set $\mathbf{\Gamma}_{jt}$. ζ_{jt} is the innovation in the Markov productivity process in ω_{jt} and uncorrelated to past input choices $\mathbf{\Gamma}_{jt-1}$. I use a joint estimation approach similar to Wooldridge (2009) to exploit these moment conditions.

This approach yields estimates for the structural output elasticities as well as the plant level productivity ω_{jt} and production shocks ϵ_{jt} .

2.2.3 An alternative for robustness checks: system GMM

To check the robustness of the estimated output elasticities from the control function approach to the invertibility condition, I also implement a dynamic panel system GMM approach following Blundell and Bond (1998, 2000). Details are reported in Appendix C.2.

2.3 Demand structure and estimation

2.3.1 Dual role of the demand model

The demand model satisfies two roles. First, it allows us to estimate the elasticities of demand needed for the identification of τ_{jt} . I want to allow for flexible heterogeneous demand elasticities across (i) producers and (ii) counterfactuals by endogenising them. This avoids attributing

uncaptured demand heterogeneity to the input distortions τ_{jt} .³⁶ As described in Section 2.1, even though producers compete with the same product, such demand heterogeneity can arise because producers cover different geographical regions. Alternatively, product quality or brand loyalty differences introduce different price sensitivities among consumer groups or downstream firms. The second role of the demand model is to provide a structure for quantitative welfare analysis.

2.3.2 *Heterogeneous consumers: mixed logit random utility model*

The buyers of output in the application of this paper – cast iron – are likely to be downstream firms, not consumers directly. However, to focus on the analysis of the cast iron sector, I abstract from modelling downstream sectors. Downstream firms are assumed to transform the outputs by segment into final products in a way that preserves product characteristics such that e.g. a high quality final product requires a high quality output. The downstream firms are assumed to operate with constant returns to scale and complete pass-through such that utility to consumers can be modelled as if they are buying the product directly. Schmalensee (1976), for example, shows that in competitive markets with constant returns to scale, consumer surplus can be estimated from the market of intermediate goods (i.e. the output of the cast-iron firms) instead of the final goods. What matters is that the demand elasticities are well estimated using the variation in output prices and quantities of all firms. I will therefore model consumers as if they are buying directly from the cast iron firms.³⁷

Heterogeneous consumers face a discrete choice problem from which firm j to buy to maximise their utility. Consumer heterogeneity in terms of price sensitivities and preferences over characteristics can be gauged by a random coefficient utility model. The seminal contribution of Berry et al. (1995), henceforth BLP, develops a random utility mixed logit approach which (i) has more realistic properties regarding demand (cross-) elasticities than either a basic logit model (independence of irrelevant alternatives) or a Kimball (1995) model (where elasticities depend only on output shares), and (ii) addresses price endogeneity. The framework is also well suited for welfare analysis. The downside is that it is not trivial to estimate this system and that algorithms for counterfactual analysis can be time-intensive to converge. Crucially, the price elasticity of demand and the markup depend on the structural

³⁶In the literature following Restuccia and Rogerson (2008) and Hsieh and Klenow (2009), studies have typically employed a simple CES model of demand with an assumed instead of estimated demand elasticity. This allows for some welfare analysis, but confounds heterogeneity in demand with input distortions. The literature estimating heterogeneous markups following De Loecker and Warzynski (2012) does not assume any demand model, which prevents welfare analysis.

³⁷As mentioned in Section 2.1.1, transport costs on the output side are not contained in the prices and implicitly captured by the estimated demand (cross-) elasticities and unobserved product characteristics.

parameters and distortions and prices and quantities of all firms. They are thus endogenous and vary across factual and counterfactual scenarios.

2.3.3 Specifying the utility function and demand elasticities

Consumers are indexed by i and choose between products j to maximise their utility:

$$U_{ijt} = (y_{it} - P_{jt}^r)\theta_{it}^p + x_{jt}\theta_{it}^x + \xi_j + \xi_t + \Delta\xi_{jt} + \mu_{ijt} \equiv V_{ijt} + \mu_{ijt} \quad (6)$$

where y_{it} is consumer income, P_{jt}^r are realised prices (which are associated with realised quantities – these are the ones that are relevant for the consumers), x_{jt} a vector of product characteristics and a constant, ξ_j average utility from unobserved time-constant product characteristics, ξ_t average unobserved market-specific utility, and $\Delta\xi_{jt}$ the unobserved deviation from a particular product in a particular market from the unobserved averages. The unobserved ξ_j can contain the quality and the location of a product and will be absorbed by fixed effects dummies.³⁸ For the baseline results I only include a constant in x_{jt} as there are few time variant product characteristics (since the time invariant characteristics are absorbed in ξ_j). The parameters θ_{it}^p and θ_{it}^x are the random coefficients that determine the heterogeneity in preferences across consumers and are allowed to vary both by consumer and by market³⁹. The set up in Equation (6) allows for heterogeneous marginal utility of income (and prices) across consumers.⁴⁰ The non-random utility can be summarised by V_{ijt} . The random utility component is μ_{ijt} , which follows an i.i.d. Type I extreme value distribution.⁴¹

Appendix D.1 describes how the parameters in the utility function are identified and estimated. The algorithm involves an inner loop the minimises the distance between the observed market shares and the theoretically derived market shares from the utility maximisation. The outer loop addresses prices endogeneity and forms the moment conditions. The price elasticity of demand is:

$$\frac{1}{\eta_{jt}} \equiv \frac{\partial Q_{jt}}{\partial P_{jt}} \frac{P_{jt}}{Q_{jt}} = \frac{\partial(s_{jt}Y_t)}{\partial P_{jt}} \frac{P_{jt}}{s_{jt}Y_t} = \frac{\partial s_{jt}}{\partial P_{jt}} \frac{P_{jt}}{s_{jt}} = \frac{P_{jt}}{s_{jt}} \frac{1}{N} \sum_i (\theta_{it}^p s_{ijt}(1 - s_{ijt})) \quad (7)$$

where s_{jt} is the market share of product j , Y_t the market size and s_{ijt} consumer i 's

³⁸See [Nevo \(2001\)](#) for a discussion of the benefits of such brand dummies. The dimensionality increases with J , and not with J^2 as in an AIDS model ([Deaton and Muellbauer, 1980](#)).

³⁹We can interpret this as different consumers in different markets (periods). This precludes dynamic demand considerations.

⁴⁰The consumer specific marginal utility of income is, however, constant with the level of income, which facilitates welfare calculation, and follows from risk-neutrality.

⁴¹This is a standard assumption in the literature because it facilitates inversion of market shares and exact welfare analysis. Note that the distributional assumption is not required for identification, but the instruments are key to identification ([Berry and Haile, 2014](#)).

expected expenditure share in product j (see Appendix D.1). I omit the notation with r for realised output (or market share) here, since the elasticities can be derived from any prices and quantities (so in the realised as well as in the counterfactual equilibria) conditional on the estimated parameters. Cross-elasticities can be calculated similarly and vary by firm-pair in each market.

2.4 Factual and counterfactual equilibria and welfare

With the estimated structural parameters, I can recover the matrix of input distortions $\boldsymbol{\tau}$ and solve for counterfactual allocations. I first discuss what the relevant counterfactual is, before I describe how I solve for it.

2.4.1 Misallocation costs: counterfactual distortions as weighted geometric average

What is the relevant counterfactual to evaluate the size of misallocation losses? The relevant “no misallocation” counterfactual for this paper is the state of the economy when the distortions are removed. This would be the allocation that would occur in the same oligopolistic setting, but without input distortions. Note that this is not necessarily the optimum that a social planner might choose, which is the counterfactual in Behrens et al. (2018), for example. The counterfactual in this paper can be interpreted as what we could achieve, if we managed to address input distortions in a market economy.

In the counterfactual, the distribution of $\boldsymbol{\tau}$ is degenerate, such that it is constant across plants. In principle, any constant $\tilde{\boldsymbol{\tau}}$ would equalise marginal revenue products of inputs across plants, adjusted for input prices (recall $\tau_{jt}^M \equiv \frac{MRPM_{jt}}{P_{jt}^M}$). A natural candidate is setting $\tilde{\boldsymbol{\tau}}$ to unity. However, with measurement error in the deflator for output and input prices, unity is no longer the appropriate counterfactual as we would artificially inflate or deflate input costs per unit across the board. Note that this would not affect production parameter estimation (quantities) or demand estimation (year fixed effects) as the measurement error is common to all firms. The measurement error multiplies prices and in- or deflates all $\boldsymbol{\tau}$ by the same proportion within each cross section.⁴² Consider a simple example. Suppose that the error in the input price deflator is a change in the unit of the input price (P_{jt}^M) from dollars into cents for one period. Since the measured $MRPM_{jt} \equiv (\eta_{jt} + 1)\alpha_{jt}^M P_{jt} Q_{jt}/M_{jt}$ remains unchanged, all measured τ_{jt}^M are scaled down by a factor of 100. As a result, the counterfactual $\tilde{\tau}_{jt}^M$ needs to be scaled down by a 100 as well.⁴³

⁴²The Hsieh and Klenow (2009) method implicitly addresses this by defining an aggregate production function with the same structure as for the plant level and then taking ratios of plant level quantities to aggregate quantities.

⁴³As another example, suppose that all true τ_{jt}^M are already unity, i.e. there is no misallocation. Again,

I take the stance that *allocative* inefficiencies between plants should be attributed to *differences* in distortions across plants alone while preserving an average of the distortions (which could be the measurement error). We can get the correct counterfactual even for τ that are polluted with measurement error, if we assume that across the economy the true τ^{true} are on average neither favourable nor adverse per input used (i.e. unity). The correct counterfactual is unity multiplied by the measurement error. This is equivalent to setting the counterfactual $\tilde{\tau}$ to each period's weighted geometric average τ , where the weights are plant expenditure on that input (i.e. materials or labour).⁴⁴ The weighted *geometric* average is taken because of the nonlinear scale of τ (the geometric average is just the exponentiated arithmetic average of $\log(\tau)$). Weights are used to account for different plant sizes such that we retrieve the average distortion per input used, not the simple average across plants.

All welfare results are robust to and qualitatively the same when using a counterfactual of unity, but inflated due to $\tilde{\tau}$ being above unity in most cases.⁴⁵

2.4.2 Equilibria with endogenous marginal costs, markups and aggregate inputs

The counterfactuals $\tilde{\tau}$ change the cost structure of firms, which in turn implies different best response prices and quantities in the counterfactual Bertrand Nash equilibrium conditions, along with changes in the endogenous markups. Both the factual and counterfactual equilibria are defined as the following set of equations and inequality constraints:

Definition of equilibrium: *An (internal) equilibrium satisfies profit maximisation of all plants. This consists of intersecting their best response functions, which yields the set of first order conditions (FOC), and a set of inequality constraints (SOC) for sufficiency of profit*

wrong price deflators would scale the measured distortions and we would wrongly infer misallocation losses since the measured distortions are not unity but some other constant.

⁴⁴If measurement error ϵ_t^{DEF} multiplicatively enters the input price deflator, then we work with $\tau_{jt}^{M,true} \epsilon_t^{DEF} P_{jt}^M$ in the firm's costs, where $\tau_{jt}^M = \tau_{jt}^{M,true} \epsilon_t^{DEF}$. This in turn means all true $\tau_{jt}^{M,true}$ are multiplicatively shifted by the same ϵ_t^{DEF} in each period. The relevant non-misallocation counterfactual is not $\tilde{\tau}_{jt}^M = 1$ but $\tilde{\tau}_{jt}^{M,true} = 1$, so $\tilde{\tau}_{jt}^M = \epsilon_t^{DEF}$. With the weighted geometric average as counterfactual we achieve this under measurement error, as long as it holds that the weighted geometric average of the true $\tau_{jt}^{M,true}$ are unity (i.e. $\exp \sum_j \ln(\tau_{jt}^{M,true}) * weight_{jt} = 1$). The counterfactual $\tilde{\tau}$ is each period's weighted geometric average τ : $\tilde{\tau}_t = \exp \sum_j \ln(\tau_{jt}^M) * weight_{jt} = \exp \sum_j \ln(\tau_{jt}^{M,true} * \epsilon_t^{DEF}) * weight_{jt} = \exp \sum_j \ln(\tau_{jt}^{M,true}) * weight_{jt} * \epsilon_t^{DEF} = 1 * \epsilon_t^{DEF}$. When presenting statistics on distortions in the rest of the paper I therefore use annually demeaned distortions.

⁴⁵Results available upon request.

maximisation, conditional on all structural parameters and the distortions $\boldsymbol{\tau}$:

$$\frac{P_{jt}}{MC_{jt}(Q_{jt}(\mathbf{P}_t), \mathbf{c}_{jt}(\boldsymbol{\tau}))} - \frac{1}{1 + \eta_{jt}(\mathbf{P}_t)} = 0 \quad (\text{FOC})$$

$$2\frac{\partial Q_{jt}}{\partial P_{jt}} + (P_{jt} - MC_{jt})\frac{\partial^2 Q_{jt}}{(\partial P_{jt})^2} - \frac{\partial MC_{jt}}{\partial Q_{jt}}\left(\frac{\partial Q_{jt}}{\partial P_{jt}}\right)^2 \leq 0 \quad (\text{SOC})$$

I provide detailed derivations for the terms in Appendix E.⁴⁶

We are thus comparing well defined equilibria when analysing misallocation losses in an attempt to gauge the full costs of misallocation. This includes potential expansion or contraction in aggregate input use. This approach provides the advantage of explicitly endogenising the key variables (prices, quantities, input use, demand elasticities, pass-through) while preserving the estimated structural parameters (production elasticities, preferences, plant TFPQ (Ω_{jt}), etc.).⁴⁷ So far, counterfactual analyses in the input misallocation literature have assumed exogenous markups that do not change in counterfactual equilibria.

Next I briefly describe how I obtain the factual equilibrium from realised prices and quantities, and how I obtain the counterfactual.

2.4.3 From realised prices to (factual) equilibrium prices

First we need to recognise that we do not observe the factual equilibrium directly. Due to the unanticipated shock to production ϵ_{jt} , we observe realised prices and quantities in the data which are different to the equilibrium quantities and prices that firms expected and chose. Yet, the stage where prices and inputs are chosen (i.e. before the shock) is the relevant stage for inferring the input distortions $\boldsymbol{\tau}$ as described in Section 2.1. Since firms are risk neutral and the shock entirely unanticipated (and mean zero), it does not influence their production input decisions, as described in Section 2.2. I assume that firms choose the next best (realised) prices that clear their shock adjusted produced output and therefore the market.

The equilibrium quantities can be easily calculated from realised quantities, $Q_{jt} = \frac{Q_{jt}^r}{\exp(\epsilon_{jt})}$ from (4), and equilibrium market shares are $\hat{s}_{jt} = \frac{\hat{s}_{jt}^r}{\exp(\epsilon_{jt})}$ from (33). Given the equilibrium quantities, I search for the equilibrium prices that solve the necessary and sufficient conditions of the Bertrand Nash framework. This timing assumption harmonises the production and

⁴⁶Existence is proved by finding an equilibrium. Uniqueness is not proved. Even if there were multiple equilibria, we do not know which one would be reached. I could not find any numerical evidence on multiple equilibria, as a set of genetic algorithms as well as multiple starting points converged to the same equilibrium.

⁴⁷Often, counterfactual analyses using the BLP approach do not estimate the production side, and marginal costs are simply assumed to be constant with respect to output (Berry et al., 1999; Nevo, 2000a; Petrin, 2002). Since I explicitly incorporate and estimate a structural model of production, I can relax this assumption and allow marginal costs to vary with output quantities according to estimated production functions.

demand estimation with equilibrium behaviour to derive the distortions τ from Equation (3). See Appendix E for more details on finding the factual equilibrium.

2.4.4 Counterfactual equilibria

Once we have obtained the factual equilibrium and τ , we can set *any* counterfactual $\tilde{\tau}$, search for the new equilibria and perform comparative statics between the factual equilibrium and a version of the counterfactual equilibria. The three main counterfactuals that I construct either eliminate the variation in distortions in material inputs to $\tilde{\tau}_{jt}^M$, in labour to $\tilde{\tau}_{jt}^L$, or both simultaneously. The counterfactual equilibrium is pinned down by a vector of (output) prices alone, given the structural parameters. All of the comparative statics are alongside the intensive margin. Some plants can operate near or at zero output in the counterfactuals, resembling firm exit, but I do not explicitly model the extensive margin of exit and entry of new firms. For the counterfactual analysis, I use a Cobb-Douglas production function, since cost and marginal cost functions as well as conditional factor demands can be derived analytically, which makes solving for equilibrium prices more tractable. Again, Appendix E provides the details.

We need to assume how much capital firms choose in the counterfactual. The simplest solution is to assume that optimal installed capital follows a static optimisation condition (i.e. the same as for labour and materials). This leaves us with the unknown distribution of the rental rate for capital. While I could assume a range of values or distribution for this rental rate, I back it out from a static optimisation condition in the factual equilibrium. The median value of this inferred rental rate is 29%. As Asker et al. (2014) show, attributing all rental rate differences to misallocation could be misleading, as capital is mainly a dynamically optimised input. The rental rate contains a mix of capital distortions and capital adjustment costs. I preserve the plant specific rental rate across factual and counterfactual equilibria, i.e. I preserve the degree of capital misallocation.⁴⁸

2.4.5 How noisy are misallocation losses? A parametric bootstrap.

A feature of this paper’s approach that is novel to the input misallocation literature is that I am able to derive confidence bands for any of the comparative statics.⁴⁹ I draw a set of

⁴⁸Alternatively, I could specify a more complicated dynamic optimisation problem modelling adjustment costs such as David and Venkateswaran (2017) and use the residual variation of this as capital distortions. My approach is more conservative in terms of total misallocation losses by maintaining any capital distortions (and adjustment costs) across counterfactual equilibria.

⁴⁹At the time of writing I am not aware of a paper that provides estimates of uncertainty around estimated input misallocation losses without calibration. The structural approach based on microdata to generate estimates of uncertainty around gains may also be useful in related counterfactual analysis studies e.g. on spatial misallocation of housing (Hsieh and Moretti, 2019) or infrastructure (Fajgelbaum and Schaal, 2017), or

parameter estimates (Θ, Σ) from their joint asymptotic normal distribution using the estimated covariance matrices and for each draw, find the factual equilibrium (since ϵ is different for each draw), calculate τ , find the counterfactuals and perform the comparative statics analysis.⁵⁰ This channels the information about the uncertainty in the structural parameters, such as plant productivities from the underlying Markov process, output elasticities, preference parameters and markups into the final comparative static of interest.

2.4.6 Calculating profit gains and average expected compensating variation

Instead of relying on the usual aggregate production functions, I calculate aggregate firm profits, compensating variation and aggregate input productivities for comparative statics. Profits as well as material and labour productivities are straightforward to aggregate, as I solve for the factual and counterfactual equilibrium prices, quantities and inputs for each plant. I use exact welfare measures for the demand side by calculating expected consumer compensating variation CV_{it} from moving from the factual equilibrium prices \mathbf{P}_t to the counterfactual equilibrium prices $\tilde{\mathbf{P}}_t$. It is an expected welfare measure because of the random utility component μ_{ijt} . For each consumer, the CV_{it} solves:

$$\max_j U_{ijt}(y_{it} - P_{jt}, x_{jt}, \mu_{ijt}; \theta_{it}^p, \theta_{it}^x, \xi,) = \max_j U_{ijt}(y_{it} - \tilde{P}_{jt} - CV_{it}, x_{jt}, \mu_{jit}; \theta_{it}^p, \theta_{it}^x, \xi) \quad (8)$$

Due to (i) additive (ii) GEV random utility disturbances, and (iii) constant marginal utility of income, I can conveniently use the [Small and Rosen \(1981\)](#) close form expression for CV_{it} :

$$CV_{it} = \frac{\ln(\sum_j \exp(\tilde{V}_{ijt})) - \ln(\sum_j \exp(V_{ijt}))}{-\theta_{it}^p}$$

where \tilde{V}_{ijt} and V_{ijt} are the counterfactual and factual utility components defined in (6). We can take the average over consumers to get average expected compensating variation for each period (market) per unit. That is $CV_t = \frac{1}{N} \sum_i CV_{it}$. Multiplying this figure by the total quantity of output yields total expected compensating variation.

misallocation losses from within-country trade distortions ([Costinot and Donaldson, 2016](#)). [Adao et al. \(2017\)](#) for example develop a structural “mixed-CES” approach for trade models based on perfectly competitive goods and factor markets that allows for bootstrapped confidence intervals around welfare gains.

⁵⁰I draw from the production side parameters Θ and demand side parameters Σ , and assuming independence between them. On the production side, ϵ_{jt} is a function of the drawn parameters Θ and data. On the demand side, I can solve for the linear parameters $(\theta^p, \theta^x, \delta_{jt}, \xi)$ by using the draws from Σ , the contraction mapping and the linear IV regression. I repeat the draws and analysis 330 times for all outcomes and then take the desired quantiles of the outcomes in order to get consistent confidence intervals.

3 Data and descriptives

I first describe the main plant and product level data. I provide some relevant background on the cast iron industry in India and descriptive statistics. Thereafter I describe the geographic data used in Section 6.

3.1 Plant data and descriptives

3.1.1 Plant and product level data

I use annual plant level panel data from the Indian Annual Survey of Industries (ASI) from 2000 to 2012. Since the Collection of Statistics Act in 1953, detailed plant level data is collected by the Ministry of Statistics and Program Implementation (MoSPI) of India, and most medium and large firms are familiar with the reporting. The mandatory nature as well as the long history of the survey makes it an arguably established and reliably data source in the developing context.⁵¹ For the purpose of this study, the most important features of the dataset are that the output and input information is provided by product codes, both in revenue (or expenditures) and in physical quantities, which allows me to disentangle quantities from price effects. The ASI was traditionally a repeated cross-section, which researchers matched throughout the years (Bollard et al., 2013; Harrison et al., 2013). Recently it has been released in panel form, which latest research has started to use (Martin et al., 2017; Rotemberg, 2014; Allcott et al., 2016; Akcigit et al., 2016).

A general shortcoming of the ASI data is that it covers only the formal manufacturing sector defined in its 1948 Factories Act, while a large share of manufacturing employment is in the informal sector (around 80% (Hsieh and Klenow, 2014)). However, since larger firms tend to be formal, the formal sector accounts for around two-thirds of output in manufacturing (Allcott et al., 2016). Informality is less of a concern for the cast iron industry. It is highly likely that output is even more skewed towards formal firms in this sector compared to e.g. the textiles industry in India.

I study single product cast iron plants. The aim is to compare plants that are as homogeneous as possible in their production technology to disentangle distortions from production heterogeneity within sectors. Cast iron is a 7-digit product, and to give a sense of the level of detail, the number of 2-digit sectors in the ASI manufacturing section (NIC08) are 24, the number of 4-digit sectors is 137, whereas the number of 7-digit product categories (NPCMS11) that are also manufactured in India is 5476. On average there are 40 different

⁵¹The data is an annual census of plants with ≥ 100 employees (until 2004 ≥ 200) and a sample (around 20%) of all plants with ≥ 10 employees with electricity and all plants with ≥ 20 employees without electricity from a 4 digit sector-state strata.

product categories within a 4-digit sector.

Finally, I use single product plants, because the ASI reports plant level outputs by product, but not inputs by product line. While there are ways to deal with multiproduct plants⁵², they are likely to introduce further misattribution of unknown input allocations to the estimated distortions.⁵³

3.1.2 *Cast iron production in India*

Cast iron is an iron-carbon alloy with high carbon content produced in different grades (e.g. hardness) by varying carbon, silicon and other components and processes.⁵⁴ It is used in many machines, automobile parts (such as gearboxes and cylinders), pipes and historically in construction. Cast iron is made from melting pig iron (which in turn is produced from smelting iron ore with coke and limestone in a blast furnace), coke, limestone and scrap steel, and small quantities of other metals into a desired grade and primary casting. It can be placed in the production chain between the rawer pig iron upstream and semi-finished and finished sheets, cables, pipes, blades or tins (which might be turned into tools, doorframes etc) downstream. Depending on the final use, the downstream production chain can be shorter or longer.

How significant is cast iron production in the Indian iron and steel sector? As Figure 16 in Appendix F shows, a declining but considerable share of plants that produce some product in the broader classification of iron alloys of primary form (ASICC 711 or NPCMS 411) produce also cast iron (from 35% of plants in 1999 down to 20% of plants in 2009). Sales of cast iron account for a slightly more stable 25% of primary iron alloys until the financial crisis of 2008. Figure 17 in Appendix F shows that around 60% of firms producing cast iron are single product firms. However, the physical output quantity produced by single product firms is typically slightly lower, as multiproduct firms tend to be larger. There is limited industry concentration amongst the cast iron plants in the sample as Figure 18 in Appendix F shows. Appendix F provides a more detailed account of India's iron and steel sector. It also gives more detailed environmental context to this sector. The substantial carbon emissions in this sector are in part determined by aggregate material productivities. One contribution

⁵²See either the simpler method in De Loecker (2011) or the more advanced method De Loecker et al. (2016), for example.

⁵³For the demand side, I can included products that are produced by multi-product firms to increase the coverage of the sample and improve precision of the demand elasticities.

⁵⁴The product codes of cast iron in the ASI data are 4111102 (NPCMS11) and 71112 (ASICC). Cast iron has at least 2% carbon content (till around 3.5%-4%), while steel has less than 2% carbon content. Sometimes cast iron is loosely included in the term steelmaking. Steel on the other hand is used in construction and infrastructure, heavy machinery, white goods and tools. The advantage of cast iron over steel is a lower melting point (and costs, as well as better machinability, i.e. cutability), but tends to be brittle and have less tensile and compressive strength than steel.

of this paper is to study whether misallocation has an effect on these aggregate material productivities.

3.1.3 Descriptive statistics of key variables

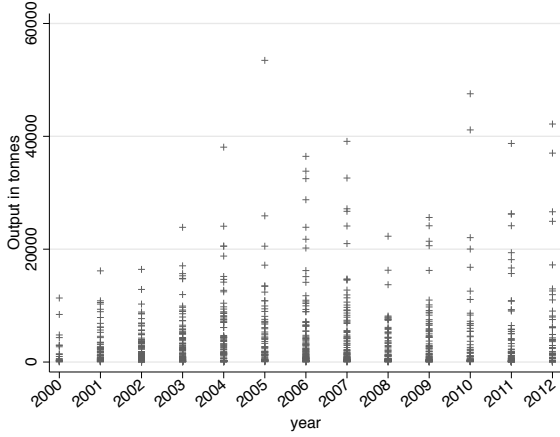
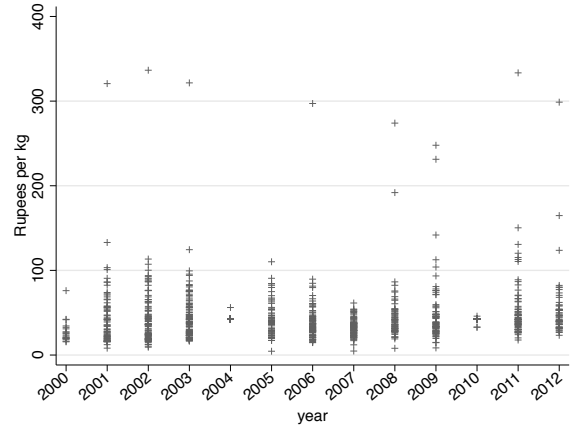
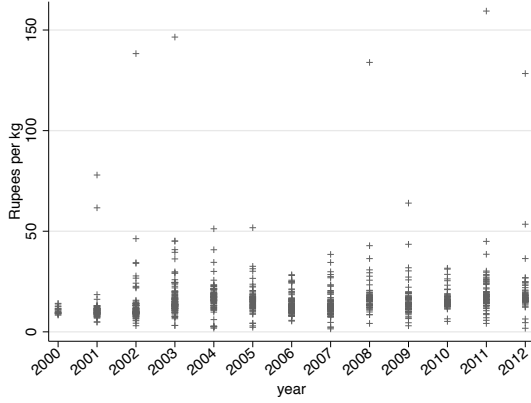
All output and input prices as well as book values of capital (at the start of the accounting period) are deflated with industry specific and capital deflators respectively from the [Office of the Economic Adviser \(2019\)](#). I use output prices net of plant level subsidies, taxes or distribution costs. Labour includes workers employed through subcontractors as well as informal labourers.⁵⁵ Wages include the salaries as well as bonuses and welfare expenses. For input materials, I use the sum of the weight of input materials, and the materials price is the corresponding average price, including shipping fees. I recover the input price by dividing total expenditure on material inputs by the total weight of material inputs. Table 1 provides some descriptive statistics on the sample of plants, after trimming the plants which are in the bottom and top 1 percentiles of either output prices, the physical output to material or labour ratio, or sales to installed capital ratio. I only keep plants that are present for at least two consecutive years in the data. The total number of plants varies by year, from around 60 to 110.

There is considerable variation in the scale in which plants operate, as shown in the plant output quantities in Figure 1. The ratio of the 75th percentile to the 25th percentile of output quantity is 15.4. The prices for output and inputs are plotted in Figure 2 and Figure 3 respectively. Output prices are mostly within the range of 20 to 70 Rupees per kg with a mean of 44, which roughly matches global average steel prices over this period.⁵⁶ This suggests a setting with differentiated products. In a typical monopolistic competition framework, quantities are negatively correlated with prices in the cross section. I find no statistically significant relationship in the cross section. This suggest that there is quality differentiation. Figure 20 in Appendix G.1 supports this story. It plots the output prices against the input prices and shows that they are positively correlated, consistent with quality differentiation where higher priced outputs require higher priced inputs.

Average materials input prices are 16 rupees per kg and 80% are between 9 and 23 rupees per kg. Table 1 converts labour L_{jt} in man-days into the number of employees (on average) with a mean of 175 across plant years. The wage rate is 203 rupees per day on average,

⁵⁵In Appendix G.10 I measure labour as expenditure on labour instead of man-days. This provides a sensitivity analysis towards measuring skill in labour, if skill is correlated with pay. Results are qualitatively robust to this.

⁵⁶Figure 2 also reveals that there is much less dispersion in output prices in 2004 and 2010 (also before any trimming), which is a feature of the underlying raw data, perhaps simply due to sample variability over the years. As I calculate annual misallocation losses, it is easy to see in Appendix G.8 that if anything, the misallocation losses for these two years are slightly smaller.

Figure 1: Plant output quantities.**Figure 2:** Plant output prices.**Figure 3:** Plant material input prices.**Table 1:** Descriptive statistics

	Mean	SD	p10	p90
Output quantity	3929	6345	140	10099
Output price	43.9	31.6	20.9	70.1
Materials quantity	5024	8932	154	12657
Materials price	15.7	11.4	8.8	23.0
Man-days (th.)	54.3	77.0	2.4	151
Employees	175	248	9.0	469
Daily wage (₹)	203	120	92.3	341
Capital (mil. ₹)	54.7	161	0.5	116
Observations	1001			

Notes: Quantities are in tonnes, prices in rupees per kg. All prices are deflated by 3-digit industry deflators. Wholesale price deflators are from the Office of the Economic Adviser from the Government of India.

which corresponds to around 4.5 USD per day. The 10th percentile at around half this figure. Despite an increase in wages from 3.3 to 5.2 USD per day over the sample period, they are still low and reflect the persistent poverty despite industrial growth described in [Bhagwati and Panagariya \(2014\)](#).

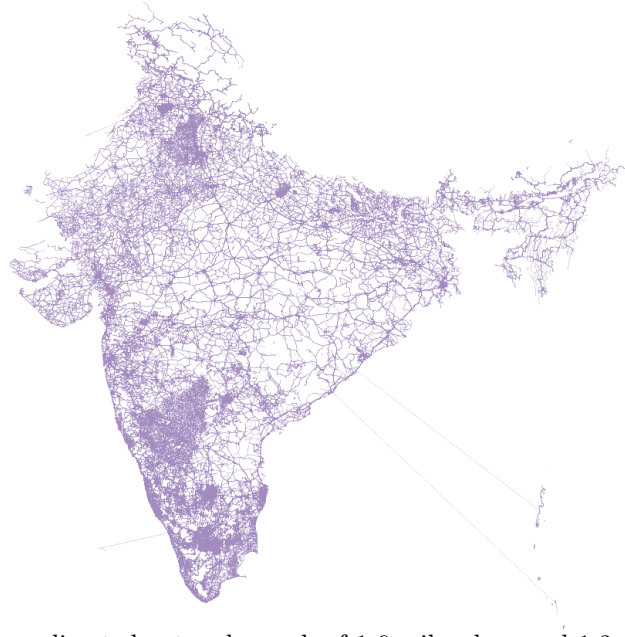
3.2 Geographic data

I use geo-located data for the analysis of supplier access and misallocation in Section 6. The geo-located data of administrative boundaries is from [Database of Global Administrative Areas \(GADM\) \(2016\)](#). The data of transport infrastructure is from [OSMF \(2016\)](#).

3.2.1 Matching plant data with the location of firms

There are no exact geographical identifiers of plants in the ASI due to confidentiality. By matching the panel and the cross-sectional versions of the ASI, I obtain the districts of plants which are only contained in the latter version. The district centroids act as plant locations

Figure 4: Indian rail and road transport network



Notes: The map shows the undirected network graph of 1.6 mil. edges and 1.2 mil. nodes, based on Indian rail and road infrastructure.

for the rest of the paper.⁵⁷ I matched the ASI districts to the 594 geo-coded district data via fuzzy string matching within states, with extensive manual matching and checks until all districts were matched.

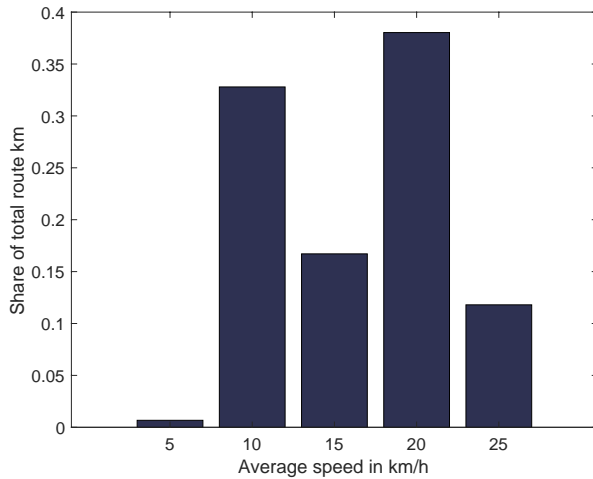
3.2.2 *Information on transport infrastructure*

I use data on railroads and roads. Transportation via inland waterways is negligible, as it is severely underdeveloped in India (NTDPC, 2014).⁵⁸ The share of imported materials in total materials is around 2% in quantity and value terms for this sample. I therefore ignore international sourcing. The transport network contains information about the type of each edge, for example broad vs. narrow gauge rails, or motorways vs. secondary roads. This information is used to assess the speed for each edge of the network.

There is, however, no temporal information on the opening of railroad tracks and roads. Figure 30 in Appendix H.1 shows that the route kilometres of railways during the sample period for 2000-2012 only increased by 3.8%, almost all of it in the last 4 years. Furthermore, the average speed of goods trains was nearly constant (see Figure 30), due to little investment in upgrading of existing infrastructure. I therefore treat the transport infrastructure as constant over time using a snapshot from the end of the sample period. For roads, the picture

⁵⁷I also repeatedly drew random points in each district as plant location and estimate the effect for each draw. The average effect is close to the reported estimates.

⁵⁸The share of transportation via waterways in India is less than 1% of tonne-km (Raghuram, 2004), at least an order of magnitude lower than in Bangladesh, the US, China or Germany (Rangaraj and Raghuram, 2007; NCAER, 2015)

Figure 5: Share of speed classes in network**Table 2:** Average speed by edge types

Rail edge type	Road edge type	Speed in km/h
Broad gauge rail	Motorway, motorway link	25
Narrow gauge rail, light rail	Primary, primary link, trunk, trunk link	20
	Secondary, secondary link	15
Funicular, yard, platform, station, freight station, turntable	Tertiary, tertiary link, tertiary unclassified, road, minor	10
	Connection of plant location to network	5

Notes: The right table shows the speed assumptions for different types of infrastructure. The left figure shows the shares of the speed classes in total route km of the network.

is slightly different, and there has been an increase in total road length ([Ministry of Road Transport and Highways, 2016](#)). Since the steel industry relies predominately on rails (see Section 6.1), I ignore this temporal variation. I provide a robustness check using only railways, with very similar results.

3.2.3 Construction of geographical network

In order to run network analysis algorithms to calculate access to suppliers, we need a weighted undirected network graph with connected plant locations (district centroids). I prepare the infrastructure data by keeping only segments that can be used for shipping, i.e. deleting abandoned rail tracks, rural bridleways etc. I then perform a series of network preparation and cleaning tasks, for example, to make sure that road intersections contain nodes and that relevant nodes are snapped to each other. I connect the plant locations (district centroids) with a straight line to the nearest point in the network.⁵⁹ The full network contains around 1.6 million edges and 1.2 million nodes, which can still be handled well with a standard computer and optimised network algorithms. Figure 4 shows the entire network graph of roads and railways that are used for the analysis.

3.2.4 Speed assumptions and edge weights

The weights for the network edges are determined by how fast goods can be shipped on a particular piece of infrastructure. This depends on length and speed. The length can be calculated, but we need assumptions for speed. I exploit the information on the type of

⁵⁹I am assigning a low speed to travelling on this edge, see Table 2.

infrastructure and assign them into speed classes. For example, travelling on railroads and motorways is faster than on tertiary roads. Table 2 shows the speed assumptions for different edge types, which are based on reported figures from the literature.⁶⁰ Only the relative speed values matter for the way that supplier access will be constructed. Figure 5 shows the prevalence of speed classes in terms of route kilometres.

4 Results for demand, production and distortions

4.1 Results from demand estimation

To address the price endogeneity in the demand estimation, I use the average plant level wage w_{jt} , and the average plant level price of a tonne of material inputs P_{jt}^M as instruments for output price P_{jt} , which tend to perform well in BLP style estimations (Armstrong, 2016). A theory guided justification for the choice of these instruments, along with first stage tests, estimation results for the structural parameters and results from using alternative instruments is provided in Appendix G.2.

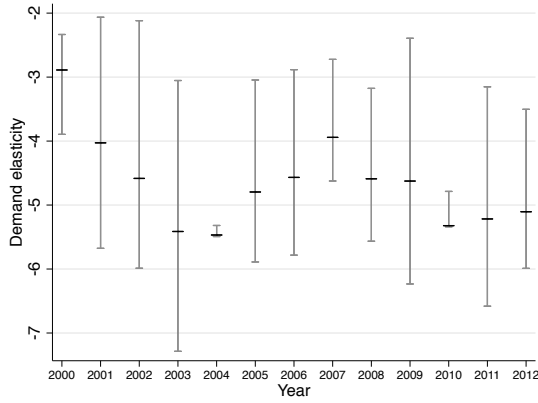
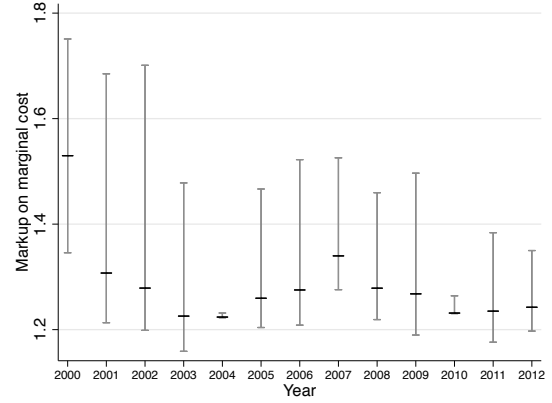
4.1.1 Estimated demand elasticities and decreasing markups over time

A more familiar parameter than the random coefficients in the utility equation are the demand elasticities ($\frac{1}{\eta_{jt}}$), which are determined by the estimated parameters and data as shown in Equation (7).⁶¹ The estimated plant-year level demand elasticities are shown in Figure 6, where the 90% bands are the percentiles of the distribution of the elasticities across plants within a year. We can use the equilibrium optimality condition (1) to express the demand elasticities as price over marginal cost markups. The median as well as the 90% bands of the cross section of markups is plotted in Figure 7.

There is considerable variation in markups across plants within the given years. Across years, different plants are sampled, so there is natural sample variation. The years 2004 and

⁶⁰The assumptions for average speed of goods trains are supported by information from the Ministry of Railways (see Figure 30). Average truck speeds on roads in India are typically a third of the counterparts in more developed countries (NTDPC, 2014). EY (2013) estimates average truck speeds at around 20 km per hour. Baum-Snow et al. (2018) assume 25 km per hour for Chinese roads. Allen and Atkin (2016) use a value of 20 miles per hour for non-highway roads for India. Alder (2017) uses speeds of 35 km per hour based on a survey by the World Bank (2005b) which argues that truck speeds are typically less than 40 km per hour in India. The (short) direct connection from the district centroid (plant location) to the closest point in the network is assigned a low speed of 5 km per hour. I compared the calculated fastest path times between a few district pairs to the duration using Google Maps, and the results are reassuringly close.

⁶¹For the rest of the analysis, I ignore the 1% of observations where I estimate a demand elasticity larger than -1 . With the standard oligopolistic models, we cannot calculate a markup or marginal costs for these observations. I do not drop these observations, but exclude them for calculating distortions τ and when comparing the factual to the counterfactuals. The 9 observations where this is the case have a median market share of 0.0004 and a maximum market share of 0.007, and their market share remains this small in all the counterfactuals.

Figure 6: Demand elasticities**Figure 7: Markups**

Notes: The figures plot the estimated plant level demand elasticities (left) and markups (right). Plotted are the 5th, 50th and 95th percentile across plants within each year.

2010, for which I found little price variation in the raw data (Figure 2), reassuringly also have less variation in demand elasticities and markups. The median elasticity and markup across all years are -4.82 and 1.26 respectively. De Loecker et al. (2016) find that in the Indian Prowess data during 1989–2003 the median markup for manufacturing firms is 1.34 and for basic metals firms 1.20 using the production side method of De Loecker and Warzynski (2012).

The markups decreased slightly over time. A linear regression of logged markups on years shows that markups decreased by 0.6% and 0.4% each year on average, in a pooled and a within-plant fixed effect regression respectively, significant with SE clustered at the plant. This is consistent with a story of increasing competition, particularly from large foreign low price producers from neighbouring China. I find that markups and plant total factor productivity are positively correlated. This correlation is driven by productivity pushing down marginal costs, as productivity and prices are also negatively correlated (see Appendix G.2). Markups and prices are negatively correlated, consistent with more elastic demand at higher price points as in Atkin and Donaldson (2015).⁶²

4.2 Results from production estimation

The results from estimating the Cobb-Douglas production function are reported in Table 3. Column (1) shows the baseline results with standard errors clustered at the plant level.⁶³ The direction of the bias in the OLS coefficients is as expected from the discussion in Appendix C.1. The material elasticity is upward biased from the simultaneity problem, and the capital

⁶²The correlation between markups and prices, and markups and market share varies across periods, and is of opposite sign in some periods. This degree of flexibility is not possible with elasticities from Kimball (1995).

⁶³The Hansen overidentification J-test for valid instruments is not rejected at the 5% level in any of the specifications. There is no standard rank test for instrument strength here as there are cross equation restrictions.

Table 3: Estimates from a Cobb-Douglas production function

	Type of correction		Comparison to literature	
	(1) Simultaneity & Selectivity	(2) None: OLS	(3) De Loecker et al. (2016)	(4) Collard-Wexler and De Loecker (2014)
α^K	.06*** (.02)	.04*** (.01)	.01 (.06)	.08*** (.02)
α^L	.22*** (.05)	.14*** (.02)	.14 (.09)	.27*** (.02)
α^M	.64*** (.05)	.80*** (.03)	.77 (.11)	.68*** (.02)
RTS	.92*** (.03)	.99*** (.01)	.92	1.03***
N	443	1001	949	1498

Notes: The first two columns show the output elasticities and returns to scale with corrections for simultaneity and selectivity and without (OLS). The second two columns show results from related studies for comparison.

coefficient is (slightly) downward biased from the selectivity problem.

I perform several robustness checks regarding these estimates and the underlying invertibility condition, as discussed in Appendix G.3. I also use a translog production function (see Table 12 in Appendix G.4) where elasticities vary by plant and year. The mean elasticities are very similar to the estimates from the Cobb-Douglas production function, with returns to scale close to unity.

4.2.1 Comparison to estimates from related studies

We can compare these results to other production function estimates from the relevant literature. In particular, De Loecker et al. (2016) estimate a translog production function for India for the period 1989-2003, however for the entire 2-digit basic metals sector, capturing other technologies. Collard-Wexler and De Loecker (2014) estimate a Cobb-Douglas production function for steel producers in the US between 1962-2002. The last two columns in Table 3 compares my estimates to their estimates. De Loecker et al. (2016) estimate higher material elasticities and lower capital and labour elasticities, but the estimates of Collard-Wexler and De Loecker (2014) are remarkably close to my Cobb-Douglas and translog estimates. Arguably, the narrow technological focus of Collard-Wexler and De Loecker (2014) on steel producers in the US is more relevant for comparing elasticities than the geographic commonality but higher technological difference in De Loecker et al. (2016). Since their study in the US captures multiple decades, and not only cutting-edge technology, the production technologies are likely to be standard and similar to Indian producers in my data.

4.2.2 Analysis of estimated total factor productivity

Since I use output and input quantities as well as a gross output function, the control function approach estimates *physical* total factor productivity Ω_{jt} (also denoted TFPQ). Total *revenue* factor productivity TFPR is simply defined as $P_{jt} \cdot TFPQ_{jt}$. Due to the large and growing interest in TFPQ and TFPR in the literature, it is worth to briefly analyse these estimates. A more detailed analysis is in Appendix G.5.

The main points are as follows. First, there is evidence that more productive firms grew faster, based on comparisons of weighted and unweighted TFPQ. Second, TFPR grew by more than TFPQ, consistent with increasing prices. Together with decreasing markups, this implies that marginal costs have increased.⁶⁴ Third, the dispersion in TFPQ is smaller than in other studies, most likely due to the much narrower industry definition in this paper. Fourth, the dispersion in TFPR is greater than the dispersion in TFPQ. This is in contrast to Hsieh and Klenow (2009). Prices and TFPQ are negatively correlated, but prices are much more dispersed than TFPQ, leading to a higher dispersion in TFPR than in TFPQ.

4.3 Results for the factual equilibrium and distortions

4.3.1 Descriptives on estimated input distortions τ : misleading SD

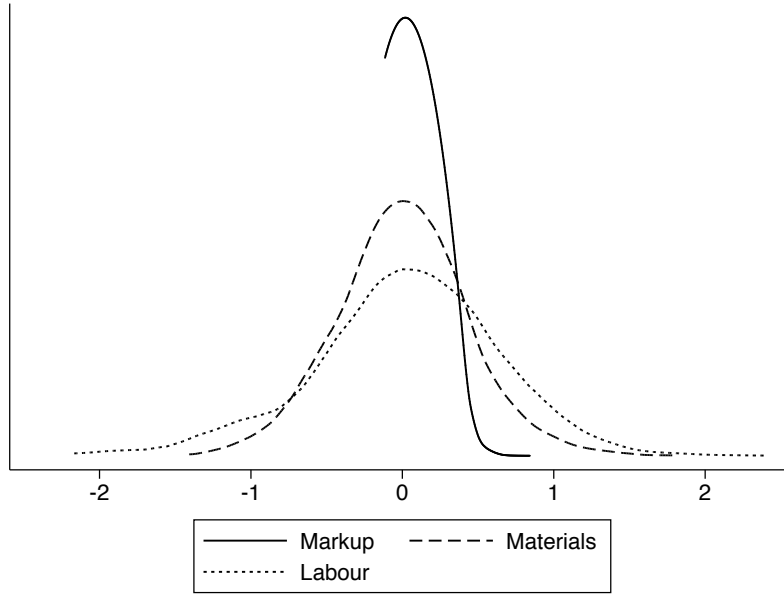
I calculate the τ_{jt}^M and τ_{jt}^L according to Equation (3), using the expected prices and quantities, the input expenditure, and the estimated output and demand elasticities.⁶⁵ For comparability, I also demean the τ , by dividing each by the within-year geometric weighted mean, where the weights are the input expenditures (see Section 2.4.1), and take logs to transform it into a linear scale. The annual empirical density of these variables is plotted for τ_{jt}^M and τ_{jt}^L in Figure 23 and Figure 24 respectively for each year in Appendix G.7, whereas Figure 8 pools the demeaned distortions across years.

There is pronounced dispersion in both labour and material distortions, and some annual densities are multi-peaked. There is no clear trend in the degree of dispersion over time, but some years appear to exhibit less dispersion than others, in part influenced by sampling

⁶⁴Material input prices have been rising at around 3% per year, consistent with the global price increases in raw metals commodity prices (see e.g. IMF). Increasing marginal costs could also be due to changes in τ .

⁶⁵In order to infer τ I first estimate the expected quantities Q_{jt} and prices P_{jt} as described in Section 2.4 by using $Q_{jt} = \frac{Q_{jt}^r}{\exp(\epsilon_{jt})}$ and solving for the prices. The estimated shocks $\exp(\epsilon_{jt})$ over the entire sample have a mean of 1.017 and the 90% range of estimates is [0.55,1.45]: on the extremes of this interval the plants have an estimated shock that decreased and increased output by 45% and 45% respectively. The log of the expected prices are plotted against the log of realised prices (i.e. after production shock ϵ) across all years in Figure 22 in Appendix G.6 and shows that they are similar. The ratio of realised to expected prices ($\frac{P_{jt}^r}{P_{jt}}$) has mean 1.01 and the 90% range of estimates is [0.91,1.15], much tighter than quantity ratio, which is consistent with a convex elastic downward-sloping demand curve.

Figure 8: Dispersion in markups, τ_{jt}^M and τ_{jt}^L across all years



Notes: Plotted are the kernel densities of the logged markup, $\ln(\tau_{jt}^M)$ and $\ln(\tau_{jt}^L)$ divided by the respective weighted means, where the weights are plant materials and labour expenditure. Used kernel is epanechnikov with bandwidth 0.2.

variability over time. A range of $[-0.5, 0.5]$ on the axis corresponds to a τ of $[0.6, 1.65]$. Such values for τ imply that the plant faces a distortion “as if” it had to pay only 60% of the input price or pay a 65% tax on the input respectively.⁶⁶

The standard deviations of the distortions, or alternatively the standard deviation of the marginal revenue product of an input, is often used as a statistic for misallocation (e.g. [Hsieh and Klenow, 2009](#); [Asker et al., 2014](#)).⁶⁷ This paper shows that, while being popular, this statistic can be misleading. The standard deviation of $\log(\tau_{jt}^L)$ is 0.60 compared to 0.40 for $\log(\tau_{jt}^M)$. The equality of variances is rejected in the robust [Levene \(1960\)](#) and [Brown and Forsythe \(1974\)](#) tests.⁶⁸ This might suggest that misallocation of labour is more costly than misallocation of materials in this industry. However, as the next section shows, the opposite is the case in terms of welfare losses. Importantly, this implies that the pure variation in τ is not a sufficient statistic to rank welfare losses, at least not across inputs. The size of the

⁶⁶Recall also that the dispersion in τ is separate and in addition to any dispersion in plant input prices and wages disparities. Input prices are likely to reflect quality. The dispersion in material input prices is slightly smaller than the dispersion in τ , and the dispersion in wages is larger.

⁶⁷Recall $\tau_{jt}^X = MRPX_{jt}/P_{jt}^X$. Often, the variation in TFPR is used as a summary of the $MRPX$ of all inputs.

⁶⁸I can only conduct the test for the unweighted densities, since statistical significance is non-trivial to expand to weighted samples here. Figure 25 in Appendix G.7, which plots the standard deviations corresponding to the plotted densities, shows that for some years, the standard deviation is greater in τ_{jt}^M than in τ_{jt}^L , but insignificantly so. For most years, the standard deviation is statistically significantly larger for τ_{jt}^L , where the statistical significance is obtained for the unweighted samples. Insignificant are the differences in the years 2001, 2002, 2006 and 2009. The hypothesis of equal distributions in the Kolmogorov–Smirnov test is strongly rejected.

welfare losses depend on which plants face the distortions and how it affects other plants through the market structure, which cannot be captured in the variation of τ or the marginal revenue products alone.

4.3.2 Statistical significance of estimated distortions

Since I can also derive confidence intervals for all individual τ_{jt} , I can test whether differences in τ_{jt} across plants are also statistically significant. I use two groups of plants, those with a τ_{jt} smaller than the 30th and those with a τ_{jt} larger than the 70th percentile. No plant is categorised in the opposite group in any of the bootstrapped versions of the τ_{jt} . Furthermore, for around 90% of all plants, the positive or negative logged demeaned distortion is significantly different from zero at the 10% level.

4.3.3 Adjusting distortions for markups and correlation between distortions

Figure 8 also depicts the density of the logged demeaned markup. There is no mass below zero since this corresponds to a markup of less than zero percent. The dispersion of markups is significantly smaller, and the tests for equal variances and distributions are strongly rejected. We can also compare the distribution of $\ln(\tau_{jt}^M)$ and $\ln(\tau_{jt}^L)$ with a version for each with constant markups (i.e. a “naive” version), as shown in Figure 26 and 27 in Appendix G.7. While the “naive” and correct distortions appear similar, there are significant differences in the inferred welfare losses as discussed in Section 5.4.2.⁶⁹

Finally, it is interesting to ask whether the inferred distortions are correlated. Figure 28 in Appendix G.7 plots the distortions against each other and the correlation is nearly zero and insignificant. Ex-ante we might expect that a firm that is constrained in one input is likely to be constrained in other inputs as well. While it is the case for some firms, the opposite is the case for other firms, being constrained in one input and having an advantage in the other input. One reason for not observing a stronger correlation between distortions could be that firms that are severely disadvantaged in both inputs are likely to be not competitive and exit or do not enter the market. The fact that the input distortions are uncorrelated also provides

⁶⁹If they looked very different, then we would likely obtain welfare *losses* from correcting the (wrong) distortions. Pooled across years, the standard deviation drops from 0.40 to 0.37 for $\ln(\tau_{jt}^M)$ in the naive version, but stays roughly constant at 0.60 for τ_{jt}^L . The decrease in the variation in the former is because the naive $\ln(\tau_{jt}^M)$ is negatively correlated with the estimated markup (-0.33***). This translates into a positive correlation with the inverse of the markup, so there is variation added to the naive $\ln(\tau_{jt}^M)$ (See Equation (3)), increasing the standard deviation to 0.40 in the correct $\ln(\tau_{jt}^M)$. Since the correlation between the naive $\ln(\tau_{jt}^L)$ and the markup is low (0.05), there is hardly any change in the standard deviation. Despite similar pooled standard deviations, the consequences of mismeasured distortions in terms of welfare are more substantial as shown in Section 5.4.2. For the same degree of dispersion, individual distortions can still be severely mismeasured. It matters which plants face which distortions for the size of the welfare losses.

evidence against potential concerns that there is a distortion on the output side instead of the input side, as that would predict a perfect correlation between the input distortions.

5 Results from the counterfactual analysis

I begin this section by analysing the welfare consequences of input misallocation for consumers and producers. Then I discuss the effects on aggregate input productivities, before I compare the welfare losses to a version with constant markups. I end this section by showing the effects of misallocation on the size distribution of plants.

5.1 Welfare and profits

The gains from removing misallocation distortions are shown in Figure 9. The first panel shows expected total compensating variation, the second aggregate profits and the third the total welfare gains, all in billions of rupees. For each panel, the gains from removing material or labour misallocation distortions individually are also reported.⁷⁰ The bootstrapped 95% confidence intervals are shown around the point estimates and derived as described in Section 2.4.5. These figures are the total across all sample years. Table 14 in Appendix G.8 shows the gains for each year individually. There is some variation in the gains across years, but without census data we cannot reliably interpret these as changes in misallocation.⁷¹

5.1.1 Higher consumer than producer incidence

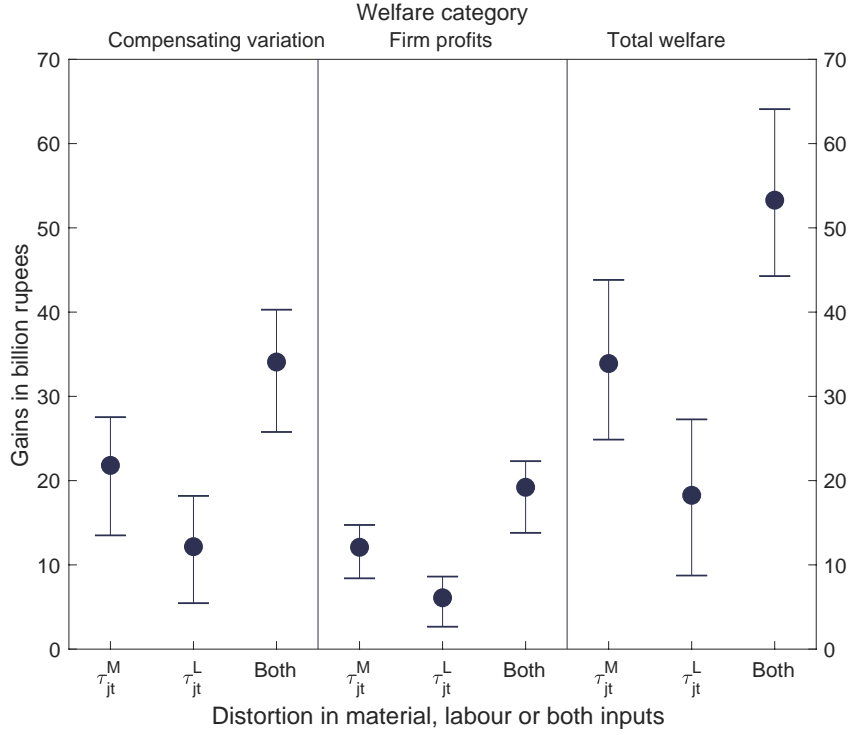
Who gains from removing misallocation – producers or consumers? The incidence of the changes in distortions depends on the pass-through rate of marginal costs to prices, the demand elasticities, market power and interdependencies between all plants. These are in turn determined endogenously by the estimated fundamental demand and production parameters. For all three counterfactuals, the compensating variation is around double the size of the profit gains, and the bootstrapped difference statistically significant.⁷²

⁷⁰Removing both input distortions can be slightly larger or smaller than the sum of gains, because the full interdependencies are taken into account through the model. The sum of the gains from removing either labour or material distortions is close to removing them jointly. This suggests that removing distortions from one input does not affect the gains from removing distortions in the other input.

⁷¹Almost all of the compensating variation, profits and total welfare estimates are positive and significant with few exceptions. The variation in the point estimates across years arise from three confounded factors, however. First, the sample size is different for each year, as some firms enter and some firms drop out of the sample, responsible for scale effects in aggregate compensating variation and profits. Second, due to the unbalanced sample, the composition of firms changes, and in some periods more dispersed τ firms may be sampled than in others, leading to higher estimated misallocation losses. Third, the actual degree of misallocation could have improved or worsened throughout the years.

⁷²I calculate the compensating variation only arising from the output produced by plants in the sample, and do not extrapolate to the whole market size, as for example in Nevo (2000a), in order to have comparable

Figure 9: Welfare gains from removing misallocation distortions



Notes: Plotted are the respective welfare gains from eliminating material distortions (τ_{jt}^M), labour distortions (τ_{jt}^L) or both, summing across all 13 years of the sample. Bootstrapped 95% confidence intervals are shown around the point estimates.

While individual plants experience increases or decreases in their marginal costs, depending on their initial level of τ , there is a more efficient allocation of inputs across plants, which benefits consumers through (average) price reductions. Indeed, the drop in average price per sold output quantity is 6%, 7% and 14% for the three counterfactuals respectively. While there are winners and losers on the firm side, the winners win more than the losers lose. Average profits are increasing despite average price declines, driven by the allocative efficiency gains.⁷³

5.1.2 Higher misallocation in materials than labour inputs

A surprising result is that the welfare losses from misallocation of input materials are higher than those from misallocation of labour. As the right panel in Figure 9 shows, the point estimate is 89% higher at almost 34 billion rupees vs. 18 billion rupees. The difference is statistically significant with a bootstrapped p-value of 0.03**.⁷⁴ This comes at a surprise for a prior that labour is a less flexible input with higher potential for misallocation, particularly

numbers for the consumer and producer gains.

⁷³Section G.14 shows that the variation in markups also decreases in the counterfactual for misallocation in materials.

⁷⁴Comparing the difference in all bootstrapped runs is the appropriate way to test the difference to account for interdependencies, rather than comparing the individual confidence intervals.

in the Indian context.

There is hardly any evidence in the literature comparing misallocation of materials and labour directly.⁷⁵ But perhaps, differences in access to materials plays a bigger role than labour market distortions in this industry, as materials are an important production input with a high estimated output elasticity. For example, the political connectedness of firms (e.g. Faccio, 2006; Akcigit et al., 2018) could be more relevant for distorted spending on material inputs than on labour inputs. In Section 6, I show that differences in geographic access to suppliers can partially explain these costly input material distortions.

The result that material misallocation costs are substantial is important for policy making for two reasons. First, targeting the allocative barriers in the material market could be easier, both politically and practically. As described in Hasan and Jandoc (2014) or Dougherty (2009) many Indian states have stricter labour firing laws for firms above a certain employee threshold. This could be a source of variation in τ_{jt}^L , but removing such policies could be politically challenging, due to valid concerns of employee protection. From a practical point of view, improving workforce mobility is challenging (Bryan et al., 2014) despite widespread regional structural mismatches in the labour market (Bryan and Morten, Forthcoming; Munshi and Rosenzweig, 2016).

Second, the (unmodelled) short term costs associated with reallocation in the materials market are likely to be much smaller than in the labour market, simply due to the fact that goods are reallocated and not people. Reallocation of labour involves hiring and firing, and even if the new equilibrium features higher employment, there are undoubtedly transitional costs for the labourers, which may be large (e.g. Walker, 2013). Of course, addressing the variation in τ_{jt}^M also reallocates market share among firms, which necessarily also involves labour reallocation. But intuitively as well as empirically⁷⁶, the degree of layoffs (and hiring) is larger when removing labour markets distortions.

There are possible concerns that the losses from material misallocation are larger than for labour. First, in terms of external validity, other industries, particularly those that primarily rely on labour, are likely to have higher misallocation losses from labour. Second, I measure labour inputs as the number of total man-days, which does not account for the impact of skill on output in the production function. Misallocation of talent can play a role (Hsieh et al., Forthcoming). I construct a robustness test that accounts for skills by measuring labour as

⁷⁵Hsieh and Klenow (2009) conclude that the misallocation in capital markets is higher than in labour markets in their study based on value added production functions. Dias et al. (2016) use the Hsieh and Klenow (2009) model for gross output with material inputs, and also find that capital misallocation is higher than labour misallocation. Slightly altering their model also sheds light on comparing labour and material misallocation in a Hsieh and Klenow (2009) style model. I analyse this in section G.15.

⁷⁶The counterfactual $\tilde{\tau}_{jt}^L$ involves more layoffs and more hiring than the counterfactual $\tilde{\tau}_{jt}^M$, despite the lower welfare gains.

the total wage bill instead of man-days. If skills are paid a premium, the wage bill captures skills as well. As Appendix G.10 shows, the gap between losses from material and labour misallocation increases, if anything.⁷⁷ Third, wage disparities (which are slightly larger than material input price disparities) may be interpreted as misallocation themselves, and reducing those could offer additional gains. I keep plant level material input prices as well as wages constant throughout the counterfactuals, because I assume that firms are price takers on the input side. While this is an arguably realistic assumption for materials, given that there are much fewer and larger upstream firms (see Appendix B), it might not be the case for labour.

5.1.3 Accounting for changes in tax revenues

The interpretation of the misallocation distortions τ is broad and could pick up any form of barriers and implicit costs. For the extreme case, that all of the τ are only plant level input tax differences, I also calculate the implied impact on government tax income. I add the difference in tax income between the counterfactual and factual equilibrium to total welfare, as reported in Appendix G.11. Across the whole period, including government tax income changes in the calculation slightly increases the welfare gains from removing misallocation distortions.

5.1.4 Correlation of distortions and TFPQ

As Restuccia and Rogerson (2008) show, the correlation between distortions τ_{jt} and plant level productivity ($TFPQ = \Omega_{jt}$) matters significantly for welfare losses.⁷⁸ In the counterfactual, previously high τ plants (constrained) tend to grow while low τ plants shrink.⁷⁹ If the constrained high τ_{jt} plants are also the more productive plants (high Ω_{jt}) the aggregate welfare effects are larger. In Appendix G.12, I show that the estimated correlations between τ_{jt} and Ω_{jt} are low, and that a higher correlation would have implied even higher welfare effects.

5.1.5 Insights from bootstraps: returns to scale are important

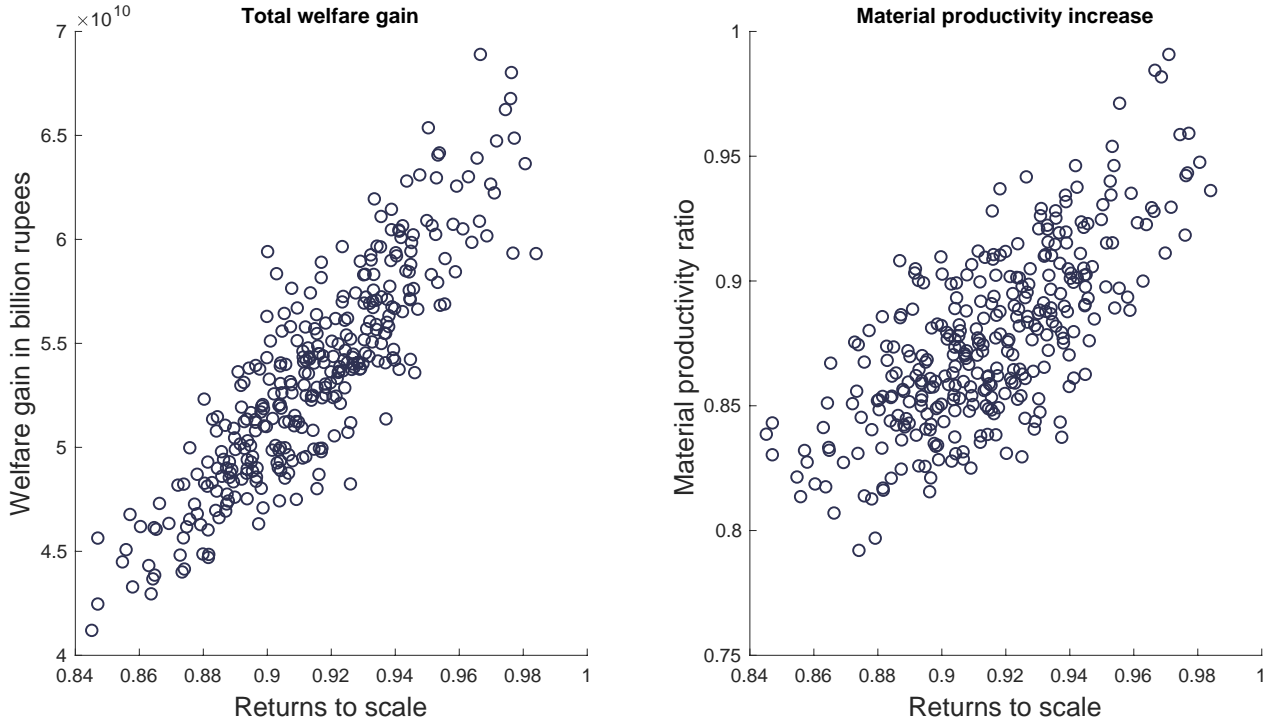
An advantage of the estimated structural model in this paper is that we can examine the sensitivity of the welfare gains with respect to specific underlying parameters that vary across

⁷⁷Note that I do not construct an analogous robustness check for materials. This is because output is measured in quantity. Higher material quality (i.e. expenditure) is likely to increase the quality and the price of output, not necessarily its quantity. Higher skilled labourers, on the other hand, more likely increase the quantity and the price of output.

⁷⁸Of course plant size also matters, so we should think of it as correlation weighted by size to be precise, see also Hopenhayn (2014b) for a theoretical exploration.

⁷⁹Whether firms shrink or grow also depends on the (estimated) interdependencies between firms.

Figure 10: Correlation between returns to scale, welfare and material productivity gains



Notes: Plotted are the respective outcomes from the bootstrapped runs, where the returns to scale is a function of the bootstrapped underlying structural parameters.

bootstraps. Figure 10 (left panel) plots the total welfare gains in each bootstrapped run against the returns to scale, which are a function of the underlying bootstrapped structural production side parameters. The returns to scale are a significant driver of the size of the estimated welfare losses from misallocation. The 95% confidence interval means that returns to scale of 0.97 vs 0.86 are associated with around 40% higher welfare losses of 65 vs. 45 billion rupees. Table 4 shows a regression of the total welfare gains on the bootstrapped returns to scale. The R^2 is high – almost 80% of the welfare variations across bootstraps can be explained by the variation in estimated returns to scale alone. The second column includes additional statistics from the factual equilibrium. As expected, the variation in plant level productivity Ω_{jt} as well as its correlation with plant level distortions are positively and significantly associated with welfare losses from misallocation, as discussed in the previous paragraph.

The size of the standardized coefficient on the returns to scale is much larger than of any of the other coefficients. Hopenhayn (2014b) shows that the original Hsieh and Klenow (2009) model is in theory highly sensitive to their constant returns to scale assumption. Indeed, when I use their model with data on the entire manufacturing sector or just my sample, and assume returns to scale of 0.92, all gains in their model are eliminated and some even turn negative, as Section 5.3 shows. This corroborates the theoretical insights from Hopenhayn

Table 4: Total welfare gains and statistics from factual equilibrium across bootstraps

	(1)	(2)
Returns to scale	0.90*** [31.4]	1.04*** [19.8]
SD(Ω_{jt})		0.21*** [4.1]
SD($\frac{1}{\eta_{jt}}$)		0.19*** [7.5]
Corr(Ω_{jt}, τ_{jt}^M)		0.15*** [3.2]
Corr(Ω_{jt}, τ_{jt}^L)		0.18*** [4.0]
Corr($\frac{1}{1+\eta_{jt}}, \tau_{jt}^M$)		-0.17*** [-4.8]
Corr($\frac{1}{1+\eta_{jt}}, \tau_{jt}^L$)		-0.08*** [-2.7]
Corr(τ_{jt}^M, τ_{jt}^L)		0.02 [0.9]
R^2	0.80	0.88

Notes: The table shows the estimates from an OLS regressions of total welfare gains (from both material and labour distortions) on statistics in the factual equilibrium. There are 330 bootstrapped runs, and each run is equivalent to one observation for this regression. Coefficients are standardized and t-statistics in square brackets based on robust standard errors.

(2014b) empirically. There are two important messages emerging from this. First, the model and estimations of misallocation losses presented in this paper – even though sensitive to the returns to scale – are much less sensitive than coarser approaches following Hsieh and Klenow (2009). Second, the uncertainty in the welfare gains is primarily driven by the uncertainty around the returns to scale. It is therefore critical to accurately estimate the returns to scale for different sectors to have reasonably unbiased welfare estimates, and if possible, construct confidence intervals around them.

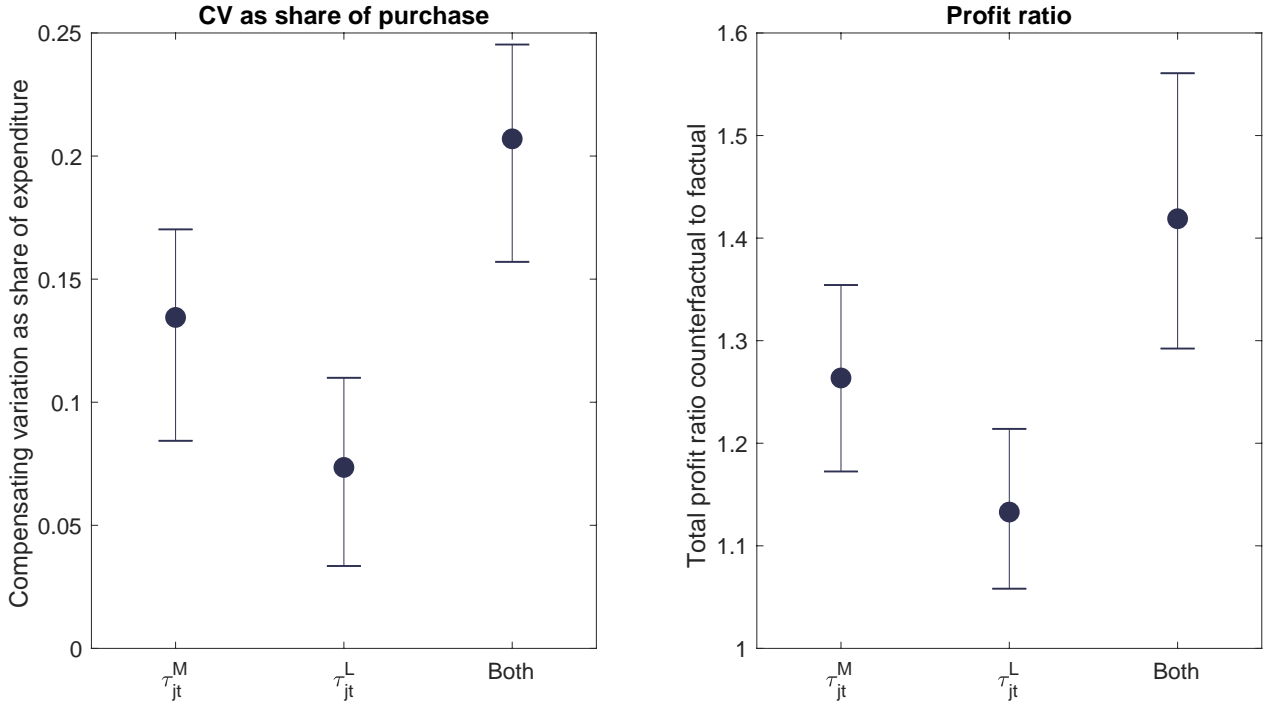
5.1.6 Large size of misallocation losses

Finally, it is left to discuss whether the reported welfare gains of 34, 18 and 53 billion rupees in the respective counterfactuals for $\tilde{\tau}_{jt}^M$, $\tilde{\tau}_{jt}^L$ and both, are large. Since I only cover the cast iron industry, it is most appropriate to set the gains into perspective of the size of this industry. Total sales are 171.5 billion rupees, so the total welfare gains are 20%, 11% and 31% of total sales of the plants in the sample.

I plot the relative levels of total compensating variation and profits in Figure 11 (for annual figures see Table 13 in Appendix G.8). I express the expected average compensating variation *per unit purchased* as share of the (factual) weighted average unit price.⁸⁰ I express the profit ratio as total profits in the counterfactual over total profits in the factual equilibrium. Removing misallocation in materials and labour increases consumer welfare equivalent to a price drop of 21%, while profits grow by 42%. Considering that these comparative statics

⁸⁰This is a useful statistic since each simulated consumer purchases one unit (see Equation (8)).

Figure 11: Interpretation of the size of welfare gains



Notes: Plotted are the respective welfare gains from eliminating material distortions (τ_{jt}^M), labour distortions (τ_{jt}^L) or both, summing across all 13 years of the sample. The left panel expresses the compensating variation per unit purchased as a share of the unit price (i.e. as share of expenditure on the products in the sample). Bootstrapped 95% confidence intervals are shown around the point estimates.

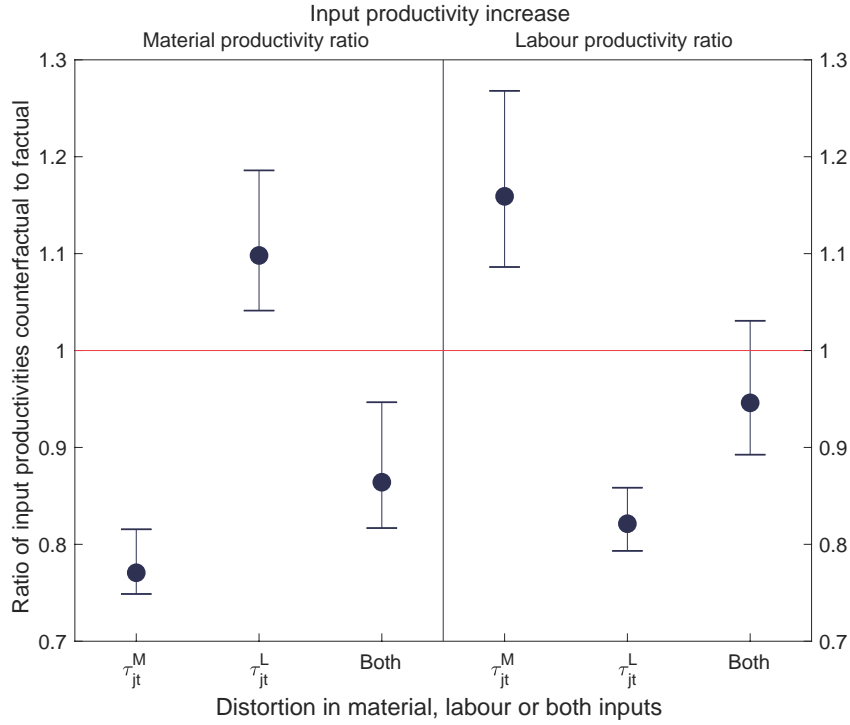
do not entail any technological innovation or diffusion, nor change the inherent factor price differences between plants and regions, the welfare gains from reallocation in material input and labour markets are sizeable. While general equilibrium effects across sectors are beyond the scope of this paper, we know that under complementarity between sectors, misallocation in one sector can have large indirect effects on the economy through an input output structure (Jones, 2011, 2013).

5.2 Input productivities

Figure 12 reports the changes in aggregate physical material productivity (physical output/physical materials input) and aggregate physical labour productivity (physical output/worker) expressed as a ratio of counterfactual to factual input productivity. The figures are pooled across all years, i.e. taking the sum of the quantities across years before calculating the ratios.⁸¹

⁸¹Table 15 in Appendix G.9 reports the annual ratios as well as pure aggregate physical output changes. Table 16 in Appendix G.9 reports the outcomes in terms of aggregate revenue productivities. The changes in physical input productivities are the more relevant metric to discuss as we do not need to deal with deflation (see Appendix G.9).

Figure 12: Aggregate input productivity gains from removing misallocation distortions



Notes: Plotted are the respective input productivity ratios (counterfactual to factual equilibria) from eliminating material distortions ($\bar{\tau}_{jt}^M$), labour distortions ($\bar{\tau}_{jt}^L$) or both, pooling across all 13 years of the sample. Bootstrapped 95% confidence intervals are shown around the point estimates.

5.2.1 Distortions increase aggregate input productivity of same input

The first result is that material productivity *decreases* when material distortions are removed, and *increases* if labour distortions are removed. The same holds analogously for labour productivity. This might be initially surprising but has an intuitive explanation. When the distortions in the material market are lifted, plants with previous constraints use relatively more materials and plants with previous preferential access use relatively fewer materials. But due to the allocative efficiency improvements, the former plants expand their material use more than the latter plants reduce their use. On aggregate, the removal of frictions amounts to higher aggregate incentives to use that input relative to other inputs. This means that improving material misallocation can actually decrease aggregate material productivity through increased incentives to use that input. This result appears much like Jevon's paradox (Jevons, 1865), where increases in energy efficiency increase aggregate energy intensity due to a rebound effect, because energy is cheaper to use.

5.2.2 No misallocation losses in aggregate input productivities

When both input distortions are removed, aggregate input productivities slightly decrease (in the case of labour productivity insignificantly). This is because both aggregate outputs

and inputs grow in the counterfactual.⁸² The aggregate analysis masks high heterogeneity in changes of input productivities across plants, which are discussed in Appendix G.13. As for the welfare gains, the returns to scale in production matter significantly for the size of the gains in input productivities as seen in the right panel of Figure 10. The results suggests that improvements in sectoral material efficiency require innovation and technology adoption, at least in this sector, as there is a limited role of improving allocative input distortions. Yet, with the introduction of new technologies, reallocation can still play an important part in the dynamics of the industry, as documented by Collard-Wexler and De Loecker (2014) for the US steel industry.

5.2.3 Emissions per welfare dollar higher from misallocation

One way we could include the environmental externalities of production (see Appendix F) into the welfare calculation is to compare the increase in emissions with the increase in welfare in a back of the envelope calculation. We know that materials use increases in the counterfactuals, as well as profits and consumer welfare. Since we do not have a baseline of consumer welfare, but only the compensating variation, we could just add it, in an admittedly simplistic way, to the profit increase. If we take the percentage increase of emissions as the percentage increase of material inputs,⁸³ and compare it to the welfare increase, we find that welfare grows 39% more than emissions. Therefore, while pure quantity input intensities are slightly increasing, the emission intensity of welfare is decreasing in the counterfactual. This is driven by the comparatively larger gains in welfare than in emissions from removing misallocation distortions.

5.3 Comparison to aggregate TFP results in the literature

The aggregate input productivity results may be somewhat surprising. Perhaps most prominently, Hsieh and Klenow (2009) found that aggregate TFP would increase by 40-60% in India if it equalised its marginal products to US levels. If sectoral TFP increased by such an extent, we would also expect the input productivities to increase in a similar fashion (with a near constant returns to scale). I can define an analogous standard aggregate Cobb Douglas production function to examine the impact on “aggregate TFP”. Using my estimated

⁸²One immediate concern is that the distortions in capital have been preserved across factual and counterfactuals. I reran the entire analysis where I also remove any differences in the inferred rental rates (i.e. “ τ^K ”) in the same way as the labour and materials distortions, and calculate counterfactuals. The conclusions hardly change: the point estimates for the ratio in physical material productivity and labour productivity are slightly below one. Full results including welfare analysis with capital distortions available on request.

⁸³Of course there might be non-linear relationships, both in the mapping from inputs to emissions as well as from emissions to welfare

elasticities from the plant level, I find that aggregate TFP is roughly the same in the factual and counterfactual equilibria. Why is there this apparent difference in this paper to some of the key literature? There are three explanations that can realign these results.

The first is of interpretive nature. The TFP results of [Hsieh and Klenow \(2009\)](#) can be regarded as welfare results, as they are equivalent to gains in the utility of a representative consumer with CES demand as in [Dixit and Stiglitz \(1977\)](#) or [Melitz \(2003\)](#). Appendix G.15 shows this explicitly. When their results are interpreted as gains in utility rather than gains in physical productivity, they are consistent with the welfare gains in this paper.

Second, the analysis of [Hsieh and Klenow \(2009\)](#) assumes that aggregate inputs do not change when removing distortions. They construct a statistic that summarises the loss from misallocation into a TFP measure, keeping aggregate inputs constant. When I use the plant level estimates of TFPQ (Ω) to calculate their efficient benchmark TFP_s^* ⁸⁴, I obtain gains in TFP_s^* that are roughly 100-200% larger than the baseline aggregate Cobb-Douglas TFP. This is similar to their 127% TFP gain estimate for India in the 90s. However, this requires aggregate inputs to remain constant.⁸⁵ When distortions are removed, there is little reason why aggregate input demand should remain constant in the counterfactual equilibrium.⁸⁶ In this paper, I can distinguish between TFP and welfare while comparing equilibria where the optimising behaviour of all plants, and potential aggregate input growth is taken into account. The results suggest that it is not aggregate physical TFP that increases, but welfare, which is a subtle but important distinction.

Third, there is a set of methodological and empirical differences to keep in mind when comparing the divergent TFP findings. For example, I observe plant specific input prices, use gross output instead of value added production functions, estimate all elasticities and returns to scale, allow for endogenous plant varying markups, and don't restrict the correlation between TFPQ and input distortions. Appendix G.15 provides more details and replicates the [Hsieh and Klenow \(2009\)](#) model and shows how the TFP results respond to changes to the above mentioned elements.⁸⁷ The returns to scale, in particular, affect TFP gains in their

⁸⁴ $TFP_s^* = \left[\sum_i TFPQ_{si}^{\sigma-1} \right]^{\frac{1}{\sigma-1}}$, the value depends on σ , here between 3 and 6.

⁸⁵This is best seen when considering their equation for TFP losses, where the inputs only drop out if they remain constant across the inefficient and efficient equilibria (see Appendix G.15 for more details):

$$\frac{Y_s}{Y_s^*} = \frac{(TFP_s K_s^{\alpha_s} X_s^{\beta_s} L_s^{1-\alpha_s-\beta_s})}{(TFP_s^* K_s^{\alpha_s} X_s^{\beta_s} L_s^{1-\alpha_s-\beta_s})} = \frac{TFP_s^{\theta_s}}{TFP_s^*}$$

⁸⁶When attempting to find a counterfactual equilibrium where I constrain aggregate inputs to factual levels, a set of algorithms with a range of starting points fail to converge to a point where firm first order conditions would be satisfied. This suggests that – at least in our case – there is no counterfactual equilibrium with the same level of aggregate inputs.

⁸⁷I cannot fully nest their approach in my approach due to the substantial difference in both demand

Table 5: Bias from constant markups

	Compensating Variation			Profits			Total welfare		
	$\tilde{\tau}_{jt}^M$	$\tilde{\tau}_{jt}^L$	<i>both</i>	$\tilde{\tau}_{jt}^M$	$\tilde{\tau}_{jt}^L$	<i>both</i>	$\tilde{\tau}_{jt}^M$	$\tilde{\tau}_{jt}^L$	<i>both</i>
Welfare losses baseline (bil. Rs.)	21.8	12.1	33.9	12.2	6.1	18.3	34.1	19.2	53.3
Bias with constant markups	-23%	-30%	-14%	-18%	-22%	-12%	-21%	-27%	-13%

Notes: The first row shows the baseline welfare losses from misallocation in inputs, where $\tilde{\tau}_{jt}^M$ refers to the counterfactual with removed material distortions, $\tilde{\tau}_{jt}^L$ with removed labour distortions and *both* to both removed. The bias in welfare losses is calculated from a counterfactual, where “naive” distortions are removed using this paper’s model. Naive distortions are inferred from ignoring the variation in markups.

model substantially and can even turn them negative.

5.4 Markup changes and ignoring markups

In this paper, I have accounted for markups that are endogenous and variable. This section first briefly examines how much the markups change endogenously. I then use a measure of distortions that ignores variation in markups to calculate counterfactual gains from mismeasured “naive” distortions.

5.4.1 Markup variation across counterfactuals

On average, the endogenous markup changes by 5% for each plant between the factual and counterfactual equilibrium.⁸⁸ This is large compared to the 7% average deviation of markups from the average markup in the factual equilibrium. This suggest that accounting for the endogeneity of markups is important. With exogenous markups, we would be over- or undercounting input misallocation losses.

5.4.2 Mismeasuring distortions with constant markups generates bias

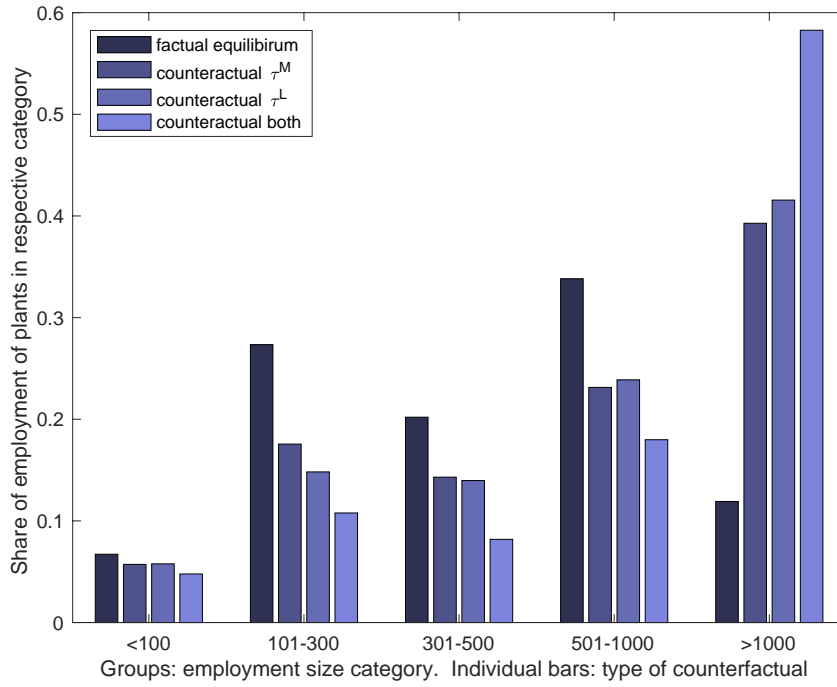
I obtain “naive” τ by setting the demand elasticities η_{jt} (and therefore markups) constant when calculating this version of τ from Equation (3).⁸⁹ I solve for the new counterfactuals, where I remove these naive τ instead. The markups are still allowed to adjust endogenously in the counterfactual. That is, I mismeasure the τ but still use the same model with the same primitives that allows for variable markups in counterfactual equilibria. As Table 5 shows, the inferred welfare costs from misallocation are substantially *lower* than in the baseline

models. [Ho and Ruzic \(2017\)](#) nest a model allowing for non-constant returns to scale, and the difference in inferred misallocation losses are substantial for the US. In recent work, [Haltiwanger et al. \(2018\)](#) reject some of the key assumptions of the [Hsieh and Klenow \(2009\)](#) model using more detailed data. With my data, I can reject the same tests of the validity of the necessary assumptions in [Hsieh and Klenow \(2009\)](#).

⁸⁸Appendix G.14 reports further statistics.

⁸⁹They are displayed in Figures 26 and 27 and briefly described in Section 4.3.3. Despite the resemblance of the standard deviation of the naive and the baseline distortions, the welfare bias is still considerable, as it matters which distortions are measured for which plants.

Figure 13: The effect of misallocation distortions on the size distribution of plants



Notes: The vertical axis is the share in total employment by the plants belonging to a category. The categories from left to right are the size category in terms of plant employment: <100, 101-300, 301-500, 501-1000 and >1000. The left bar in each grouping is the factual equilibrium, the second bar the counterfactual with removed material distortions $\tilde{\tau}_{jt}^M$, the third bar the counterfactual with removed labour distortions $\tilde{\tau}_{jt}^L$ and the fourth bar the counterfactual where both distortions are removed.

version.⁹⁰ Mismeasuring distortions can considerably bias welfare conclusions. Disentangling distortions from fundamental heterogeneity can thus be important to avoid detecting allocative inefficiencies where there are in fact none, or failing to detect them where they in fact exist.

5.5 The effect of misallocation on the size distribution of plants

Before I analyse the role of supplier access, I briefly discuss the effects of the input distortions on the size distribution of firms. [Hasan and Jandoc \(2014\)](#) show that on average, 84% of India's manufacturing workers are in small firms, but 50% of China's workers are in large firms. They hypothesise that this is one of the proximate reasons for lower growth in India, as small firms tend to be less productive.⁹¹ Figure 13 shows that input misallocation is part of the reason for skewed firm distributions. In the counterfactuals, a much larger share of workers is in large firms. This can be viewed as complementary microlevel evidence to [Bento](#)

⁹⁰[Ho and Ruzic \(2017\)](#) present evidence that the [Hsieh and Klenow \(2009\)](#) model understates misallocation losses for the US manufacturing sector when markups are industry specific instead of common to all industries. Their markups are constant within industries.

⁹¹[Kothari \(2014\)](#) argues that developing countries typically have a thick left tail in the firm size distribution, in part driven by lower demand for high quality products predominately produced in larger firms, with some evidence for India. See more recent discussions on firm size distributions in the developing context e.g. in [Cirera et al. \(2018\)](#).

and Restuccia (2017) who document a positive relationship between average productivity and average firm size across countries, and show that distortions keep average firm size small.⁹²

6 Supplier access and misallocation

What is causing the costly material input distortions? I next test the hypothesis of whether differences in access to input suppliers drive wedges into the efficient allocation of materials. I first describe the freight transport issues in India that are likely to be captured in the estimated input distortions. Then I construct a measure of supplier access. Finally I provide the empirical strategy and specification before presenting the results.

6.1 Freight transport issues in India

India's freight transport infrastructure has often been criticised from industry and policy makers alike. Its poor state has been identified as a key constraint for the efficient running and expansion of heavy industry and steel in particular (NCAER, 2015). Inadequate road quality and severe congestions result in high and uncertain transit times, with average truck speeds at around a third of those in developed countries (NTDPC, 2014).

While the share of freight traffic on rail is at around 30% overall (NTDPC, 2014), it is more important for the steel industry with a rail share of around 70% (EY, 2014). There are issues with rail shipping that mirror the problems with road shipping. Passenger trains and freight trains share the same tracks, leading to congestion. Further delays are frequent due to outdated infrastructure operating above capacity limits, breakdowns, different rail gauges requiring different wagons, and numerous state border checkpoints for tax purposes which can take days or weeks to clear (EY, 2013; NTDPC, 2014; EY, 2014; NCAER, 2015). Freight trains are only travelling at an average speed of 25km/h (Appendix H.1) and the Government of India is investing heavily to increase the speed and reliability with current rail infrastructure projects (NTDPC, 2014). Van Leemput (2016) estimates that India faces higher internal than international trade barriers.

Steel plants require to ship a large amount of heavy and bulky inputs.⁹³ The above described issues suggest that there are two types of costs to shipping. One type are the

⁹²In line with Hsieh and Klenow (2014), I find that larger plants are more negatively affected from the input distortions. In terms of age, older plants are more adversely affected from material input distortions but benefit from labour distortions. The results on distortions and size distribution are also consistent with the finding of Martin et al. (2017). They show that dismantling small scale reservations by removing restrictions on firm size (India's SSI policy) led to output growth driven by the expansion of previously size constrained firms. See also Alfaro and Chari (2014) who analyse the firm size distribution in India following the end of the licence Raj.

⁹³One tonne of steel requires the transportation of more than four tonnes of input materials (NCAER, 2015).

Table 6: Transport obstacles as distortions: Some evidence using [World Bank \(2005a\)](#)

	(1)
Transport obstacle?	0.24**
	(0.10)
N	27
R^2	0.19

Notes: The regression is at the district level in 2005. The dependent variable is τ_{jt}^M demeaned by the weighted geometric mean and in logs. The independent variable is the district average of the survey question whether transportation is an obstacle to firm growth from [World Bank \(2005a\)](#). Year is 2005, and only respondents from the metals or minerals industry are used. Robust standard errors are in parentheses.

shipping fees, the direct trade costs. The other types are indirect trade costs, such as delay and uncertainty. In fact, we know from the trade literature, that these “indirect” costs of trade are large ([Anderson and Van Wincoop, 2004](#)). They can also include search costs and contract enforcement costs ([Startz, 2018](#)). Using a specific railway line in India, [Firth \(2017\)](#) presents evidence that the *variance* of shipping time causes the bulk of costs to firms and constrains their operation.

Crucially, the shipping fees are explicitly included in the (factory gate) input prices and are thus observed. The indirect trade costs, however, are not accounted for. Differences in indirect trade costs would therefore be captured by the estimated input distortion. Plants that need to source from further away are more likely to experience any of these issues in freight transportation and likely have higher indirect trade costs and distortions. The aim of this section is to test this hypothesis.

I use the Enterprise Survey in India from [World Bank \(2005a\)](#) for some motivating evidence. The survey asks firms whether transportation is an obstacle for their growth. Around a third of the firms answer that transportation is an obstacle. I regress the logged demeaned input distortion τ_{jt}^M , which I will derive below, on the district average of the responses in Table 6. Plants that claim that they face transportation obstacles also have higher estimated material input distortions τ_{jt}^M .

6.2 Measuring supplier access

A first measure for access to suppliers is the district rail penetration, which is simply the total rail km divided by the district size in square km. [Adamopoulos \(2011\)](#) uses a similar measure on a country level for international comparisons, and I use this measure for some additional results. Since it only varies at the district level, we cannot condition on district fixed effects and it is thus likely endogenous. It also does not capture access to suppliers through the transportation network outside of the district.

To obtain a measure of a plant’s potential to reach suppliers, I construct a measure of

supplier access SA_{dt} for each district d in period t . Redding and Venables (2004) provide a theoretical foundation of both market access on the output side and supplier access on the input side. This is the access to *potential* suppliers, not necessarily the actual supplier choices made by plants.⁹⁴ It is similar to the measure of market access in Donaldson and Hornbeck (2016), but on the input side and only considering relevant input suppliers (as opposed to population size):

$$SA_{dt} = - \sum_h T_{dh} N_{ht} \quad (9)$$

where T_{dh} are the costs in district d to source from district h and a function of the fastest path through the transport network. N_{ht} is the share of the relevant suppliers in district h in the country-wide value of the supplier industry at time t .⁹⁵ The relevant supplier industries are mainly pig iron and coking coal, derived from the detailed information of input use of the cast iron plants in the data. I also use a version of SA_{dt} for robustness checks where I exclude the same district such that $h \neq d$. For T_{dh} I follow the literature and use a function that is concave in the fastest path FP_{dh} from h to d , which also captures mobilisation costs:

$$T_{dh} = 1 + FP_{dh}^{0.8} \quad (10)$$

where the value 0.8 as well as the structure of the function is in line with recent studies relating travel time to costs.⁹⁶ I also use a linear version $T_{dh} = FP_{dh}$ for robustness checks. I calculate the fastest route FP_{dh} using network algorithms. Since we have an undirected graph with positive weights, I can use Dijkstra’s 1959 algorithm using the distance divided by the speed as edge weights.⁹⁷ The resulting histogram of bilateral shipping times is plotted in Figure 31 in Appendix H.2. The median shipping time is 62 hours, and manual inspection yields shipping times between district pairs that are close to estimates using Google Maps for

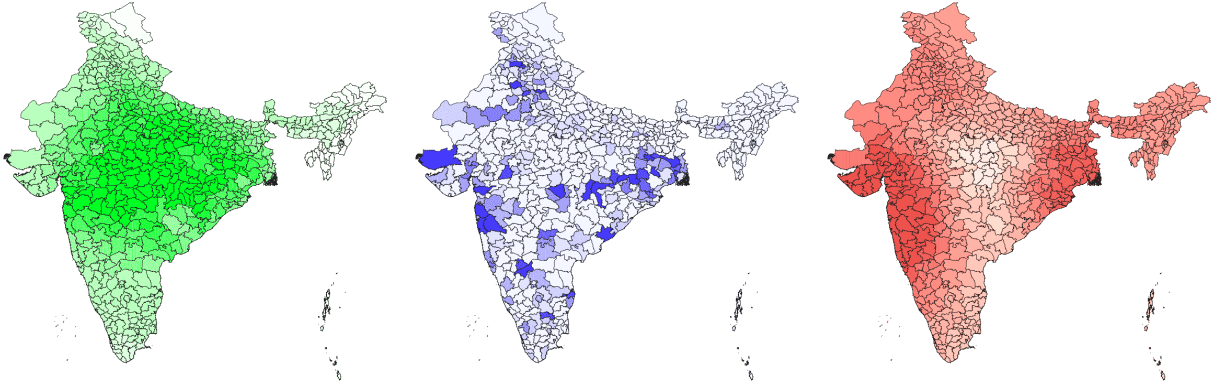
⁹⁴I do not observe plant to plant trade. The measure should therefore be interpreted as access to potential suppliers. This is standard in the literature, dating back to the measure of “market potential” in Harris (1954).

⁹⁵For each district I multiply the plant level values with their sampling multiplier, where the sampling multiplier constitutes how many plants are represented by each plant, to recover a measure based on the universe of plants. Note that when using logged SA_{dt} , it does not matter whether we take N_{ht} as the share or the absolute size of supplier industries, as it will be absorbed by the year fixed effects.

⁹⁶See e.g. Baum-Snow et al. (2018), Alder (2017), Roberts et al. (2012). Also Au and Henderson (2006) use a concave relationship between distance and shipping costs. I also performed some robustness checks by varying parameters (γ_1, γ_2) in $T_{dh} = 1 + \gamma_1 FP_{dh}^{\gamma_2}$, with similar results.

⁹⁷With my network of 1.6 million edges and 1.2 million nodes, it takes only around a minute to calculate the fastest path matrix with an optimised algorithm. The maximum running time for Dijkstra’s algorithm with a Fibonacci heap (Fredman and Tarjan, 1987) for one source node to all other nodes is $O(|E| + |V| \log |V|)$ where $|E|$ is the number of edges and $|V|$ the number of nodes. Other studies that used Dijkstra’s 1959 algorithm to analyse economic outcomes are e.g. Faber (2014) or Donaldson and Hornbeck (2016). If an exact vector based network is not available but only rasterized data then a fast marching algorithm can be used (Allen and Arkolakis, 2014; Faber, 2014; Allen and Atkin, 2016; Alder, 2017).

Figure 14: Average supplier access, supplier presence and change in supplier access



Notes: The left map shows the average supplier access of districts. The middle map shows the average size of supplier industry. The right map shows the average of the absolute deviation of the supplier access from its average within districts. Darker shading mean higher values.

the same district pairs.

As mentioned in the introduction, this paper stands out from the literature by combining (i) directly measured costs T_{dh} with (ii) only including *relevant* potential supplier industries that I recover from the input product codes. Based on the calculated fastest path FP_{dh} , Figure 14 plots the average SA_{dt} over the sample period for all districts in India in the left map. The middle map plots the location of the supplier industries (average value over time). The right map plots the average of the absolute deviation of the SA_{dt} to its within district average, i.e. a measure of how much it changed over time. This is a summary of the variation that I use for identification to which I turn next.

6.3 Identification and estimation

There are two sources of variation in supplier access SA_{dt} , the time-invariant network component T_{dh} and the time-variant geography of shares of the relevant potential suppliers N_{ht} . There are two endogenous location decisions. One is the location decision of plants, which likely depends on the transport infrastructure. The other is the location decision of infrastructure, which is placed strategically, not randomly. Both give rise to omitted variables and reverse causality concerns. Despite these concerns, others have used changes in infrastructure (e.g Donaldson and Hornbeck, 2016)⁹⁸, but I cannot use this variation even if I wanted to, as there were hardly any changes in railway infrastructure for this sample period in India (see Appendix H.1).

My strategy is to use variation in the growth or decline of (distant) supplier industries as the identifying source for variation in SA_{dt} by using district fixed effects. For example, if the

⁹⁸Donaldson and Hornbeck (2016) state that this is their main endogeneity concern. They run robustness checks by controlling for the presence of nearby railroad tracks.

supplier industry grows in a distant district A, and district A is better connected to district B than to C, then the supplier access for plants in district B improves compared to those in C. The main specification is:

$$\log(\tau_{jt}^M) = \phi SA_{dt} + \mathbf{X}_{jt}\boldsymbol{\chi} + \boldsymbol{\lambda}_d + \boldsymbol{\kappa}_t + \iota_{jt} \quad (11)$$

where τ_{jt}^M is the material input distortion estimated from the structural model and SA_{dt} the supplier access for district d at time t .⁹⁹ \mathbf{X}_{jt} is a vector of control variables, such as plant age or legal form, $\boldsymbol{\lambda}_d$ are district fixed effects, $\boldsymbol{\kappa}_t$ year fixed effects and ι_{jt} an error.

If the growth in supplier industries in other districts is uncorrelated to the shocks to distortions ι_{jt} , the effect is identified. There may be further reverse causality concerns warranted if current plant distortions affect supplier industries in other districts, or endogeneity issues when plants and supplier industries are hit by correlated shocks. To at least partially address these concerns, I use lagged supplier access SA_{dt-1} and obtain similar estimates. I also include plant fixed effects in reported robustness checks, which along with further robustness checks are reported in Section 6.5.

By focusing on a single product, the elasticities and the recovered distortions are relatively well measured. They are disentangled from fundamental heterogeneity across plants within sectors, both in terms of production technique and demand conditions. This puts us in a position to explain a distortion with a signal to noise ratio that is much higher than usually found in the literature. They are still estimated, however. Since they are used as a dependent variable there is no classical measurement attenuation bias problem. The advantage of the method in this paper is that I can recover a distribution of τ_{jt}^M for every single plant, based on the parametric bootstrap from estimated production and demand parameters. This allows me to perform robustness checks with respect to the uncertainty in the dependent variable.

6.4 Results

6.4.1 Access to suppliers decreases material input distortions

The distortions τ_{jt}^M do not capture shipping fees, as those are accounted for in the model and observed in P_{jt}^M . In Appendix H.3 I show that input prices inclusive of shipping fees P_{jt}^M and supplier access SA_{dt} are negatively correlated as expected.¹⁰⁰ The indirect costs of trade, such as uncertainty or delays in shipping or search costs, on the other hand, will be picked up by the distortions τ_{jt}^M . I test the hypothesis that with longer routes to suppliers, i.e. a lower

⁹⁹I demeaned τ_{jt}^M within every year.

¹⁰⁰The relationship is marginally insignificant, possibly due to heterogeneous input quality having a bigger effect on input prices than shipping fees.

Table 7: Input material distortion and supplier access

	(1)	(2)	(3)	(4)	(5)	(6)
Supplier access	-0.27*** [-3.92]	-0.23*** [-3.19]	-0.28** [-2.05]		-0.19** [-2.43]	-0.26** [-2.17]
Supplier access (lagged)				-0.31** [-2.15]		
Rail km/sqkm					-0.25* [-1.70]	-0.21 [-1.47]
Plant level controls	No	Yes	Yes	Yes	Yes	Yes
District level controls	No	Yes	No	No	Yes	Yes
State level controls	No	Yes	No	No	Yes	Yes
Year FE	No	Yes	Yes	Yes	Yes	Yes
District FE	No	No	Yes	Yes	No	No
N	926	882	926	926	882	882
R^2	0.07	0.25	0.46	0.46	0.26	0.26

Notes: The dependent variable is the logged demeaned (by year) distortion in material inputs τ_{jt}^M . Coefficients are standardised. t-statistics in brackets are based on standard errors clustered at the district level. Plant level controls include firm age, dummies on ownership type, and whether plants are part of the census section. District level controls include population, population density and gender ratio in 2001, district population growth from 1991-2001, and whether the district was subject to left wing extremism (who sometimes target infrastructure) in the years leading up to 2009. State level controls include the male literacy rate and the share of male industrial workers in 2001. Data for district and state level controls is based on the population census and retrieved from indiastat.com.

SA_{dt} , the distortions τ_{jt}^M are higher, most likely because the indirect trade costs are higher.

Table 7 shows the results from estimating regression (11). The first two columns are without district effects λ_d .¹⁰¹ The main result is in Column (3), controlling for district fixed effects. A one standard deviation increase in supplier access reduces the material distortion by 0.28 of its standard deviation.

For the interpretation of the results, it is relevant whether the distortions are likely to rather capture input market power (monopsony power) or indirect trade costs. My results are robust to including year and district or plant fixed effects which eliminates monopsony power that is constant along these dimensions.¹⁰² Furthermore, monopsony power is proportional to the measured input distortions as shown in Appendix B. Therefore, the results in Table 7 would only be consistent with a monopsony power story if worse access to suppliers *increased* monopsony power. We would expect, however, that if supplier access is poor, there are few suppliers around, which would suggest that there is less monopsony power. In this sense, the monopsony power story would work against the reported results. In Appendix I.1, I construct two proxies for monopsony power, one based on market share, one based on directly estimating the input price elasticity. The results are robust to controlling for either. In addition, I construct a measure of input distortion net of monopsony power as a dependent variable, which confirms the main results (all additional results in Table 25). Overall, the

¹⁰¹The OLS estimates without district fixed effects are slightly upward biased. This suggest that plants with higher distortions τ_{jt}^M tend to locate in better connected areas.

¹⁰²That is, cast iron plants are allowed to have monopsony power ($dP_{jt}^M/dM_{jt} \neq 0$), as long as they share a common materials price elasticity of material input consumption ($d \log P_t^M/d \log M_t$) or if the plant specific elasticity is fixed over time ($d \log P_j^M/d \log M_j$).

results strongly suggest that indirect costs of trade are captured by the estimated distortions. These distortions in turn lead to misallocation of input materials with adverse aggregate consequences.

For this industry, the welfare costs of materials misallocation are large, and differences in supplier access are a significant contributor. Taking the results at face value, it is worth making the following four points. First, it is important to note that the relationship between distortions and differences in supplier access are not simply features of our spatial reality. It is costly to move goods across space. However, since shipping costs are accounted for, the part in distortions that is due to differences in supplier access can be addressed without necessarily reducing shipping fees.¹⁰³ These are not the transport costs of moving goods across space, but differences in costs generated by the multitude of freight transport issues described in Section 6.1.

Second, this analysis has policy implications, as it provides us with a margin that we can address to improve allocative efficiency. Reducing costs of uncertainty and delays, e.g. by strengthening the transportation infrastructure network, or reducing border checkpoints due to differential tax systems is likely to improve allocative efficiency. In fact, the GST (good and sales tax) reform in 2017 unifies the tax system and substantially reduces border checkpoints within India. Loosely speaking, if indirect trade costs are reduced, we would start to see the relationship between supplier access and input distortions disappear.

Third, while the analysis provides evidence that differences in supplier access generate misallocation, I cannot distinguish between its components, such as information, search or uncertainty costs. Fourth, while there is a significant relationship between supplier access and material input distortions, there is still unexplained variation left in the material distortions as the R^2 show.

In Columns (5) and (6), I add a measure of district rail penetration, which is total rail km per district area. Both omit district fixed effects. In Column (6), I additionally instrument supplier access with a measure of supplier access that is demeaned at the district level to account for the lack of district fixed effects. The supplier access coefficient is robust to controlling for district rail penetration. The effect of district rail penetration is marginally significant as well, which suggest that the shipping constraints within districts may matter as well. I next explore the robustness of the results and provide placebo analyses to test the identification assumption.

¹⁰³Shipping fees can also be regarded as additional spatial frictions, quantified by Behrens et al. (2017) for the US in a setting with endogenous markups, for example.

Table 8: Input material distortion and supplier access: robustness checks

	Excl. own dist.		Linear costs		Plant FE		Rail only	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Supplier access	-0.28**		-0.32**		-0.17*		-0.30*	
	[-2.00]		[-2.32]		[-1.76]		[-1.79]	
Supplier access (lagged)		-0.40**		-0.36**		-0.22**		-0.33*
		[-2.55]		[-2.39]		[-2.00]		[-1.93]
Plant level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	No	No	Yes	Yes
Plant FE	No	No	No	No	Yes	Yes	No	No
N	926	926	926	926	924	924	926	926
R ²	0.46	0.46	0.46	0.46	0.73	0.73	0.45	0.46

Notes: The dependent variable is the logged demeaned (by year) distortion on material inputs τ_{jt}^M . The supplier access in the first two columns excludes the own district of a plant when calculating supplier access $SA_{dt} = -\sum_h T_{dh} N_{ht} \forall h \neq d$. Columns (3) and (4) assume linear costs, i.e. $T_{dh} = FP_{dh}$. Columns (5) and (6) include plant fixed effects. Columns (7) and (8) calculated the fastest path FP_{dh} based on the railway network alone instead of using roads and rails. Coefficients are standardised. t-statistics in brackets are based on standard errors clustered at the district level. Plant level controls include firm age, dummies on ownership type, and whether plants are part of the census section.

6.5 Robustness and three placebo tests

6.5.1 Robustness checks

I perform a number of robustness checks. First, there might be some correlation between simultaneous shocks that affect the cast iron industry, but also its suppliers. In Column (4) of Table 7, I use lagged supplier access, and the effect, if anything is slightly larger.¹⁰⁴ Second, I exclude the own districts of plants when calculating supplier access (Equation (9)) in Column (1) and (2) of Table 8. Third, I use the fastest path directly as cost of shipping (Equation (10)) in Columns (3) and (4) of Table 8. Fourth, Columns (5) and (6) include plant fixed effects. Fifth, for Columns (7) and (8) I use the railway network only, dropping all roads before calculating the fastest path.¹⁰⁵ Overall, the effect of supplier access differences remains robust.

To assess the effect that the dependent variable, the distortion, is estimated, I make use of the estimated covariance matrices of the underlying production and demand parameters. I obtain a different set of τ_{jt}^M for every draw from the distribution of the estimated fundamental parameters. For every set of τ_{jt}^M , I run regression (11) and obtain the point estimates and t-statistics of supplier access. In total, I run 330 regressions and plot the point estimates and t-statistics in Figure 32 in Appendix I.2. The average point estimate is -0.295 with a minimum to maximum of -0.34 to -0.22. All estimates are significant at the 10% level at least.

I next run three placebo tests to provide further support for the causality of the estimated

¹⁰⁴This suggests an upward bias in the current period supplier access. Suppose a plant receives a shock which reduces some components in its input distortion. If it increases input demand, then the well connected suppliers could benefit, introducing a reverse causality problem, which biases the current period supplier access in the reported direction.

¹⁰⁵In another robustness check, I use the geographic distance instead of the geodesic distance (shortest path) of the network, with similar results.

Table 9: Placebos: labour distortion, *irrelevant* supplier access, or market access

	τ_{jt}^L (1)	Irrelevant inputs			Market access
		(2)	(3)	(4)	(5)
Supplier access	-0.11 [-1.30]				
Supplier access (textiles)		-0.15 [-1.16]			
Supplier access (food)			0.46 [1.55]		
Supplier access (rubber)				-0.33 [-1.26]	
Market access					-0.36 [-1.47]
Plant level controls	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes
N	926	926	926	926	926
R^2	0.42	0.45	0.45	0.45	0.45

Notes: The dependent variable is the logged demeaned (by year) labour distortion τ_{jt}^L in Column (1) and the material distortion τ_{jt}^M in all other columns. The supplier access in the Columns (2) -(4) is based on access to the textiles, food, and rubber industries respectively. The market access variable in Column (5) is access to output markets, i.e. to those industries that buy cast iron. Coefficients are standardised. t-statistics in brackets are based on standard errors clustered at the district level. Plant level controls include firm age, dummies on ownership type, and whether plants are part of the census section.

relationship.

6.5.2 Placebo 1: No effect on labour distortions

The first placebo takes the estimated logged *labour* distortion τ_{jt}^L as dependent variable. The access to materials suppliers should not affect these distortions. Because commuting happens predominately within districts, the access to distant suppliers should also not pick up access to labour markets, which is likely contained in τ_{jt}^L . As Column (1) in Table 9 reports, the association is not statistically significant.

6.5.3 Placebo 2: No effect of access to irrelevant supplier

For a second placebo test, I construct a measure of supplier access to *irrelevant* supplier industries. In particular, access to textiles, food or rubber industries. Columns (2) to (4) of Table 9 show that there is no effect. This shows that it is only differences in access to *relevant* suppliers that is contained in the input misallocation distortions τ_{jt}^M .

6.5.4 Placebo 3: No effect from access to output markets

Finally, a concern is that the input distortions τ_{jt}^M capture additional costs from shipping the outputs, rather than inputs. To test this, I also construct a market access variable. It is based on the size of the industries that *buy* from the cast iron plants in the sample, i.e. the

downstream firms such as engine manufacturing that use cast iron as input. Column (5) in Table 9 reports that the estimate, while negative, is not statistically significant. The data I use provides a measure of distributional (i.e. shipping costs) on the output side. I show in Table 26 in Appendix I.3, that this measure of output shipping costs is significantly correlated to market access. This suggests that the measure of market access captures what we think it should capture, access to buyers, but is not significantly driving distortions on the *input* side.

On the whole, it is only the access to relevant input suppliers that affect material input distortions, and consequently misallocation losses.

7 Conclusion

This paper develops an approach that disentangles input misallocation distortions from fundamental heterogeneity in demand and markups across plants. I can distinguish between effects on producers and consumers, and effects on aggregate productivities. This provides a nuanced picture for the Indian cast iron industry. Removing input distortions in one input decreases the aggregate input productivity of the same input while improving the aggregate input productivity of the other input. This is in part driven by substitution to the input that is becoming more efficiently allocated. I find no evidence that removing misallocation distortions in both inputs would lead to improvements in input productivities. Since aggregate input productivities are determined by aggregate outputs and inputs, I allow aggregate inputs to adjust in counterfactual equilibria, which is in contrast to previous studies in this literature. This result is relevant for policies aimed at improving aggregate material efficiency. At least for the Indian cast iron industry, the results suggest that there are no allocative gains, and that within-firm innovations and technology diffusion are a more promising way to improve aggregate material efficiency.

I find that input misallocation significantly affects the size distribution of plants, keeping plants artificially smaller. There are also significant welfare losses from misallocation. The welfare losses are higher for consumers than for producers, driven by the price effects of input misallocation. The welfare losses from misallocation of materials are larger than those from labour. Even though I ignore any direct welfare costs on the employee side, this is a surprising result. Despite a lack of studies on misallocation of input materials, the results suggest that these distortions could play a bigger role in explaining differences in performance of materials dependent sectors across countries.

In the last part of the paper I ask what these costly material input distortions represent. I find that differences in access to suppliers through the transportation network drives the estimated input distortions. I show that it is only the access to relevant input suppliers, as

access to irrelevant suppliers or access to output markets is not significantly related to the distortions on the input side.

I emphasise that the input misallocation distortions should be interpreted in terms of differences in *indirect* trade costs. Differences in shipping fees represent the spatial reality, as it is inherently costly to ship goods across space, and therefore hard to eliminate. Since input shipping fees are observed and accounted for in the model, the distortions are net of input shipping fees. On the other hand, any indirect trade costs associated with sourcing inputs, which are lower for better supplier access, are captured in the estimated distortions. This sheds more light into the black box of misallocation losses.

The policy implications are that there are aggregate reallocation gains from reducing differences in indirect trade costs, without necessarily decreasing shipping costs. These include, for example, costs of delay, search and uncertainty. State border checkpoints for goods within India, for example, create shipping delays and are, for the purposes of this study, policy distortions that create input misallocation. The described relationship between supplier access and misallocation distortions is likely to have external validity in contexts of industries that require substantial input shipping, but face unreliable transport infrastructure.

While I cannot distinguish between different types of indirect trade costs that increase with remoteness, this paper provides an important insight into the drivers of misallocation. Especially misallocation of input materials has received little attention and we have known even less about underlying determinants. Future research aims to quantify the misallocation costs of differences in supplier access as well as other potential drivers in India and other countries.

The paper provides methodological novelties to the literature. By combining production and demand into a full structural model I disentangle endogenous markups from input distortions. In the case of the Indian cast iron industry, ignoring variation in markups biases the estimated welfare costs from misallocation downwards. I can provide confidence intervals around welfare cost and any other outcome, as I estimate all parameters in the model. This, for the first time in this literature, provides measures of uncertainty around aggregate misallocation losses. The developed approach can be applied to products and countries where quantity and price data on outputs and inputs is available. While there are disadvantages such as higher data requirements and computationally more intensive procedures, the benefits are detailed insights that admit a rich set of outcomes, which hopefully can be useful both for tailoring further research on misallocation as well as informing policy. Shortcomings are that I focus on misallocation in a static sense without dynamic considerations, ignore general equilibrium implications, and focus on the intensive margin ignoring firm entry. These provide

interesting avenues for future research.

References

- Akerberg, D. A., Caves, K. and Frazer, G. (2015), ‘Identification properties of recent production function estimators’, *Econometrica* **83**(6), 2411–2451.
- Akerberg, D., Benkard, C. L., Berry, S. and Pakes, A. (2007), ‘Econometric tools for analyzing market outcomes’, *Handbook of econometrics* **6**, 4171–4276.
- Adamopoulos, T. (2011), ‘Transportation costs, agricultural productivity, and cross-country income differences’, *International Economic Review* **52**(2), 489–521.
- Adamopoulos, T., Brandt, L., Leight, J. and Restuccia, D. (2017), ‘Misallocation, selection and productivity: A quantitative analysis with panel data from china’, *NBER working paper*.
- Adamopoulos, T. and Restuccia, D. (2014a), ‘Land reform and productivity: A quantitative analysis with micro data’, *mimeo, University of Toronto*.
- Adamopoulos, T. and Restuccia, D. (2014b), ‘The size distribution of farms and international productivity differences’, *The American Economic Review* **104**(6), 1667–1697.
- Adao, R., Costinot, A. and Donaldson, D. (2017), ‘Non-parametric counterfactual predictions in neoclassical models of international trade’, *American Economic Review* **107**(3), 633–89.
- Akcigit, U., Alp, H. and Peters, M. (2016), ‘Lack of selection and imperfect managerial contracts: Firm dynamics in developing countries’, *MIT working paper*.
- Akcigit, U., Baslandze, S. and Lotti, F. (2018), ‘Connecting to power: political connections, innovation, and firm dynamics’, *mimeo*.
- Alder, S. (2017), ‘Chinese roads in india: The effect of transport infrastructure on economic development’, *Work. Pap., Univ. North Carolina, Chapel Hill*.
- Alfaro, L. and Chari, A. (2014), ‘Deregulation, misallocation, and size: Evidence from india’, *The Journal of Law and Economics* **57**(4), 897–936.
- Allcott, H., Collard-Wexler, A. and O’Connell, S. D. (2016), ‘How do electricity shortages affect industry? evidence from india’, *The American Economic Review* **106**(3), 587–624.
- Allen, T. (2014), ‘Information frictions in trade’, *Econometrica* **82**(6), 2041–2083.
- Allen, T. and Arkolakis, C. (2014), ‘Trade and the topography of the spatial economy’, *The Quarterly Journal of Economics* **129**(3), 1085–1140.
- Allen, T. and Arkolakis, C. (2019), ‘The welfare effects of transportation infrastructure improvements’, *NBER working paper*.
- Allen, T. and Atkin, D. (2016), ‘Volatility and the gains from trade’, *NBER working paper*.
- Anderson, J. E. and Van Wincoop, E. (2004), ‘Trade costs’, *Journal of Economic literature* **42**(3), 691–751.
- Angrist, J. D. and Pischke, J.-S. (2008), *Mostly harmless econometrics: An empiricist’s companion*, Princeton university press.
- Arellano, M. and Bond, S. (1991), ‘Some tests of specification for panel data: Monte carlo evidence and an application to employment equations’, *The review of economic studies* **58**(2), 277–297.
- Arkolakis, C., Costinot, A., Donaldson, D. and Rodríguez-Clare, A. (2018), ‘The elusive pro-competitive effects of trade’, *The Review of Economic Studies*.
- Armstrong, T. B. (2016), ‘Large market asymptotics for differentiated product demand estimators with economic models of supply’, *Econometrica* **84**(5), 1961–1980.
- Asker, J., Collard-Wexler, A. and De Loecker, J. (2014), ‘Dynamic inputs and resource (mis) allocation’, *Journal of Political Economy* **122**(5), 1013–1063.
- Asturias, J., García-Santana, M. and Ramos, R. (2018), ‘Competition and the welfare gains from transportation infrastructure: Evidence from the Golden Quadrilateral of India’, *Journal of the European Economic Association*.
- Atkin, D. and Donaldson, D. (2015), ‘Who’s getting globalized? the size and implications of intra-national trade costs’, *NBER working paper*.
- Au, C.-C. and Henderson, J. V. (2006), ‘Are chinese cities too small?’, *The Review of Economic Studies* **73**(3), 549–576.
- Baily, M. N., Hulten, C., Campbell, D., Bresnahan, T. and Caves, R. E. (1992), ‘Productivity dynamics in manufacturing plants’, *Brookings papers on economic activity. Microeconomics* **1992**, 187–267.
- Banerjee, A. V., Duflo, E. and Qian, N. (2012), ‘On the road: Access to transportation infrastructure and economic growth in china’, *NBER Working Paper* (w17897).
- Baptist, S. and Hepburn, C. (2013), ‘Intermediate inputs and economic productivity’, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **371**(1986).
- Baqae, D. R. and Farhi, E. (2017), ‘Productivity and misallocation in general equilibrium.’, *NBER working paper*.
- Baum-Snow, N., Henderson, J. V., Turner, M. A., Zhang, Q. and Brandt, L. (2018), ‘Does investment in national highways help or hurt hinterland city growth?’, *NBER working paper*.
- Bayer, C., Mecikovsky, A. M. and Meier, M. (2018), ‘Misallocation, markups, and technology’, *mimeo*.

- Behrens, K., Mion, G., Murata, Y. and Suedekum, J. (2017), ‘Spatial frictions’, *Journal of Urban Economics* **97**, 40–70.
- Behrens, K., Mion, G., Murata, Y. and Suedekum, J. (2018), ‘Quantifying the gap between equilibrium and optimum under monopolistic competition’, *Mimeo*.
- Bento, P. and Restuccia, D. (2017), ‘Misallocation, establishment size, and productivity’, *American Economic Journal: Macroeconomics*.
- Berndt, E. R. and Christensen, L. R. (1973), ‘The translog function and the substitution of equipment, structures, and labor in us manufacturing 1929-68’, *Journal of econometrics* **1**(1), 81–113.
- Berry, S., Levinsohn, J. and Pakes, A. (1995), ‘Automobile prices in market equilibrium’, *Econometrica* **63**(4), 841–890.
- Berry, S., Levinsohn, J. and Pakes, A. (1999), ‘Voluntary export restraints on automobiles: Evaluating a trade policy’, *American Economic Review* **89**(3), 400–430.
- Berry, S. T. (1994), ‘Estimating discrete-choice models of product differentiation’, *The RAND Journal of Economics* **25**(2), 242–262.
- Berry, S. T. and Haile, P. A. (2014), ‘Identification in differentiated products markets using market level data’, *Econometrica* **82**(5), 1749–1797.
- Bhagwati, J. and Panagariya, A. (2014), *Reforms and Economic Transformation in India*, Oxford University Press.
- Bils, M., Klenow, P. J. and Ruane, C. (2017), Misallocation or mismeasurement?, in ‘2017 Meeting Papers’, number 715, Society for Economic Dynamics.
- Bloom, N., Eifert, B., Mahajan, A., McKenzie, D. and Roberts, J. (2013), ‘Does management matter? evidence from india *’, *The Quarterly Journal of Economics* **128**(1), 1–51.
- Blundell, R. and Bond, S. (1998), ‘Initial conditions and moment restrictions in dynamic panel data models’, *Journal of econometrics* **87**(1), 115–143.
- Blundell, R. and Bond, S. (2000), ‘Gmm estimation with persistent panel data: an application to production functions’, *Econometric reviews* **19**(3), 321–340.
- Boehm, J. and Oberfield, E. (2018), ‘Misallocation in the market for inputs: Enforcement and the organization of production’, *Mimeo*.
- Bollard, A., Klenow, P. and Sharma, G. (2013), ‘India’s mysterious manufacturing miracle’, *Review of Economic Dynamics* **16**(1), 59–85.
- BP (2016), *Statistical review of world energy*, BP.
- Brown, M. B. and Forsythe, A. B. (1974), ‘Robust tests for the equality of variances’, *Journal of the American Statistical Association* **69**(346), 364–367.
- Bryan, G., Chowdhury, S. and Mobarak, A. M. (2014), ‘Underinvestment in a profitable technology: The case of seasonal migration in bangladesh’, *Econometrica* **82**(5), 1671–1748.
- Bryan, G. and Morten, M. (Forthcoming), ‘The aggregate productivity effects of internal migration: Evidence from indonesia’, *Journal of Political Economy*.
- Byrd, R. H., Gilbert, J. C. and Nocedal, J. (2000), ‘A trust region method based on interior point techniques for nonlinear programming’, *Mathematical Programming* **89**(1), 149–185.
- Catherine, S., Chaney, T., Huang, Z., Sraer, D. A. and Thesmar, D. (2018), ‘Quantifying reduced-form evidence on collateral constraints’, *Mimeo*.
- Chamberlain, G. (1987), ‘Asymptotic efficiency in estimation with conditional moment restrictions’, *Journal of Econometrics* **34**(3), 305–334.
- Chari, A., Liu, E. M., Wang, S.-Y. and Wang, Y. (2017), ‘Property rights, land misallocation and agricultural efficiency in china’, *NBER working paper* (24099).
- Chari, V. V., Kehoe, P. J. and McGrattan, E. R. (2007), ‘Business cycle accounting’, *Econometrica* **75**(3), 781–836.
- Chen, C., Restuccia, D. and Santaaulàlia-Llopis, R. (2017), ‘The effects of land markets on resource allocation and agricultural productivity’, *NBER working paper* (24034).
- Cheng, W. and Morrow, J. (2018), ‘Firm productivity differences from factor markets’, *The Journal of Industrial Economics* **66**(1), 126–171.
- Christensen, L. R., Jorgenson, D. W. and Lau, L. J. (1971), ‘Conjugate duality and the transcendental logarithmic production function’, *Econometrica* **39**(4), 255–256.
- Christensen, L. R., Jorgenson, D. W. and Lau, L. J. (1973), ‘Transcendental logarithmic production frontiers’, *The Review of Economics and Statistics* **55**(1), 28–45.
- Cirera, X., Jaef, R. F., Gonne, N. and Maemir, H. (2018), ‘The elusive missing middle: The size distribution of firms in sub-saharan africa’, *mimeo*.
- Collard-Wexler, A. and De Loecker, J. (2014), ‘Reallocation and technology: evidence from the us steel industry’, *The American Economic Review* **105**(1), 131–171.
- Costinot, A. and Donaldson, D. (2016), ‘How large are the gains from economic integration? theory and evidence from us agriculture, 1880-1997’, *NBER working paper*.
- Da-Rocha, J.-M., Tavares, M. M. and Restuccia, D. (2017), ‘Policy distortions and aggregate productivity with endogenous establishment-level productivity’, *NBER working paper*.
- Database of Global Administrative Areas (GADM) (2016), ‘Database of global administrative areas’.
URL: <https://gadm.org/>
- David, J. M., Hopenhayn, H. A. and Venkateswaran, V. (2016), ‘Information, misallocation, and aggregate productivity’, *The Quarterly Journal of Economics* **131**(2), 943–1005.
- David, J. M. and Venkateswaran, V. (2017), ‘The sources of capital misallocation’, *NBER working paper* (23129).

- Davis, P. and Schiraldi, P. (2014), ‘The flexible coefficient multinomial logit (fc-mnl) model of demand for differentiated products’, *The RAND Journal of Economics* **45**(1), 32–63.
- Davis, S. J., Grim, C., Haltiwanger, J. and Streitwieser, M. (2013), ‘Electricity unit value prices and purchase quantities: Us manufacturing plants, 1963–2000’, *Review of Economics and Statistics* **95**(4), 1150–1165.
- De Loecker, J. (2011), ‘Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity’, *Econometrica* **79**(5), 1407–1451.
- De Loecker, J. (2014), ‘Firm performance in a global economy’, *Annual Review of Economics* **6**, 201–227.
- De Loecker, J., Goldberg, P. K., Khandelwal, A. K. and Pavcnik, N. (2016), ‘Prices, markups, and trade reform’, *Econometrica* **84**(2), 445–510.
- De Loecker, J. and Scott, P. T. (2016), ‘Estimating market power evidence from the us brewing industry’, *NBER working paper*.
- De Loecker, J. and Warzynski, F. (2012), ‘Markups and firm-level export status’, *The American Economic Review* **102**(6), 2437–2471.
- Deaton, A. and Muellbauer, J. (1980), ‘An almost ideal demand system’, *The American economic review* **70**(3), 312–326.
- Dhingra, S. and Morrow, J. (Forthcoming), ‘Monopolistic competition and optimum product diversity under firm heterogeneity’, *Journal of Political Economy*.
- Dias, D. A., Marques, C. R. and Richmond, C. (2016), ‘Misallocation and productivity in the lead up to the eurozone crisis’, *Journal of Macroeconomics* **49**, 46–70.
- Dijkstra, E. W. (1959), ‘A note on two problems in connexion with graphs’, *Numerische mathematik* **1**(1), 269–271.
- Dixit, A. K. and Stiglitz, J. E. (1977), ‘Monopolistic competition and optimum product diversity’, *American Economic Review* **67**(3), 297–308.
- Donaldson, D. (2018), ‘Railroads of the raj: Estimating the impact of transportation infrastructure’, *American Economic Review* **108**(4-5), 899–934.
- Donaldson, D. and Hornbeck, R. (2016), ‘Railroads and american economic growth: A “market access” approach’, *The Quarterly Journal of Economics* **131**(2), 799–858.
- Doraszelski, U. and Jaumandreu, J. (2013), ‘R&d and productivity: Estimating endogenous productivity.’, *Review of Economic Studies* **80**(4).
- Dougherty, S. M. (2009), ‘Labour regulation and employment dynamics at the state level in india’, *Review of Market Integration* **1**(3), 295–337.
- Dubé, J.-P. H., Fox, J. T. and Su, C.-L. (2012), ‘Improving the numerical performance of blp static and dynamic discrete choice random coefficients demand estimation’, *Econometrica* **80**(5), 2231–2267.
- Duranton, G., Ghani, E., Grover, A. and Kerr, W. (2015), ‘The misallocation of land and other factors of production in india’, *World Bank Policy Research Working Paper* (7221).
- Eberhardt, M., Helmers, C. et al. (2010), ‘Untested assumptions and data slicing: A critical review of firm-level production function estimators’, *Department of Economics, University of Oxford, Working paper series*.
- EC (2016), in European Commission, Joint Research Centre (JRC)/PBL Netherlands Environmental Assessment Agency, ed., ‘Emission Database for Global Atmospheric Research (EDGAR)’, release version 4.3.2. <http://edgar.jrc.ec.europa.eu>.
- Edmond, C., Midrigan, V. and Xu, D. Y. (2015), ‘Competition, markups, and the gains from international trade’, *The American Economic Review* **105**(10), 3183–3221.
- Edmond, C., Midrigan, V. and Xu, D. Y. (2018), ‘How costly are markups?’, *NBER working paper*.
- EIA (2015), *International energy statistics*, EIA.
- Ericson, R. and Pakes, A. (1995), ‘Markov-perfect industry dynamics: A framework for empirical work’, *The Review of Economic Studies* **62**(1), 53–82.
- Eslava, M. and Haltiwanger, J. (2019), ‘The life-cycle growth of plants: The role of productivity, demand and distortions’, *Mimeo*.
- European Commission (2013), *Sectoral Resource Maps*, European Commission, DG Environment.
- EY (2013), *Movement of goods in India*, Ernst and Young.
- EY (2014), *Indian steel. Strategy to ambition*, Ernst and Young.
- Faber, B. (2014), ‘Trade integration, market size, and industrialization: evidence from china’s national trunk highway system’, *Review of Economic Studies* **81**(3), 1046–1070.
- Faccio, M. (2006), ‘Politically connected firms’, *The American economic review* **96**(1), 369–386.
- Fackler, P. L. and Goodwin, B. K. (2001), ‘Spatial price analysis’, *Handbook of agricultural economics* **1**, 971–1024.
- Fajgelbaum, P. D. and Schaal, E. (2017), ‘Optimal transport networks in spatial equilibrium’, *NBER working paper*.
- Firth, J. (2017), ‘I’ve been waiting on the railroad: The effects of congestion on firm production’, *MIT mimeo*.
- Fogel, R. W. (1964), *Railroads and American economic growth*, Johns Hopkins Press Baltimore.
- Forlani, E., Martin, R., Mion, G. and Muûls, M. (2016), ‘Unraveling firms: demand, productivity and markups heterogeneity’.
- Foster, L., Haltiwanger, J. C. and Krizan, C. J. (2001), Aggregate productivity growth. lessons from microeconomic evidence, in ‘New developments in productivity analysis’, University of Chicago Press, pp. 303–372.

- Foster, L., Haltiwanger, J. and Syverson, C. (2008), ‘Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?’, *American Economic Review* **98**(1), 394–425.
- Fredman, M. L. and Tarjan, R. E. (1987), ‘Fibonacci heaps and their uses in improved network optimization algorithms’, *Journal of the ACM (JACM)* **34**(3), 596–615.
- Gandhi, A. and Houde, J.-F. (2016), ‘Measuring substitution patterns in differentiated products industries’, *mimeo*.
- Gandhi, A., Navarro, S. and Rivers, D. (2016), ‘On the identification of production functions: How heterogeneous is productivity?’, *Working paper*.
- Gandhi, A., Navarro, S., Rivers, D. et al. (2017), ‘How heterogeneous is productivity? a comparison of gross output and value added’, *Mimeo*.
- Garicano, L., LeLarge, C. and Van Reenen, J. (2016), ‘Firm size distortions and the productivity distribution: Evidence from france’, *American Economic Review* **106**(11), 3439–79.
- Goldberg, P. (1995), ‘Product differentiation and oligopoly in international markets: The case of the us automobile industry’, *Econometrica* **63**(4), 891–951.
- Gopinath, G., Kalemli-Özcan, Ş., Karabarbounis, L. and Villegas-Sanchez, C. (2017), ‘Capital allocation and productivity in south europe’, *The Quarterly Journal of Economics* **132**(4), 1915–1967.
- Griliches, Z. and Mairesse, J. (1999), ‘Production functions: The search for identification’, in S. Strøm, ed., ‘Econometrics and Economic Theory in the 20th Century’, Cambridge University Press, pp. 169–203.
- Griliches, Z. and Regev, H. (1995), ‘Firm productivity in israeli industry 1979–1988’, *Journal of econometrics* **65**(1), 175–203.
- Griliches, Z. and Ringstad, V. (1971), *Economies of Scale and the Form of the Production Function*, North-Holland Publishing Company.
- Hall, R. E. (1986), ‘Market structure and macroeconomic fluctuations’, *Brookings papers on economic activity* **1986**(2), 285–338.
- Hall, R. E. (1988), ‘The relation between price and marginal costs in us industry’, *Journal of Political Economy* **96**(5).
- Haltiwanger, J., Kulick, R. and Syverson, C. (2018), ‘Misallocation measures: The distortion that ate the residual’, *NBER working paper* (24199).
- Harberger, A. C. (1954), ‘Monopoly and resource allocation’, *The American Economic Review* **44**(2), 77–87.
- Harris, C. D. (1954), ‘The market as a factor in the localization of industry in the united states’, *Annals of the association of American geographers* **44**(4), 315–348.
- Harrison, A. E., Martin, L. A. and Nataraj, S. (2013), ‘Learning versus stealing: how important are market-share reallocations to india’s productivity growth?’, *The World Bank Economic Review* **27**(2), 202–228.
- Hasan, R. and Jandoc, K. R. (2014), ‘Labor regulations and firm size distribution in indian manufacturing’, in J. Bhagwati and A. Panagariya, eds, ‘Reforms and Economic Transformation in India’, Oxford University Press, pp. 15–48.
- Hausman, J., Leonard, G., Zona, J. D. et al. (1994), ‘Competitive analysis with differentiated products’, *Annals of Economics and Statistics* (34), 143–157.
- Ho, S.-J. and Ruzic, D. (2017), ‘Returns to scale, productivity measurement, and trends in us manufacturing misallocation’, *Mimeo*.
- Hopenhayn, H. A. (1992), ‘Entry, exit, and firm dynamics in long run equilibrium’, *Econometrica* **60**(5), 1127–50.
- Hopenhayn, H. A. (2014a), ‘Firms, misallocation, and aggregate productivity: A review’, *Annu. Rev. Econ.* **6**(1), 735–770.
- Hopenhayn, H. A. (2014b), ‘On the measure of distortions’, *NBER working paper* (20404).
- Hopenhayn, H. A. and Rogerson, R. (1993), ‘Job turnover and policy evaluation: A general equilibrium analysis’, *Journal of Political Economy* **101**(5), 915–38.
- Hornbeck, R. and Rotemberg, M. (2019), ‘Engines of productivity growth: Railroads, reallocation, and the rise of american manufacturing’, *Presentation (at LSE)*.
- Hottman, C. J., Redding, S. J. and Weinstein, D. E. (2016), ‘Quantifying the sources of firm heterogeneity’, *The Quarterly Journal of Economics* **131**(3), 1291–1364.
- Hsieh, C.-T., Hurst, E., Jones, C. I. and Klenow, P. J. (Forthcoming), ‘The allocation of talent and us economic growth’, *Econometrica*.
- Hsieh, C.-T. and Klenow, P. J. (2009), ‘Misallocation and manufacturing tfp in china and india’, *The Quarterly Journal of Economics* **124**(4), 1403–1448.
- Hsieh, C.-T. and Klenow, P. J. (2014), ‘The life cycle of plants in india and mexico’, *The Quarterly Journal of Economics* **129**(3), 1035–1084.
- Hsieh, C.-T. and Moretti, E. (2019), ‘Housing constraints and spatial misallocation’, *American Economic Journal: Macroeconomics* **11**(2), 1–39.
- Huang, Y. and Xiong, W. (2018), ‘Geographic distribution of firm productivity and production: A “market access” approach’, *Mimeo Harvard University*.
- IPCC (2007), in B. Metz, O. Davidson, P. Bosch, R. Dave and L. Meyer, eds, ‘Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change’, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

- Jevons, W. (1865), The coal question: can britain survive?, in A. Flux, ed., ‘The Coal Question: An Inquiry Concerning the Progress of the Nation, and the Probable Exhaustion of Our Coal-mines.’, Augustus M. Kelley, New York.
- Jones, C. (2013), Misallocation, economic growth, and input-output economics misallocation, economic growth, and input-output economics, in D. Acemoglu, M. Arellano and E. Dekel, eds, ‘Advances in Economics and Econometrics: Tenth World Congress’, Vol. 2, Cambridge University Press, p. 419.
- Jones, C. I. (2011), ‘Intermediate goods and weak links in the theory of economic development’, *American Economic Journal: Macroeconomics* **3**(2), 1–28.
- Katayama, H., Lu, S. and Tybout, J. R. (2009), ‘Firm-level productivity studies: illusions and a solution’, *International Journal of Industrial Organization* **27**(3), 403–413.
- Kimball, M. (1995), ‘The quantitative analytics of the basic neomonetarist model’, *Journal of Money, Credit and Banking* **27**(4), 1241–77.
- Klenow, P. J. and Willis, J. L. (2016), ‘Real rigidities and nominal price changes’, *Economica* **83**(331), 443–472.
- Kmenta, J. (1967), ‘On estimation of the ces production function’, *International Economic Review* **8**(2), 180–189.
- Kothari, S. (2014), ‘The size distribution of manufacturing plants and development’, *IMF Working Paper* **14**(236).
- Kugler, M. and Verhoogen, E. (2012), ‘Prices, plant size, and product quality’, *Review of Economic Studies* **79**(1), 307–339.
- Lenzu, S. and Manaresi, F. (2018), ‘Do marginal products differ from user costs? micro-level evidence from italian firms’, *Mimeo*.
- Levene, H. (1960), Robust tests for equality of variances, in I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow and H. B. Mann, eds, ‘Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling’, Stanford University Press, pp. 278–292.
- Levinsohn, J. and Petrin, A. (2003), ‘Estimating production functions using inputs to control for unobservables’, *The Review of Economic Studies* **70**(2), 317–341.
- Liang, Y. (2017), ‘Imperfect competition and misallocations’, *mimeo*.
- Marin, A. G. and Voigtländer, N. (Forthcoming), ‘Exporting and plant-level efficiency gains: It’s in the measure’, *Journal of Political Economy*.
- Marschak, J. and Andrews, W. H. (1944), ‘Random simultaneous equations and the theory of production’, *Econometrica, Journal of the Econometric Society* **12**(3/4), 143–205.
- Martin, L. A., Nataraj, S. and Harrison, A. E. (2017), ‘In with the big, out with the small: Removing small-scale reservations in india’, *American Economic Review* **107**(2), 354.
- Mayer, T., Melitz, M. J. and Ottaviano, G. I. (2014), ‘Market size, competition, and the product mix of exporters’, *The American Economic Review* **104**(2), 495–536.
- Melitz, M. J. (2003), ‘The impact of trade on intra-industry reallocations and aggregate industry productivity’, *Econometrica* **71**(6), 1695–1725.
- Midrigan, V. and Xu, D. Y. (2014), ‘Finance and misallocation: Evidence from plant-level data’, *American Economic Review* **104**(2), 422–458.
- Ministry of Road Transport and Highways (2016), *Basic Road Statistics of India*.
- Morlacco, M. (2019), ‘Market power in input markets: Theory and evidence from french manufacturing’, *mimeo*.
- Munshi, K. and Rosenzweig, M. (2016), ‘Networks and misallocation: Insurance, migration, and the rural-urban wage gap’, *The American Economic Review* **106**(1), 46–98.
- NCAER (2015), *The Indian Steel Industry: Key Reforms for a Brighter Future*, The National Council of Applied Economic Research.
- Nevo, A. (2000a), ‘Mergers with differentiated products: The case of the ready-to-eat cereal industry’, *RAND Journal of Economics* **31**(3), 395–421.
- Nevo, A. (2000b), ‘A practitioner’s guide to estimation of random-coefficients logit models of demand’, *Journal of Economics & Management Strategy* **9**(4), 513–548.
- Nevo, A. (2001), ‘Measuring market power in the ready-to-eat cereal industry’, *Econometrica* **69**(2), 307–342.
- Nishida, M., Petrin, A. and Polanec, S. (2014), ‘Exploring reallocation’s apparent weak contribution to growth’, *Journal of Productivity Analysis* **42**(2), 187–210.
- Nishida, M., Petrin, A., Rotemberg, M. and White, T. K. (2015), ‘Are we undercounting reallocation’s contribution to growth?’, *Harvard working paper*.
- Nocedal, J. and Wright, S. J. (2006), *Sequential quadratic programming*, Springer.
- NTDPC (2014), *India Transport Report: Moving India to 2032*, National Transport Development Policy Committee. Routledge.
- OECD (2015), *Material Resources, Productivity and the Environment*, OECD Green Growth Studies, OECD Publishing, Paris.
- Office of the Economic Adviser (2019), *Index Files for WPI Series, various years*, Ministry of Commerce and Industry, Government of India.
- Olley, G. S. and Pakes, A. (1996), ‘The dynamics of productivity in the telecommunications equipment industry’, *Econometrica* **64**(6), 1263–1297.
- OSMF (2016), ‘Openstreetmap’.
URL: <https://www.openstreetmap.org>

- Pakes, A. (1996), Dynamic structural models, problems and prospects: mixed continuous discrete controls and market interaction, in 'Advances in Econometrics, Sixth World Congress', Vol. 2, C. Sims, pp. 171–259.
- Peters, M. (2013), 'Heterogeneous mark-ups, growth and endogenous misallocation', *Working paper. The London School of Economics and Political Science*.
- Petrin, A. (2002), 'Quantifying the benefits of new products: The case of the minivan', *Journal of Political Economy* **110**(4), 705–729.
- Petrin, A. and Levinsohn, J. (2012), 'Measuring aggregate productivity growth using plant-level data', *The RAND Journal of Economics* **43**(4), 705–725.
- Pozzi, A. and Schivardi, F. (2016), 'Demand or productivity: what determines firm growth?', *The RAND Journal of Economics* **47**(3), 608–630.
- Raghuram, G. (2004), 'Integrating coastal shipping with the national transport network in india', *The International Association of Maritime Economists: Annual Conference 2004* pp. 1389–99.
- Rangaraj, N. and Raghuram, G. (2007), 'Viability of inland water transport in india', *Asian Development Bank Policy Brief*.
- Redding, S. J. (2010), 'The empirics of new economic geography', *Journal of Regional Science* **50**(1), 297–311.
- Redding, S. J. and Rossi-Hansberg, E. (2017), 'Quantitative spatial economics', *Annual Review of Economics* **9**, 21–58.
- Redding, S. and Venables, A. J. (2004), 'Economic geography and international inequality', *Journal of international Economics* **62**(1), 53–82.
- Restuccia, D. and Rogerson, R. (2008), 'Policy distortions and aggregate productivity with heterogeneous establishments', *Review of Economic Dynamics* **11**(4), 707–720.
- Restuccia, D. and Rogerson, R. (2013), 'Misallocation and productivity', *Review of Economic Dynamics* **16**(1), 1–10.
- Restuccia, D. and Rogerson, R. (2017), 'The causes and costs of misallocation', *Journal of Economic Perspectives* **31**(3), 151–74.
- Restuccia, D. and Santaaulalia-Llopis, R. (2017), 'Land misallocation and productivity', *NBER working paper*.
- Reynaert, M. and Verboven, F. (2014), 'Improving the performance of random coefficients demand models: the role of optimal instruments', *Journal of Econometrics* **179**(1), 83–98.
- Roberts, M., Deichmann, U., Fingleton, B. and Shi, T. (2012), 'Evaluating china's road to prosperity: A new economic geography approach', *Regional Science and Urban Economics* **42**(4), 580–594.
- Roeger, W. (1995), 'Can imperfect competition explain the difference between primal and dual productivity measures? estimates for us manufacturing', *Journal of political Economy* **103**(2), 316–330.
- Rotemberg, M. (2014), Equilibrium effects of firm subsidies, Technical report, Mimeo, Harvard University.
- Rotemberg, M. and White, T. K. (2017), 'Measuring cross-country differences in misallocation', *mimeo*.
- Schmalensee, R. (1976), 'Another look at the social valuation of input price changes', *The American Economic Review* **66**(1), 239–243.
- Shenoy, A. (2015), 'Market failures and misallocation', *Mimeo, UC Santa Cruz*.
- Shenoy, A. (2016), 'Estimating the production function when firms are constrained', *Mimeo, UC Santa Cruz*.
- Small, K. and Rosen, H. (1981), 'Applied welfare economics with discrete choice models', *Econometrica* **49**(1), 105–30.
- Startz, M. (2018), 'The value of face-to-face: Search and contracting problems in nigerian trade', *Mimeo*.
- Stone, R. (1954), 'Linear expenditure systems and demand analysis: an application to the pattern of british demand', *The Economic Journal* **64**(255), 511–527.
- Syversen, C. (2004a), 'Market structure and productivity: A concrete example', *Journal of Political Economy* **112**(6), 1181–1222.
- Syversen, C. (2004b), 'Product substitutability and productivity dispersion', *Review of Economics and Statistics* **86**(2), 534–550.
- Syversen, C. (2011), 'What determines productivity?', *Journal of Economic Literature* **49**(2), 326–365.
- Tortarolo, D. and Zarate, R. D. (2018), 'Measuring imperfect competition in product and labor markets. an empirical analysis using firm-level production data', *CAF – Working paper N° 2018/03*.
- UNFCCC (2016), *Greenhouse Gas Inventory Data*, UNFCCC.
- UNIDO (2016), 'Industrial statistics database', *Indstat4: Industrial Statistics Database*.
- Van Leemput, E. (2016), 'A passage to india: Quantifying internal and external barriers to trade', *Mimeo*.
- Vries, G. J. (2014), 'Productivity in a distorted market: the case of brazil's retail sector', *Review of Income and Wealth* **60**(3), 499–524.
- Walker, W. R. (2013), 'The transitional costs of sectoral reallocation: Evidence from the clean air act and the workforce', *The Quarterly journal of economics* **128**(4), 1787–1835.
- Wooldridge, J. M. (2009), 'On estimating firm-level production functions using proxy variables to control for unobservables', *Economics Letters* **104**(3), 112–114.
- World Bank (2005a), *Enterprise Survey India*.
- World Bank (2005b), *India: Road Transport Service Efficiency Study.*, World Bank, Energy and Infrastructure Operations Division.

World Coal Association (2014), *Coal and Steel Statistics*.

WSA (2010), *Steel Statistical Yearbook*, World Steel Association.

WSA (2016a), *Steel and coal factsheet*, World Steel Association.

WSA (2016b), *Steel Statistical Yearbook*, World Steel Association.

WSA (2017), *World Steel Association FAQ*, Vol. retrieved from <http://www.worldsteel.org/faq>, World Steel Association.

A Further underlying and related literature

This paper relates to five different types of literature. The theoretical and empirical misallocation literature in economic growth and development, the production and demand estimation literature from industrial organisation, a more policy oriented environmental material efficiency literature, the literature on Indian economic development of manufacturing industries and the literature on market and supplier access.

In their surveys, Restuccia and Rogerson (2013, 2017) categorise the recent input misallocation literature into indirect and direct approaches.¹⁰⁶ Indirect approaches use wedges that capture a bundle of distortions, and typically aim to answer the question of how severe misallocation is (using data) or could be (using simulation). Direct approaches typically model or evaluate a particular distortion, often borrowing constraints, to analyse a particular cause of misallocation. This paper contributes to both types of literature.

Two of the most influential papers in this literature have been the theoretical analysis by Restuccia and Rogerson (2008), and Hsieh and Klenow (2009) using micro data. Restuccia and Rogerson (2017) argue that the literature attributes a significant role to misallocation from a variety of sources. Much of the focus in the literature is on misallocation in a static sense, but a few papers, such as Peters (2013); Da-Rocha et al. (2017) also emphasise the dynamic consequences of such static misallocation. For more detailed surveys of this literature, the reader is referred to the dedicated surveys by Syverson (2011); Hopenhayn (2014a); Restuccia and Rogerson (2013, 2017). There is a literature that decompose aggregate productivity changes into within-firm and across-firm components based on Olley and Pakes (1996) or Petrin and Levinsohn (2012), for example.¹⁰⁷ The nuanced difference in this literature is that these approaches only identify realised reallocation gains over the years, but not the level of misallocation compared to an optimum.

Second, this paper is related to an empirical industrial organisation literature to identify parameters while distinguishing different margins of heterogeneity. On the side of estimating production functions, this paper follows the control function approach.¹⁰⁸ The literature on this topic dates back to Marschak and Andrews (1944) and is summarised in Griliches and

¹⁰⁶One of the earliest articles on resource misallocation dates back to the study of monopoly power in the US by Harberger (1954). The more recent literature is based on new trade theory models with an emphasis on heterogeneous firms. Most papers analyse the manufacturing sector, some the agricultural sector (e.g. Adamopoulos and Restuccia, 2014b; Restuccia and Santaaulalia-Llopis, 2017; Adamopoulos et al., 2017), and few the service/retail sector (e.g. Vries, 2014).

¹⁰⁷See also decompositions by Baily et al. (1992), Foster et al. (2001), Griliches and Regev (1995) or Baqaee and Farhi (2017).

¹⁰⁸This approach has been introduced by Olley and Pakes (1996) and further developed by Levinsohn and Petrin (2003); Wooldridge (2009); Akerberg et al. (2015). See also De Loecker et al. (2016) for a recent implementation with some innovations and Gandhi et al. (2016) or Forlani et al. (2016) for some criticism and alternative suggestions.

Mairesse (1999) or Eberhardt et al. (2010). On the demand side, I implement a discrete choice random utility mixed model approach of Berry et al. (1995), which is based on the characteristics of products in order to address the representative consumer restriction and dimensionality problem of more traditional demand systems (e.g. AIDS by Deaton and Muellbauer (1980)), and allows for more realistic cross elasticities than more basic random utility logit models. For a survey, see e.g. Akerberg et al. (2007).¹⁰⁹ Combining these two approaches on the production and demand side are novel to the misallocation literature.¹¹⁰ This paper is also related to price cost markup estimation. While there is literature to estimate markups from the demand side¹¹¹ and the production side¹¹², this paper combines both.¹¹³ I take the estimated markups from the demand system and use the identifying equation from the production side to disentangle the markups from input distortions that drive input misallocation.

This paper is also relevant for the literature on material efficiency. In the policy sphere, there has been growing attention for sustainable material use due to environmental and economic considerations, see e.g. OECD (2015); European Commission (2013) or the creation of the dedicated Indian Resource Panel in late 2015. While emphasis is often on within-firm innovation, there is little evidence on whether across-firm misallocation could complement efforts in this respect. Baptist and Hepburn (2013) offer descriptive evidence that higher intermediate input intensity is correlated with lower TFP, which could point towards misallocation. This paper is a first rigorous analysis of the impact of misallocation on aggregate material resource efficiency.

Fourth, this paper contributes to the literature on misallocation in manufacturing sectors in India. Bollard et al. (2013) and Harrison et al. (2013) find little *changes* in misallocation in India's manufacturing sector over time. This is a counterintuitive result as many economists and policy makers thought that the Indian reforms would impact allocative efficiency substantially. However, Nishida et al. (2014) and Nishida et al. (2015) show that these previous approach may be misleading and the method based on Petrin and Levinsohn (2012) yields opposite results. All these studies also use the Indian Annual Survey of Industries data.

¹⁰⁹Adao et al. (2017) for example apply a Berry et al. (1995) inspired demand system to a gravity trade model to depart from CES.

¹¹⁰Hsieh and Klenow (2009) for example assume values for the parameters on a demand side CES model, and take production side parameters from US ratios of aggregate data.

¹¹¹See e.g. Stone (1954), Deaton and Muellbauer (1980), Goldberg (1995) and Berry et al. (1999).

¹¹²See e.g. Hall (1986, 1988), Roeger (1995) or De Loecker and Warzynski (2012)

¹¹³Forlani et al. (2016) unravel productivity and markup variation, but not input distortions. Pozzi and Schivardi (2016) or De Loecker and Scott (2016) also estimate supply and demand parameters, but no input distortions. In the bounding exercise of David and Venkateswaran (2017) to separate capital distortions from markups, the upper bounds ignore materials input distortions, as they are assumed to be absent for estimation.

Finally, this paper is relevant to the new economic geography literature on market and supplier access. Apart from the main literature cited in the introduction and main body, there is a related literature estimates the intra-national costs of trade from price differences instead of using fastest path algorithms (Fackler and Goodwin, 2001; Anderson and Van Wincoop, 2004). Donaldson (2018) infers trade costs from price differences of single origin goods and finds that railroads decreased trade costs in colonial India. Atkin and Donaldson (2015) use price differentials at the barcode level for Ethiopia and Nigeria. Importantly, they adjust for markups which would otherwise distort trade costs estimates when using prices. Asturias et al. (2018) examine the impact of road construction on allocative efficiency in India using price gaps. The misallocation in their model comes from dispersion of markups. In contrast, this paper accounts for variable endogenous markups as fundamental differences in demand, and misallocation stems from distortions on the input side. I provide some evidence that the findings are robust to controlling for proxies for monopsony power as well.

B Input distortions and monopsony power

If inputs are not elastically supplied, i.e. plants are not price takers and have some monopsony power, the cost minimisation problem changes to:

$$\begin{aligned} \min_{L_{jt}, M_{jt}} \quad & r_{jt}K_{jt} + \tau_{jt}^{L_{adj}} w_{jt}(L_{jt})L_{jt} + \tau_{jt}^{M_{adj}} P_{jt}^M(M_{jt})M_{jt} \\ \text{s.t.} \quad & F(K_{jt}, L_{jt}, M_{jt})\Omega_{jt} \geq Q_{jt} \end{aligned}$$

where $\tau_{jt}^{M_{adj}}$ is the new input distortion adjusted for monopsony power, and the input prices are some functions of the input quantities. The first order condition with respect to materials (and the analogue can be derived for labour) is:

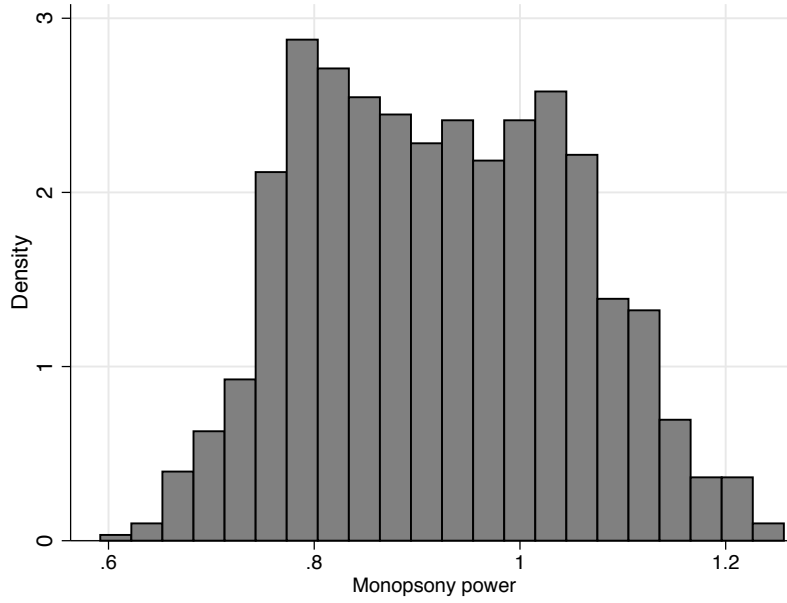
$$\underbrace{\tau_{jt}^{M_{adj}}(\psi_{jt} + 1)}_{\equiv \tau_{jt}^M} = (\eta_{jt} + 1)\alpha_{jt}^M \frac{P_{jt}Q_{jt}}{P_{jt}^M M_{jt}} \quad (12)$$

where $\psi_{jt} \equiv \frac{\partial P_{jt}^M M_{jt}}{\partial M_{jt} P_{jt}^M}$ is the inverse input price elasticity of input demand. Note that if we ignore $\tau_{jt}^{M_{adj}}$, and use that $MRPM \equiv (\eta_{jt} + 1)\alpha_{jt}^M \frac{P_{jt}Q_{jt}}{P_{jt}^M M_{jt}}$, we can write $(\psi_{jt} + 1) = \frac{MRPM}{P_{jt}^M}$. That is $(\psi_{jt} + 1)$ is the ability to pay an input a lower price than its marginal revenue product, a common definition of market power on the input side, or monopsony power. The measured input distortion τ_{jt}^M captures input market power as well as other input distortions $\tau_{jt}^{M_{adj}}$.¹¹⁴

How likely is it that the measured input distortion τ_{jt}^M represent monopsony power in

¹¹⁴See also Morlacco (2019), who studies monopsony power based on similar first order conditions.

Figure 15: Histogram of estimated input market power ($\psi_{jt} + 1$)



Notes: The figure plots the histogram of monopsony power ($\psi_{jt} + 1$). A ratio larger than one suggest that the input price is below the marginal revenue product for that input.

this case? First, if plants that are close to each other and operate in the same input market have similar monopsony power, and τ_{jt}^M is primarily driven by monopsony power, then the τ_{jt}^M should vary more across states than within states. Using a variance decomposition as in Davis et al. (2013), I find that on average, the variance between states is 31% and 39% of the total variance for the logged τ_{jt}^L and τ_{jt}^M respectively, so most of the variance is within states. Second, for material inputs, monopsony power is likely higher if there are many upstream suppliers and few cast iron plants using that particular input. One of the main inputs is pig iron. We can compare the number of plants using pig iron as an input and the number of plants producing pig iron in the raw data. The number of plants producing pig iron is between 24 and 60 throughout the years. The number of plants using pig iron is between 280 and 800 throughout the years. It is unlikely that the comparatively large number of smaller plants can exert input market power over the few large plants. Third if monopsony power is plant specific and applies to both inputs, labour and materials, then the τ_{jt}^L and τ_{jt}^M should be correlated. Figure 28 in Appendix G.7 shows that they are not correlated, however.

Fourth, I construct two, admittedly heuristic, plant specific proxies for monopsony power. If relatively larger plants can exert more market power on the input side as well, a larger market share of a plant in a given state can proxy for monopsony power. The correlations between the market share within a state and the material or labour distortions are small (<0.02) and statistically insignificant. The second proxy for monopsony power is based on a direct estimate of $\frac{\partial P_{jt}^M M_{jt}}{\partial M_{jt} P_{jt}^M}$. I recover those heuristically by regressing logged input prices on a

second order polynomial in input quantities controlling for district and time fixed effects:

$$\log(P_{jt}^M) = f(\log(M_{jt})) + \boldsymbol{\lambda}_d + \boldsymbol{\kappa}_t + v_{jt} \quad (13)$$

Based on the coefficients, I can compute ψ_{jt} . I plot the histogram of input market power ($\psi_{jt} + 1$) in Figure 15. The majority of plants has a negative elasticity ψ_{jt} and therefore an input market power ($\psi_{jt} + 1$) below one. A negative ψ_{jt} means that input prices decrease for larger quantities, which can be related to quantity discounts instead of input market power. The correlation between $\log(\psi_{jt} + 1)$ and $\log(\tau_{jt}^L)$ and $\log(\tau_{jt}^M)$ is (-0.12 and -0.19) respectively, which is inconsistent with distortions capturing monopsony power.

C Estimation details for the production side

C.1 Control function approach to production side estimation

C.1.1 The production side identification problem

The fundamental problem is that we do not observe total factor productivity ω_{jt} , which is likely correlated with inputs and causes endogeneity problems.¹¹⁵ Furthermore, we need to pin down plant total factor productivities for the counterfactual analysis.

C.1.2 Unexpected output shock

In order to avoid adding an ad-hoc error term just for the estimation, I incorporate an additional error term ϵ into the entire structural model, so that it is consistent with firm behaviour and the Nash competition framework throughout, as detailed below. Splitting up the combined error term into a so-called transmitted (to inputs) component ω_{jt} and untransmitted component ϵ_{jt} is common in the productivity literature (Griliches and Mairesse, 1999). The way we can interpret this in the context of the conduct model in this paper is that the equilibrium prices and output are treated as expected prices and output. Firms maximise profits by choosing the expected prices in line with Bertrand-Nash competition. They base their production input decision on achieving the desired expected output by minimising costs. During or after production, an unanticipated multiplicative shock to expected firm output

¹¹⁵In traditional production function estimation total factor productivity ω_{jt} has often been treated as regression error term. This has been recognised as problematic for a long time (Marschak and Andrews, 1944), as it is very likely correlated with the input choices. Researchers often resorted to using an index number approach, essentially retrieving the output elasticities from the mean or median of the first order condition (3). If we assume that on the mean or median, the associated τ and $(\eta + 1)$ are unity, then we could use this approach, at least for constant production elasticities. In the estimation strategy I use, the mean or median τ and $(\eta + 1)$ vary by year and are not always close to unity, which would bias the index number estimates.

occurs ($\exp(\epsilon_{jt})$) and defines realised, observed output Q_{jt}^r :

$$Q_{jt}^r = Q_{jt} \exp(\epsilon_{jt}) \quad (14)$$

I assume that the input decisions have been made by the time this shock materialises and that this shock is entirely unpredictable by the firm. It could likewise also be interpreted as measurement error in the output variable. The firm productivity ω_{jt} on the other hand fully enters into the decision of the input variables. The realised, observed plant output in logs is therefore:

$$q_{jt}^r = f(k_{jt}, l_{jt}, m_{jt}) + \omega_{jt} + \epsilon_{jt} \quad (15)$$

This is the basic equation I want to estimate but in order to implement it I need to make further identifying and functional assumptions.

C.1.3 Functional form assumption

For the baseline estimation and counterfactual analysis, I follow the key literature and assume a Cobb-Douglas production function:

$$q_{jt}^r = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \beta_m m_{jt} + \omega_{jt} + \epsilon_{jt} \quad (16)$$

With Cobb-Douglas, as opposed to a translog production function for example, we can derive a closed form analytical solution for the conditional input demand functions which dramatically eases the search for equilibria. However, in Appendix C.3, I make a second order Taylor approximation around the unknown production function which results in a translog specification and can be viewed as a generalised approximation to a CES production function. The production estimates from this more flexible translog approximation are reported in the results as well, and are on average reassuringly close to the Cobb-Douglas estimates.¹¹⁶

C.1.4 Simultaneity and selection biases

The main identification challenge is that we are unlikely to get consistent estimates when running OLS on Equation (16) since we do not observe productivity ω_{jt} . While the shock ϵ_{jt} is assumed to be unexpected and unknown to the firm, and therefore uncorrelated with input choices, the productivity ω_{jt} is known to the firm, and highly likely to influence input

¹¹⁶Essentially the Cobb-Douglas is a simplification of the translog specification by enforcing some parameter restrictions.

choices. If a firm experiences a positive productivity shock, it is likely to use more variable inputs, creating a positive bias in the coefficient. This problem is commonly referred to in the productivity literature as simultaneity or transmission bias (Marschak and Andrews, 1944; Griliches and Mairesse, 1999).

A second identification issue comes from sample selection. Similar to Heckman’s selection problem, we only observe firms that are in production. Firms’ survival is positively correlated with productivity. But firms’ decisions to exit are negatively correlated with installed capital, conditional on unobserved productivity. As Ericson and Pakes (1995) argue, capital serves as buffer to shocks. Therefore, surviving firms have an expected productivity that is decreasing in installed capital. This creates a downward bias in the capital coefficients in an OLS regression omitting productivity. Moreover, since I only use single-product firms for estimating the production function, there is an additional self-selection problem of single-product firms turning multi-product which is positively related to productivity. Conditional on productivity, firms with higher installed capital (or labour) are more likely to introduce a second product, as in the model of Mayer et al. (2014). Again, “surviving” single product firms (i.e. those not turning multiproduct) have an expected productivity that is decreasing in installed capital (and perhaps labour).

Note that in the selection problem the bias arises in the more persistent variable (capital), becoming more severe with more dynamics, whereas in the simultaneity problem, the bias arises in the flexible inputs (material), becoming more severe with higher flexibility.¹¹⁷ Both can cause inconsistent and biased estimates of all coefficients, so we should address both.

C.1.5 Addressing simultaneity and selectivity

I assume that material inputs are a function of several observed variables and a scalar unobserved variable, productivity:

$$m_{jt} = m(k_{jt}, l_{jt}, \mathbf{z}_{jt}, \omega_{jt}) \quad (17)$$

where \mathbf{z}_{jt} are additional variables which I discuss below. Additional to this scalar unobservable assumption, it is assumed that this function is (conditionally) monotonically increasing in productivity ω_{jt} .¹¹⁸ Therefore, the material demand function can be inverted for productivity,

¹¹⁷This is one reason why value added production functions have traditionally often been used in the literature, to avoid highly flexible materials as input in the estimation.

¹¹⁸Only relatively mild conditions are necessary that the marginal product of materials is increasing in ω_{jt} (Levinsohn and Petrin, 2003). This is easier to prove in the case where investment acts as the proxy variable (Pakes, 1996).

which we can later flexibly include in the estimating equation:

$$\omega_{jt} = h(k_{jt}, l_{jt}, \mathbf{z}_{jt}, m_{jt}) \quad (18)$$

where $h(\cdot)$ is an unknown function, which we can approximate with polynomials or semi-parametrically. The choice of variables in \mathbf{z}_{jt} could be important, but was omitted in the pioneering applications (Olley and Pakes, 1996; Levinsohn and Petrin, 2003). Gandhi et al. (2016) argue that having no additional variables in \mathbf{z}_{jt} leaves the production function non-parametrically non-identified.¹¹⁹ The second reason why \mathbf{z}_{jt} may be important is that we need to be comfortable with the assumption that productivity is the only unobservable driving material demand, which is more likely if we control for factors such as input prices. For robustness checks, I include in \mathbf{z}_{ijt} :

$$\mathbf{z}_{jt} = (p_{jt}^M, IMP_{jt}, \eta_{jt}, p_{jt}, s_{jt}, \mathbf{G}_t)$$

that is log material input prices p_{jt}^m , import status of material goods IMP_{jt} , the inverse demand elasticity η_{jt} , log output prices p_{jt} , market share s_{jt} and firm location dummies \mathbf{G}_t .¹²⁰ The last three variables have also been used in the proxy equation in De Loecker et al. (2016). However, they have neither observed input prices nor a measure of the demand elasticity. As Forlani et al. (2016) argue, variations in demand elasticity or market power are likely to drive material demand and thus are important to include. Note that the framework of De Loecker and Warzynski (2012) retrieves markups after production side estimation and can therefore not include markups in the estimation. The advantage of the approach in this paper is that we can recover demand elasticities from independent demand side estimations.

A central concern is that the misallocation wedges τ_{jt} are likely to influence material demand. Since they are identified from the input prices, demand elasticity, output prices and market shares (Equation 3)), it is sensible to include them in \mathbf{z}_{jt} . The monotonicity assumption needs to hold only conditional on \mathbf{z}_{jt} . Figure 21 plots the relationship between material inputs and productivity and shows a monotonic relationship. For concerns that the τ_{jt} still violate the scalar unobservable assumption, I furthermore implement a Blundell-Bond system-GMM estimator for the production function as a robustness check, as detailed further below.

In practice, I address the self-selection problem of firms from being single-product into exit and into multi-product firms by following and augmenting the strategies of Olley and

¹¹⁹As otherwise only the shock in productivity from Equation 20 identifies it, which is unobserved and later assumed to be orthogonal to all input choice lags.

¹²⁰As indicated in the results, I use a full \mathbf{z}_{jt} for robustness checks, but not for the baseline results

Pakes (1996) and De Loecker et al. (2016). Essentially, I estimate the probability of being in the sample $Prob_{jt}$ with a discrete model, and the predicted probability from this estimation will be included in the final estimation.¹²¹

C.1.6 Markov productivity process

A common convenient¹²² assumption in the production function and productivity literature, including the proxy approach, is that productivity follows a first order Markov process (see influential early papers of Hopenhayn (1992); Hopenhayn and Rogerson (1993)):

$$\begin{aligned}
\omega_{jt} &= E(\omega_{jt} \mid \omega_{jt-1}, IMP_{jt-1}, sp_{jt} = 1) + \zeta_{jt} \\
&= \Psi\left(\omega_{jt-1}, IMP_{jt-1}, \tilde{d}(\underline{\omega}_{jt}, \overline{\omega}_{jt})\right) + \zeta_{jt} \\
&= \Psi\left(h(k_{jt-1}, l_{jt-1}, \mathbf{z}_{jt-1}, m_{jt-1}), IMP_{jt-1}, g^{-1}(\omega_{jt-1}, Prob_{jt})\right) + \zeta_{jt} \\
&= \tilde{\Psi}\left(h(k_{jt-1}, l_{jt-1}, \mathbf{z}_{jt-1}, m_{jt-1}), IMP_{jt-1}, Prob_{jt}\right) + \zeta_{jt}
\end{aligned} \tag{20}$$

¹²¹In more detail: In Ericson and Pakes (1995), productivity follows a Markov process and the exit decision depends on a threshold value of productivity ω_{jt} . A draw below this threshold value makes it more profitable to sell the firm since its sell-off value is higher than discounted net profits based on the current productivity draw. However, the productivity threshold also depends on installed capital, and is decreasing in it since discounted profits are higher for higher capitalised firms. The sell-off value is assumed to increase less in capital than discounted profits increase in capital. In short, the firm exits if $\omega_{jt} < \omega_{jt}(k_{jt})$. In the model of Mayer et al. (2014), the number of products is increasing (as a step function) in productivity draws. The multiproduct threshold productivity $\overline{\omega}_{jt}$ is again decreasing in capital, since bigger firms are more likely to be able to set up new product lines. In short, the firm becomes multi-product if $\omega_{jt} > \overline{\omega}_{jt}(k_{jt})$. Putting these elements together, the conditional probability of being in the single-product sample, indicated by $sp_{jt} = 1$ is:

$$\begin{aligned}
&Pr[sp_{jt} = 1 \mid \underline{\omega}_{jt}(k_{jt}), \overline{\omega}_{jt}(k_{jt}), \omega_{jt-1}] \\
&= Pr\left[\underline{\omega}_{jt}(k_{jt}) < \omega_{jt} < \overline{\omega}_{jt}(k_{jt}) \mid \underline{\omega}_{jt}(k_{jt}), \overline{\omega}_{jt}(k_{jt}, l_{jt}), \omega_{jt-1}\right] \\
&= \tilde{g}\left(\tilde{d}(\underline{\omega}_{jt}(k_{jt}), \overline{\omega}_{jt}(k_{jt})), \omega_{jt-1}\right) \\
&= g\left(k_{jt-1}, i_{jt-1}, l_{jt-1}, \mathbf{z}_{jt-1}, m_{jt-1}\right) \equiv Prob_{jt}
\end{aligned} \tag{19}$$

where i_{jt-1} is investment. Conditionally on knowing the thresholds and previous period productivity, the probability that current productivity lies within the thresholds can be written as an unknown function of these elements. The reason why I use the notation of a function \tilde{d} to summarise both thresholds will become apparent below. Since capital is a function of previous period capital and previous period investment, and productivity a function of given variables from the invertibility condition, we can write the survival probability as an unknown function of these previous period variables. I estimate this probability with a discrete model, and the predicted probability $Prob_{jt}$ will be included in the final estimation. In this estimation I include whether the plant belongs to the census or the sampled sector, as this is additional critical information whether the plant is contained in the sample. I do not empirically restrict the threshold productivities to be decreasing in its arguments, but estimate the function $g(\cdot)$ flexibly.

¹²²It is convenient, as with higher order Markov processes, we need a longer history of the data and effectively lose observations.

The productivity process is not completely exogenous¹²³, but is allowed to depend on the firm's import status, because of potential international technology spillovers and depends on the firm being in the sample and single product ($sp_{jt} = 1$) which ultimately depends on its survival probability from (19). Therefore the productivity is an unknown function of the elements in the last equation and the shock to productivity ζ_{jt} .¹²⁴

The Markov assumption implies that the CDF of ω_{jt} is a decreasing function of ω_{jt-1} , i.e. that high ω_{jt-1} firms stochastically dominate low ω_{jt-1} firms. By construction:

$$E[\zeta_{jt} \mid \omega_{jt-1}, IMP_{jt-1}, \tilde{d}(\underline{\omega}_{jt}, \overline{\omega}_{jt}), sp_{jt} = 1] = 0 \quad (21)$$

Note that the Markov assumption implies that ζ_{jt} is not only uncorrelated with all lagged variables in the function $h(\cdot)$, but through the capital accumulation equation also with current capital k_{jt} , if current capital is only a function of previous period capital and previous period investment and depreciation, which have all been realised before the productivity shock ζ_{jt} is incurred, a common assumption in the literature. Similarly, if labour hiring and firing takes enough adjustment time, ζ_{jt} could also be uncorrelated with current labour inputs. I check my results for either assumption on labour timing. These orthogonality assumptions in the productivity Markov process are crucial for the identification of the production function parameters.

¹²³It can also be allowed to depend additionally on R&D expenditures as in [Doraszelski and Jaumandreu \(2013\)](#), but I have no data on this.

¹²⁴The second equality states that the conditional expectation is a function of its conditioning variables including the function of the productivity thresholds since they define the range for $sp_{jt} = 1$ if $\underline{\omega}_{jt}(k_{jt}) < \omega_{jt} < \overline{\omega}_{jt}(k_{jt}, l_{jt})$, where ω_{jt-1} enters the function $\Psi(\cdot)$ twice. For the third equality, I use the control function for productivity, and inverting the unknown function $\tilde{g}(\cdot)$ from Equation (19) to write $\tilde{d}(\underline{\omega}_{jt}(k_{jt}), \overline{\omega}_{jt}(k_{jt}, l_{jt})) = \tilde{g}^{-1}(\omega_{jt-1}, Prob_{jt})$. I assume that this inversion exists. A sufficient condition would be that there is indeed a function $\tilde{d}(\cdot)$ in which $\tilde{g}(\cdot)$ is monotonous, despite $\tilde{g}(\cdot)$ being increasing in $\overline{\omega}_{jt}$ and decreasing $\underline{\omega}_{jt}$ individually, so e.g. the gap $\tilde{d}(\cdot) = \overline{\omega}_{jt} - \underline{\omega}_{jt}$. For more discussion on the assumptions on an inversion involving one threshold, see [Olley and Pakes \(1996\)](#). The last equation shows that I address the selection problem in the productivity process by conditioning on the probability or propensity of being in the sample, i.e. between the thresholds.

C.1.7 Estimated equations and moments

Following Wooldridge (2009), we can write down two equations for the production function, where we substitute in for ω_{it} from Equation (18) and from Equation (20):¹²⁵

$$q_{jt}^r = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \beta_m m_{jt} + h(k_{jt}, l_{jt}, \mathbf{z}_{jt}, m_{jt}) + \epsilon_{jt} \quad (22)$$

$$q_{jt}^r = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \beta_m m_{jt} + \tilde{\Psi} \left(h(k_{jt-1}, l_{jt-1}, \mathbf{z}_{jt-1}, m_{jt-1}), IMP_{jt-1}, Prob_{jt} \right) + \zeta_{jt} + \epsilon_{jt} \quad (23)$$

For estimation we need to specify the unknown functions $h(\cdot)$ and $\tilde{\Psi}(\cdot)$. For $h(\cdot)$ I use a third order polynomial with all interactions in its arguments.¹²⁶ For the Markov process in productivity I use an AR(1) process, so $\tilde{\Psi}(\cdot)$ becomes a linear function.¹²⁷

For consistent estimates, we need to specify the instrument matrix for each of the two equations, which requires assumptions on timing. For the first Equation (22), the shock to production ϵ_{jt} is unexpected and incurred during or after production and therefore not linked to current (or past) firm input choices. We can use the full set of current and past variables as instruments for themselves, which I denote as the information set $\mathbf{\Gamma}_{jt}$.

However, for the second Equation (23), the joint error term contains ζ_{jt} , which is part of ω_{jt} , which the firm is assumed to know before the beginning of production. So clearly, this is correlated with current input choices. Since ζ_{jt} is the non-anticipated innovation in the Markov productivity process, it is not correlated with past input choices, however. It depends on the assumption of the flexibility of inputs, which current input choices are problematic from an econometric point of view. If we believe that current capital is set in the last period (last period investment, depreciation and capital stock), then current period capital is not correlated with ζ_{it} . I follow the literature in assuming this. For labour, it depends how flexibly hiring and firing takes place. Most likely it is partially dynamic, so there is not complete digression on the size of the labour force each period (multi-year contracts). I allow current labour choices to be correlated with current productivity, but also check robustness with a version where it is fully dynamic, i.e. determined in the previous period. I denote the set of instruments for the second Equation (23) as the information set $\mathbf{\Gamma}_{jt-1}$, where k_{jt} is contained since it is determined in the previous period, and depending on the labour assumption, l_{jt} is contained or only l_{jt-1} .

¹²⁵See Appendix C.3 for the translog version.

¹²⁶As in De Loecker et al. (2016) for example. I also check the results with higher-order polynomials. One can alternatively use non-parametric methods as in Levinsohn and Petrin (2003), at the expense of a much more complicated estimation procedure.

¹²⁷Similar to the assumption in Forlani et al. (2016). I also check the results' robustness with higher order polynomials.

By rearranging the equations we can formulate the set of population moment equations, where the errors are a function of all parameters Θ :

$$E \left(\begin{array}{c|c} \epsilon_{jt}(\Theta) & \mathbf{\Gamma}_{jt} \\ \hline (\epsilon_{jt} + \zeta_{jt})(\Theta) & \mathbf{\Gamma}_{jt-1} \end{array} \right) = 0$$

We can form the analogous stacked sample moments and write the criterion function $\tilde{Q}(\Theta)$ to be minimised:

$$\begin{aligned} \text{Define: } \mathbf{r}_{jt}(\Theta) &\equiv \begin{pmatrix} \epsilon_{jt}(\Theta) \\ (\epsilon_{jt} + \zeta_{jt})(\Theta) \end{pmatrix} \text{ and } \tilde{\mathbf{\Gamma}}_{jt} \equiv \begin{pmatrix} \mathbf{\Gamma}_{jt} & 0 \\ 0 & \mathbf{\Gamma}_{jt-1} \end{pmatrix} \\ \text{Set of sample moment conditions: } &\frac{1}{JT} \sum_j \sum_t [\tilde{\mathbf{\Gamma}}_{jt}' \hat{\mathbf{r}}_{jt}(\Theta)] = \mathbf{0} \\ \hat{\Theta} = \min_{\Theta} \tilde{Q}(\Theta) &= \frac{1}{JT} \left[\sum_j \sum_t [\tilde{\mathbf{\Gamma}}_{jt}' \hat{\mathbf{r}}_{jt}(\Theta)]' \mathbf{W} \sum_j \sum_t [\tilde{\mathbf{\Gamma}}_{jt}' \hat{\mathbf{r}}_{jt}(\Theta)] \right] \end{aligned}$$

where the weighting matrix \mathbf{W} is clustered on plants accounting for non-identically distributed and autocorrelated errors.

C.1.8 Production elasticities

Having estimated the vector of parameters $\hat{\Theta}$, I simply use the residual of the first stage Equation (22) to get the estimate $\hat{\epsilon}_{it}$. I recover productivities $\hat{\omega}_{jt}$ by subtracting the production function with plugged in estimates $\hat{\Theta}$ from the predicted values \hat{q}_{jt} in the first stage equation. The estimate for the production elasticity of inputs α_{jt} is simply the corresponding coefficient, for example for materials:¹²⁸

$$\hat{\alpha}_{jt}^M = \hat{\beta}_m$$

C.2 An alternative for production function estimation: dynamic panel system GMM

As an alternative to the proxy approach, I also implement a quasi-differenced dynamic panel system GMM approach (Blundell and Bond, 2000). This serves as a robustness check, as the material demand equation depends on τ , which we do not observe. Everything else equal, firms with higher τ demand less materials. I include the factors that drive τ in \mathbf{z}_{jt} (such as input prices or demand elasticity), but the system GMM approach serves as a test

¹²⁸For the translog it varies by plant: $\hat{\alpha}_{jt}^M = \hat{\beta}_m + \hat{\beta}_{lm}l_{jt} + \hat{\beta}_{km}k_{jt} + \hat{\beta}_{mm}m_{ijt}$, see Appendix C.3.

whether this is enough for the scalar unobservable and invertibility condition required for the proxy approach. [Shenoy \(2015\)](#) uses a dynamic panel method in his analysis of input misallocation for Thai rice farmers due to a similar concern. He also develops a test for the scalar unobservable assumption, and argues that with input constraints, the dynamic panel approach tends to perform better in his setting ([Shenoy, 2016](#)).

I maintain the first order Markov assumption for the productivity process, which I further specify into an AR(1) process. But I allow for a firm specific time-invariant component of productivity ν_j :

$$\omega_{jt} = \rho\omega_{jt-1} + \nu_j + \zeta_{jt}$$

Analogous to Equation 20, we could condition the productivity process on sample selection and import status, or R&D expenditures. Quasi-first differencing the log production function (15) by ρ eliminates unobserved ω_{jt} :

$$q_{jt}^r = -\rho q_{jt-1}^r + f(k_{jt}, l_{jt}, m_{jt}) - \rho f(k_{jt-1}, l_{jt-1}, m_{jt-1}) + \nu_j + \zeta_{jt} + \epsilon_{jt} - \rho\epsilon_{jt-1}$$

By first differencing this equation, I can implement an [Arellano and Bond \(1991\)](#) estimator by using lagged independent variables as instrument. I use a more efficient [Blundell and Bond \(2000\)](#) system GMM estimator, where I use lagged differenced independent variables as instruments for the level equation in addition.¹²⁹ This estimator is implemented for both the Cobb-Douglas and the translog versions.

C.3 Estimation with translog production function

I also provide the production function estimates for a more flexible translog specification. To repeat, the realised, observed output in logs by firm is:

$$q_{jt}^r = f(k_{jt}, l_{jt}, m_{jt}) + \omega_{jt} + \epsilon_{jt} \tag{24}$$

Since we don't know the functional form of the production function we can form a second order Taylor approximation¹³⁰ with approximation error ν_{jt} around the point $\mathbf{X} = \mathbf{1}$ (so

¹²⁹The appropriate instruments need to take into account that q_{jt-1}^r is correlated with ϵ_{jt-1} , and l_{jt-1} and m_{jt-1} and k_{jt} with ζ_{jt-1} .

¹³⁰And Young's theorem of equal cross-partials.

$\mathbf{x} = \mathbf{0}$):

$$\begin{aligned}
q_{jt}^r = & f(\mathbf{0}) + \frac{\partial f}{\partial k_{jt}|_{\mathbf{x}=\mathbf{0}}} k_{jt} + \frac{\partial f}{\partial l_{jt}|_{\mathbf{x}=\mathbf{0}}} l_{jt} + \frac{\partial f}{\partial m_{jt}|_{\mathbf{x}=\mathbf{0}}} m_{jt} \\
& + \frac{1}{2} \frac{\partial^2 f}{(\partial k_{jt})^2}|_{\mathbf{x}=\mathbf{0}} k_{jt}^2 + \frac{1}{2} \frac{\partial^2 f}{(\partial l_{jt})^2}|_{\mathbf{x}=\mathbf{0}} l_{jt}^2 + \frac{1}{2} \frac{\partial^2 f}{(\partial m_{jt})^2}|_{\mathbf{x}=\mathbf{0}} m_{jt}^2 \\
& + \frac{\partial^2 f}{\partial k_{jt} \partial l_{jt}|_{\mathbf{x}=\mathbf{0}}} k_{jt} l_{jt} + \frac{\partial^2 f}{\partial k_{jt} \partial m_{jt}|_{\mathbf{x}=\mathbf{0}}} k_{jt} m_{jt} + \frac{\partial^2 f}{\partial l_{jt} \partial m_{jt}|_{\mathbf{x}=\mathbf{0}}} l_{jt} m_{jt} \\
& + \omega_{jt} + \epsilon_{jt} + \nu_{jt}
\end{aligned}$$

Since the derivatives evaluated at $\mathbf{x} = \mathbf{0}$ are constant across time and firms within the same product category, we can interpret them as the regression coefficients,¹³¹ which yields a second order translog¹³² production function:

$$q_{jt}^r = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \beta_m m_{jt} + \beta_{kk} k_{jt}^2 + \beta_{ll} l_{jt}^2 + \beta_{mm} m_{jt}^2 \quad (25)$$

$$+ \beta_{kl} k_{jt} l_{jt} + \beta_{km} k_{jt} m_{jt} + \beta_{lm} l_{jt} m_{jt} + \omega_{jt} + \epsilon_{jt} \quad (26)$$

In this equation, the additional error term ν_{ijt} which comes from approximation is assumed to be zero and omitted. We cannot identify this term (which can in principle be “transmitted” to inputs) separately to the shock ϵ_{jt} , so assume it is zero and the approximation is perfect. This is of course a silent assumption in all of the production estimation literature. Compared to Cobb-Douglas or CES functional form assumptions, the translog specification is more flexible allowing for significant amount of curvature, and is thus less likely to suffer from functional form assumption or approximation bias.

Analogous to the main text, we can write down two equations for the production function, where I substitute in for ω_{it} from Equation 18 and from Equation 20:

$$\begin{aligned}
q_{jt}^r = & \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \beta_m m_{jt} + \beta_{kk} k_{jt}^2 + \beta_{ll} l_{jt}^2 + \beta_{mm} m_{jt}^2 \\
& + \beta_{kl} k_{jt} l_{jt} + \beta_{km} k_{jt} m_{jt} + \beta_{lm} l_{jt} m_{jt} + h(k_{jt}, l_{jt}, \mathbf{z}_{jt}, m_{jt}) + \epsilon_{jt}
\end{aligned} \quad (27)$$

$$\begin{aligned}
q_{jt}^r = & \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \beta_m m_{jt} + \beta_{kk} k_{jt}^2 + \beta_{ll} l_{jt}^2 + \beta_{mm} m_{jt}^2 \\
& + \beta_{kl} k_{jt} l_{jt} + \beta_{km} k_{jt} m_{jt} + \beta_{lm} l_{jt} m_{jt} \\
& + \tilde{\Psi} \left(h(k_{jt-1}, l_{jt-1}, \mathbf{z}_{jt-1}, m_{jt-1}), IMP_{jt-1}, Prob_{jt} \right) + \zeta_{jt} + \epsilon_{jt}
\end{aligned} \quad (28)$$

¹³¹The factor of a half is incorporated in the coefficient for the quadratic terms.

¹³²The transcendental logarithmic function was introduced by Christensen et al. (1971, 1973); Berndt and Christensen (1973). See also Griliches and Ringstad (1971) who propose a similar generalisation of the approximation for estimating CES functions by Kmenta (1967).

The estimate for the production elasticity of material inputs is, for example:

$$\hat{\alpha}_{jt}^M = \hat{\beta}_m + \hat{\beta}_{lm}l_{jt} + \hat{\beta}_{km}k_{jt} + \hat{\beta}_{mm}m_{ijt}$$

For the Cobb-Douglas version, we apply the parameter restrictions that $\beta_{lm} = \beta_{km} = \beta_{mm} = 0$, so that $\hat{\alpha}_{jt}^M = \hat{\beta}_m$.

D Estimation details for the demand side

D.1 Estimation of the demand model

To repeat for convenience, consumers are indexed by i and need to decide to buy from a firm j to maximise their utility from using product j :

$$U_{ijt} = (y_{it} - P_{jt}^r)\theta_{it}^p + x_{jt}\theta_{it}^x + \xi_j + \xi_t + \Delta\xi_{jt} + \mu_{ijt} \equiv V_{ijt} + \mu_{ijt} \quad (29)$$

where y_{it} is consumer income, P_{jt}^r realised prices (which are associated with realised quantities – these are the ones that are relevant for the consumers), x_{jt} a vector of product characteristics and a constant, ξ_j average utility from unobserved time-constant product characteristics, ξ_t average unobserved market-specific utility, and $\Delta\xi_{jt}$ the unobserved deviations from a particular product in a particular market from the unobserved averages. The unobserved ξ_j can contain the quality and the location of a product and ξ_j and ξ_t will be absorbed by fixed effects dummies. For the baseline results I only include a constant in x_{jt} as there are few time variant product characteristics (since the time invariant characteristics are absorbed in ξ_j). The non-random utility can be summarised by V_{ijt} . The random utility component is μ_{ijt} , which follows an i.i.d. Type I extreme value distribution.

We can further specify the random parameters into a mean and variance component:

$$\begin{pmatrix} \theta_{it}^p \\ \theta_{it}^x \end{pmatrix} = \begin{pmatrix} \theta^p \\ \theta^x \end{pmatrix} + \begin{pmatrix} \sigma^p & 0 \\ 0 & \sigma^x \end{pmatrix} \begin{pmatrix} \nu_{it}^p \\ \nu_{it}^x \end{pmatrix}, \quad \boldsymbol{\nu}_{it} \sim P(\boldsymbol{\nu})$$

where ν_{it} are draws from a multivariate normal distribution. Therefore the consumer heterogeneity has three dimensions, the random utility shock μ_{ijt} as well as the two ν_{it} draws. I estimate the means $\begin{pmatrix} \theta^p \\ \theta^x \end{pmatrix}$ and the variances $\Sigma \equiv \begin{pmatrix} \sigma^p & 0 \\ 0 & \sigma^x \end{pmatrix}$ of the random coefficients. We can rewrite the utility function with a mean ($\equiv \delta_{jt}$) and individual consumer part, which

simplifies the estimation algorithm:

$$U_{ijt} = \theta_{it}^p y_{it} \underbrace{-\theta_{jt}^p P_{jt}^r + \theta_{jt}^x x_{jt} + \xi_j + \xi_t + \Delta \xi_{jt}}_{\equiv \delta_{jt}} - \sigma^p \nu_{it}^p P_{jt}^r + \sigma^x \nu_{it}^x x_{jt} + \mu_{ijt}$$

D.1.1 Derived theoretical market shares

Consumer i purchases from firm j if it yields the highest utility, compared to the products from all other firms or the outside option $j = 0$. The outside good also serves to normalise the utility by setting the mean and individual components in the outside good utility to zero.¹³³ Define as set A_{jt} the set of consumers which strictly prefer product j .¹³⁴ The integral over the consumers that belong to this set is the theoretical (realised) market share s_{jt}^r of firm j in period t :

$$\begin{aligned} s_{jt}^r &= \int_{A_{jt}} dP(\boldsymbol{\nu}, \boldsymbol{\mu}) = \int_{A_{jt}} dP(\boldsymbol{\mu} \mid \boldsymbol{\nu}) dP(\boldsymbol{\nu}) = \int_{A_{jt}} dP(\boldsymbol{\mu}) dP(\boldsymbol{\nu}) \\ &= \int_{A_{jt}} \frac{\exp(\theta_{it}^p y_{it}) \exp(\delta_{jt} - \sigma^p \nu_{it}^p P_{jt}^r + \sigma^x \nu_{it}^x x_{jt})}{\exp(\theta_{it}^p y_{it}) \left[\exp(\delta_{0t} - \sigma^p \nu_{it}^p P_{0t}^r + \sigma^x \nu_{it}^x x_{0t}) + \sum_{j=1} \exp(\delta_{jt} - \sigma^p \nu_{it}^p P_{jt}^r + \sigma^x \nu_{it}^x x_{jt}) \right]} dP(\boldsymbol{\nu}) \\ &= \int_{A_{jt}} \frac{\exp(\delta_{jt} - \sigma^p \nu_{it}^p P_{jt}^r + \sigma^x \nu_{it}^x x_{jt})}{1 + \sum_{j=1} \exp(\delta_{jt} - \sigma^p \nu_{it}^p P_{jt}^r + \sigma^x \nu_{it}^x x_{jt})} dP(\boldsymbol{\nu}) \end{aligned} \quad (30)$$

where the third equality in the first row follows from assuming that the random coefficient and the random utility shocks are independent. The fourth equality uses the Type I extreme value distributional assumption about the random utility shocks, and the fifth equality uses that I normalise the components of the utility of the outside good ($j = 0$) to zero.

D.1.2 Minimising the distance between theoretical and observed market shares

The theoretically predicted market shares can be used to find parameter values that match them to empirically observed realised market shares \hat{s}_{jt}^r . The problem is that the parameters enter in a nonlinear fashion into the market shares, which is a difficult minimization problem and more importantly, cannot address price endogeneity concerns in the usual linear way. The main contribution of BLP and Berry (1994) is to show how we can estimate the parameters while taking price endogeneity into account in a linear fashion.¹³⁵ For logit models without

¹³³So everything except the terms $\theta_{it}^p y_{it}$ and μ_{i0t} , see e.g. Nevo (2000b) for more details.

¹³⁴So $A_{jt} = \{(\boldsymbol{\nu}_{it}, \boldsymbol{\mu}_{it}) \mid U_{ijt} > U_{ilt} \forall l = 0, 1, \dots, J\}$

¹³⁵Estimating demand systems has been the focus of a large literature over decades. While allowing flexible substitution patterns between products, the drawbacks of the popular classic Almost Ideal Demand System (Deaton and Muellbauer, 1980) are the number of parameters required to be estimated and the requirement of a representative consumer. Therefore, parts of the relevant IO literature have moved from a product space approach towards a characteristic space approach, which I employ in this paper as well. The BLP model has been further extended (e.g. Nevo (2001)) and used in a variety of contexts, often for merger analysis (e.g.

random coefficients, this is easily achieved by an analytic relationship between market share ratios and mean utility δ_{jt} in (30). Berry (1994) and BLP solve the integral in (30) by simulating consumers and using a contraction mapping. The following sketches the procedure and algorithm. For a more detailed account, see BLP, Berry (1994) or Nevo (2000b).

The algorithm operates on an inner and an outer loop. The inner loop first solves for δ_{jt} , and then linearly estimates the mean coefficients (θ^p, θ^x) and ξ . The outer loop solves for Σ . The inner loop finds a δ_{jt} for a given Σ that sets the observed (\hat{s}_{jt}^r) and the theoretical market shares (s_{jt}^r) equal: $\min_{\delta_{jt}} \|s_{jt}^r - \hat{s}_{jt}^r\|$. The theoretical market shares are calculated via numerical integration by simulation by drawing a number of consumers N (a consumer is defined by ν_{it} after integrating the random utility component μ_{ijt} out):

$$s_{jt}^r \approx \frac{1}{N} \sum_i \frac{\exp(\delta_{jt} - \sigma^p \nu_{it}^p P_{jt}^r + \sigma^x \nu_{it}^x x_{jt})}{1 + \sum_{j=1} \exp(\delta_{jt} - \sigma^p \nu_{it}^p P_{jt}^r + \sigma^x \nu_{it}^x x_{jt})} \quad (31)$$

For the baseline I simulate $N = 2000$ consumers, but also check robustness of the point estimates with $N = 10000$ consumers. Berry (1994) proves that there exists a unique δ_{jt} which matches the theoretical and empirical market shares under mild regularity conditions. Based on this, BLP employ a contraction mapping (nested fixed point algorithm) where for each step h , the new δ_{jt}^{h+1} is found conditional on Σ by:

$$\delta_{jt}^{h+1} = \delta_{jt}^h + \ln(\hat{s}_{jt}^r) - \ln(s_{jt}^r)$$

which is iterated until the change in δ_{jt} , so $\ln(\hat{s}_{jt}^r) - \ln(s_{jt}^r)$, is below a tolerance level.¹³⁶

D.1.3 Identifying the linear preference parameters

Thereafter, I obtain the linear parameters (θ^p, θ^x) through a linear IV regression of δ_{jt} from its definition on:

$$\delta_{jt} = -\theta^p P_{jt}^r + \theta^x x_{jt} + \xi_j + \xi_t + \Delta \xi_{jt} \quad (32)$$

where I instrument the endogenous price P_{jt}^r with plant cost shifters and use the appropriate product and time dummies. The price endogeneity arises from correlation with the unobserved

Nevo (2000a)), but also for welfare consequence due to e.g. trade policy changes (Berry et al., 1999) or the introduction of a particular product (minivan) (Petrin, 2002).

¹³⁶I use a tolerance level of on average 10^{-13} with a maximum tolerance of 10^{-12} for an individual jt . Davis and Schiraldi (2014) provide a faster convergence to the unique vector of fixed points via a Newton-Raphson algorithm. An alternative to the inner loop contraction mapping is to use a MPEC (mathematical program with equilibrium constraints) algorithm that takes the market shares as constraints to the GMM objective function, see Dubé et al. (2012).

taste shocks $\Delta\xi_{jt}$, which might allow changing prices without consequences for quantities sold, for example.

D.1.4 Moment conditions in outer loop

From the IV regression I also calculate $\Delta\xi_{jt}$ with which I form the objective GMM function to be minimised to obtain a solution for Σ :

$$\hat{\Sigma} = \arg \min_{\Sigma} \Delta\xi' Z W Z' \Delta\xi$$

where Z is the instrument matrix and $W = (Z'Z)^{-1}$ is a weighting matrix. The underlying moment conditions are that the unobserved deviation in mean utility $\Delta\xi$ are orthogonal to the instrument matrix. I will further discuss the choice of the instrument matrix when I present the results in Section 4.1. The outer loop searches over the parameter space of the nonlinear parameters in Σ , and for each iteration, the inner loop and linear IV regression are performed. This procedure solves for all structural demand side parameters. The estimation performs better with analytical Jacobians, which are provided, along with further estimation details and the analytic robust standard errors of the estimates in Appendix D.2.

D.1.5 Outside good and observed market shares

The estimation uses data on market shares (of sold quantities), rather than quantities themselves. Since the size of the market also includes the outside good, we need to quantify the outside good. BLP, which analyse the car market, for example, take as total market the population that *can* buy a vehicle. Here, I take as market size Y_t the total amount of the particular product sold by Indian firms, both by the firms in the sample and outside the sample by using the plant specific sampling multiplier in the data. Therefore the plant level quantity sold is:

$$Q_{jt}^r = \hat{s}_{jt}^r Y_t \tag{33}$$

and $\sum_{j \geq 1} Q_{jt}^r < 1$ due to the outside good. An increase in the production of an in-sample firm would therefore not increase Y_t , but $\sum_j Q_{jt}^r$.

D.1.6 Price elasticities of demand

The price elasticity of demand is:

$$\frac{1}{\eta_{jt}} \equiv \frac{\partial Q_{jt}}{\partial P_{jt}} \frac{P_{jt}}{Q_{jt}} = \frac{\partial(s_{jt}Y_t)}{\partial P_{jt}} \frac{P_{jt}}{s_{jt}Y_t} = \frac{\partial s_{jt}}{\partial P_{jt}} \frac{P_{jt}}{s_{jt}} = \frac{P_{jt}}{s_{jt}} \frac{1}{N} \sum_i (\theta_{it}^p s_{ijt}(1 - s_{ijt})) \quad (34)$$

where $s_{ijt} \equiv \frac{\exp(\delta_{jt} - \sigma^p \nu_{it}^p P_{jt} + \sigma^x \nu_{it}^x x_{jt})}{1 + \sum_{j=1} \exp(\delta_{jt} - \sigma^p \nu_{it}^p P_{jt} + \sigma^x \nu_{it}^x x_{jt})}$. I omit the notation with r for realised output (or market share) here, since the elasticities can be derived from any prices and quantities in any equilibrium conditional on the estimated parameters. Cross-elasticities can be calculated similarly and vary by firm-pair in each market.

D.2 Demand side algorithm details, Jacobian and standard errors

For the outer loop that searches over Σ , both an interior-point (see e.g. [Byrd et al. \(2000\)](#)) as well as sequential quadratic programming (SQP) algorithm perform well (see e.g. [Nocedal and Wright \(2006\)](#)). As starting values for Σ , I use a vector of zeros, but checked robustness with various positive and negative starting values. I supply a starting value for δ_{jt} from a model without random coefficients where it has an analytical solution: $\delta_{jt}^{start} = \log(\hat{s}_{jt}^r) - \log((\text{outside market share})_t)$.

The Jacobian of the objective function can be solved for analytically and can be supplied to the optimisation algorithm for speed improvements. The derivative of the objective function (let us call it $f()$ for simplicity) with respect to the two elements of Σ on the diagonal (σ) is, with a slight abuse of notation:

$$\frac{\partial f(\Delta \xi(\delta(s^r(\sigma))))}{\partial \sigma} = \frac{\partial f(.)}{\partial \Delta \xi} \frac{\partial \Delta \xi}{\partial \delta} \frac{\partial \delta}{\partial s^r} \frac{\partial s^r}{\partial \sigma} = 2ZWZ' \Delta \xi \left(\frac{\partial s^r}{\partial \delta} \right)^{-1} \frac{\partial s^r}{\partial \sigma}$$

The second last component of the Jacobian $\left(\frac{\partial s^r}{\partial \delta} \right)^{-1}$ is a square matrix with a size equal to the number of observations (so products times markets or periods). The elements of the matrix can be calculated by taking the derivative of Equation (31), which I repeat for convenience:

$$s_{jt}^r \approx \frac{1}{N} \sum_i \frac{\exp(\delta_{jt} - \sigma^p \nu_{it}^p P_{jt}^r + \sigma^x \nu_{it}^x x_{jt})}{1 + \sum_{j=1} \exp(\delta_{jt} - \sigma^p \nu_{it}^p P_{jt}^r + \sigma^x \nu_{it}^x x_{jt})} \equiv \frac{1}{N} \sum_i s_{ijt}^r$$

$$\begin{aligned} \frac{\partial s_{jt}^r}{\partial \delta_{jt}} &= \frac{1}{N} \sum_i s_{ijt}^r (1 - s_{ijt}^r) \quad \forall j \\ \frac{\partial s_{jt}^r}{\partial \delta_{mt}} &= -\frac{1}{N} \sum_i s_{ijt}^r s_{imt}^r \quad \forall m \neq j \end{aligned}$$

The last component of the Jacobian $\frac{\partial \mathbf{s}^r}{\partial \boldsymbol{\sigma}}$ is again obtained by taking the derivative of the market share equation with respect to each of the k diagonal elements σ^k of Σ (with associated draw ν_{it}^k), so:

$$\begin{aligned}\frac{\partial s_{jt}^r}{\partial \sigma^p} &= \frac{1}{N} \sum_i \nu_{it}^p s_{ijt}^r (\sum_j P_{jt}^r s_{ijt}^r - P_{jt}^r) \\ \frac{\partial s_{jt}^r}{\partial \sigma^x} &= \frac{1}{N} \sum_i \nu_{it}^x s_{ijt}^r (x_{jt} - \sum_j x_{jt} s_{ijt}^r)\end{aligned}$$

The standard errors of $\hat{\Sigma}$ are obtained by taking the square root of the diagonal terms of its covariance matrix. The covariance matrix of the GMM estimate $\hat{\boldsymbol{\sigma}}$ is:

$$VC\hat{O}V(\hat{\boldsymbol{\sigma}}) = N (G'ZWZ'G)^{-1} \left(G'ZW\hat{V}WZ'G \right) (G'ZWZ'G)^{-1}$$

where G is the gradient of the moment conditions, for which we can use part of the Jacobian of the objective function above:

$$G \equiv \left(\frac{\partial \mathbf{s}^r}{\partial \boldsymbol{\delta}} \right)^{-1} \frac{\partial \mathbf{s}^r}{\partial \boldsymbol{\sigma}}$$

and W is the 2SLS weighting matrix and \hat{V} a consistent heteroskedasticity robust estimator of the moment conditions:

$$\begin{aligned}W &= (Z'Z)^{-1} \\ \hat{V} &= \frac{1}{N} \sum_{jt} \Delta \xi_{jt} z'_{jt} z_{jt}\end{aligned}$$

The linear structural parameters $\{\theta^p, \theta^x\}$ depend on the non-linear structural parameters Σ and are solved for via the inner loop in the algorithm. In order to obtain a covariance matrix of the linear parameters I bootstrap from the estimated $VC\hat{O}V(\hat{\boldsymbol{\sigma}})$ and solve the inner loop for each draw with an associated $\{\theta^p, \theta^x\}$. I recover the standard errors of the linear parameters from the resulting sampling distribution of $\{\theta^p, \theta^x\}$.

E Details for estimating equilibria

Both the factual and the counterfactual equilibria are determined by a vector of prices, since the market shares (and output quantities) are a function of prices and the structural parameters. The input quantities for the counterfactual can be derived from the cost minimisation conditions once the equilibrium prices (and quantities) are found. Note that the contraction

mapping in the inner loop was only needed to identify the structural demand parameters. For the equilibria I only search over prices taking the structural demand and production parameters as given. The strategy is to (1) first find the factual equilibrium and associated equilibrium prices (i.e. before shock ϵ introduces noise into realised observed quantities and prices), then (2) calculate the implied factual $\boldsymbol{\tau}$, and (3) use a counterfactual $\tilde{\boldsymbol{\tau}}$ to obtain prices that solve for the counterfactual equilibrium.

The equilibrium conditions are:

$$\frac{P_{jt}}{MC_{jt}(Q_{jt}(\mathbf{P}_t), \mathbf{c}_{jt}(\boldsymbol{\tau}))} - \frac{1}{1 + \eta_{jt}(\mathbf{P}_t)} = 0 \quad (\text{FOC})$$

$$2 \frac{\partial Q_{jt}}{\partial P_{jt}} + (P_{jt} - MC_{jt}) \frac{\partial^2 Q_{jt}}{(\partial P_{jt})^2} - \frac{\partial MC_{jt}}{\partial Q_{jt}} \left(\frac{\partial Q_{jt}}{\partial P_{jt}} \right)^2 \leq 0 \quad (\text{SOC})$$

For both, the factual and counterfactual equilibria, I use the Hessian (SOC) as a constraint in the optimisation to ensure profit maximisation (and not minimisation). We can rewrite the Hessian with markets shares instead of quantities using Equation (33): $Q_{jt} = s_{jt}Y_t$. Note that I do not use superscript r in this section since I am now using equilibrium, not realised, quantities and prices. The Hessian is:

$$H_{jt} = Y_t \left(2 \frac{\partial s_{jt}}{\partial P_{jt}} + (P_{jt} - MC_{jt}) \frac{\partial^2 s_{jt}}{(\partial P_{jt})^2} - Y_t \frac{\partial MC_{jt}}{\partial s_{jt}} \left(\frac{\partial s_{jt}}{\partial P_{jt}} \right)^2 \right)$$

where

$$\begin{aligned} \frac{\partial s_{jt}}{\partial P_{jt}} &= \frac{1}{N} \sum_i (\theta_{it}^p s_{ijt} (1 - s_{ijt})) \\ \frac{\partial^2 s_{jt}}{(\partial P_{jt})^2} &= \frac{1}{N} \sum_i ((\theta_{it}^p)^2 s_{ijt} (1 - s_{ijt}) (1 - s_{ijt} - s_{ijt})) \\ \text{where } s_{ijt} &\equiv \frac{\exp(\delta_{jt} - \sigma^p \nu_{it}^p P_{jt} + \sigma^x \nu_{it}^x x_{jt})}{1 + \sum_{j=1} \exp(\delta_{jt} - \sigma^p \nu_{it}^p P_{jt} + \sigma^x \nu_{it}^x x_{jt})} \\ MC_{jt} &= \frac{P_{jt}}{1 + \eta_{jt}(\mathbf{P}_t)} \end{aligned}$$

When using a Cobb-Douglas production function, the marginal cost function¹³⁷ has an analytical closed form and is a function of output Q_{jt} , output elasticities and distortions τ_{jt}^L, τ_{jt}^M :

$$MC_{jt} = \frac{(s_{jt}Y_t)^{-1}}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M} \left(\frac{s_{jt}Y_t}{\Omega_{jt}} \right)^{\frac{1}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} \left[\begin{aligned} & \left(\frac{r_{jt}\alpha_{jt}^M}{\alpha_{jt}^K} \right)^{\frac{\alpha_{jt}^K}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} \left(\frac{w_{jt}\tau_{jt}^L\alpha_{jt}^M}{\alpha_{jt}^L} \right)^{\frac{\alpha_{jt}^L}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} (P_{jt}^M \tau_{jt}^M)^{\frac{\alpha_{jt}^M}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} \\ & + \left(\frac{r_{jt}\alpha_{jt}^L}{\alpha_{jt}^K} \right)^{\frac{\alpha_{jt}^K}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} \left(\frac{\alpha_{jt}^L P_{jt}^M \tau_{jt}^M}{\alpha_{jt}^M} \right)^{\frac{\alpha_{jt}^M}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} (w_{jt}\tau_{jt}^L)^{\frac{\alpha_{jt}^L}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} \\ & + \left(\frac{\alpha_{jt}^K w_{jt}\tau_{jt}^L}{\alpha_{jt}^L} \right)^{\frac{\alpha_{jt}^L}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} \left(\frac{\alpha_{jt}^K P_{jt}^M \tau_{jt}^M}{\alpha_{jt}^M} \right)^{\frac{\alpha_{jt}^M}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} (r_{jt})^{\frac{\alpha_{jt}^K}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} \end{aligned} \right]$$

Hence the remaining component of the Hessian is:

$$\frac{\partial MC_{jt}}{\partial s_{jt}} = \left(\frac{1 - (\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M)}{(\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M)^2} \right) (s_{jt}Y_t)^{\frac{1}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M} - 2} \Omega_{jt}^{-\frac{1}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} \left[\begin{aligned} & \left(\frac{r_{jt}\alpha_{jt}^M}{\alpha_{jt}^K} \right)^{\frac{\alpha_{jt}^K}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} \left(\frac{w_{jt}\tau_{jt}^L\alpha_{jt}^M}{\alpha_{jt}^L} \right)^{\frac{\alpha_{jt}^L}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} (P_{jt}^M \tau_{jt}^M)^{\frac{\alpha_{jt}^M}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} \\ & + \left(\frac{r_{jt}\alpha_{jt}^L}{\alpha_{jt}^K} \right)^{\frac{\alpha_{jt}^K}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} \left(\frac{\alpha_{jt}^L P_{jt}^M \tau_{jt}^M}{\alpha_{jt}^M} \right)^{\frac{\alpha_{jt}^M}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} (w_{jt}\tau_{jt}^L)^{\frac{\alpha_{jt}^L}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} \\ & + \left(\frac{\alpha_{jt}^K w_{jt}\tau_{jt}^L}{\alpha_{jt}^L} \right)^{\frac{\alpha_{jt}^L}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} \left(\frac{\alpha_{jt}^K P_{jt}^M \tau_{jt}^M}{\alpha_{jt}^M} \right)^{\frac{\alpha_{jt}^M}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} (r_{jt})^{\frac{\alpha_{jt}^K}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} \end{aligned} \right]$$

For the factual equilibrium, I search over prices that minimise the summed squared

¹³⁷The cost function is:

$$C_{jt} = \left(\frac{Q_{jt}}{\Omega_{jt}} \right)^{\frac{1}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} \left[\begin{aligned} & \left(\frac{r_{jt}\alpha_{jt}^M}{\alpha_{jt}^K} \right)^{\frac{\alpha_{jt}^K}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} \left(\frac{w_{jt}\tau_{jt}^L\alpha_{jt}^M}{\alpha_{jt}^L} \right)^{\frac{\alpha_{jt}^L}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} (P_{jt}^M \tau_{jt}^M)^{\frac{\alpha_{jt}^M}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} \\ & + \left(\frac{r_{jt}\alpha_{jt}^L}{\alpha_{jt}^K} \right)^{\frac{\alpha_{jt}^K}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} \left(\frac{\alpha_{jt}^L P_{jt}^M \tau_{jt}^M}{\alpha_{jt}^M} \right)^{\frac{\alpha_{jt}^M}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} (w_{jt}\tau_{jt}^L)^{\frac{\alpha_{jt}^L}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} \\ & + \left(\frac{\alpha_{jt}^K w_{jt}\tau_{jt}^L}{\alpha_{jt}^L} \right)^{\frac{\alpha_{jt}^L}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} \left(\frac{\alpha_{jt}^K P_{jt}^M \tau_{jt}^M}{\alpha_{jt}^M} \right)^{\frac{\alpha_{jt}^M}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} (r_{jt})^{\frac{\alpha_{jt}^K}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} \end{aligned} \right]$$

distances between the markets shares that are a function of these prices \mathbf{s} and the equilibrium market shares $\hat{\mathbf{s}}$ that we know from Equation (33), $\hat{s}_{jt} = \frac{\hat{s}_{jt}^r}{\exp(\epsilon_{jt})}$:

$$\begin{aligned} \mathbf{P}_t^{fact} &= \arg \min_{\mathbf{P}_t} -\hat{\mathbf{s}}' \mathbf{s}(\mathbf{P}_t) \\ s.t. \quad &H_{jt}(\mathbf{P}_t) \leq 0 \quad \forall \quad jt \end{aligned}$$

Using realised prices as starting values for the search typically is most efficient, but I also check the robustness with alternative starting values. I use the first order equilibrium conditions above to infer the $\boldsymbol{\tau}$. Once we set the counterfactual $\tilde{\boldsymbol{\tau}}$, we can search for the counterfactual equilibrium prices. For the counterfactuals, I use the first order equilibrium conditions as the objective function. I minimise the squared distances between the variable marginal costs inferred from the prices and demand elasticity, and the variable marginal costs from the derivative of the cost function:

$$\begin{aligned} MC_{jt}^D &= \frac{P_{jt}}{1 + \eta_{jt}(\mathbf{P}_t)} = \\ MC_{jt}^C &= \frac{1}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M} \left(\frac{s_{jt} Y_t}{\Omega_{jt}} \right)^{\frac{1}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M} - 1} \\ &\quad \left[\left(\frac{r_{jt} \alpha_{jt}^M}{\alpha_{jt}^K} \right)^{\frac{\alpha_{jt}^K}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} \left(\frac{w_{jt} \tau_{jt}^L \alpha_{jt}^M}{\alpha_{jt}^L} \right)^{\frac{\alpha_{jt}^L}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} (P_{jt}^M \tau_{jt}^M)^{\frac{\alpha_{jt}^M}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} \right. \\ &\quad + \left(\frac{r_{jt} \alpha_{jt}^L}{\alpha_{jt}^K} \right)^{\frac{\alpha_{jt}^K}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} \left(\frac{\alpha_{jt}^L P_{jt}^M \tau_{jt}^M}{\alpha_{jt}^M} \right)^{\frac{\alpha_{jt}^M}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} (w_{jt} \tau_{jt}^L)^{\frac{\alpha_{jt}^L}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} \\ &\quad \left. + \left(\frac{\alpha_{jt}^K w_{jt} \tau_{jt}^L}{\alpha_{jt}^L} \right)^{\frac{\alpha_{jt}^L}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} \left(\frac{\alpha_{jt}^K P_{jt}^M \tau_{jt}^M}{\alpha_{jt}^M} \right)^{\frac{\alpha_{jt}^M}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} (r_{jt})^{\frac{\alpha_{jt}^K}{\alpha_{jt}^K + \alpha_{jt}^L + \alpha_{jt}^M}} \right] \end{aligned}$$

Therefore:

$$\begin{aligned} \mathbf{P}_t^{counterfact} &= \arg \min_{\mathbf{P}_t} -\mathbf{MC}^C(\boldsymbol{\tau})' \mathbf{MC}^D(\mathbf{P}_t) \\ s.t. \quad &H_{jt}(\mathbf{P}_t) \leq 0 \quad \forall \quad jt \end{aligned}$$

Using the factual equilibrium prices as starting values for the counterfactual prices is typically most efficient, but I checked a range of alternative starting values.

All plant-years where I estimate an elasticity larger than -1 in the original demand estimation are ignored in terms of finding any equilibrium and are ignored for the comparative

statics as well.¹³⁸ Furthermore I constrain the elasticity to be smaller than -1 in the factual and counterfactual estimation such that the relationship between prices and marginal cost markups is defined. This also improves stability in finding the equilibria since the algorithm does not move over the discontinuity. I also supply a lower bound of zero on prices to the algorithm.

For both, the factual and the counterfactual equilibria, I solve separately for each market (i.e. time period) since they are independent. This reduces the dimension over which prices are searched and speeds up the algorithm considerably. I use a sequential quadratic programming (SQP) algorithm (Nocedal and Wright, 2006) which almost always performs faster than an interior point algorithms for these purposes.

While neither the existence nor the uniqueness of equilibria is proven analytically, the algorithm always finds an equilibrium, including for the bootstrapped equilibria where I have different simulated structural parameters for each draw. This at least proves existence for this sample. I perform some checks on the global nature (and uniqueness) of the minimum by using a range of different starting values which converge to the same equilibrium.

F Details on the Indian iron and steel industry

Iron and steel industry in India

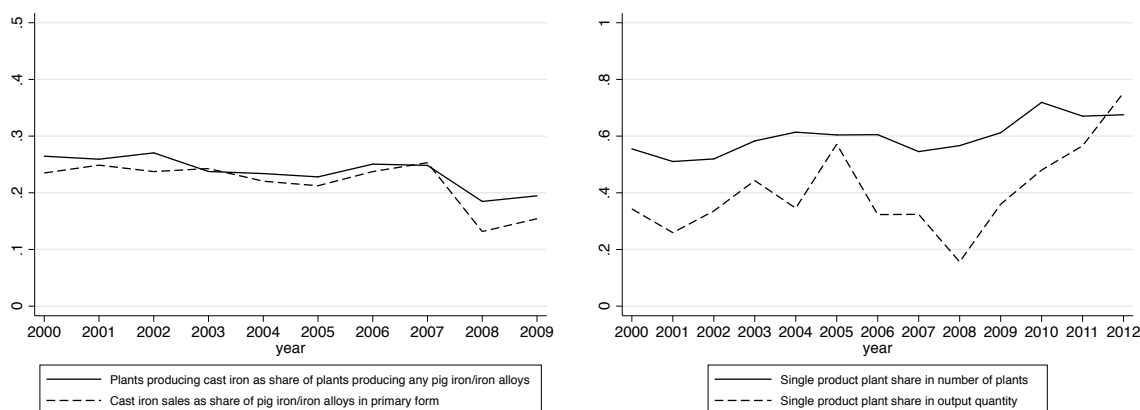
The iron and steel sector in India is an interesting sector in itself. In 2015 India has been the third largest producer both of total crude steel and pig iron after China and Japan (WSA, 2016b), up from ninth place for both in 2000 (WSA, 2010). The share of basic iron and steel (2710 in ISIC3) in value added of total manufacturing was 15% in 2007 (latest year available in UNIDO (2016) Indstat). After Bahrain, which have a small total manufacturing sector, this is the largest value added share any basic iron and steel sector has in their national manufacturing value added in the world (see Figure 19). Figure 16, 17 and 18 show more descriptive statistics on the cast iron sample described in the main text.

Raw materials in the iron and steel industry and carbon emissions

In terms of raw materials used in the iron and steel sector, India was the fourth largest producer of iron ore after China, Australia and Brazil (WSA, 2016b) in 2015, and the third largest coal producer (EIA, 2015). Even after the raw material mining stage, the production chain and process in the iron and steel industry has substantial environmental significance due to its heavy use of coal. The main raw materials iron ore and coal as well as alloying

¹³⁸As noted in the main text, there are only 9 observations with a median market share of 0.0004.

Figure 16: Cast iron producing plants as share of all iron alloy producing plants
Figure 17: Share of single product plants in cast iron manufacturing

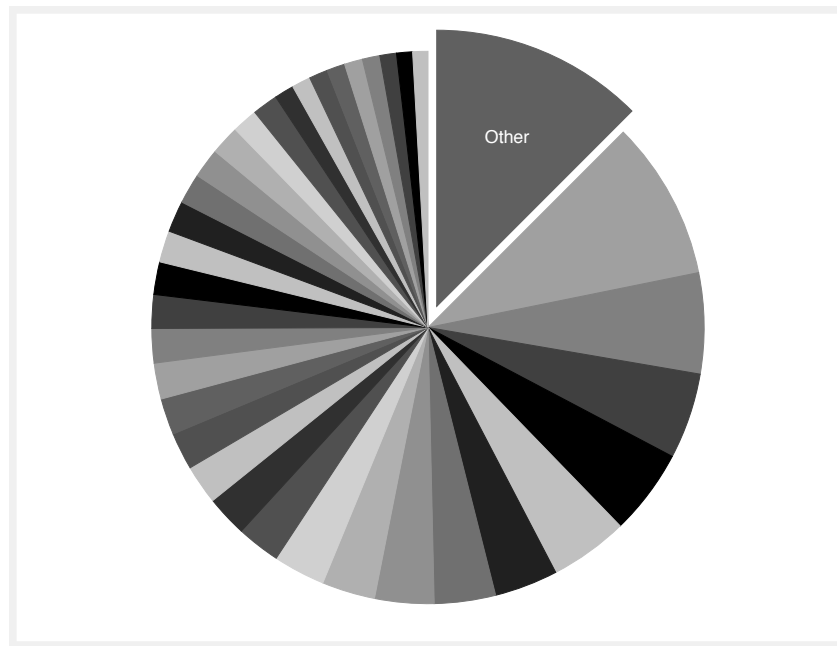


elements such as nickel or chromium are relatively abundant resources. Coal has typically the dual role of providing the heating for smelting and melting in the production chain, but is also directly required to adjust the carbon content of the products. Carbon is often burnt out in the melting process and needs to be re-added accordingly. Carbon emissions come therefore from the heat generation as well as the process of production directly. For the upstream production of pig iron, coke is the main reducing agent to turn iron ore into pig iron saturated with carbon in the smelting process.

Globally, in 2013 around 15% of total coal consumption is accounted for by the iron and steel industry ([World Coal Association, 2014](#)), more than twice the entire consumption of the EU ([BP, 2016](#)). India accounts for around 9% of global coal consumption, primarily through its electricity generating sector which is heavily reliant on coal ([BP, 2016](#)). Therefore even with substitution to a different low-carbon fuel for electricity generation and heating, coal is likely to remain a necessary ingredient in iron and steel manufacturing, with accompanying process emissions. According to the [UNFCCC \(2016\)](#), the iron and steel sector accounts for more than a third of all process emissions in Annex I countries. Per tonne of produced steel in India around 3.1 to 3.8 tonne of CO₂ are emitted ([IPCC, 2007](#)). With India's steel production at 89 million tonnes in 2015 ([WSA, 2016b](#)), this implies around 328 million tonnes of total CO₂ emissions of the Indian iron and steel sector, around 82% of UK's or 13% of India's total emissions in 2015 ([EC, 2016](#)).

Reduction of process emissions can be achieved through process and product innovation. In terms of process innovation, pulverised coal injection techniques can save around 30% of coal ([WSA, 2016a](#)), and emissions can be reduced through ex-post carbon capture and storage. In terms of product innovation, it is often cited that 75% of steel types have been introduced in the past 20 years. The Eiffel Tower would only require a third of today's material and old

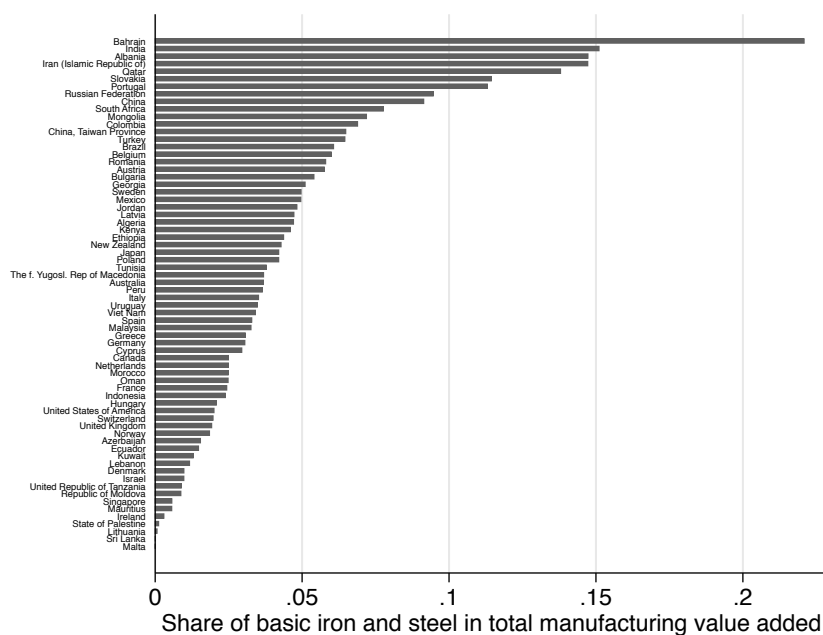
Figure 18: Industry concentration: 35 biggest players in 2004



Notes: The market shares of the 35 largest plants and "other" plants are shown in 2004. The calculation is based on the single product firms in the final sample.

automobiles would only require two thirds of today's steel ([WSA, 2017](#)). However, if there are barriers to reallocation from less efficient to more efficient firms, then removing barriers and reallocating inputs could in principle decrease aggregate process emissions, without any plant level process or product innovations, which this paper can shed light on.

Figure 19: Share of basic iron and steel in total manufacturing value added

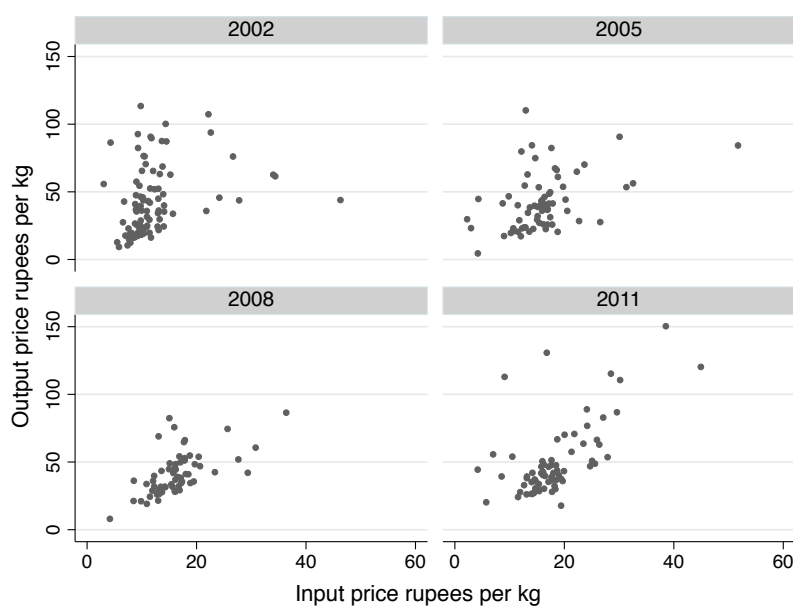


Notes: Calculations based on data from [UNIDO \(2016\)](#).

G Further results and robustness checks

G.1 Output and input prices

Figure 20: Output and input prices for selected years



Notes: The figure plots deflated output and input prices of the cast iron plants in rupees per kg. Input prices are recovered by dividing total expenditure on material inputs by the total weight of material inputs.

G.2 Detailed results from demand estimation

G.2.1 Instrument choice and first stage

In order to consistently estimate the mixed logit demand model we need instruments for the endogenous price. Taste shocks captured in $\Delta\xi_{jt}$, for example, are likely to be positively correlated to prices in the mean utility regression (32). If higher utility can be derived from a product, producers are likely to be able to raise prices without compromising on sales. Note that including product and year fixed effects ξ already goes a long way by accounting for time or firm invariant taste characteristics as well as other sources of endogeneity bias that do not vary across these dimensions.

There are several candidates for instruments which have been used in demand estimation. The profit maximisation condition for plants is a useful guide for instrument choice around which the literature can be structured. Rewriting condition (1) yields $P_{jt} = \frac{1+\eta_{jt}(\mathbf{P}_t)}{MC_{jt}(Q_{jt}(\mathbf{P}_t), \mathbf{c}_{jt})}$. Much of the literature relies on an internal instrument that drives the numerator $1+\eta_{jt}(\mathbf{P}_t)$.¹³⁹ However, since I have endogenous marginal costs that depend on observed marginal cost shifters, I can use these external instruments for prices. Shocks to input prices are assumed to be uncorrelated with shocks to taste $\Delta\xi_{jt}$, conditional on average product quality which is controlled for by product fixed effects. [Armstrong \(2016\)](#) shows that internal BLP-style instruments, in particular in the Bertrand Nash structure, tend to perform poorly in small samples and also lose identifying power asymptotically. He recommends cost shifters which are consistent over a broad range of cases.

I use the average plant level wages w_{jt} , and the average plant level prices of a tonne of material inputs P_{jt}^M as instruments for output prices P_{jt} . At the solution of the model, the first stage Kleibergen-Paap F statistic is 21.81, rejecting a weak instrument hypothesis. The Hansen overidentification Chi-Square J statistic is smaller than 0.001, with a p-value close to unity so the hypothesis of valid instruments can not be rejected. The point estimate of the IV regression of mean utility on price is -17.15*** as shown in Table 10. A plain OLS regression in (32) yields an estimate of -14.60***, which is significantly positively biased as expected.

An alternative instrument was proposed by [Foster et al. \(2008\)](#) who estimate productivity (TFPQ) and a demand function. They propose to use TFPQ (Ω_{jt}) or the innovation in the

¹³⁹[Hausman et al. \(1994\)](#) estimate a nested logit demand system for beer and use other cities' beer prices to instrument for a city's beer price. BLP themselves use other products' characteristics values of the same firm and the same product's characteristics values of other firms (via a competition channel). Both do not rely on additional data, but require stronger assumptions, as discussed therein and in [Nevo \(2001\)](#). [Reynaert and Verboven \(2014\)](#) suggest an improved set of BLP instruments based on [Chamberlain \(1987\)](#) optimal instruments (see also more recent work on this in [Gandhi and Houde \(2016\)](#)). The rank condition of these types of instruments can be rationalised by noting that they affect the demand elasticities, upon which the price choice depends in the firm's optimisation condition.

Table 10: Estimates of demand parameters

	Point estimate	SE
θ^p	-17.15***	0.40 (bootstrapped)
σ^p	-5.83***	0.13 (robust)
σ^x	0.02	1.19 (robust)
ξ	Yes	.

Notes: The table shows the estimates for the structural parameters on the demand side. The number of observations is 989. The standard errors of the linear parameter θ^p depends on the non-linear parameters. Details for calculation of the standard errors are in Appendix D.2.

productivity process (ζ_{jt}) as instruments for prices. In my data, the correlation between Ω_{jt} and P_{jt} is -0.18***, and between ζ_{jt} and P_{jt} -0.09*. This makes intuitive sense, considering that higher productivity leads to higher output quantity which is associated with lower prices. However, Forlani et al. (2016) find in their joint estimation of demand and productivity that shocks in consumer taste and shocks to productivity are negatively correlated, which would make both TFPQ (Ω_{jt}) and the innovation in it (ζ_{jt}) unsuitable instruments.¹⁴⁰ Including TFPQ (Ω_{jt}) as an additional instrument for estimating Equation (32) yields a point estimate of -17.20*** for the instrumented price, very similar to my baseline result of -17.15*** in Table 10. The first stage F statistic is 16.05, and Hansen’s J test of valid instruments cannot be rejected with a p-value of 0.70.

G.2.2 Estimated demand parameters

I omit the estimates for the dummies (θ^x, ξ_j, ξ_t) in Table 10. The estimate for the mean price coefficient θ^p (-17.15***) is negative and highly statistically significant, as is the standard deviation of the mean price coefficient σ^x (-5.83***).¹⁴¹ This means that there is significant variation in the random coefficient on price. On the other hand, the variance on the constant (σ^x) is small and insignificant, so presents little evidence of a random intercept. See Appendix D.2 for details on the calculation of the standard errors.

G.3 Further results and robustness checks for the production side

Table 11 shows the baseline estimates in Column (1) and the OLS results in Column (2). Since I am only using observations that are part of consecutive spells of data because of the

¹⁴⁰I also find that productivity shocks ζ_{jt} and unobserved average utility shocks $\Delta\xi_{jt}$ are negatively correlated (-0.04), but insignificantly. I avoid using ζ_{jt} as instrument also for practical reasons. I would need to estimate the production side to generate the instrument for the demand side. But when estimating the demand side first, I can use demand side estimates as instruments for the invertibility condition in the production estimation.

¹⁴¹This translates into a variance of 33.9. The negative sign of the standard deviation is irrelevant as the square, the variance, is always positive. If I take positive starting values for σ^p instead of zero, I get an equal size “positive” standard deviation as solution.

Table 11: Estimates from a Cobb-Douglas production function

	Type of correction				
	(1) Simultaneity & Selectivity	(2) None: OLS	(3) Simultaneity only	(4) Sim. & Selec. w. augmented z	(5) BB system- GMM
α^K	.06*** (.02)	.04*** (.01)	.10*** (.02)	.05** (.02)	.08 (.09)
α^L	.22*** (.05)	.14*** (.02)	.19*** (.05)	.20*** (.05)	.06 (.10)
α^M	.64*** (.05)	.80*** (.03)	.68*** (.06)	.65*** (.05)	.73*** (.07)
RTS	.92*** (.03)	.99*** (.01)	.97*** (.02)	.90*** (.03)	.86*** (.06)
N	443	1001	512	443	512

Notes: The columns show the output elasticities and returns to scale for different type of corrections for simultaneity and selectivity. Column (4) includes an additional variable in the material demand equation $(\hat{\eta}_{jt} + 1) \frac{P_{jt}^r Q_{jt}^r}{P_{jt}^M M_{jt}} = \frac{\tau_{jt}^M P_{jt}^r Q_{jt}^r}{\alpha_{jt}^M P_{jt} Q_{jt}}$. Column (5) is based on a Blundell-Bond system GMM estimator. Clustered standard errors at the plant level are in parentheses.

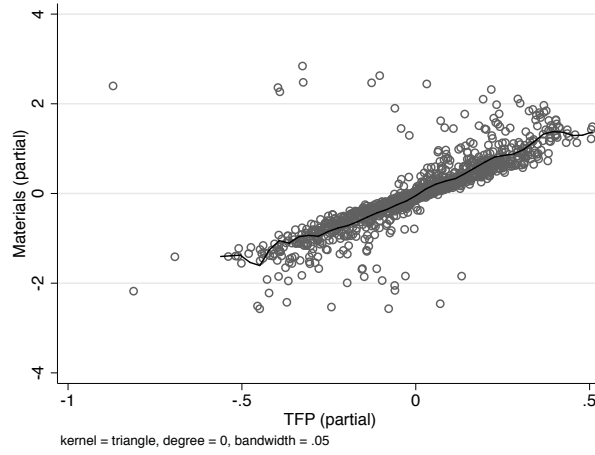
assumed timing structure of the model, I have fewer observations than for the plain OLS result in Column (2). Observations that belong to a plant that has consecutive spells but are in years without consecutive spells are not used for the estimation. The OLS results are robust to only using the same consecutive-spells sample of Column (1) as well. Importantly, since there are no lags in the first equation of this GMM system, I can calculate ϵ_{jt} and Ω_{jt} also for the non-consecutive observations (and can use them for the counterfactual exercise).

Column (3) controls for simultaneity only, which reduces the material elasticity compared to Column (2) in the Cobb-Douglas specification. In the translog version the mean increases, but with a much higher standard deviation.

I perform several robustness checks for the crucial scalar unobservable and invertibility condition in the control function approach. Column (4) includes a variable in the invertibility condition that is aimed to capture variation in the unobserved τ . I construct a variable $(\hat{\eta}_{jt} + 1) \frac{P_{jt}^r Q_{jt}^r}{P_{jt}^M M_{jt}}$ which is similar to the definition of τ and include it in \mathbf{z}_{jt} .¹⁴² Empirically, the estimates are very close to the main specification in Column (1) suggesting no violation of the invertibility condition (18). In Column (5), I use the Blundell-Bond system GMM estimator as a further check for the invertibility. Compared to OLS it also reduces the bias in the OLS material elasticity in the same direction as my main specification, particularly for Cobb-Douglas, but yields less precise estimates. We can also inspect the monotonicity required for the invertibility condition, similar to Levinsohn and Petrin (2003). Figure 21 plots material use against productivity, where I partialled out a polynomial of the other

¹⁴²It is the same apart from using realised rather than expected revenue, and omitting the output elasticity, which is constant across plants in the Cobb Douglas case.

Figure 21: Monotonicity of material demand in productivity



Notes: Plotted are the residuals from a regression of log materials m_{it} on a third order polynomial in labour and capital, against the residuals of a regression of log productivity $\hat{\omega}_{jt}$ on the same polynomial.

input variables, and fitted a local kernel with a tight bandwidth. For most of the density, the smoothed mean is indeed monotonically increasing in productivity (as expected), with a couple of outliers.

Finally, the autoregressive parameter estimate in the productivity process is 0.87*** for Cobb-Douglas and 0.81** for translog, very similar to the annual persistence parameter of 0.8 in Foster et al. (2008). The coefficient for the predicted probability of being in the sample is positive and significant for predicting productivity as expected and in line with the corrected selection bias in the capital elasticity.

G.4 Translog estimates

The estimates from a translog production function are presented in Table 12. The translog elasticities vary by plant and year, but the mean elasticities are very similar to the estimates from the Cobb-Douglas production function, with returns to scale close to one. For the translog elasticity, the standard deviation across all plant-years are reported in parentheses. In Column (6) I additionally use an investment function instead of a material demand function for the invertibility condition, following the original Olley and Pakes (1996). These estimates are also less precise because of the lumpiness and zeros in investment data.

G.5 Analysis of estimated plant total factor productivities

How has productivity evolved and how dispersed is it? I recover *physical* total factor productivity Ω_{jt} (also denoted TFPQ). Total *revenue* factor productivity TFPR is simply

Table 12: Estimates from a translog production function

	(1) Simultaneity & Selectivity	(2) None: OLS	Type of correction		(5) BB system- GMM	(6) Sim. & Selec. w. investment
			(3) Simultaneity only	(4) Sim. & Selec. w. augmented		
α^K	.07 (.14)	.05 (.05)	.08 (.10)	z .07 (.11)	.09 (.07)	.05 (.14)
α^L	.28 (.15)	.13 (.05)	.08 (.21)	.28 (.14)	.04 (.15)	.14 (.34)
α^M	.60 (.22)	.81 (.13)	.82 (.34)	.59 (.19)	.79 (.17)	.76 (.39)
RTS	.95 (.08)	.98 (.06)	.97 (.15)	.94 (.06)	.92 (.05)	.96 (.16)
N	443	1001	512	443	511	410

Notes: The columns show the output elasticities and returns to scale for different types of corrections for simultaneity and selectivity. Column (4) includes an additional variable in the material demand equation $(\hat{\eta}_{jt} + 1) \frac{P_{jt}^r Q_{jt}^r}{P_{jt}^M M_{jt}} = \frac{\tau_{jt}^M P_{jt}^r Q_{jt}^r}{\alpha_{jt}^M P_{jt} Q_{jt}}$. Column (5) is based on a Blundell-Bond system GMM estimator. Column (6) is based on an investment function instead of a material demand function for the invertibility condition, following the original [Olley and Pakes \(1996\)](#). Standard deviations across the entire sample are in parentheses.

defined as $P_{jt} \cdot TFPQ_{jt}$.¹⁴³

I find that TFPQ has not significantly changed over time within firms. A linear within-plants fixed effects regression of logged TFPQ on years yields small and insignificant results. The same goes for a pooled (across plants) regression suggesting that average TFPQ stagnated. This also suggests that entry of productive and exit of unproductive firm did not play a major role either. However, when I weight the pooled regression by output quantity, TFPQ has increased by 1% per year, significant with SE clustered at the plant. This comes from changes in the weights, and likely due to more productive firms growing faster compared to less productive firms, or from larger firms becoming more productive than smaller firms. When we interpret the weighted TFPQ as a form of aggregate TFPQ, the results suggest that despite stagnating average TFPQ, aggregate TFPQ seems to have slightly increased over the sample period.

TFPR increased over the years (using deflated prices), highly significant with SE clustered at the plant. On average, each year the TFPR increases by 2%, both in the pooled and the within estimation. This rise in inflation adjusted prices, together with the result on decreasing markups (Section 4.1.1) suggests increasing marginal costs. Indeed, marginal costs have been rising by a little over 2% per year for the pooled and within specification. This is mainly due to an increase in input prices. Material input prices have been rising at around 3% per year, consistent with the global prices increases in raw metals commodity prices (see e.g. IMF). Increased marginal costs could also have been driven by changes in τ .

The estimated *dispersion* in TFPQ is smaller than in some other studies in the literature.¹⁴⁴

¹⁴³The reported numbers are based on my baseline CD specification of Column (1) that I also use for the counterfactual analysis.

¹⁴⁴The dispersion of productivity (both TFPQ and TFPR) across plants is of interest in itself as it has

The ratio in TFPQ of the 90th percentile plant to the 10th percentile plant is 1.83 in this sample, much smaller than the ratio reported for India in [Hsieh and Klenow \(2009\)](#), which is over 20. This is likely due to three aspects. First, [Hsieh and Klenow \(2009\)](#) don't observe prices and quantities and cannot estimate TFPQ directly. Second, they use a value added instead of gross output production function. Third, I look at a much narrower industry. My ratio is more in line with the ready-mix concrete producers in the US of around 1.91 reported in [Syverson \(2004a\)](#). Finally, the higher dispersion in these studies could also arise, from lumping Ω_{jt} and ϵ_{jt} together, which I disentangle. Indeed the 90th to 10th percentile ratio of the comparable ($\Omega_{jt} \exp(\epsilon_{jt})$) in my data is 2.5.

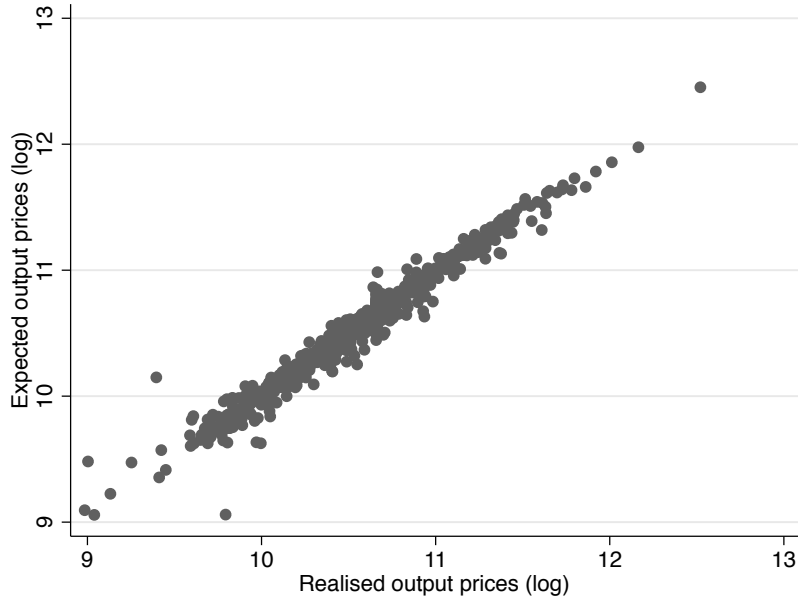
The dispersion in TFPQ is smaller than the dispersion in TFPR. The 90th to 10th TFPR ratio is 3.3, larger than the average ratios of 1.92 in the US within 4-digit sectors as reported in [Syverson \(2004b\)](#), but smaller than the ratio of 5 reported for India in [Hsieh and Klenow \(2009\)](#). Interestingly, I estimate a lower dispersion for TFPQ than for TFPR, which is the opposite for [Hsieh and Klenow \(2009\)](#) and for [Foster et al. \(2008\)](#). I also find a robust and significant negative correlation between TFPQ and prices in the data, rationalised by a standard downward-sloping demand curve. However, because the dispersion in prices is much larger than the dispersion in TFPQ, combining both leads to a dispersion in TFPR that is smaller than that of prices but larger than that of TFPQ.

G.6 Realised and expected prices

Figure 22 plots the realised (after shock ϵ) prices against the expected prices.

become an important feature and subject of analysis in various disciplines, as reviewed in [Syverson \(2011\)](#).

Figure 22: Expected and realised prices

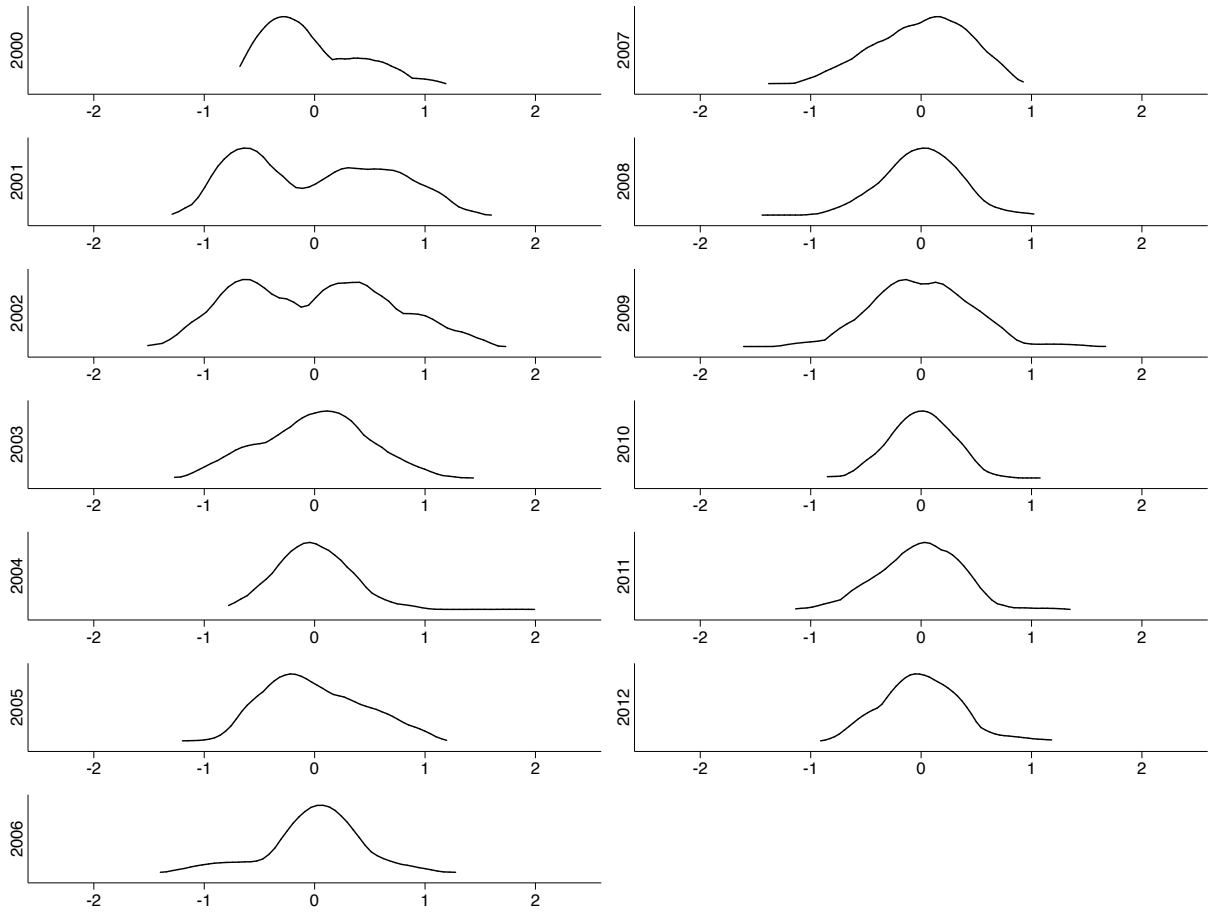


Notes: Plotted is the the realised observed prices $\log(P_{jt}^r)$ against the equilibrium prices $\log(P_{jt})$.

G.7 Additional descriptive figures on distortions

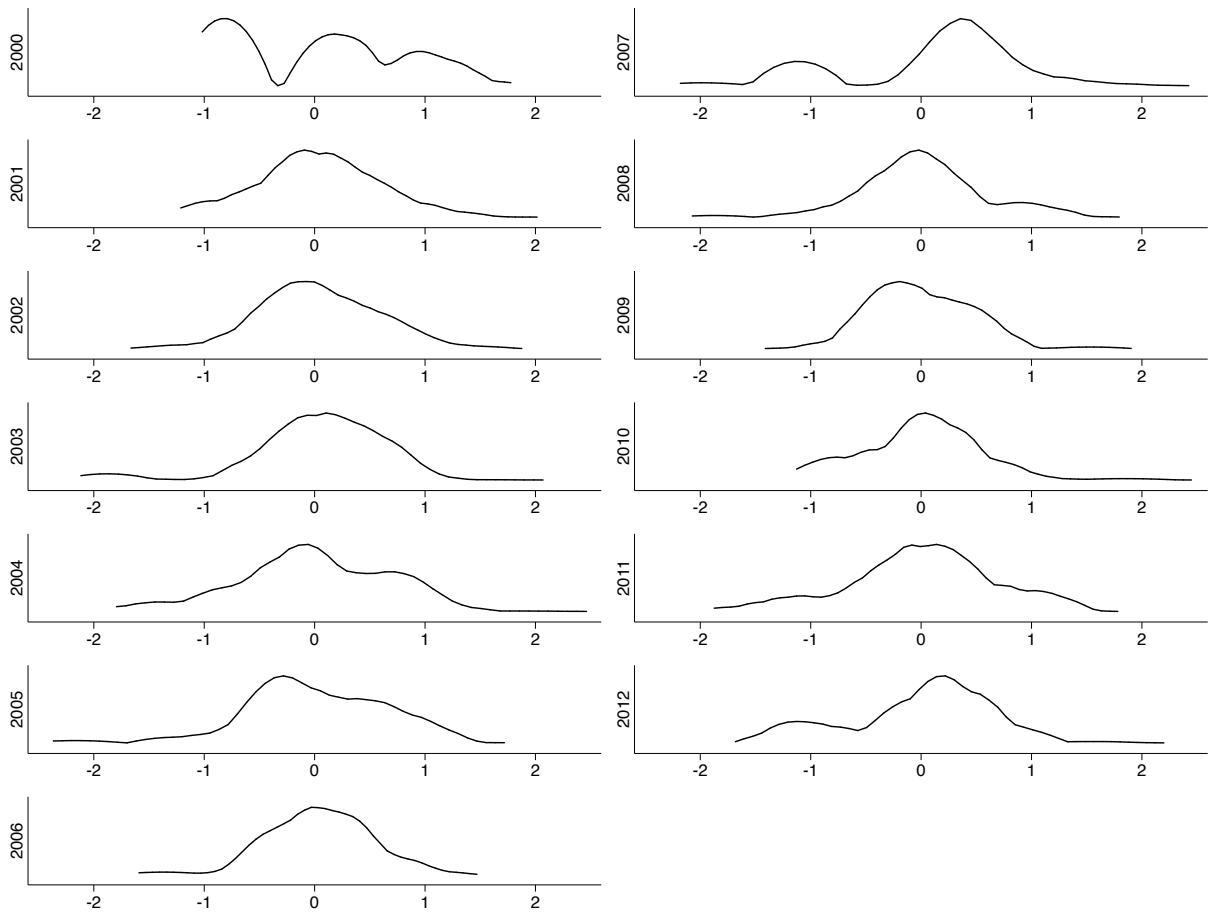
Figures 23 and 24 plot the distortions by year. Figure 25 plots the annual standard deviations. We can also compare the distribution of $\ln(\tau_{jt}^M)$ and $\ln(\tau_{jt}^L)$ in their baseline versions with variable markups with a version for each with constant markups as shown in Figure 26 and 27. Figure 28 plots $\ln(\tau_{jt}^M)$ against $\ln(\tau_{jt}^L)$.

Figure 23: Dispersion in τ_{jt}^M by year



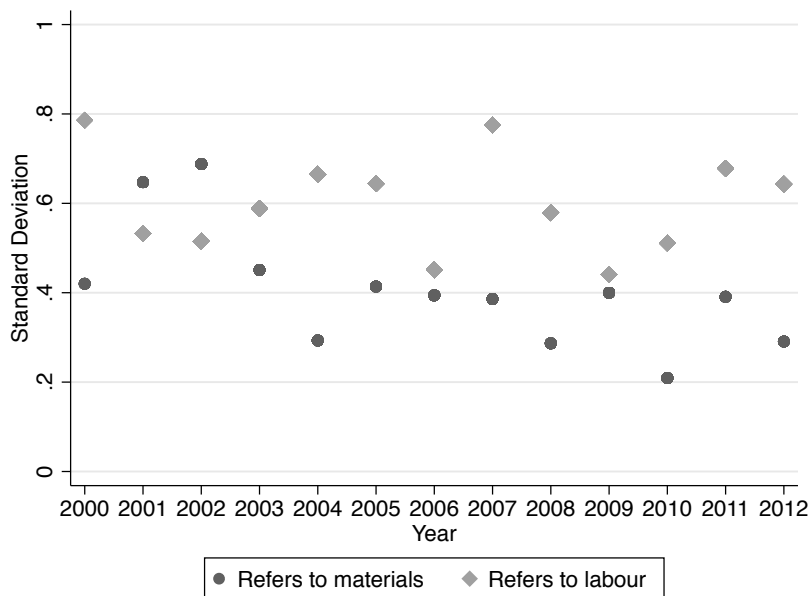
Notes: Plotted is the kernel density of $\ln(\tau_{jt}^M)$ divided by the weighted geometric mean of τ_{jt}^M , where the weights are plant material expenditure. Used kernel is epanechnikov with bandwidth 0.2.

Figure 24: Dispersion in τ_{jt}^L by year



Notes: Plotted is the kernel density of $\ln(\tau_{jt}^L)$ divided by the weighted geometric mean of τ_{jt}^L , where the weights are plant labour expenditure. Used kernel is epanechnikov with bandwidth 0.2.

Figure 25: Standard deviation for τ_{jt}^M and τ_{jt}^L by year



Notes: Plotted are the standard deviations of the demeaned and weighted $\ln(\tau_{jt}^M)$ and $\ln(\tau_{jt}^L)$.

Figure 26: Markup correction for τ_{jt}^M

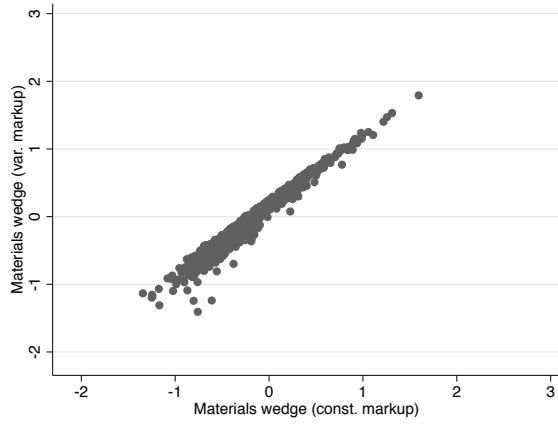
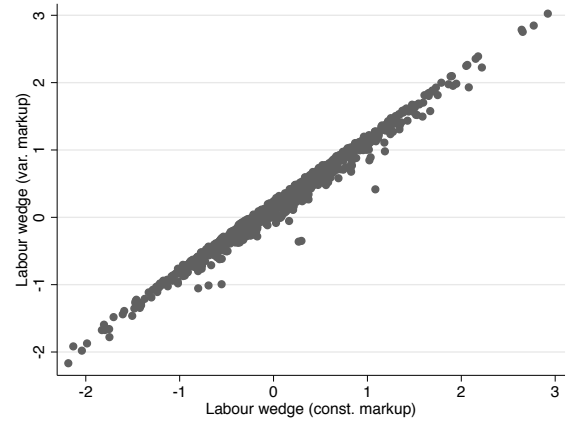
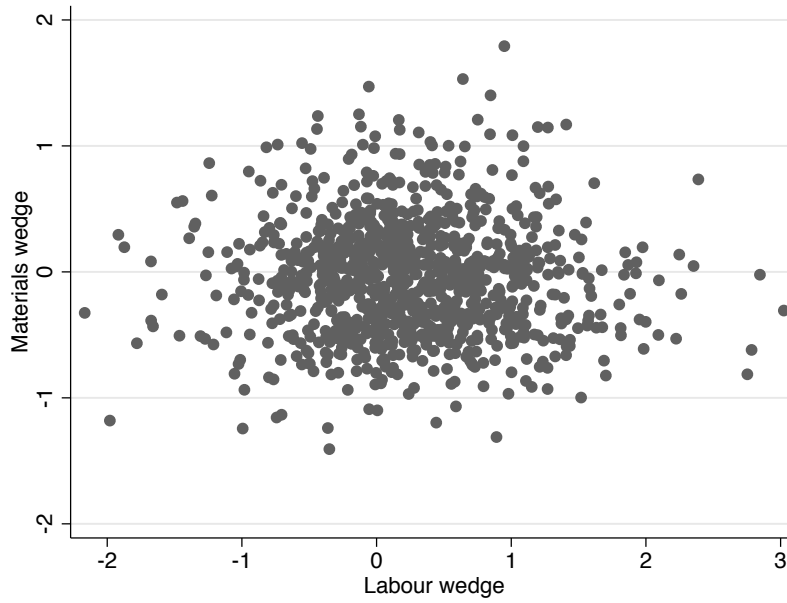


Figure 27: Markup correction for τ_{jt}^L



Notes: Plotted are the pooled demeaned $\ln(\tau_{jt}^M)$ and $\ln(\tau_{jt}^L)$. The vertical axis corresponds to the input distortions corrected for markups and the horizontal axis corresponds to a “naive” version where an average markup is used to calculate input distortions instead.

Figure 28: Correlation between τ_{jt}^M and τ_{jt}^L



Notes: Plotted are $\ln(\tau_{jt}^M)$ and $\ln(\tau_{jt}^L)$ divided by the respective weighted means, where the weights are plant materials and labour expenditure. Pooled across all years.

Table 14: Welfare gains in billion rupees

	Compensating Variation			Profits			Total welfare		
	$\tilde{\tau}_{jt}^M$	$\tilde{\tau}_{jt}^L$	both	$\tilde{\tau}_{jt}^M$	$\tilde{\tau}_{jt}^L$	both	$\tilde{\tau}_{jt}^M$	$\tilde{\tau}_{jt}^L$	both
2000	0.15 [0.11,0.21]	0.14 [0.07,0.2]	0.27 [0.23,0.34]	0.1 [0.07,0.13]	0.08 [0.03,0.1]	0.18 [0.11,0.2]	0.25 [0.19,0.33]	0.21 [0.1,0.3]	0.45 [0.37,0.53]
2001	1.57 [1.13,2.25]	0.39 [0.2,0.62]	2.12 [1.66,2.76]	1 [0.72,1.19]	0.21 [0.09,0.3]	1.32 [0.98,1.43]	2.57 [1.94,3.37]	0.6 [0.29,0.91]	3.43 [2.78,4.15]
2002	3.27 [2.36,4.2]	0.56 [0.28,0.92]	4.08 [3.22,4.81]	2.04 [1.45,2.41]	0.32 [0.15,0.48]	2.48 [1.79,2.77]	5.31 [3.92,6.51]	0.88 [0.43,1.4]	6.56 [5.33,7.45]
2003	2.49 [1.76,3.57]	0.89 [0.45,1.41]	3.29 [2.57,4.26]	1.1 [0.72,1.38]	0.39 [0.18,0.54]	1.41 [1.02,1.64]	3.59 [2.5,4.9]	1.28 [0.64,1.92]	4.7 [3.83,5.77]
2004	2.49 [1.93,3.02]	1.66 [0.89,2.34]	3.68 [3.22,4.02]	0.86 [0.48,1.12]	0.49 [0.22,0.66]	1.22 [0.68,1.38]	3.35 [2.53,4.1]	2.15 [1.1,2.97]	4.9 [4.06,5.34]
2005	2.07 [1.53,2.87]	1.4 [0.66,2.4]	4.07 [3.08,5.34]	1.17 [0.81,1.42]	0.8 [0.35,1.29]	2.5 [1.71,3.29]	3.23 [2.42,4.24]	2.2 [0.99,3.73]	6.57 [5.8,5.9]
2006	2.1 [1.52,3.03]	0.97 [0.48,1.51]	3.25 [2.55,4.28]	1.13 [0.78,1.4]	0.45 [0.2,0.63]	1.72 [1.22,2.09]	3.23 [2.43,4.32]	1.42 [0.69,2.13]	4.97 [3.97,6.4]
2007	1.3 [1.04,1.77]	1.47 [0.81,2.18]	2.6 [2.13,3.34]	0.79 [0.59,0.93]	0.82 [0.35,1.12]	1.6 [1.04,1.83]	2.1 [1.68,2.64]	2.28 [1.14,3.24]	4.2 [3.4,5.02]
2008	0.07 [-0.1,0.33]	0.6 [0.28,0.98]	0.59 [0.3,0.95]	0.14 [0.04,0.24]	0.37 [0.16,0.55]	0.44 [0.26,0.56]	0.21 [-0.06,0.57]	0.97 [0.44,1.51]	1.03 [0.6,1.49]
2009	2.43 [1.66,3.81]	0.81 [0.38,1.31]	3.06 [2.33,4.27]	1.4 [0.98,1.81]	0.41 [0.18,0.61]	1.7 [1.27,2]	3.83 [2.68,5.62]	1.22 [0.56,1.89]	4.76 [3.76,6.22]
2010	0.51 [0.39,0.68]	0.87 [0.44,1.35]	1.11 [0.79,1.44]	0.19 [0.13,0.23]	0.34 [0.15,0.5]	0.44 [0.26,0.54]	0.7 [0.54,0.91]	1.22 [0.59,1.84]	1.55 [1.07,1.98]
2011	1.8 [1.18,2.95]	1.18 [0.61,2.01]	3.52 [2.84,5.45]	1.55 [0.96,2.23]	0.76 [0.35,1.07]	2.98 [2.14,4.86]	3.35 [2.18,5.08]	1.95 [0.96,2.98]	6.5 [5.06,10.16]
2012	1.54 [1,2.3]	1.22 [0.6,1.99]	2.44 [2,3.18]	0.63 [0.38,1.01]	0.66 [0.28,1.03]	1.22 [0.86,1.54]	2.17 [1.41,3.29]	1.88 [0.88,2.96]	3.66 [2.99,4.74]
Total	21.81 [16.08,30.12]	12.16 [6.13,18.87]	34.08 [27.86,42.38]	12.09 [8.4,14.74]	6.09 [2.66,8.61]	19.21 [13.8,22.31]	33.9 [24.87,43.82]	18.26 [8.74,27.27]	53.28 [44.27,64.09]
Per year	1.68 [1.24,2.32]	0.94 [0.47,1.45]	2.62 [2.14,3.26]	0.93 [0.65,1.13]	0.47 [0.2,0.66]	1.48 [1.06,1.72]	2.61 [1.91,3.37]	1.4 [0.67,2.1]	4.1 [3.41,4.93]

Notes: The table shows the respective welfare gains from eliminating material distortions ($\tilde{\tau}_{jt}^M$), labour distortions ($\tilde{\tau}_{jt}^L$) or both in billion rupees. The last three columns sum the consumer side compensating variation and the profits for total welfare gains. The last two rows report the total across all years and the implied average per year. Bootstrapped 95% confidence intervals in brackets (see Section 2.4.5).

G.8 Tables with annual welfare gains

Table 13: Compensating variation as share of consumer expenditure and profit growth

	Compensating Variation			Profit growth		
	$\tilde{\tau}_{jt}^M$	$\tilde{\tau}_{jt}^L$	<i>both</i>	$\tilde{\tau}_{jt}^M$	$\tilde{\tau}_{jt}^L$	<i>both</i>
2000	0.13 [0.09,0.17]	0.11 [0.06,0.17]	0.23 [0.19,0.29]	1.22 [1.14,1.29]	1.16 [1.06,1.24]	1.38 [1.24,1.46]
2001	0.24 [0.17,0.34]	0.06 [0.03,0.09]	0.32 [0.25,0.42]	1.51 [1.35,1.63]	1.11 [1.05,1.17]	1.66 [1.48,1.82]
2002	0.31 [0.22,0.4]	0.05 [0.03,0.09]	0.39 [0.31,0.46]	1.71 [1.47,1.87]	1.11 [1.05,1.18]	1.86 [1.6,2.03]
2003	0.16 [0.11,0.23]	0.06 [0.03,0.09]	0.21 [0.16,0.27]	1.29 [1.18,1.39]	1.1 [1.04,1.16]	1.37 [1.25,1.47]
2004	0.15 [0.11,0.18]	0.1 [0.05,0.14]	0.22 [0.19,0.24]	1.2 [1.1,1.3]	1.11 [1.05,1.17]	1.28 [1.15,1.38]
2005	0.13 [0.1,0.19]	0.09 [0.04,0.16]	0.26 [0.2,0.35]	1.28 [1.19,1.37]	1.19 [1.08,1.36]	1.61 [1.38,1.94]
2006	0.1 [0.08,0.15]	0.05 [0.02,0.08]	0.16 [0.13,0.21]	1.21 [1.14,1.28]	1.08 [1.04,1.13]	1.32 [1.21,1.46]
2007	0.09 [0.07,0.12]	0.1 [0.06,0.15]	0.18 [0.15,0.23]	1.18 [1.13,1.23]	1.19 [1.08,1.28]	1.36 [1.24,1.47]
2008	0.01 [-0.01,0.04]	0.06 [0.03,0.11]	0.06 [0.03,0.1]	1.05 [1.02,1.1]	1.14 [1.06,1.23]	1.17 [1.1,1.24]
2009	0.15 [0.1,0.24]	0.05 [0.02,0.08]	0.19 [0.15,0.27]	1.33 [1.22,1.5]	1.1 [1.04,1.16]	1.4 [1.29,1.56]
2010	0.05 [0.04,0.06]	0.08 [0.04,0.13]	0.1 [0.08,0.14]	1.07 [1.04,1.1]	1.13 [1.06,1.21]	1.16 [1.09,1.22]
2011	0.1 [0.06,0.16]	0.06 [0.03,0.11]	0.19 [0.15,0.29]	1.34 [1.2,1.52]	1.17 [1.08,1.24]	1.65 [1.44,2.2]
2012	0.1 [0.06,0.14]	0.08 [0.04,0.12]	0.15 [0.12,0.2]	1.15 [1.09,1.26]	1.15 [1.06,1.26]	1.29 [1.19,1.43]
Total	0.13 [0.1,0.18]	0.07 [0.04,0.11]	0.21 [0.17,0.26]	1.26 [1.17,1.35]	1.13 [1.06,1.21]	1.42 [1.29,1.56]

Notes: The table shows the respective welfare gains from eliminating material distortions ($\tilde{\tau}_{jt}^M$), labour distortions ($\tilde{\tau}_{jt}^L$) or both. The first three columns express the compensating variation per unit purchased as a share of the unit price (i.e. as share of expenditure on the products in the sample). The profit ratio is total profits in the counterfactual divided by total profits in the factual equilibrium. Bootstrapped 95% confidence intervals in brackets (see Section 2.4.5).

G.9 Tables with annual input productivity gains

Table 15 and 16 report the ratio in the physical and revenue productivities respectively, i.e. the ratio between the input productivity in the counterfactual and the factual equilibria.¹⁴⁵

¹⁴⁵The first three columns in both tables show that output increases more than revenues (except in one case in 2008), as the average price decreases which contributes to the consumer welfare gains. Therefore the ratios of the counterfactual and factual productivities is lower in the revenue productivity outcomes. Since we have decreasing prices across the counterfactuals, we would need to correct for this and inflate the revenue productivity accordingly. But since I can measure output in weight, the physical productivity is a directly suited metric for deflated value per unit.

Table 15: Physical output and productivity ratios

	Output ratio			Material productivity			Labour productivity		
	$\tilde{\tau}_{jt}^M$	$\tilde{\tau}_{jt}^L$	<i>both</i>	$\tilde{\tau}_{jt}^M$	$\tilde{\tau}_{jt}^L$	<i>both</i>	$\tilde{\tau}_{jt}^M$	$\tilde{\tau}_{jt}^L$	<i>both</i>
2000	1.27 [1.16,1.35]	1.26 [1.11,1.37]	1.52 [1.32,1.63]	0.89 [0.87,0.93]	1.19 [1.09,1.31]	1.03 [0.97,1.13]	0.9 [0.87,0.96]	0.55 [0.52,0.58]	0.55 [0.53,0.59]
2001	1.61 [1.37,1.83]	1.15 [1.06,1.24]	1.87 [1.56,2.1]	0.75 [0.72,0.79]	1.1 [1.04,1.17]	0.81 [0.78,0.88]	1.1 [1.02,1.24]	0.8 [0.79,0.83]	0.87 [0.83,0.95]
2002	1.98 [1.59,2.2]	1.16 [1.07,1.26]	2.27 [1.82,2.42]	0.65 [0.59,0.75]	1.09 [1.04,1.18]	0.74 [0.65,0.86]	1.55 [1.32,1.86]	0.85 [0.81,0.89]	1.18 [1.05,1.35]
2003	1.38 [1.21,1.53]	1.15 [1.06,1.23]	1.52 [1.34,1.65]	0.89 [0.85,0.95]	1.1 [1.05,1.18]	0.98 [0.93,1.07]	1.04 [0.99,1.15]	0.88 [0.85,0.93]	0.98 [0.93,1.09]
2004	1.36 [1.22,1.44]	1.21 [1.1,1.3]	1.53 [1.32,1.58]	0.65 [0.59,0.76]	1.12 [1.05,1.25]	0.77 [0.68,0.92]	1.42 [1.24,1.72]	0.81 [0.76,0.86]	1.08 [0.97,1.23]
2005	1.36 [1.23,1.49]	1.24 [1.1,1.42]	1.79 [1.49,2.05]	0.82 [0.8,0.85]	1.12 [1.04,1.26]	0.95 [0.88,1.1]	1.31 [1.18,1.53]	0.81 [0.8,0.84]	0.95 [0.9,1.08]
2006	1.26 [1.16,1.36]	1.12 [1.05,1.19]	1.42 [1.27,1.56]	0.87 [0.86,0.9]	1.05 [1.02,1.1]	0.92 [0.9,0.98]	1.12 [1.05,1.23]	0.88 [0.86,0.91]	0.95 [0.9,1.02]
2007	1.26 [1.18,1.33]	1.28 [1.12,1.42]	1.56 [1.35,1.71]	0.92 [0.89,0.95]	1.15 [1.07,1.25]	1.05 [0.99,1.14]	1.1 [1.06,1.17]	0.62 [0.59,0.65]	0.69 [0.66,0.74]
2008	1.04 [0.98,1.11]	1.21 [1.08,1.34]	1.21 [1.1,1.32]	0.93 [0.93,0.95]	1.1 [1.04,1.2]	1.01 [0.97,1.09]	0.98 [0.96,1]	0.92 [0.87,0.98]	0.93 [0.9,0.98]
2009	1.38 [1.22,1.6]	1.14 [1.06,1.22]	1.51 [1.33,1.72]	0.74 [0.73,0.78]	1.06 [1.02,1.12]	0.8 [0.77,0.87]	1.05 [1.02,1.17]	0.95 [0.91,1.01]	0.99 [0.95,1.1]
2010	1.12 [1.08,1.16]	1.21 [1.1,1.33]	1.28 [1.17,1.37]	0.95 [0.94,0.97]	1.06 [1.02,1.13]	1.02 [0.98,1.08]	1.01 [1,1.04]	0.83 [0.8,0.89]	0.86 [0.84,0.9]
2011	1.35 [1.2,1.53]	1.22 [1.1,1.34]	1.72 [1.49,2.12]	0.71 [0.66,0.78]	1.09 [1.03,1.17]	0.78 [0.71,0.89]	1.35 [1.21,1.64]	0.86 [0.83,0.91]	1.17 [1.04,1.38]
2012	1.3 [1.17,1.49]	1.25 [1.1,1.41]	1.53 [1.36,1.73]	0.71 [0.68,0.82]	1.11 [1.04,1.21]	0.79 [0.74,0.9]	1.08 [1.01,1.22]	0.79 [0.77,0.83]	0.92 [0.85,1.07]
Total	1.35 [1.22,1.47]	1.2 [1.08,1.3]	1.58 [1.4,1.73]	0.77 [0.75,0.82]	1.1 [1.04,1.19]	0.86 [0.82,0.95]	1.16 [1.09,1.27]	0.82 [0.79,0.86]	0.95 [0.89,1.03]

Notes: The table shows the respective gains from eliminating material distortions ($\tilde{\tau}_{jt}^M$), labour distortions ($\tilde{\tau}_{jt}^L$) or both. Outcome variables are the ratio of the counterfactual to the factual. Bootstrapped 95% confidence intervals in brackets (see Section 2.4.5).

Table 16: Revenue and revenue productivity ratios

	Revenue ratio			Material productivity			Labour productivity		
	$\tilde{\tau}_{jt}^M$	$\tilde{\tau}_{jt}^L$	both	$\tilde{\tau}_{jt}^M$	$\tilde{\tau}_{jt}^L$	both	$\tilde{\tau}_{jt}^M$	$\tilde{\tau}_{jt}^L$	both
2000	1.28 [1.19,1.33]	1.13 [1.04,1.16]	1.39 [1.23,1.45]	0.9 [0.89,0.92]	1.06 [1.03,1.12]	0.95 [0.93,0.99]	0.91 [0.89,0.94]	0.49 [0.47,0.51]	0.51 [0.49,0.55]
2001	1.69 [1.49,1.83]	1.11 [1.05,1.16]	1.81 [1.56,1.92]	0.78 [0.78,0.8]	1.05 [1.03,1.1]	0.78 [0.78,0.81]	1.15 [1.1,1.23]	0.77 [0.76,0.78]	0.84 [0.83,0.87]
2002	1.7 [1.46,1.75]	1.11 [1.05,1.16]	1.73 [1.46,1.76]	0.56 [0.54,0.59]	1.05 [1.02,1.1]	0.56 [0.54,0.63]	1.32 [1.24,1.47]	0.82 [0.79,0.84]	0.9 [0.87,0.99]
2003	1.34 [1.21,1.4]	1.1 [1.04,1.14]	1.4 [1.26,1.45]	0.86 [0.81,0.88]	1.06 [1.03,1.1]	0.9 [0.87,0.94]	1.01 [0.98,1.07]	0.85 [0.83,0.87]	0.9 [0.88,0.95]
2004	1.02 [0.93,1.04]	1.06 [1.01,1.08]	1.03 [0.91,1.06]	0.49 [0.42,0.6]	0.99 [0.96,1.04]	0.51 [0.44,0.64]	1.07 [1.02,1.17]	0.71 [0.7,0.72]	0.73 [0.7,0.78]
2005	1.32 [1.21,1.4]	1.2 [1.09,1.32]	1.55 [1.34,1.63]	0.79 [0.77,0.81]	1.08 [1.04,1.18]	0.83 [0.8,0.88]	1.27 [1.15,1.43]	0.78 [0.77,0.8]	0.83 [0.81,0.87]
2006	1.24 [1.16,1.3]	1.07 [1.03,1.1]	1.31 [1.2,1.36]	0.86 [0.84,0.88]	1 [1,1.02]	0.85 [0.84,0.88]	1.11 [1.06,1.18]	0.84 [0.83,0.85]	0.88 [0.86,0.91]
2007	1.15 [1.08,1.17]	1.15 [1.06,1.21]	1.28 [1.13,1.32]	0.84 [0.8,0.86]	1.03 [1.01,1.06]	0.86 [0.82,0.89]	1 [0.99,1.03]	0.55 [0.54,0.57]	0.56 [0.55,0.59]
2008	1.07 [1.03,1.11]	1.12 [1.05,1.18]	1.17 [1.1,1.23]	0.96 [0.94,0.99]	1.02 [1.01,1.05]	0.98 [0.94,1.03]	1.01 [0.99,1.03]	0.85 [0.84,0.87]	0.9 [0.89,0.93]
2009	1.41 [1.26,1.55]	1.07 [1.03,1.09]	1.44 [1.29,1.57]	0.76 [0.73,0.8]	0.99 [0.98,1.01]	0.77 [0.73,0.81]	1.08 [1.05,1.13]	0.89 [0.87,0.91]	0.95 [0.91,1.03]
2010	1.05 [1.02,1.06]	1.09 [1.03,1.12]	1.12 [1.04,1.14]	0.9 [0.87,0.92]	0.95 [0.93,0.98]	0.89 [0.86,0.92]	0.95 [0.93,0.97]	0.75 [0.73,0.77]	0.75 [0.73,0.78]
2011	1.34 [1.23,1.44]	1.15 [1.07,1.23]	1.52 [1.36,1.66]	0.7 [0.63,0.75]	1.03 [1.01,1.07]	0.69 [0.61,0.78]	1.34 [1.22,1.54]	0.81 [0.79,0.84]	1.03 [0.94,1.13]
2012	1.13 [1.07,1.17]	1.14 [1.06,1.21]	1.24 [1.13,1.3]	0.62 [0.58,0.69]	1.02 [1.01,1.05]	0.64 [0.58,0.74]	0.94 [0.88,1.04]	0.73 [0.69,0.76]	0.75 [0.71,0.8]
Total	1.27 [1.17,1.32]	1.11 [1.05,1.16]	1.36 [1.22,1.42]	0.72 [0.7,0.76]	1.02 [1.01,1.06]	0.74 [0.71,0.79]	1.09 [1.05,1.14]	0.77 [0.76,0.77]	0.81 [0.8,0.85]

Notes: The table shows the respective gains from eliminating material distortions ($\tilde{\tau}_{jt}^M$), labour distortions ($\tilde{\tau}_{jt}^L$) or both. Outcome variables are the ratio of the counterfactual to the factual. Bootstrapped 95% confidence intervals in brackets (see Section 2.4.5).

G.10 Measurement of labour input

Instead of using man-days as labour input L_{jt} we could also use the wage bill as labour input, so $L_{jt}^{alt} = L_{jt} * w_{jt}$. If higher skill correlates with higher salaries, this alternative measurement of labour input accounts for difference in quality of labour across plants. Rerunning the entire analysis with L_{jt}^{alt} yields the results reported in Table 17, 18 and 19. I omit confidence intervals for simplicity, but almost all point estimates are within the range of the confidence intervals of the baseline version.

Table 17: Welfare gains in billion rupees using L_{jt}^{alt}

	Comp. Variation			Profits			Total welfare		
	$\tilde{\tau}_{jt}^M$	$\tilde{\tau}_{jt}^L$	<i>both</i>	$\tilde{\tau}_{jt}^M$	$\tilde{\tau}_{jt}^L$	<i>both</i>	$\tilde{\tau}_{jt}^M$	$\tilde{\tau}_{jt}^L$	<i>both</i>
2000	0.15	0.08	0.22	0.12	0.04	0.16	0.28	0.12	0.38
2001	1.79	0.20	2.09	1.32	0.12	1.51	3.11	0.31	3.60
2002	3.53	0.25	3.94	2.49	0.17	2.74	6.01	0.42	6.68
2003	2.65	0.47	3.04	1.36	0.23	1.52	4.01	0.70	4.56
2004	2.95	0.93	3.63	0.91	0.28	1.08	3.86	1.21	4.71
2005	2.39	0.62	3.29	1.56	0.39	2.16	3.96	1.01	5.45
2006	2.55	0.47	3.09	1.56	0.24	1.87	4.11	0.71	4.96
2007	1.57	0.66	2.06	1.01	0.38	1.34	2.58	1.04	3.39
2008	0.25	0.25	0.70	0.28	0.18	0.70	0.53	0.43	1.40
2009	2.83	0.42	3.18	2.00	0.23	2.19	4.83	0.65	5.37
2010	0.61	0.41	0.87	0.25	0.17	0.37	0.86	0.58	1.24
2011	2.44	0.63	3.48	2.20	0.42	3.21	4.64	1.05	6.69
2012	1.99	0.58	2.38	0.93	0.33	1.24	2.93	0.91	3.62
Total	25.70	5.96	31.97	16.00	3.18	20.09	41.70	9.14	52.07
Per year	1.98	0.46	2.46	1.23	0.24	1.55	3.21	0.70	4.01

Notes: The table shows the respective welfare gains from eliminating material distortions ($\tilde{\tau}_{jt}^M$), labour distortions ($\tilde{\tau}_{jt}^L$) or both in billion rupees. The last three columns sum the consumer side compensating variation and the profits for total welfare gains. The last two rows report the total across all years and the implied average per year.

Table 18: Compensating variation as share of consumer expenditure and profit growth using L_{jt}^{alt}

	Comp. Variation			Profit growth		
	$\tilde{\tau}_{jt}^M$	$\tilde{\tau}_{jt}^L$	<i>both</i>	$\tilde{\tau}_{jt}^M$	$\tilde{\tau}_{jt}^L$	<i>both</i>
2000	0.13	0.06	0.19	1.25	1.09	1.34
2001	0.27	0.03	0.32	1.60	1.05	1.69
2002	0.34	0.02	0.38	1.77	1.05	1.85
2003	0.17	0.03	0.19	1.31	1.05	1.34
2004	0.17	0.05	0.21	1.18	1.05	1.21
2005	0.16	0.04	0.22	1.34	1.08	1.47
2006	0.13	0.02	0.15	1.25	1.04	1.31
2007	0.11	0.05	0.14	1.21	1.08	1.27
2008	0.03	0.03	0.08	1.10	1.06	1.25
2009	0.18	0.03	0.21	1.44	1.05	1.48
2010	0.06	0.04	0.08	1.08	1.06	1.12
2011	0.13	0.03	0.19	1.43	1.08	1.62
2012	0.13	0.04	0.15	1.20	1.07	1.26
Total	0.16	0.04	0.19	1.31	1.06	1.39

Notes: The table shows the respective welfare gains from eliminating material distortions ($\tilde{\tau}_{jt}^M$), labour distortions ($\tilde{\tau}_{jt}^L$) or both. The first three columns express the compensating variation per unit purchased as a share of the unit price (i.e. as share of expenditure on the products in the sample). The profit ratio is total profits in the counterfactual divided by total profits in the factual equilibrium.

Table 19: Physical output and productivity ratios using L_{jt}^{alt}

	Output ratio			Material productivity			Labour productivity		
	$\tilde{\tau}_{jt}^M$	$\tilde{\tau}_{jt}^L$	<i>both</i>	$\tilde{\tau}_{jt}^M$	$\tilde{\tau}_{jt}^L$	<i>both</i>	$\tilde{\tau}_{jt}^M$	$\tilde{\tau}_{jt}^L$	<i>both</i>
2000	1.29	1.15	1.43	0.90	1.09	0.96	0.79	0.78	0.74
2001	1.74	1.08	1.88	0.78	1.04	0.81	1.05	0.85	0.83
2002	2.12	1.07	2.27	0.67	1.03	0.71	1.45	0.86	1.13
2003	1.40	1.07	1.46	0.89	1.04	0.93	0.93	0.84	0.86
2004	1.40	1.11	1.49	0.61	1.04	0.66	1.63	0.88	1.32
2005	1.45	1.10	1.64	0.84	1.03	0.89	1.20	0.80	0.88
2006	1.31	1.05	1.39	0.90	1.01	0.92	1.04	0.89	0.91
2007	1.31	1.11	1.42	0.93	1.04	0.96	1.08	0.83	0.93
2008	1.11	1.09	1.31	0.94	1.03	0.94	0.96	0.88	0.90
2009	1.51	1.07	1.59	0.76	1.02	0.78	0.97	0.89	0.88
2010	1.14	1.10	1.22	0.96	1.01	0.97	1.01	0.92	0.98
2011	1.50	1.12	1.74	0.71	1.04	0.74	1.23	0.82	0.98
2012	1.41	1.11	1.53	0.71	1.04	0.74	1.12	0.86	1.07
Total	1.43	1.09	1.55	0.77	1.03	0.81	1.11	0.86	0.95

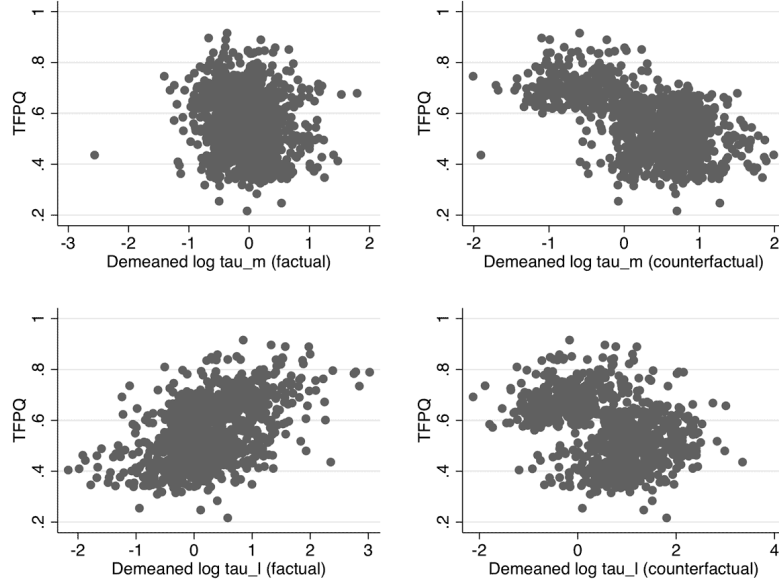
Notes: The table shows the respective gains from eliminating material distortions ($\tilde{\tau}_{jt}^M$), labour distortions ($\tilde{\tau}_{jt}^L$) or both. Outcome variables are the ratio of the counterfactual to the factual.

G.11 Wedges as tax income

When the τ are interpreted as taxes, removing differences in τ also affects government revenues. In Table 20, I add the tax revenue changes to the welfare calculations. The difference to the baseline version is statistically significant when using bootstrapped differences.

G.12 A counterfactual of more negatively correlated TFPQ and τ

Figure 29: Correlation between TFPQ and τ



Notes: Plotted are the plant TFPQ (Ω_{jt}) and the annually demeaned $\ln(\tau_{jt}^M)$ or $\ln(\tau_{jt}^L)$. The left panels correspond to the factual equilibrium. The right panel corresponds to an alternative counterfactual, where the τ are reduced for all above average productivity plants and increased for all below average productivity plants until the correlation is -0.5.

Figure 29 shows the correlations between plant level TFPQ (Ω) and τ . In the factual equilibrium (the two left panels), there is no significant correlation between Ω and τ^M and a slight positive correlation with τ^L . I construct an alternative counterfactual where the τ are set such that their weighted geometric average is preserved, but the τ are reduced for all above average productivity plants and increased for all below average productivity plants. This does not remove misallocation. The two right panels depict this counterfactual and the resulting correlation between Ω and τ . There are substantial welfare gains from moving from the left to the right panels. The size of the welfare gains depends on how strongly the (artificial) correlation differs in the counterfactual compared to the factual. In this case (correlation to both around -0.5) the welfare gains are roughly half of the welfare gains of the baseline results where the τ are removed instead.

Table 20: Welfare gains in billion rupees with tax income adjustments

	Total welfare			Taxes on materials			Taxes on labour			Total welfare with taxes		
	$\bar{\tau}_{jt}^M$	$\bar{\tau}_{jt}^L$	both	$\bar{\tau}_{jt}^M$	$\bar{\tau}_{jt}^L$	both	$\bar{\tau}_{jt}^M$	$\bar{\tau}_{jt}^L$	both	$\bar{\tau}_{jt}^M$	$\bar{\tau}_{jt}^L$	both
2000	0.25 [0.19,0.33]	0.21 [0.1,0.3]	0.45 [0.37,0.53]	-0.08 [-0.09,-0.05]	0.01 [0.01,0.03]	-0.08 [-0.1,-0.05]	0.05 [0.03,0.07]	0.02 [0,0.03]	0.07 [0.03,0.09]	0.23 [0.15,0.32]	0.24 [0.12,0.35]	0.44 [0.33,0.54]
2001	2.57 [1.94,3.37]	0.6 [0.29,0.91]	3.43 [2.78,4.15]	0.05 [-0.38,0.5]	0.19 [0.1,0.29]	0.13 [-0.35,0.55]	0.85 [0.45,1.18]	0.12 [0.03,0.24]	0.97 [0.47,1.33]	3.48 [2.29,4.68]	0.91 [0.42,1.42]	4.53 [3.27,5.76]
2002	5.31 [3.92,6.51]	0.88 [0.43,1.4]	6.56 [5.33,7.45]	0.27 [-0.4,0.86]	0.48 [0.28,0.75]	0.26 [-0.38,0.83]	1.26 [0.61,1.65]	0.2 [0.05,0.4]	1.24 [0.54,1.58]	6.85 [4.68,8.14]	1.56 [0.77,2.51]	8.05 [6.14,8.99]
2003	3.59 [2.5,4.9]	1.28 [0.64,1.92]	4.7 [3.83,5.77]	-0.14 [-0.66,0.34]	0.21 [0.06,0.35]	-0.06 [-0.63,0.42]	1.01 [0.43,1.4]	0.28 [0.07,0.54]	1.15 [0.53,1.62]	4.46 [2.73,6.15]	1.76 [0.84,2.75]	5.8 [4.3,7.46]
2004	3.35 [2.53,4.1]	2.15 [1.1,2.97]	4.9 [4.06,5.34]	-0.43 [-0.69,-0.35]	0.03 [-0.13,0.15]	-0.43 [-0.69,-0.02]	-0.11 [-0.47,0]	0.13 [-0.04,0.19]	-0.18 [-0.55,0]	2.81 [1.9,3.35]	2.31 [1.22,3.09]	4.29 [3.22,4.75]
2005	3.23 [2.42,4.24]	2.2 [0.99,3.73]	6.57 [5.8,59]	-0.37 [-0.76,0.09]	0.72 [0.33,1.31]	-0.22 [-0.9,0.31]	0.9 [0.49,1.25]	0.55 [0.13,1.12]	1.42 [0.62,1.91]	3.77 [2.41,5.1]	3.47 [1.52,6.02]	7.77 [5.43,10.17]
2006	3.23 [2.43,4.32]	1.42 [0.69,2.13]	4.97 [3.97,6.4]	-0.1 [-0.56,0.33]	0.1 [-0.01,0.21]	0.02 [-0.53,0.45]	0.88 [0.46,1.25]	0.22 [0.05,0.42]	1.09 [0.49,1.53]	4.01 [2.66,5.56]	1.73 [0.79,2.69]	6.08 [4.47,7.97]
2007	2.1 [1.68,2.64]	2.28 [1.14,3.24]	4.2 [3.4,5.02]	-0.44 [-0.81,-0.3]	-0.11 [-0.3,0.01]	-0.43 [-0.82,-0.24]	0.3 [0.11,0.41]	0.3 [0.06,0.54]	0.54 [0.13,0.8]	1.96 [1.41,2.53]	2.47 [1.23,3.54]	4.31 [3.21,5.32]
2008	0.21 [-0.06,0.57]	0.97 [0.44,1.51]	1.03 [0.6,1.49]	0.13 [0,0.27]	0 [-0.11,0.08]	0.19 [-0.01,0.39]	0.13 [0.06,0.2]	0.18 [0.04,0.34]	0.29 [0.11,0.48]	0.47 [0.17,0.89]	1.16 [0.51,1.83]	1.51 [1.01,2.13]
2009	3.83 [2.68,5.62]	1.22 [0.56,1.89]	4.76 [3.76,6.22]	0.21 [-0.43,0.9]	0.1 [0,0.17]	0.25 [-0.41,0.89]	1.22 [0.63,1.8]	0.15 [0.03,0.27]	1.28 [0.63,1.89]	5.26 [3.18,7.93]	1.47 [0.67,2.29]	6.3 [4.34,8.79]
2010	0.7 [0.54,0.91]	1.22 [0.59,1.84]	1.55 [1.07,1.98]	-0.06 [-0.11,0.01]	-0.01 [-0.14,0.06]	0.03 [-0.09,0.19]	0.08 [0.01,0.1]	0.14 [0.02,0.24]	0.19 [0.03,0.27]	0.72 [0.5,0.93]	1.35 [0.68,1.99]	1.76 [1.2,2.27]
2011	3.35 [2.18,5.08]	1.95 [0.96,2.98]	6.5 [5.06,10.16]	0.68 [-0.24,1.58]	0.91 [0.46,1.46]	1.39 [-0.07,2.53]	1.16 [0.62,1.66]	0.51 [0.13,1.01]	1.61 [0.73,2.28]	5.18 [2.99,7.52]	3.37 [1.59,5.31]	9.51 [6.99,12.81]
2012	2.17 [1.41,3.29]	1.88 [0.88,2.96]	3.66 [2.99,4.74]	-0.03 [-0.22,0.18]	0.14 [-0.07,0.31]	0.17 [-0.16,0.35]	0.35 [0.12,0.49]	0.4 [0.1,0.77]	0.63 [0.2,1.03]	2.49 [1.49,3.77]	2.42 [1.1,3.87]	4.47 [3.27,5.81]
Total	33.9 [24.87,43.82]	18.26 [8.74,27.27]	53.28 [44.27,64.09]	-0.3 [-4.74,3.59]	2.78 [1.1,4.64]	1.2 [-4.28,5.8]	8.09 [3.89,11.07]	3.19 [0.72,5.84]	10.31 [4.36,14.29]	41.68 [26.9,55.02]	24.23 [11.37,37.39]	64.8 [48.41,81.18]
Per year	2.61 [1.91,3.37]	1.4 [0.67,2.1]	4.1 [3.41,4.93]	-0.02 [-0.36,0.28]	0.21 [0.08,0.36]	0.09 [-0.33,0.45]	0.62 [0.3,0.85]	0.25 [0.06,0.45]	0.79 [0.34,1.1]	3.21 [2.07,4.23]	1.86 [0.87,2.88]	4.98 [3.72,6.24]

Notes: The table shows the respective welfare gains from eliminating material distortions ($\bar{\tau}_{jt}^M$), labour distortions ($\bar{\tau}_{jt}^L$) or both in billion rupees. Bootstrapped 95% confidence intervals in brackets (see Section 2.4.5).

Table 21: Determinants of plant level changes in input productivities

<i>Dependent variable:</i>	Δ plant mat. prod. (log)			Δ plant lab. prod. (log)		
<i>Counterfactual:</i>	$\tilde{\tau}_{jt}^M$	$\tilde{\tau}_{jt}^L$	<i>both</i>	$\tilde{\tau}_{jt}^M$	$\tilde{\tau}_{jt}^L$	<i>both</i>
<i>Dep. var. (exp) 10th - 90th percentiles:</i>	[.8, 1.4]	[.9, 1.2]	[.8, 1.5]	[.8, 1.4]	[.4, 1.7]	[.4, 1.8]
τ_{jt}^M (log demeaned)	-1.03*** (0.00)	-0.01** (0.04)	-0.89*** (0.00)	0.96*** (0.00)	-0.00** (0.04)	0.36*** (0.00)
τ_{jt}^L (log demeaned)	0.00 (0.33)	1.00*** (0.00)	0.48*** (0.00)	0.00 (0.32)	-1.00*** (0.00)	-0.91*** (0.00)
TFPQ (log)	-0.01** (0.04)	-0.00 (0.33)	-0.01** (0.02)	-0.01** (0.04)	-0.00 (0.33)	-0.00** (0.02)
Markup	-0.07*** (0.00)	-0.01** (0.01)	-0.05*** (0.00)	-0.06*** (0.00)	-0.00** (0.01)	-0.02*** (0.00)
N	979	979	979	979	979	979
R^2	0.99	0.99	0.99	0.99	1.00	1.00

Notes: Coefficients are standardized. p-values in parentheses are based on clustered standard errors at the plant level. Dependent variables are the log of the ratio of the input productivities in the counterfactual to the factual (i.e. the difference Δ in logs). Only observations with an initial demand elasticity < -1 are included.

G.13 Heterogeneity in plant input productivity changes

Table 21 reports the 10th and 90th percentile of the input productivity ratios (between counterfactual and factual) in the table header. There is substantial heterogeneity in the comparative statics across plants. For the third column, for example, where both distortions are removed, the 10th percentile is a decrease in material productivity of 20% while the 90th percentile is an increase of 50%.

I run regressions where the dependent variable is the change in the log input productivities between the counterfactual and the factual equilibrium at the level of plants. I regress this on the two τ , plant productivity TFPQ (Ω) and initial markups. The table reports standardized coefficients. The R^2 is extremely high, and the variation in input productivity growth across plants can be well explained by the initial input distortions.

Removing distortions in one input has large effects on the changes in plant level input productivity of the same input. This is intuitive. Plants with a high τ^M face costs using materials, and when these costs are reduced in the counterfactual, the plant has incentives to use the previously constrained materials relatively more intensively, thereby decreasing the ratio of output to materials. This also provides intuition of why aggregate input productivities do not increase when distortions in both markets are removed. The plants that grow (i.e. previously constrained through high τ) also use the input relatively more intensively and addressing both distortions compensates each distortion's effects on the input productivities.

Table 22: Change in markup variation

	Factual	Counterfactual $\tilde{\tau}_{jt}^M$	Counterfactual $\tilde{\tau}_{jt}^L$	Counterfactual $\tilde{\tau}_{jt}^M$ and $\tilde{\tau}_{jt}^L$
Median	1.26	1.25	1.27	1.26
95th/5th	1.30	1.20	1.32	1.23
90th/10th	1.21	1.14	1.22	1.16
75th/25th	1.09	1.07	1.10	1.07

Notes: The table shows the median and ratios of percentiles of the markups in different equilibria, pooled across years.

G.14 Changes in markup dispersion

Table 22 reports the median and the ratios of markups of the the 95th to th 5th percentile, the 90th to the 10th percentile and the 75th to the 25th percentile across all years. The first column reports the ratios for the factual equilibrium and the other columns for the corresponding counterfactual equilibria. The median markup is similar, but the variation in markups is lower in the counterfactuals except for the one where only labour misallocation is removed ($\tilde{\tau}_{jt}^L$).

G.15 Modifying and discussing Hsieh and Klenow (2009)

The way Hsieh and Klenow (2009) calculate losses from misallocation is by calculating the output gap $\frac{Y}{Y^*}$:

$$\frac{Y}{Y^*} = \prod_{s=1}^S \frac{(TFP_s K_s^{\alpha_s} X_s^{\beta_s} L_s^{1-\alpha_s-\beta_s})^{\theta_s}}{(TFP_s^* K_s^{\alpha_s} X_s^{\beta_s} L_s^{1-\alpha_s-\beta_s})^{\theta_s}} = \prod_{s=1}^S \frac{TFP_s^{\theta_s}}{TFP_s^{*\theta_s}} = \prod_{s=1}^S \left[\sum_i \left(\frac{A_{si}}{\overline{TFP}_s^*} \frac{\overline{TFPR}_s}{TFPR_{si}} \right)^{\sigma-1} \right]^{\frac{\theta_s}{\sigma-1}} \quad (35)$$

where I follow the notation in their paper, and adding X_{si} as material input for firm i in sector s with corresponding output elasticity β_s . The other components are:

$$\frac{A_{si}}{\overline{TFP}_s^*} = \frac{(P_{si} Y_{si})^{\frac{\sigma}{\sigma-1}} / K_{si}^{\alpha_s} (v X_{si})^{\beta_s} (w L_{si})^{1-\alpha_s-\beta_s}}{\left[\sum_i [(P_{si} Y_{si})^{\frac{\sigma}{\sigma-1}} / K_{si}^{\alpha_s} (v X_{si})^{\beta_s} (w L_{si})^{1-\alpha_s-\beta_s}]^{\sigma-1} \right]^{\frac{1}{\sigma-1}}} \quad (36)$$

$$\frac{TFPR_{si}}{\overline{TFPR}_s} = \frac{P_{si} Y_{si}}{P_s Y_s} \frac{(\sum_i K_{si})^{\alpha_s} (\sum_i v X_{si})^{\beta_s} (\sum_i w L_{si})^{1-\alpha_s-\beta_s}}{K_{si}^{\alpha_s} (v X_{si})^{\beta_s} (w L_{si})^{1-\alpha_s-\beta_s}} \quad (37)$$

As in their paper, I calibrate the demand elasticity σ to 3, $\theta_s = \frac{P_s Y_s}{P Y}$ because of the perfect competition assumption in the final good sector, and I take the output elasticities from the CES-NBER database for each sector.¹⁴⁶

¹⁴⁶Trimming of outliers is analogously done to their paper by pooling sectors in a year and trimming top

Table 23: Output gains from replication and extension of Hsieh and Klenow (2009)

	2 factor value added model			3 factor gross output model		
	CRS	RS=0.95	RS=0.92	CRS	RS=0.95	RS=0.92
Manufacturing	95 [93,98]	37 [36,40]	12 [10,15]	41 [39,42]	2 [1,3]	-16 [-16,-15]
Basic metals	146 [127,175]	71 [64,87]	37 [30,53]	54 [51,59]	10 [9,12]	-9 [-11,-7]
Basic metals with estimated elasticities	-	-	-	32 [29,34]	-3 [-5,-1]	-19 [-20,-17]

Notes: Calculated is the average gain from the four periods 2000-2003, in percent. The square brackets contain the minimum and maximum of the four years. The columns have different returns to scale assumptions, and the elasticities are scaled to fit them accordingly. The elasticities corresponding to the first two rows are from the NBER-CES database. The last row uses the estimated elasticities from this paper, and they are scaled to fit the returns to scale accordingly. The last column of this row just uses the estimated returns to scale (0.92).

The first point to note is that aggregate sectoral output Y_s can be written in two ways in their paper:

$$Y_s = TFP_s^* K_s^{\alpha_s} X_s^{\beta_s} L_s^{1-\alpha_s-\beta_s} \quad (38)$$

$$Y_s = \left(\sum_{i=1}^{M_s} Y_{si}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \quad (39)$$

where the first equation is the aggregate production function, and the second is the utility function of the representative consumer.¹⁴⁷ The output gap in Equation (35) can not only be interpreted as pure production side TFP gap, but equally as gap in utility. The latter interpretation is in line with the results of this paper.

Second, I replicate their analysis and adjust two assumptions. First I either use their 2 factor value added production model,¹⁴⁸ or add materials as a third factor for gross output production functions. The second adjustment are the assumed returns to scale, where I use their constant returns to scale, 0.95, or my estimated 0.92 returns to scale. The results are reported in Table 23. I replicate the analysis for the entire manufacturing sector (first row), or for the basic metals sector only (second and third row). In addition, I use the estimated input elasticities instead of the CES-NBER elasticities in the third row.¹⁴⁹

In their paper they report output gains between 100 and 128 percent for the years 1987

and bottom 1% of both ratios, $\frac{A_{si}}{TFP_s^*}$ and $\frac{TFPR_{si}}{TFPR_s}$, and then recalculating all sector level variables.

¹⁴⁷The demand function is derived through a cost minimisation of a representative purchaser in industry s , who minimises the expenditure of all varieties in industry s ($\sum_i P_{si} Y_{si}$) subject to $Y_s \geq \bar{Y}_s$. That is subject to some minimum level of CES output or “utility” where $Y_s = \left(\sum_{i=1}^{M_s} Y_{si}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$, which describe the preferences over varieties with an elasticity of substitution σ .

¹⁴⁸That is dropping X_{si} from the analysis and using value added instead of revenues for $P_{si} Y_{si}$.

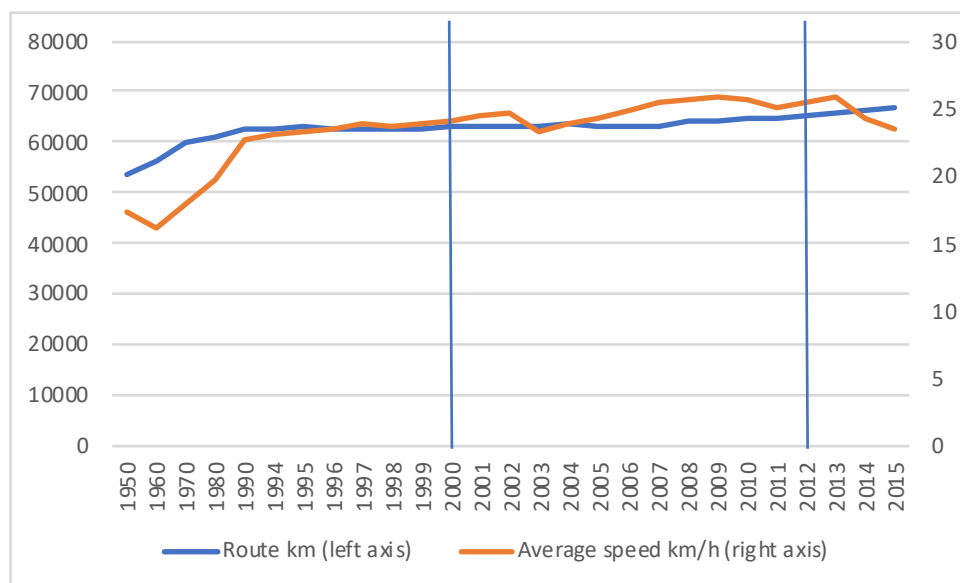
¹⁴⁹When I use my sample of cast iron plants only and apply my estimated elasticities to their model, I obtain a positive correlation between the implied logged TFPQ from their model, and my estimated logged TFPQ. The R^2 is 0.27, so there is a substantial difference between the productivities.

and 1994 respectively. I use the years 2000 to 2003 and calculate gains of 95 percent, close to their estimates from earlier years. When I reduce the assumed returns to scale to 0.95 or 0.92, the hypothetical gains fall dramatically to only a tenth of the gains under CRS.¹⁵⁰ Moving to a gross output model shrinks the gains as well.¹⁵¹ With 0.92 returns to scale, some of the gains are actually negative – India would be better off with misallocation than without. Part of the problem why these strange results arise in their model lies with the definition of the counterfactual $\tilde{\tau}$. In this paper, I use a geometric average instead of unity, which addresses measurement error in the distortions that is constant across plants, for example in the output elasticities (that are scaled by the returns to scale), as discussed in Section 2.4.1.

H Additional descriptives for the construction of supplier access

H.1 Railway route kilometres and speed

Figure 30: Total route kilometres of Indian railways and average speed of goods trains



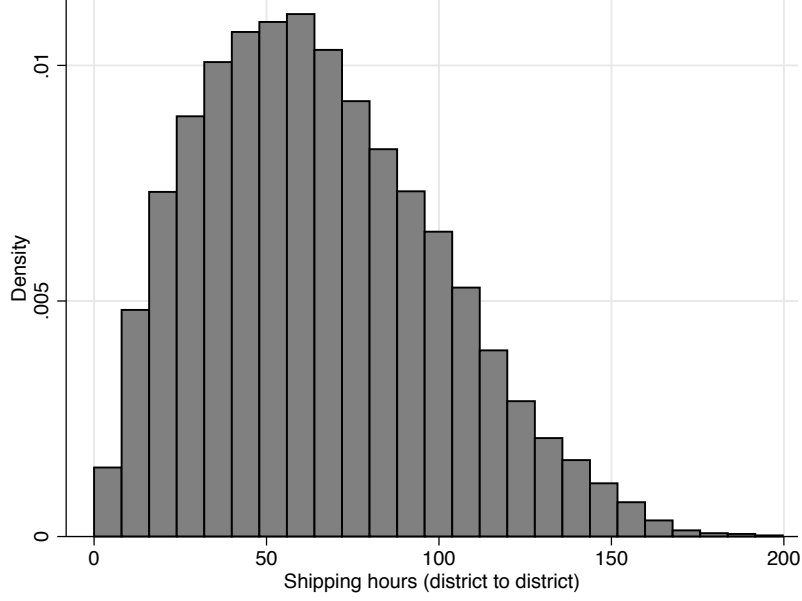
Notes: Vertical lines indicate sample period. Source: Calculated based on information from Ministry of Railways India, retrieved through indiastat.com

¹⁵⁰For a theoretical analysis of the sensitivity of these types of models to the returns to scale assumption see [Hopenhayn \(2014b\)](#).

¹⁵¹[Gandhi et al. \(2017\)](#) explore the theoretical and empirical relationships between value added and gross output production functions. See also [Dias et al. \(2016\)](#) which find larger gains in a [Hsieh and Klenow \(2009\)](#) 2 factor value added type model (their Table A1) than in a gross output 3 factor model (their Table 4).

H.2 Estimates of the fastest path FP_{dh}

Figure 31: Histogram of bilateral fastest paths FP_{dh}



Notes: The figure plots the histogram for bilateral estimated shipping times FP_{dh} between districts using the road and railroad transport network. Shipping times are trimmed at 200 hours. Shipping times are estimated using [Dijkstra's 1959](#) algorithm, with speed assumptions as edge weights as shown in Table 2.

H.3 Correlation between input price and supplier access

Table 24: Input prices and supplier access

	Input price (log)			
	(1)	(2)	(3)	(4)
Supplier access	-0.22 [-1.32]	-0.20 [-1.31]	-0.19 [-1.41]	-0.18 [-1.26]
Plant level controls	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes
District FE	Yes	Yes	No	No
Plant FE	No	No	Yes	Yes
N	946	946	946	946
R^2	0.32	0.33	0.64	0.65

Notes: The dependent variable is the logged material input price $\log P_{jt}^M$. Coefficients are standardised. t-statistics in brackets are based on standard errors clustered at the district level. Plant level controls include firm age, dummies on ownership type, and whether plants are part of the census section.

I Robustness checks for Section 6

I.1 Monopsony power

Appendix B shows that the measured input distortion τ_{jt}^M captures input market power $(\psi_{jt} + 1)$ as well as other input distortions $\tau^{M_{adj}}$, where $(\psi_{jt} + 1)$ is the ability to pay an input a lower price than its marginal revenue product, a common definition of market power on the input side, or monopsony power. If ψ_{jt} does not vary across plants, or does not vary across

time for individual plants, than it will be absorbed by year or plant fixed effects respectively (see Table 8). In Appendix B I recover a heuristic estimate of ψ_{jt} and plot the histogram of input market power ($\psi_{jt} + 1$).

I use the estimated $(\psi_{jt} + 1)$ as control variables in Column (3) and (4) in Table 25. In Columns (5) and (6) I instead use $(\psi_{jt} + 1)$ to recover the adjusted distortion $\tau^{M_{adj}}$ as dependent variable. I construct a second proxy for input market power. If larger plants can exert more market power on the input side as well, a larger market share of a plant in a given district can proxy for monopsony power. In Column (1) and (2) of Table 25 I control for plant market shares within a district.¹⁵² Overall, the results are robust to these additional test, and the input distortions are significantly higher for plants with worse supplier access, supporting the interpretation that this is driven by indirect trade costs.

Table 25: Monopsony power: additional proxy controls and adjusted input distortion

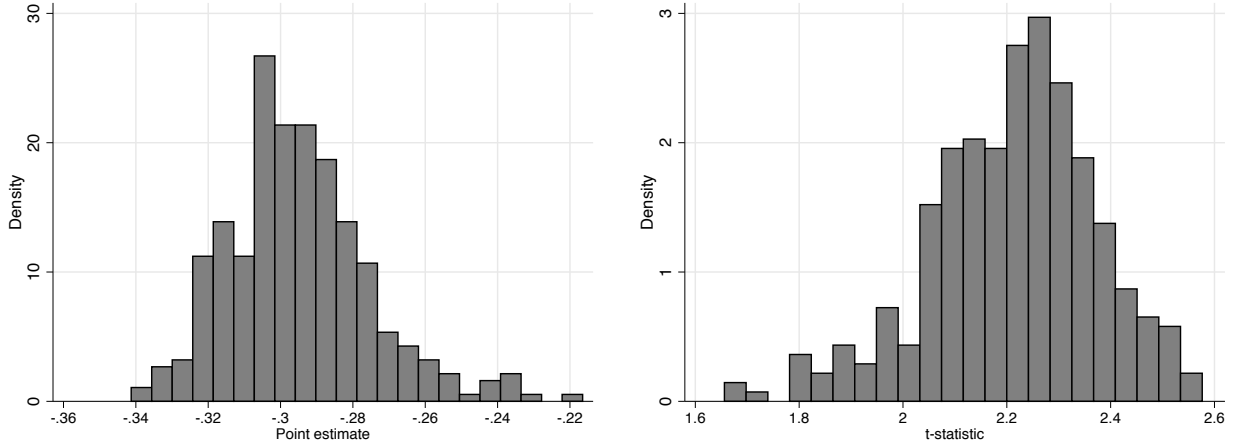
	Input distortion τ_{jt}^M (log)				Adjusted $\tau^{M_{adj}}$ (log)	
	(1)	(2)	(3)	(4)	(5)	(6)
Supplier access	-0.27*** [-4.76]	-0.28** [-2.11]	-0.27*** [-4.77]	-0.25** [-2.06]	-0.25*** [-4.45]	-0.19 [-1.57]
Market share in district	0.03 [0.48]	-0.04 [-0.69]				
Inv. inp price elasticity + 1 (log)			0.06 [0.61]	0.14 [1.08]		
Plant level controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
District FE	No	Yes	No	Yes	No	Yes
N	926	925	926	925	926	925
R^2	0.20	0.46	0.20	0.46	0.31	0.54

Notes: The dependent variable is the demeaned (by year) logged material input distortion τ_{jt} in the first four columns, and the adjusted logged $\tau^{M_{adj}}$ in the last two columns. The market share is calculated at the district year level in terms of value of cast iron sold. The input price elasticity is estimated from a regression of logged input prices on a second order polynomial in logged material input use, controlling for plant and year fixed effects. I take the log of the inverse plant level input price elasticities plus one $\log(\psi_{jt} + 1)$, i.e. the log of the markdown. Plant level controls include firm age, dummies on ownership type, and whether plants are part of the census section. Coefficients are standardised. t-statistics in brackets are based on standard errors clustered at the district level.

¹⁵²These controls may be bad controls in the sense of Angrist and Pischke (2008), but the point is to show that the coefficient of supplier access is reasonably robust to those.

I.2 Accounting for estimated input distortions

Figure 32: Distribution of point estimates and t-statistics



Notes: The left panel plots the histogram of 330 point estimates of supplier access of Equation (11). Each individual regression has a different τ_{jt}^M as dependent variable, based on 330 draws of the underlying production and demand parameters from their estimated covariance matrices. The right panel shows the histogram of t-statistics of the supplier access estimates. All estimates are significant at the 10% at least.

I.3 Market access and output shipping costs

Table 26 shows the result from a regression of the share of output shipping costs in revenues on market access (i.e. access to buyers). Column (2) shows that the access to suppliers is in turn not relevant for *output* shipping costs.

Table 26: Market access and share of output shipping costs in revenue

	(1)	(2)
Market access	-0.11*** [-3.50]	
Supplier access		0.00 [0.04]
Year FE	Yes	Yes
N	946	946
r2	0.04	0.02

Notes: Standardized coefficients. t-statistics in brackets based on SE clustered on districts.